

Laura Alagna  
Nat Wilson  
Sarah Dorpinghaus  
Doug Boyd  
Michael Shallcross  
Dan Noonan  
Cinda May

# When the Rubber Hits the Road: Real-World Digital Preservation



Midwest Archives Conference

May 23, 2018

#MAC2018 #s302

# Digital Preservation (in the real world)

Laura Alagna, Northwestern University

 @Digitized\_Laura

## Results from the 2<sup>nd</sup> NDSA Storage Survey

- The NDSA Infrastructure Working Group surveyed NDSA membership in 2011 and 2013 on digital preservation practices
- Trends revealed:
  - Memory institutions are continually increasing the amount of data preserved – NDSA members surveyed nearly doubled content preserved between 2011 and 2013
  - Organizations generally underestimate growth in digital content
  - Surveys indicated that respondents have a strong record of mitigating against risk of disasters, but mitigating against everyday threats such as bit rot is “an opportunity for improvement”
- **“General preservation practices are not always reflective of community best practice standards”**  
(<https://doi.org/10.1045/july2017-gallinger>)

## “Beyond the Repository”

- Team from Northwestern and UC-San Diego conducted a survey on digital preservation in 2017 as part of an IMLS grant
- We widely distributed the survey and received 170 complete responses
- Trends revealed:
  - Significant variety in how digital preservation is conducted
  - More than 90% reported preserving a terabyte or more
  - No one digital preservation system was used by a majority of respondents
  - Survey responses cite a number of barriers to achieving better digital preservation practices, including limited funding, staff, or expertise, as well as lack of buy-in from administrators and limitations in technology
- **“This is so messy...”**
- (<https://doi.org/10.21985/N28M2Z/>)

If only...



---

---

# When the Rubber Hits the Road: Real-World Digital Preservation

— Nat Wilson —  
Carleton College  
#MAC2018 #s302

---

---

# Digital Preservation Planning - 5 years later

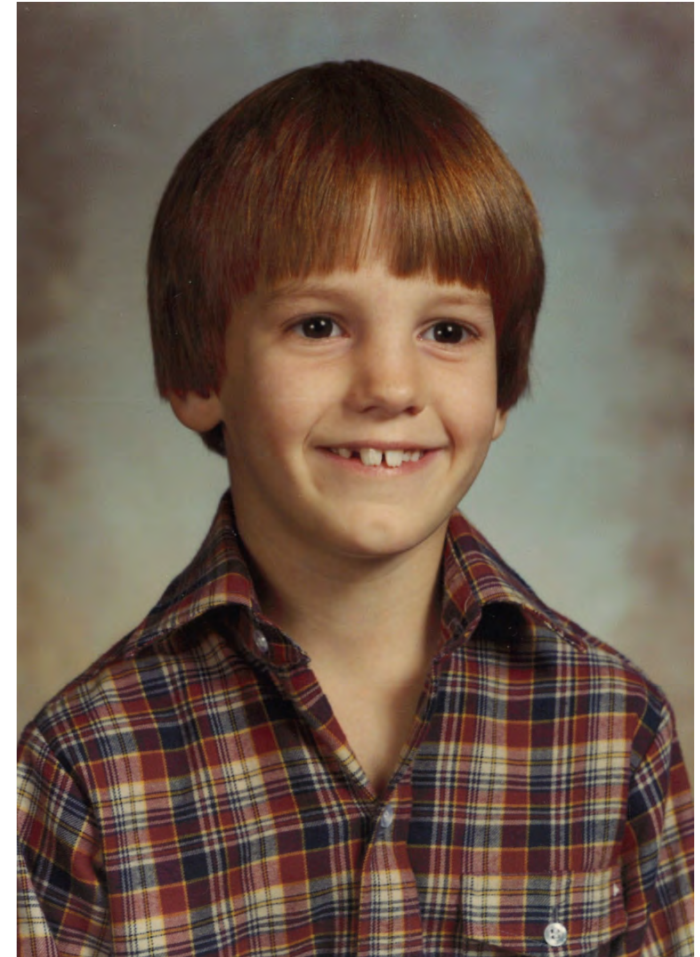
Ideal vs realistic goals

Started work at Carleton in 2010

# Digital Preservation Planning - 5 years later

Ideal vs realistic goals

Started work at Carleton in 2010





## **Digital Preservation Planning - 5 years later**

No program in place for digital archives at Carleton

Framework for preservation planning in 2012

Implementing the plan for a little over 5 years

# First Things First - Policy Planning in 2012

Starting from scratch

Strong support from institution

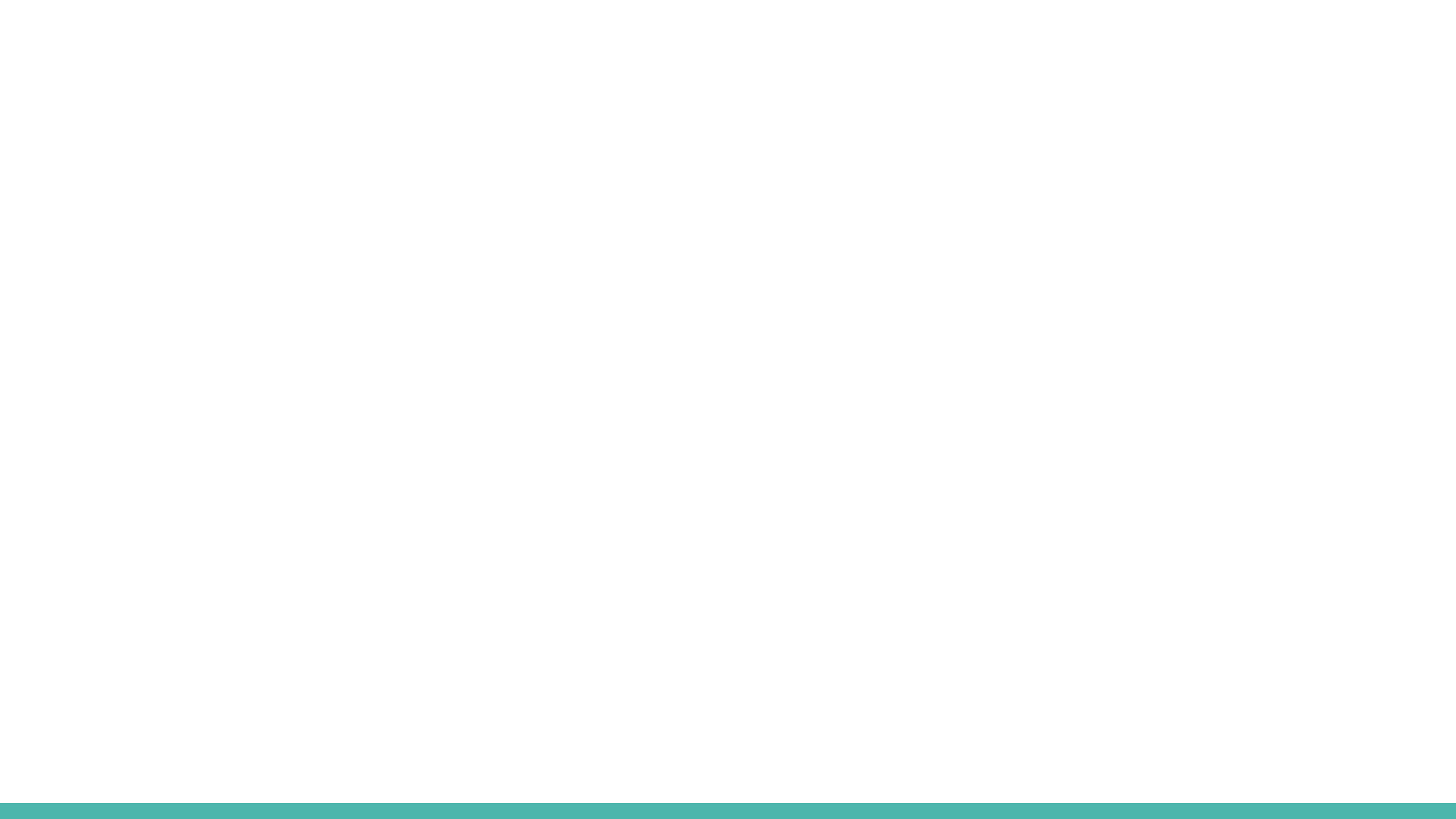
Worked closely with Library Director from the beginning

- Director was very invested in the process and supportive of the Archives
- Extensive experience with strategic planning

# First Things First - Policy Planning in 2012

## Goals:

- Widely applicable
- Scalable
- Prioritizes records and adjusts practices accordingly
  - Efficient allocation of resources



**CARLETON COLLEGE**  
**CONCEPTUAL FRAMEWORK**  
**FOR DIGITAL ASSET MANAGEMENT**  
**OF ARCHIVAL RECORDS**  
**(CCCFDAMAR)**



**CARLETON COLLEGE**



**CONCEPTUAL FRAMEWORK**

**FOR DIGITAL ASSET MANAGEMENT**



**OF ARCHIVAL RECORDS**

**(CCCFDAMAR)**



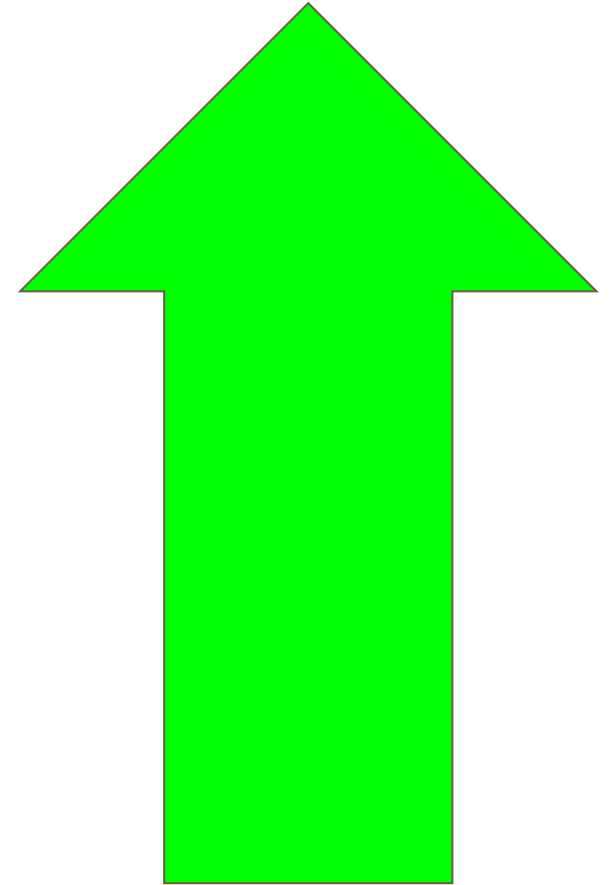
## Tiered approach / Triage

- 3 Levels for records in the archive
- Higher tiered items received more care
- Placement determined by numerous factors

## Tiered approach / Triage

Factors that increased a record's ranking.

- High value to institutional record
- Cost of replacement in the event of loss
- Heavily used

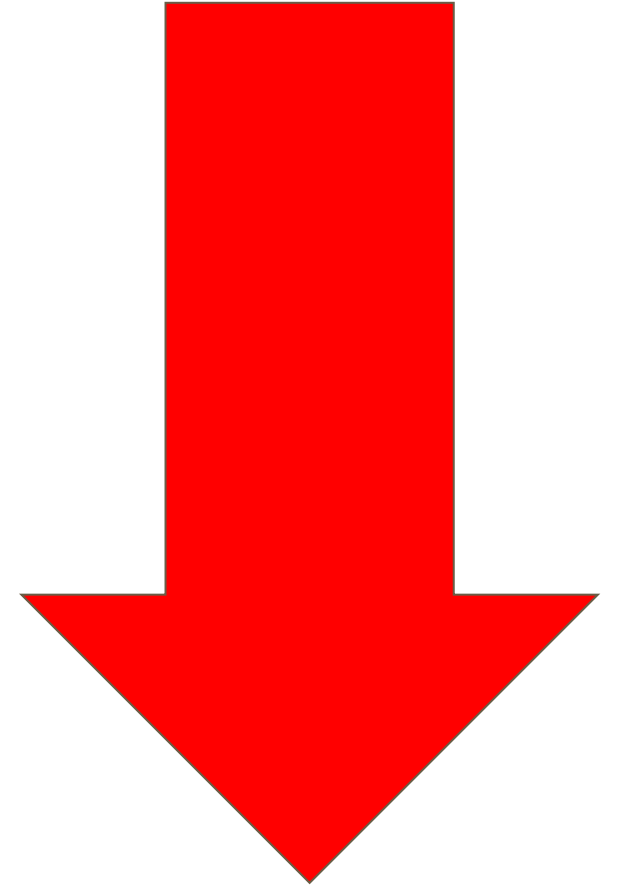




## Tiered approach / Triage

Factors that lowered a record's ranking.

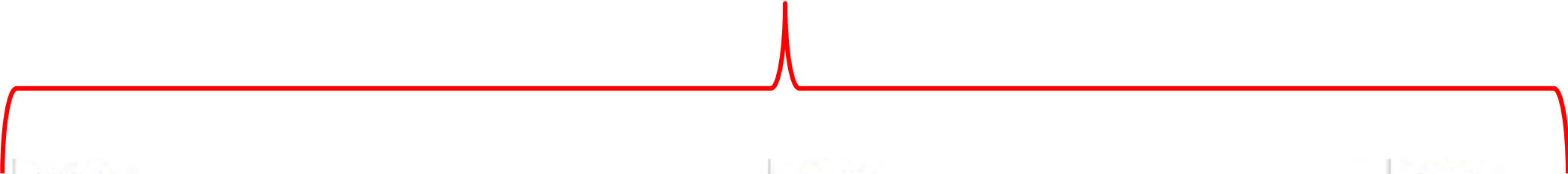
- Cost of preservation
- Low to moderate value to institution
- Hardware or software dependencies



Tiers/Perservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

Tiers/Perservation Actions	Tier 1	Tier 2
General Description	<p>Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.</p>	<p>Important records of cultural heritage of the college. Loss would compromise our future history of the college. As for Tier 1 records, every effort would be made to recover lost or corrupted files.</p>
Fixity Checks	<p>Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.</p>	<p>Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.</p>
Preservation Metadata	<p>High level of detail. Keep full PREMIS records.</p>	<p>Moderate level of detail. Regular verification, creating and updating records at the level.</p>
Backup cycle	<p>Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.</p>	<p>Average frequency for all campus. Backups and active server directories under ITS program.</p>

Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration unless original is in our list of supported formats. If original is not a supported format, it will not be accepted into archives, we will migrate to the most preferred preservation format.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to system structure. 1 time/ 5 years regardless of system changes.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted for migration. Some unsupported formats may be acceptable with the understanding that migration to a preservation copy will be necessary and some information may be lost.



