Accessioning and Stabilization Workflow

The accessioning workflow also includes what I refer to as the stabilization workflow (see figure 4.1). By stabilization, I mean that all the steps taken in this workflow make it so that the digital materials can survive in a state of benign neglect for as long as it takes your institution to move from accessioning to processing. However, benign neglect for digital materials is slightly different from what is done with paper-based materials. It requires that the digital materials be periodically audited and preservation storage be maintained.

The steps in this workflow could be individually done by human actors working with separate tools for each part of the workflow. At the opposite end of the spectrum, the workflow could be almost entirely automated, where the only necessary intervention from a human would be metadata creation and upload of data into a digital asset management system. Most institutions will fall somewhere between these two extremes. The guidance I provide in this section is meant to be high-level enough for you to adapt the strategies for your institution but practical enough for an institution just starting out to have a detailed enough road map to feel confident in moving forward building out its own workflows.

For each section of this and subsequent workflows, there are a variety of tools and services available to complete the steps. A community-built and -maintained directory of these tools, COPTR, is available for you to look through and determine the best possible avenue for your particular institution to take. There is also a more detailed and organized chart of these tools produced by POWRR.[1] However, this chart is a little out of date, so it could include tools that are no longer available, or the tools listed could have improved functionality that is not included in the discussion section of the chart.

Create Accession Record

The creation of an accession record is a process in and of itself, unique to each institution. There are commonalities in all the methods because the accession process is meant to capture intellectual and physical control over the materials acquired by your archives. The accessioning process starts with your donor engagement. All of the documentation that you generate while developing a relationship with the donor should be gathered together in a file.

My institution errs on the side of caution. Our practice is to gather the donor correspondence, the deed of gift, and printouts of any digital correspondence, which goes into an accession master file that is typically generated for the first accession in a collection. For those accessions that include digital materials, the accession master file should also include the donor survey or interview and any notes made by the archivist during a site visit. For additions to the collection that are separate accessions, either any documentation gets added to the original master file or a new accession master file is created. All of this documentation forms the legal evidence of custody, the written and signed agreements regarding what can and cannot be done with the materials, any restriction requirements, and the reference material that helps generate a description record later on in the processing workflow.

Beyond this accession master file, institutions typically maintain an accession database that contains an accession record for every set of materials acquired by the institution. This database could be a Microsoft Access or Microsoft Excel table, an ArchivesSpace instance, or a homegrown institutional web database.[2] In general, this piece of the accession record includes who donated the material, when it was donated, who

