



ELBOW GREASE AND FLEXIBILITY:


PRACTICAL INSIGHTS FROM MIGRATIONS BIG AND SMALL

@mac_omaha2017 | #omamac17 | #s402


SETTING THE STAGE...




Laura Alagna: @Digitized_Laura

 Optimized for mobile: Migrating data extracted from mobile devices to preservation formats

Stephanie Bricking

 Migration for Access: Lessons Learned from Moving Wiki Collections into CONTENTdm

Sara Dorpinghaus: @SMDorpinghaus

 Tiered: UK Libraries' Approach to Migrating Born-Digital Content

Lori Myers-Steele

 Elbow Grease and Flexibility: Just What Our Multifaceted Migration Project Needed

Dan Noonan: @DannyNoonan1962

 MOM: MASTER OBJECTS MIGRATION PLANNING AT TOSU

Q&A

@mac_omaha2017 | #omamac17 | #s402

Good Afternoon and welcome to: “Elbow Grease and Flexibility: Practical Insights from Migrations Big & Small!” Ever since Sarah suggested that title for our session, I felt we should be up here in coveralls with our names on them. As our abstract surmised: Some institutions are at the point of migrating to their second (or maybe third or more) digital preservation environment. Some may be migrating their digital objects to a collection management system for the first time, while others are attempting to tackle migrating content off of obsolete media. Some may be a combination of all the above, depending upon the type of digital materials involved.

Today we will roll up our sleeves, to get our hands dirty discussing the migration our digital materials from multiple points-of-view. For simplicity sakes we decided to proceed in an alphabetical manner today.

Laura Alagna: Laura Alagna the Digital Preservation Librarian at Northwestern University Libraries will present. “Optimized for mobile: migrating data extracted from mobile devices to preservation formats.” As cloud storage has become nearly ubiquitous, it’s tempting to think that obsolete media are a thing of the past. However, significant amounts of digital records are still stored on “less obsolete” media - mobile devices and tablets are the new floppy disks. Her presentation will describe the trials and

tribulations of extracting data from mobile devices and migrating it to preservation-friendly formats, as well as thoughts on future challenges in archiving mobile data.

Laura has developed and implemented policies and workflows for managing digital content, and currently works on integrating digital preservation into Northwestern's "next generation" repositories. She is an active member of the Chicago Area Archivists and the BitCurator Consortium, and is an appointed member of the Illinois State Historical Records Advisory Board.

Stephanie Bricking: Stephanie Bricking the Metadata Librarian at the Public Library of Cincinnati & Hamilton County, in Cincinnati, Ohio will present, "Migration for Access: Lessons Learned from Moving Wiki Collections into CONTENTdm." In November 2014, the Public Library of Cincinnati and Hamilton County acquired CONTENTdm, in order to make their digitized materials more accessible. The content was originally displayed in a home-grown CMS and a wiki. Her presentation will primarily focus on the migration from the wiki platform, which was completed a year ago this month. The lessons they learned from migrating content from both the home-grown CMS and the wiki can be applicable to any migration, regardless of the platform.

Prior to joining the public library, Stephanie was the Albert B. Sabin project archivist at the University of Cincinnati.

Sarah Dorpinghaus: Sarah Dorpinghaus, the Director of Digital Services at the University of Kentucky Libraries Special Collections Research Center, will present "Tiered: UK Libraries' Approach to Migrating Born-Digital Content." The University of Kentucky Libraries Special Collections Research Center recently set up a workstation running BitCurator and other tools to migrate content off 3.5" diskettes, zip disks, CDs, DVDs, and more. After a pilot project focused on already-processed collections, they re-evaluated their workflows and created a tiered approach to migrating and stabilizing born-digital content. Her presentation will summarize the insight gained from the pilot project and review their new and more efficient approach to migrating born-digital content.

Sarah, as Director of Digital Services, develops long-term strategies and oversees workflows for digital content creation, management, and preservation. Under the invaluable tutelage of several MAC members, Sarah earned her MLIS from the University of Iowa and held project archivist positions at the College of Charleston and the Chicago History Museum before

landing in Kentucky.

Lori Myers-Steele: Lori Myers-Steele the Collections Archivist at Berea College, will present “Elbow Grease and Flexibility: Just What Our Multifaceted Migration Project Needed,” which will provide practical insight from her department’s efforts in centralizing their collection and digital asset management activities from an array of systems and schemes, into an environment constructed around Archon and Preservica. They are thirty-three months into a thirty-six month grant funded project aimed at gaining intellectual control over their collections and providing better and increased access to such collections, this project has taken many unforeseen twists and turns while producing unanticipated outcomes.

Originally from Texas, Lori received her MA from the National University of Ireland, Galway, before landing in Berea and obtaining a graduate certificate in Archival Studies from East Tennessee State University.

Dan Noonan: Finally, I am Dan Noonan, an assistant professor and the Digital Resources Archivist at THE Ohio State University. My presentation, “MOM: Master Objects Migration Planning at tOSU,” will delve into our efforts to migrate (or not) nearly 2 million digital files--from 8 collecting units with minimum, scattered or unknown metadata—from a pseudo-DarkArchive to a robust “light archive” constructed on a FEDORA/Hydra preservation and access environment.

We have planned to have sufficient time at the end of this session for Q&A, so please hold your questions until then of all the panel.

Please note if you are not already following this year’s conference on Twitter, the handle is @mac_omaha2017 and when tweeting the conference hashtag is #moamac17 and this session is #s402.

Thank you and let’s get our hands dirty!



OPTIMIZED FOR MOBILE: MIGRATING DATA EXTRACTED FROM MOBILE DEVICES TO PRESERVATION FORMATS

LAURA ALAGNA
@DIGITIZED_LAURA

@mac_omaha2017 | #omamac17 | #5402

Optimized for Mobile:

Migrating data extracted from mobile devices to preservation formats

Laura Alagna

Digital Preservation Librarian
Northwestern University Libraries

 @Digitized_Laura



@mac_omaha2017 | #omamac17 | #s402

First, some background...

We jumped in!



[About Archives](#) • [About SAA](#) • [Careers](#) • [Education](#) • [Publications](#) •

[HOME](#) » [GROUPS](#) » [MANUSCRIPT REPOSITORIES SECTION](#) » [JUMP IN INITIATIVE](#)

Jump In Initiative

[Roster](#) | [Description](#) | [Steering Committee Roster](#) | [Join/Leave](#)

Jump In

2013 Results are in!

The steering committee of the Society of American Archivists' Manuscript Repositories Section introduces a new initiative encouraging archivists and their institutions to *Jump In* to managing born-digital content.

[@Digitized_Laura](#) | [@mac_omaha2017](#) | [#omamac17](#) | [#s402](#)

An interesting development...

395 CD-ROM

266 DVD

106 floppy disks (all types)

29 DAT

7 USB flash drives

1 external hard drive

→ 1 mobile phone



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Are mobile devices



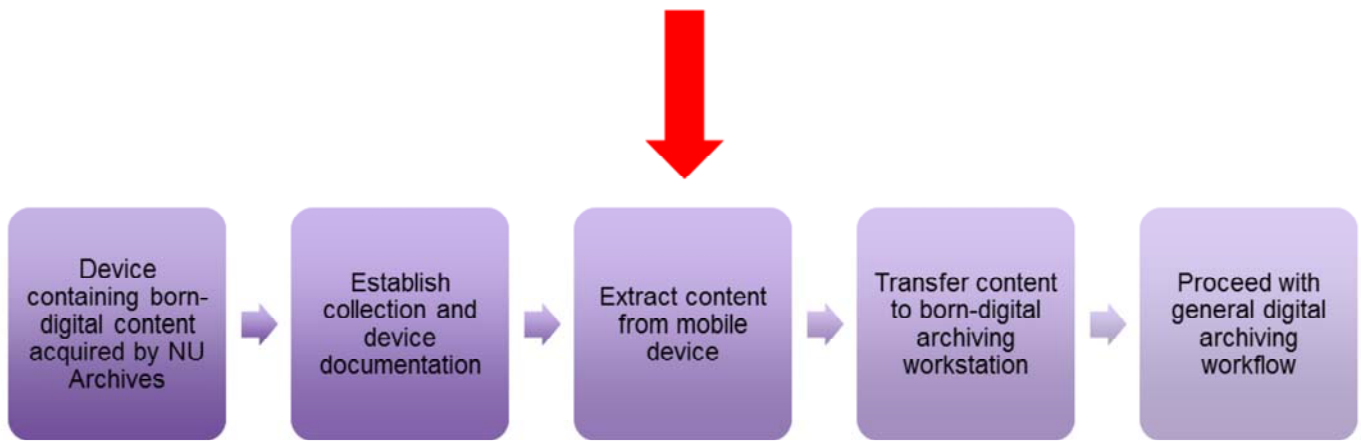
artifacts or media?

