Visualizing Objects, Places, and Spaces: A Digital Project Handbook

Tools & Resources for Archival Projects

Beth Fischer, Hannah Jacobs, Claire Cahoon, Noah Huffman

Published on: Oct 14, 2019

DOI: 10.21428/51bee781.40b4d7e0

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

When choosing tools for an archival project, it is easy to jump to a system you are already familiar with or to assume you need a custom solution, but there are other factors that need to be considered. What does the final archive include? How are pieces of information within it related to each other? Who uses it? How do they access it? How frequently does it need to be updated? Will you have funding for upfront and/or ongoing costs?

All of these factors can affect your decisions. Doing some initial work to assess your project's scope and work on wireframes will help you choose tools or platforms that will accommodate your needs.

Tools for Acquiring Content

If you begin with content in a physical format, whether it is documents or three-dimensional objects, you need a way to turn it into something that can be stored and read by a computer.

Scanners

For documents and printed content, you will need access to a scanner. There are a lot of different kinds of scanners. In general, the cheapest type is a simple flatbed scanner, but these take a lot of time to use as you have to put each page in individually, and they can leave shadows if you have bound material to scan. They are also not ideal if you have things like photographs, as they're often designed to emphasize clarity and contrast over details like precise colors.

If you have a lot of unbound pages of the same size, consider looking for a sheetfed scanner, which allows you to put in a stack of paper and have it all scanned without needing someone to stand there positioning each page. Many institutional copy machines have this feature, so you may already have access to one in an office or academic department. Public libraries also sometimes have this type of scanner or copier, as do copy-service companies like Kinko's. Most of these scanners scan everything in the stack into a single file, typically a .pdf, so you may have to do additional processing to divide the file into separate files or to change the file types

into something you need. The other types of scanners all require you to place each document individually, so they take a lot longer.

If you have photos, there are scanners that are designed specifically for handling photos and color images that have higher resolution options. These are also good for digitizing photo negatives or slides. Academic libraries often have these, as do copy centers.

If you have delicate materials that might be damaged by being run through a sheetfed scanner or from pressing against the glass of a flatbed scanner, you can look for an overhead scanner. This type of scanner has a lens on a mount that holds it over the document instead of touching it directly. If you have a book to scan, you may need special weights or attachments to keep the book open and still while you scan. Like photo scanners and simple flatbed scanners, you do have to scan each document or page individually, and overhead scanners often have a more complex interface with fewer automation options than the other scanner types.

If you have text or images in bound volumes, you should look for a scanner that allows the page to be scanned without harming the spine of the book or leaving a shadow. Some types of overhead scanners have special cameras or cradles to help with this, or have software that can eliminate the curve of the page and shadows -- a particularly common type in academic libraries is called a Kic scanner. There are also flatbed scanners that come with a special cradle to position your book properly.

There are a lot of smartphone or tablet applications that can scan documents directly; these are of limited value in most cases because they are slow, have significant limitations on quality and format, and the potential for user error is high. You will likely have to redo a lot of scans and will end up with content of limited quality. However, this can be useful if you primarily need the scans for reference, perhaps to use for checking text that has been manually entered, or if the content cannot be moved to a location with a scanner. If you go with a smartphone scanner, it is a good idea to get a smartphone-compatible tripod and a remote shutter release so that your scan will be as stable as possible.

Scanning can take a long time, especially if you have content of different sizes or need high-resolution scans. Using a sheetfed scanner when appropriate will speed up the process. Many scanner software types also allow you to set and save a set of preferences for things like scan file type, resolution, and OCR (text recognition) options, which will save you a lot of time and require less attention.

Cameras and Tools for 3D Objects

The most common way to “digitize” a three-dimensional object is to take one or more photos of it using a camera, tripod, and sometimes a turntable. This is often a more specialized practice than scanning printed documents and images, and it is also harder to find appropriate equipment to borrow. Given the learning curve and equipment expense, it is often a good use of time and resources to hire a professional to handle this. However, if the documentation needs are relatively limited and the number of things to be processed is small, it is feasible to do without extensive experience. In addition to the camera itself, you will want to consider tools to help with lighting the object and stabilizing the camera.

Typically, you will want access to a dSLR camera. These are consumer-friendly cameras with settings that can be adjusted in significant ways to make the most of your images. If you aren’t familiar with the settings on such a camera, there are a lot of tutorials online for topics like focal length, exposure, and white balance. Look up a tutorial or video on your individual camera model as well to get a better sense of its particular idiosyncrasies. You also might need a special lens, especially if you need to take detail shots close to an object.

If you are limited to something like a smartphone camera for your images, look for an app that will let you manually override some of the camera’s automatic settings. Smartphone cameras have been getting very good and may be suitable for some projects, especially if you take the time to get the lighting right and stabilize the camera.

The biggest problem most people have using smartphone cameras for archival images is that it is very difficult to keep your hand stable as you focus and take the image. A smartphone-compatible tripod and a remote shutter release (both very affordable) help a lot with this and are fully worth the cost.

