Digital Collection Processing, Access, Preservation: solutions for a small budget
Lynn Klundt and Rachel Senese
Northern State University
Beulah Williams Library
Aberdeen, SD
Digital Collection Processing, Access, Preservation: solutions for a small budget
Lynn Klundt and Rachel Senese
Northern State University
Beulah Williams Library
Aberdeen, SD
Library and Archives History

Northern State University

Beulah Williams Library

History of the archives

Beulah Williams Library
http://www.northern.edu/library/pages/default.aspx
Digital Library of South Dakota  
http://dslblsdin.net/ 

- Composed of a consortium of seven South Dakota universities 
- The members of the consortium were chosen to make a variety of unique collections available in a digital format 
- Created for the purpose of creating a statewide digital repository 
- Each of the six libraries received a piece of unique digitizing equipment and supportive funding from the BOR 

NSU Digital Goals 

- Continually research and refine programs and workflows for greater efficiency 
- Enhance access to our digital collections 
- Expand our collections through scanning and collecting born digital materials 
- Provide storage and back up of digital collections for campus and surrounding communities 

Technology Acquired 

- Funding to help pay for a portion of our OCLC CONTENTdm instance 
- Epson Expression 10000XL photo flatbed automatic document feeder 
- Nikon Super CoolScan 9000ED slide and film scanner 
- Two Gateway computers 
- North Light HD copy light (2) 
- Vacuum Board 
- Nikon SLR camera 
- Better Light camera
Digital Library of South Dakota

http://dlsd.sdln.net/

• Composed of a consortium of seven South Dakota universities

• The members of the consortium were chosen to make a variety of unique collections available in a digital format

• Created for the purpose of creating a statewide digital repository

• Each of the six libraries received a piece of unique digitizing equipment and supportive funding from the BOR

• Funding to help pay for our OCLC CONTENT

• Epson Expression 10 flatbed automatic document scanner

• Nikon Super CoolScan slide and film scanner

• Two Gateway computers

• North Light HID copy light

• Vacuum Board
Technology Acquired

- Funding to help pay for a portion of our OCLC CONTENTdm instance
- Epson Expression 10000XL photo flatbed automatic document feeder
- Nikon Super CoolScan 9000ED slide and film scanner
- Two Gateway computers
- North Light HID copy light (2)
- Vacuum Board
- Nikon SLR camera
- Better Light camera
NSU Digital Goals

• Continually research and refine programs and workflows for greater efficiency

• Enhance access to our digital collections

• Expand our collections through scanning and collecting born digital materials

• Provide storage and back up of digital collections for campus and surrounding communities
• A digital repository created by OCLC
  
  • Handles the storage, management, and delivery of collections to users across the web. (CONTENTdm.org)

• Purchased with funding from the Digital Library of South Dakota Funding and the Knight Foundation Grant.

• All of the library's digital projects are uploaded and described in CONTENTdm.

• Software used with CONTENTdm
  
  • Adobe Acrobat
  • Photoshop

digitalcollections.northern.edu
NSU Digital Goals

- Continually research and refine programs and workflows for greater efficiency
- Enhance access to our digital collections
- Expand our collections through scanning and collecting born digital media
NSU History

Constantly updated and augmented, the NSU History collection contains items from every aspect of Northern's past. It contains photos, news clippings, Northern publications, and programs from events that happened here. Also included is a collection of memorabilia such as letter jackets, bumper stickers, Gypsy Day costumes and scrapbooks. Mostly provided by donors, these items are often used for alumni displays at events.
**Eureka Pioneer Museum**

This collection contains various items from the Eureka Pioneer Museum in Eureka, SD. Many of the items pertain to the local Germans from Russia population. With individual photographs, photo albums, and documents, the collection gives fascinating insight into the Eureka area.

**Frederick Historical Collection**

This collection includes photographs and documents from the museum in Frederick, SD.