Getting to work

Establishing a process for archiving content from mobile devices

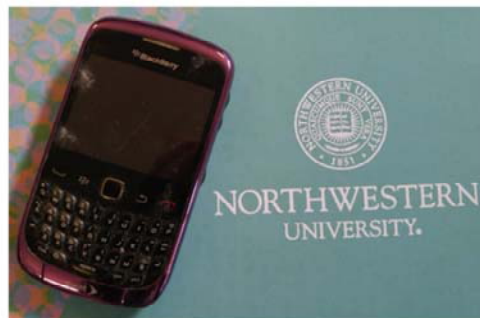


@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402



“Extract content from mobile device”

1. The device has a dead battery



“Extract content from mobile device”

1. The device has a dead battery
2. How to copy data from it?



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

“Extract content from mobile device”

1. The device has a dead battery
2. How to copy data from it?
3. How to turn that into a preservation-friendly format?



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

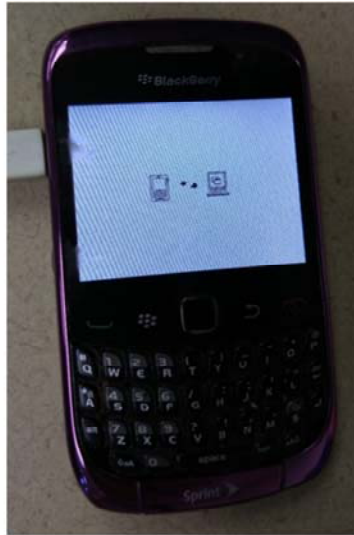
Problems within problems

Problem 1: dead battery



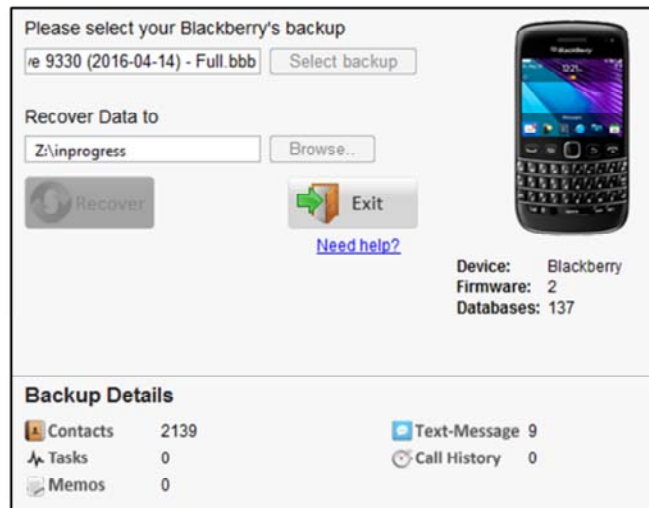
@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Problem 2: how to copy the data



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Problem 3: preservation for mobile data



The screenshot shows a software interface for recovering BlackBerry data. It includes a 'Please select your BlackBerry's backup' section with a dropdown menu showing 're 9330 (2016-04-14) - Full.bbb' and a 'Select backup' button. Below this is a 'Recover Data to' section with a text input field containing 'Z:\inprogress' and a 'Browse...' button. There are 'Recover' and 'Exit' buttons, along with a 'Need help?' link. A small image of a BlackBerry phone is shown on the right. At the bottom right, device information is listed: 'Device: Blackberry', 'Firmware: 2', and 'Databases: 137'. A 'Backup Details' section at the bottom lists: 'Contacts: 2139', 'Text-Message: 9', 'Tasks: 0', and 'Call History: 0'. 'Memos' is listed with a count of 0 but no icon is visible.

Please select your BlackBerry's backup
re 9330 (2016-04-14) - Full.bbb Select backup

Recover Data to
Z:\inprogress Browse..

Recover Exit
[Need help?](#)