<b>Tiers/Perservation Actions</b>	<b>Tier 1</b>	<b>Tier 2</b>	<b>Tier 3</b>
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to the history of the college. Loss would be unfortunate but not catastrophic. Every effort would be made to retain permanent copies. Significant effort or expense would be incurred to recover lost or corrupted files.

# Framework from 2012

Tiers/Preservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

# Progress as of 2018

Tiers/Preservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

# Points of failure

- Some things are beyond our control
  - Storage method and backup cycle provided by IT
- Some things are beyond our skill
  - Checksum validation across entire archive



# Compromise

Backup cycle and medium - whatever is provided by our IT department.

Checksum validation on 10% of holdings every 1-2 years.

- Not dependant on value.

# Progress as of 2018

Tiers/Perservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

# Progress as of 2018

Tiers/Perservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

# Progress as of 2018

Tiers/Perservation Actions	Tier 1	Tier 2	Tier 3
General Description	Critically important records to the institutional record and cultural heritage of the college. In some cases there may be legal requirements to keep materials indefinitely. Loss would constitute a major compromise in future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.	Important records to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not be as devastating as for Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.	Records useful to understanding the history of the college. Intent is to retain permanently, but loss would be unfortunate, but not a matter necessitating significant effort or expense to recover lost or corrupted files.
Fixity Checks	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.	Frequent checks, 2-3 times the life expectancy of the storage media. If any copies are stored on magnetic tape, check tape readability as well as fixity.
Preservation Metadata	High level of detail. Keep full PREMIS records.	Moderate level of detail. Fixity digest, format verification, creating software, compression level.	Minimal level of detail. Fixity digest, format verification.
Backup cycle	Frequent, redundant and geographically distributed. Redundant dark archive as part of a consortium (LOCKSS), frequency based on consortiums requirements. Backups under ITS program for current and active server directories.	Average frequency and distributed across campus. Backups under ITS program for current and active server directories.	Occasional cross campus backups.
Backup media	Spinning disk, as part of the dark archive. Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.	Spinning disk, on the Carleton network.
Migration	Aggressive with the ability to backtrack. If original copy is not in our most preferred preservation format, we will migrate one copy to this preservation format.	Moderate. No migration upon ingest if the original is in our list of supported digital formats. If original is not a supported format, and is accepted into archives, we will migrate original to the most preferred preservation format.	Passive. Migration of originals to most preferred preservation format only if there is a justifiable reason to do so.
Disposition of Original	Retain original, all derivative versions and all migrated versions.	Retention of most recently migrated version.	Retain only original or most recent migrated copy, whichever is newest.
Recovery Tests	Once with every major change to ITS's system structure. 1 time/year regardless of system changes.	Once with every major change to ITS's system structure. 1 time/ 5 years regardless of system changes.	Once with every major change to ITS's system structure.
Format Support	Most formats of original copy supported for as long as possible. Migration copies created with as much fidelity to the original as possible. Archives will work closely with creating departments to find solution that best preserves the integrity of the original file and the intellectual content of any migration copies in the future.	Prescribed list accepted formats based on preferred formats. Some unsupported formats acceptable with the understanding that migration to a preservation copy will probably be necessary and some information may be lost.	Prescribed list accepted formats based on preferred formats. Unsupported formats not accepted.

# When the Rubber Hits the Road: Real-World Digital Preservation

Nat Wilson  
Carleton College  
nwilson@carleton.edu

Conceptual Framework for Digital Preservation  
<https://goo.gl/Gmrjxv>

#MAC2018 #s302

# What Goes Where?: Measures to Decrease the Costs of Digital Storage

Sarah Dorpinghaus  
University of Kentucky



# Storage options



- Campus storage



- LTO tape



- AWS



- Google Drive



- DPN



- Internet Archive

# Partnerships & discounts



- Campus IT
- AWS
- Google Drive
- Internet Archive



# Preservation

A system based on decisions of what, where, and why?

## What goes where?

- Digitized for access or preservation?
- How many resources went into creating the digital copy?
- What quality should be kept?
- How will the materials be used?
- What are the risks?

# What goes where?



- Campus storage



- LTO tape



- AWS



- Google Drive



- DPN



- Internet Archive

# What goes where?



- AIP + DIP
- Complete backups
- Partial backups
- Complete backups
- Partial backups
- Newspaper + media

# Document

- Document policies, workflows, locations
- Develop growth estimates
- Build storage costs into annual budget

sarah.dorpinghaus@uky.edu  
@SMDorpinghaus

# Preserving Digital Oral Histories



## **DOUG BOYD**

Louie B. Nunn Center for Oral History

University of Kentucky Libraries

[doug.boyd@uky.edu](mailto:doug.boyd@uky.edu)

Twitter: [douglasaboyd](https://twitter.com/douglasaboyd)



# Louie B. Nunn Center for Oral History

In 2008

- 6000 interviews
- Majority still analog





# Louie B. Nunn Center for Oral History

In 2018

- 11,000+ interviews
- 100% digitized



**UK** Interview with Gilda Wabbit (Sam Themer), April 5, 2017

Louie B. Nunn Center for Oral History, University of Kentucky Libraries  
Lance Poston, Interviewer | 2017oh232\_lgbtq001  
Kentucky LGBTQ Oral History Collection | [Kentucky LGBTQ Oral History Project](#)

1:53 / 32:02

0:01 - Connection to Kentucky through the arts  
2:32 - Coming out  
5:55 - University of Kentucky vocal music program  
8:05 - Safe spaces in the Lexington community  
13:26 - The creation of "Gilda Wabbit"  
19:11 - Involvement in the "Future That Liberals Want" meme  
26:01 - Current and future work

Transcript  Index

Search this Index

keyword  Go



# Louie B. Nunn Center for Oral History

In 2017

- **36** concurrent **interviewing projects**
- **917** new **interviews** accessioned in 2017
  - 733 Audio
  - 176 **Video**
- **116** interviews conducted in the Nunn Center **studio**



# Audio AND Video





# Challenges

- Storage

Rapid Growth

100% digitized

Video





## Audio AND Video

- Master  
AVCHD  
**30 gb / hour**
- Preservation Master  
Apple ProRes HQ  
**100 gb / hour**





## Audio AND Video

- SIP Size for 1 interview could be **250 gb**



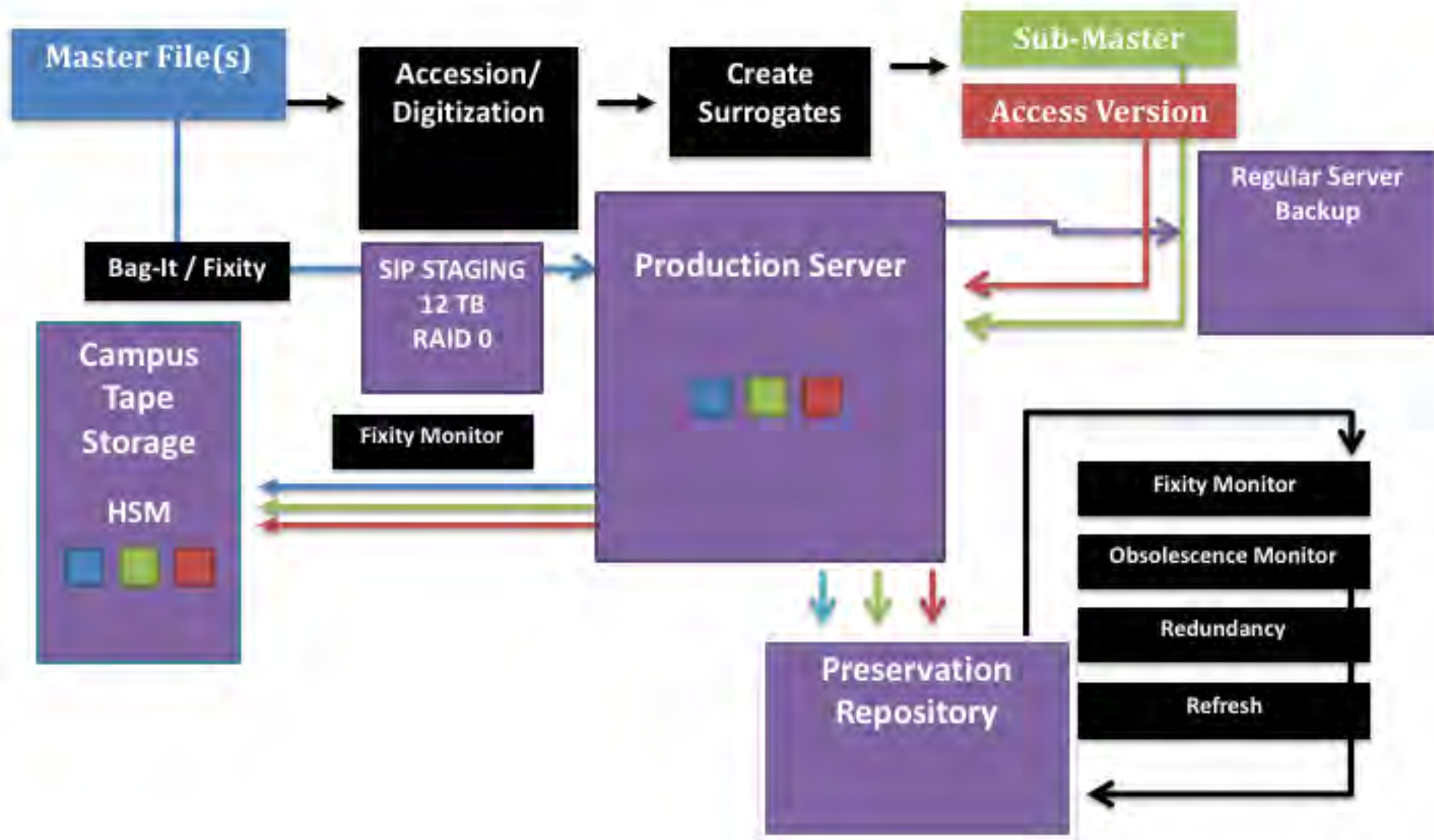


# Locations

- Storage
  - Servers
  - Preservation Repository
  - Campus Tape System (HSM)

