ARCHIVAL CHEAT SHEETS

TO EASY

A QUICK REFERENCE FOR THE ARCHIVAL NEW & COOL

JAN, 2024

TABLE OF CONTENTS
ARCHIVAL CHEAT SHEETS

OAIS REFERENCE MODEL [OAIS final v3 draft with changes wrt OAISv2 20190924-rl.docx (ccsds.org)](ccsds.org)

This is your digital archive's ecosystem interaction for long-term preservation and access needs.

FADGI 2023 3rd edition [FADGI Technical Guidelines for Digitizing Cultural Heritage Materials_3rd Edition_05092023.pdf (digitizationguidelines.gov)](digitizationguidelines.gov)

This manual sets the standards that you will follow for digitizing materials! Not all digitization efforts

are created   equally, so always follow the four-star recommendations herein and you can't go wrong!

NDSA LoP https://ndsa.org/publications/levels-of-digital-preservation/

After digitization, after-care is needed aka digital preservation so be mindful that this is on-going!

SOCIETY OF AMERICAN ARCHIVISTS GLOSSARY

Glossary Terms | Society of American Archivists

Although not a manual, per se, this website will help you understand the technical jargon

that the fields of archives and digital curation uses. When in doubt, look it up!

OAIS

REFERENCE MODEL

The Open Archival Information Systems (OAIS) Reference Model is a conceptual model for digital archives. It became an ISO standard in 2002. The goal of OAIS is to show general requirements for an archive to provide long-term preservation of digital information.
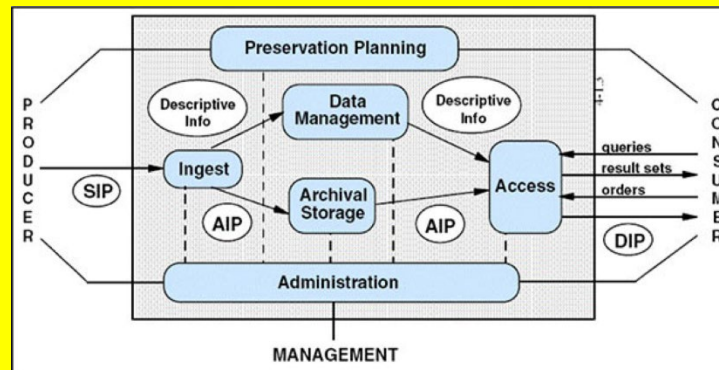
The OAIS identifies stakeholders in planning, creating, caring for, and using digital archives: management, producer, and the consumer.

- **Management:** Responsible for the strategic planning and policy development for the OAIS archive.
- **Producer:** Responsible for submitting and transferring items, collections, and knowledge to the OAIS archive.
- **Consumer:** Individuals, organizations or systems that locate, request and use the digital materials stored by the OAIS archive.

Jerrid Lee Miller developed the **NDN Country OAIS Reference Model** at the Cherokee Nation Language Department.

NDN COUNTRY

OAIS REFERENCE MODEL

Thinkin' long-term — Preservation planning — Supports Everybody

Locates the NDN Taco aka metadata.

Descriptive information

Gets every-one — Data management — together

Descriptive information

Grocery Store or Factory

Makin' it & — Ingest — Gettin' it out

Comes & — Access — Gets it

Like IHS — Archival Or storage hold'in

AIP

AIP

SIP

PRODUCER

Do'in this day to day — Administration — Skoden! Stoodis!

DIP

CONSUMER

Guy buying the NDN Taco

A DIP is like ordering (requesting) and then eating an Indian Taco aka Navajo Taco from the frybread stand. It has all of the fixins' and you enjoy it. The DIP is generated from the AIP, but it is more compressed for access.

An SIP is like frybread ingredients. You have to add them together and cook them before there is frybread. These digital materials are the same way.

An AIP is like making the perfect NDN Taco, you cooked the frybread to perfection and pieced together all of the ingredients.

NDN COUNTRY OAIS REFERENCE Model

WHAT THE HECK IS A "SIP" "AIP" AND "DIP"?

SIP = Submission Information Package

A SIP is the package of files and information created by the *producer* and brought into the *archive*. Like frybread ingredients from the grocery store that haven't been combined and cooked yet.

