The Convergence of Records Management & Digital Preservation

Excerpt from: Howard Loos, CRM & Chris Erickson. ARMA. 2014
Brigham Young University

- Private University. Research Universities (ranked 37th in the US)
- Students: 34,101. Full time faculty and staff: 4,000
- Offers 190 bachelor’s, 68 master's and 25 doctoral programs.

University Includes:
- Five Campuses: Provo, Idaho, Hawaii, Salt Lake, Jerusalem
- Harold B Lee Library. Howard W. Hunter Law Library
- Four Museums: Art, Natural History, Paleontology, Anthropology
- Three TV stations, Two Radio stations

Technology Transfer. Assoc. University Technology Managers:
- Ranked number 1 in invention disclosures
- Ranked number 1 in new patent applications; in license income
- Ranked number 1 in startup companies spun out
Harold B. Lee Library

- Library collection: 11,674,388
- Employees: 360
- Top 25 most used digital libraries in U.S.
- Top Ten University Libraries in the United States (The Princeton Review)
- Daily library use: 10,810
- Shelving: 100 miles
- University Archives:
  - 5,200 linear feet of records
  - Acquires 200+ linear feet per year
THE PROGRESSION TO E-RIM

- BYU Established - 1875
- University Archives - 1956
- University Records Management - 1974
- Digital Preservation - 2002
- Rosetta Digital Archive (installed) - 2012
- Electronic Records Program - 2013
RIM’s Mission

To assist departments in fulfilling their responsibility to identify and manage records and information in accordance with legal, regulatory and operational requirements:

- Legal – Federal and state laws/statutes
- Regulatory – Regulating entities
- Operational – Meet business need
- Archival or historical Value
RIM Life Cycle to DP Life Cycle

Capture • Store • Retrieve • Protect

Information Assets

Preservation & Disposition

Inspection & Migration

Retention & Storage

Creation

Maintenance & Use

Paper

Electronic

Film

Collection

Preservation and Transfer

Appraisal

Metadata Prep, File Conversion, Cataloging

Technical Analysis and Prep
RIM and DP Common Objectives

- Work with departments to identify and preserve records of archival or historical value

Then What?

- Ensure paper and digital records of historical value make their way to the University Archives
- Preserve and provide access to digital records over time
**RIM Challenges**

- Electronic records are everywhere
  - PC
  - Shared Drives
  - Outlook
  - DVD/CD/Removable Media

- Electronic records are not identified and managed like their paper counterpart

- Policy does not address electronic records

- Various format – no standards
RIM Challenges – A Few Examples

- Boxes of DVDs – marked as PERM
- Boxes of VHS, floppy discs and other media
RIM Challenges – A Few More Examples

- Using office copy/scanners for digitization projects
- Retiring employees
- Discontinued departments or programs
- Decommissioned systems (with records)
HOW LONG WILL YOUR FILES LAST?

1-5 YRS 3-7 YRS 5-8 YRS

HARD DRIVE DVD FLASH DRIVE
**SUCCESSFUL APPROACHES**

- Update RIM Policy and RRS
- Establish Guidelines
  - File Formats
    - Documents – PDF, PDF/A
    - Photos – JPG or JP2
  - Media (M-Disc)
  - Scanners – file compression (TIFF/G4)
- RIM Plan/Inventory
- Recommended Records Repository – SP2013
HOW LONG WILL YOUR FILES LAST?

1-5 YRS
3-7 YRS
5-8 YRS
1000 YRS

HARD DRIVE
DVD
FLASH DRIVE
M-DISC
There is a way to store permanently
The M-DISC™ is certified and proven to last over 1000 years, offering the best archival data storage solution anywhere. No more worrying about losing something you consider irreplaceable.
What About Microfilm/Microfiche?

- Discontinued microfilming service in 2011
- Long-term storage is moving to:
  - M-Disc
  - Digital Archive
Departments work with RIM to complete a Department RIM Plan. Steps include:

- Develop records inventory
- Map records to RRS
- Identify if paper or electronic (or both)
- Identify where records are stored (location or system)
- Develop an action plan
# Inventory/Department RIM Plan

