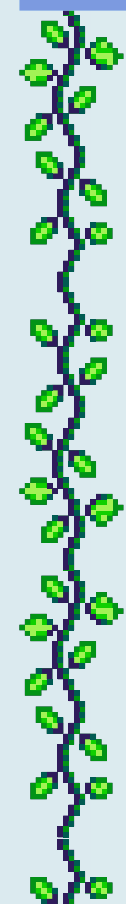
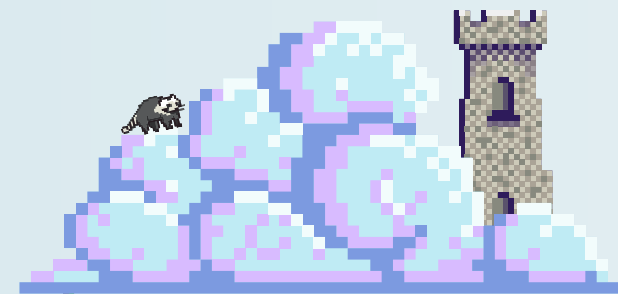


> champing at the **bit**

Experiences Designing a User-Centered Born-Digital Processing Workflow

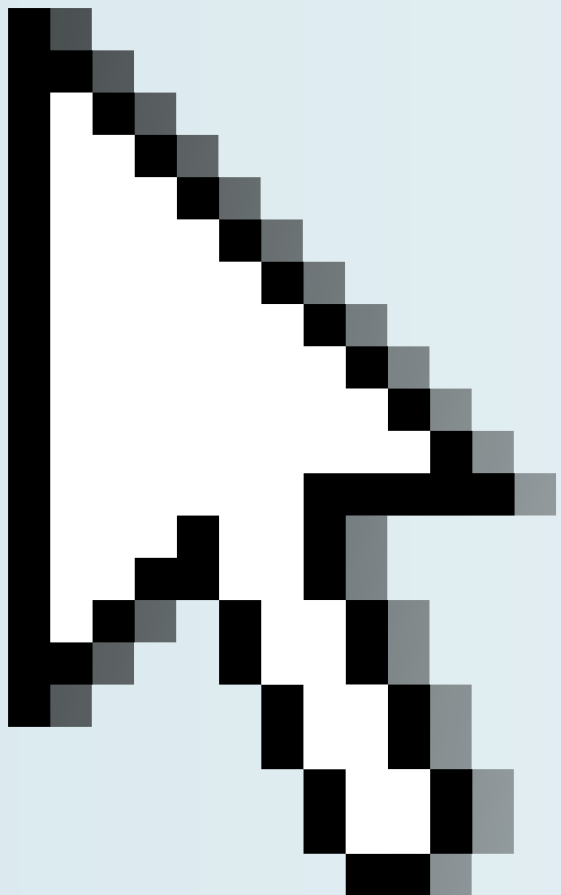
Presenters: Kayla Cavanaugh and Emma Eubank
Graduate student interns, Duke University Medical Center Archives

Society of North Carolina Archivists Annual Meeting
May 21, 2025



> agenda

1. About us & about Duke University Medical Center Archives
2. Contextualizing our guide – the whys and goals
3. A walk-through of the pre-ingest workflow at DUMCA
4. Questions that guided our decision-making
5. Time for sharing of digital archiving experiences among audience members



> about Emma and Kayla



Both interned at DUMCA, 2023-2025

Both graduated in May 2025 with MSLS from UNC Chapel Hill School of Information & Library Science



Both adore Sandy the hamster and assorted pets found in DUMCA collections.



Collage artist! →



← While we do have a Golden Shovel of Truth, DUMCA's collections unfortunately do not include the papers of Dr. Jacoby.

