

Digital Preservation: An Historical Perspective and Overview



PETER D. VERHEYEN
HEAD OF PRESERVATION AND CONSERVATION
26 MARCH 2009



What is digital preservation?

2

Digital preservation is the management of digital assets over time to ensure their continued accessibility.

Historical Perspective

3

- **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
 - Standards developed to ensure quality
- **Despite problems, all remain readable**

Peter D. Verheyen, Head of Preservation

21 July 2009

Digital Content is Fragile

4

- **Media and storage devices degrade and/or become obsolete quickly.**
 - Replacement costs...
 - Current appeal for old hardware/software at CoOL (Stanford)
- **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.’**

Peter D. Verheyen, Head of Preservation

21 July 2009

Digital Preservation Strategies

5

- **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

Peter D. Verheyen, Head of Preservation

21 July 2009

Planning for Digital Preservation

6

- **Develop a plan for managing each file format in a digital archive.**
 - May include specific strategies based upon data format risk assessments.
 - ✦ How likely support software/hardware is 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 .

Peter D. Verheyen, Head of Preservation

21 July 2009

Digital Preservation Plan

7

- Subject to regular review/revision.
- May be built into application to varying degrees.
- 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

Peter D. Verheyen, Head of Preservation

21 July 2009

Preservation Storage Requirements

8

- **Standard back-up practices are insufficient for long-term preservation. Preservation requirements include:**
 - **Standards:** OAIS, ensures accessibility and compatibility
 - **Audit/reporting.** Ability to run reports on data integrity for ingested digital objects.
 - **Automated back-up.** Regular back-ups to multiple physical locations at defined time intervals.
 - **Automated technical/preservation metadata capture.**
 - **Data integrity.** Any means of ensuring that data is whole, complete, or authentic. The term “fixity” is also used.
 - Continued...

Peter D. Verheyen, Head of Preservation

21 July 2009

Preservation Storage Requirements

9

- **Preservation requirements, cont:**

- **File versioning.** Saves “snapshot” of file at the point it is amended, thereby helping to ensure data integrity.
- **Migration tools.** Convert data from one type of storage material to another to ensure continued access to the information as the material becomes obsolete or degrades over time.
- **Object verification.** Checks a submission for schema errors, file format problems, and ingest parameter inconsistencies that might affect its suitability for preservation.
- Continued...

Peter D. Verheyen, Head of Preservation

21 July 2009

Preservation Storage Requirements

10

- **Preservation requirements, cont:**

- **Security.** Tools and behaviors to ensure write access only to authenticated users.
- **Self-healing.** In information technology, self-healing describes any device or system that has the ability to perceive that it is not operating correctly and, without human intervention, make the necessary adjustments to restore itself to normal operation.
- **Virus check.** At the point of ingest and at regular intervals.

Peter D. Verheyen, Head of Preservation

21 July 2009

Trusted Digital Repository (TDR)

11

- 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.

Peter D. Verheyen, Head of Preservation

21 July 2009

Examples of DP Initiatives and Tools

12

Co-operative Projects

- LOCKSS & CLOCKSS
- Portico

Tools

- Fedora Commons
- D-Space

Hosted Options

- Digital Archive @ OCLC

Peter D. Verheyen, Head of Preservation

21 July 2009

Examples of DP Initiatives and Tools

13

Co-operative Projects

- **LOCKSS (Lots of Copies Keep Stuff Safe):** “Grass roots” international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized “open access” e-content. LOCKSS is OAIS-compliant; the 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.

Peter D. Verheyen, Head of Preservation

21 July 2009

Examples of DP Initiatives and Tools

14

Co-operative Projects

- **CLOCKSS (*Controlled* LOCKSS)** is a joint venture between the world’s leading scholarly publishers and research libraries whose mission is to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of Web-based scholarly publications for the benefit of the greater global research community. Files are preserved as originally published (PDF, HTML, ...) and migrated on access.