AIP = Archival Information Package

An AIP is a complete package of files and metadata created by the *archive* for preservation/access. Like the perfect NDN Taco, you cooked the frybread to perfection, taste-tested for quality, and pieced together all the ingredients.

DIP = Dissemination information Package

A DIP is derived from the AIP and delivered from the *archive* to the *consumer* in response to a request. Often compressed for access. Like a person ordering (requesting) an NDN Taco from the frybread stand. It has all of the fixin's and you enjoy it.

NDSA LOP

WITH ARCHIVIST-IN THE-WILD WISDOM

Functional Area	Level			
	Level 1 (Know your content)	**Level 2 (Protect your content)**	**Level 3 (Monitor your content)**	**Level 4 (Sustain your content)**
Storage	Have two complete copies in separate locations Document all storage media where content is stored Put content into stable storage **Digital Storage is needed!**	Have three complete copies with at least one copy in a separate geographic location Document storage and storage media indicating the resources and dependencies they require to function **3-2-1 Rule!**	Have at least one copy in a geographic location with a different disaster threat than the other copies Have at least one copy on a different storage media type Track the obsolescence of storage and media	Have at least three copies in geographic locations, each with a different disaster threat Maximize storage diversification to avoid single points of failure Have a plan and execute actions to address obsolescence of storage hardware, software, and media
Integrity	Verify integrity information if it has been provided with the content Generate integrity information if not provided with the content Virus check all content; isolate content for quarantine as needed	Verify integrity information when moving or copying content Use write-blockers when working with original media Back up integrity information and store copy in a separate location from the content	Verify integrity information of content at fixed intervals Document integrity information verification processes and outcomes Perform audit of integrity information on demand	Verify integrity information in response to specific events or activities Replace or repair corrupted content as necessary
	------Be transparent: You're word is your bond and you're only as good as what you last put out-----			
Control	Determine the human and software agents that should be authorized to read, write, move, and delete content **Who & what is doing this?**	Document the human and software agents authorized to read, write, move, and delete content and apply these **Put it down to pen!**	Maintain logs and identify the human and software agents that performed actions on content **Add to it.**	Perform periodic review of actions/access logs **Look back on it!**
Metadata	Create inventory of content, also documenting current storage locations Backup inventory and store at least one copy separately from content **Inventory time!**	Store enough metadata to know what the content is (this might include some combination of administrative, technical, descriptive, preservation, and structural) **Put it all together.**	Determine what metadata standards to apply Find and fill gaps in your metadata to meet those standards **Dublin Core? METS? MODS? Etc.**	Record preservation actions associated with content and when those actions occur Implement metadata standards chosen **Just do it!**
Content	Document file formats and other essential content characteristics including how and when these were identified **What? When? How?**	Verify file formats and other essential content characteristics **FADGI & Best Practices** Build relationships with content creators to encourage sustainable file choices **Professional Development**	Monitor for obsolescence, and changes in technologies on which content is dependent **Change with the times!**	Perform migrations, normalizations, emulation, and similar activities that ensure content can be accessed **Perpetual motion.**

N D S A L O P C H A R T

*W/ Short explanation in red

FADGI:

DIGITIZATION QUICK LOOK

Audio Digitization Standards (think Oral History Project)

Use for Digitized File	File Format	Bit Depth	Sample Rate
Master Preservation File (think AIP)	Wav.	24-bit	96KHz
Access File (think DIP)	MP3	192 kbps	44.1-96KHz

Image Digitization Standards (think photos and documents)

Master Preservation File (think AIP)	TIFF	16-bit grayscale (for black and white) or 48-bit color (for color)	600 dpi
Access File (think DIP)	JPEG 2000, PDF/A	16-bit grayscale (for black and white) or 48-bit color (for color)	400 dpi+

Video Digitization Standards (think events or interviews)*

Use for Digitized File	File Format/ Wrapper/Codec	Sample Rate	Resolution	Frames Per Second
Master Preservation File (think AIP)	AVI	4:2:2	480x720	60 for interlaced, (480x720i)
Access File (think DIP)	mp4		480x720 for SD 720xx1280p for HD or Film	60 for interlaced, 30 for HD, 24 for film

