Chapter 3

Acquisition Workflow

Building Relationships with Donors and Creators

In many cases donors and creators are one and the same, but that is not a universal truth. In terms of digital material, engaging with creators is almost more important than working with a donor, especially if they are separate entities. The creators provide vital contextual clues that are much more hidden in digital materials than traditional paper materials. Creators control how files are created, named, and arranged in their original operating environment. All of this has a direct effect on if and how digital materials can be preserved and provided to researchers in the future. The earlier you can build a relationship with a creator, the more likely it is that digital materials will survive to be used by a future researcher.

Beyond the technical and contextual aspects of digital preservation, there are strong ethical reasons for engaging donors and content creators frequently during the entire process. This engagement can prevent archival silences and the biases of archivists affecting the appraisal, arrangement, and description of the materials. Engaging donors and giving them direct power over how their materials are preserved, described, and provided to researchers ensures that trust is maintained with the donors and that more perspectives and voices are represented in the archives.[1]

Also, due to emerging post-custodial concepts, it may be that materials never directly enter your archive. That does not mean that these workflows are made obsolete. Instead, it will be your responsibility to adapt the workflows to fit a model of shared custody or to use them as a tool to teach community archives how to maintain and provide access to their digital materials. The key to digital preservation is adaptability. At this stage, you will have to adapt to the needs, cultural and personal, of your donors to ensure that the digital materials they create are integrated into the archival record.

The Importance of Informed Consent

Engaging with donors and creators allows you to ensure that they are fully informed about the ramifications of donating digital materials that will eventually be available to researchers. Informed consent is a concept that I first encountered in medicine, but it is also commonly found in research. It has also been an issue more recently in disclosure of personal information gained by companies like Facebook when users interact with their services. In medicine in many countries, informed consent is a legal right that patients have. It mandates that all the risks and benefits of medical intervention be explained to a patient in a manner that the patient understands before the patient legally agrees to the procedure. I will say that this is the intent of informed consent. In practice, some explanations patients are given are similar to the terms and conditions forms consumers are asked to agree to before downloading an app, seemingly written in a different language. There has been a strong push by patient and consumer advocates to mandate that the way in which informed consent is achieved be simplified to make sure that those affected truly understand what they are risking (or giving away) in return for the service received.

Informed consent applies to digital preservation in many different ways, including how materials are retrieved from donors, preserved by archivists, and made available to researchers. Personally identifiable information and clearly private information that can

be easy to spot in paper materials can be hidden in the metadata and system files of digital materials. Donors are often not aware of this hidden data or even the amount of data that can be pulled off a device through digital forensics or that is present in underlying metadata in files that were directly copied from the donor's operating system.

Gaining informed consent from your donors about what they are donating and potentially providing future researchers access to is a critical aspect of the donor engagement process and the acquisition workflows. There are various strategies for ensuring that donors clearly understand what they are agreeing to. These strategies are adaptable to a donor's demonstrated level of understanding of the technical aspects of digital preservation. The ultimate goal is to disclose to donors all pertinent information regarding the donation of digital materials, including the possibility of disk image creation and digital forensic analysis, in a manner the donors can understand. Ideally, you would have a donor demonstrate this understanding in some manner before signing a deed of gift that explicitly includes sections related to digital materials that the donor must agree to separately.

I have developed a list of goals for ensuring that I have done everything possible to achieve informed consent:

- The donor is aware that digital materials will be transferred as part of the donation and that it is possible that a disk image may need to be created and digital forensic analysis performed.
- The donor is aware that they can restrict access to all or a portion of the digital material, disk image, and digital forensic reports.
- The donor is aware that they can refuse to allow a disk image to be created or digital forensics to be performed.
- The donor has shown me that they understand the information that can be found within digital materials, disk images, and digital forensic reports.
- The donor consents to the final agreed-upon digital preservation plans for their materials, in writing.

The mechanisms that help me achieve these goals include

- a deed of gift with a section dedicated to the unique permissions required for digital materials
- a donor interview, a donor survey, or both regarding the creation and use of their digital materials
- a demonstration of where information is hidden in individual file metadata and file system metadata overall and the kinds of information that can be found within the reports generated by digital forensic analyses
- providing donors with all policies and procedures regarding digital materials and carefully walking through how those policies and procedures would apply to the donor's materials

It may seem like a lot of up-front work with donors before materials are ever acquired by your institution. However, these interactions only increase the trust your donors will have in your institution and in your abilities to provide long-term access to the materials. This work, especially the donor interview and survey documentation, will also help you when you later arrange and describe the materials. Finally, the documentation provides you with legal protection because you have generated evidence of the donor's knowledge and consent to all the agreed-upon preservation actions.

