Digital Preservation @

The Harold B. Lee Library
How we got here?

1. Understanding Digital Preservation
2. Search for Content
3. Maintain Optical Disc Storage
4. In House Preservation System
5. M-Disc Storage
6. Rosetta System
How much library data?

- **Overall Library**
  - Total ~ 80.0 TB

- **Special Collections:**
  - CD 15 TBs
  - DVD 30 TBs
  - HD 7 TBs

- **University DataSets**
  - Total ??? PB
Long Term Digital Vision

- To preserve BYU digital items:
  - Digitized images, audio, video,
  - Electronic documents, articles, university records, web sites, and other materials

- To manage these materials over time:
  - Manage Formats
  - Maintain Standard Objects and Metadata

- To provide long term access
  - Digital Preservation Repository
  - There is no preservation without access
Rosetta Implementation Goals

Rosetta selected as a digital preservation system

1. Identify the **people** needed for digital preservation;

2. Recommend **policies** to be created or updated

3. Outline **procedures** and workflows for staff
Project Goals: 1. People

What we learned about people for Rosetta:

1. Library Administration
2. Subject Specialists
3. Rosetta Team:
   ● System Administration
   ● Content Specialist
   ● Technical:
     ● Utilities Development & Data Programmer
     ● Technical support
Project Goals: 2. Policies

What we learned about policies:

1. Updated our overall digital preservation policy
2. Created selection and approval policies
   - Created a selection decision form
   - Collection Management Committee
   - Board of Curators
3. Created preservation metadata schema
4. Outlining procedures and workflows for staff
5. Creating procedures for campus partners
Content 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: [ ] Permanently archive content [ ] Do not archive content

   The Final Preservation Priority is: [ ] 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 [ ] Internet Archive [ ] Other

   b. Are the 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 ID 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-Discs.

4. Additional Considerations. Add documentation as needed.

Preserving Content in the BYU Permanent Digital Archive:
Decision Chart

Assessment and Selection Questions:
Circle your response and follow the arrows:

1. Does the Library have a responsibility to keep this collection for 20+ years?
   [ ] Yes [ ] No
   [ ] If yes to any question, go to section 2.
   [ ] If no to all questions:
      [ ] Go to Preservation Decision below

2. Does the Library have legal permission to:
   [ ] Make digital preservation copies?
   [ ] Change the digital format if needed?
   [ ] Make objects accessible to others, or on a specified date?
   [ ] Date available

3. Are the objects available elsewhere?
   [ ] Are these objects commercial items preserved by someone else?
   [ ] If BYU has a contract or permission to preserve the commercial objects, wish the preservation process go forward?
   [ ] Is there a physical copy in good condition that will be available long term?
   [ ] Is there a digital copy permanently archived in a Trusted Digital Repository?
   [ ] Which repository?

4. What is the need/priority to preserve items?
   [ ] Is the content deteriorating or at risk?
   [ ] Possible risk conditions include: [ ] Deteriorating or poor condition
      [ ] Near-obsolesce media or format
      [ ] Projected object lifespan: 1 - 3 years
      [ ] Content is ephemeral or transitory
      [ ] Media viewer unavailable or discontinuing
      [ ] 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

Name: 
Ext. 
Date: 
Selected Digital Content

Rosetta to include:
- University Archives & Records
- CONTENTdm Collections
- Special Collections Requests
- Curators & specialists: Selections
- Campus Departments
- Various file types and formats

Rosetta to exclude:
- Digital files without copyright permissions
- Digital files that aren’t a preservation copy
- Commercial products
- Restricted content

Examples:
- HBLL Digital Lab images and files
- Scanned books
- CONTENTdm Collections
- Audio and Video
- BYU Digital Journals
- Institutional Repository
### Digital Preservation Metadata