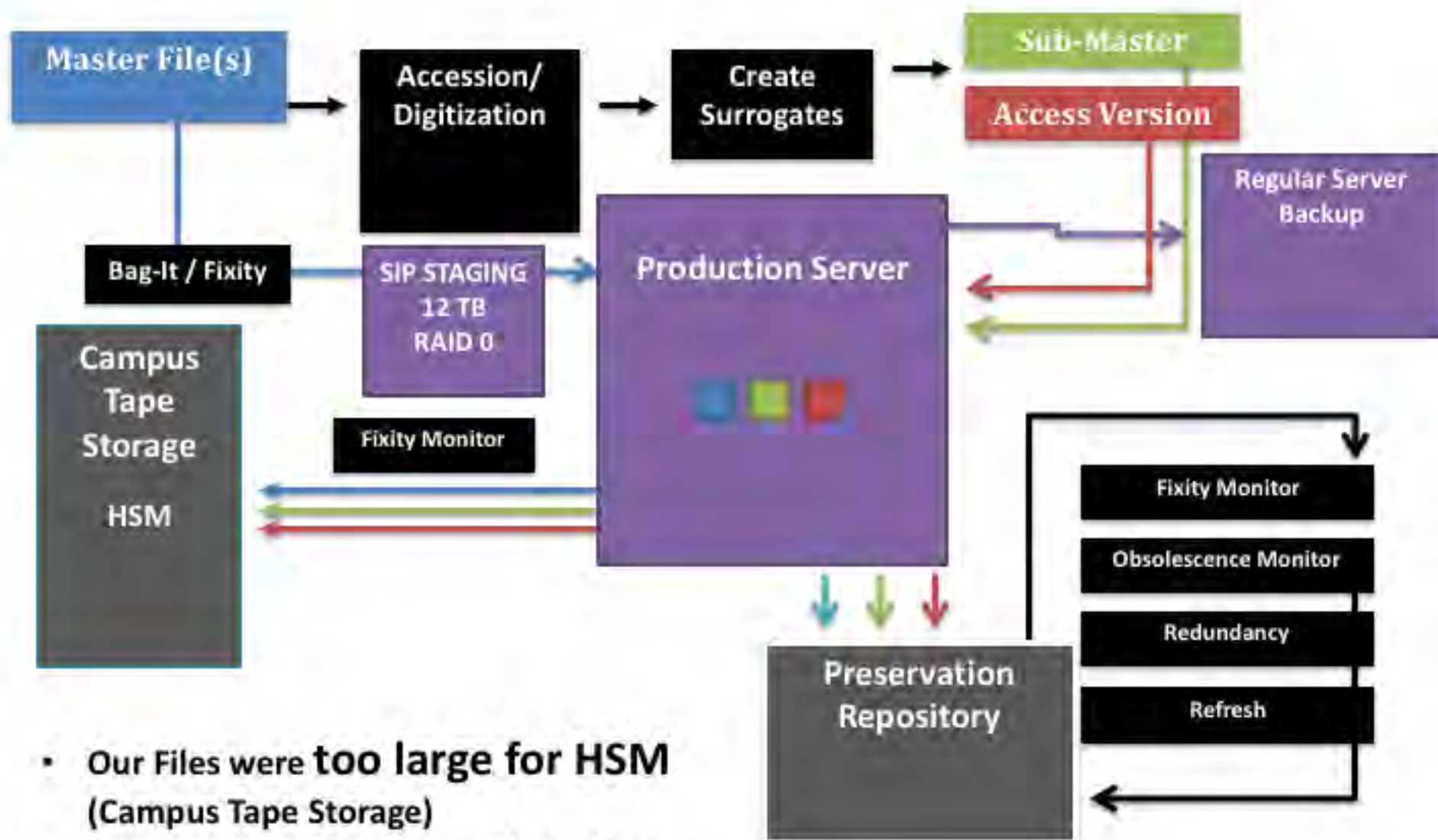


# PRESERVING DIGITAL ORAL HISTORY



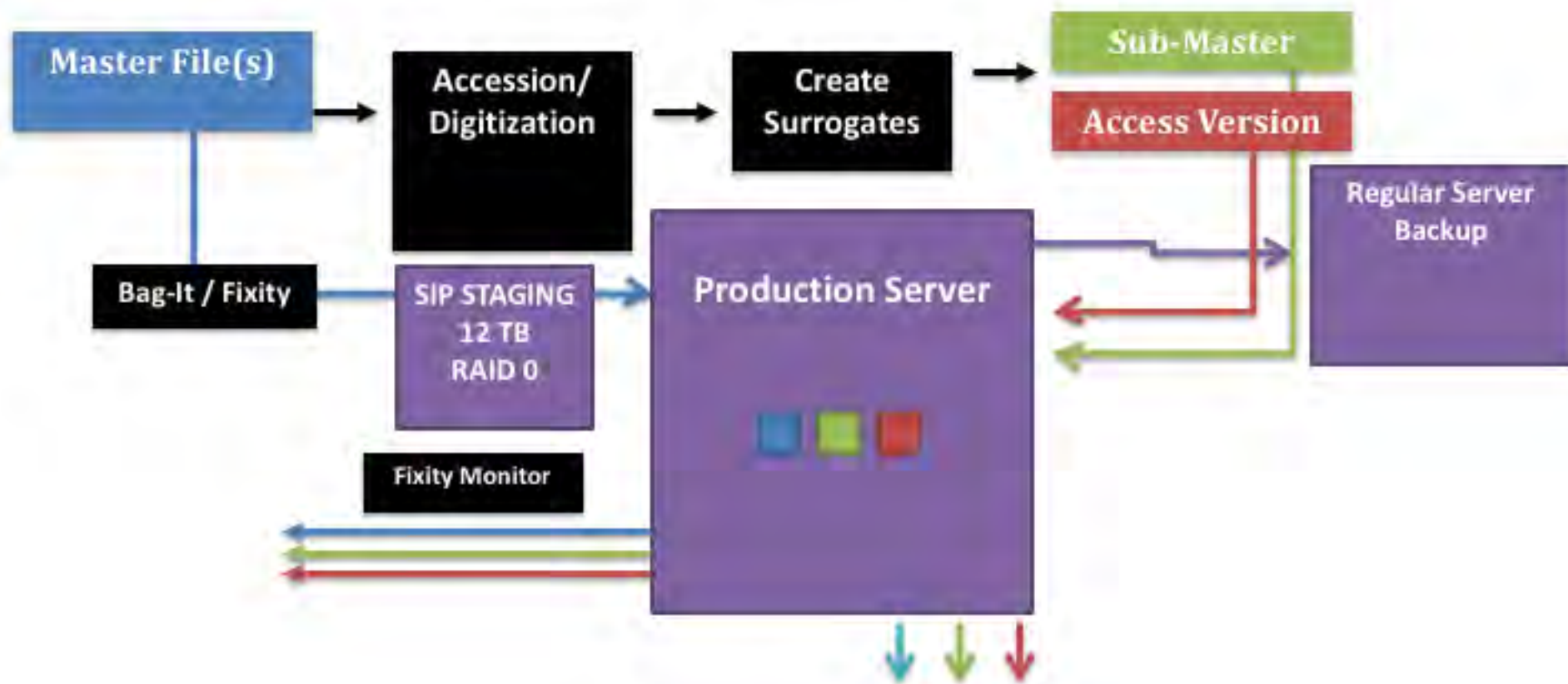


## PRESERVING DIGITAL ORAL HISTORY

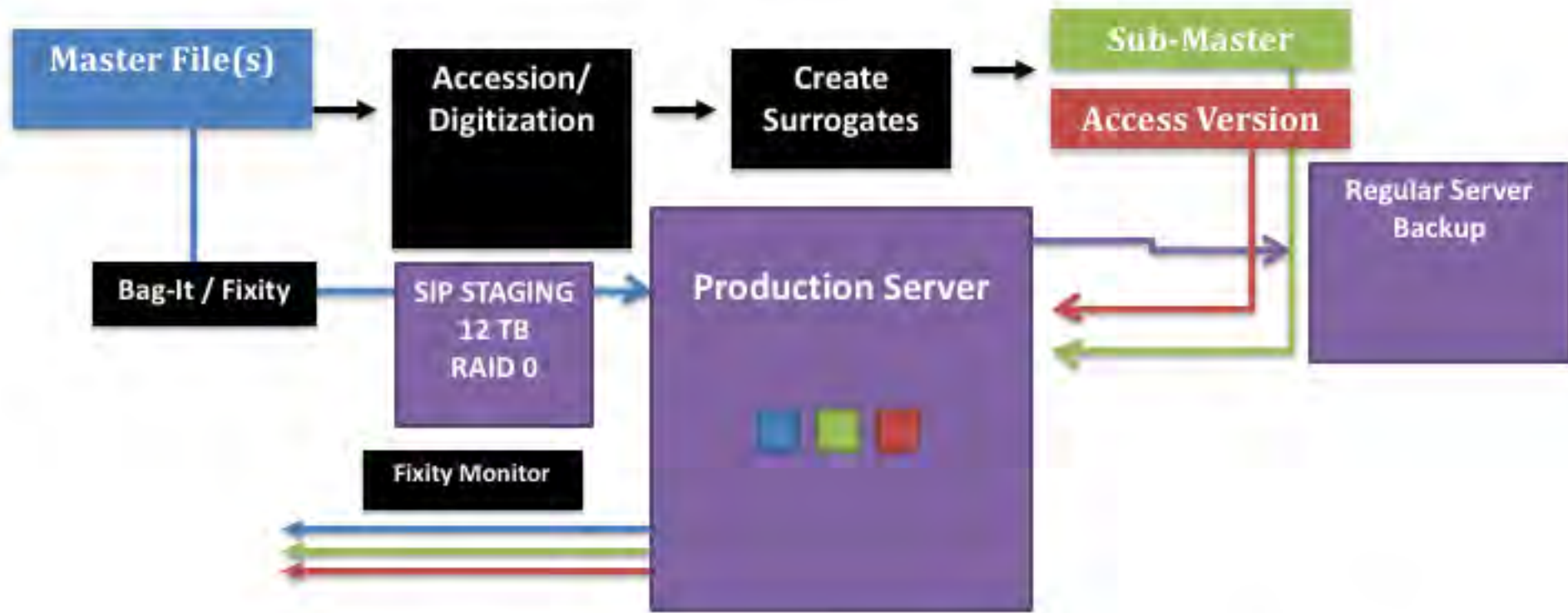


- Our Files were **too large** for HSM (Campus Tape Storage)
- Our Files were **too large** for the Preservation Repository

# PRESERVING DIGITAL ORAL HISTORY



## PRESERVING DIGITAL ORAL HISTORY

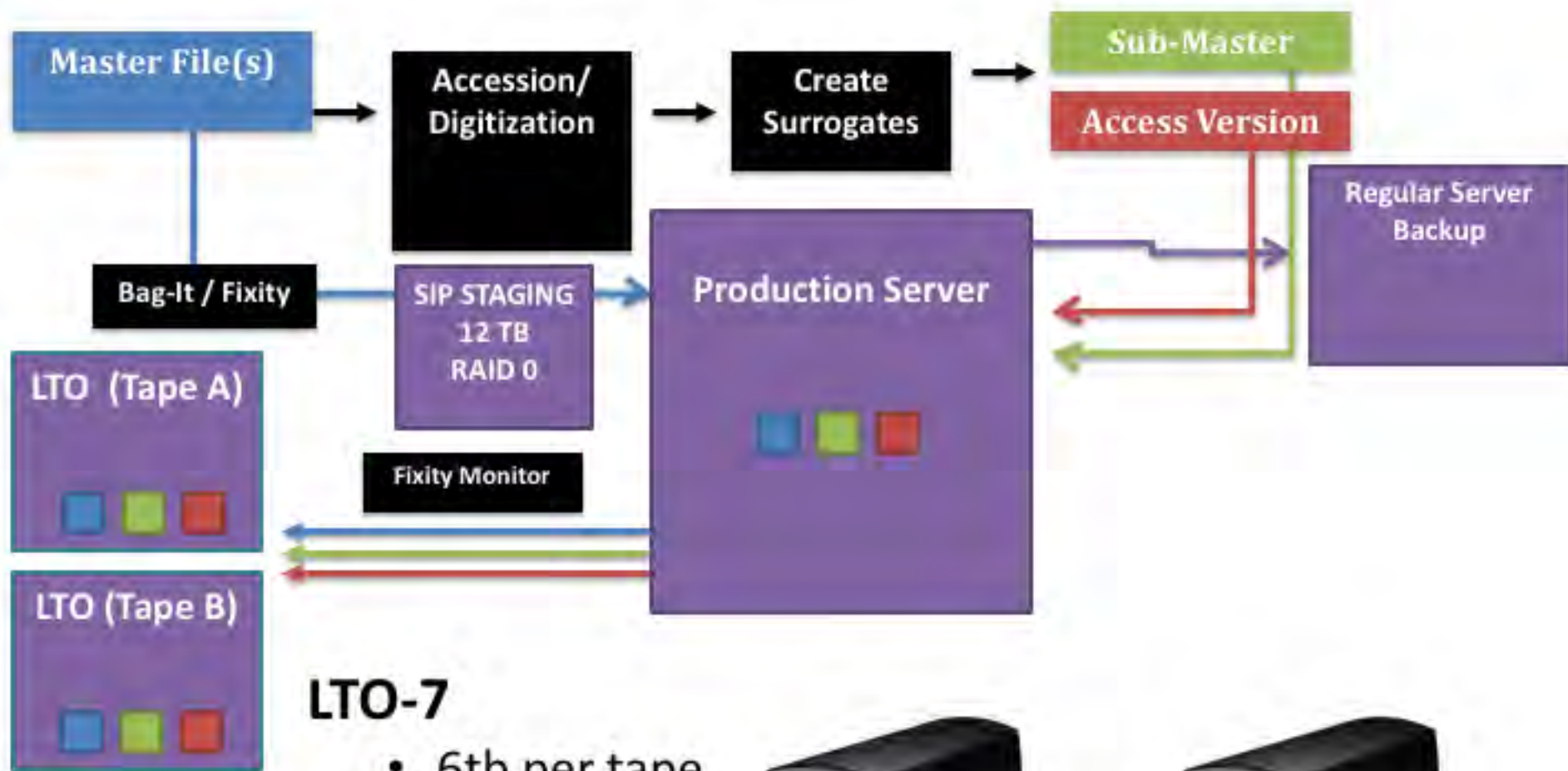


## LTO-7

- 6tb per tape



# PRESERVING DIGITAL ORAL HISTORY



## LTO-7

- 6tb per tape





# Louie B. Nunn Center for Oral History

## LTO Tape Storage

- Snapshot of entire collection (2 sets as of 12/2017)
- Each new accession
- 2 Simultaneous Versions Created
  - 1 stored in archival storage
  - 1 stored offsite





# Louie B. Nunn Center for Oral History

## LTO Tape Storage

- Metadata: Tracking what & where
- Refresh schedule
- LTO-8 (obsolescence cycle)





# Louie B. Nunn Center for Oral History

## LTO Tape Storage

- Metadata: Tracking what & where
- Refresh schedule
- LTO-8 (obsolescence cycle)
- \$150 / 6 TB x 2