## Department Records Management Plan/Inventory

<table>
<thead>
<tr>
<th>Item no.</th>
<th>Records Series Title</th>
<th>Records Category (Not yet used)</th>
<th>Medium</th>
<th>Retention Event + Retention Period</th>
<th>Final Disposition</th>
<th>Procedure/Special Instructions/Systems Used (if electronic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Department Meeting Minutes</td>
<td>TBO</td>
<td>Paper/Electronic</td>
<td>PERM</td>
<td>Archives</td>
<td>Retain current year + 5 years in the department. Transfer to the Records Center for 5 years and then transfer to the University Archives.</td>
</tr>
<tr>
<td>2</td>
<td>Kronos Reports</td>
<td>TBO</td>
<td>Electronic</td>
<td>3 Years</td>
<td>Destroy</td>
<td>Retain in department for 3 years and then destroy.</td>
</tr>
<tr>
<td>3</td>
<td>Undergraduate Student Records (department)</td>
<td>TBO</td>
<td>Paper/Electronic</td>
<td>GRA+10 Years</td>
<td>Destroy</td>
<td>Retain in department for 3 years. Transfer to the Records Center for 7 years and then destroy. Retain electronic records in department for 10 years and then destroy.</td>
</tr>
</tbody>
</table>
Digital Preservation Integrated Tools
DIGITAL PRESERVATION (DP)
Preservation Overview

What is Digital Preservation?
Challenges of Long Term Preservation
Best Practices & Processes
Where Do I Begin?
Practical Processes & Technology
Digital Preservation Workflow
What is Digital Preservation?

Digital preservation combines policies, strategies, and actions that ensure access to digital content over time.

Digital preservation is an ongoing obligation.

ALA Annual Conference, Washington, D.C., June 24, 2007
Long Term Challenges

- Obsolescence
  - Hardware
  - Software & Formats
  - Media
- Media Failure
- Storage & Handling
- Metadata and Access
- Organizational Challenges
Hardware / Software
# Format Sustainability

**Sustainability of Digital Formats**  
**Planning for Library of Congress Collections**

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Sustainability Factors</th>
<th>Content Categories</th>
<th>Format Descriptions</th>
<th>Contact</th>
</tr>
</thead>
</table>

Format Descriptions >> Format Description Categories >> Browse Alphabetical List >> Format Descriptions as XML

## Format Descriptions

### Still Image
- SVG 1.1
- TIFF 6
- All still image format descriptions

### Sound
- WAV
- MP3
- All sound format descriptions

### Moving Image
- MPEG-4
- AVI
- All moving image format descriptions

### Textual
- NITF
- XML
- All text format descriptions

### Web Archive
- ARC
- WARC
- All Web archive format descriptions

### Datasets
- DBF
- HDF5
- All dataset format descriptions

### Geospatial
- ESRI_shape
- GeoPDF 2.2
- All geospatial format descriptions

### Generic
- ASF
- RIFF
- All generic format descriptions
The National Archives

The technical registry
PRONOM

Search: By format

1. File formats
   Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.
   *:  
   Search

To search for a particular file format, enter the name of the file format and then click 'search'. Leave the file format name blank to find all file formats in the database.

2. Compatible software
   Enter a file extension and click 'search' to find all software which can process in any way files with that extension.
   *:  
   Search

   Enter a file format name and click 'search' to find all software which can process in any way files of that format.

3. File format risk
   Click Search to find All formats with a risk over 
   Search
Media Obsolescence
Average Lifespan of Digital Media - Years

- Floppy disks
- Hard drives
- Flash media
- CD/DVD/Blu-ray
- Gold CD/DVD/Blu-ray
- Magnetic tape
Average Lifespan of Digital Media - Years

- Floppy disks
- Hard drives
- Flash media
- CD/DVD/Blu-ray
- Gold CD/DVD/Blu-ray
- Magnetic tape

- Actual
- Estimated
# Data Storage Costs

<table>
<thead>
<tr>
<th>Digital Storage Costs</th>
<th>1 TB</th>
<th>50 TBs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Year</td>
<td>Yearly Charge</td>
</tr>
<tr>
<td>Data Center Storage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Center</td>
<td>$1,200</td>
<td>$1,200</td>
</tr>
<tr>
<td>Cloud Storage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon S3 - Regular</td>
<td>$360</td>
<td>$360</td>
</tr>
<tr>
<td>Amazon S3 - Dark copy in Glacier</td>
<td>$480</td>
<td>$480</td>
</tr>
<tr>
<td>Amazon S3 - Reduced &amp; Glacier</td>
<td>$288</td>
<td>$288</td>
</tr>
<tr>
<td>DuraSpace - Preservation</td>
<td>$1,800</td>
<td>$1,800</td>
</tr>
<tr>
<td>DuraSpace - Dark copy in Glacier</td>
<td>$1,925</td>
<td>$1,925</td>
</tr>
<tr>
<td>DuraSpace - Enterprise Plus</td>
<td>$5,625</td>
<td>$5,625</td>
</tr>
<tr>
<td>M-Discs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD (@4.7 GB = 250 Discs / TB)</td>
<td>$638</td>
<td>$0</td>
</tr>
<tr>
<td>BD (@25GB = 40 Discs / TB)</td>
<td>$200</td>
<td>$0</td>
</tr>
</tbody>
</table>
METADATA: Access / Discovery