* Video master preservation files can be very large, if uncompressed video is not possible, use lossless instead of lossy compression.
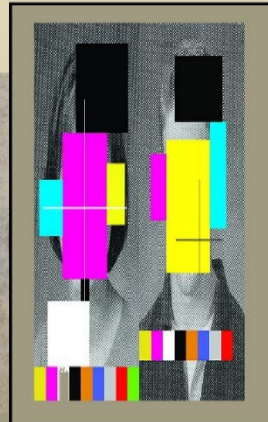
FREE:

Bitcurator 4.4.3 (open-source) Releases · BitCurator/bitcurator-distro Wiki · GitHub

For digital forensics to include pre-imaging, data triage, forensic disk imaging &, file system analysis.

DROID 6.7.0 (open-source) https://tna-cdn-live-uk.s3.eu-west-2.amazonaws.com/documents/droid-binary-6.7.0-bin.zip

Performs automate batch identification of file formats.

Bagger 2.8.1 (open-source) GitHub - LibraryOfCongress/bagger: The Bagger application packages data files according to the BagIt specification.

Creates, manipulates & validates AIP information from one computer system to another

YAH, THESE COST MONEY, BUT THEY'RE NEEDED:

TeraCopy Pro: $29.95 per year TeraCopy for Windows - Code Sector

This **clones** digital files so **no bit data is lost!** Also it generates & validates checksums.

Fixity Pro: $47.90 per year Fixity Pro

This is an integrity checking tool. Basically it audits and verifies Checksums are still good.

Total AV Antivirus Pro: $119.00 per year Total Antivirus 2023 – Save Up to 80% Off Antivirus Today (totalav.com)

Acts as a primary antivirus scanner so your digital collections do not become infected by a rogue file!

Norton 360 Delux: $49.99 per year Norton 360 Deluxe | Powerful protection for your devices

This is your second line of defense in the event your primary one didn't pickup it up on their scan!

3 ~~xxiex~~ of your data!
clones

X Master Preservation file
X Access clone
X Backup file

clones
2 ~~copies~~ stored locally

x Quarantine/Processing Computer
x Hard drive
x NAS/RAID

clone
1 ~~copy~~ stored off-site

x in the Cloud
x Bank Vault
x Annex building or outlier office
x at your house (if you must)

3-2-1 Backup Strategy

3
2
1

R
U
L
E

NAS/RAID

A.K.A.

Network Attached Storage

or Redundant Array Inexpensive Disk

Cheapest bang for your buck:

Asustor Drivestor 2 as1102t (2 bay NAS)

*4 bay option also available. If this is bought, buy 4 Seagate internal hard drives instead of 2.

Cost: $179.00

Amazon.com: Asustor Drivestor 2 AS1102T - 2 Bay NAS, 1.4GHz Quad Core, Single 2.5GbE Port, 1GB RAM DDR4, Network Attached Storage, Personal Private Cloud (Diskless) : Electronics

+

Seagate - IronWolf 8TB Internal SATA NAS Hard Drive

Cost: $159.99 x2 (because the Asustor has 2 bays)

Amazon.com: Seagate IronWolf 8TB NAS Internal Hard Drive HDD – 3.5 Inch SATA 6Gb/s 7200 RPM 256MB Cache for RAID Network Attached Storage – Frustration Free Packaging (ST8000VNZ04/N004) : Electronics

OR Top-shelf spending:

Promise Pegasus32 R6 24 TB RAID System

Cost: $2,299.00

Only for Mac users PROMISE Pegasus32 R6 24TB RAID System - Apple

Hard Drive

Better Performance in its class:

LaCie d2 Professional 8TB External Hard Drive
Cost: $249.99

Amazon.com: LaCie d2 Professional 8TB External Hard Drive Desktop HDD – USB-C USB 3.0 7200 RPM Enterprise Class Drives, 5 Year Warranty and Recovery Service (STHA8000800) : Clothing, Shoes & Jewelry

This hard drive comes with 8TB of storage, so for most small to medium archives, this will meet one component of the 3-2-1 rule and offers plenty of storage (although this is nowhere near the end all for your digital storage needs). It also comes with a 5 year warranty for recovery service. The maker also has 4TB, 10TB, 14TB, & 16TB models available.

