



**Migrating from OCLC's Digital Archive
to DuraCloud**

State Library of North Carolina

- Part of the North Carolina Department of Cultural Resources
- Work closely/pool resources with the State Archives
- Digital Information Management Program

CONTENT

State Publications

Genealogy Research
North Carolina

STAFF

~ 4.75 FTE

CONTENTdm

Connexion Digital Import

SYSTEMS

Local server

(state-supported)

Offsite storage

(vendor)

STORAGE

Born-Digital 3.25

Digitized .75

CONTENTdm
Connexion

CONTENTdm
Project Client

Local Storage

Remote Storage

.75

CONTENT

- We preserve access and master copies
- 1.27 TB, 162,000+ files
- Mostly .tif, .pdf, .jpg, .txt

File Types

File Types	Gigabytes	Files
image/tiff	960.310	36063
application/pdf	150.998	29401
image/jpeg	10.815	22377
text/plain	1.806	6443
application/octet-stream	35.794	5010
image/png	0.553	547
application/xml	0.039	182
image/gif	0.000	59
application/postscript	0.211	15
text/html	0.001	15
application/rtf	0.003	2
audio/x-wav	0.550	1
video/mpeg	0.457	1

CONTENT

File structure by “project”

admindocs

fulltext

images_access

images_master

images_processed

metadata

Naming convention

pubs_serial_annualreportclean2005.pdf

gen_statefair_lifcharacterthomasruffin1871_0001.tif

Local storage

- managed by department-wide IT
- includes working & preservation content
- server is shared, but our directory is restricted
- daily incremental backups

STORAGE

OCLC's Digital Archive

- Began using in 2008
- Web interface for access
- FTP or automatic uploads
- Integrated with CONTENTdm
- Detailed reporting, broken out by CONTENTdm collection
- Fixity checks, virus checks

Monthly Activity & Storage Report

Report Date: 2011-01-01

Reporting Period: 2010-12-01 through 2011-01-01

Institution Name: STATE LIBR OF NORTH CAROLINA

Institution Symbol: NCS

Content Integrity Summary

Number of Files in the Archive: 882551 files

Number of Files with viruses: 0 files (exceptions below)

Number of Files with Valid Checksums: 882551 files (exceptions below)

Content Storage Summary

Storage in Use - Beginning of the Month: 334.027 GB

Storage Encumbered this Month: 63.757 MB

Storage in Use - End of the Month: 334.089 GB

Storage Purchased: 16000 GB

Number of Files in the Archive - Beginning of the Month: 882394 files

Number of Files Added this Month: 157 files

Archive Accesses

Number of Online Accesses this Month: 34 (details available as a separate report upon request)

Number of Bulk Disseminations: 0 (details below)

Bulk Dissemination Details

Dissemination Unit	ID (coll name or vol ID)	File Count	Volume
No bulk disseminations reported this period			

File Integrity Details

Collection Alias	Volume ID	File Name	Fixity Status	Date Last Fixity Check	Virus Check Status	Date Last Virus Check
No exceptions reported this period						

File Type Frequency

File Type	Occurrences	Percentage
Institution-wide (882558 Files)		
text/plain\0118bit	2	0.000226614
image/tiff	4903	0.5555442

Archive Accession Report

Report Date: 2011-01-06

Institution Name: STATE LIBR OF NORTH CAROLINA

Institution Symbol: NCS

Accession Summary

Collection Name: p249901coll22

Collection ID: p249901coll22

Archival Volume ID: 1294329577464

Accession Date: 2011-01-06

Number of Files Recieved: 10492

Number of Files in Manifest: 1

Number of Viruses Suspect: 0

Number of File Outside Archival Volume: 10491

Archival Volume Storage Size: 429.482 KB

File Details

File Name	File Size	In Manifest	Fixity Status	Virus Status	MIME Type
pubs_parentreportabout2010.pdf	429.482 KB	yes	OK	OK	application/octet-stream

Warning: File extension does not match any of the expected extensions for this MIME type.