Importance of Metadata
- Find and use
- Manage resources over time
- Understand, provide context
- Provenance, authenticity

Metadata Schema
- MARC, Dublin Core, PREMIS
- Structure (XML, METS)

Dublin Core
Growing the vocabulary
ORGANIZATIONAL CHALLENGES

Institutional Commitment
- Mission
- Funding
- Staffing
- Training

Trusted Digital Repository

Policies & Procedures

Metadata and Access
1 Trusted Digital Repositories

A Definition

A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. Trusted digital repositories may take different forms; some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance. Whatever the overall infrastructure, however, to meet expectations all trusted digital repositories must:

- 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
- meet the responsibilities detailed in Section 3 of this paper.
Best Practices & Processes

- OAIS Functional Model
- File Creation and Handling
- Institutional Policies
- Metadata Creation
OAIS Functional Model
Institutional Policies

- Create a Digital Preservation plan
- Include all records (email, social media, etc.)
- Follow your Retention Schedule
- Create a Metadata Standard
- Staffing and training policies

Technical Management

- Determine formats to allow
- Run virus scans and fixity checks
- Verify copies regularly
Best Practices - Files

File Creation
- Unique, descriptive file names (annual_rpt_2010.pdf)
- Choose formats carefully (prefer open, non-proprietary)
- Backup up files regularly; store in a safe place

File Management
- Create a digital preservation workflow
- Create good metadata
- Make multiple copies and manage them
- Refresh or migrate media (every 3 – 5 years)
- Monitor formats and migrate as needed
How Do I Start?

- Start Where You Are Now
- Identify The Digital Content You Have
- Select The Content To Preserve
- Establish Metadata Standards
- Decide How To Store Digital Content
- Choose Practical Practices...
Practical Processes

Formats

- Library of Congress website: Sustainability of Formats
- Text: Original + XML, PDF/A, PDF
- Audio: BWAV, FLAC, MP3, at highest resolution
- Images: TIFF, JPEG2000, CR2, PNG, uncompressed, lossless
- Video: Original, AVI, MOV, MP4, reviewing MXF
- Data: CSV, XML

Metadata

- Dublin Core
- METS XML
- Title, creator, date, place, publisher, identifiers, edition, subject ...
Harold B. Lee Library Permanent Digital Archive:
Digital Preservation Decision Tool

Digital Content: ________________________________

Collection Description and Extent: ________________________________

Content Reviewer: __________________________ Position: __________________________

1. Preservation Decision and Preservation Priority:
   (Based on Decision Chart from page 2)

   The final Preservation Decision is: ________________________________
   The final Preservation Priority is: ________________________________

   - Permanently archive content
   - High Preservation Priority
   - Normal Preservation Priority
   - Do not archive content

2. Digital Content and Access / Presentation Options:
   a. What is the quality of the objects to be preserved?
      - High Resolution
      - Low Resolution
      - Internet Archive
      - Other

   b. Are objects in the Digital Archive currently accessible to anyone other than the
      archivist and the object owner? Select relevant options below and list who should view:
      - Access only for archivist and those with permission from the content owner.
      - BYU campus access only.
      - Specific Net IDs or computer IP addresses.
      - May be accessible through library discovery tools.
      - The content is an active website.

   c. Public Access version of objects in the Digital Archive?
      - Do these objects also have publicly accessible versions?
      - Where are the accessible versions located?
      - Are high resolution copies likely to be requested from the Archive?
      - What languages will the metadata be in?

   d. Presentation / Representation. Select the option that is most important:
      - The Intellectual content must retain the original look and feel.
      - The Intellectual content should retain the format if possible.
      - The Intellectual content is more important than the format.

3. Local Storage Options for Long Term Content:
   Long term digital storage outside of the Digital Preservation Archive is also available by using
   Millennia's M-Disks.