Device: Blackberry
Firmware: 2
Databases: 137

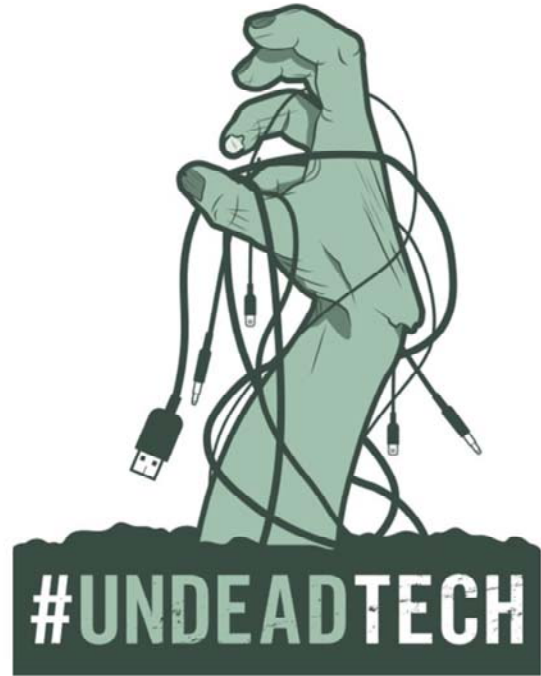
Backup Details

Contacts	2139	Text-Message	9
Tasks	0	Call History	0
Memos	0		

@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Time to get creative

Solving the dead battery and connection problems



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

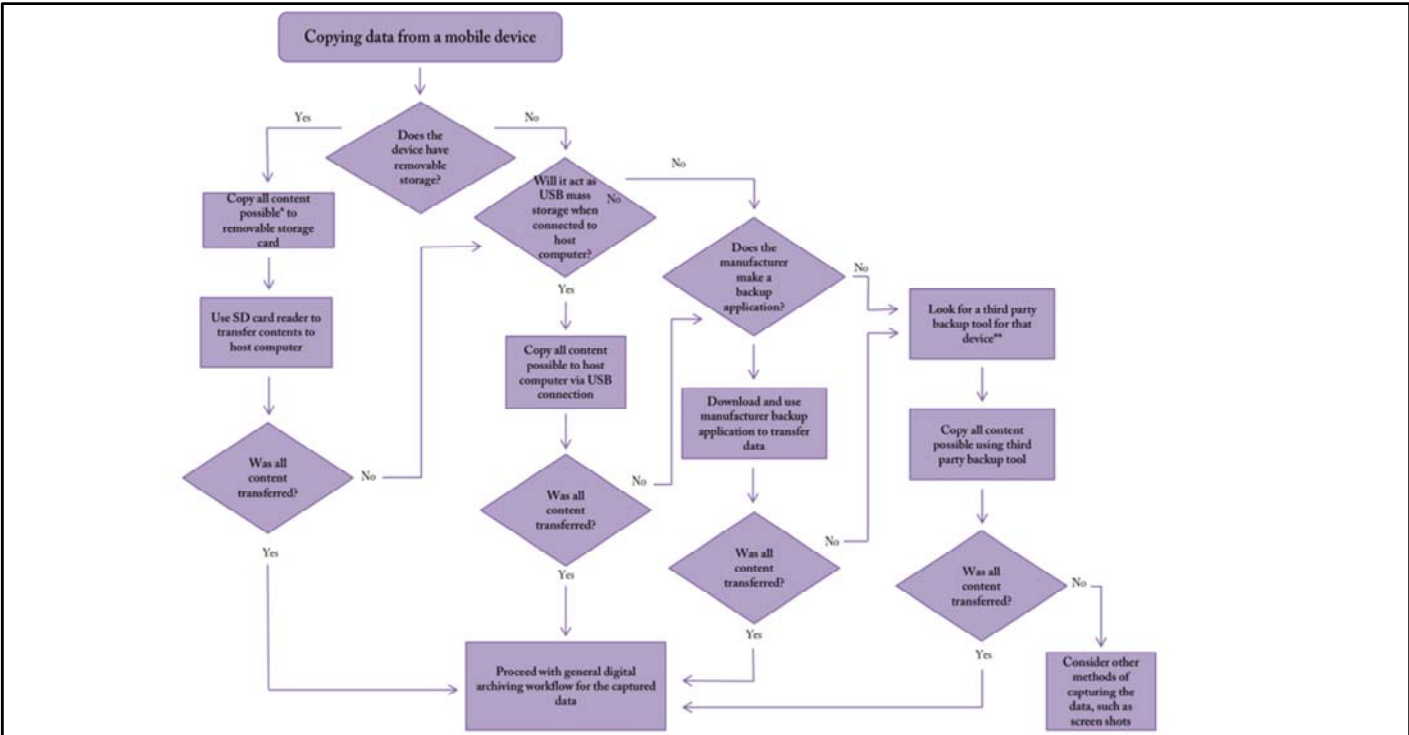


@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Device-specific dependencies



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402



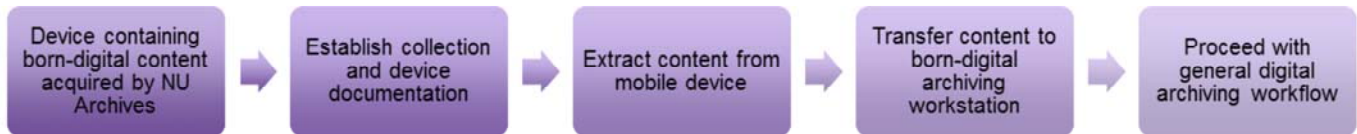
Alternatives

- Screenshots
- Photos
- Transcription



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Process for archiving content from mobile devices



Final thoughts



On mobile devices as media for archival records



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

On the future of digital preservation for mobile content



@Digitized_Laura | @mac_omaha2017 | #omamac17 | #s402

Thank you!



Laura Alagna

Digital Preservation Librarian

Northwestern University Libraries

laura.alagna@northwestern.edu

[@Digitized_Laura](https://twitter.com/Digitized_Laura)

@mac_omaha2017 | #omamac17 | #s402



MIGRATION FOR ACCESS: LESSONS LEARNED FROM MOVING WIKI COLLECTIONS INTO CONTENTDM

STEPHANIE BRICKING

@mac_omaha2017 | #omamac17 | #s402



THE
PUBLIC
LIBRARY
of Cincinnati
and
Hamilton County

Migration for Access :

Lessons Learned from Moving Wiki Collections into
CONTENTdm

Stephanie Bricking, Metadata Librarian
@mac_omaha2017 | #omamac17 | #s402

The Public Library of Cincinnati & Hamilton County
Connecting people with the world of ideas and information.

Background Information

- Digital Services department formed in November 2007
- Equipment purchased with the support of multiple LSTA grants administered by the State Library of Ohio
- Digitized items displayed in:
 - Virtual Library (2007-2008)
 - Wiki collections (Summer 2007)



@mac_omaha2017 | #omamac17 | #s402

Sample Item in Virtual Library

THE PUBLIC LIBRARY
of Cincinnati
and
Hamilton County

Virtual Library

Virtual Library Collection | Entire Library Collection

Search

Homepage : Old & Rare Books : 900 - History, Genealogy & Geography : The Ohio railroad guide, illustrated : Cincinnati to Erie via Colu...

Find Downloadable Old and Rare Books:

Title: **The Ohio railroad guide, illustrated : Cincinnati to Erie via Columbus and Cleveland.**
Call Number: 917.71 O36
Publication Date: 1854

PDF (24 MB)
PDF Extra (28 MB)

@mac_omaha2017 | #omamac17 | #s402

Sample Photos in the Wiki


Page: Discussion Read · View source · View history

Delta Queen


ABCDEFGHIJKLMNOPQRSTUVWXYZ

• From the Way Collection


Boat Information
Name: Delta Queen
Origin: Cincinnati, OH
Years Active: 1925-



Delta Queen on a river with people on the bank.



Delta Queen's pilothouse with a large steering wheel.



Delta Queen's interior staircase.

• Neville Island, Pa., March, 1948. Pilothouse of the Delta Queen.

@mac_omaha2017 | #omamac17 | #s402

Why Migrate?

- Needed to provide:
 - Better access to collections
 - Better searching and browsing capabilities
- Needed to bring collections together
- Needed ability to expose metadata through OAI-PMH



@mac_omaha2017 | #omamac17 | #s402

Lessons Learned during Preparation


1. Don't try to reinvent the wheel.
2. Take the time to establish an initial workflow.
3. Use what's available in your library.
4. Don't be afraid to jump right in!



@mac_omaha2017 | #omamac17 | #s402

Use What's Available

Title Views of points on the Ohio River during the flood of 1884.			
Publication Info. [Cincinnati? : L.C. Goodale, 1884?]			
Call No. 551.57 q1884			
Rating ☆☆☆☆☆			
Copies	Location	Call No.	Note
	Main - 3rd Floor - Genealogy & Local History - Cincinnati Rm-River	551.57 q1884	



Views of Points on the Ohio River during the Flood of 1884.
Photographs of these places were secured by Levi C. Goodale at the time, later he had these photogravures made from the photographs.

Description [48] leaves of plates : all ill. ; 17 x 26 cm.

Note Caption title.
"Photographs of these places were secured by Levi C. Goodale at the time, later he had these photogravures made from the photographs"--First leaf.

Summary Photogravures, mounted on thin paper and bound by cloth hinges in black cloth. Title and captions beneath each photograph are printed slips which are glued to the pages. Includes scenes in Ohio, Indiana, Kentucky, and West Virginia.

Note Copy in Genealogy & Local History Dept., Cincinnati Collection bound in black cloth. OC

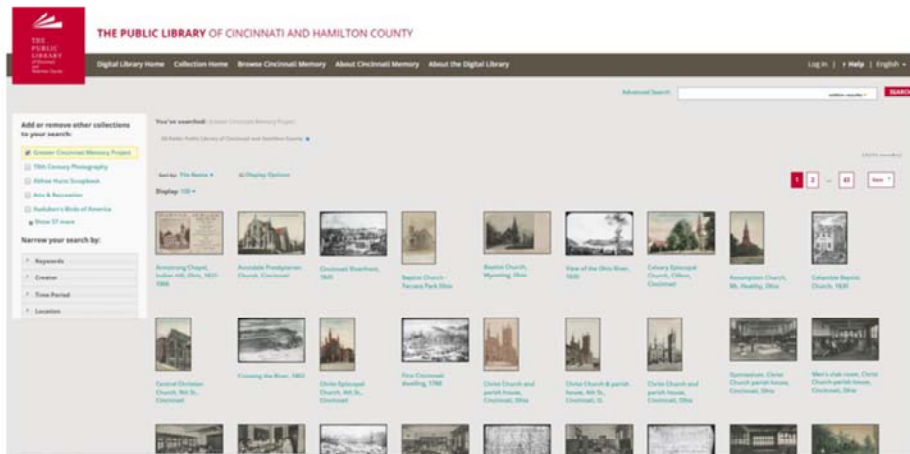
Subject Floods -- Ohio River -- Pictorial works.

Local Subject Imprints (in rare books) -- Ohio -- Cincinnati -- 1884.