#### Preservation Metadata
- Based on Dublin Core
- Mountain West Digital Library Standard

#### Original Cataloging
- Harvested into Rosetta
- METS XML structure

---

<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>Odysseus, the Blind, ca. 313-ca. 308; Unknown</td>
</tr>
<tr>
<td>Contributor</td>
<td>Compiler/contributor, etc.</td>
<td>Cono, 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, date digital, date created, date modified</td>
<td>ca. 1867, 2003-07, 2007-09-25</td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td>Item description, abstract edition, physical description, dimensions, medium</td>
<td>Portrait of Lt. Beecher, nephew of Henry Ward Beecher. Electronic reproduction. 3 x 4, silver print. 3 x 4 in. silver print</td>
</tr>
<tr>
<td>Format</td>
<td>The content and the file format of the digital object.</td>
<td>JPEG 100; Audio/MP3 From <a href="http://www.lib.byu.edu/aws/assignments/media-type/index.html">http://www.lib.byu.edu/aws/assignments/media-type/index.html</a></td>
</tr>
<tr>
<td>IDENTIFIER</td>
<td>call number, persistent id, link to source system, page url / centpath, link to finding aid, file name</td>
<td>MSS Pitem 1 handle://1877/234543 byu1235646 e.g. <a href="http://libcontentdm.oclc.org/u7/camp.216">http://libcontentdm.oclc.org/u7/camp.216</a> e.g. <a href="http://lib.byu.edu/IAD/XMU/MSSP16.xml">http://lib.byu.edu/IAD/XMU/MSSP16.xml</a> 001_mssp16.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, publisher digital</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td>Relation **</td>
<td>isPartOf / collection name</td>
<td>Wulff M. Camp Photograph Collection</td>
</tr>
<tr>
<td>RIGHTS</td>
<td>patron use instructions, copyright status/owner, access level</td>
<td>Public Domain; Courtesy L. Tom Perry Special Collections, Harold B. Lee Library, Brigham Young University; Public</td>
</tr>
<tr>
<td>Source</td>
<td>learning institution</td>
<td>Brigham Young University</td>
</tr>
<tr>
<td>Subject</td>
<td>subject names, geographic, topical, names genre</td>
<td>Beecher, Frederick Henry; Photographs; Portraits; Beecher Island, Battle of, 1868;</td>
</tr>
<tr>
<td>Type</td>
<td>Broad description of the CONTENT in the digital object, Conversion Specifications</td>
<td>Collection; still images; From DCMIP Metadata Types; <a href="http://dublincore.org/documents/dcmip-terms/">http://dublincore.org/documents/dcmip-terms/</a></td>
</tr>
<tr>
<td>Conversion</td>
<td>technical specifications</td>
<td>64 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
Digital Preservation Workflow

BYU Harold B. Lee Library
Digital Preservation Education

Presented Library of Congress Workshop:

**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 to protect your digital content?
**Manage** - what provisions for long-term management?
**Provide** - what considerations for long-term access?
Project Goals: 3. Procedures

What we learned about procedures and workflows:

- Use existing workflows where possible
  - New collection approval
  - Metadata updates
  - Static and Dynamic collections
- Data harvest and ingest
- University Records
Accessing Rosetta Content

- Rosetta is primarily a preservation repository
- Rosetta will be available to subject specialists
- Rosetta content can be available in ScholarSearch
  - Limited access to some Rosetta content (university records, etc.)
  - Policies will determine the level of access
Rosetta Implementation Plan

- **Phase 1:**
  - Setup, Install, Training
  - Preservation metadata
  - Preservation workflows
  - Selection / ingest policies
  - Content: images, PDFs, documents, audio / video

- **Phase 2:**
  - Media and mixed formats
  - Digital Lab content
  - Internet Archive files
  - Standards & best practices
  - Redundant backup
  - CONTENTdm ingest

**Full Production: 2013**

- Install Rosetta Production environment
- Training and testing environment
- Library Employee Training and Documentation
Questions?

Chris Erickson
2013

chris_erickson@byu.edu