# Louie B. Nunn Center for Oral History

## LTO Tape Storage

Software

PreRollPost

Itfs

thunderbolt

inventory + exported csv

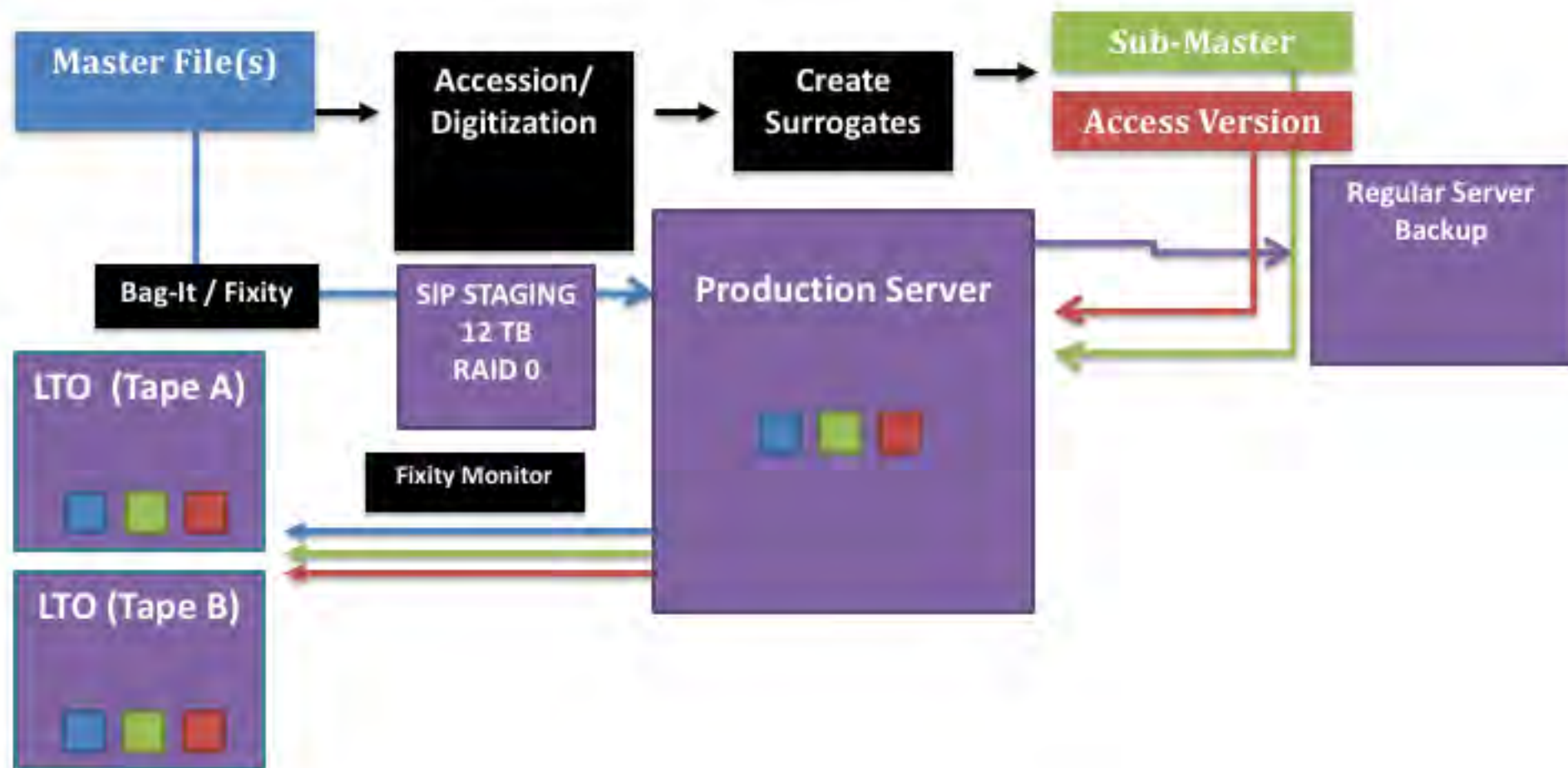
fixity

tapes *not* software dependent

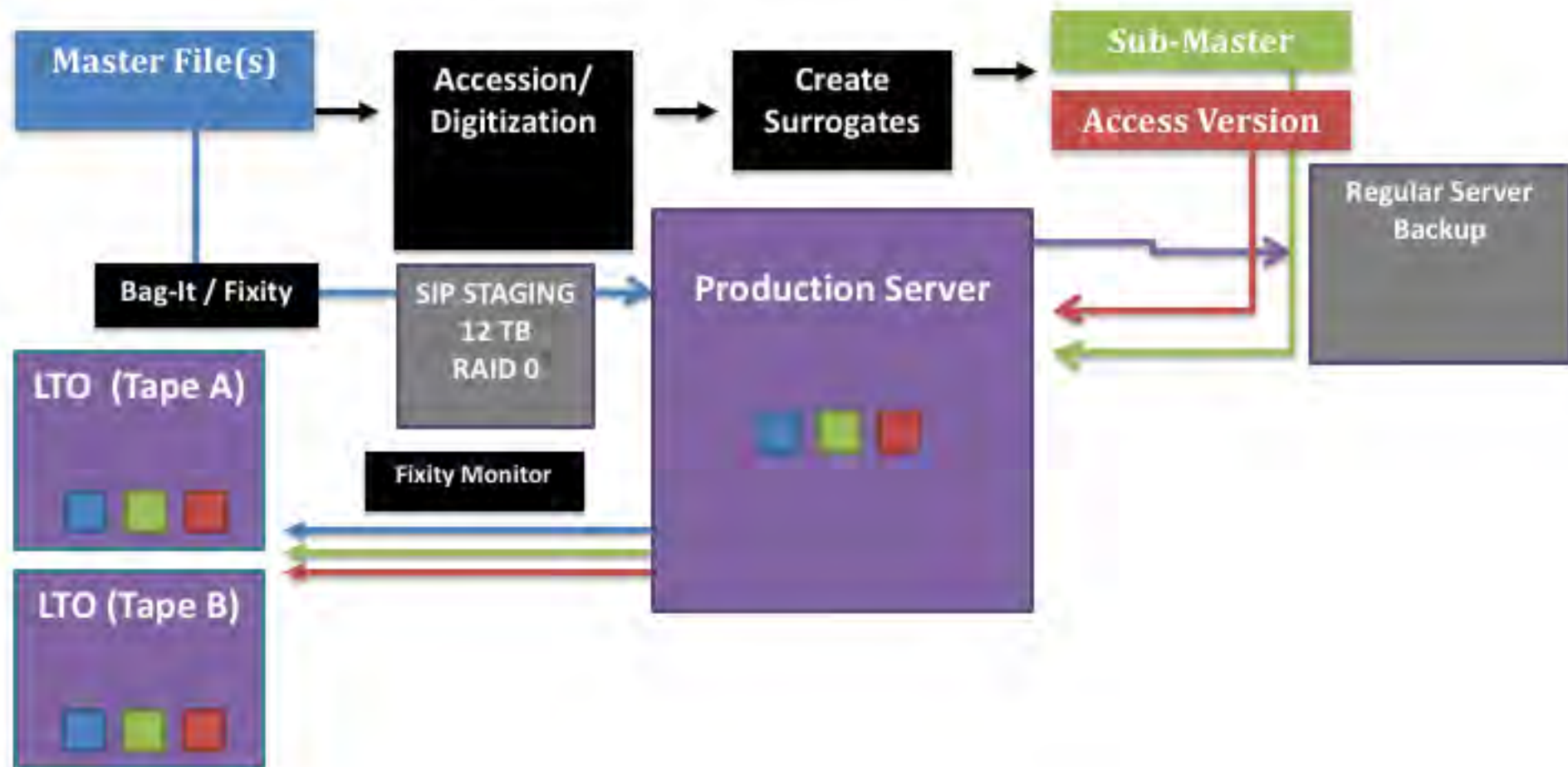




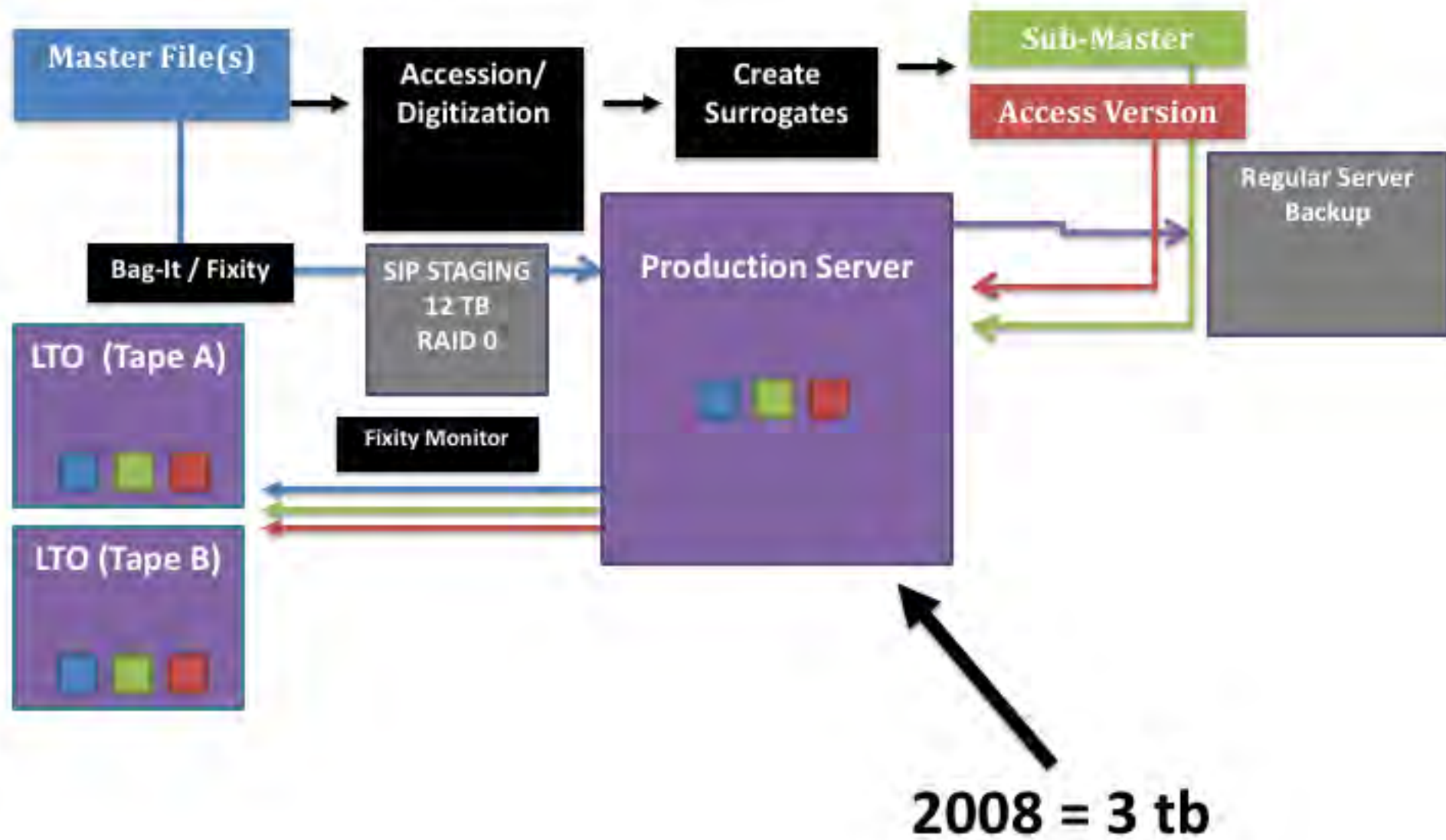
# PRESERVING DIGITAL ORAL HISTORY



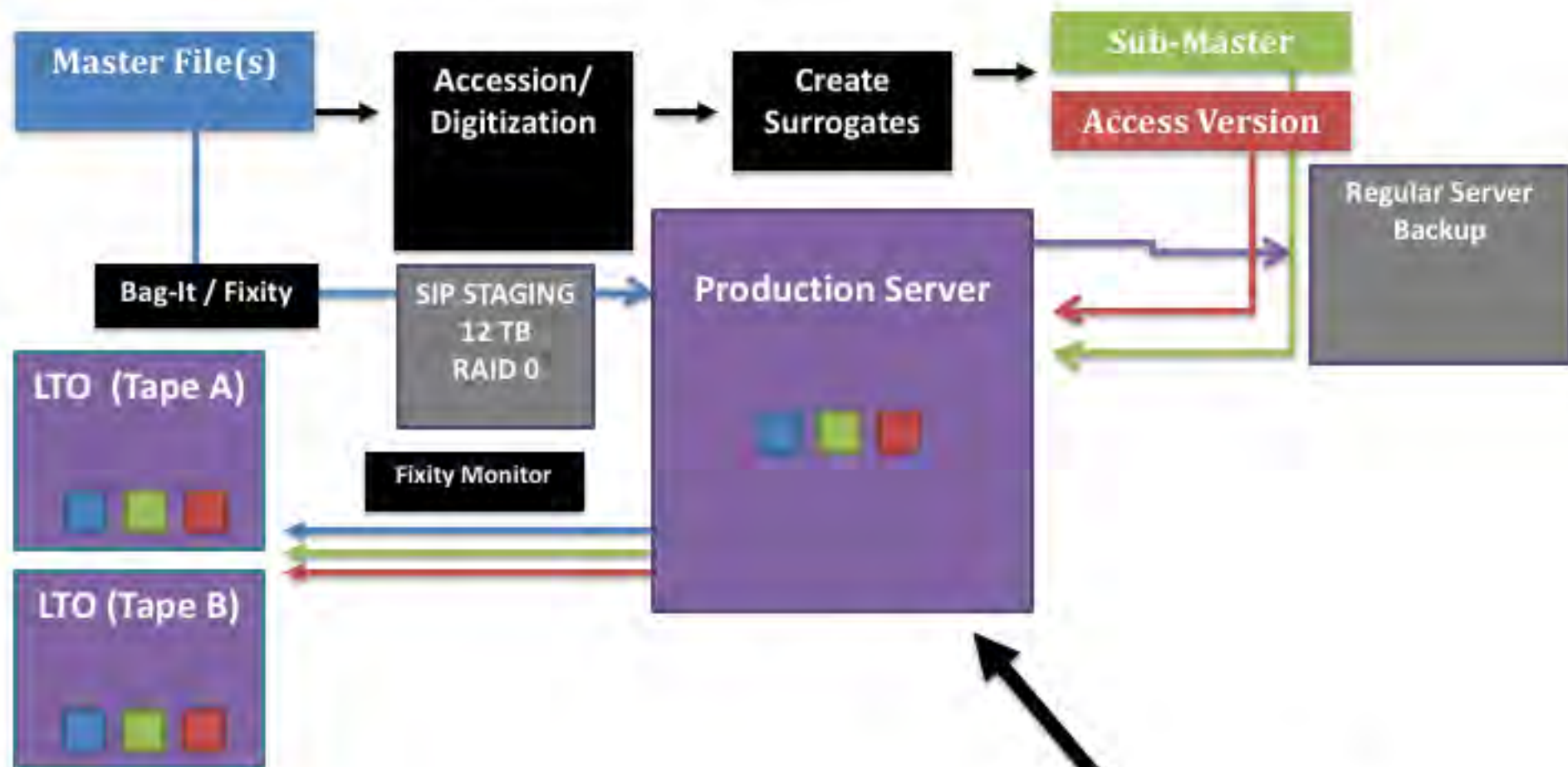
# PRESERVING DIGITAL ORAL HISTORY



# PRESERVING DIGITAL ORAL HISTORY

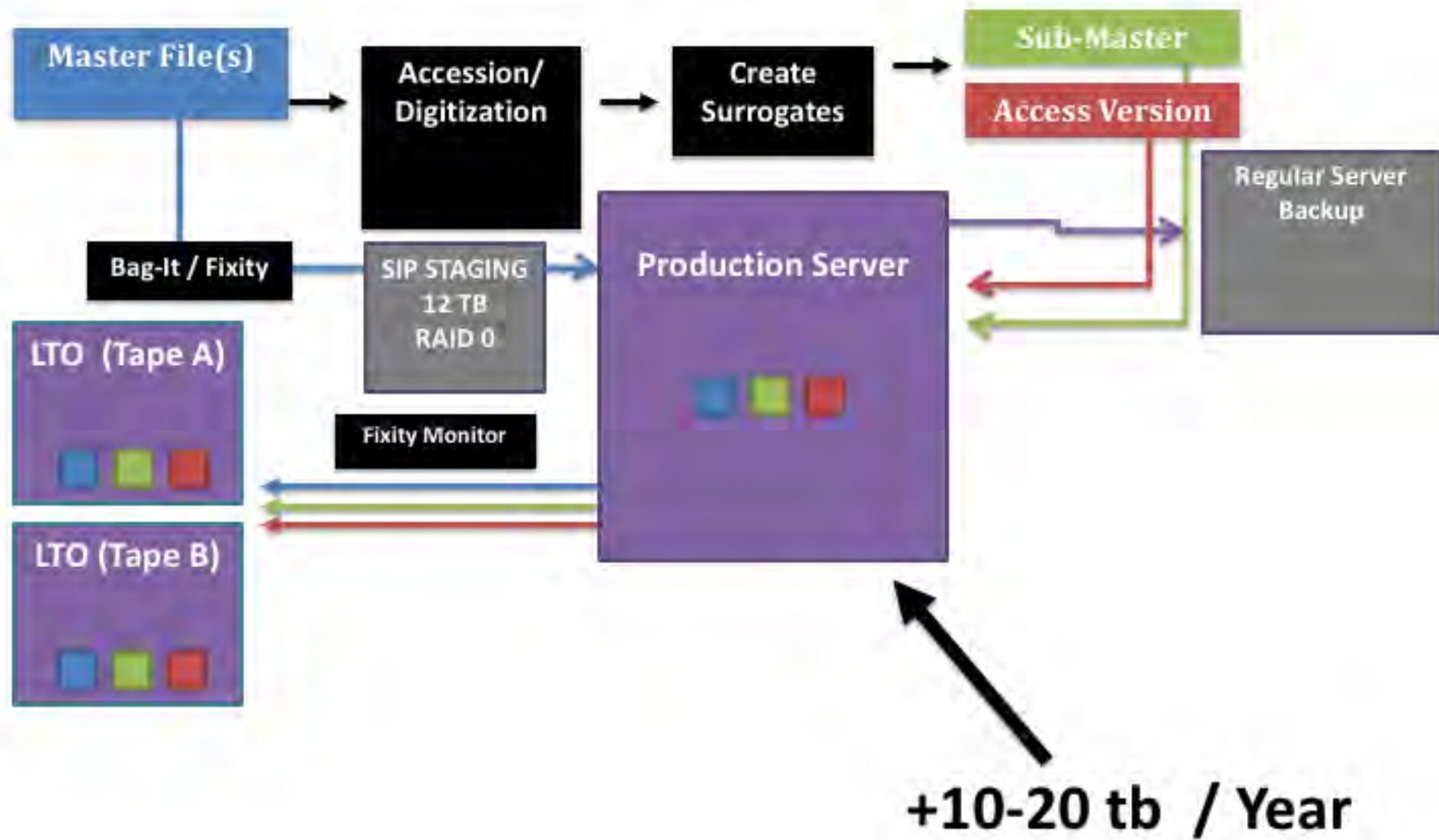


# PRESERVING DIGITAL ORAL HISTORY

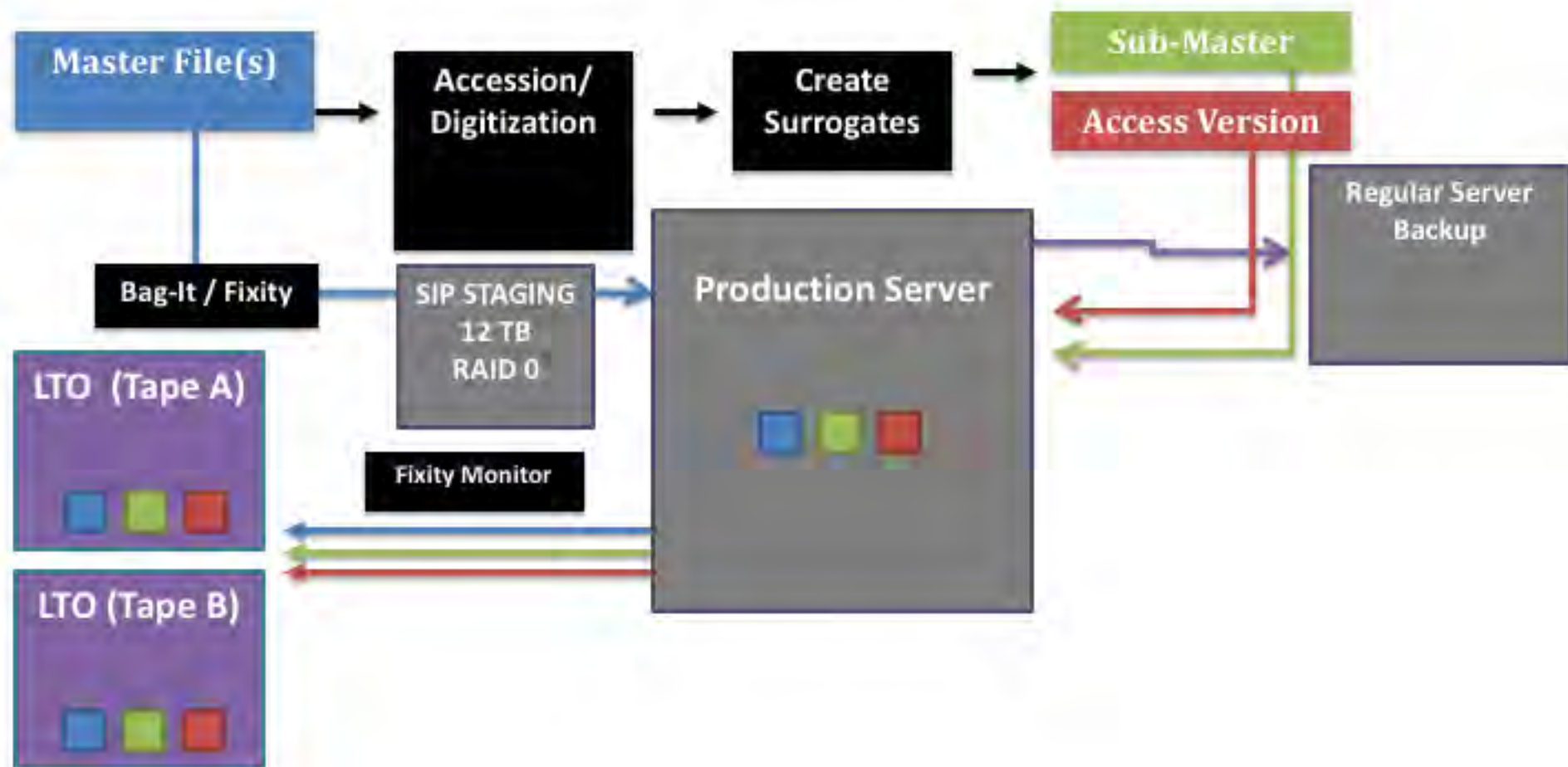


2018 = 117 tb

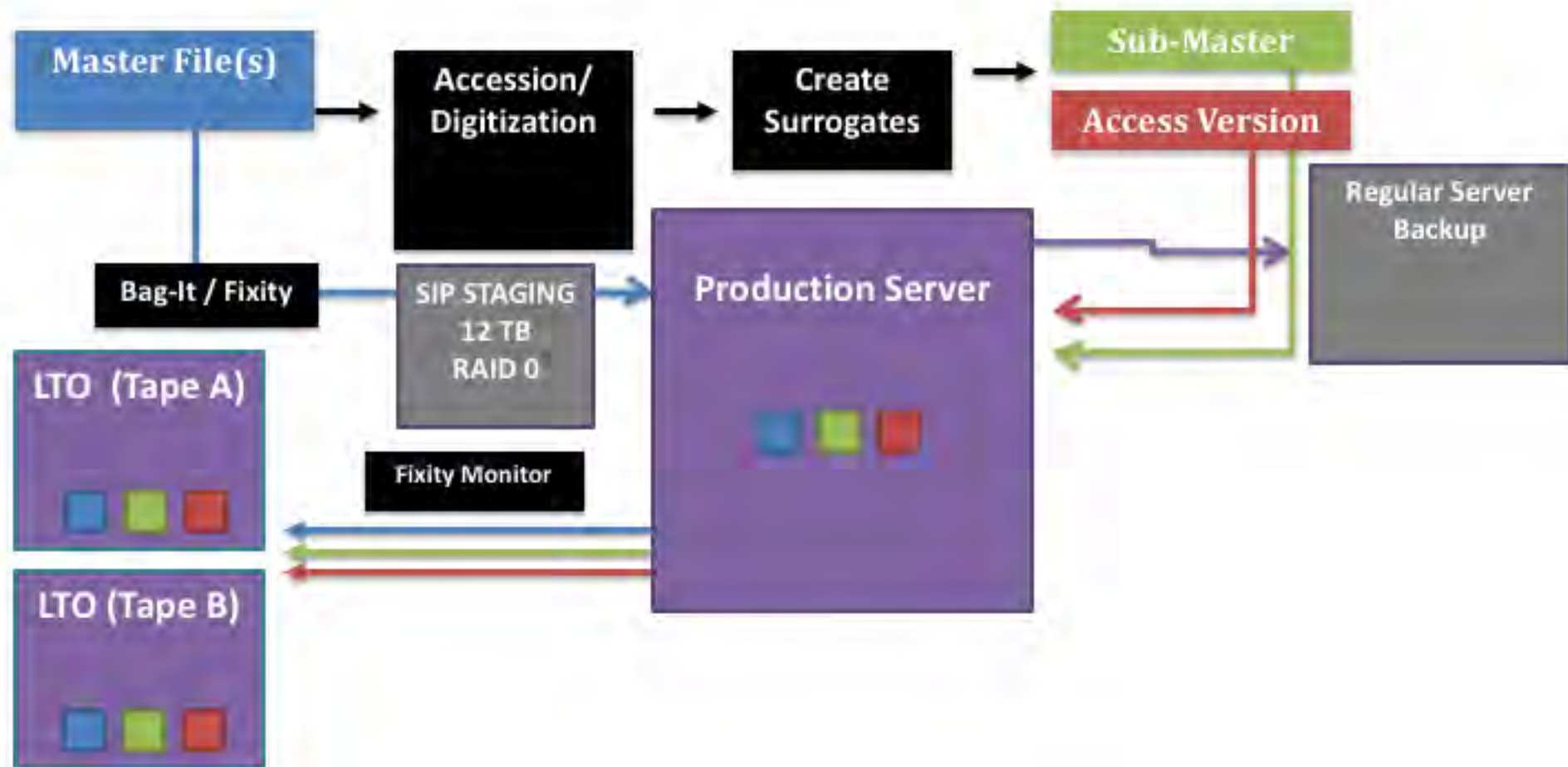
# PRESERVING DIGITAL ORAL HISTORY



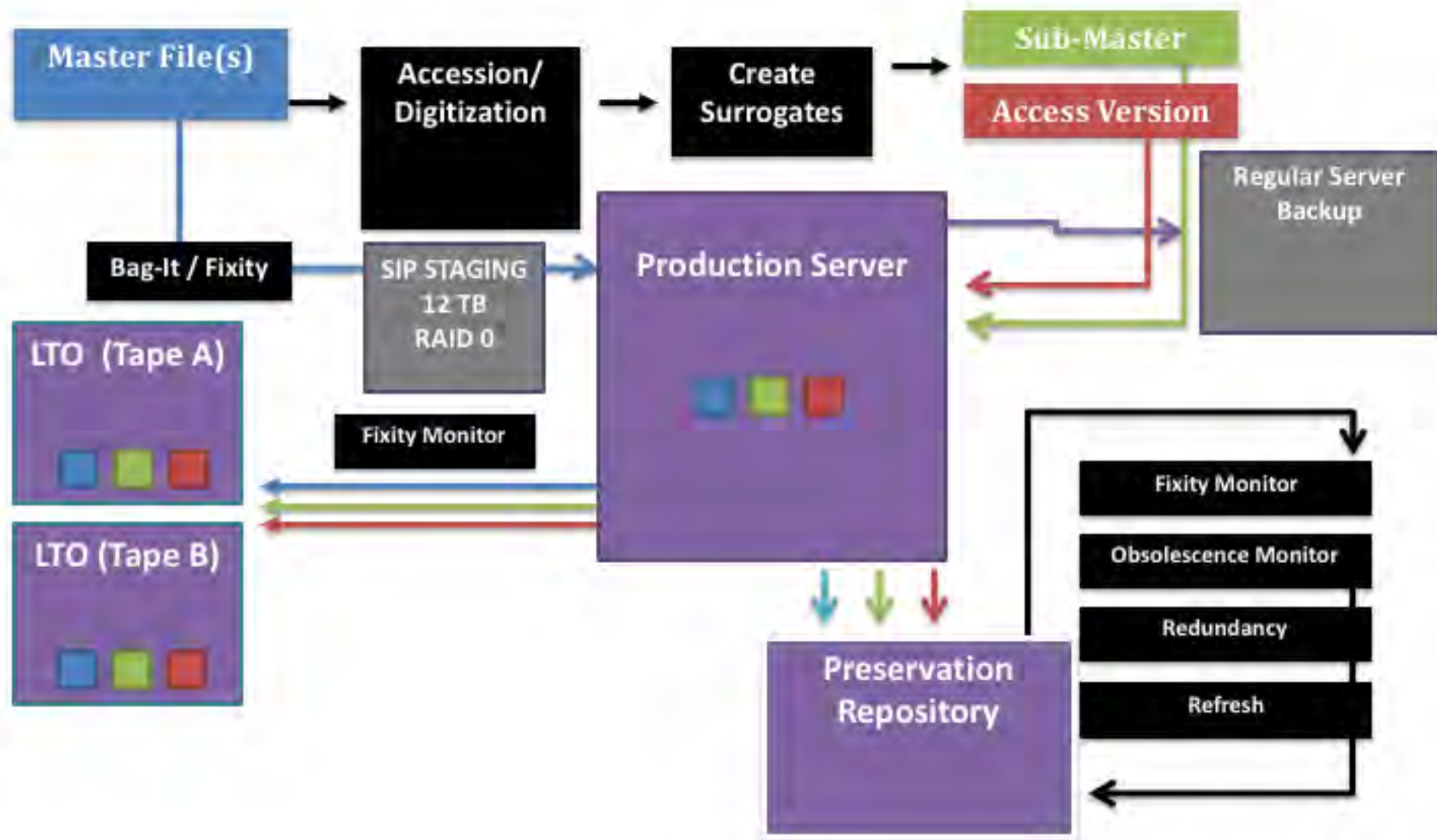
# PRESERVING DIGITAL ORAL HISTORY



# PRESERVING DIGITAL ORAL HISTORY

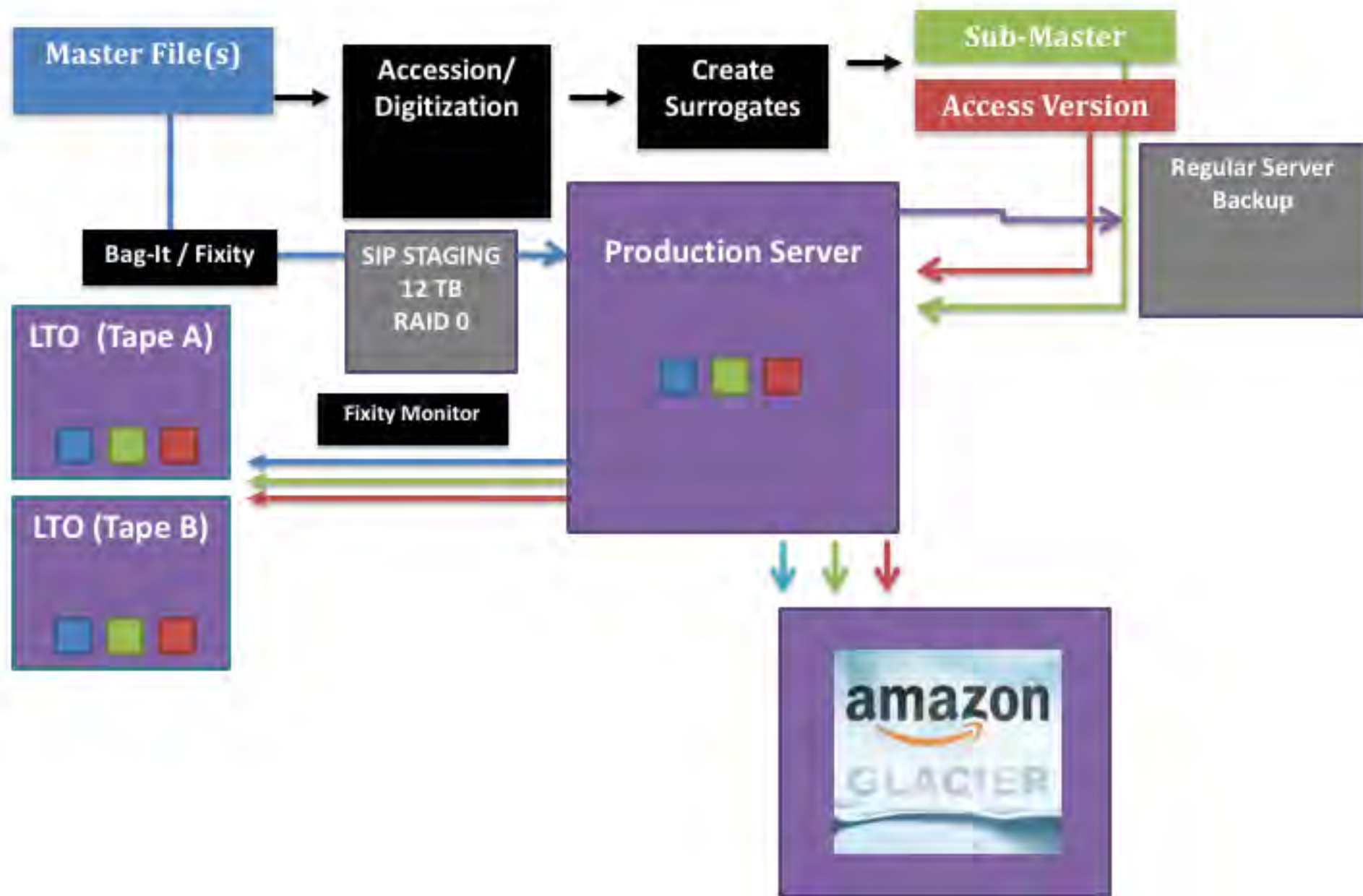


# PRESERVING DIGITAL ORAL HISTORY

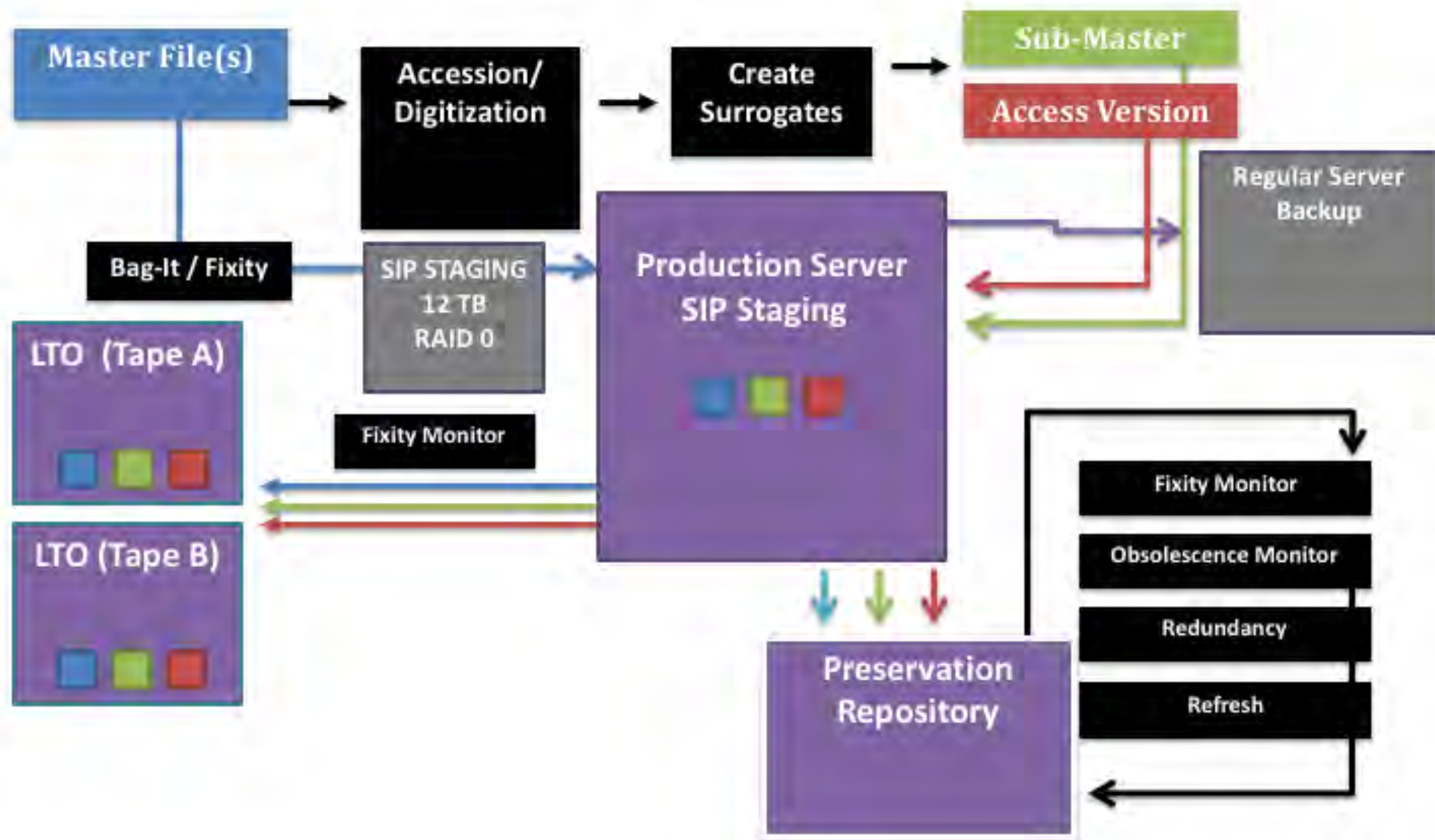




# PRESERVING DIGITAL ORAL HISTORY



# PRESERVING DIGITAL ORAL HISTORY



# Preserving Digital Oral Histories

## **DOUG BOYD**

Louie B. Nunn Center for Oral History

University of Kentucky Libraries

[doug.boyd@uky.edu](mailto:doug.boyd@uky.edu)

Twitter: [douglasaboyd](https://twitter.com/douglasaboyd)



# From Here to There:

Integrating Open Source Tools in Digital  
Preservation Workflows

Mike Shallcross

[http://bentley.umich.edu/  
shallcro@umich.edu](http://bentley.umich.edu/shallcro@umich.edu)

Midwest Archives Conference Annual Meeting  
March 23, 2018

# Overview

- Key assumptions for BHL digital preservation workflow development.
- Iterations of workflow development, 1997-
  - Staffing
  - Processes and workflows
  - Infrastructure
- Some lessons learned