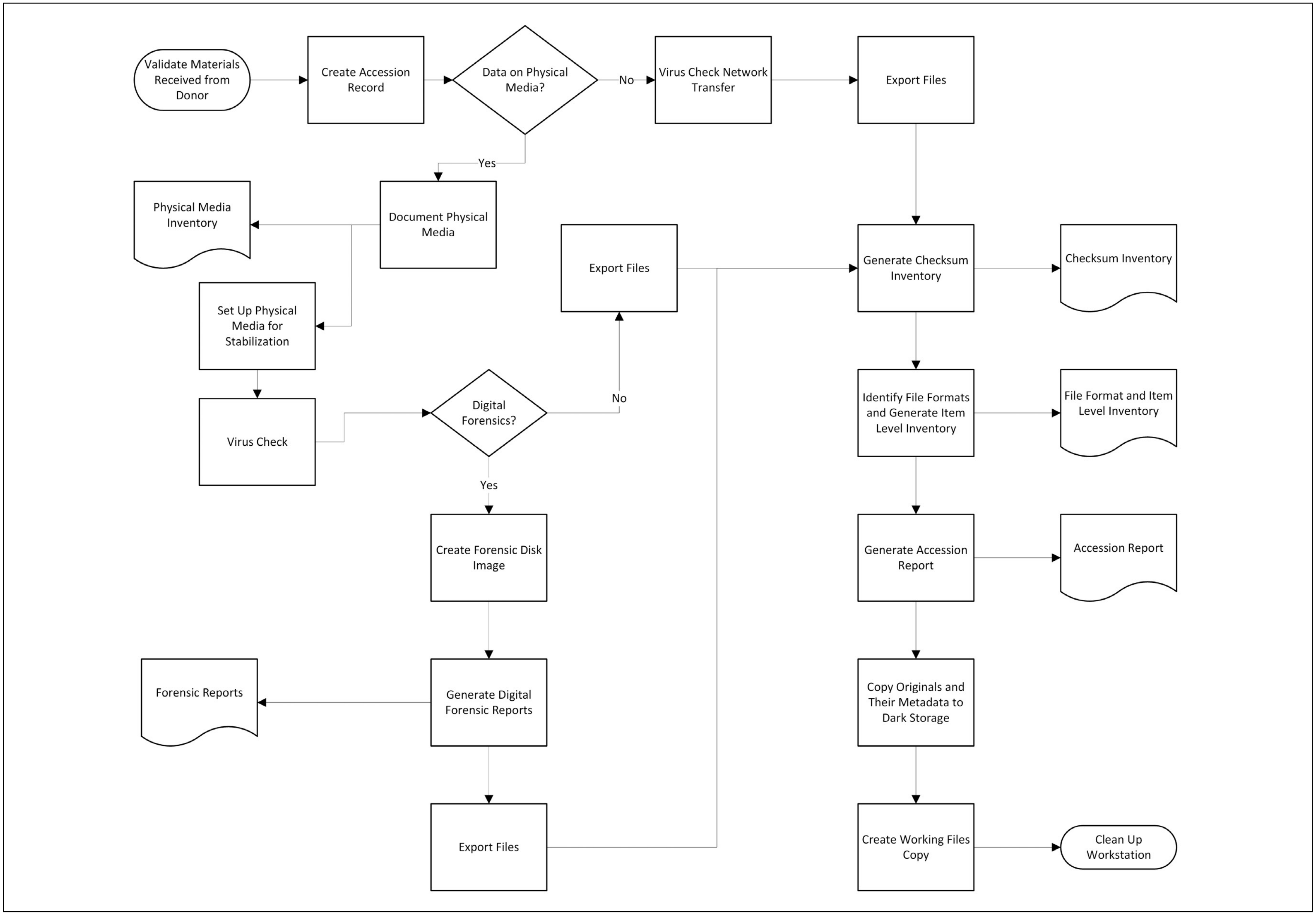
Validate Materials Received from Donor
Create Accession Record
Data on Physical Media?
No
Virus Check Network Transfer
Export Files
Yes
Document Physical Media
Physical Media Inventory
Set Up Physical Media for Stabilization
Virus Check
Digital Forensics?
No
Export Files
Yes
Create Forensic Disk Image
Generate Digital Forensic Reports
Forensic Reports
Export Files
Generate Checksum Inventory
Checksum Inventory
Identify File Formats and Generate Item Level Inventory
File Format and Item Level Inventory
Generate Accession Report
Accession Report
Copy Originals and Their Metadata to Dark Storage
Create Working Files Copy
Clean Up Workstation

Figure 4.1
Diagram of a high-level accession and stabilization workflow

created the material, when it was created, approximately how much material was donated, and what kinds of materials were included in the donation. This information could be more detailed and broken out into subcategories depending on institutional processes, but the metadata generation process is typically the same regardless of what type of material is donated: paper, digital, or a mixture of the two.

Physical Media Pathway

Depending on how much involvement the archivist responsible for stabilizing the digital materials has with the donor, it may not be clear if an accession includes digital materials until the acquisition arrives at the archives. This is why an initial rough inventory is essential, preferably before the items are retrieved from the donor, but at the latest when materials arrive at the archives. Physical media carriers, such as hard drives and floppy disks, are much more sensitive to environmental changes and being dropped or jostled than paper records. I encourage you to separate these physical media carriers from their place among the analog records during this initial inventory. Depending on the circumstances, you may place a flag or separation sheet indicating that a floppy disk, CD, USB drive, or other physical media carrier was removed from a folder or box, but if there is no clear order to the materials, there may not be a need for such a flag or separation sheet.

Document Physical Media

When physical media carriers are part of an accession, try to make sure you have found all of the items before moving forward with the workflow. It is possible to add materials later, but some of the stabilization steps that are done at the very end of the workflow are meant to be done on the entirety of the accession at once. Once you have collected all the physical media carriers, give each media carrier an identifier. At my institution the identifier is very simple: the accession number underscore running item number. If there are fifty physical media carriers in an accession, the first would have the identifier AccessionNumber_001 and the last would have the identifier AccessionNumber_050. Other institutions include the type of physical media carrier as part of the identifier, for example, AccessionNumber_DVD_001. Optionally, this identifier can be directly written on the physical media carrier, or a sticky note with the number could be placed on the physical media carrier.

