

Digital Preservation Strategies in Practice at IWU

Meg Miner, University Archivist
Illinois Wesleyan University

POWRR Institute 3
Noble Science Library
Arizona State University
June 14, 2018



Meg's preservation philosophy

(aka local-centric decisions)

- Not everything created must be saved.
- Analog originals in good condition count as one preservation copy.
- Not everything saved as a digital object needs bit-level preservation immediately on acceptance.
- Digital objects that are saved/preserved will be accessible but not necessarily 24/7 on demand.
- Digitized and born-digital objects can be reformatted and/or edited at the archivist's discretion.

Digital collection file inventory

yellow = not for long term pres; green=long term and bit level					
AMES_09Share Folder	Format	# Files	Size, Collection	Size, File Ave.	Coll. Creation Date
19thCentStuPub	Tif	5,776	829 MB	6.97 MB	1-Feb-10
19thCentStuPub	Pdf	279	868 MB	0.32 MB	1-Feb-10
2009_Honors	Pdf/Word Doc	54	17.9 MB	3.02 MB	31-Mar-10
All Research Honors Theses	Pdf	658	7.55 GB	87.15 GB	18-Mar-10
Arends_Leslie_chair_restore	jpg	14	5.75 MB	2.43 MB	14-Jul-10
Argus_all_by_year	Pdf	2,981	14.3 GB	208.46 GB	30-Mar-10
Argus_fulltext_pdfs	Pdf	2,831	13.8 GB	205.14 GB	4-Feb-10
A:\audio_reformatting_project_20090805\16_441 Derivative	Wave sound	36	6.46 GB	5.57 GB	20-Aug-09
A:\audio_reformatting_project_20090805\24_96 Archival	MRK, XML, Wave sc	144	21.1 GB	6.82 GB	20-Aug-09
A:\audio_reformatting_project_20090805\Metadata	jpg	112	66.3 MB	1.69 MB	20-Aug-09
BSN50_completed	jpg, txt, tif	729	597 MB	1.22 MB	29-Nov-10
Bushnell_Civil_Rights	mpeg, mpeg4	8	5.15 GB	1.55 GB	16-Apr-10
Faculty Minutes	Pdf	1531	286 MB	5.35 MB	3-Nov-09
from Archives on S	Pdf, jpg	979	2.51 GB	390.04 GB	3-Nov-09
Gearhart Collection	text	8	222 KB	0.04 KB	8-Jul-09
Gearhart Collection	jpg	96	12.5 MB	7.68 MB	8-Jul-09
Greenstone files	html	5,663	80.1 MB	70.7 MB	31-Mar-10
historical photos	jpg	343	136 MB	2.52 MB	8-Jul-09
IWU Journal Projects	jpg	78	126 MB	0.62 MB	22-Nov-10
library art	jpg	91	1.02 GB	89.22 GB	19-May-10
Mandala Sand Painting	jpg	56	93.7 MB	0.6 MB	27-Jan-10
Non-res_Registrars_book2	Pdf	2	489 MB	0.004 MB	2-Jul-09
Non-res_Registrars_book2	Tif	1,422	17 GB	83.65 GB	2-Jul-09
Non-res_registration_books	Pdf files	31	1.30 GB	23.85 GB	23-Mar-10
orallist files	audio, txt, jpg	288	4.44 GB		16-Nov-10
student art purchase	jpg	13	5.17 MB	2.51 MB	3-Mar-10
Titan TV	AVI, Qtime, jpg, Res	816	148 GB	5.51 GB	11-Feb-10
Univ_comm_discs	mp3, jpg	252	825 MB	.31 MB	14-Jun-10

Plan for this session

- Updates on processing actions
- Updates on storage activities
- Continuing worries
- Mental health tips
- Questions/discussions

Updates on processing actions

- Clarification of two Case Study points:
 - DA developer has not updated the tool to include the functions of the DA-MetadataTransformer (DA-MT)
 - I misunderstood the “exclude” function in DataAccessioner (no record kept of that, but Archivematica does record that type of decision)

“Exclude” in DA

DataAccessioner v. 1.0

File FITS Tools

Your Name Meg Miner

Accession Number AE2015-03_SeniorDinner

Collection Title Senior Dinner 2015

Accession to Disk Z:\copied media\processed_with_data_accessioner\MasterCopies_DO NOT OPEN FILES

Source/Directory Exclude Include

Source Name/Identifier SeniorDinner

SeniorDinner	Date
SeniorDinner	Apr
001.jpg	Apr
002.jpg	Apr
003.jpg	Apr
004.JPG	Apr
005.JPG	Apr
006.jpg	Apr
007.jpg	Apr
008.jpg	Apr
009.jpg	Apr

File/Folder Dublin Core Metadata

Dublin Core Element dc:identifier

Metadata Value

Add New Remove Selected

Element	Value
dc:contributor	Carney-Hall, Karla
dc:date	2015-04-07
dc:format	jpg
dc:format	docx
dc:description	Primarily contains images
dc:identifier	RG 18-1/8 Special Files
dc:identifier	RG 11 - 1/8/6 Senior

Migrate Cancel Clear Source Information Clear All

Running FITS on Z:\copied media\processed_with_data_accessioner\MasterCopies_DO NOT OPEN FILES\AE2015

Benefits of DA + DA-MT

- Portability! Example: author's papers.
- Non-invasive insights into bulk transfers on media.
- Outputs hold *potential* for future auditing (good stewardship) but immediately useful for
 - Calculating quantities accumulated.
 - Identifying formats, especially at-risk content.





