

Building Digital Preservation Infrastructure: Partners, Tools and Services

Kevin Comerford & Karl Benedict

University of New Mexico Libraries

Antonio Guillermo Martínez Largo

Libnova

UNM Libraries Data Storage Improvements 2009-2010

- At first, completely decentralized data storage
- Finding funding
- Acquired centralized data storage 2010
 - Data replication of critical file stores
- Windows-accessible data storage
- Centralizing data storage – the “V:\” Drive
- File management protocols formalized
- Initiated routine backup services

First Forays into Preservation

2011-2015

- Content Assessment
- Generated Asset Inventories
- Functional Preservation Policies
 - Project- and Service-Based
- Data Clean-up & Harmonization
 - Format conversion
 - Rescanning
- End of Semester DP Backup
- Redundant Copies/Offsite Storage

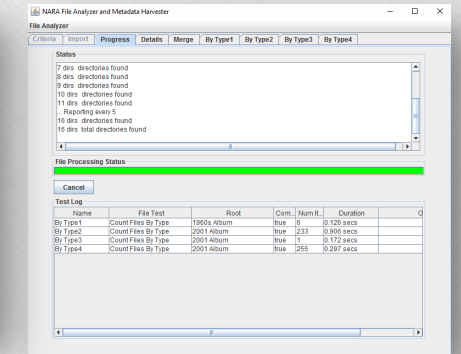
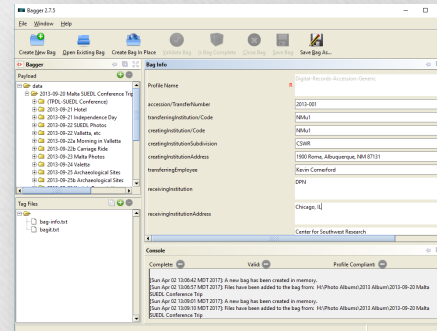
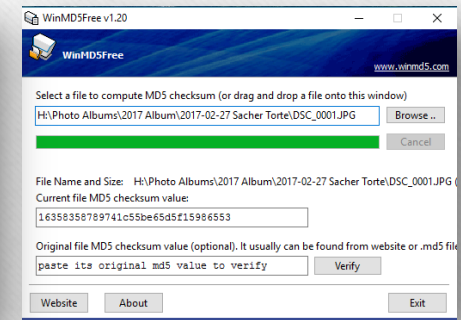
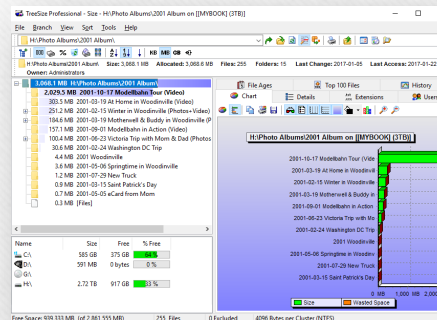


Digital Preservation Priorities

Collection	Priority	Rationale
Historic Pictorial High Resolution Scans	1	Replacement Cost & Time
Historic Manuscript High Resolution Scans	2	Replacement Cost & Time
Research Data Collections (Various)	3	Difficulty in Re-Acquiring – Value in participation
Institutional Repository	4	Difficulty in Re-Acquiring
Archives & Architecture High Resolution Scans	5	Already exist in multiple media

Trials and Tribulations: Preservation Tools and Utilities

- File Movers
 - *MS RichCopy*
 - *TeraCopy*
 - *RoboCopy*
- Reporting Tools
 - *TreeSize Professional*
 - *NARA File Analyzer*
- Fixity Checking Tools
 - *WinMD5*
- Packaging Tools
 - *LOC Bagger*
- Custom Scripts – Data extraction, file moving, reporting