After an identifier has been assigned, you are now able to add the details about these carriers to an existing inventory for all the physical media carriers in the archive or create an inventory for the accession alone. Your institution's documentation workflows and systems will determine how you approach this task. The documentation for each physical media carrier and network transfer should include

- identifier
- accession number of the accession the materials belong to
- transfer type: network, physical media, e-mail, or digitized
- date the materials were acquired
- who the materials were received from
- who in the institution received the materials
- where the media is stored: physical location or server location
- media format: optical, flash, or magnetic
- media subtype: CD-R, DVD, USB flash drive, and so on
- manufacturer
- model number
- approximate age of the physical media carrier
- condition of the physical media carrier
- media label text (if any)
- if the media has been photographed

In some institutions, this inventory is also where the steps taken to stabilize and process the digital materials are documented. However, at the bare minimum, the items listed above should be recorded during the accessioning process. As you fill in the metadata for each physical media carrier, photograph the item, front and back, and include those photographs with the metadata for the accession so that you do not have to retrieve the physical media carrier for reference whenever you are working with the digital files for the accession. These photographs are also of interest to researchers because they provide researchers with a way to view how the creator organized and documented the physical media carriers without having to look at the actual items.

I want to make clear that the above documentation does not need to occur if you are retrieving digital materials from a donor using your own external hard drive or other temporary physical media storage device because that is just a transfer mechanism, not a permanent addition to the archive. Additionally, the above documentation will not need to occur if the donor expects you to return the physical media carrier. Again, the device is a method of transfer and not a permanent addition. However, all the subsequent steps for physical media will need to be performed.

Set Up Physical Media for Stabilization

First, gather any external hardware that you need to access the content on the physical media you are stabilizing. You will probably need to have on hand a hub that allows you to access different types of flash media, such as SD cards, microSD cards, and proprietary digital camera memory cards. You will also likely need two or three different 3.5-inch floppy drives. Interestingly, even if there is no appreciable difference between floppy disks, sometimes switching between drives will increase the likelihood that you can access the content on the floppies. There is also a possibility that you will need to acquire a system to access 5.25-inch floppy disks and Zip disks. These are the most common materials I have found at the different institutions I have worked at. After setting up the external hardware to access the media items, you will need to set up your write blocker.

Anytime a computer's operating system interacts with physical media carriers, there is the potential for the technical metadata attached to each digital file in that physical media carrier to be inadvertently altered. To prevent this, physical media needs an additional safeguard of a write blocker setup before media is connected to your digital materials processing station. As an aside, I encourage you, if you have the resources, to have a standalone station dedicated to the stabilization of digital materials. This is a major safeguard against viruses that not only protects the unstabilized physical media and newly transferred digital materials, but also protects those materials already in your preservation system.

A write blocker can be a physical device or a piece of software that is turned on and off as needed. My institution uses a combination of both. The software is depended upon only if the hardware write blocker is incapable of interacting with the physical media. This often occurs with physical media older than a 3.5-inch floppy disk. After the write blocker has been set up, you now need to choose if you will be doing digital forensics or simply making a logical copy of the files on the physical media device. I mention digital forensics solely in the physical media section because many of the software tools used by archivists that perform digital forensic analyses require disk images, which are most commonly acquired from physical media items.

Virus Scan

Before doing anything, check the materials for viruses. I am using the term virus to cover any malicious programming that could be found on materials you are bringing into the archive. Most computers now come with built-in security that includes the ability to scan discrete sections of files for viruses and even allow for the quarantine of affected files. Generally, when viruses are found, you will need to involve your information technology department to determine if the virus can be dealt with or if you should deaccession the affected material. There are open source virus resolution solutions available, one of which is integrated in the BitCurator suite of software. However, it may be that the information technology department in your institution has a workflow regarding virus checking and remediation that it would prefer you to follow, so be in constant communication with your information technology experts.[3]

Forensic Disk Image

If you have decided that a physical media item warrants digital forensic analysis and you have the donor's informed consent to perform this analysis, you may then proceed with this section of the workflow. Using the digital forensic software suite of your choice—BitCurator, KryoFlux, Forensic Toolkit (FTK), or some other tool—create a forensic disk image for each physical media item.[4] This is a bit-for-bit copy of the physical media including empty space. Part of the creation process will include embedding metadata about who is creating the disk image, the identifier the disk image is associated with, and so on, into the disk image file. After you have created the disk image, you will then export the digital files from the disk image and export the technical metadata associated with the disk image, specifically the file system metadata that includes how a computer was used by the creator, deleted files that have not yet been overwritten by the operating system, and more. Be careful to make sure that all the metadata files are named in such a way that they remain associated with the disk image file. This is most easily done by using the physical media identifier as part of the file and folder names.

