Libraries and the Preservation of Cultural Heritage

Building on the Past to Develop Our Future

Chris Erickson. Fortaleza 2017
Cultural heritage:

- Consists of tangible and intangible, natural and cultural, movable and immovable assets inherited from the past.
- Extremely valuable for the present and the future of communities.
- Access, preservation, and education of cultural heritage are essential elements for people and their culture.
Preserving Cultural Heritage

Importance of cultural heritage:

- Preservation is not just for its own sake, nor even just to record the past, though that it valuable.
- Preserving the past is important because of the effects it may have on the present and future.
- Preservation is about keeping alive the knowledge, wisdom, and heritage of the past so that it may enlighten us today, tomorrow and forever.
Digital Preservation Concepts

- Foundational concepts of digital preservation
  - Digital Stewardship
  - Trusted Repository
  - OAIS Reference model
Five Organizational Stages of Digital Preservation

1. Acknowledge: Digital preservation is a local concern
2. Act: Initiate digital preservation projects
3. Consolidate: Transition from projects to programs
4. Institutionalize: Incorporate the larger environment
5. Externalize: Embrace inter-institutional collaboration

Anne R. Kenney & Nancy Y. McGovern
OAIS Digital Preservation Model

Reference Model for an Open Archival Information System (OAIS)
**Digital Preservation Models**

Three-legged Stool

- Organizational infrastructure (the "what")
- Technological infrastructure (the "how")
- Resources framework (the "how much")

Adapted from: Kenney and McGovern, 2003
Digital Preservation Models

Inter-related Activities: The Triad

- Management-related activities
- Technological activities
- Content-centered activities
Digital Preservation Practices

Managing Digital Content over Time

- 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 digital content?
- Manage - what is needed for long-term management?
- Provide - what plans are there for long-term access?
Digital Preservation Practices

Identify Your Digital Content

- A detailed inventory is the best way to identify content
- Preservation requires a commitment of resources
- Identifying content part of making informed decisions
- Effective planning is based on knowing preservation needs
- Not all of digital content will be preserved
Digital Preservation Practices

Select - what content will be preserved?

- Matching mission to content
- Document selection decisions
- Determine Selection Criteria
- Selection is an ongoing process
Digital Preservation Practices

Storage - what are the long term issues?

- Determine the storage that fits your organization
  - How many copies to store?
  - Storage costs?
  - Institutional constraints and restrictions
- Preservation content includes objects and metadata
- Storage may be cheap, management is not
- 3 – 2 – 1 Storage Rule
Digital Preservation Practices

Protect – what is the problem?
- Change and loss – accidental and intentional
- Obsolescence – as technology evolves
- Inappropriate access – e.g. confidential or restricted data
- Disasters – emergencies of all kinds
Manage - what is needed long-term?

A balanced approach will address:

- Organizational requirements and objectives
- Technological opportunities and change
- Resources – funding, staff, equipment, etc.
Digital Preservation Practices

Provide – plans for long-term access?

- Purpose: ensure future access
- Preservation makes long-term access possible
- Implement proven technologies
- Create new versions of content over time
Brigham Young University
Harold B. Lee Library

- Library collection: 11,674,388
- Employees: 376
- Top 25 most used digital libraries in U.S.
- Daily library patrons: 10,810
- Shelving: 100 miles

University Archives:
- 5,200 linear feet of records
- Acquires 200+ linear feet per year

Digital Preservation:
- Program established 2002
- Rosetta Archive began 2012
Digital Preservation Implementation

Digital Preservation **Mission** Statement:
- To preserve and maintain readability and accessibility for all selected digital materials produced within or acquired by Brigham Young University

Created a Digital Preservation **Plan**
- Review regularly with Library Administration

Established Digital Preservation **Workflow**
- Directed by the Digital Preservation Committee
- Rosetta Digital Archive – a dark archive for preservation
- Access to files is through On Line Systems
To preserve University digital items:
- Media: images, audio, video
- Texts: Documents, university records and other materials