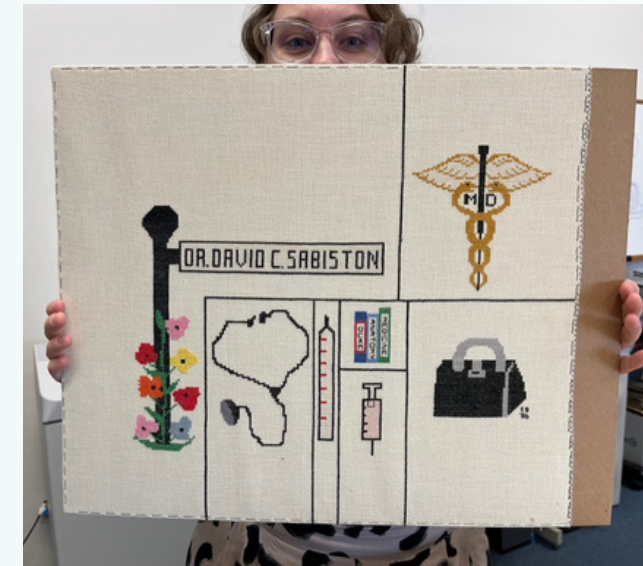
> about

Duke University Medical Center Archives

Located in Durham, NC, overlooking beautiful I-85!

“Serves Medical Center customers and the global community by preserving, maintaining, documenting, and making available for research the permanent records of the functions of the DUMC and Duke University Health System (DUHS), including health care delivery methodology, research, education and training.”

Offers access to **born-digital files**, including documentation of clinical trials, hyperbaric chamber research data, medical library website iterations, and more.



Duke University
Medical Center Library & Archives

> about

Duke University Medical Center Archives

From the Medical Center Library & Archives 2024 Snapshot...

We maintain administrative and historical records for Duke Health, including the records of 2 Nobel Laureates, World War II surgical teams, School of Medicine and School of Nursing departments, and student groups.

In fiscal year 2024, we:

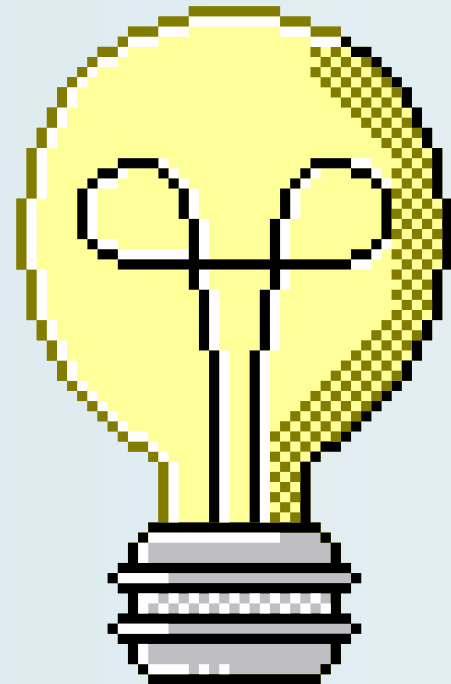
- Added over **5,500 digital files** and over 151 linear feet of print items
- Provided 4,597 scans of Archives documents which allowed remote access to collections
- Answered a total of 886 Archives reference questions
- Conducted 146 research consultations



> why?

Our internship experience provided us the opportunity to design a pre-ingest born-digital processing workflow.

This experience got us thinking about user interactions with born-digital files at DUMCA.



Are users champing at the bit to access digital materials?

How do you translate archival decision-making in a clear and transparent way, while not quite knowing who will want to access materials?

What does it mean to create a pre-ingest guide for high-turnover archivists in our roles as interns?

> context

How we began developing the Pre-Ingest Guide

Envisioning our goal: A guide which empowers the archivist to make decisions on born digital files and prepare them for ingest. This guide had to flow into DUMCA's Electronic Records Processing Guide, which outlines bagging and ingesting files.

Institutional buy-in & support: Our guide builds directly on work by former DUMCA intern Carter Hulinsky, and shows DUMCA's willingness to let interns learn and contribute.

Finding our born-digital community: We welcomed information & guidance from fellow practitioners, knowledge-holders, and collaboratives.

- Resources developed by the Digital Archival traNsfer, iNgest, and packagiNg Group (DANNNG), Digital Preservation Coalition, and LibGuides created by institutions
- Joining the TRLN Born-Digital Special Collections Interest Group
- Thank you to Farrell, Tom, and Lulu!