Digital Collections Inventory

(the POWRR example)

Category	Title & Description	Date	Location	Extent	Format
<i>(locally defined: project name? content creation method?)</i>	<i>(donor applied and/or yours...what's your practice?)</i>	<i>(YYYYMMDD or other locally defined format)</i>	<i>(storage place of choice-- networked server recommended)</i>	<i>(quantity of folders, files by type or total size)</i>	<i>what extensions are involved: .jpg, .tif, .xls?)</i>
Special Collections, mixed: digitized and born digital	A Digital Dog Collection. Donated by Jane (nee Pennypincher) and John Moneybags, Class of 2006. Canine Health research.	20160310	C:\User\Public\NewAccess ions\Master Copies		

IWU Accession record in Access

Accessions

 **Accessions**   

Accession No. AccessionType:

Date Received Date Rec'd (OLD)

Donor

Title

Phys. Description **Condition**

Record Group No. Location

Volume Amount Volume (OLD)

E-Media Type Media Size: **Media Retained** ☐

Source Notes

Updates on storage activities

- Redundant Array of Independent Disks (RAID)
 - Location changed back to physical library building.
 - Plan continued use for near-line storage needs.
- Fixity an open source tool for monitoring digital files in long-term storage locations.
 - One caution: in a Windows environment, Fixity needs an uninterrupted port location.
- Checksum comparisons
 - Still no storage system that will ingest checksums previously created and automatically compare them to ones generated during ingest.

Updates on storage activities

DuraCloud™

- Requested Level 3 so that streaming media from storage would be possible but received funding for Level 2.
- Storage costs are now separate from subscription.
- Provider routinely analyzes reports about corruption in stored content.
 - As of two weeks ago, staff report that replacements occur when files don't complete the initial transfer.

1,067 Temp objects, 290.5 GB

The screenshot displays a digital archive interface. On the left, a 'Content Items' panel shows a list of audio files, with the first 200 of 905 items visible. The files are named with a prefix 'ASC_Audio_Ship1-2/' followed by a date and a description. A 'Manifest' button is highlighted with a red circle. On the right, the 'Space Detail' panel for 'audio original' shows 905 items, a creation date of 2017-04-14T21:02:56, and a last health check of Fri Mar 30 06:46:08 UTC 2018 - success [report]. Below this, a 'History' section contains a 'Cumulative Byte and File Counts Over Time' chart. The chart shows a sharp increase in both bytes and files around May 2018, reaching approximately 290.5 GB and 1,067 files.

Content Items

Showing 1 - 200 of 905 show more>

Refresh Upload

- ASC_Audio_Ship1-2/1957-03-25_Titan_Time_Radio_Program.wav
- ASC_Audio_Ship1-2/1965_Performance_Problems_in_the_music_of_Vare
- ASC_Audio_Ship1-2/1968_Inaugural_Convocation_a.wav
- ASC_Audio_Ship1-2/1969-03-18_Astronaut_Neil_Borman_Luncheon_a.wav
- ASC_Audio_Ship1-2/1969-03-18_Astronaut_Neil_Borman_Luncheon_edit.
- ASC_Audio_Ship1-2/1969-03-18_Astronaut_Neil_Borman_Observatory_a
- ASC_Audio_Ship1-2/1969-03-18_Founders_Day_a.wav
- ASC_Audio_Ship1-2/1969-07_Admissions_Radio_tape_a.wav
- ASC_Audio_Ship1-2/1970-04-22_Loring_C_Merwin_a.wav
- ASC_Audio_Ship1-2/1970-04-22_Loring_C_Merwin_b.wav
- ASC_Audio_Ship1-2/1970_Radio_Wesleyan.wav
- ASC_Audio_Ship1-2/1971-05-03_Quinton_Snook_Oral_History_a.wav
- ASC_Audio_Ship1-2/1971-05-03_Quinton_Snook_Oral_History_b.wav
- ASC_Audio_Ship1-2/1971-05-23_Sen_Stevenson_Commencement_Addre
- ASC_Audio_Ship1-2/1971-05-23_Sen_Stevenson_Commencement_Addre
- ASC_Audio_Ship1-2/1971-05-27_William_Matheson_a.wav

Space Detail

audio original

Manifest

Items: 905

Created: 2017-04-14T21:02:56

Last Health Check: Fri Mar 30 06:46:08 UTC 2018 - success [report]