4. Additional Considerations. Add documentation as needed.

Preserving Content in the BYU Permanent Digital Archive:
Decision Chart

1. Does the Library have a responsibility to keep this collection for 30+ years?
   - Is this important for the library collection?
     - Yes
     - No
   - Does this fit current collecting levels?
     - Yes
     - No
   - Is this an archival collection?
     - Yes
     - No
   - Is there a legal requirement to keep it?
     - Yes
     - No
   - Is there a contract to keep it?
     - Yes
     - No
   - Responsibility to preserve:
     - Check Anarchival content.
     - Check Normal Preservation Priority.

2. Does the Library have legal permission to:
   - Make digital preservation copies?
     - Yes
     - No
   - Change the digital format if needed?
     - Yes
     - No
   - Make objects accessible to others now, or on a specified date?
     - Yes
     - No
   - Date available available:
   - Copyright permissions:
     - Check BYU Copyright Office.

3. Are the objects available elsewhere?
   - Are these objects commercially licensed by someone else?
     - Yes
     - No
   - If BYU has a contract or permission to preserve the commercial objects, circle No
     - Yes
     - No
   - Is there a physical copy in good condition that will be available long term?
     - Yes
     - No
   - Is there a digital copy permanently archived in a Trusted Digital Repository?
     - Yes
     - No
   - 3. Other copies:
      - Check Permanently Archive content.

4. What is the need / priority to preserve items?
   - Is the content deteriorating or at risk?
     - Yes
     - No
   - Possible risk conditions include:
     - Deterioration or poor condition
     - Near-extinct media format
     - Projected object lifespan: 3 - 10 years
     - Content is ephemeral or transient
     - Media viewer unavailable or discerning
     - Would be difficult or impossible to recreate the digital objects if lost?
     - Yes
     - No
   - Need / Priority:
     - Check High Preservation Priority

5. Preservation Decision:
   - Permanently archive content in the Digital Archive
   - Do not archive content in the Digital Archive

6. Preservation Priority:
   - High Preservation Priority
   - Normal Preservation Priority

Name: __________________________ Ext. __________________________ Date: __________________________
Digital Preservation Metadata

- Based on Dublin Core
- Mountain West Digital Library Standard
- Simple Crosswalk
- Required Fields

Original Cataloging
- Harvested into Rosetta
- METS XML structure

Preservation Metadata

- Policies and Procedures
  http://sites.lib.byu.edu/digitalpreservation/

---

**Preservation Metadata Schema**

<table>
<thead>
<tr>
<th>Dublin Core label</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Author/photographer/creator, etc.</td>
<td>Obinhyms, the Blind, ca. 313-ca. 388:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown</td>
</tr>
<tr>
<td>Contributor</td>
<td>Compiler/contributor, etc.</td>
<td>Como, Walter Mason, 1867-1925:</td>
</tr>
<tr>
<td>TITLE</td>
<td>title</td>
<td>Lt. Beecher</td>
</tr>
<tr>
<td>DATE</td>
<td>date original</td>
<td>ca. 1867</td>
</tr>
<tr>
<td></td>
<td>date digital</td>
<td>2009-07</td>
</tr>
<tr>
<td></td>
<td>date created</td>
<td>2007-09-25</td>
</tr>
<tr>
<td></td>
<td>physical description</td>
<td>Electronic reproduction</td>
</tr>
<tr>
<td></td>
<td>dimensions</td>
<td>3 x 4, silver print</td>
</tr>
<tr>
<td></td>
<td>medium</td>
<td>3 x 4, silver print</td>
</tr>
<tr>
<td>Format</td>
<td>The content and the file format of</td>
<td>Image/jpeg; Audio/mp4</td>
</tr>
<tr>
<td>IDENTIFIER</td>
<td>call number</td>
<td>METS 1151</td>
</tr>
<tr>
<td></td>
<td>persistent id</td>
<td>handle:/1877/240453</td>
</tr>
<tr>
<td></td>
<td>Link to source system</td>
<td>byu123456</td>
</tr>
<tr>
<td></td>
<td>page-url / filename</td>
<td>e.g. <a href="http://lib.contestion.odcr.org/s/jCamp.236">http://lib.contestion.odcr.org/s/jCamp.236</a></td>
</tr>
<tr>
<td></td>
<td>Link to finding aid file name</td>
<td>e.g. <a href="http://lib.byu.edu/ead/vmu/MSS16.xml">http://lib.byu.edu/ead/vmu/MSS16.xml</a> 001_mmp155.jpg</td>
</tr>
<tr>
<td>Provenance</td>
<td>statement of ownership, etc.</td>
<td>Converted from TIFF file</td>
</tr>
<tr>
<td>Publisher</td>
<td>publisher original</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td></td>
<td>publisher digital</td>
<td></td>
</tr>
<tr>
<td>Relation **</td>
<td>isPartOf / collection name</td>
<td>Walter M. Camp Photograph Collection</td>
</tr>
<tr>
<td>RIGHTS</td>
<td>patron use instructions</td>
<td><a href="http://www.lib.byu.edu/uc_copyright.html">http://www.lib.byu.edu/uc_copyright.html</a></td>
</tr>
<tr>
<td></td>
<td>copyright status/owner</td>
<td>Public Domain; Courtesy L. Tom Perry Special Collections, Harold B. Lee Library, Brigham Young University</td>
</tr>
<tr>
<td></td>
<td>access level</td>
<td>Public</td>
</tr>
<tr>
<td>Source</td>
<td>lending institution</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td>Subject</td>
<td>subject names</td>
<td>Beecher, Frederick Henry</td>
</tr>
<tr>
<td></td>
<td>subject geogrcaphic, topical, names</td>
<td>Photographs; Portraits</td>
</tr>
<tr>
<td></td>
<td>genre</td>
<td>Beecher Island, Bottle of, 1868;</td>
</tr>
<tr>
<td>Type</td>
<td>Broad description of the CONTENT in</td>
<td>Collection; stillImage;</td>
</tr>
<tr>
<td></td>
<td>the digital object</td>
<td>From DCMI Metadata Types;</td>
</tr>
<tr>
<td></td>
<td>Conversion Specifications</td>
<td>jpeg Image created using Photoshop software version 3</td>
</tr>
<tr>
<td>Coverage</td>
<td>coverage</td>
<td>portrait</td>
</tr>
</tbody>
</table>