> workflow summary

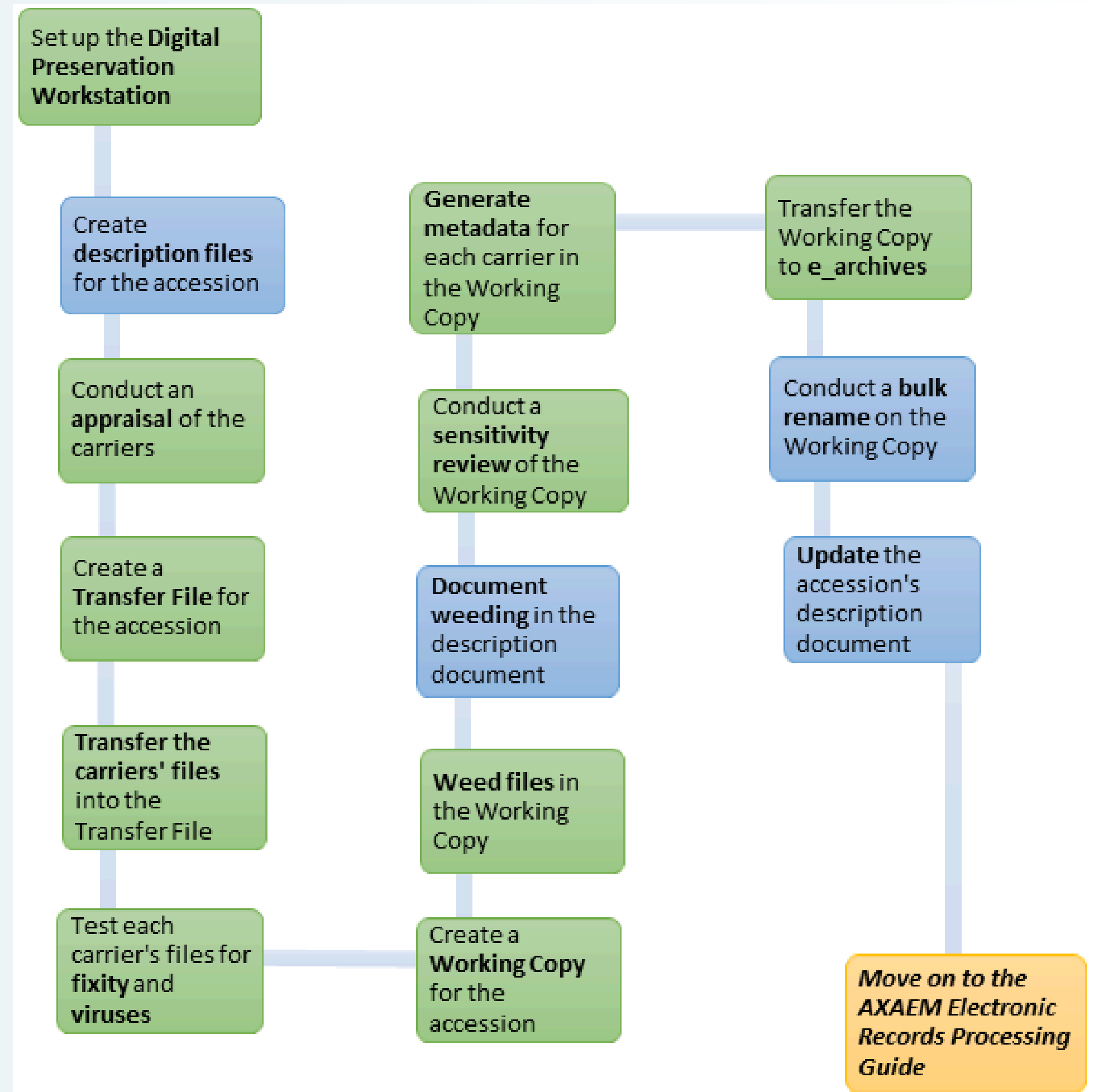
A bird's-eye view

Green: completed on DUMCA's Digital Preservation Workstation

Blue: completed on any Windows computer

Yellow: marks the end of the pre-ingest stage

How do these steps affect a user's experience using files?