Recount

History

Cumulative Byte and File Counts Over Time

300.0 GB
250.0 GB
200.0 GB
150.0 GB

May 2017 Jul 2017 Sep 2017 Nov 2017 Jan 2018 Mar 2018 May 2018

Bytes Files

1000 files
900 files
800 files
700 files
600 files
500 files

DuraCloud upload manifest

space-id	content-id	MD5			
audio-original	Brown_Jared/Alan_Pakula/AlixGordon_pt2_GeorgeJustin_BobWoodward_MartinStargerDor	550f1be72fe6576caeb5eaeb7ea7baf	e		
audio-original	Brown_Jared/Alan_Pakula/AlixGordon_pt2_GeorgeJustin_BobWoodward_MartinStargerDor	11c507b7b0161a1924855454b70450e			
audio-original	Brown_Jared/Alan_Pakula/AnnaBoorstin_ChrisMurray_pt1_AssetFront.JPG	78ea99393f0fef252981ae715df0b82d			
audio-original	Brown_Jared/Alan_Pakula/BobBoorstin_pt2_MartinStarger_AssetFront.JPG	7d71272b6af2a3f8913592750e77fb4d			
audio-original	Brown_Jared/Alan_Pakula/AnnaBoorstin_ChrisMurray_pt1_b_access.wav	9313d828cc09ac06e2f65fff07f47933			
audio-original	Brown_Jared/Alan_Pakula/BobBoorstin_pt2_MartinStarger_a_access.wav	f4cd0b1b187dab26f147e6add312cdca			
audio-original	Brown_Jared/Alan_Pakula/AnnaBoorstin_ChrisMurray_pt1_a_access.wav	b69b651d2cbb752fb377aa57b52937a2			
audio-original	Brown_Jared/Alan_Pakula/BobWoodward_9_17_04_CyleKessler_12_8_04_HarrisonFord_12_9	46149654d28aa16fc993a417ae65f9b1			
audio-original	Brown_Jared/Alan_Pakula/AnnetteLindorf_AlanaPakula_at92ndSt_interview_11_27_90_b_acc	ad126f00152a62c2672f9fe353b73a24			
audio-original	Brown_Jared/Alan_Pakula/AnnetteLindorf_AlanaPakula_at92ndSt_interview_11_27_90_Asset	11216d076440257b05e3ab817b5f4052			
audio-original	Brown_Jared/Alan_Pakula/AlixGordon_pt2_GeorgeJustin_BobWoodward_MartinStargerDor	a232054ff132fea2a68c828a9ce82adc			
audio-original	Brown_Jared/Alan_Pakula/BobBoorstin_pt2_MartinStarger_b_access.wav	c0326b9cec5b2576c3f37f8cffe473b4			
audio-original	Brown_Jared/Alan_Pakula/CandiceBergen_MichaelSmall_AssetFront.JPG	44690e8dc9e1de098d4ce46e5a59b35b			
audio-original	Brown_Jared/Alan_Pakula/CandiceBergen_MichaelSmall_b_access.wav	21625701e6707ba2588e96e0f20a609e			
audio-original	Brown_Jared/Alan_Pakula/BobWoodward_9_17_04_CyleKessler_12_8_04_HarrisonFord_12_9	dad6c66c584e62785fd868d38d88ad56			
audio-original	Brown_Jared/Alan_Pakula/BobWoodward_9_17_04_CyleKessler_12_8_04_HarrisonFord_12_9	002b299a6a605a5a850dcfc64f09be89			
audio-original	Brown_Jared/Alan_Pakula/CathySolt_pt2_BobBoorstin_pt1_AssetFront.JPG	255bc86d91675b4aec90b9c3a37f81d9			
audio-original	Brown_Jared/Alan_Pakula/CandiceBergen_MichaelSmall_a_access.wav	25fd1aca117efa3e1ba2a8795c9deec3			
audio-original	Brown_Jared/Alan_Pakula/Celia_Costas_PattyDohertyHess_HannahPakula_pt1_AssetFront.JI	8f76993876b9cbbb2a3d516cac17293b			

Retrieval in Command Line

```
Administrator: C:\Windows\system32\cmd.exe - java -jar retrievaltool-4.2.0-driver.jar -h iwu.duraclo...
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Meg Miner>java -jar retrievaltool-4.2.0-driver.jar -h iwu.duracloud.org
-u mminer -c retrieved-files -f file-list.txt -s video
DuraCloud password:

Starting up the Retrieval Tool ...

Retrieval Tool 4.2.0 - Configuration
-----
Retrieve spaces: video
DuraStore Host: iwu.duracloud.org
DuraStore Port: 443
DuraStore Username: mminer
Retrieval Tool Content Directory: C:\Users\Meg Miner\retrieved-files
Retrieval Tool Work Directory: C:\Users\Meg Miner\duracloud-retrieval-work
Retrieval Tool Overwrite Local Files: false
Retrieval Tool Retain File Time Stamps: true
Retrieval Tool List Content Only: false
-----

... Startup Complete
The Retrieval Tool will exit when processing is complete. Status will be printed
every 10 minutes.

Retrieval Tool 4.2.0 - Status
-----
Start Time: 2017-04-12 14:16:13.531
Current Time: 2017-04-12 14:26:13.565
Retrievals In Process: 0
Successful Retrievals: 5
No Change Needed: 0
Failed Retrievals: 0
-----
```

File names for object retrieval

SpColl_Brown_Jared/LuntFontanne/Sargent_Mary.wav			
SpColl_Brown_Jared/LuntFontanne/Schaffer_George Interview.wav			
SpColl_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-chunk-0000			
SpColl_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-chunk-0001			
SpColl_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-chunk-0002			
SpColl_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-manifest			
SpColl_Brown_Jared/LuntFontanne/Terkel_Studs.wav			
SpColl_Brown_Jared/LuntFontanne/Texaco_MetOpera.wav			
SpColl_Brown_Jared/LuntFontanne/TheLunts_Speech.wav			
SpColl_Brown_Jared/LuntFontanne/Valency_Maurice.wav			
SpColl_Brown_Jared/LuntFontanne/VanPattern_Complete.wav			
SpColl_Brown_Jared/LuntFontanne/Wolcott_Alexander Letters.wav			
SpColl_Brown_Jared/LuntFontanne/Zolotows Letters and Notes.wav.dura-chunk-0000			
SpColl_Brown_Jared/LuntFontanne/Zolotows Letters and Notes.wav.dura-chunk-0001			
SpColl_Brown_Jared/LuntFontanne/Zolotows Letters and Notes.wav.dura-manifest			
SpColl_Brown_Jared/Mostel/Abbott_George.wav			
SpColl_Brown_Jared/Mostel/Ackerman_Bettye.wav			
SpColl_Brown_Jared/Mostel/Allen_Steve.wav			
SpColl_Brown_Jared/Mostel/Berkman_Aaron.wav			
SpColl_Brown_Jared/Mostel/Bernstien_Walter.wav			
SpColl_Brown_Jared/Mostel/Billops_Hatch.wav			
SpColl_Brown_Jared/Mostel/Burgen_Sidney.wav			
SpColl_Brown_Jared/Mostel/Chodorin_Jerome.wav			
SpColl_Brown_Jared/Mostel/Cole_Toby.wav			
SpColl_Brown_Jared/Mostel/Corey_Jeff.wav			
SpColl_Brown_Jared/Mostel/Debuskey_Merle.wav			
SpColl_Brown_Jared/Mostel/Faye_Joey.wav			
SpColl_Brown_Jared/Mostel/Fishko_Bella.wav			
SpColl_Brown_Jared/Mostel/Garfield_Julie.wav			
SpColl_Brown_Jared/Mostel/Gelbart_Larry.wav			
SpColl_Brown_Jared/Mostel/Gilbert_Lou.wav			
SpColl_Brown_Jared/Mostel/Gillford_Jack_Maddeline.wav.dura-chunk-0000			