Tripods and remote shutter releases are also useful when using a standalone camera. In addition to stabilizing the camera so that it is easier to get a focused photo, they can also be a helpful part of your workflow, so that you can move the objects you are photographing with minimal trips back and forth between the camera position and the object. This is also helpful for making sure your workflow is consistent, that you are taking all your images from the same height and distance.

If you are taking photos of objects from multiple angles, also consider getting a turntable. This allows you to move your object quickly and consistently, and minimizes how much you need to directly touch the object.

You may also need tools to help with lighting. If you are photographing buildings or outdoor locations, there is probably a limited amount you can do without professional-level equipment and skills. Indoors, however, a light tent is an easy way to create a consistent background for objects and to make sure the light around an object is even. Depending on how much natural light you have and how consistent the lighting in your space is, you might also want to get 2 clip-on lamps to help create even light around the object and minimize dark shadows.

Automated Tools for Formatting or Cleaning Content

Once you acquire your data in a digital format, you will likely need to make additional changes to its file type or format. There are two goals with this: putting your data into a format that is consistent across your content, and putting it into a file type that is best for your project.

In some cases, you will be able to do this at the same time as you are acquiring content; other times, you will need to clean up and organize content that is in a variety of forms. There are automated programs and workflows that can help you do this quickly.

In many cases, you can use a program you already have. Database or spreadsheet programs like Excel and image processing programs like Photoshop have advanced functions that can save you a lot of time by performing the same set of simple but time-consuming tasks on groups of content. Some common examples:

- Photoshop can take a folder of images, open each one, save it to certain dimensions, change the format, and save it with a particular file name.
- Excel or google spreadsheets can take a bunch of text that is all in one block and divide it into separate rows and columns following a simple formula.
- Your operating system on your computer can rename a bunch of files to the same formula and save them in a particular place

The options tend to be specialized and project specific, but a quick search on what you want to do and the name of the program or tool you hope to use can get you step-by-step instructions. Using the search term “batch” will help you find ways to perform the same operation on a large group of files, while “macro” will help you find instructions for how to create a set of instructions that will run sequentially. For moving open text to a structured format, start with “text to columns” instructions, and look up “regular expressions” or “regex” for more complicated patterns in your data. If you have data that is particularly complicated or where multiple types of content are jumbled together, check out OpenRefine, an open source application for filtering, cleaning, and organizing content.

Archival projects often present a special problem in dealing with scanned text that the computer treats as an image. This is especially true when you are working with historical documents. There are a number of tools that you can try to automatically convert the image to text so that you don’t need to type the content in by hand. For this process, you are looking for “Optical character recognition” software, usually abbreviated OCR. Basic OCR tools are built into a lot of tools and programs that you may be using for the rest of your project, like many scanners, Adobe Acrobat, and even Google Docs, but documents with irregular content, page damage, or other complexities may require a specialized tool. Some common choices include ABBYY FineReader and Tesseract. If you have audio content, you might also want to consider a program for automated audio transcription.

Tools to Store and Share your Content

Once your data is in the form you need, you need to store it and you need to have a way for people to access it in the right way. Depending on the project, you need tools to store, back up, and host your data, and you may also need programs to design and display an interface or site for your material.

If the project is relatively small and has a narrow audience, you might not need to store your content in a place that is accessible by a website. For example, you might use the data for an archive that you only access onsite, or the goal may be to create a repository that stores data as a backup in case data in another format is lost. In such cases, you may start by purchasing standalone storage drives, which come in a wide range of sizes, or you may want to buy a subscription to an online file storage service, which is typically cheaper than normal server space but limits features like download speed or the frequency or number of user requests.

If the final audience is very limited, you might simply have a folder structure with files that are named in a standard way; or you might use a spreadsheet application like Excel that records your content and structures it into tables. If your content is stored locally in a hard drive, it's a good idea to supplement this with an online backup service like BackBlaze or iDrive. These programs are cheaper than regular cloud storage, but they limit how frequently and how quickly you can access your information, so you don't want to use those as your primary storage solution, and you wouldn't want to use them to store material that needs to be publicly accessible on a website.

You might create an online exhibition or website in a tool that stores your content as well, like Omeka or Wordpress. This is a good solution for archives with a few hundred records, or that have a limited purpose like supporting class projects. With larger projects, though, you will likely run out of the storage that comes free with the basic level of website management tools or on a hard drive, and then you need to start thinking about a more comprehensive file storage plan. You might use a server-based file storage solution through your institution, or through a separate service like Dropbox, Box, or OneDrive.

As you have larger or more complicated interfaces for sharing your data, you also may need to move to a separate web host. Many website management platforms, like Wordpress, include both basic file storage and hosting, while others, like Omeka and custom-built websites, require separate support from a service like Reclaim Hosting. Many educational institutions provide these services to their students, faculty, and some staff, and they can often also help you find the right combination of services to complement what they have available. It is also common to migrate your data from one service to another when your project grows larger or more complex, so you aren't stuck in the system you start using.

Continue reading: [Teams & Expertise for an Archival Project](#)