To be clear, this engagement is not limited to donors from outside your institution or institution members who are providing personal materials to your archives. It also applies to content creators within your institution who are transferring records to your archive according to a records retention schedule or other internal policies and procedures. In the case of internal records creators, the engagement should happen as early as possible so that the records are being created using consistent file formats and standardized file naming and file organization practices whenever possible. These creators also need to be aware of what can be hidden in files. Despite the best intentions and efforts of organizations and employees, often a creator's personal material can be accidentally included in organizational materials transferred to the archive. Constant communication before and during transfer can help reduce the chances of this happening.

Depending on your institution's staffing and organizational makeup, it may be that you are not working with the donor directly. In that case, you will need to work closely with the curator, archivist, or representative who interfaces with the donor so that as much of this process can be done as possible. How you approach this will depend upon your organization's culture and priorities. You may need to lean on the legal risks associated with not following these procedures. You could emphasize the contextual documentation generated through these procedures that would make the eventual processing of the collection more efficient. You know what avenues of advocacy work best for your organization. Leverage that knowledge.

Acquisition Workflow

The acquisition workflow (see figure 3.1) determines if materials should be acquired, documents interactions with donors before materials are transferred,

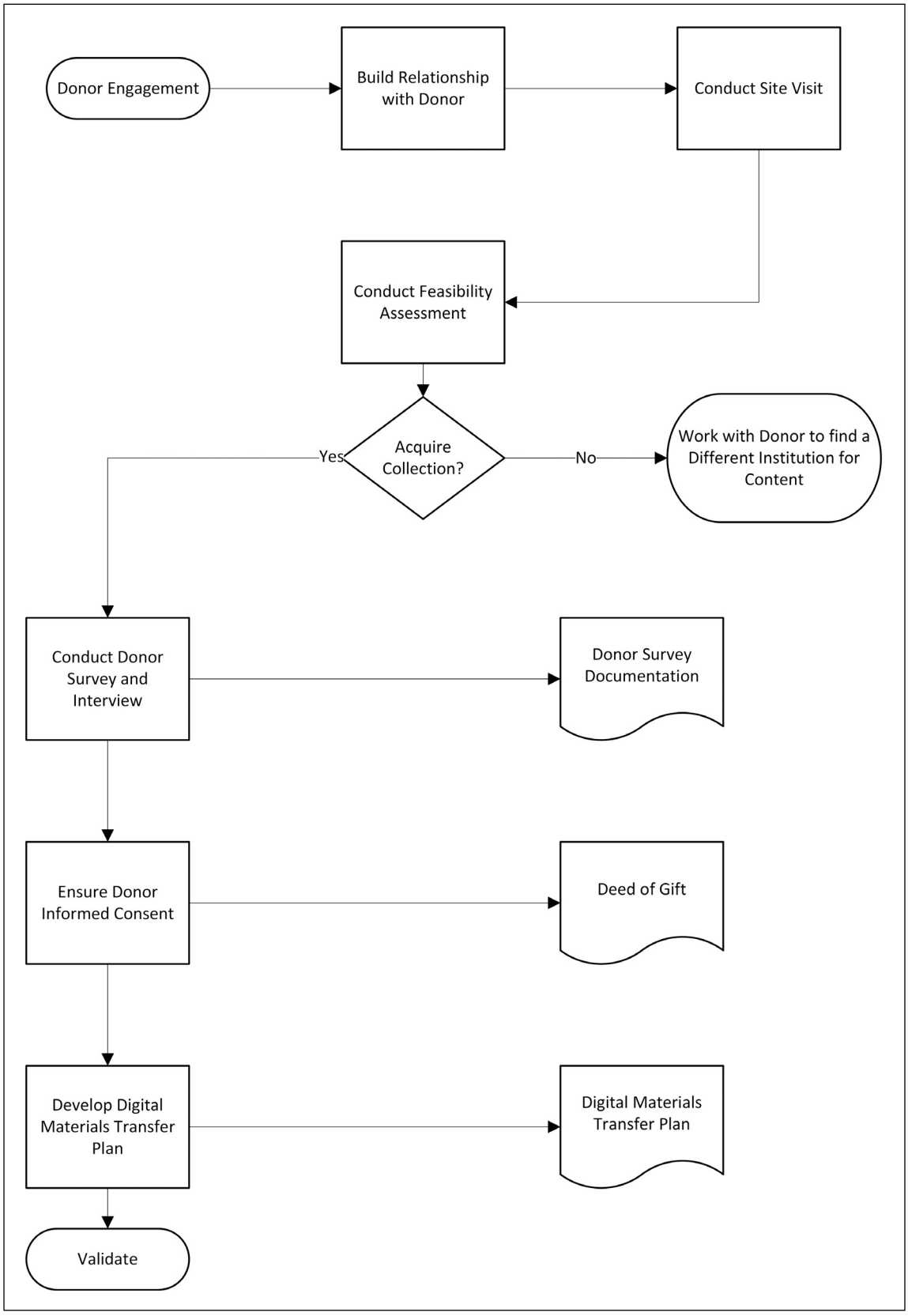

Figure 3.1
High-level workflow diagram of the acquisition process

and describes the actual transfer of materials from donor to your institution.

Feasibility Assessment

When your donor is ready to start discussing a transfer or donation of materials to your institution, before moving forward with any other steps, you need to determine if it is feasible for you to accept the materials. The first half of this feasibility assessment is determining that the digital materials are within the collecting scope of your institution. The second half of the feasibility assessment focuses on answering this question: "Do we have the time, expertise, and resources to responsibly curate the materials being offered to us?" If you currently have all the necessary resources, then you move along in the workflow. If you do not, you must then determine if it possible to advocate for and realistically receive the additional resources needed to preserve and make available the offered digital materials in the long term. If not, I encourage you to work with the donor to find another institution that has the interest and ability to preserve and provide access to the materials.

The feasibility assessment is solely focused on your institution's *current* infrastructures and resources in relation to the digital material being offered by the donor. Remember, this assessment is being done without doing a deep dive into the materials, just with the general information provided by the donor or discovered in a brief site visit.