INSTITUTION=NCS

SERVERURL=http://server16062.contentdm.oclc.org/

COLLECTION=/p249901col122

VOLUMEID=pubs_ag_20120604

ITEMLIST=admindocs/pubs_ag_20120604_hash.txt

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1887.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1903.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1904.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1906.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1908.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1909.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1912.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1915.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1916.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1917.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1918.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1919.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1920.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1921.pdf

ITEMLIST=images_access/pubs_ag_aragriculturalalexperiment1922.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1887.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1903.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1904.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1906.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1908.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1909.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1912.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1915.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1916.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1917.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1918.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1919.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1920.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1921.pdf

ITEMLIST=images_master/pubs_ag_aragriculturalalexperiment1922.pdf

ITEMLIST=metadata/pubs_ag_aragriculturalalexperiment_metadata.txt

ITEMTOTAL=32



- Integration with CONTENTdm
- Fixity checks and virus scans
- Responsive support
- Extensive reports



- Integration with CONTENTdm
- Finding and retrieving items
- Manifest/batch upload requirement
- Vendor-side error reporting
- Verifying storage contents

DuraSpace's



- Began using in 2012
- Web interface for access
- Web interface or client-side tools for upload
- Content Management System-agnostic
- Fixity checks

- Spaces + Add Space
-
- arc-geomapp
 - arc-testing
 - lib
 - lib-nao
 - tst-test
 - x-duracloud-admin
 - x-service-out

Content Items Refresh + Add Items

Showing 1 - 200 of ? show more> StatePubs

- StatePubs/metadata/pubs_metadataparentlevel_20120313.txt
- StatePubs/metadata/pubs_metadataparentlevel_20120313.xlsx
- StatePubs/metadata/pubs_metadataparentlevel_20120731.txt
- StatePubs/metadata/pubs_metadataparentlevel_20120731.xlsx
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt01.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt02.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt03.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt04.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt05.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt06.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt07.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt08.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt09.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt10.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt11.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt12.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt13.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt14.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt15.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt16.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt17.txt
- StatePubs/metadata/pubs_txtmetadatapagelevel_20120801_pt18.txt

Space Detail

lib

Delete Space Make space publicly readable Streaming: On Off

Permissions

Edit Add

User/Group	Read	Write
library	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Igregory	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Details

Items: 100134 Recount

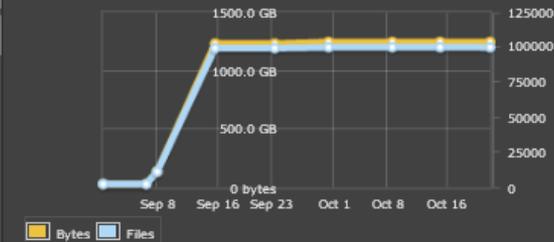
Created: Tue, 28 Aug 2012 16:24:18 UTC

Size:

Last Health Check: Sat Sep 01 02:45:58 UTC 2012 - success [Report](#)