If you can afford it and your budget permits, buy two, if you can!

Meanwhile, back in the FADGI Section 5.2, the Archivist has moved on from metadata in the workflow and is now taking on Production Scheduling ....

Selection of what to digitize. Evaluating its condition. Also metadata creation. Yes, all of this........

It's designed for for later processing actions. Right, Archivist?

Right you are, Red Jacket Guy. Production Scheduling flexibility means having a buffer of project work ready for digitization. Production scheduling means efficiantly creating a flow of work not creating a log jam.

5.

Now the Archivist moves on to the next step in the workflow, Digitization Prep.......

Materials are assembled in the work space, Archivist..........

Should I clearly organize information about file naming, shooting order, project specifications, and anything else????

Yes, address any special handling needs & meet me in the next step......
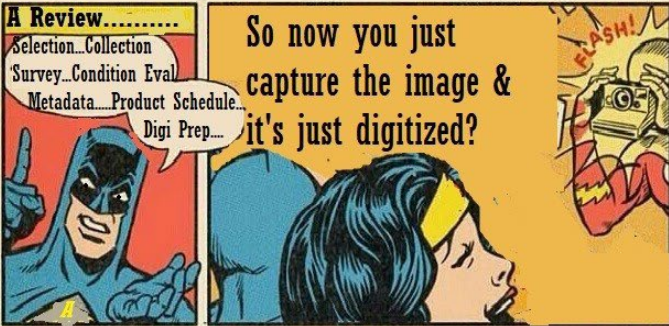
6.

Workflow Part 7: Digitization

A Review...........
Selection...Collection
Survey...Condition Eval
Metadata.....Product Schedule...
Digi Prep.....

So now you just capture the image & it's just digitized?

FLASH!

No! It varies!

What?!?

Appropriate targets!

Hmm...

It varies based off of.... type & scale......

CALLEN

FADGI

Do it right!

Determine appropriate spatial resolution....

And ICC color space...

Small targets are placed in.... monitors color... later edited out.

Test once a day.... fine tune.... monitor.... repeat.... Oh yes, repeat daily!!!!

Many materials will be housed.... handle accordingly... Position.... Avoid glare... unwanted reflections!!!!

SMAK
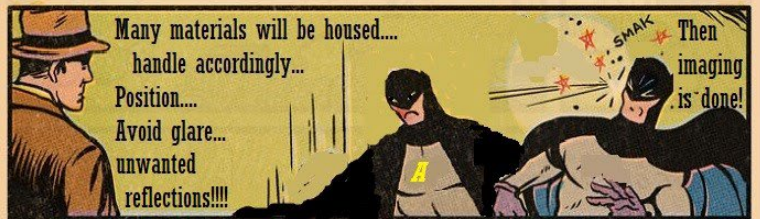
Then imaging is done!

Processes are then grouped together... Label the housing folder as imaged... The main concern is to read the image correctly, as it represents! Some materials will need multiple imaging because of size

HA HA

But there's more in the next step!

Panel 8: "So now we move on to post-processing as a workflow step.... -Thinks-" — "I know what that is."

"It involves the adjustment to the digital images we just created. It is also highly specific. Standards......Requirements!"

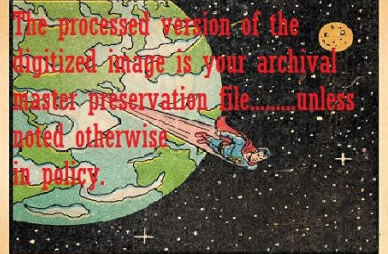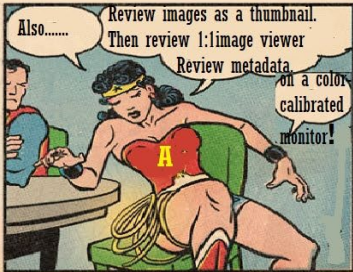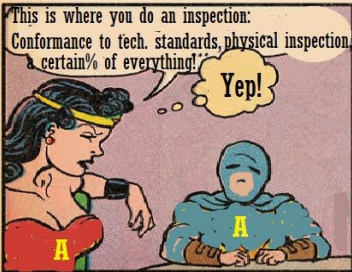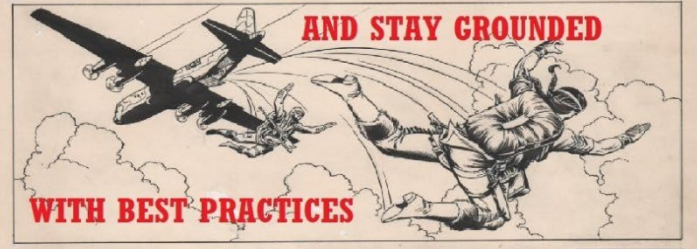"Archivists reading this.... think needed editing!"