Errors!

```
er_George Interview.wav.dura-manifest' in 'video' due to: Response code was 404,
expected value was 200. Response body value: Error attempting to get properties
for content 'SpColl_Brown_Jared/LuntFontanne/Schaffer_George Interview.wav.dura
-manifest' in 'video' due to: Could not find content item with ID SpColl_Brown_J
ared/LuntFontanne/Schaffer_George Interview.wav.dura-manifest in S3 bucket akiaizkbpv
ilxluax5a.video. S3 error: Not Found (Service: Amazon S3; Status Code: 404
; Error Code: 404 Not Found; Request ID: A54653B97E0B2F57)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/Seawell
_Donald.wav.dura-manifest' after 5 attempts. Last error message was: Error getti
ng content: Error attempting to get content 'SpColl_Brown_Jared/LuntFontanne/Sea
well_Donald.wav.dura-manifest.dura-manifest' in 'video' due to: Response code wa
s 404, expected value was 200. Response body value: Error attempting to get prop
erties for content 'SpColl_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-mani
fest.dura-manifest' in 'video' due to: Could not find content item with ID SpColl
_Brown_Jared/LuntFontanne/Seawell_Donald.wav.dura-manifest.dura-manifest in S3
bucket akiaizkbpv ilxluax5a.video. S3 error: Not Found (Service: Amazon S3; Stat
us Code: 404; Error Code: 404 Not Found; Request ID: C321D1FA6028DE0F)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/Terkel_
Studs.wav' after 5 attempts. Last error message was: Error getting content: Error
attempting to get content 'SpColl_Brown_Jared/LuntFontanne/Terkel_Studs.wav.dura
-manifest' in 'video' due to: Response code was 404, expected value was 200. R
esponse body value: Error attempting to get properties for content 'SpColl_Brown
_Jared/LuntFontanne/Terkel_Studs.wav.dura-manifest' in 'video' due to: Could not
find content item with ID SpColl_Brown_Jared/LuntFontanne/Terkel_Studs.wav.dura
-manifest in S3 bucket akiaizkbpv ilxluax5a.video. S3 error: Not Found (Service:
Amazon S3; Status Code: 404; Error Code: 404 Not Found; Request ID: 0A2AE6AB0DC
CA51D)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/Texaco_
MetOpera.wav' after 5 attempts. Last error message was: Error getting content: E
rror attempting to get content 'SpColl_Brown_Jared/LuntFontanne/Texaco_MetOpera.
wav.dura-manifest' in 'video' due to: Response code was 404, expected value was
200. Response body value: Error attempting to get properties for content 'SpColl
_Brown_Jared/LuntFontanne/Texaco_MetOpera.wav.dura-manifest' in 'video' due to:
Could not find content item with ID SpColl_Brown_Jared/LuntFontanne/Texaco_MetOp
era.wav.dura-manifest in S3 bucket akiaizkbpv ilxluax5a.video. S3 error: Not Fou
nd (Service: Amazon S3; Status Code: 404; Error Code: 404 Not Found; Request ID:
F66348B800493DD9)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/TheLunt
s_Speech.wav' after 5 attempts. Last error message was: Error getting content: E
rror attempting to get content 'SpColl_Brown_Jared/LuntFontanne/TheLunts_Speech.
wav.dura-manifest' in 'video' due to: Response code was 404, expected value was
200. Response body value: Error attempting to get properties for content 'SpColl
_Brown_Jared/LuntFontanne/TheLunts_Speech.wav.dura-manifest' in 'video' due to:
Could not find content item with ID SpColl_Brown_Jared/LuntFontanne/TheLunts_Spe
ech.wav.dura-manifest in S3 bucket akiaizkbpv ilxluax5a.video. S3 error: Not Fou
nd (Service: Amazon S3; Status Code: 404; Error Code: 404 Not Found; Request ID:
973049E5814DD122)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/Valency
_Maurice.wav' after 5 attempts. Last error message was: Error getting content: E
rror attempting to get content 'SpColl_Brown_Jared/LuntFontanne/Valency_Maurice.
wav.dura-manifest' in 'video' due to: Response code was 404, expected value was
200. Response body value: Error attempting to get properties for content 'SpColl
_Brown_Jared/LuntFontanne/Valency_Maurice.wav.dura-manifest' in 'video' due to:
Could not find content item with ID SpColl_Brown_Jared/LuntFontanne/Valency_Maur
ice.wav.dura-manifest in S3 bucket akiaizkbpv ilxluax5a.video. S3 error: Not Fou
nd (Service: Amazon S3; Status Code: 404; Error Code: 404 Not Found; Request ID:
243D2F1B94438068)
Failed to retrieve DuraCloud file 'video/SpColl_Brown_Jared/LuntFontanne/VanPatt
ern_Complete.wav' after 5 attempts. Last error message was: Error getting conten
t: Error attempting to get content 'SpColl_Brown_Jared/LuntFontanne/VanPattern_C
omplete.wav.dura-manifest' in 'video' due to: Response code was 404, expected va
lue was 200. Response body value: Error attempting to get properties for content
'SpColl_Brown_Jared/LuntFontanne/VanPattern_Complete.wav.dura-manifest' in 'vid
eo' due to: Could not find content item with ID SpColl_Brown_Jared/LuntFontanne/
VanPattern_Complete.wav.dura-manifest in S3 bucket akiaizkbpv ilxluax5a.video. S
3 error: Not Found (Service: Amazon S3; Status Code: 404; Error Code: 404 Not Fo
und; Request ID: 3B6BBFF3AF79259A)
```


Continuing worries

- Capture of Web content
 - Good news: help from staff!
 - Labor intensive and expect that some things are missed.
- Media dependency and Normalization
 - Good news: Budget allocation and *Summer of Old Media!* (digital and analog)
 - Labor intensive both for removable media and for staying aware of content loaded to DuraCloud (no normalizing in this storage system).