Peter D. Verheyen, Head of Preservation

21 July 2009

Examples of DP Initiatives and Tools

15

Co-operative Projects

- **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. The archive is migrated forward en-masse as formats change. A standardized format Similar mission as CLOCKSS but focused on non-open access publications.**

Peter D. Verheyen, Head of Preservation

21 July 2009

Examples of DP Initiatives and Tools

16

Co-operative Projects

When are files released?

Files are released when a “trigger event” occurs.

A “trigger event” is when 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.

Peter D. Verheyen, Head of Preservation

21 July 2009

Tools: Fedora Commons

17

- **Flexible Extensible Digital Object Repository Architecture**
- **Divided into four areas – repository services; preservation services; semantic services; enterprise services. Provides structure to facilitate long-term digital preservation**
 - Can form the core of a trusted, preservation-enabled repository
 - ✦ Metadata
 - ✦ Versioning
 - ✦ ...

Peter D. Verheyen, Head of Preservation

21 July 2009

Tools: OCLC Digital Archive

18

- **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

Peter D. Verheyen, Head of Preservation

21 July 2009

Tools: DSpace

19

- **Repository**
 - Secure managed storage of a wide variety of standard file formats
 - Automated monitoring and workflows
 - Open source
 - Can form the core of a trusted, preservation-enabled repository
 - × **Metadata**
 - × **Versioning**
 - × ...

Peter D. Verheyen, Head of Preservation

21 July 2009

Tools: CONTENTdm

20

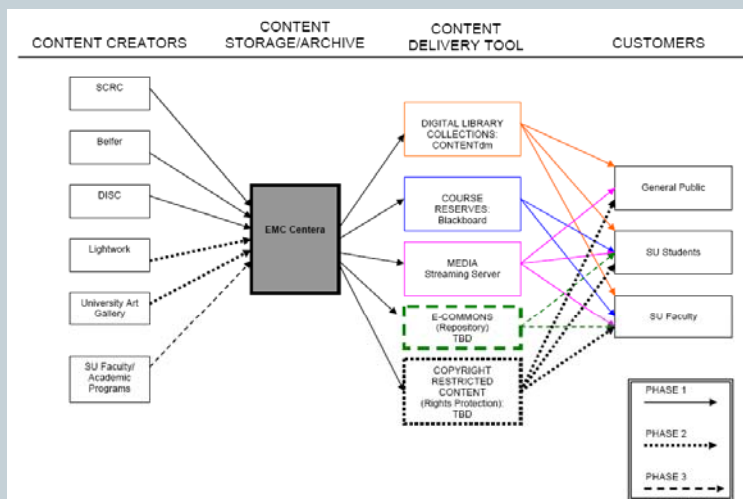
- **Repository**
 - Secure managed storage of a wide variety of standard file formats
 - Can form the core of a trusted, preservation-enabled repository
 - × **Metadata**
 - × **Master images/files**

Peter D. Verheyen, Head of Preservation

21 July 2009

Proposed SU Library Digital Content Management Solution

21



Peter D. Verheyen, Head of Preservation

21 July 2009

What Can We Do – Easy Steps

22

- **Know what you have**
 - Content
 - 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
- **Keep and maintain obsolete hardware as needed... you just might**
- **Find partners to share resources and costs**

Peter D. Verheyen, Head of Preservation

21 July 2009

Further Readings

23

- For a vivid statement of the problem, see <http://www.popularmechanics.com/technology/industry/4201645.html>
- *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
- Digital preservation at the Library of Congress:
<http://www.digitalpreservation.gov/>
- For more on attributes of Trusted Digital Repositories, see <http://www.oclc.org/programs/ourwork/past/trustedrep/default.htm>
- For information on Preservation Metadata, see <http://www.oclc.org/research/projects/pmwg/>
- For examples of collaborative digital preservation efforts, see <http://www.hathitrust.org/> and <http://www.lockss.org/lockss/Home>
- For examples of digital preservation tools, see <http://www.fedora.info/about/> and <http://www.dspace.org/>

Peter D. Verheyen, Head of Preservation

21 July 2009