INTEGRATING ROSETTA INTO UNIVERSITY WORKFLOWS

FOR LONG TERM PRESERVATION

CHRIS ERICKSON. JERUSALEM 2014
Overview

Overview Of Brigham Young University
Rosetta: Our long term preservation
Harvest & Ingest Processes
Digital Content & Repositories
Digital Preservation Workflow
**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
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
Associations and Consortia related to digital preservation:
Digital Preservation Network (DPN)

Digital Preservation Network:
- 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.
- Pilot with Rosetta this year.
Portal to digital resources in **159** organizations
- Universities, colleges, public libraries, museums, archives, government agencies, and historical societies
- Utah, Nevada, Idaho, Arizona, Montana, and Hawaii.
- Primo Search platform.
- Digital Public Library of America partner.

Top 10 digital hubs
ARL is a nonprofit organization of 125 research libraries at comprehensive, research-extensive institutions in the US and Canada.
Rosetta in the Library

- Rosetta installed in March 2012
  - Production Environment
  - Development / Testing Environment
- Integrate with Library Workflows & Processes
- Selection Process & Decision Form
- Custom Harvest Tool & Ingest Processes
Digital Content

Location of Digital Objects
- Various Discs & Drives
- On Line Library Repositories
- Campus partners

Digital File Characteristics
- Digital Files: images, text, audio, video, office documents, etc.
- Single and compound objects
- Special handling, e.g. file headers
- Metadata
Harvest Process

Create SIPS from Digital Collections
- Complete metadata – standard METS xml file
- Rosetta data – Dublin Core xml file
- Include original cataloging information in xml (METS Source)
- Metadata: collect from repositories, csv or Excel files
- Extract objects from web services or file upload

Harvest Tool Parameters for Rosetta
- Collection
- Retention Policy
- Access Rights
- Harvest selection by Date / Time
Ingest Process

Ingest SIPS into Rosetta
- Rosetta xml files + Object files
- Deposits SIPS in Rosetta
- Stores deposited SIPS and log files until verified

Ingest Parameters for Rosetta
- Producer
- Material Flow
- Rosetta user
- Provide Status Report
Library Repositories

- Historical Image Collections (CONTENTdm)
- Institutional Repository
- Open Journal Systems (OJS)
- University Archives and Records
- Internet Archive
- Others . . .
HISTORICAL IMAGE COLLECTIONS

- CONTENTdm System; Hosted at OCLC
- Approximately 4.5 million items
- Harvest is the final step in workflow
  - Retrieve CONTENTdm metadata with web services
  - Include derivative files or master files
  - Single and compound objects
- Store additional copy on Millenniata M-disc
INSTITUTIONAL REPOSITORY

Scholarly and creative content of Campus

- Scholarly articles, posters, research, monographs, student publications, audio, video, source code, etc.
- ETDs, Herbarium, Historical clothing collection, etc.

ScholarsArchive currently in CONTENTdm

Other content from Colleges & Departments

- Subject Guides
- Educational & instructional content
- Audio / Video files (campus streaming)
- Web sites
Electronic Campus Journals

- Archive On-Campus digital journals
- Some reformatted, some digital-only
- Currently use Open Journal Software (OJS)

Harvesting into Rosetta

- Web services; created OJS export plugin
- Objects organized in the XML structure in page order
- Each journal is a collection in Rosetta
- Each issue is an IE in the collection
- Every article and section is part of the IE
Internet Archive scanners

**Process:**
- Access through Web services
- Harvest in entry date order
- Create one IE per title

**Harvest:**
- Multiple metadata files
- Raw and processed images
Internet Archive

Internet Archive scanners

Process:
- Access through Web services
- Harvest in entry date order
- Create one IE per title

Harvest:
- Multiple metadata files
- Raw and processed images
University Content

Historical, Institutional, Archival content
- University records (SharePoint)
- Institutional historical files
- University Photographs, documents, audio, video ...
- Newspapers (digitized and digital born)

Harvest & Ingest process
- Metadata in spreadsheet (csv, .xls, .xlsx)
- Objects in folder
- Create SIP with METS xml & Ingest into Rosetta
Digital Preservation Workflow
OTHER PROJECTS

Special Collections & Manuscripts
- Folder discovery and harvest tool
- Acquisitions workstations – In Library and On The Road
- ArchivesSpace interface

University Content
- Harvest tool for camera images with metadata
- Ingest tool for permanent records from SharePoint

Academic Consortia
- Pilot project with DPN
- Mountain West Digital Library
Digital Preservation Integrated Tools

DPN

Harold B Lee Library (CONTENTdm)
University Communications (csv ingest)
Special Collections (Internet Archive, Folder ingest)

University Photographer (photo ingest)
Howard W Hunter Law Library (csv ingest)
University Archives (ArchiveSpace)

Colleges & Departments (csv ingest)
University Records (Sharepoint)
Institutional Repository (OJS, CONTENTdm)

MWDL

Primo

ExLibris

Rosetta
Questions?
cle@byu.edu