Decision help from the Digital Preservation Coalition

The 'Bit List' of Digitally Endangered Species

The DPC's 'Bit List' of Digitally Endangered Species is a crowd-sourcing exercise to discover which digital materials our community thinks are most at risk, as well as those which are relatively safe thanks to digital preservation. By compiling and maintaining this list over the coming years, the DPC aims to celebrate great digital preservation endeavors as entries become less of a 'concern,' whilst still highlighting the need for efforts to safeguard those still considered 'critically endangered.'

[Listen to William Kilbride talking to the BBC Click programme about the Bit List](#)

Click on each of the Risk Classifications to see the entries for each category.

LOWER RISK



Digital materials are listed as *Lower Risk* when it does not meet the requirements for other categories but where there is a distinct preservation requirement. Failure or removal of the preservation function would result in re-classification to one of the threatened categories.

VULNERABLE



Digital materials are listed as *Vulnerable* when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs. This classification includes *Lower Risk* materials in the presence of aggravating conditions.

ENDANGERED



Digital materials are listed *Endangered* when they face material technical challenges to preservation or responsibility for care is poorly understood, or where the responsible agencies are poorly equipped to meet preservation needs. This classification includes *Vulnerable* materials in the presence of aggravating conditions.

CRITICALLY ENDANGERED



Digital materials are listed *Critically Endangered* when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs. This classification includes *Endangered* materials in the presence of aggravating conditions.

PRACTICALLY EXTINCT



Digital materials are listed as *Practically Extinct* when the few known examples are inaccessible by most practical means and methods. This classification includes *Critically Endangered* materials in the presence of aggravating conditions.

CONCERN



Digital materials are listed as of *Concern* when an active member of the digital preservation community has expressed a legitimate concern but the concern has not yet been assessed by the BitList jury. They will be assessed for inclusion in the subsequent year.

Sanity savers

- Don't hesitate to psych yourself up!
 - Periodically assess progress to make your efforts visible to yourself/others: annual reports and timelines.

Meg's Baby Steps

Timeline 2015

- *Spring* Case Study published.
- *Summer* Tested Fixity, continued content inventory and preservation priorities, started cross-training for Sp2016 sabbatical coverage.
- *Fall* [Homecoming!] Advocate for DuraCloud subscription, workflows for Sp2016 sabbatical coverage (text, spreadsheet, and then a flowchart), started negotiations for extensive hybrid (paper & electronic) manuscript collection.
- *Winter* Request for online access to collection of audio cassette recordings; accepted hybrid mss collection from author (privacy implications = policy creation needed).

Meg's Baby Steps

Timeline 2016

- *Spring* Sabbatical resulting in lots of digital content!
- *Summer* Informed that a subscription to level 2 of DuraCloud would start in the fall.
- *Fall* [Homecoming!] Started using DuraCloud, hired digitized audio editor (see philosophy).
- *Winter* Documenting ingest decisions, learning about DuraCloud's "chunking" (aka 1GB increments) process, working through Command Line (@\$#&) after 30 years!

Meg's Baby Steps

Timeline 2017-2018

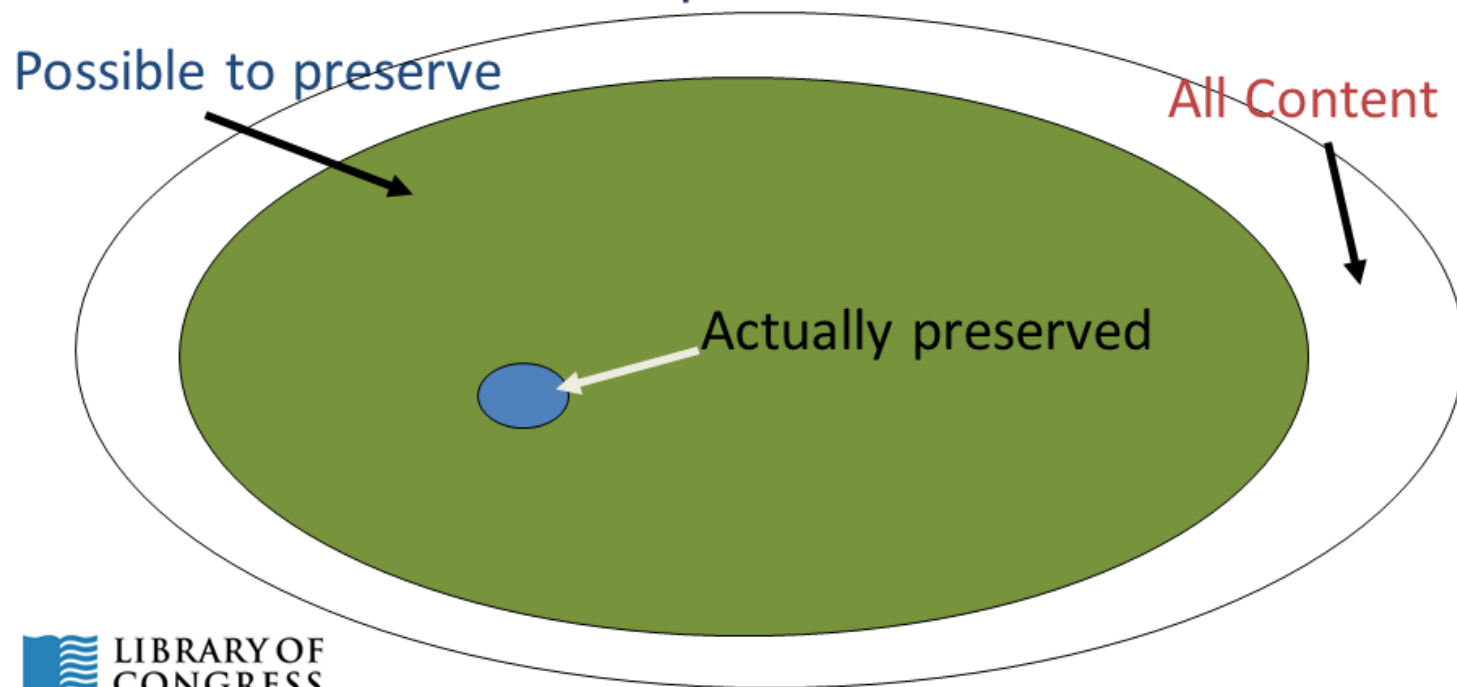
- *Spring* Documenting ingest decisions, reappraisal.
- *Summer* Comprehensive analog & digital A/V media assessment, digitized video editing, DuraCloud ingest, inventory (Finding Aid) with storage and access points as needed.
- *Fall* [Homecoming!] – *Spring 2018* Working with digitized content—more decisions! Learned about using DuraCloud as temp storage.
- *Summer* Two students assistants editing and conducting quality control on digitized content; two others creating and reviewing transcripts.