> terminology

How information in our guide is communicated

Our guide is written with future interns in mind, who are early in their career, as well as archivists who are new to digital archiving practices.

- A human-created glossary
- Basic Linux commands, screen captures, and step-by-step instructions

```
5. Use the following command: tree [carrier directory] > [metadata output directory]/[carrier directory]_tree.txt
   a. For example, tree CD_1 > A2020_035_metadata/CD_1_tree.txt
```

- Language that emphasizes the collaborative nature of archiving born-digital records
- Encouragement to try new things and think strategically about user access

> glossary

Example of a glossary entry:

Master copy (preservation copy)

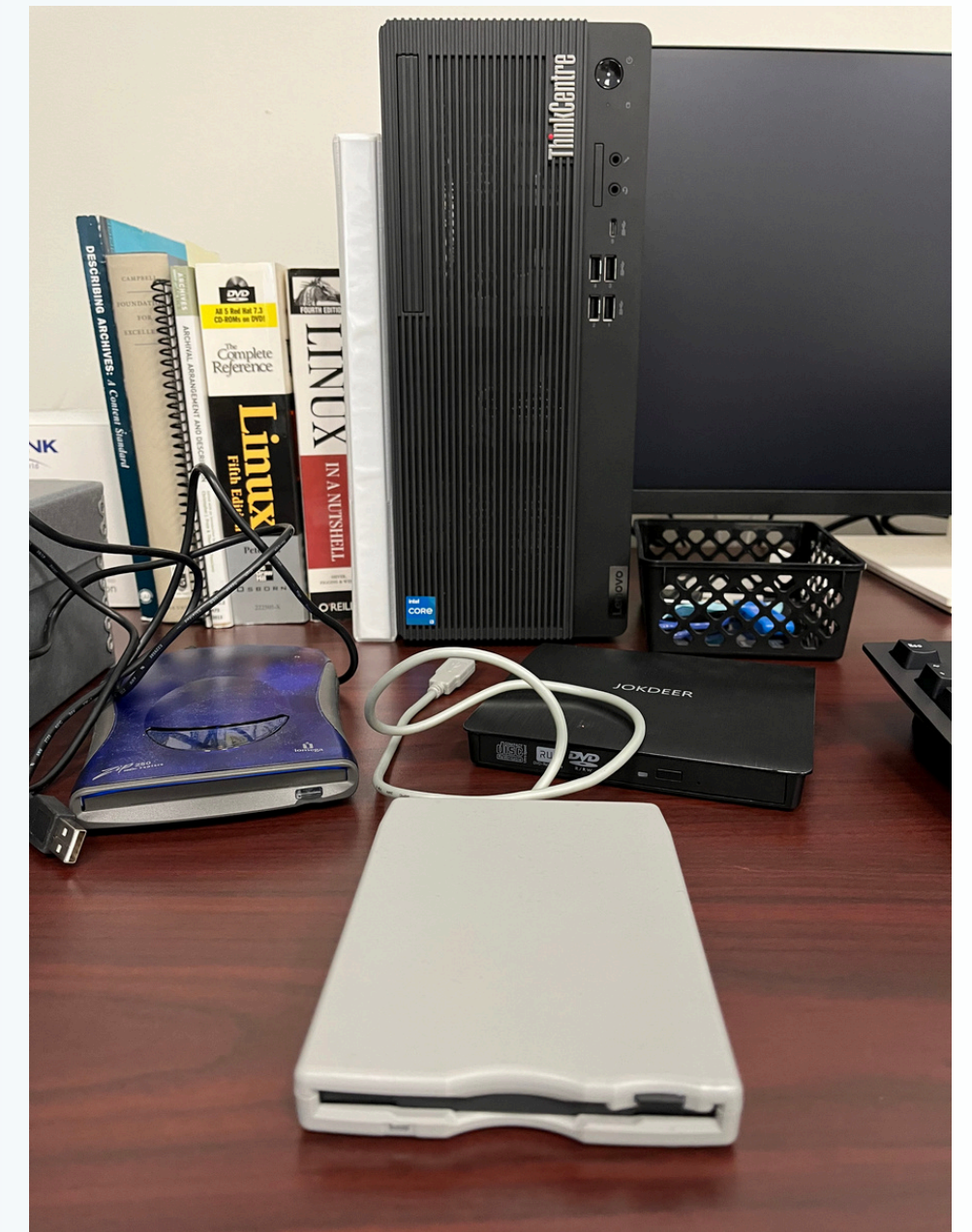
In the past, this term has been used to refer to the **preservation copy** of an accession's files, meaning copies of files that present the most accurate picture of a carrier's contents at the time of accessioning. The term "master" carries historically laden meanings that can be harmful. Ideally, "preservation copy" should be used whenever possible. Be aware that "master" likely will come up in older instances of DUMCA digital preservation workflows.