To manage these materials over time
- Multiple versions
- Manage formats
- Maintain standard objects and metadata

To ensure long term access
- Rosetta: a Digital Preservation Repository
- Access through other library systems
The OAIS Digital Preservation Model

Reference Model for an Open Archival Information System (OAIS)
Our Digital Preservation Model

- University Records
- University Archives
- Faculty Research
- Curriculum content
- Campus Depts.
- Digital Collections
- Library Resources
- Special Collections
- ... Others

- Preservation Research and Planning
- Standards Metadata
- Selection & Appraisal
- Cataloging
- LIS / Data Management
- Processing Testing
- Archival Storage
- Distribution Info
- Access Services
- User Access
- Faculty & Students

Administration of Information System

Digital Preservation Committee
- Records Management
- Archives
- Administration
- Library
- Legal
- IT Department

Modeled on OAIS Standard
Digital Preservation Archive

Rosetta Digital Archive Implemented in 2012

Central to Preserving University Content

- University Library & Special Collections
- Cultural & Historical Collections
- University Archives & Records
- Institutional Repository
Rosetta Digital Preservation System
Digital Preservation Network (DPN):

- Includes 60+ research libraries throughout the United States.
- Formed to ensure that the complete scholarly record is preserved.
- DPN is a dark archive with a federated approach to preservation.
- Harold B. Lee Library is a charter member.

Rosetta Implementation
- Export Rosetta content
- Ingest in DuraCloud node
Digital Preservation Implementation

Inventory, Selection and Acquisition
- Preservation Decision / Selection Form to complete
  - Required by Special Collections Coordinating Committee

Selection by Established Standards
- Digital Preservation Priorities
- Levels of Digital Preservation

Preservation of the “Intellectual Entities”
- Best quality digital object in original format +
- Full metadata from cataloging records
Preservation Selection Form

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

2. Digital Content and Access / Presentation Options:
   a. What is the quality of the objects to be preserved?
      - High Resolution
      - Low Resolution
      - Informal Archive
      - Other: ____________________________
   b. Are the objects in the Digital Archive currently accessible only 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. Is the archived website content restricted?
   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 language(s) 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 M-Disks.

4. Additional Considerations. Add documentation as needed.

Preserving Content in the BYU Permanent Digital Archive:

Decision Chart

Appraisal and selection questions: Circle your response and follow the arrow:

1. Does the Library have a responsibility to keep this collection for 20+ years?
   - Is this important for the library collection? Yes No
   - Is this an archival collection? Yes No
   - Is there a legal requirement to keep it? Yes No

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: ____________________________

3. Are the objects available elsewhere?
   - Are these objects commercially being preserved by someone else? Yes No
   - If BYU has a contract or permission to preserve the commercial objects, circle No
   - Is there a physical copy in good condition that will be available long-term? Yes No
   - Is there digital copy permanently archived in a Trusted Digital Repository? Yes No
   - Which repository?

4. What is the need/priority to preserve items?
   - Is the content deteriorating or at risk? Yes No
     - Possible risk conditions include:
       - Deteriorating or poor condition
       - Near-obsolescent media or format
       - Projected object lifespan: 1 - 3 years
       - Content is ephemeral or transitory
       - Media viewer unavailable or disconcerting
     - Would it be difficult or impossible to recreate the digital objects if lost? Yes No

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

Preservation Priority:
   - High Preservation Priority
   - Normal Preservation Priority

Responses and directions: Indicate the decision / answer:

1. Responsibility to preserve:
   - If Yes to any question, go to section 2.
   - If No or all questions, go to Preservation Decision below:
     - Check Do not archive content
     - Do not store digital master files

2. Copyright permissions:
   - If yes, contact the BYU Copyright Office before continuing:
     - If Yes to any question, go to Preservation Decision below:
       - Check Do not archive content
       - Do not store digital master files

3. Other copies:
   - If No to all questions, go to Preservation Decision below:
   - Check Permanently Archive content
   - Go to section 4 to decide the priority:
   - Check High Preservation Priority