Sanity savers

- POWRR's Baby Steps = actions are local-centric
 - Adapt “best practices” into actions that make sense in your world (XML vs. Excel example).
 - Make priorities based on your collections and needs.

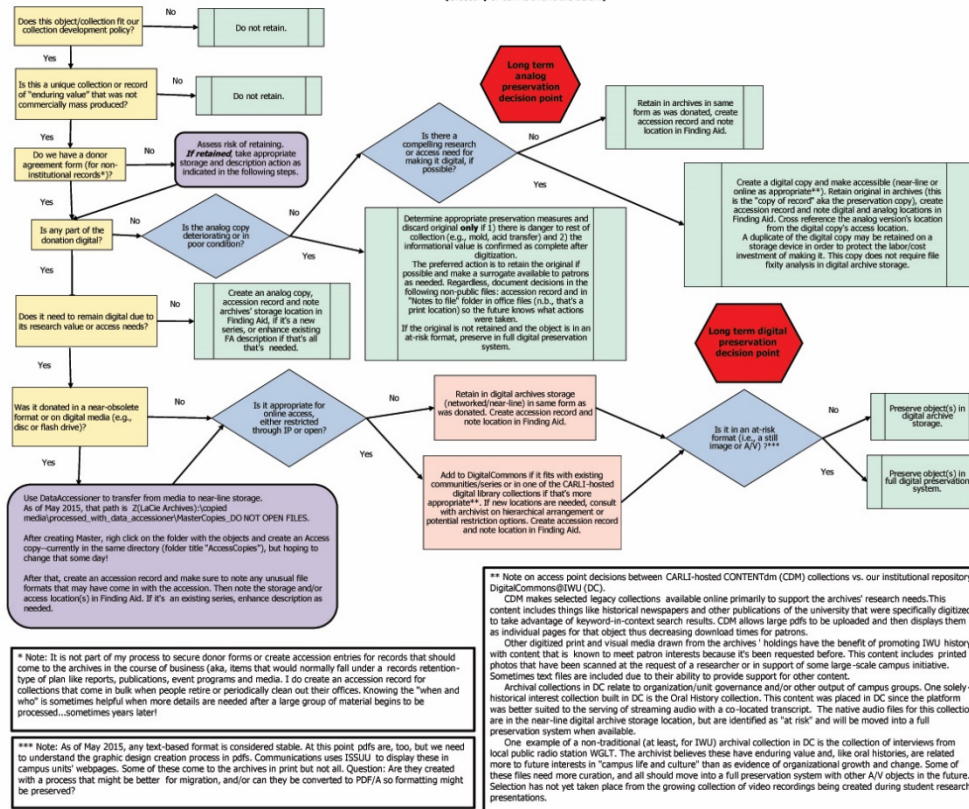
Decision making visualized

Good preservation decisions are based on a deep *understanding* of the **possible** content to be preserved



IWU Archives Flowchart: From Selection through Storage

(Glossary of terms available below)



* Note: It is not part of my process to secure donor forms or create accession entries for records that should come to the archives in the course of business (aka, items that would normally fall under a records retention-type of plan like reports, publications, event programs and media. I do create an accession record for collections that come in bulk when people retire or periodically clean out their offices. Knowing the "when and who" is sometimes helpful when more details are needed after a large group of material begins to be processed—sometimes years later!

*** Note: As of May 2015, any text-based format is considered stable. At this point, pdfs are, too, but we need to understand the graphic design creation process in pdfs. Communications uses ISSU to display these in campus units' webpages. Some of these come to the archives in print but not all. Question: Are they created with a process that might be better for migration, and/or can they be converted to PDF/A so formatting might be preserved?

Digital Preservation Terminology relevant to this flowchart

Digital Preservation: set of processes, activities and management of digital information over time to ensure long-term accessibility. Because of the relatively short lifecycle of digital information, preservation is an ongoing process.

Enduring Value: Tate Archives & Special Collections in The Ames Library has defined enduring value (aka, archival value) as unique materials concerning IUW history and/or materials such as those described in the archives' collection policy.

Digital Archive: a managed collection of primary source historical records in the highest resolution digital format available. Digital Archives are managed by archivists just as physical archives are; their storage options include local back up locations, networked drives with fixity checks (see Digital Archive Storage), and full preservation systems.

Digital Library: a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and are accessible by computers. The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system.

Digital Archive Storage: includes maintaining onsite and offsite backup copies, fixity-checking, and periodic refreshment by copying files to new storage media. The intent of this preservation level is to ensure that the integrity of the original file is assessed periodically so that later dissemination is possible and that the object is an authentic representation of the original.