DUMCA Pre-Ingest Electronic Records Processing Guide 31

Other terms include **bit rot**, **disk imaging**, **file fixity**, **terminal**, and more.

> tools

Digital Preservation Workstation



Our set-up runs Linux and supports a variety of physical media readers.



> documentation

Create description files for the accession

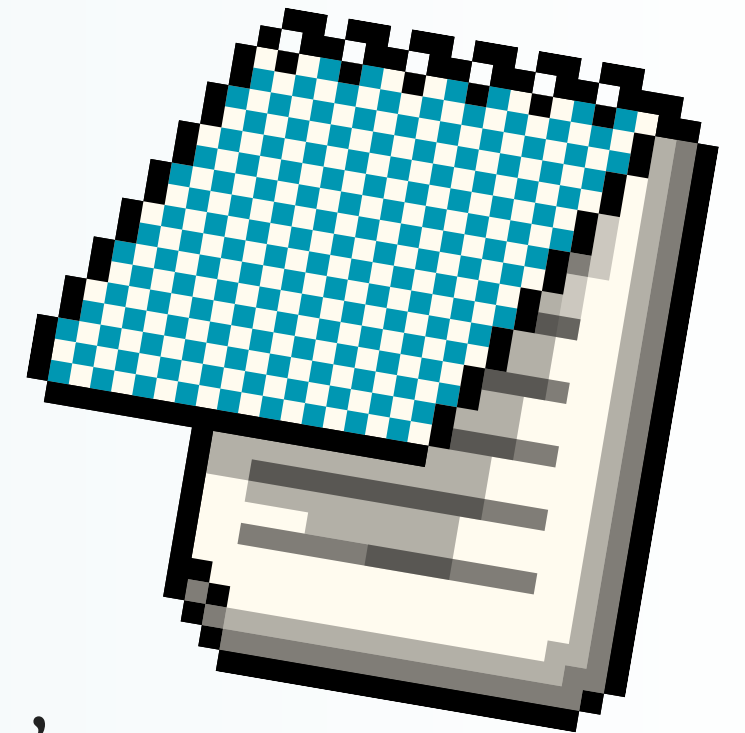
Documentation ensures accountability is built into the workflow and allows us to communicate indirectly with future processing archivists.

Media inventory (.xlsx): records details about actions taken upon an accession's digital files as they are processed

- Carrier information
- File name changes
- Restricted files
- Initialing of completed steps

Description document (.docx): records information about the accession's digital files as they approach their final form, and (sub)series description