Added Author Goodale, Levi C.

@mac_omaha2017 | #omamac17 | #s402

Jump Right In



@mac_omaha2017 | #omamac17 | #s402

Lessons Learned during Wiki Migration

5. Communication is key.
6. Work systematically.
7. If something doesn't work, it's okay to start over.
8. Get feedback from potential users and adjust accordingly.



@mac_omaha2017 | #omamac17 | #s402

Communication is Key

- Collaboration between two departments
- Open line of communication
- Document processes
 - Make sure to update documentation



@mac_omaha2017 | #omamac17 | #s402

Work Systematically



Description

Rating: ☆☆☆☆ Based on 0 rating(s)

Title: [Boy rowing makeshift raft at flooded Knowlton's Corner, Cincinnati, Ohio](#)

Photographer: [Ross, Betty T.](#)

Date (Original): 1937

Date: 1937

Time Period: 1931-1945

Keywords: [Floods](#), [Stores, Retail](#), [Rafts](#)

Location: [Northside \(Cincinnati, Ohio\)](#)

Format (Original): [Digital Images](#)

Type (Original): [Still Image](#)

Collection: [Photos by Betty T. Ross](#)

Format (Digitized): [jpg](#)

File Size: **449083 Bytes**

Date (Digitized): 2012-01-10

Publisher (Digital): [Public Library of Cincinnati and Hamilton County, Digital Services Department](#)

Copyright Notice: [The materials in this collection are made available for use in research, teaching and private study. Reproduction rights are not available for this image.](#)

File Name: [OHRiverFlood022_036.jpg](#)

@mac_omaha2017 | #omamac17 | #s402

It's Okay to Start Over

The screenshot displays the website for The Public Library of Cincinnati and Hamilton County. The header includes the library's name and navigation links: Home, Exhibit Home, Browse Exhibit, About the Digital Library, Log In, Help, and English. A search bar is located at the top right with the text 'Advanced Search' and 'with results' and a 'SEARCH' button. The main content area shows a search results page for 'Green Acres'. On the left, there is a sidebar with the heading 'Add or remove other collections to your search:' and a list of categories: 19th Century Photography, Adobe Manual Scrapbook, Arts & Recreation, Audubon's Birds of America, and Show 50 more. Below this is a 'Narrow your search by:' section with filters for Title, Keywords, and Format (Digital). A 'Browse suggested topics' section is also present. The main area displays a grid of book covers with titles such as 'A local history of the Cincinnati Horticultural Society in relation to the city', 'A notice on the cultivation of the grape in vineyards', 'Annual catalogue, 1892', 'Annual catalogue, 1892', 'Annual catalogue, 1892', 'Annual report of the Department of Gardening', 'Catalogue of plants cultivated in the garden of Joseph Dicks arranged', 'Catalogue of plants cultivated and for sale by E.S. Jackson, nurseryman and florist', 'Catalogue of the flowering plants, ferns and fungi growing in the vicinity of Cincinnati', 'Chemical composting', 'Cincinnati sea gardens, V. 01', 'Cincinnati sea gardens, V. 02', 'Flowers from the Holy Land', and 'Herbals of the continent: 50 original books from German, French, Dutch, and Italian sources'. The page indicates '137 results'.

@mac_omaha2017 | #omamac17 | #s402

Get Feedback



Rating ★★★★☆ Based on 8 ratings

Title Delta Queen construction

Real Name and/or Subject Delta Queen

Date (Original) 1925, 1926

Date 1925, 1926

Date Period 1921-1930

Description Directory No. P 574
Frame # 1022 (captioned...)
From: Frank J. Burns.

Note Photographs prefaced with "P" or "C" are listed and described in Way's Packet Directory, 1924 edition. Those prefaced with "T" are listed and described in Way's Tourist Directory, 1933 edition.

Keywords Delta Queen (Description)
Passenger
Shipbuilding
Tourism

Location Cincinnati (Ohio)

Format (Original) Photographs

Type (Original) Still Image

Collection Public Library of Cincinnati and Hamilton County, Inland River Photographs

Sub Collection Public Library of Cincinnati and Hamilton County, Inland River Photographs, Frederic Way, Jr. collection

Public Name Delta Queen

Access to this item [Open Access](#)

Format (Digital) jpg

File Size 272773 Bytes

Date (Original) 2007-09-07

Publisher (Digital) Public Library of Cincinnati and Hamilton County, Digital Services Department

Copyright Notice Please contact the Public Library of Cincinnati and Hamilton County for information regarding copyright for this item.

Custom Information For information regarding image use permissions, contact the Public Library of Cincinnati and Hamilton County's Genealogy & Local History Department at <http://www.cincinnatihistory.org/contact.aspx?for=105>.

File Name [DeltaQueen007a-50.jpg](#)

@mac_omaha2017 | #omamac17 | #s402

Looking Back



@mac_omaha2017 | #omamac17 | #s402

Thank You!



THE
PUBLIC
LIBRARY
of Cincinnati
and
Hamilton County

Contact Information:

Stephanie Bricking, MA, MLS

Metadata Librarian

Stephanie.Bricking@cincinnati-library.org

@mac_omaha2017 | #omamac17 | #s402



TIERED: UK LIBRARIES' APPROACH TO MIGRATING BORN-DIGITAL CONTENT

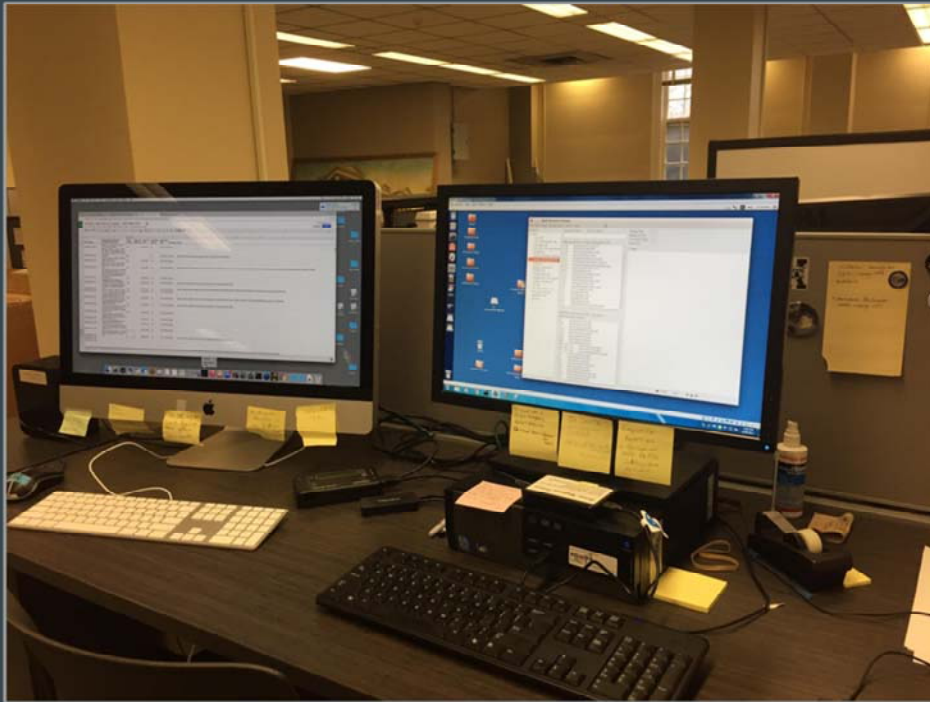
SARAH DORPINGHAUS
@SMDORPINGHAUS

@mac_omaha2017 | #omamac17 | #s402

UK Libraries Tiered Approach to Migrating Born-Digital Content



Sarah Dorpinghaus, Director of Digital Services
Special Collections Research Center



Background information:

UK Libraries SCRC born-digital workstation includes readers for CDs, DVDs, zip drives, 3.5" diskettes, and multi-card reader.

The workstation is running a VM with a BitCurator suite of tools for creating disk images and analysis of files.