Full Preservation: includes bit-level preservation of the originally submitted files, as well as services intended to ensure that the content of the files will remain usable into the indefinite future. These services vary according to the file type but may include the creation of normalized forms of the file and/or the reforming of obsolete formats to reasonably comparable successor formats. It is not guaranteed, however, that normalized or migrated versions of any file will be identical in functionality or in look and feel to the original file. Note also that if a complex or compound digital object (aka a logical object) is comprised of individual files in both supported and unsupported formats, there is no guarantee that the object will remain usable as intended by the creator. A judgment call on how to save some objects may be needed and absent any currently known ways of "normalizing" or "accessing" digital preservation systems that are capable of preserving content will win over formatting choices of creators in most cases. The obvious exception is where artistic/graphic design elements are integral to the reason for keeping the object. In those cases, bit-level assessments may be all that is possible.

The concept of a digital preservation flowchart and the terminology used here originated in a draft document by Tawnya Keller and Lisa Chaufy, University of Utah, dated 9/20/2012 and available at http://campusguides.lib.utah.edu/it.php?content_id=5929064. Keller attributes her idea to Chris Erickson at Brigham Young University. The author secured Keller's permission to re-use the original flowchart for DigitalPOWRR workshops on small scale digital preservation strategies.

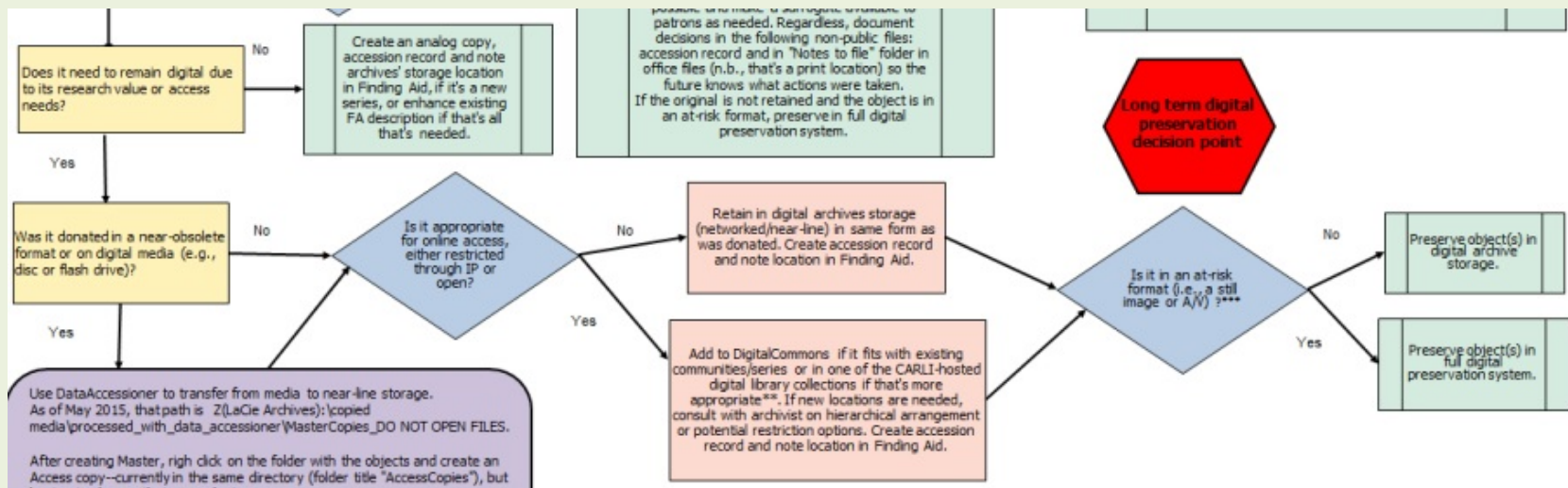
The original flowchart was extensively revised by Meg Miner, Illinois Wesleyan University, in May 2015, for the purpose of illustrating the decisionmaking process used when accessioning content into the University's Archives and in determining storage locations and access points. Contact archives@iuw.edu with questions or comments.



Last updated July 20, 2015

Post-
POWRR
emphasis:
decision
making,
documenta-
tion &
advocacy.

The DP part of the flowchart



Sanity savers (cont'd)

- POWRR's Baby Steps = actions are local-centric
 - Adapt “best practices” into actions that make sense to you (XML vs. Excel).
 - Make priorities based on your collections and needs.
 - Engage in whatever stewardship activities are possible for you.

Stewardship is incremental!

Blue = my standing at the end of POWRR.

Green = my progress since then.
Handwriting = my exceptions to these suggestions 😊

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	<ul style="list-style-type: none"> - Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system 	<ul style="list-style-type: none"> - At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them 	<ul style="list-style-type: none"> - At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media 	<ul style="list-style-type: none"> - At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	<ul style="list-style-type: none"> - Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content 	<ul style="list-style-type: none"> - Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content 	<ul style="list-style-type: none"> - Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content 	<ul style="list-style-type: none"> - Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	<ul style="list-style-type: none"> - Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files 	<ul style="list-style-type: none"> - Document access restrictions for content 	<ul style="list-style-type: none"> - Maintain logs of who performed what actions on files, including deletions and preservation actions 	<ul style="list-style-type: none"> - Perform audit of logs
Metadata	<ul style="list-style-type: none"> - Inventory of content and its storage location - Ensure backup and non-collocation of inventory 	<ul style="list-style-type: none"> - Store administrative metadata - Store transformative metadata and log events 	<ul style="list-style-type: none"> - Store standard technical and descriptive metadata 	<ul style="list-style-type: none"> - Store standard preservation metadata
File Formats	<ul style="list-style-type: none"> - When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs 	<ul style="list-style-type: none"> - Inventory of file formats in use 	<ul style="list-style-type: none"> - Monitor file format obsolescence issues 	<ul style="list-style-type: none"> - Perform format migrations, emulation and similar activities as needed

Counts stable analog originals & online access copies of born-digital A/V

Not Automated in any known pres. system

Manually!