ASSESSMENT QUESTIONS

- Are there file format types within the donation outside your current abilities to preserve?
- Is the size or volume of the collection beyond your current storage and preservation management capacities?
- What is the likelihood that content is corrupted, unstable, unreliable, or incomplete?
- Is there content that requires specific software platforms to render and make accessible that you do not currently have access to? If you do not have those resources, is it feasible to acquire and maintain the software needed?
- Is preservation of original physical media carriers required or necessary?
- How feasible and practical are ongoing transfers of data, if needed?
- What are the potential needed migrations and transformations? What are the anticipated costs of those migrations and transformations?

It is your overall evaluation of all these questions that will determine if it is feasible for you to acquire the collection. Each question will be weighted differently depending on your own institution's resources and priorities. For those acquisitions you choose to move forward with, the answers to the questions above can be incorporated into your donor survey or donor interview documentation.

Donor Survey and Interview

The ideal scenario is to pair the donor survey with a follow-up interview so you can tease out the most information possible. A conversation, instead of simply reading and responding to the text on the page, can help jar memories loose and help donors better understand the questions in the survey. The survey could be shorter or longer than the example I provide. It all depends on the types of materials your institution collects, the priorities you have regarding documenting collections in context, and what each donor is offering.

CREATION

Determine copyright and intellectual property ownership and dates of creation:

- Does your collection contain e-mail, documents, or other materials produced by others?
- If so, who else created content included, and what are their roles?
- What are the earliest dates of file creation?
- What are the latest dates of file creation?

CONTENT

Determine content types, file format types, and duplication of content in paper and digital formats:

- What types of content do you create: correspondence, journals, research notes, preservations, reports, photographs, sound recordings, videos, research databases, others?
- What types of digital files did you create: word processing, spreadsheets, images, databases, websites, others?
- Is there content that exists in both paper and digital form (such as print outs of word processing files)? Can you identify this content?

ORGANIZATION

Determine naming scheme for files, organization and ordering of files, frequency of file destruction, and storage of files on multiple devices:

- How are digital files named?
- Do filenames indicate if the file is a draft or final version? How is this indicated?
- How are digital files organized?

- Are personal files stored separately from work files?
- Are digital files destroyed on a regular basis?
- Is more than one computer used to create and store digital files? Remember that a computer could be a tablet, phone, or an actual PC.

