

Access Workflow

As with all other aspects of archiving and librarianship, the answer to how users gain access to digital materials will always be “It depends.” It depends on institutional policy, on the resources the institution has to provide equitable online access to materials, on whether the institution has the ability to provide mediated access to materials that require special software to access, and many more variables that determine how and when a user can access digital materials.

The access workflow (see figure 6.1) shows two clearly defined pathways for providing access to digital materials: direct access and mediated access. In direct access, users are able to find and view digital materials without any need to contact the institution of origin. Mediated access *requires* the user to contact the institution to gain access to materials, either because the materials are not available online or because the user needs credentials to access the materials that are available online. Your institution’s policies will determine if direct access to born-digital materials is viable or if only digitized materials will be made directly available to users.

There is also the question of whether users will be able to gain access to unprocessed materials. There is a clear precedent for this with physical materials, especially if institution members wish to gain access to institution-produced content or a creator’s family members wish to access materials by or about themselves. Therefore, you must be prepared for users to ask about digital materials mentioned in finding aids that are not yet processed or to request access to unprocessed materials that users know to be at the archives but are not yet part of an existing collection. Realistically, due to privacy concerns and the exponentially larger volume of digital material compared to physical material, an archivist may need to review materials before providing a user access to unprocessed digital content. A risk-averse institution will simply restrict digital materials until they are processed. However,

for low-risk collections, where all the accession documentation clearly indicates there should be no privacy problems and the materials are fairly easy to review using file viewer and analysis tools, it will not always be an automatic no.

Unprocessed Materials

If you do not currently have a policy regarding user access to unprocessed materials, start there. The policy will support archivists who interact with users requesting access, will create transparency about who has access to materials and why, and will clearly delineate who is responsible for making the decisions on a case-by-case basis. Your policy may very well state that no access to unprocessed digital collections material will be allowed. As long as that policy is consistently applied to *all* users, there should be no problems with implementing it. If the policy allows access to some unprocessed materials, you must very carefully delineate the terms of that access and how those decisions will be made and then consistently apply that policy.

Access to unprocessed material, digital or physical, is always mediated by an archivist in some way. Typically, the user does not gain access to the entirety of the unprocessed materials (unless the accession is very small). Instead, the user clearly states their research question and what they are hoping to find within the unprocessed materials. An archivist then uses the guidelines to carefully extract a subset of the materials and provides them to the user. I would suggest, when providing these materials to users, it not be through e-mail. Instead, access should be through more secure means, such as requiring a user to come to the archives and access the materials via a reading room computer or through a cloud storage environment that requires some form of authentication to access.

After consulting your policy to determine if access to an unprocessed accession is warranted, you must then determine how much of the accession to provide the researcher. Is it a fairly small accession that would be easy to remove any restricted material from? Or is

it a large collection that would take more time than is feasible to sort through? For a large collection, I suggest having the researcher be as specific as possible about what content they are looking for, including specific keywords for you to use when examining the