Migration Project

Nicole Reynolds: GA & practicum

- Identify
- Migrate
- Document
- Workflow revision

UK Libraries Graduate Assistant Nicole Reynolds worked to develop workflows for migrating born- digital materials. In an additional practicum experience, she focused specifically on our born-digital backlog.

In general, trying to create a disk image for everything and capture content as completely as we can.

We search our content management system for keywords relating to digital materials, including alternate spellings. We found over 170 collections with some mention of digital materials.

Nicole updated documentation (accession or resource records) as collections were migrated.

Nicole also revised, enhanced, and created new workflows for media or challenges we had not yet encountered.

Conclusions

- Balance of comprehensiveness and efficiency
- Tiered approach

Over half way through her practicum project and we've realized a few things:

1. We don't know how to describe the reports and disk images or make them available; in fact maybe we don't even need all this information
2. It's taking much too much time to do full forensic migrations for all of this content
3. Need to reevaluate our policies and procedures -- tiered approach (not a new idea)

	MINIMAL	INTERMEDIATE	FULL
TRANSFER	<ul style="list-style-type: none"> Virus scan as needed/batch Robocopy when appropriate 	<ul style="list-style-type: none"> Virus scan Robocopy, VLC, Windows Media Player, or Handbrake 	<ul style="list-style-type: none"> Virus scan Disk image- Guymager File extraction- BitCurator Disk Image Access

Part 1: Defining the Tiers

		informative, capture information/artwork	<ul style="list-style-type: none"> Capture/digitize PII with descriptive or visual info
REVIEW		<ul style="list-style-type: none"> Review files for challenging formats 	<ul style="list-style-type: none"> Review files for challenging formats and no extension/formats Reformat copy of files with Quick View Plus if needed Search for PII and other sensitive information at a general level

	MINIMAL	INTERMEDIATE	FULL
TRANSFER	<ul style="list-style-type: none"> Virus scan as needed/batch Robocopy when appropriate 	<ul style="list-style-type: none"> Virus scan Robocopy, VLC, Windows Media Player, or Handbrake 	<ul style="list-style-type: none"> Virus scan Disk image- Guymager File extraction- BitCurator Disk Image Access
STORE	<ul style="list-style-type: none"> BagIt standard Multiple backups 	<ul style="list-style-type: none"> BagIt standard Multiple backups 	<ul style="list-style-type: none"> BagIt standard Multiple backups
DOCUMENT	<ul style="list-style-type: none"> Update accession record 	<ul style="list-style-type: none"> Collection-level data in "born digital" spreadsheet Item-level media log File Info Exporter 3000 documentation file Update accession record If PM is especially visual or informative, capture information/artwork 	<ul style="list-style-type: none"> Collection-level data in "born digital" spreadsheet Item-level media log Reports: Bulk Extractor Viewer and BitCurator Reporting Tool Update accession record Capture/digitize PM with descriptive or visual info
REVIEW		<ul style="list-style-type: none"> Review files for challenging formats 	<ul style="list-style-type: none"> Review files for challenging formats and no extension/formats Reformat copy of files with Quick View Plus if needed Search for PII and other sensitive information at a general level

Some aspects of the work remain the same no matter what. For example, we will always run virus scans on our PM. We will also store all our files in the BagIt standard and have multiple copies, regardless of what level of work was done while migrating the content.

But there are significant differences.

Minimal: Goal is to essentially transfer the content quickly and do minimal updating of the existing documentation.

Full: Full forensic migration: disk image, extract the files, generate the report, investigate for PII, capture the visual info on the PM, item-level media log

There's an in between area- Intermediate: When a full disk image isn't required but we still want to extract the embedded metadata via File Information Exporter 3000 (directory listing with last date modified, file type and size, file name, etc.); everything else (normalizing files, capturing the info on the PM) is a case by case basis.

Tier 1: Full Forensic Migration

Part 2: Decision Tree

- Does the **research value** of the collection warrant a full forensic migration?
- Is there likely **PII or other sensitive information** stored in the files?

How do we know what level of migration is appropriate for a collection or a particular PM?

Created a set of questions and decision points will dictate the amount of information gathered at time of migration and ultimately the level of access that can be provided.

Answers to the following questions are based on conversations with the creator and/or donor, the deed of gift, a visual survey of the physical media, and a general understanding of the collection's contents.

Tier 1: Full Forensic Migration

- Is the PM more than a vehicle for transport or storage? Were the files **created and edited on the PM?** (ex: laptop, phone)
- Is it most **efficient** to create a disk image of the PM? (ex: laptop, computer)
- Does the **research value** of the collection warrant a full forensic migration?
- Is there likely **PII or other sensitive information** stored in the files?

Tier 1: Determine if a **full forensic migration** is appropriate

- Was the PM used more than a vehicle for transport or storage? Were the files created and edited on the PM? (Example: laptop or phone)
- Is it most efficient to create a disk image of the PM? (Example: laptop or computer)
- Does the research value of the collection warrant a full forensic migration?
- Is there likely PII or other sensitive information stored in the files?

Examples: judicial collection

Tier 1: Full Forensic Migration

- Is the PM more than a vehicle for transport or storage? Were the files **created and edited on the PM?** (ex: laptop, phone)
- Is it most **efficient** to create a disk image of the PM? (ex: laptop, computer)
- Does the **research value** of the collection warrant a full forensic migration?
- Is there likely **PII or other sensitive information** stored in the files?

If no, then on to Tier 2 

Some collections will have PM that warrants a full forensic migration while others may warrant only intermediate or minimal attention.

If a collection or individual PM do not meet the requirements for a full forensic migration then move on to Tier 2.

Caveat!

- Is the PM **damaged**? (determined via a physical survey or other audio/visual signs during transfer)
- Are the files or the PM **corrupted**? (determined during transfer)

Some materials originally designated as minimal or intermediate that may be reclassified as warranting a full migration

- Is the PM damaged? (determined via a physical survey or other audio/visual signs during transfer)
- Are the files or the PM corrupted? (determined during transfer)

Tier 2: Intermediate Migration

- Does the PM potentially have **useful** descriptive or preservation **metadata** embedded in the files?

Tier 2: Determine if an **intermediate migration** is appropriate

- Does the PM potentially have *useful* descriptive or preservation metadata embedded in the files?
- Were the files transferred onto the PM from a machine used to create the files close to the time of donation?

Example: organizational records

Tier 2: Intermediate Migration

- Does the PM potentially have **useful descriptive or preservation metadata** embedded in the files?

If no, then on to Tier 3 

If a collection or individual PM do not meet the requirements for an intermediate migration then move on to Tier 3.

Tier 3: Minimal Migration

Typical characteristics

- Embedded metadata on files is **not useful**
- Embedded metadata is **not accurate**
- Files are an **internal transfer** from administrative units within the university

Tier 3: Determine if a **minimal migration** is appropriate

Typical characteristics of content appropriate for minimal migrations:

- Embedded metadata on files is not useful
- Embedded metadata is no longer accurate
- Files traveled through several “carriers” before arriving at SCRC
- Files are an internal transfer from administrative units within the university

Example: university publications

What if we can't read the media?



PM is kept in collection and files are not migrated when SCRC cannot acquire the readers to access the files.

- ① *Research value*
- ① *Quality of embedded metadata*
- ① *Efficiency*

It boils down to three questions.

Looking forward >>>

- Continue migrations
- Continue research & workflow revision
- Description and access

sarah.dorpinghaus@uky.edu
@SMDorpinghaus



@mac_omaha2017 | #omamac17 | #s402





ELBOW GREASE AND FLEXIBILITY: JUST WHAT OUR MULTIFACETED MIGRATION PROJECT NEEDED

LORIMYERS-STEELE

@mac_omaha2017 | #omamac17 | #s402

Elbow Grease and Flexibility: Just What Are Multifaceted Migration Project Needed

Lori Myers-Steele
Berea College Special Collections and Archives

@mac_omaha2017 | #omamac17 | #s402

BEREA COLLEGE
SPECIAL COLLECTIONS & ARCHIVES

Berea College Externally-Funded Special Collections and Archives Preservation and Access Grant



Project's purpose:

- Digitally preserve, centralize, and make accessible the collections gathered from previous grants
- Gain intellectual control over collections, centralize data, and make information more accessible for both archives staff and patrons



Produce a customized and integrated digital asset management system and collection-wide discovery tool for improved searching and access to SC&A collection records, digital objects, library guides, and digital exhibits.