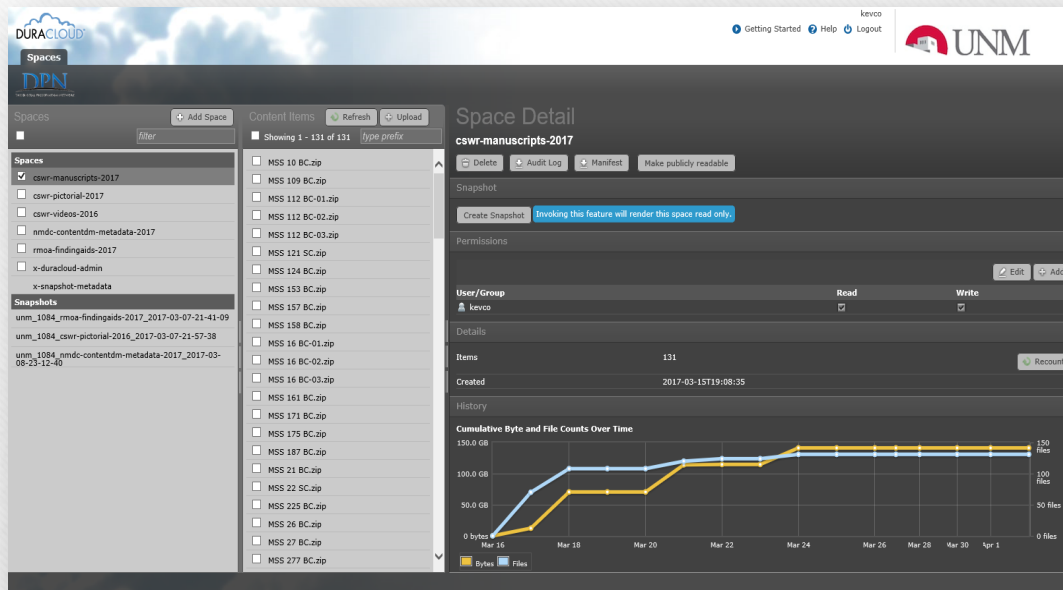


Preservation Metadata Sources

- Tapped Existing Data Sources
- Archival Finding Aids
 - RMOA Repository
 - XSLT Transforms
 - Collection to Item Data
- ContentDM Metadata
 - Simple Export Tools
 - Easy TSV & XLS extracts

The screenshot displays the Rocky Mountain Online Archive (RMOA) website. The top navigation bar includes links for Home, Browse, Advanced search, About, Help, and Contact. The main content area is titled 'Manuscripts' and 'PUBLISHED WORKS', listing various documents with their descriptions and container information (box and folder numbers). A detailed view of a 'Palenque corridor' item is shown, featuring a collection of photographs and a description. The description includes the text: 'A corridor in the Casa de las Leyes at the archaeological site of Palenque, Chiapas, Mexico. Caption reads: Fotografía del corredor donde se...'. The website also includes a section for 'About this collection' with the UNM logo and a brief overview of the collection's contents.

Digital Preservation Network 2016



- Using DuraCloud UI/ Sync Tool to ingest content
- LOC Bagger to package collections of objects and metadata
- Starting with Pictorial and Manuscript scans

Libnova Libsafe DPS 2017

The screenshot displays the Libsafe DPS web interface. The top navigation bar includes links for Home, Catalog, Ingestion, Retrieve, Active preservation, Documentation site, and Advanced options. The user is identified as Kevin Comerford, with links for Notifications panel and Logout.

Storage and capacity:
System capacity: 10TB
Free space: 98%
Preserved objects: 246

Current running jobs:
Ingestion: 1 job
Retrieve: 0 jobs
Audit: 0 jobs
Evolution: 0 jobs
Digital signatures: 0 jobs

Last system activity:
Ingestion:
2017-04-03 - ING0000020
2017-04-03 - ING0000021
2017-03-28 - ING0000019
Retrieve:
2017-03-14 - RET0000003
2017-03-14 - RET0000002
2017-03-14 - RET0000001

Active tasks:
Ingestion Task 20 [View](#)
[View more](#)

Catalog - Browse by tree:
The University of New Mexico
Pictorial
NOT DESCRIBED
Manuscripts
Research Data

Objects preserved into the catalog (246 objects):

Object name	LUOI	Ingestion	Preservation area	Preservation plan	Created on	Thumbnail	View
<input type="checkbox"/> PICT 000-561	19.42	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-560	19.41	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-556	19.40	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-546	19.39	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-542	19.38	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-540	19.37	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View
<input type="checkbox"/> PICT 000-535	19.36	ING0000019	Pictorial Collection	Pictorial.01	2017-03-28 23:36:02		View

- Hosted Solution
- Azure Cloud Storage
- Ingest Metadata with Digital Objects
- Mirroring DPN archive with Pictorial and Manuscript scans
- Looking forward to DPN/Libsafe collaboration

Digital Preservation 3-Year Schedule*

Collection	DPN	LIBSAFE
Pictorial High Res Scans (4.6 TB)	2016	2017
Manuscript High Res Scans (6.2 TB)	2017	2017
Research Datasets (2 TB)	2017	2017
Institutional Repository (1.5 TB)	2017	2017
Archives & Architecture High Res Scans (2 TB)	2018	2018
Other Special Collections (2 TB)	2018	2018
Statewide Image Collections (2 TB)	2018	2018
Other UNM Agencies (5TB+)	2019	2019
<i>* Includes estimates for incremental additions from all collections</i>		

Digital Preservation Archive Growth Estimates

4-Year Archive Growth Estimate