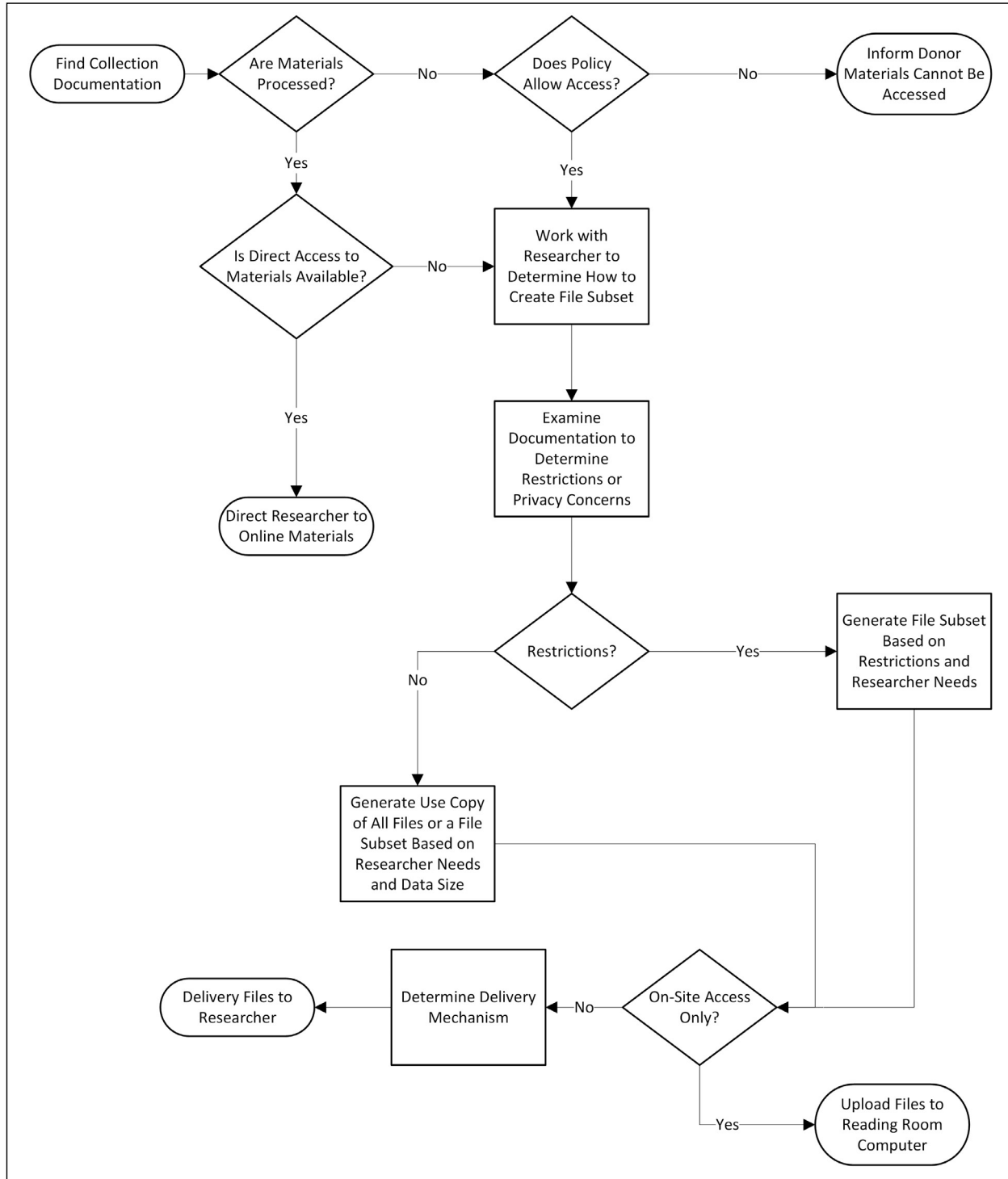


Figure 6.1
Diagram of a high-level access workflow

accession. Then, you can use either your computer's built-in file explorer or a tool like TreeSize to search for items in the accession that might be useful to your researcher.¹ Whether you are providing the entire accession, minus restricted material, or a small subset of files, you will need to create a staging location to copy files into for the researcher before moving on. By having a staging location, you will prevent potential accidental alterations to your working files copy.

Then you need to determine if there are materials in the accession that are restricted or at least too sensitive to grant researcher access to at this time. For this, I use a three-step process. First, I examine the accession master file to determine if there are any materials the donor has specifically requested be restricted and to see if there is the potential for personally identifiable information in the accession. With those guidelines in mind, I use a tool such as Bulk Extractor to search for personally identifiable information if this process has not already been completed in a previous workflow.² Finally, I remove any restricted material from the staging set of files.

Now that you have a file set ready to deliver to the researcher, you must determine how you will do so. Typically, there are two options for unprocessed material, reading room-only access or providing the files through a cloud storage service that requires the researcher to go through an authentication process before accessing the materials. The reading room option is the most restrictive and security-conscious. It requires the researcher to come to the institution and use the reading room computer to access the files. In some cases, these reading room computers have been specially modified so that the external ports that allow flash memory devices to be inserted have been deactivated, access to the internet and internal institutional networks has been removed, and other security measures have been put in place so that the researcher has access only to the materials that have been preloaded on the machine with no ability to download or remove those materials from the machine. In other cases, the machine has not been altered and the researcher is on their honor to follow the rules set by the archives regarding copying of the materials that have been preloaded on the machine.

The second option is to provide the materials to your researcher through a cloud storage service provider such as Google Drive, Box, Microsoft OneDrive, and so many more. This option does not require a researcher to come to the institution, which allows you to provide this service to a much more geographically diverse set of researchers, and if a researcher has accessibility needs, they can use their own computer setup where the tools they need are available. All of these services allow you to set limits on what researchers can do with the materials, so depending

on your policies, the researcher may have only view access, or they may be able to download the items. Also, these services allow you to automate when permissions are rescinded so that the researcher does not have perpetual access to the materials. Finally, the sharing mechanisms on these services almost always automatically include authentication steps. This is due to the ways access to shared material is linked to specific e-mail accounts or user accounts so that access given to a researcher would not be able to be shared beyond that person.

Processed Materials

For processed materials, the mediated pathways described above are also an option, albeit often with fewer restrictions on researcher behavior. However, there are also multiple unmediated ways for users to access materials. Examples include clicking on a link in a finding aid, searching through an institutional repository, landing on the public interface of a digital asset management system, or accessing items through your institution's online library catalog. The key difference between mediated and unmediated access is that, for unmediated access, while an archivist, librarian, or curator may have helped researchers find the material, the intervention of these professionals is not required for researchers to gain access to the material.

Often the decision about whether materials will be available through unmediated access occurs during processing because the mechanisms to provide that kind of access generally need to be built into the final description of the materials. When these decisions are made after processing has been completed, you need to modify any existing descriptions to point to the access location in some way, either through building collection-, folder-, or item-level links to the access location, by adding text to a notes field about how to access the materials, or by building out the description in the digital asset management system's public user interface. During or after processing, the workflow for building out these descriptions will be dictated by your institution's policies on description and existing workflows regarding the tools used to create those descriptions, be it encoding the Encoded Archival Description or using a tool like ArchivesSpace to create a resource record that will eventually become the finding aid.³

Notes

1. "TreeSize," JAM Software, <https://www.jam-software.com/treesize/>.
2. Simson Garfinkel, "bulk_extractor," GitHub, https://github.com/simsong/bulk_extractor.
3. ArchivesSpace home page, <https://archivesspace.org/>.