@mac_omaha2017 | #omamac17 | #s402

BEREA COLLEGE
SPECIAL COLLECTIONS & ARCHIVES



Where we started:

- Utilizing ContentDM for our digital collections and presence
- 100+ terabytes of audio visual materials in a variety of Sound Archives collections with no central storage
- Using a variety of software and databases for different functions
- Disparate online presence for finding aids and guides
- Numerous collections partially processed or without complete finding aids
- Recent staffing changes with a great loss of institutional and collection knowledge



But...

We knew we had great collections reflecting our unique history and collection mission – we needed to make them more accessible and manageable.

@mac_omaha2017 | #omamac17 | #s402

BEREA COLLEGE
SPECIAL COLLECTIONS & ARCHIVES



**Where the elbow grease and flexibility came in handy
(a.k.a. “the challenges”):**

- Strict timelines set by grant that —in most cases— could not be changed
- Organizational issues with large *Preservica* digital records upload, frustrations with new platform and *Universal Access*
- Data migration and metadata mapping problems
- Continual clean up and quality checking of data, imports and access issues
- Changes in staffing throughout 3-year grant period

@mac_omaha2017 | #omamac17 | #s402

BEREA COLLEGE
SPECIAL COLLECTIONS & ARCHIVES



Takeaways (or what we've learned):

- Flexibility is key – develop and follow work flows; however be ready and willing to change course
- Learn from your mistakes
- It takes a village (or team) and lots of elbow grease
- If you build it, they will come



@mac_omaha2017 | #omamac17 | #s402

BEREA COLLEGE
SPECIAL COLLECTIONS & ARCHIVES



MOM: MASTER OBJECTS MIGRATION PLANNING AT TOSU

DAN NOONAN
@DANNYNOONAN1962

@mac_omaha2017 | #omamac17 | #5402



THE OHIO STATE
UNIVERSITY

**MOM:
MASTER OBJECTS MIGRATION PLANNING AT TOSU**

go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/s402)





WHERE ARE WE GOING?



- Setting the stage: A very brief history...
- Project planning & migration preparation...
- Metadata, a love note...
- Challenges for our further adventures...

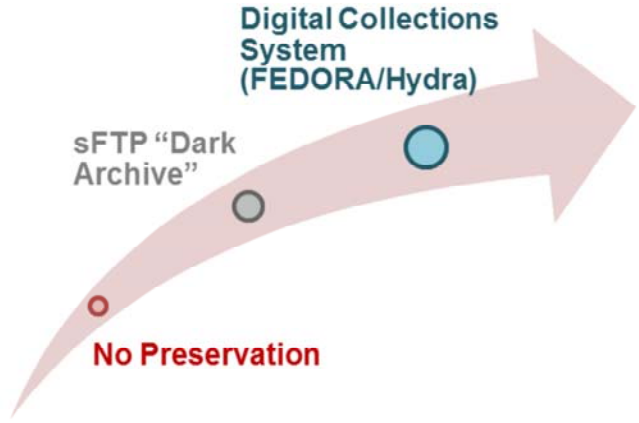
go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/#s402)

Good Afternoon, again! I will be leading you on a brisk trip, exploring OSU's adventures in migration—from organized chaos to a FEDORA/Hydra-based preservation environment. I will set the stage with a brief history of our digital preservation efforts and then provide an overview of our project planning and migration prep activities, that includes getting ourselves elbow-deep in file and collection identification, deduplication, prioritization, metadata and metadata transformation prior to ingest. Lastly, I will identify our existing challenges as our migration activities move forward under full steam.

A BRIEF HISTORY...

- Digital Projects drive
- “Dark Archive”
- Digital Preservation Framework (The Framework)
- Master Objects Repository (MOR)



go.osu.edu/archives | go.osu.edu/noonan | @dannynoonan1962 | @mac_omaha2017 | #omamac17 | #s402

So a long time ago in a library far, far away...a story whose origins are lost in the mists of time (or at least more than a decade ago). No one can say definitively that this is what happened, but it is what I have been able to piece together, along with my own first-hand knowledge over the past ten years. Outside of a couple of big projects dedicated to brittle books and theses and dissertations, the bulk of digitization efforts were conducted by our Special Collections and Archives personnel. As they began to fill departmental share-drive space with their projects, the Libraries began to run out of storage space. To accommodate this growing digital mass, the Libraries IT department pulled an old web server out of mothballs to create a shared drive known simply as K.

Sometime around a decade ago that drive became unstable and its contents were unceremoniously placed on a server known as dspace04. This gave rise to the myth that their collections were being preserved in DSpace. NO, dspace04 was just a sever used for staging DSpace upgrades and ingests, and just happened to have excess capacity. With the dumping of the K-drive server space contents onto dspace04 (and not into DSpace itself), we ended up with all sorts of digital materials on this server. We had inadvertently created the “Dark Archive” with little or no consideration for what we were putting there—not only did we get digital master objects, but derivatives, working files and various detritus.

However, a key benefit of the Dark Archive is that it did/does provide controlled access through sFTP. As such, the Libraries did subsequently begin to take more prescribed steps in deciding what it put onto this server, while carving out a new K-drive space to actively

work on projects; however, this did not curtail excessive amounts of duplication of master and derivative objects, nor was there any official policy around its use; and lastly good file management policies/techniques/processes were not used, nor in many cases basic metadata created.

In 2012 and 2013 a team from the OSU Libraries participated in the DigCCur Institute. Our project was the development of a digital preservation policy framework (or The Framework) that began to set the stage for migration to a true preservation environment. This effort dovetailed with the hiring of our Head of Digital Initiatives, Terry Reese, who is the chief architect of our new FEDORA/Hydra preservation environment. In 2014 he spearheaded the Master Objects Repository Task Force, which laid out a plan for our digital preservation activities including:

- defining Master and Derivative Objects
- defining the environment and high-level management processes for the M-O-R
- recommending procedures for proper deposit and registration of appropriate objects in the M-O-R including workflows ,metadata for management & identification purposes, and interactions with other systems as appropriate.

The recommendations were software/hardware agnostic to allow Master Objects to be migrated to and preserved on future storage platforms.

Subsequently in 2015 the Libraries decided to implement a FEDORA repository solution using Hydra-based Sufia for our user interface.