# Key Assumptions in BHL Workflow Development

- The most pressing/immediate issues should be addressed first.
- Workflows should be informed by standards (esp. OAIS) if/when possible.
- Communities of practice provide value and support.
- Solutions should allow for alternative options in the future.
- We are going to make mistakes, but will hopefully improve!

# 1997-2009: First Steps

- Staffing:
  - .5 FTE, limited external IT support
  - Limited expertise: GUI applications
- Processes and workflows:
  - Boutique projects
  - Highly specialized/manual processes
- Infrastructure:
  - Workspace: personal computer
  - Storage: optical media, website

# Integrations, Pt. I (2010-2011)

- Staffing:
  - 2 FTE filled by grant funding (Mellon Foundation)
  - Developing skills; explored 60+ tools and reviewed peer institutions' workflows
- Processes and workflows:
  - Standardize and scale digital preservation strategies (format characterization, defined migration pathways, PII detection, etc.)
  - Developed a more robust—but still manual—workflow
- Infrastructure
  - Workspace: Personal computers
  - Storage: UM network drive (short term) & DSpace (long term)



# Integrations, Pt. II (2011-2013)

- Staffing:
  - New Digital Curation Division (formed April 2011); 2 dedicated FTE.
  - Limited tech skills: GUI/CLI applications & basic shell scripting.
- Processes and workflows:
  - Automated key steps with Windows CMD.EXE shell scripts.
  - Microservice design (influenced by Archivematica).
    - Easy to adapt scripts or introduce custom workarounds.
- Infrastructure:
  - Processing and backlog storage: NAS box.
  - Homegrown systems (EAD creation, Filemaker database for accessions/locations).

# Integrate This!

```