**South Dakota Germans from Russia**

This collection was donated by the Riechel family whose ancestors came to the United States from Russia in 1896 and settled in Strasburg, S. D. The collection includes documents, photographs, and recipes from various German cooks. Among the recipes are those of the Riechel family. It is the largest of its kind in the United States. An additional collection was donated by the Strasburg German High School.
Aberdeen Area Historical Resources Collection

This collection contains resources pertaining to the history of Aberdeen and the surrounding area. Gathered from many different sources in the community, the collection contains photographs, manuscripts, documents, maps and other items.

http://digitalcollections.northern.edu/cdm/landingpage/collection/local
- Open-source web-based publishing platform

- Provides pre-made content manager for collections, library and museum metadata

- Omeka allows for limited usage of Dublin Core metadata

- Have a server and hosting site (DreamHost)

**Software used with Omeka**

- Cyberduck (free)  
  [http://cyberduck.info](http://cyberduck.info)

- Global Mapper  
  [www.globalmapper.com](http://www.globalmapper.com)

- Komodo Edit (free)  
  [http://komodoide.org/komodo-edit](http://komodoide.org/komodo-edit)

- Microsoft Excel

**Plugins**

- Dropbox
- Geolocating
- CSV import
- Simple content form
- Social bookmarking
- Doc viewer
• Have a server and hosting site (DreamHost)

**Plugins**
- Dropbox
- Geolocating
- CSV import
- Simple content form
- Social bookmarking
- Doc viewer
- Facebook comments

**Themes**

**Software used with Omeka**
- Cyberduck (free)
  [https://cyberduck.io/?fb=on](https://cyberduck.io/?fb=on)
- Global Mapper
  [www.globalmapper.com](http://www.globalmapper.com)
- Komodo Edit (free)
- Microsoft Excel
- Wordpress (free)
  [https://wordpress.com/website/](https://wordpress.com/website/)
Aberdeen Area History

http://aberdeenareahistory.org/
Outreach and Publicity

- Small community outreach
- Newspaper stories
- PBS Segment
- Radio shows
- Other outreach
What We've Done

Currently we have: 3 T of digitally born materials; 900 GB of scanned materials

- Scan as we go
  - no organization to scans
  - no naming systems

- No methodology to storing scans

- Periodically back up scans to external hard drives kept on site

- Transfer born digital materials with no integrity or virus checks/scans

- Have not been monitoring Fixity
Education

**Training courses:** SAA's DAS courses: Digital Curation, Digital Archives and Libraries, Digital Preservation, Digital Forensics, Thinking Digital, Standards of Digital Archives

**Conferences:** All Things Digital

**Listservs:** Archives and Archivists, College and University Section Discussion List, Metadata and Digital Object Roundtable, Digital Curation Google Group

**Readings:** *Digital Curation, Practical Digital Preservation*, Provenance Journal (Chris Prom's article "Making Digital Preservation Practical"

**Websites:** digitalpreservation.gov, digipres.org/tools/ubergrid, avpreserve.org
Preservica

Digital Preservation

"Preservica includes a comprehensive suite of OAIS (Open Archival Information System) compliant workflows for ingest, data management, storage, access, administration and preservation, as well as our new Universal Access module that allows you to safely share open content in your archive or collection within the public – making digital preservation a natural part of your information life-cycle."

Pricing: up to 100GB = $3,950, up to 250GB = $4,950, up to 500GB = $8,950, up to 1-10TB = $11,950 for the 1st TB then $1,950/TB in Amazon S3 or $750/TB in Amazon Glacier

Northern's cost: $14,950

Archivematica

"Archivematica is a free and open-source digital preservation system that is designed to maintain long-term access to digital memory. Archivematica is packaged with the web-based content management system AtOM for access to your digital objects."

Archivematica partnered with DuraSpace to offer ArchivesDirect-a hosted Archivematica instance with storage in DuraCloud.