- Weeding changes and ingest information



> appraisal

Conduct an initial survey before creating the Transfer File

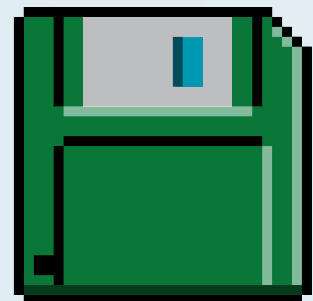
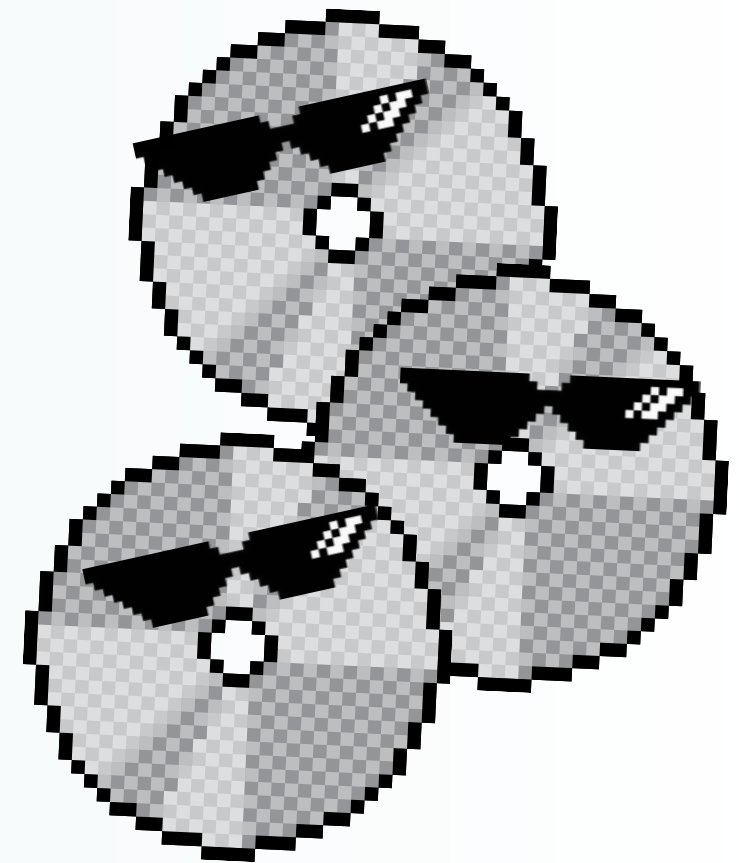
Does the general subject matter of the files **align** with the larger collection?

Do any of the files look like they fall under **HIPAA** or should be restricted?

Are any of the files **uplicated**? Are there duplicates across carriers?

Are any of the carriers **mass-produced**, or do they hold original content?

Do any of the files appear to contain **malware**?



→ Follow a naming convention to name each carrier.

→ Document all appraisal decisions.

> creating a transfer file and a working copy

Copy & test files from the carrier

Transfer File: preserves carriers' files in their original state

- Organized by carrier, preserving original order of each carrier

→ **Conduct a fixity check and a virus scan, and then create the Working Copy.**

Working Copy: testing ground for archivist to decide on file relevance, duplicity and sensitivity, and then holds generated metadata outputs

- Maintains carrier separation of Transfer File

→ **Later (beyond the pre-ingest workflow), a Preservation Copy will be made from the Working Copy.**