PROJECT PLANNING & PREPARATION



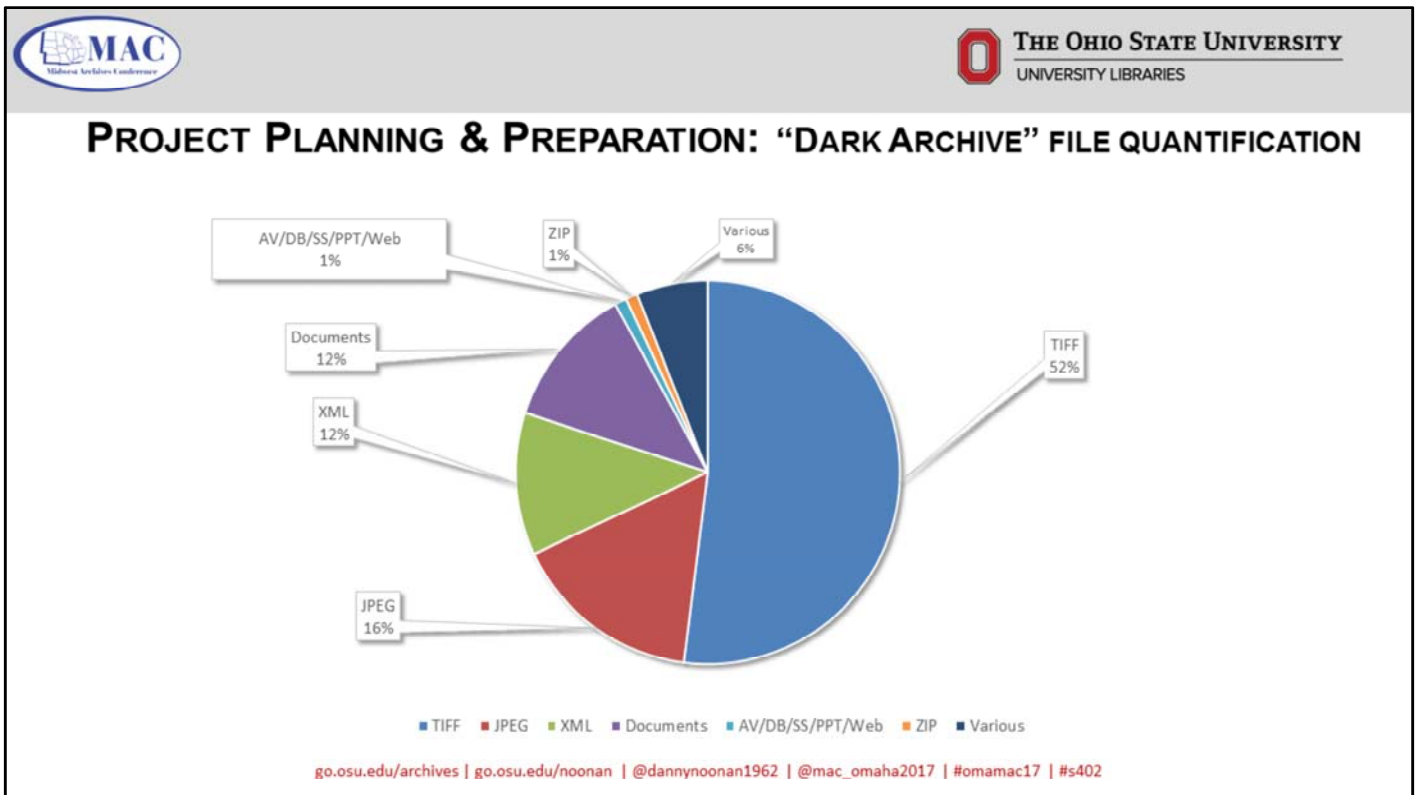
- De-duplication effort
- “Dark Archive” file quantification
- “Collection” Analysis Tool
- Prioritization

go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/#s402)

So where to start. As early as late 2011, the Libraries engaged a retired librarian to conduct a rudimentary inventory of our digital stuff. I inherited this inventory that covered not only items in our Dark Archive, but also our DSpace repository know as the KnowledgeBank, our shared drives and items on loose media. Through some educated interpretation and SWAGging I estimated that we had upwards of 14TBs that likely needed to find a home in a true preservation environment. One of the things this inventory lacked was a comprehensive look at our Dark Archive and its contents.

As we began to examine the Dark Archive, one thing we were certain of was that there was/is a significant amount of duplication within the it and with the replacement K-drive and the departmental/committee shared J-drive. In conjunction with the development “The Framework”, we started to conduct a de-duplication effort on the Dark Archive, where we identified over 215,000 duplicates. This was driven by the fact that we were running out of digital storage space at the time. Working with our IT Infrastructure Support group, we developed spreadsheets that identified file-paths for duplicate pairs (and sometimes triplicate, quadruplicate or more). In sharing these with the responsible collection archivist or curator, we discovered that they also likely had copies on the K-Drive. So we did a Dark Archive vs K-drive analysis with the intention of retiring the use of the K-drive and making certain all masters were in the Dark Archive and derivatives distributed to their appropriate access point. By mid-2014 we had made significant headway of de-duping the Dark Archive and had finally retired the K-drive (or so we thought, but more on that later).

2015 saw the implementation of our FEDORA/Sufia platform whose pilot content was Libraries' collections content migrated from an external system that another campus entity no longer supported.



In preparation for migration of content from the Dark Archive, we identified more than 85 files types and nearly 2,000,000 objects that needed to be considered for migration. The good news was that 52% were TIF images that for the most part should be a no-brainer for migration. The next largest quantity of files were JPEGs which may be masters or derivatives; documents, the bulk of which are PDFs; XML which may be metadata, but the bulk of which are poorly formed faux-xml; various AV objects, Databases, Spreadsheets, PowePoints and web-files; and zip files, whose internal contents will need to be examined.

There remaining 6% are obscure file types that may or may not need to be migrated or are the result of poor file naming practices.

We now knew how many things we had, but who do they belong to and how do we prioritize the migration of more than a million items?

And oh, what about all the metadata that will be needed, because one thing is absolutely clear: NOTHING GOES INTO THE MASTER OBJECTS REPOSITORY WITHOUT A MINIMUM AMOUNT OF METADATA!!!!!!



PROJECT PLANNING & PREPARATION



- De-duplication effort
- “Dark Archive” file quantification
- “Collection” Analysis Tool
- Prioritization

go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/#s402)

So how did we approach the prioritization?

Fortunately, the Dark Archives’ folder structure is set to coincide with collection owners. Right off the bat, we put 47% of the files on the back-burner as they are either master objects or support files for items in the KnowledgeBank—those will be the last files we will examine when we determine our strategy for interaction between DSpace and FEDORA.

Nearly a quarter of the files account for 11 collections within the University Archives, which are mostly from the Office of the President’s document management system, which will need a secure instance of the M-O-R due to FERPA/privacy issues.

The remaining files—just shy of 30%—belong to 6 groups spread over approximately 150 collections, which means a lot more detailed analysis.

I constructed an Access database based upon the file-paths in the Dark Archive that was then shared with the appropriate archivists and curators that:

- Presented the file-path and quantification of file types within, and then
- asked them to Identify:
 - What collection the items belonged to?
 - Are there other objects that belong to this collection, and where are they located?
 - Whether the objects should be migrated or disposed, or are need of further processing or assessment?
 - What type of object are these?
 - Preservation Master
 - Provisional Master
 - Working copy
 - Access Copy
 - Reproduction Copy
 - Are these preservation formats?
 - Does collection level and individual metadata exist, and if so, where?
 - What are the intellectual property rights?
 - Public Domain
 - OSU Owned
 - Donor Owned

- Mixed
 - Unknown
- And what type of access we are allowed to provide?
 - Public
 - Reading room
 - Private
 - Closed



PROJECT PLANNING & PREPARATION



- De-duplication effort
- “Dark Archive” file quantification
- “Collection” Analysis Tool
- Prioritization

go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/#s402)

This allowed us to begin to prioritize the migration, by identifying those collections with the most available metadata, and in which we can provide public access.

An additional wrinkle, that recently began to go away, is that the only digital object our data model could handle was that of images, and only as simple/individual objects. Therefore, our highest priority were those collections that were TIFF or mostly TIFF, and of a single object nature. Secondary considerations were given to other image types, followed by documents, complex objects and audio-visual objects. With the upgrades to FEDORA and Sufia that are currently being deployed, we are beginning to be able to accommodate ingest of other data models that included complex objects and other non-image file types.



METADATA IS A LOVE NOTE TO THE FUTURE



- Application Metadata profile
- Minimum Metadata guidelines
- Content Standards & Controlled Vocabularies
- Multiple metadata streams

go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/#s402)

[Derived from prior presentation of work of Darnelle Melvin]

So you remember I said: one thing is absolutely clear: NOTHING GOES INTO THE MASTER OBJECTS REPOSITORY WITHOUT A MINIMUM AMOUNT OF METADATA!!!!!!