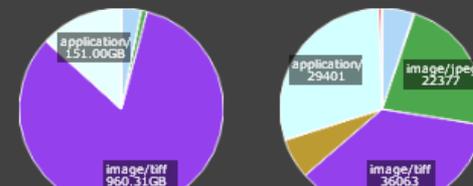
History

Byte and File Counts Over Time



Mime Types as of Oct 21, 2012

Data



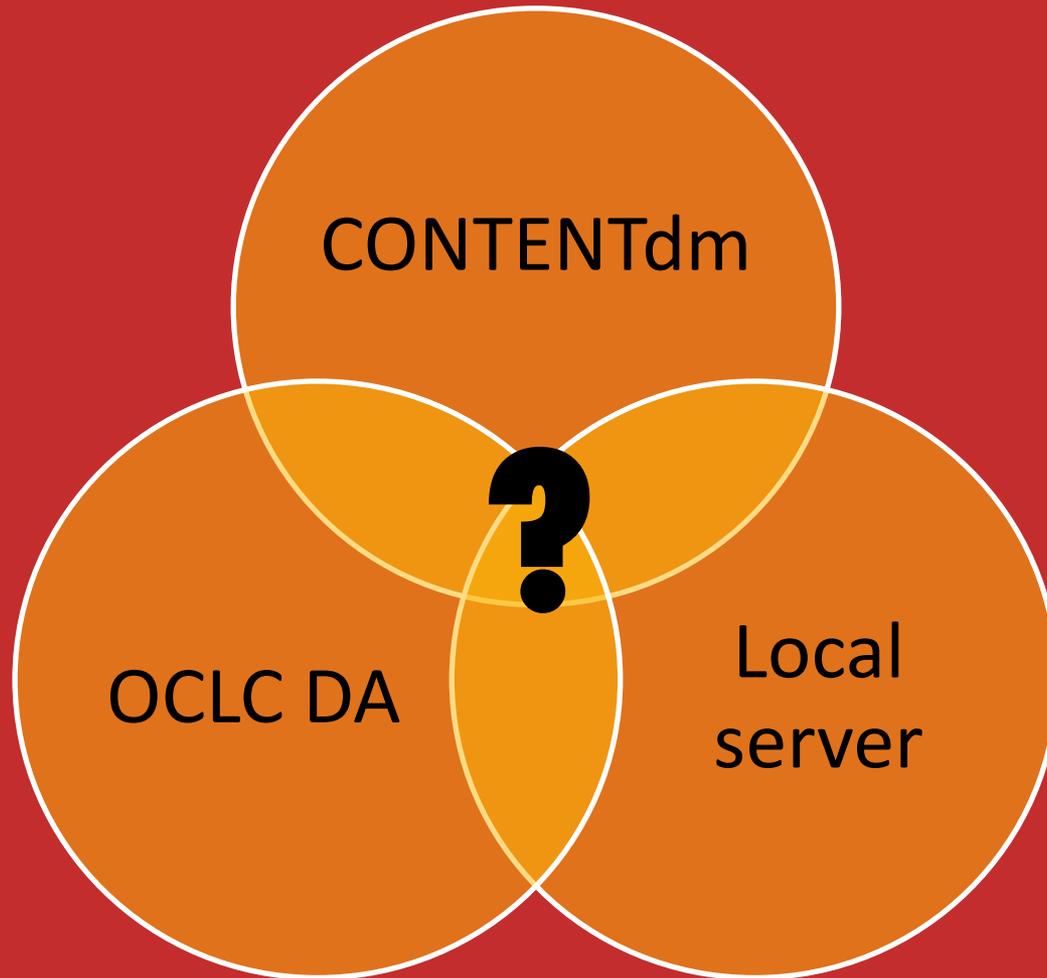


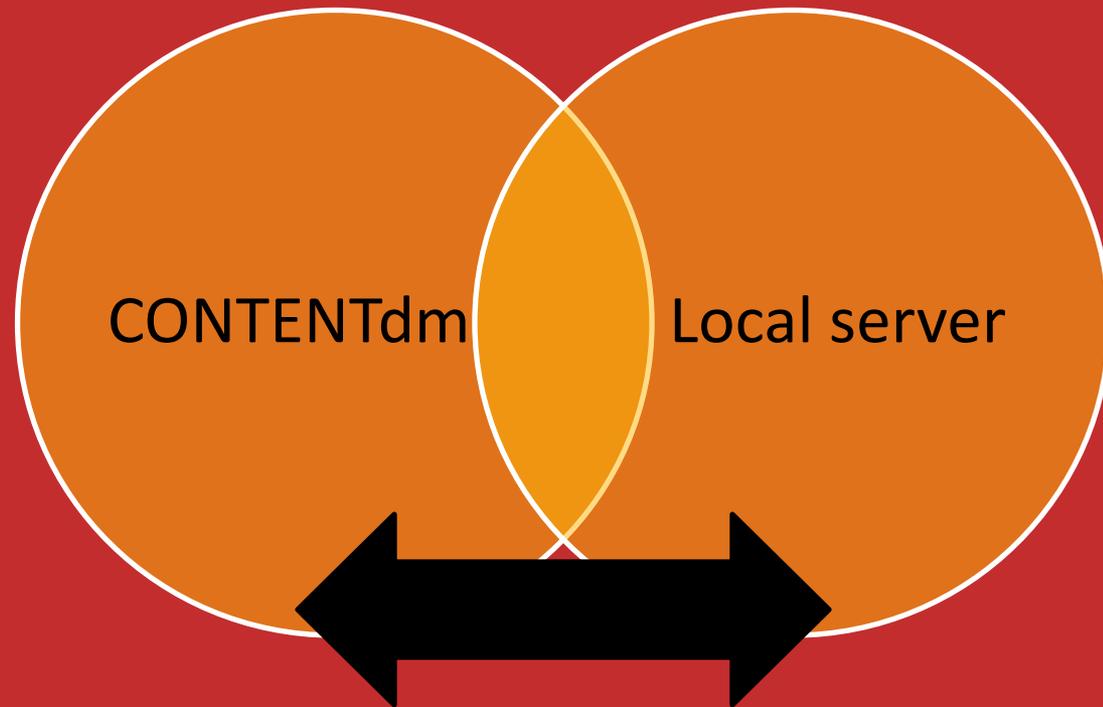
- Presentation is like a traditional gui file manager
- Can designate spaces, permissions
- Can make a space public
- Powerful upload tools
- Fixity scans
- Robust reporting
- Easy to get content out
- Choice of storage services
- VERY collaborative support
- Non-profit