*****
Bentley Historical Library * * * Digital Curation Division
AutomatedProcessor v. 1.0 (c) 2014
*****

MAIN MENU OF PROCEDURES:

1. INITIAL SURVEY
2. PII SCAN
3. IDENTIFY MISSING FILE EXTENSIONS
4. FORMAT CONVERSION
5. ARRANGEMENT, PACKAGING, and DESCRIPTION
6. TRANSFER and CLEAN UP

Q: Quit the BHL Automated Batch Processing Sequence

*****

AutoPro may be completed in one session or may be stopped at any point.
Quitting an automated procedure mid-operation may result in loss of data.
Please complete a procedure before exiting the sequence.

*****

PROCESSING STATUS--DEPOSIT 943_0001 (as of Mon 03/30/2015):

Completed: Preliminary Procedures
Completed: 1. INITIAL SURVEY procedure 1 of 6
Completed: 3. IDENTIFY MISSING FILE EXTENSIONS procedure 3 of 6
Completed: 4. FORMAT CONVERSION procedure 4 of 6
Completed: 5. ARRANGEMENT, PACKAGING, and DESCRIPTION procedure 5 of 6

*****

Enter an option:
```

# Integrations, Pt. III (2014-2016)

- Staffing
  - 3 FTE for Mellon grant (and additional support for ASpace implementation).
  - Development of technical skills (Python/Ruby, GitHub, XML parsing, etc.).
  - Increased collaboration and improved communication with IT staff.
- Processes and workflows:
  - Facilitate creation/reuse of metadata.
  - Streamline the ingest and deposit of content into a repository.
  - Introduce new appraisal and review functionality in Archivematica.
- Infrastructure
  - Storage (short/long term): UM network drives (with tape backup)
  - Web-based interfaces: ArchivesSpace (collections management, authoritative metadata), Archivematica (ingest and packaging of content), DSpace (preservation/access).
  - Exploit native APIs and SWORD protocol to transfer data/metadata among systems.