* Fields with uppercase LABELS are required for preservation.
** If objects are part of a collection to be grouped in Rosetta, the collection name is required.

January 29, 2013
Estimated Average Life of Storage Media - Years

- Floppy disks
- Flash media
- Hard drives
- CD/DVD/Blu-ray
- Gold CD/DVD
- Magnetic tape
- Microfilm
- M-Discs
Preservation Workflow

Receive content
- Identification, appraisal, selection

Technical analysis
- Virus / malware scan
- Generate checksums: MD5 / SHA1
- Evaluate format and file: PRONOM, DROID, FITS
- Manage File and format: conversion, normalize, remove personal information

Description
- Create metadata, catalog, finding aid

Preparation & Disposition
- Prepare Information Packet
- Generate copies and transfer files
- Ingest into repository
Rosetta installed in March 2012
- Production Environment
- Development / Testing Environment

Integrate with Library Workflows & Processes
Selection Process & Decision Form
Harvest Tool & Ingest Processes
Library of Congress Services

Interact with knowledgeable librarians and subject specialists on-site & online.

RESEARCH & REFERENCE SERVICES
- Acquisitions
- Archival Description
- Ask a Librarian
- Cataloging, Classification
- Cooperative Cataloging Programs
- Copyright
- Duplication Services
- Interlibrary Loan
- Law Library
- Library Card (Reader Registration)
- Library Standards

SERVICES
- Linked Data Service
- Preservation
- Publishing
- Resource Description & Access

PROGRAMS
- American Folklife Center
- Center for the Book
- Concerts
- Copyright Royalty Board
- Digital Preservation
- FEDLINK
- John W. Kluge Center
- NLS: Services to the Blind & Physically Handicapped
- Poetry & Literature Center
- Veterans History Project
- World Digital Library

Discover how the Library of Congress acquires materials for its collections and for other research institutions.

Search & browse examples, standards, and help for creating catalog records and finding aids for discovery and use of personal papers, organizational records, and other archival collections.

Specialists -- online and in our reading rooms in Washington, D.C. Go beyond the digital collections.
Digital Preservation Outreach & Education Network

DPOE People
The DPOE Network is made up of the DPOE National Trainer Network, the DPOE Steering Committee, and a community of Digital Preservation Education Advocates.

DPOE National Trainer Network

The DPOE National Trainer Network is made up of individuals who have attended a DPOE Train-the-Trainer workshop. The network covers six regions: Midwest, Northeast, Northwest, Southcentral, Southeast, and Southwest.
Questions?

http://sites.lib.byu.edu/digitalpreservation/
http://preservationmatters.blogspot.com/

howard_loos@byu.edu
chris_erickson@byu.edu