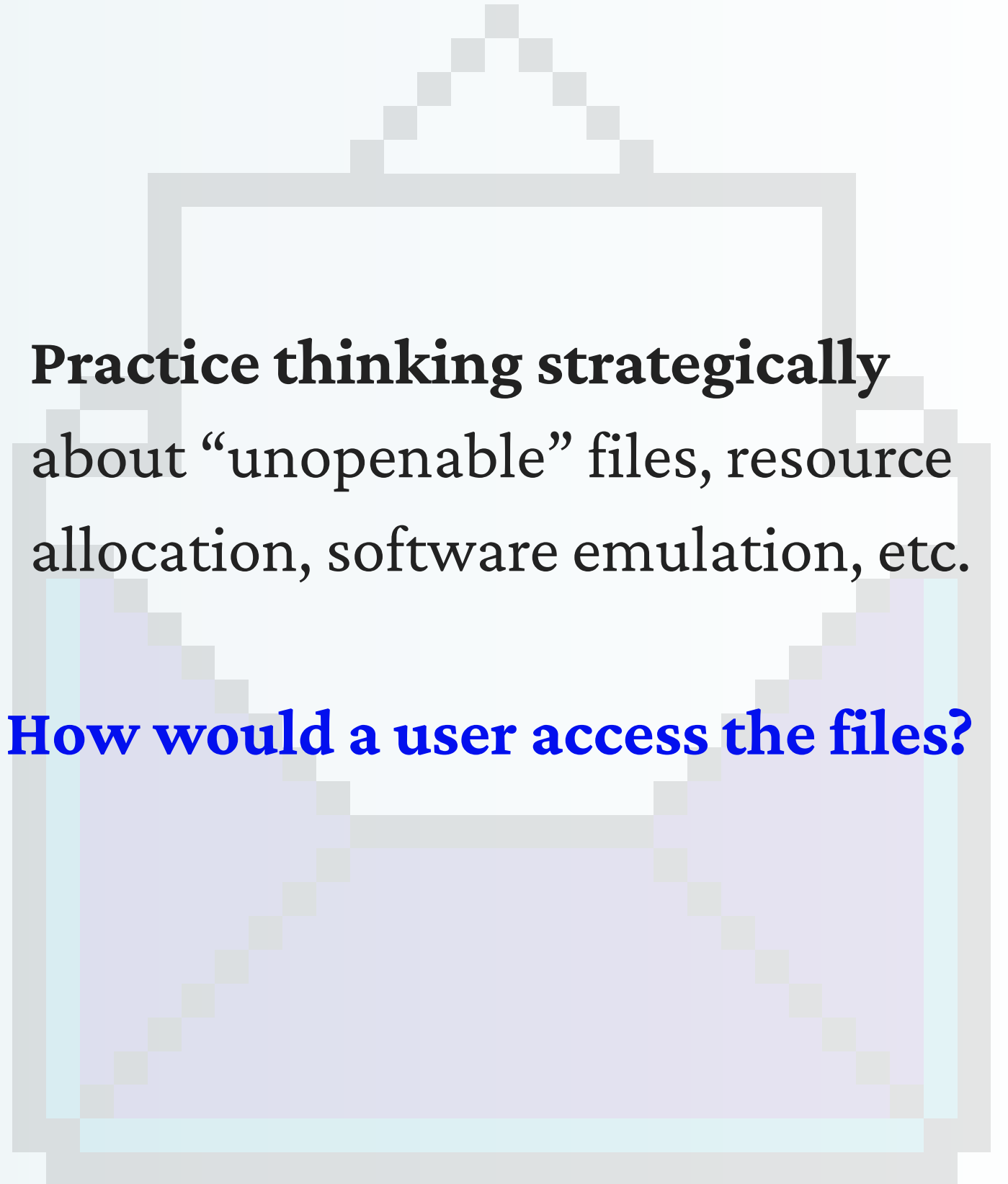
> appraisal

An iterative process

Weeding

We keep files that are:

- Created by humans
- Are not mass-produced
- Contain original content
- Able to be opened with DUMCA resources



Practice thinking strategically
about “unopenable” files, resource
allocation, software emulation, etc.

How would a user access the files?

> appraisal

An iterative process

Conducting a sensitivity review

We use human eyes and Bulk Extractor.

Common contexts for PHI and PII:

- Bulk data
- Grant materials
- Internal reports
- Internal correspondence
- CVs

“[HIPAA] protects all individually identifiable health information held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral.”