Hey! You got appraisal functionality in my Archivemática!

The screenshot displays the Archivemática web interface. At the top, the navigation bar includes 'Transfer' (with a red badge '6'), 'Backlog', 'Appraisal' (highlighted), 'Ingest' (with a red badge '12'), 'Archival storage', 'Preservation planning', 'Access', 'Administration', and 'shallcro'. Below the navigation bar, there are search filters and an 'Add New' button. A 'Tags' dropdown menu is set to 'All'. The main content area is divided into several sections:

- Backlog:** Shows '0 transfers, 5673 objects selected'. It includes a search box, an 'Add tag to selected files' button, and buttons for 'Deselect all', 'Collapse all', and 'Expand all'.
- File list:** Displays a tree view of a folder structure:
  - 2008155\_0001-391eb6a8-ea72-4cfd-a7a2-9bf2971e4627
    - logs
    - metadata
    - objects
      - BANCO2008
        - BANCO.whtt
        - BANCO
        - backblue.gif
        - fade.gif
        - index.html
      - BANCO2009

- Visualizations:** A pie chart titled 'Formats (by total number of files)' and 'Formats (by total size of files)'. The legend includes:
- 1989a (fmt/4)
- Generic CSS (x-fmt/224)
- Generic LOG (x-fmt/62)
- Generic TXT (x-fmt/111)
- HTML pre-2.0 (fmt/96)
- Hypertext Markup Language (fmt/100)
- JPEG (fmt/44)
- JavaScript file (x-fmt/423)
- PNG (fmt/11)
- Text
- Configuration file (x-fmt/421)
- XHTML (fmt/102)
- XML (fmt/101)
- ZIP file (x-...)

# Integrations, Pt. IV (2017-)

- Staffing
  - Archivist for Digital Curation, Archivist for Metadata and Digital Projects, and two project archivists.
  - Increasing expertise with Python (<https://github.com/bentley-historical-library>); biweekly, .5 hr technical skills workshops for Curation staff.
- Processes and Workflows:
  - Developing alternative workflows for legacy removable media, A/V materials, and donor-digitized materials as well as accessions with donor-supplied metadata.
  - Using ASpace as system of record for Archive-It metadata.
- Infrastructure
  - New ASpace API endpoints for use in digitization and location management.
  - Select backups in DPN.

# Lessons Learned

- Staffing
  - Find opportunities for staff to grow and develop.
  - “No one cooks the bacon alone” (Erin O’Meara)
- Processes and workflows
  - Iterative approach
  - Scalability and flexibility
  - Avoid unnecessary complexity: loosely-coupled, fine-grained services.
- Infrastructure
  - What do you and users need? What can you sustain?
  - Foster relationships with IT support (internal/external and formal/informal)
  - Open Source Software: interoperability, customizability, community support, flexibility

InDiPres:  
Distributed  
Digital Preservation  
for the Hoosier State



“When the Rubber Hits the Road”

MAC s302 / Chicago / March 23, 2018

Cinda May, Indiana State University Library

# The Problem



Henry J. Schroder, 1914. Courtesy of Helen (Fox) Julian, Vincennes, IN

- A plethora of digitization projects supported by LSTA grants, 2006-2017
  - Indiana Memory: 484,736 items
  - Hoosier State Chronicles: 500,000+ newspaper pages
- The need for an affordable, cost-effective digital preservation solution for small to mid-sized under-resourced cultural memory organizations



# The Solution, Part 1: The MetaArchive Cooperative

## What is the MetaArchive Cooperative?

- A digital preservation network created and hosted by and for memory organizations
- Established in 2004 by 6 academic libraries in cooperation with the Library of Congress NDIIPP Program
- Currently incorporates 15 secure, closed-access preservation nodes and preserves more than 200 TB of content for 60+ members

## Collaborative Membership Level

- Allows new or existing consortia to join as one entity operating a single LOCKSS server
- Cost based on 20 participating institutions
- Annual fee \$2,500 + \$100 per institution + storage (\$0.59/GB ; \$585/TB)



<https://metaarchive.org>

# The Solution, Part 2: Hoosiers for Distributed Digital Preservation

- Form a membership based group of cultural memory organizations
- The group joins the MetaArchive Cooperative as a Collaborative Member
- Indiana State Library serves as the lead institution
- Indiana State Library Foundation serves as the fiscal agent
- Indiana State University Library serves as the site of LOCKSS box (i.e. network cache)



Courtesy of the Sullivan County Public Library

# Creating Indiana Digital Preservation (InDiPres)

<http://www.indipres.org>

- Initial funding through IMLS/LSTA grants
- Working group to develop membership agreement & outline governance elements
- 8 Open Forums on digital preservation
- Foundational governance meeting
- MetaArchive membership & server purchase
- Inaugural membership meeting
- Ingest Pathways Working Group
- InDiPres Guidance Document & Technical Appendix
- Digital Preservation Policy Creation Workshop
- Ingest of members' content into the MetaArchive Digital Preservation Network



# InDiPres Membership Fee Schedule Based on a Minimum of 20 Participants

Individual Participation Fee	\$100.00/year
Share of Server Cost : (3 year replacement cycle, \$6,000/3yr/20=\$100)	\$100.00/year
Share of MetaArchive Collaborative Membership Fee (\$2,500/yr/20 participants=\$125)	\$125.00/year

Total \$325.00/year

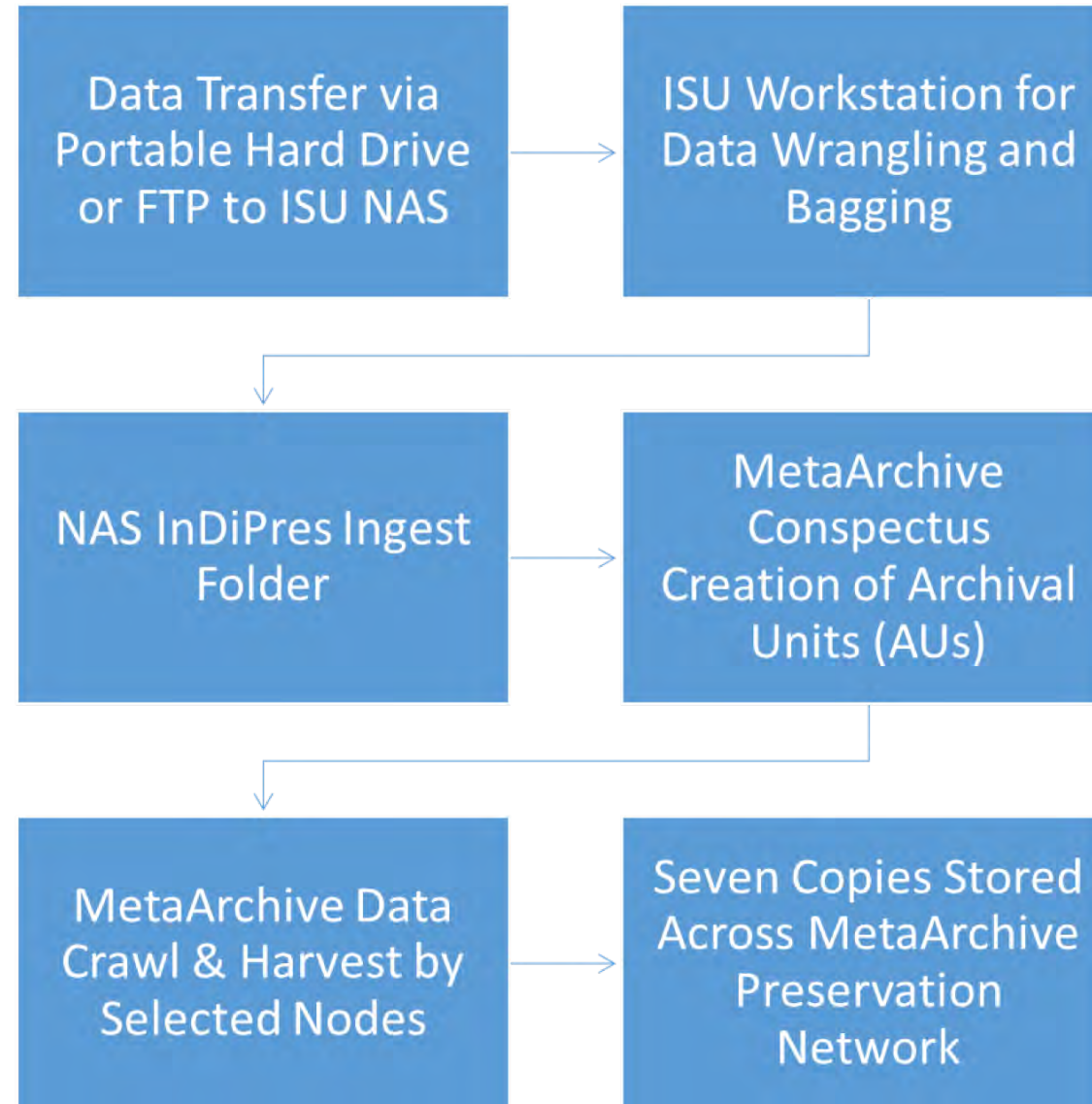
+ Individual storage fee = \$0.59/GB/year

# Current InDiPres Membership



- American Legion Auxiliary
- Bartholomew County Public Library
- Butler University Libraries
- DePauw University Libraries
- Indiana State Library (Host)
- Indiana State University Library (Host)
- Indianapolis Public Library
- Knox County Public Library
- Lebanon Public Library
- Private Academic Library Network of Indiana
- Parke County Public Library
- Logan Library, Rose Hulman Institute of Technology
- Sisters of Providence of Saint Mary-of-the-Woods Archives
- Sullivan County Public Library
- Vigo County Historical Society
- Vigo County Public Library

# InDiPres Members' Content Ingest Pathways & Workflow



# Challenges



Courtesy of the Vigo County Historical Society

- Raising awareness of the difference between preservation for access and digital preservation
- Forming a self-sustaining, collaborative organization with a governance structure
- Building the membership
- Involving members in governance committees and activities
- Providing timely ingests
- Providing skill development training
- Creating local digital preservation policies and plans

Thank You for Your Interest!



Iva Allison, ca1915, courtesy of Willoughby Steckley, Vincennes, IN

Cinda May  
Chair, Special Collections  
Indiana State University Library  
812-237-2534  
[Cinda.May@indstate.edu](mailto:Cinda.May@indstate.edu)



Laura Alagna  
Nat Wilson  
Sarah Dorpinghaus  
Doug Boyd  
Michael Shallcross  
Dan Noonan  
Cinda May

# Thank you!

Please fill out the survey feedback form:



THE OHIO STATE  
UNIVERSITY

## BALANCING THE IDEAL VS. THE REAL



#MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

Good Morning! We have been in a prolonged process at THE OSU to create a preservation and access environment for our University Libraries born-digital and converted content.

**THE OHIO STATE UNIVERSITY**  
UNIVERSITY LIBRARIES

## WELL HOW DID WE GET HERE?

**Knowledge Bank**  
The Ohio State University

**MEDIA MANAGER**  
College of Arts and Sciences

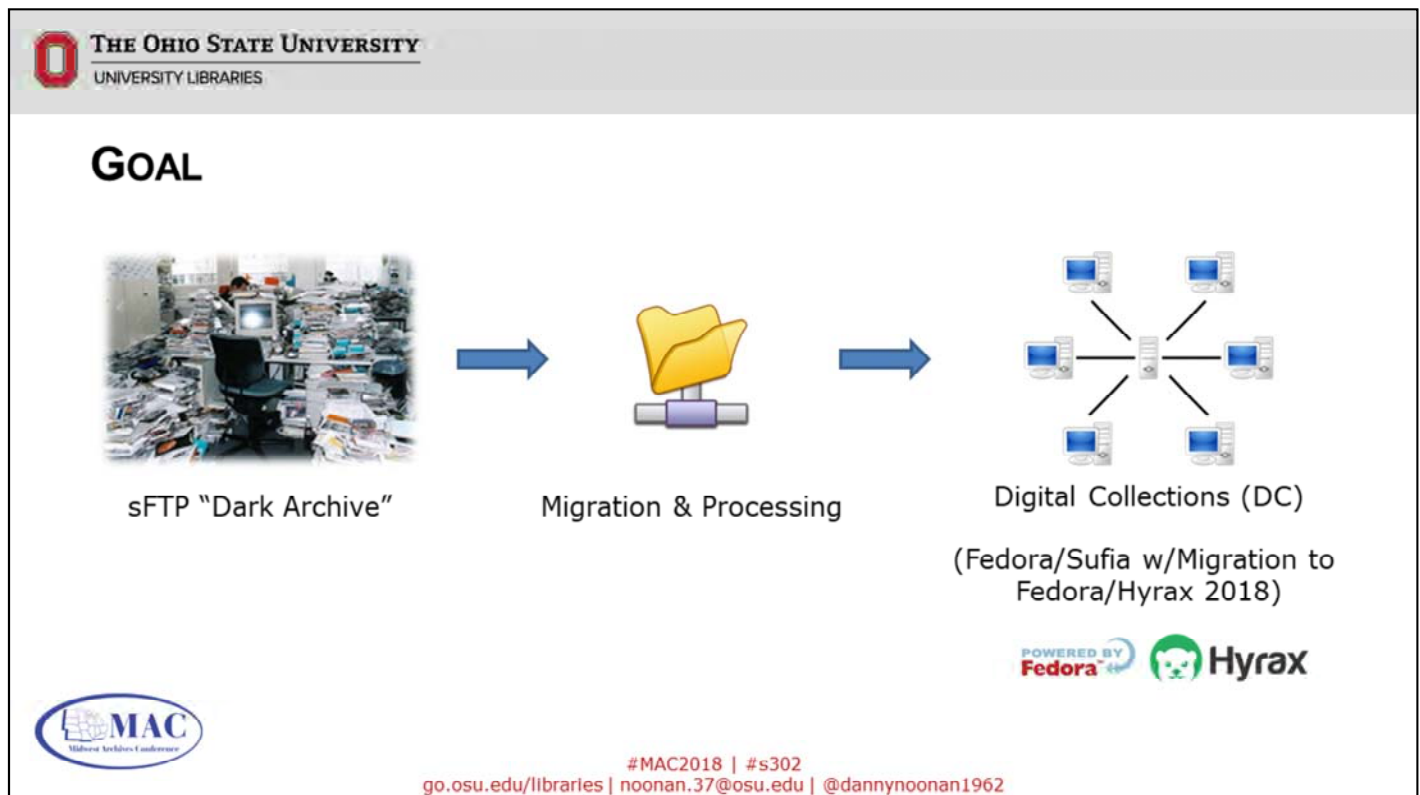
**MAC**  
Midwest Archives Conference

#MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

For years we had been reliant upon our DSpace instance known as the KnowledgeBank or KB., a home-grown OSU—albeit, a College of Arts and Sciences owned—platform, The Media Manager, and what we lovingly refer to as our “Dark Archive”.