Pricing: ArchivesDirect Standard: $11,900 + $1,000/TB
Northern’s Cost: ~$15,900

Ex Libris Rosetta

"Designed in collaboration with the National Library of New Zealand and reviewed by an international peer group of recognized leaders and innovators, Ex Libris Rosetta enables institutions to preserve and provide access to the collections in their care, now and in the future."

Pricing: Contact
Systems We've Looked At
“Preservica includes a comprehensive suite of OAIS (Open Archival Information System) compliant workflows for ingest, data management, storage, access, administration and preservation, as well as our new Universal Access module that allows you to safely share open content in your archive or collection with the public – making digital preservation a natural part of your information life-cycle.”

Pricing: up to 100GB = $3,950, up to 250GB = $6,950, up to 500GB = $8,950, up to 1-10TB = $11,950 for the 1st TB then $1,950/TB in Amazon S3 or $750/TB in Amazon Glacier

Northern's cost: $~14,950
“Archivematica is a free and open-source digital preservation system that is designed to maintain long-term access to digital memory. Archivematica is packaged with the web-based content management system AtoM for access to your digital objects.”

Archivematica partnered with DuraSpace to offer ArchivesDirect—a hosted Archivematica instance with storage in DuraCloud.

Pricing: ArchivesDirect standard: $11,900 + $1,000/TB
Northern's Cost: ~$15,900
"Designed in collaboration with the National Library of New Zealand and reviewed by an international peer group of recognized leaders and innovators, Ex Libris Rosetta enables institutions to preserve and provide access to the collections in their care, now and in the future."

Pricing: Contact
Software and Storage
(current and future)

Software
- CONTENTdm
- Adobe Suite
- OpenRefine
- FITS
- Bagit
- SobekCM
- METS Editor
- Fixity
- Windows Defender
- Malwarebytes

Storage:
- SAN Drive
- Hard drives
- Additional internal hard drive
- OCLC hosting of CONTENTdm
Metadata

- Decided to use OCLC's "Best Practices for CONTENTdm and Other OAI-PMH Compliant Repositories: Creating Sharable Metadata"

- Also use "Digital Library South Dakota Metadata Guidelines" for certain collections

- Controlled Vocabularies:
  - Format: AAT-Art and Architecture Thesaurus
  - Subject: TGN-Thesaurus of Geographic Names, TGM-Thesaurus of Graphic Materials, SAA Thesaurus for use in College and University Archives

- Use OpenRefine to fix any inconsistencies
Scans Workflow

• Scan to the network drive

• While scanning input metadata into excel document

• Combine files if necessary using Adobe Acrobat/crop or adjust files if necessary using Adobe Photoshop

• Once scanning and metadata are done, run metadata through OpenRefine

• Upload to CONTENTdm

• Select subjects within controlled vocab

• Run FITS on scans folder on network drive, output into file the program is running on
Scans Workflow

- Once successful upload to CDM, download metadata to a CSV file
- Run CSV file through SobekCM METS editor and create individual xml files of metadata
- Save the METS xml files to the corresponding scans folder
- Back up files to external hard drive
- Move all files (individual scans and compilations in applicable) with their metadata (METS and FITS) to their corresponding topical folder
- Create folder with final file name as folder title
- Run Fixity to monitor file integrity
- Save metadata excel and CSV files to collection metadata folder on external hard drive *

*eventually upload to the cloud
Born Digital Workflow

- Run Windows Defender and Malwarebytes on storage media
- Once the files are safe, use Bagit to either move/copy files to additional internal hard drive
- Rename the bag based on collection/title of the media and date of files (if known) or number of grouping within collection
- Create output file for FITS records within the bag
- Run FITS on the bag
- While running Bagit and/or FITS, write a brief descriptive summary of the files along with any information about the accession and save in the bag as description
- Back up all files to corresponding external hard drive *

*eventually upload to the cloud
Future Plans

- Cloud Storage
- Create formal born digital records station
- Upload more content to CONTENTdm
- OCLC Gateway and DPLA
- Utilize OHMS (Oral History Metadata Synchronizer)
- Assess the preservation needs of our digital video