"Now we move on to Quality Review!" — 9.

"This is where you do an inspection: Conformance to tech. standards, physical inspection. A certain% of everything!" — "Yep!"

"Also....... Review images as a thumbnail. Then review 1:1 image viewer. Review metadata on a color calibrated monitor!"

"Next step......Archiving" — ARCHIVIST — 9.

"Here comes some archival wisdom for this suggested step......"

"Images will be uploaded to staff's hard drive...... Images will also be kept on the processing computer........... Images will also be put on the organization's NAS/RAID........"

"The processed version of the digitized image is your archival master preservation file........unless noted otherwise in policy."

"One last step......."

WORKFLOW STEP 10: PUBLISHING

"Master file, you aren't suitable for public viewing!" — Master File

"An AIP is turned into a DIP."

"TIFF to JPEG / WAV to MP3 / API to MP4"

"Now you can be downloaded at home!"

"That's the shortened version"

"What? You want to know More?"

"-BE COOL-" — "FOR YOU!"

"WHAT WORKS"

"ARCHIVES IS"

"SO JUMP IN WITH A GOOD PLAN"

"AND STAY GROUNDED"

"WITH BEST PRACTICES"

"Make it work for your needs!" — REPEAT
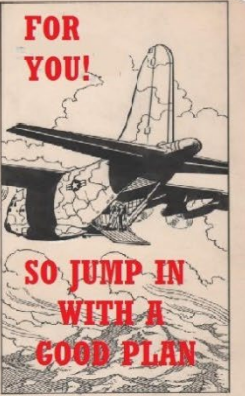
Note: A **Collections Management System (CMS)** is a software that creates, manages, & modifies archival content for public consumption. A **Digital Asset Management System (DAMS)** software stores, organizes, & shares digital media. One is accessibility driven, the other more content management and preservation driven.

Remember: Depending on your needs and capacity, you can use different tools and systems for different parts of your digital lifecycle, OR you can use systems that address multiple parts of your digital lifecycle.

These questions (developed by Jerrid Lee Miller) can be used to evaluate a DAMS:

1. Is the CMS or DAMS TERO (Tribal Employee Rights Office) certified with the tribe?

2. Is it multilingual capable? (only applies if your language has its own Unicode script)

3. Does it have a Loan Management feature- can it actually process loaning out materials?

4. Does it give user permission to other staff or public when requested?

5. Does it store the physical location of digital assets?

6. Does it have a reporting feature that can export statistical reports to interested parties.

7. Does it document copyright related information?

8. Can you assign workflow tasks through the system-who is doing what & can it be edited?

9. Can data be changed by using a batch-editing feature?

10. Can it manage different Metadata schema?

11. Can it **Create, Read, Update, Delete** (CRUD) in collection management operations?

12. Does it function as a true DAMS?

13. Does it make AIPs for preservation-will it do ingest?

14. Does it manage stored digital materials for long-term digital preservation?

15. Do you retain control of all digital materials –or- is this granted to them or 3rd person party?

16. Is this open-source (free) or proprietary (company that charges)?

17. Can you get data in and out or digitally migrate data from it if you have to?

18. Does it have an API (think integrating Apps with it) or support other like systems?

19. Is there support for IIIF (Image Interoperability Framework)-allows for zooming, etc.?

20. Does it offer support services-do they offer tech. support?

21. Is there an established community of users that you can draw advice from?

22. Is it SOC2 certified-is it certified as being certified with the American Certified Public Accountants for being

   compliant with their cybersecurity framework?