4. If yes to any question:
   - Check Normal Preservation Priority
   - Go to Preservation Decision below:

5. Check Preservation Decision:

Name: ____________________________ Ext. ____________________________ Date: ____________________________
Preservation Priorities

1. BYU created content, unique, with no physical copy
2. BYU owned, unique, At-risk items
3. Items with an At-risk physical copy
4. Digital content difficult or costly to reproduce
5. Items digitized for convenience
Levels of Digital Preservation

**Preservation Levels**

0. No preservation.
   - Library IT shared servers
   - Scheduled backups only

1. Basic preservation
   - Access copy: M-Disc or Content Management Repository
   - No other processing

2. Full preservation
   - Rosetta system
   - Preservation actions: format migration, fixity checks, etc.
   - Metadata management
   - Copies: Rosetta; Campus Data Center; Tape in offsite vault;

3. Extended preservation
   - Full preservation actions: format migration, fixity checks, etc.
   - Remote online copy: DPN, Amazon
Digital Preservation Metadata

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

Original Cataloging
- Harvested into Rosetta
- METS XML structure

Policies and Procedures
http://sites.lib.byu.edu/digitalpreservation/
Preservation Formats & Standards

**Formats**

  - Text files: Original + XML, PDF/A, PDF
  - Audio files: B WAV, FLAC, MP3, at highest resolution
  - Image files: TIFF, JPEG*, CR2, PNG; uncompressed, lossless
  - Video files: Original format, AVI, MOV, MP4; reviewing these
  - Data files: CSV, XML

**Metadata**

- Dublin Core
  - Title, creator, date, place, publisher, identifiers, edition, subject ...
- METS XML (includes structure map and file relationships)
Harvest and Ingest Tool

**Harvest content to preserve from sources**
- CONTENTdm repository
- SharePoint University Record System
- Internet Archive
- bePress Digital Commons / OJS Electronic Journals
- Unstructured Data files: metadata in csv, xls, or xlsx templates

**Metadata**
- Harvest metadata from source repository in xml format
- Or we accept ‘user created’ metadata in spreadsheets
- Create SIP: combine original object and Mets Xml
- Ingest into Rosetta Digital Archive
Digital Preservation Workflow

Receive content
- Identification, appraisal, selection

Technical analysis & Processing
- Virus / malware scan
- Generate checksums: MD5 / SHA1 / CRC32
- Evaluate format and file: PRONOM, DROID, Extract Technical Metadata
- 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
Workflow Overview

HBL Library
Long Term Digital Archive Program

Selection, Acquisition
- HBL
- Digital Content Producers
  - Digital Lab
  - Special Collections & Archives
  - Multi-Media Unit
  - AV Presentation
  - Public Services
  - Institutional Repository
  - University Records

Processing
- Committee Approval
  - Decision
  - Priority
  - Level
- Processing Formatting Bit Curator
  - Digitization
  - Metadata
  - Transcription
- Public Access Systems
  - ContentDM
  - BePress
  - Internal Archive

Disposition
- Archival Copy: Tape Offsite
- Archival Copy: Data Center
- Online Data Repository
Importance of Digital Preservation

“We have to get this right. If we don't, then not much else that we’re doing in research libraries matters. If we don’t fully develop a sustainable digital preservation program, we could negatively impact whole areas of research, because materials created right now could just disappear. I think about gaps that exist in records because of manmade events and natural disasters. This could be a disaster of our own making.”

Jennifer Paustenbaugh. University Librarian
Brigham Young University
What is Your Next Step?

What is your current situation?

What 2 or 3 items are needed most?

For each of them, write down:

- Who would you work with to accomplish it?
- How would you address the challenges you might encounter?
- How long will it take?
- What do you expect the outcome will be?

Which item will you start now?
For More Information:

Digital Preservation Matters
http://preservationmatters.blogspot.com/

Digital Preservation at BYU
https://sites.lib.byu.edu/digitalpreservation/
Muito obrigado!