However, there were (are) several issue with each of these solutions:

- The KB was designed more as an institutional repository for university scholarly output; to which we only guarantee bit level preservation. Further, it houses derivatives; and the archival and curatorial staff never warmed to the way it stored and rendered their content.
- The archival and curatorial staff liked the Media Manager and the ease with which the could upload and manage content. But, the College of Arts & Sciences decided to no longer support the system and as of just about three years ago, turned it off. Additionally, the Media Manager was constructed at a time that web browsers were not rendering TIF images, and therefor all the content was once again derivatives, and there was ZERO preservation activity going on.
- And the there is the infamous “Dark Archive” originally located on a server named DSpace4, which led to the myth that some preservation activity was happening, when in reality it was just a Secure FTP server, where a staging instance of DSpace had once been located.

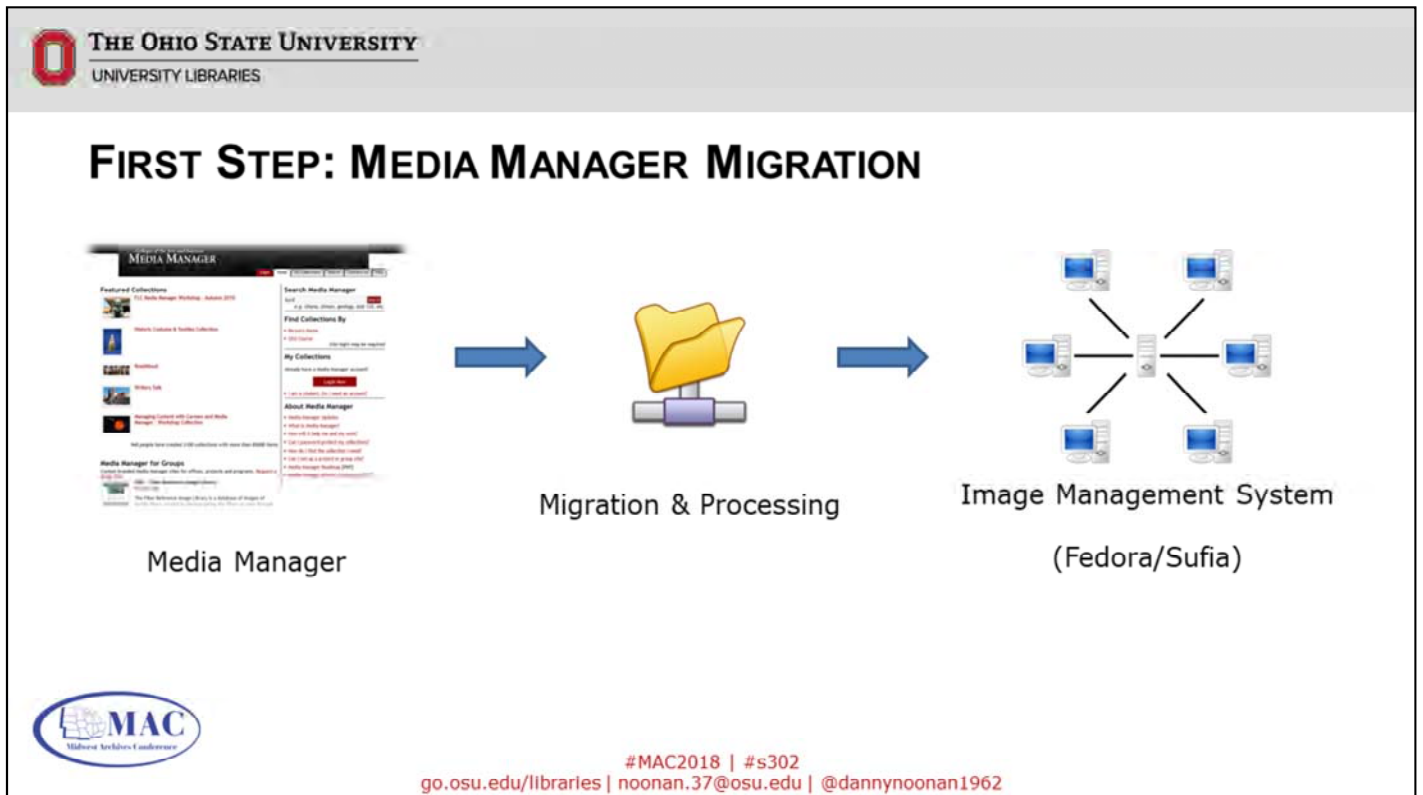


- And this image of the "Dark Archive" is a very apt analogy, as it was created by the dumping of a failing special collections projects share-drive into this environment. It being an sFTP server eventually led to more controlled access and its use for preservation masters, but by then it was filled with upwards of 2M items, of which a significant portion could at best be described as "provisional masters" with no real preservation activity happening.


Now nearly half of these item are "masters" to things that are in the KB – which we have yet to decide how to deal with.

In the meantime, we needed to decide on a preservation and access environment, especially to consume the content that was going to go away once the Media Manager was turned off.





This all occurred at about the same time we developed our Digital Preservation Policy Framework that led to a Master Objects Repository Task Force, which in turn led to our decision to build our preservation and access platform using Fedora and Hydra (specifically Sufia). A platform that we will be migrating this year to a Fedora-Hyrax environment.




By the summer of 2015, we had the first iteration of the system—the Image Management System— that had been populated solely with content from the now defunct Media Manager. While I could spend the whole session and then some discussing the challenges that arose with this first attempt, the lingering pain-points come from decisions made to narrowly address the needs/desires of the two collecting units whose content was being migrated.



## DARK ARCHIVE MIGRATION: 2013 > CURRENT

  →  → 


- De-duplication Analysis
- “DarKive” file quantification
- “Collection” Analysis Tool
- Prioritization 1.0 > 2.0 > 2.1



#MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)


Over the past couple of years I have reported at various conferences and in writing some of the processes we have gone through starting in 2013, including:

- The de-duping of content in the “Dark Archive”
- A quantification of the “Dark Archive” content by format
- The development of an Access database to analyze the content, engaging the archivists and curators in determining
  - What to keep and what to toss
  - Where the metadata is or if it exists
  - Intellectual property rights, and
  - What type of access we are allowed to provide
- Finally, we have been through several iterations of a migration prioritization that has changed based on issues that arose from the Digital Collections or DC system, itself.


 **THE OHIO STATE UNIVERSITY**  
UNIVERSITY LIBRARIES

## METADATA DESIRE VS. REALITY

Abstract or Summary	Format	Related URL
Access Rights	Identifier	Resource Type
Alternative Title	Is Format Of	Rights
Bibliographic Citation	Keyword	Source
cid	Language	Spatial
Collection	Location	Staff Notes
Collection Identifier	Materials	Sub-Collection
Contributor	Measurements	Subject
Creator	pid	Time Period
Date Copyrighted	Preservation Level Rationale	Title
Date Created	Preservation Level Value	Unit
Date Issued	Preservation Level Value	Work Type
Description	Date Assigned	Other (any other defined element)
Extent	Provenance	
	Publisher	



<http://go.osu.edu/MOMmetadata>

 #MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

We developed a robust metadata application profile that serves as an example of that ideal vs real world balancing act.

In the ideal world we would include all of these 38 metadata elements or additionally as the last entry says "other"


Note I do not include the pid & cid (parent and child ID) fields in this count, as they are required only for complex objects.

(Wait a minute, with "Other" that is 39 and pid and cid make it 41; if we only had 1 more it would truly be ideal and answer all the questions in the universe – any Doug Adams fans out there?)

 **THE OHIO STATE UNIVERSITY**  
UNIVERSITY LIBRARIES

## METADATA DESIRE VS. REALITY

Abstract or Summary	Format	Related URL
<b>Access Rights</b>	<b>Identifier</b>	<b>Resource Type</b>
Alternative Title	Is Format Of	Rights
Bibliographic Citation	Keyword	Source
<b>cid</b>	Language	Spatial
<b>Collection</b>	Location	Staff Notes
Collection Identifier	Materials	Sub-Collection
Contributor	Measurements	Subject
Creator	<b>pid</b>	Time Period
Date Copyrighted	<b>Preservation Level Rationale</b>	<b>Title</b>
Date Created	<b>Preservation Level Value</b>	<b>Unit</b>
Date Issued	Preservation Level Value	Work Type
Description	Date Assigned	Other (any other defined element)
Extent	Provenance	
	Publisher	



<http://go.osu.edu/MOMmetadata>


 #MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

In reality all we need is 8 metadata fields that the archivists and curators need to identify (along with pid and cid if complex object).


One way to look at this, is that this is ideal and real world for archivist and curators who have to worry about how to find time to get 1M objects moved into the DC preservation and access environment.

However, this may not be so ideal for patrons and users trying to access the information without the richer metadata context.




 THE OHIO STATE UNIVERSITY  
UNIVERSITY LIBRARIES

## CHALLENGES



- Staffing
- Competing priorities
- Detritus
- Fedora performance
- Object complexity & Data Models
- Hardware migrations
- Lack of user documentation
- Lack of born-digital accession workflow





#MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

The balancing act on the “digital preservation and access” high-wire that we continue to proceed on include:


- Lack of staffing devoted to archival and special collections metadata, although we are in the process of hiring a metadata strategist for the Libraries
- Competing priorities for archivists’, curators’ and system developers’ time
- Wading through the detritus in the “Dark Archive” so that we are truly only migrating preservation or provisional masters...even with all the de-duping and prioritization exercises we still end up with IDing stuff that should just be thrown away when we sit down and “dig-in” to a collection.
- Ongoing performance issues with the way Fedora works and how our layers of software interact with it.
- We see this specifically with complex objects where the number items within object seems to max out around 10 before we cannot access the item even to QC it. This is further exacerbated by using a data model that currently does not allow us to construct ordered complex objects. These we are hoping to

ameliorate with the upcoming migration to Hyrax and other system modifications.

- A library-wide server infrastructure migration over the holidays had us locked out of uploading content to the system for more than 2 months, while our systems folks dug through three different issues that were blocking permissions. In an ideal world, we would not have waited that long, but if we loop back to the first and second bullets we were dealing with competing priorities for the time of a limited staff pool.
- As it is a somewhat homegrown system, our user documentation is minimal, and as my metadata assistant and I wade through all of this, we are discovering a variety of issues that have never been documented, but are using it as an opportunity to develop more robust user documentation.
- Finally, as we have been so focused on the “Dark Archive” migration, we have yet to develop born-digital accessioning work flows to get that type of content into the DC.



**DAN NOONAN**  
Associate Professor  
Digital Preservation Librarian  
University Libraries | Information Technology  
320B 18<sup>th</sup> Avenue Library | 175 West 18<sup>th</sup> Avenue  
Columbus, OH 43210  
614.247.2425 Office  
[noonan.37@osu.edu](mailto:noonan.37@osu.edu) [go.osu.edu/noonan](http://go.osu.edu/noonan)



#MAC2018 | #s302  
[go.osu.edu/libraries](http://go.osu.edu/libraries) | [noonan.37@osu.edu](mailto:noonan.37@osu.edu) | [@dannynoonan1962](https://twitter.com/dannynoonan1962)

I'd like to finish with the following quote that I think sums up the idea of "Balancing the ideal vs. the real when the rubber hits the road in real-world digital preservation":

A colleague of ours, Sofia Becerra (from the Berklee College of Music) posted a link to Facebook yesterday from Sara Allain's blog "Letters to a Young Librarian" The post *10 Things I Didn't Learn in Archives School*, included this at #9:

"Email, social media, digital preservation - we're still figuring it out.

I regularly feel lost when it comes to these topics, but I've realized over time that it's okay to feel lost because we're all lost, as a profession.

It's easy to focus on the small majority of people and institutions that are making headway - they're the folks who present at conferences and write papers and tweet about their amazing work.

They're wonderful! They're truly doing some exceptional work. But it's also okay to be the person who is doing the little things.

You want to be ahead of the game on digital preservation? Make sure that your content isn't stored on a hard-drive and you'll be doing more than many.

As we continue to push the boundaries of what archiving comprises in the 21st Century, it's okay to take an inch rather than a mile. Positive incremental change can be as powerful as the big leaps."

(<https://letterstoayounglibrarian.blogspot.com/2018/03/10-things-i-didnt-learn-in-archives.html>)

Thank you!