Logical Copy

If you have decided that in-depth digital forensics is not warranted, there are several methods you can use to make a logical copy of the files on the physical media devices you are using. If you have a digital asset management system, homegrown or subscription service, you will most likely be using the tools built into that system to copy the files directly into the system. If your protocols require the transfer of files to a local machine before they are moved into the digital asset management system or if you do not currently have a digital asset management system, you need to use a piece of software that will perform multiple functions at once. Ideally the software will copy over files without changing the internal file metadata such as creation date or creator, will generate a report containing the technical metadata for each file, and will verify that the files were transferred without any loss or change to the data of each file. I use DataAccessioner for this, but there are other options.

Network Transfer Pathway

By network transfer, I mean any receipt of materials not on a physical media carrier, so this could include e-mail, materials donated through a web form, those placed in a cloud storage service for you to download, or an actual network transfer using file transfer protocols. When possible, communicate and test out the transfer methodology you and the donor have agreed to. The tests will help reduce the chance of technical problems due to hardware or software incompatibilities, and the donor will feel more confident in using the method to transfer materials to your institution. This confidence is especially important if you expect that the process will need to include multiple transfers. The best-case scenario for a network transfer is to use a tool that will package the data and will verify that the package was unchanged after the transfer is complete, such as Exactly.[5] The packages can then be sent via any method that works well for you and the donor.

Stabilize Accession

After you have transferred all the materials for an accession to your working space, it is time to perform the stabilization process. This is accomplished by generating technical and administrative documentation for the entirety of the accession, not just the discrete parts. Then move the accession as a whole into your preservation storage system.

Generate Technical Metadata

While you may have generated technical metadata, such as file creation date, file format type, or a checksum, using discrete tools for the individual pieces of the accession, it is a good idea to generate one single listing of all of this information for the accession as whole. This allows you to have all the information in one place and also doubles as a file-level inventory of the materials, including their current location in the overall directory for the entire accession. This one document is essential to later appraisal, preservation, and arrangement and description work. Even though it seems like repeating steps you have already completed, it is really worth doing. The tool we use at the University of Montana for this is DROID.[6]

Generate Reports

The final documentation step here is to generate some kind of accession report. This report is meant to be a brief overview of the entire accession and will be used later when you are processing the collection. The accession report is also a way to quickly remind yourself of what is in the accession. This could include any potential preservation issues that may need to be remediated in the future, such as unique file types that require specialized software to access or older file formats that will need to be normalized into a standard preservation or access format. Our accession report asks the archivist to document the following:

- Overview
  - accession number
  - deposit date
  - transfer type
  - collection number (if known)
  - creator
- Physical details
  - number of media
  - extent/data size
  - number of files
  - file format types
  - preservation issues
  - preservation recommendations
- Intellectual details
  - current organization
  - date range
  - content summary
  - privacy issues
  - donor restrictions
- Report author
- Report date

Move Accession to Storage

After all the data has been transferred, stabilized, and documented, you can now transfer the accession to your preservation storage environment. My recommendation is to transfer the digital materials into two separate locations. One set will go to preservation storage, and one set will go into working files storage. The preservation storage set should be the unaltered originals in your "dark archive," the place that very few people have access to and is accessed only to remediate cases of accidental or deliberate corruption of your working files copy. The working files are what you will process. These will eventually become the files your end users have access to. At this point, your materials can wait however long is needed until you are ready to process them.

Notes

1. "Community Owned Digital Preservation Tool Registry (COPTR)," main page, last modified June 4, 2020, http://coptr.digipres.org/Main_Page; "Tool Grid," Digital POWRR: Digital Preservation Research, https://digitalpowrr.niu.edu/digital-preservation-101/tool-grid/.
2. ArchivesSpace home page, https://archivesspace.org/.

3. BitCurator home page, http://bitcurator.net/.
4. BitCurator; KryoFlux home page, https://www.kryoflux.com/; "Forensic Toolkit (FTK)," AccessData, https://accessdata.com/products-services/forensic-toolkit-ftk.
5. "Exactly," AVP, https://www.weareavp.com/products/exactly/.
6. "DROID (Digital Record and Object Identification)," GitHub, https://digital-preservation.github.io/droid/.