Does not mean you will be listened to but we all must try!

Sanity savers

- POWRR's Baby Steps = actions are local-centric
 - Adapt “best practices” into actions that make sense to you (XML vs. Excel).
 - Make priorities based on your collections and needs.
 - Engage in whatever stewardship activities are possible for you.
 - Documenting decisions saves both you & your future-self some grief!

Editing decision documentation

A/V editing decisions

Content related to IWU history that is transferred to the archives is appraised for long term value by the archivist. Digital files that contain more than just a few minutes of unnecessary content* are opened in an uncompressed format (whenever possible), edited to the length needed to preserve the content with archival value and saved in uncompressed formats. Compressed versions are derived from this new "master." Master and access copies are loaded to [DuraCloud](#) and Vimeo or [SoundCloud](#) as applicable and notes are made (in S:\Archives\ARCHFIND\Record Groups\ Group 18-3 Digital Audio-Visual Files) about locations and editing decisions. When rendering from an editing platform, embed metadata in the uncompressed file if possible and then derive an access copy. In a case where editing actions will be evident to future users, title cards or some other visual cue is added directly into the compressed access file.

Files received from a vendor like [The MediaPreserve](#) come in multiple formats. Edit from the medium sized uncompressed version. If no editing is required, save the medium sized uncompressed and the access copy in their appropriate locations. The [MediaPreserve](#) also sends [jpgs](#) and xml reports of the digitized media. Save these in the LaCie location for that transfer project.

[DuraCloud](#) uploads will be placed in "spaces" (audio, images, text, video) with a folder structure originating in the upload computer's "to-be-synced" folder as follows:

Audio: [oral_history](#) {then [LastName_FirstName](#) as individual file inside folder}, [SpColl](#) {with subfolder for specific collection name like [CollAlumniClub](#), [STTI_Theta_Pi](#), [Brown_Jared](#) or [Kindred_Dave](#)}.

Images: IWU Founding Documents, [Depts](#), [SpColl](#) {with subfolder for specific collection name like [CollAlumniClub](#)}.

Video: [Academic_Affairs](#) {with subfolder for specific [dept](#)}; Athletics, Convocations, [Depts](#) (includes Programs/Schools), Events, Faculty (includes Staff), Guests, Honors, Presidents, Students. No subfolder is created for Commencement—these files are added as [Commencement_YYYY](#). Other naming conventions: {recurring event name}_YYYY, [LastName_FirstName](#)

Finding aid

Record Group 18 - 3: Digital Audio and Visual files

18-3/1 : Board of Trustees

Series (and DuraCloud prefix)	DATE (YYYY-MM-DD)	File Name (with DuraCloud link)	File type	Location(s) (Pres, Access)	Access URL	Details
Academic_Affairs	2002	Library/PembrokeRestoration	avi	DuraCloud, Vimeo	https://vimeo.com/219107470	Transferred from DVD
Alumni		Bedford_Luther	mov	DuraCloud, Vimeo	https://vimeo.com/258872652	Transferred from VHS
		Rodgers_Edelbert	mov	DuraCloud, Vimeo	https://vimeo.com/258639425	Transferred from VHS
Athletics						
	2012-04-17	TommyAwards_2012	avi	DuraCloud		
	2013-04-15	TommyAwards_2013	avi	DuraCloud		
	2015-04-13	TommyAwards_2015	avi	DuraCloud		
	2016-04-11	Tommy Awards_2016	avi	DuraCloud		
Commencement						
		Commencement_1994	avi	DuraCloud, Vimeo	https://vimeo.com/91415948	
		Commencement_1999	avi	DuraCloud, Vimeo	https://vimeo.com/91295831	
		Commencement_2004	avi	DuraCloud, Vimeo	https://vimeo.com/90760377	
		Commencement_2008	wav	DuraCloud, Sound	https://soundcloud.com/ames20	Transferred from CD
		Commencement_2009	wav	DuraCloud, Sound	https://soundcloud.com/ames20	Transferred from CD
		Commencement_2011	avi	DuraCloud, Vimeo	https://vimeo.com/90979983	
		Commencement_2012	avi	DuraCloud, YouTu	https://www.youtube.com/watch?v=siMPAFVg80k	
		Commencement_2013	mp4	DuraCloud		
	2014-05-04	Commencement_2014	avi	DuraCloud, YouTu	https://www.youtube.com/watch?v=K0HqI7wmmBs	
		Commencement_2015	avi	DuraCloud, Vimeo	https://vimeo.com/222242527	
		Commencement_2016	avi	DuraCloud, Vimeo	https://vimeo.com/228871734 AND https://www.youtube.com/watch?v=K0HqI7wmmBs	
		Commencement_2017	avi	DuraCloud, Vimeo	https://vimeo.com/228871685 AND https://www.youtube.com/watch?v=K0HqI7wmmBs	
Convocations						
	2017	FoundersDayConvocation_2017	mov	DuraCloud		
	2015	FoundersDayConvocation_2015	mov	DuraCloud		
	1969	FoundersDay/FrankBorman_Eva	avi	DuraCloud		Transferred from 16mm
	1969	FoundersDay/FrankBorman_Pre	avi	DuraCloud, Vimeo	https://vimeo.com/231786031	Transferred from 16mm

Thoughts to share?

- Do any of these actions ring true for your world?
- What decisions have you made?
- Successes?
- Worries?
- Other questions?

Thank You!

Meg Miner

Illinois Wesleyan University

mminer@iwu.edu

309-556-1538

<https://www.iwu.edu/library/archives>