E-MAIL

Determine use and organization of e-mail (if included in the donation):

- Do you have multiple e-mail accounts?
- What e-mail programs and services do you use: Microsoft Outlook, Mac Mail, Hotmail, Gmail, Yahoo! Mail, others?
- How is e-mail organized?
- Do you create folders or labels to organize?
- How is e-mail saved: stays in the e-mail program, copy saved to computer, paper printout, others?

PRIVACY AND SECURITY

Determine files that may require restriction, need for passwords to view files, use of other encryption methods:

- Do some files contain sensitive or confidential information?
- Are there specific files that you would want temporarily restricted or permanently removed or destroyed?
- Do any files require passwords to open?
- Where are usernames and passwords located? Do you use an external service to manage usernames and passwords?
- Are any other encryption methods used to protect files?

STORAGE AND BACKUP

Determine existence of backup procedures, storage of files on different media, and incidents of lost or damaged files:

- Do you regularly back up your files?
- Does someone assist you with technical support?
- Are your files automatically backed up? By your institution?
- What media are used for backing up files: optical disk (CD/DVD/Blu-ray), hard drive, file server, web backup service, other?

TECHNICAL

To be documented by archivists after conducting physical review of technical environment:

- What are the hardware configurations for each computer or device?
  - manufacturer
  - model no.
  - CPU
  - RAM
  - hard drive capacity
  - video card
- What operating systems are used?
- What other system software is used?
- What are the main software applications used to create digital files?
- Is "user" for software applications set to name of creator/donor?
- Are computers connected to a network file server? Is file server space used by creator/donor?
- Are login username and password required to access computers?
- What is the total size of digital files to be donated?

Deed of Gift

Depending on your organization's acquisition and accessioning procedures, this piece of the process may come earlier or later than where I have it listed. However, I encourage you to have a separate section in your deed of gift or transfer agreement document to directly address digital materials. Here is some language to consider or use as an example for additions or modifications to your existing deed of gift or transfer agreement templates.[2]

SENSITIVE INFORMATION AND ACCESS RESTRICTIONS

Some or all of the collection may contain sensitive materials and require access restrictions. It is *Donor's* responsibility to outline all restrictions that must be placed on which specific materials, who is allowed to have access to the materials during the restriction period, and when the restriction period expires. All restrictions *must* have an expiration date, or the materials will not be accepted. Either below or in an attached document, please list the specific materials (device, folders, and/or individual files) to be restricted and the conditions regarding the restrictions.

CREDENTIALS AND PERMISSIONS

If the collection contains digital materials that are protected by passwords, logins, encryptions, or other restrictions, *Donor* grants *Institution* permission to use to use passwords, logins, or other access keys *Donor* will provide in order to access the collections. If *Donor* declines (does not remember/have the ability) to supply passwords, logins, or other access keys for *Institution* to access digital materials that are protected by passwords, logins, encryption, or other restrictions,

Donor authorizes *Institution* to decrypt passwords or encryption systems, if any, to gain access to data received as part of the collections. If *Donor* does not authorize *Institution* to decrypt passwords or encryption systems to gain access to data received as part of the collection, then *Donor* agrees that *Institution* will discard these materials.

DISK IMAGING

Disk imaging is one of many established practices used by archivists to preserve materials. A disk image is a sector-by-sector copy of data that replicates the structure and content of data. *Donor* acknowledges that forensic imaging procedures may uncover information that was once deleted or overwritten by *Donor* and that imaging procedures may be used by *Institution* to preserve the collection in accordance with standard archival practices. By donating the collection, *Donor* grants *Institution* permission to use imaging procedures in order to preserve the collection.

Disk imaging may recover deleted files, log files, system files, and other files that document use of computers or systems. Does *Institution* have your permission to perform disk imaging?

If disk imaging is performed, does *Institution* have your permission to provide access to deleted files if they are recovered?

If disk imaging is performed, does *Institution* have your permission to provide access to log files, system files, and other files that document use of computers or systems if they are recovered?

Develop a Digital Materials Transfer Plan

To develop a digital materials transfer plan, you will pull heavily from the donor survey or interview mechanism. The digital materials transfer plan is meant to document the decisions made by you and the donor regarding what materials will be transferred into the custody of your institution and how that transfer will take place. This transfer plan may be determined through a series of e-mails, a simple verbal conversation, or a surprise drop-off at your institution. Any agreements should be documented in some way and included in the master file for the donation, whether through printouts of e-mails or written summary of phone or in-person conversations. Alternatively, you could document the transfer using a formal mechanism similar to the digital materials transfer plan. However transfer occurs, all of the following (or as much as is possible to collect) should be documented.