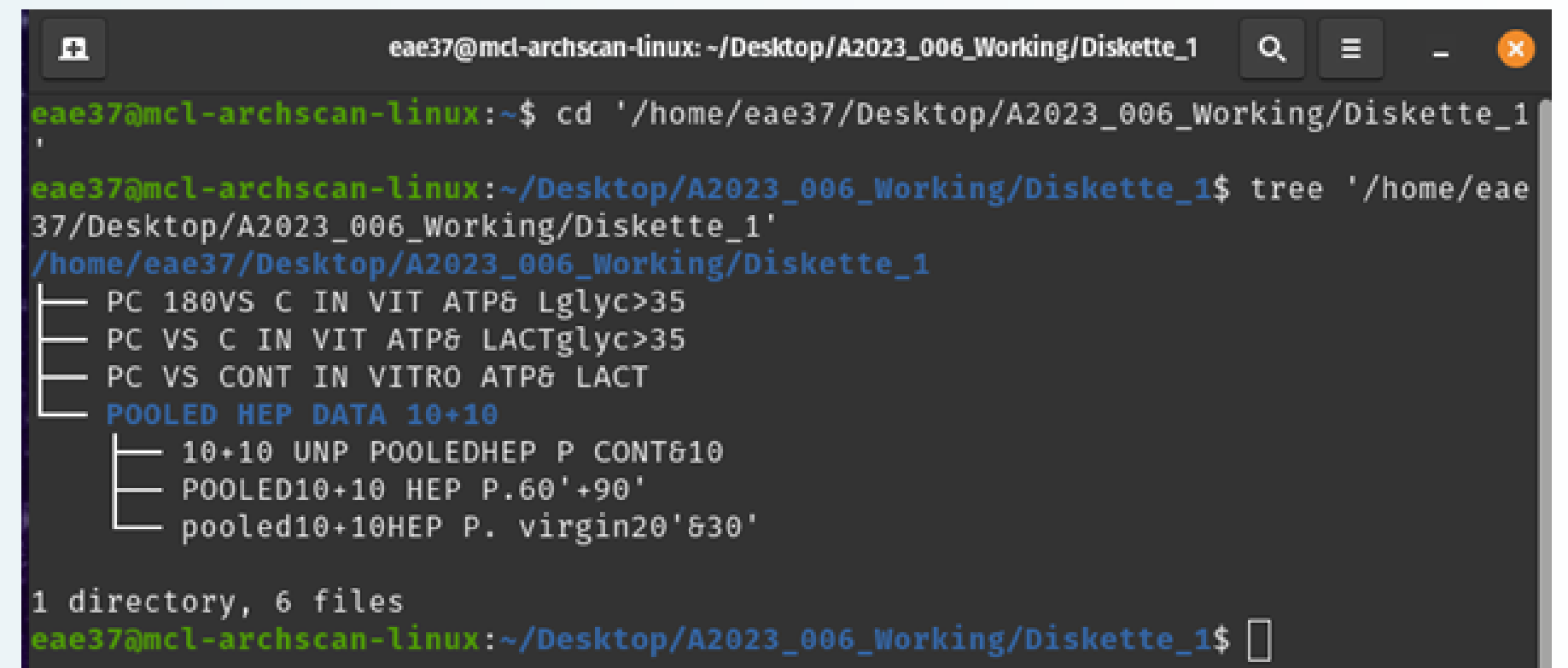
*US Department of Health and Human
Services*

> metadata

What might future users want to know?

Users need a way to navigate an accession's digital materials, so the metadata we generate **by carrier** includes:

- Tree
- File Type
- Last Modified Date
- Optional: Files by Size, Directory Size



```
eae37@mcl-archscan-linux: ~/Desktop/A2023_006_Working/Diskette_1
eae37@mcl-archscan-linux:~$ cd '/home/eae37/Desktop/A2023_006_Working/Diskette_1'
eae37@mcl-archscan-linux:~/Desktop/A2023_006_Working/Diskette_1$ tree '/home/eae37/Desktop/A2023_006_Working/Diskette_1'
/home/eae37/Desktop/A2023_006_Working/Diskette_1
├── PC 180VS C IN VIT ATP& Lglyc>35
├── PC VS C IN VIT ATP& LACTglyc>35
├── PC VS CONT IN VITRO ATP& LACT
└── POOLED HEP DATA 10+10
    ├── 10+10 UNP POOLEDHEP P CONT&10
    ├── POOLED10+10 HEP P.60'+90'
    └── pooled10+10HEP P. virgin20'&30'

1 directory, 6 files
eae37@mcl-archscan-linux:~/Desktop/A2023_006_Working/Diskette_1$
```

→ Save the metadata outputs as .txt files in the Working Copy folder.

> a return to our guiding questions

Are users champing at the bit to access digital materials?

How do you translate archival decision-making in a clear and transparent way, while not quite knowing who will want to access materials?

What does it mean to create a pre-ingest guide for high-turnover archivists in our roles as interns?

> space to share



> keep in touch

Kayla Cavanaugh, MSLS

PhD Student, Information and Library Science

UNC Chapel Hill

kecavena@unc.edu

Emma Eubank, MSLS

Contractor, Library of Congress

emmaeubank2000@gmail.com