Metadata can sometimes be messy, complex, incomplete, inconsistent, non-existent, and the quality can vary from resource to resource and collection to collection. Good metadata can increase discoverability, enhance the user experience, and identify the parties involved in the creation or contributions of a given work. Good metadata can also, act as an access point to the collection or resource once ingested into the repository; and as the slide suggests, metadata truly is a love note to the future.

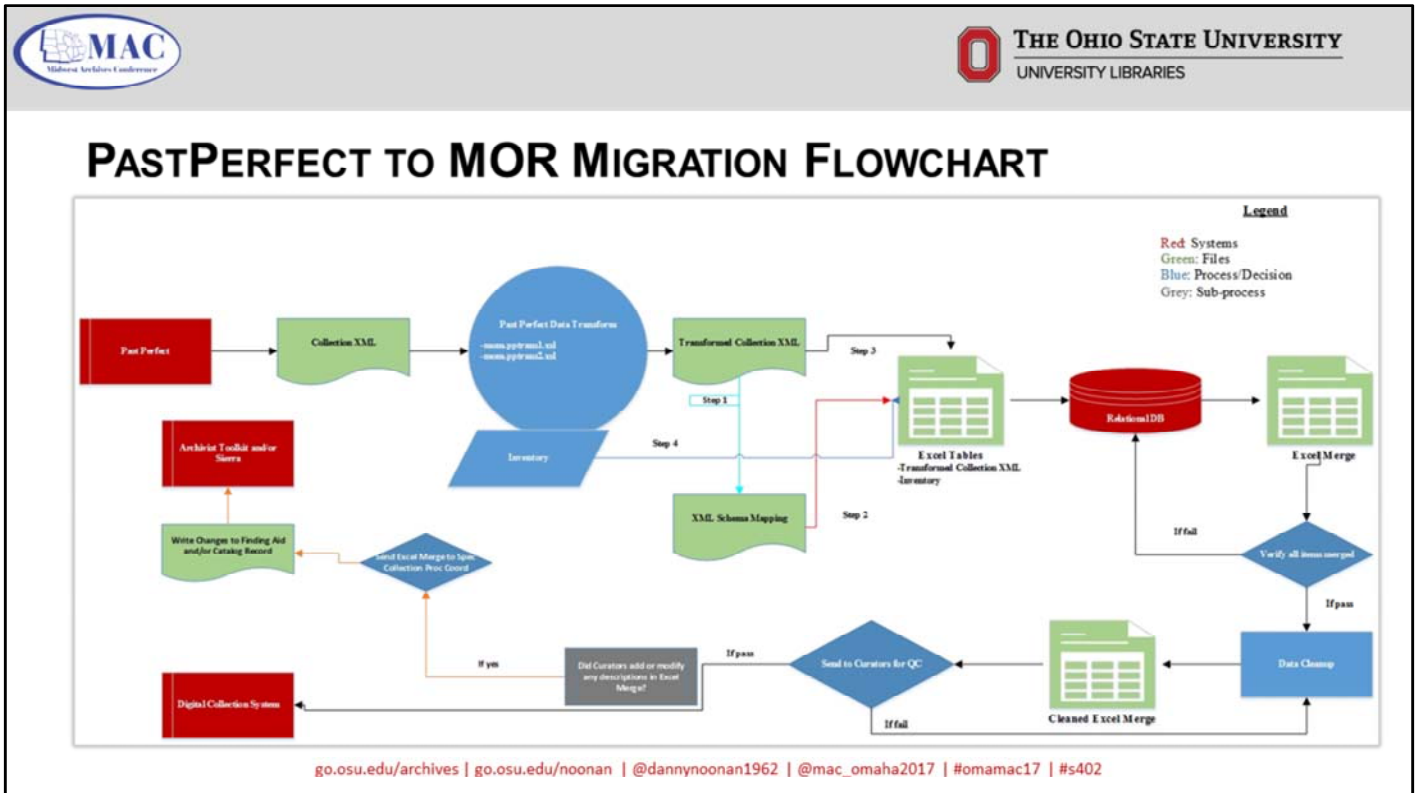
Our Metadata Working Group developed a Metadata Application Profile, a group of metadata elements, attributes, and data values which includes policies, guidelines and examples for resources intended to be stored in the M-O-R. The elements defined in the application profile were drawn from established metadata schemas such as: Dublin Core, VRA Core, and PREMIS.

From this application profile, we established subset of elements that are the minimum required fields for any item to be ingested:

- Unit owner
- Collection name

- Access Rights
- A unique Identifier
- Preservation Level & Rationale
- Resource Type
- Ownership Rights
- Title
- And if applicable:
 - Creator
 - Language
 - Sub-collection

In addition to the Application Profile, we recommended the use of established content standards and controlled vocabularies, such as CCO, DACS, RDA, Art and Architecture Thesaurus, DCMI Type Vocabulary, LCSH, and the Getty Thesaurus of Geographic Names among others.



[Derived from prior presentation of work of Darnelle Melvin]

Since multiple special collections units are ingesting content into the M-O-R, we needed an application profile that was robust but also flexible for a wide variety of stakeholders and actions.

The metadata is being transformed from PastPerfect, Access, Excel, EAD, MARC and other sources.

This is one example—and not for the weak of heart—of a metadata transformation workflow for migrating objects for which we have metadata in PastPerfect.



CHALLENGES

- Staffing
- Initially images only
- Dependence on FEDORA/Hydra/Sufia communities
- Spotty metadata



go.osu.edu/archives | go.osu.edu/noonan | [@dannynoonan1962](https://twitter.com/dannynoonan1962) | [@mac_omaha2017](https://twitter.com/mac_omaha2017) | [#omamac17](https://twitter.com/omamac17) | [#s402](https://twitter.com/s402)

In conclusion: I'd just like to point out a few of the challenges we have had/still have:

Staffing: Both for the development effort, as well as the curatorial and metadata efforts, we just haven't had enough staff to make it happen as fast as anyone would like to see it accomplished. Developer churn created issues early on, but we are relatively stable and staffed-up now. As for metadata transformation, our metadata transformation librarian is a term appointment and we will lose him within the next six months. And the archivists and curators try to find time to create metadata, while doing the rest of what they do—hence a need for clear prioritization.

Complicating the migration prioritizations was our self-imposed limitation for images only. That limitation recently has been lifted, but we are still just beginning to experiment with documents and complex objects. We will not have true ordered complex object capabilities until sometime this summer. Additionally, we have curator pushback that the system does not support enough archival hierarchy for appropriate contextual search and retrieval.

Another challenge has been a dependence on the FEDORA, Hydra and Sufia communities. This is not meant to throw slings and arrows at them, its just that without the expertise on staff we are reliant on community development for all the cool stuff we want to do—or sometimes basic functionality that our archival/curatorial staff expect.

Lastly, the one thing that will keep us adrift is chasing down that elusive treasure,

METADATA.

Thank you.



CONTACT INFO:

Dan Noonan

Digital Resources Archivist, Assistant Professor
University Libraries | University Archives
The Ohio State University
320B 18th Avenue Library
175 W. 18th Avenue
Columbus, OH 43210
noonan.37@osu.edu
614.247.2425
@dannynoonan1962
go.osu.edu/noonan

go.osu.edu/archives | go.osu.edu/noonan | @dannynoonan1962 | @mac_omaha2017 | #omamac17 | #s402

Thank you.



Q&A

Laura Alagna

Northwestern University Libraries
laura.alagna@northwestern.edu

Stephanie Bricking

Public Library of Cincinnati and Hamilton County
Stephanie.Bricking@cincinnatilibrary.org

Sarah Dorpinghaus

University of Kentucky Libraries
sarah.dorpinghaus@uky.edu

Lori Myers-Steele

Berea College Hutchins Library
Lori_Myers-Steele@berea.edu

Dan Noonan

The Ohio State University Archives
noonan.37@osu.edu

ANNUAL MEETING AND SESSION EVALUATION FORM AVAILABLE ONLINE:

bit.ly/OMAMAC2017

@mac_omaha2017 | #omamac17 | #s402