DEFINE INFORMATION

- Content types to be transferred.
- Types of software, operating systems, and other technical infrastructure that were used in the creation and management of the digital materials.
- Indicate if, and in what form, descriptive information about the digital materials exists.
- Indicate formats to be transferred. (Do not try to create a detailed list; just provide enough information to give a general idea.)
- If there are physical media items to transfer, create an inventory that indicates each type of electronic media and how many of each type are to be transferred.
- If only data is being transferred, indicate the total data size to be transferred.
- Describe the specific file-naming conventions or rules used to identify materials. Indicate if filenames are based on specific best practices or standards.

DEFINE TRANSFER PROCESS

- Packaging
  - Determine who will be packaging—donor, archivist, or donor and archivist together.
  - Determine if this is a one-time transfer or if there will be additional future transfers, scheduled or otherwise.
  - Determine type of transfer: physical only, logical only, combination of physical and logical.
- Transfer method
  - Determine how packages will be transferred:
    - Physical transport of electronic media as-is. No data or files copied from original electronic media.
    - Data transferred from local machine or network to archivist's portable device and physically transported to archives.
    - Data transferred via network or internet.
    - Data transferred from local machine or network to donor's or creator's portable drive or electronic media and transported by donor or creator to archives.
  - If data is being transferred via network or internet, test the transfer method and schedule a time for the transfer, making sure to allow for potential problems during transfer that might necessitate starting the process over.
- Tools
  - If retrieving data from the donor's or creator's device or machine to a portable drive, determine the hardware you will use for the transfer and what software tools you will use to do the data transfer. Potential transfer software tools include
    - Exactly
    - Data Accessioner
    - TeraCopy[3]
  - If the data is being transferred via network or

internet, determine what software will be used to transfer the data and that the donor understands how to use the software and has practiced using it. Ideally the data will be packaged using a software tool such as Exactly and then sent via secure network transfer protocols or a secure online file-sharing service such as Box.[4]

DEFINE VALIDATION

- Verify
  - All components (digital objects and metadata and all electronic media) were transferred.
  - Components are well-formed and were not corrupted during transfer.
  - Components are free of viruses.
- Acceptance
  - All initial validation requirements are met. This requires that components have been inventoried and (when data is transferred along with physical media carriers) checksums generated before transfer, which may not be possible for all transfer scenarios.
  - Some validation requirements are met. Acceptance criteria for transfers where full packaging (inventory, checksums generated) has not occurred before transfer may need to be flexible in response to different transfer scenarios.
  - Repeat transfer. There are likely to be transfer scenarios where some amount of data has been corrupted or files infected by a virus. Determining whether to accept such a transfer and report errors or to not accept and attempt a retransfer should be decided on a case-by-case basis.
- Tools
  - Determine validation software. Validation software should have the ability to
    - verify package contents contain all components
    - verify package contents are well-formed
    - verify package contents are free of viruses
    - produce validation report
  - Many transfer tools also act as validation tools, such as Exactly, DataAccessioner, and TeraCopy.

Transfer Materials

Follow your digital materials transfer plan. Each transfer will be more or less complicated depending on the amount of material to be transferred and the type of transfer occurring. The simplest transfers are of physical media carriers alone: you inventory the items before transfer and verify that the same number and type of items arrive at the archive. The most complicated transfers are hybrid collections, where paper materials, physical media carriers, and data are all transferred as one accession. These transfers must be carefully documented, and post-transfer verifications must be thorough so as not to disassociate any data from the other parts of the transfer. It is crucial to move directly to the accessioning and stabilization workflow at this point. Digital materials are too fragile to sit on physical media carriers or in transfer data storage for long. There are too many ways they can be irrevocably altered or corrupted.

Notes

1. David Thomas, *The Silence of the Archive* (Chicago: Neal-Schuman, 2017).
2. The suggested deed of gift language was influenced by the Georgia Tech Library's deed of gift form.
3. "Exactly," AVP, https://www.weareavp.com/products/exactly/; "Data Accessioner," http://dataaccessioner.org/; "TeraCopy for Windows," Code Sector, https://www.codesector.com/teracopy.
4. Box home page, https://www.box.com/home.