23. <u>Will it break/brake your archival budget?</u>

Tip: We strongly recommend creating your own customized list of criteria to evaluate CMS and/or DAMS!

For more resources on evaluating and comparing CMS and DAMS:

- Introduction to Digital Cultural Heritage Management Platforms by Michael Wynne on the SHN

- The Collection Management System Collection, a collaborative spreadsheet created by Ashley Blewer containing Basic, Administration, Interface, Technical, and Social considerations - link

DIGITAL ASSET MANAGEMENT SYSTEM (DAMS)

Best bang for your buck:

Preservica Starter
Cost: Free (for under 5GB of digital materials)

Cost of Digital Preservation Software | Preservica

Preservica is arguably the top-shelf, heavy weight contender for Digital Asset Management Systems out there. The Starter packet is an unconditional, no-hassle free service that allows up to 5 GB and 1 user (aka administrator) access to this DAMS for free. Upgrades to the starter pack run from $25.00-$105.00 monthly if you are needing to store anywhere from 25GB-1TB of digital materials.

The starter pack lists to many features to include here, but some of its numerous features include integrity checking with self-healing, multiple data copies stored across multiple locations (automated), preservation actions once uploaded, transformation of files (without bit loss) for preservation and access formats, drag & drop features, set files & folders to public or private (based on needs), invite & manage external content submissions, and upload multiple files at once.

Preservica was designed with being user-friendly and compliant with NDSA and OAIS specifications in mind. While it lacks a true CMS component, Archives Space (a free downloadable CMS) can easily be integrated with Preservica for free. Once those two are combined, the entire OAIS Reference Model is brought to life.

*SPECIAL NOTE: Preservica can be learnt in a few minutes, ArchivesSpace not so much.

Before committing to this DAMS, please do a sandbox demo first to see if it fits your needs.

Collection Management System

Cheapest bang for your buck:

ArchivesSpace ArchivesSpace
Cost: Free (with hosting costs)
(Tech support/membership $300 annually)

From the ArchivesSpace website: "Built for archives by archivists, ArchivesSpace is the open source archives information management application for managing and providing web access to archives, manuscripts and digital objects." You can use it to manage and displaying finding aids, as well as digital objects.

ArchivesSpace is used by anywhere from 25%-33% of archives out there. Considering there is over 42 commercially available CMS out there, that says a lot about the system. What it lacks in user-friendliness for beginners, it makes up for in a well-round performance CMS with a hugh community of users to draw on from their experiential knowledge.

If going the route of using Preservica Starter for a DAMS, please consider ArchivesSpace for its easy integration into that system and to maintain a CMS that can live up to professional standards.

*SPECIAL NOTE: Before settling on ArchivesSpace, please conduct a free sandbox demo to see if ArchivesSpace is right for you and your archival needs.

They also have a list of examples: Who's Using ArchivesSpace?

https://archivesspace.org/community/whos-using-archivesspace

CMS with focus on displaying digital items

CollectiveAccess CollectiveAccess *(used by Chugachmiut Heritage Preservation - more examples)*

Cost: Free (with hosting costs)

From the CollectiveAccess website: "CollectiveAccess is free, open-source software for cataloguing and publishing museum and archival collections." CollectiveAccess has metadata standards pre-loaded which are also customizable, workflow management features, web publishing options, handles multiple types of media, and offers granular access control. Not suitable as a complete DAMS.

Mukurtu Mukurtu (*used by Chickaloon Village Traditional Council – more examples*)

Cost: Free (with hosting costs)

From the Mukurtu website: "The free, mobile, and open source platform built with Indigenous communities to manage and share digital cultural heritage. Mukurtu is a grassroots project aiming to empower communities to manage, share, narrate, and exchange their digital heritage in culturally relevant and ethically-minded ways. We are committed to maintaining an open, community-driven approach to Mukurtu's continued development. Our first priority is to help build a platform that fosters relationships of respect and trust." Features include: flexible MukurtuCore metadata, TK Labels, granular access control using Cultural Protocols, Community Records, and Roundtrip batch import/export. Not suitable as a complete DAMS.

Keep in mind there are many more options for CMS with digital access options, as well as DAMS/CMS with options for displaying finding aids and/or digital items online.