- Searching
- Sorting
- Verifying storage contents
- Overwriting isn't hard to do
- Batch delete
- MD5

PREPARATION





1. Exported metadata from CONTENTdm
2. Exported file names from local server
3. Bashed preservation file names, checksums
4. Identified and recovered missing files

- | | |
|--|-------------------------|
| 1. Exported metadata from CONTENTdm | Onerous to impossible |
| 2. Exported file names from local server | Easy |
| 3. Bashed preservation file names, checksums | Easy but time consuming |
| 4. Identified and recovered missing files | Easy-ish |

1. Exported metadata from CONTENTdm

Onerous to
impossible

- OCLC had to provide export for largest & most critical collection
- 363 MB tar file -> 18 x 100+ MB csv files
- Added frustration: metadata for compound objects v. multi-page pdfs

2. Exported file names
from local server

Easy

1. Bashed preservation
file names, checksums

Easy but time
consuming

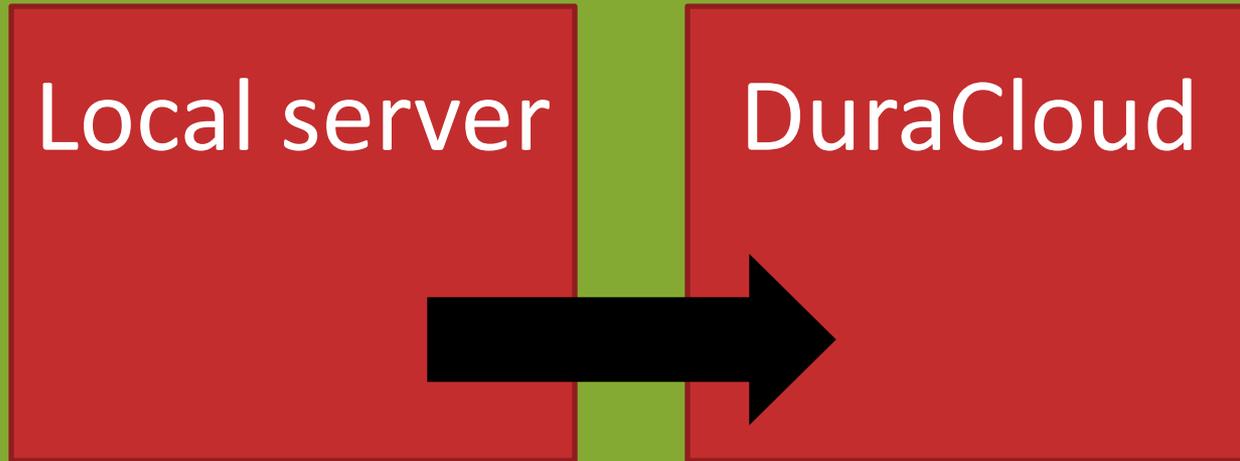
- Spreadsheet gymnastics
- Manual review for filename/checksum inconsistencies

4. Identified and recovered missing files

Easy-ish

- Missing from CONTENTdm? Added by librarians
- Missing from local server? Request to OCLC or re-download from CONTENTdm

THE MOVE



1. Tested sync and upload tools
2. Discussed spaces
3. Ran sync tool on local preservation storage
4. Ongoing maintenance: upload tool

1. Tested sync and upload tools

Easy

- Helped determine flags to manage computer resources during sync
- Verified logging output, permissions
- Helped flesh out local workflow

2. Discussed spaces

Easy, and
Interesting

- Many spaces or few, to accommodate different workflows?
- Assignment of permissions

3. Ran sync tool on local preservation storage

Easy

- Ran continuously for 5 2/3 days
- 94,177 items

4. Ongoing maintenance: upload tool Easy-ish

- Uploads done weekly and monthly
- Upload tool used to avoid accidental overwriting
- Have to create “mock” file structure

**Working
directory**

**Working
directory**

**Working
directory**

Staging

Staging – Limited Access

Local server

DuraCloud

Insights

- Room for preservation metadata improvement
 - Working with full metadata dumps is problematic
 - Need for more automated monitoring for local storage
 - Integration with CMS not helpful unless FULL integration
- in other words:
- Streamlined ingest = streamlined preservation

Still more thoughts

- No, really: manual management and auditing is getting less feasible
- What is acceptable content loss?
- What is acceptable preservation metadata error rate?
- Responsiveness to enhancement requests should be figured into vendor choice
- At 5 years out, PREMIS lite is just fine



Lisa Gregory
lisa.gregory@ncdcr.gov