

Preserving Digital Content at Scale

CNI
Albuquerque, NM
April 4, 2017

Carolyn Caizzi
Head of Repository and Digital Curation
Northwestern University Libraries

Northwestern



State of the Northwestern Libraries' Local Digital Content

Data Storage

Isilon EMC2 at 2 data centers
Daily syncs & 45-day snapshots
Tape backups
Different geographic region

How Much

Date	Actual TB	Increase TB	Percentage Increase from previous year
9/2010	2	2	100%
9/2011	21	19	950%
9/2012	38	17	81%
9/2013	75	37	97%
9/2014	125	50	67%
9/2015	200	75	60%
9/2016	267	67	34%

What makes up the 267 TB

- Preservation and Mezzanine Master Files managed by Fedora repository: **52 TB**
- Streaming Derivatives A+V Repository: **19 TB**
- Digitization & Born Digital Project Files Manually Managed: **172 TB**
- Misc. and Outdated Content: **24 TB**

What makes up the 267 TB

- Preservation and Mezzanine Master Files managed by Fedora repository: **52 TB**
- Streaming Derivatives A+V Repository: **19 TB**
- **Digitization & Born Digital Project Files Manually Managed: 172 TB**
- Misc. and Outdated Content: **24 TB**

Repository (ies)

Active Repository(ies)

- Images Repository
- A+V Repository
- Self-Deposit IR “Arch”
- Books Repository
- Output of digitization & born digital projects

Inactive Repositories and Websites

- Special collections websites
- Faculty project websites
- Books repository

Digital Preservation Scorecard

Active Preservation

NDSA Levels	Storage & Geo Location	File Fixity & Data Integrity	Info Security	Metadata	File Formats
Protect	B	F	A	A	A
Know	A	F	C	F	A
Monitor	C	F	F	A	C
Repair	F	F	F	F	C

C Average

NDSA Levels	Storage & Geo Location	File Fixity & Data Integrity	Info Security	Metadata	File Formats
Protect	B	F	A	A	A
Know	A	F	C	F	B
Monitor	C	F	F	A	C
Repair	F	F	F	F	C

Tipping Point

Next Generation

Before and Now

Before: Staffing and Structure

2 Departments

- **Digital Collections:** production staff; project managers; product owners
- **Enterprise Systems:** developers; sysadmins; devops; tech support
- **Missing:** metadata coordinator for digital content; preservation focused positions

Now: Staffing and Structure

Library Wide Re-Organization 2014-15

- 2 new AULs; 1 new department head
- 1 department: Repository & Digital Curation
- Digital preservation & digital curation librarians; 6 developers with 1 focused on digital preservation; metadata coordinator

Better Management

- Published a library digital preservation policy
- Created an inventory of our Isilon shares
- Created plans for migration of legacy content

Before: Repository Architecture

- Focus on showing content, not on management or digital preservation actions
- Based on format or for a specific collection
- Added on functionality through time

Now: Repository Architecture

Current project to **re-imagine** our repository

- Hyrax/Fedora 4 with focus on digital preservation & staff management tools
- De-couple front-end from staff tools
- Bring together objects of all formats
- Focus on batch import and batch editing
- Multiple tiers for storage, including sending most valued content to DPN

Now: IMLS Planning Grant

Lg-72-16-0135-16

- How does one curate objects to ingest into a long-term dark preservation system?
- How does versioning of objects and metadata play out in a long-term dark preservation systems and how to automate these actions?
- How can systems that store data differently be made more interoperable?

Now: Enterprise Level Digital Preservation Strategy

- Consultant engagement: digital preservation assessment of Northwestern
- Recommendations will be presented to the Provost this spring
- Recommendations include: firm organizational structure; infrastructure; education of PIs, faculty, and graduate students

Summary: Planning for More Active Preservation

- +Strategic re-imagining of vacancies
- +Review policies: hit publish
- +Storage solutions: tier
- +Repository architecture: staff tools & decoupled front end design
- +Workflows: start fresh & design with batch in mind

Thank You!

carolyn.caizzi@northwestern.edu