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

Planning Archival Projects

Beth Fischer, Hannah Jacobs, Claire Cahoon, Noah Huffman

Published on: Oct 14, 2019

DOI: 10.21428/51bee781.8ceef098

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

An archival project typically involves these elements, though different types of digital collection projects will execute them in different orders, and it is likely that you will need to revisit some stages more than once.

- Evaluating the content you want to include
- Deciding on a structure for the archive and finding or building the system to store it
- Testing a prototype version of your archival structure
- Gathering content in a digital format
- Editing content for the archive structure and entering it into the system
- Error-checking and filling in gaps in content
- Creating a process for accessing the archive
- Arranging a system for maintaining the archive

To determine the scope of your project, consider the following questions.

These will give you the information you need to develop a budget, choose tools or programs, and create a workflow for building your archive.

What content will be included? How much content is there and how much space will it take up? Will that content grow over time?

- **Archival projects are often organized around *records*¹**, a set of primary objects, documents, people, or ideas stored with *related information*² about them. For example, you might create an archive of documents, and want to include not only images of the documents themselves, but also dates, transcriptions, names, or geographical information relating to each document. Both the number of records you have and the amount of information you want to record relating to each record will

affect the tools you choose, the structure of your archive, and the cost of storing your content.

- **Make a list that includes the approximate number of records** (Hundreds? Thousands?) **and what you expect to record about each** (for example, three date fields, two latitude/longitude location pairs, 1-3 photos, a catalogue number, and a description field of 200 characters). In many cases, you will want to include several different types of records in a database, each with different kinds of information attached. An example might be a database of concert performance information that includes records of concerts, but also needs to include records about composers, performers, or individual pieces of music. Make sure your list includes these different types of records and estimates for how much content will be stored for each.
- **Then, try to make an estimate of how much digital space your content will take up initially and how much it might grow.** Many off-the-shelf archive programs have size limits, and digital storage costs money. There are some useful guidelines for data that you can search online, but it is better to use some of your actual data to answer this question. A rough estimate at the level of how many gigabytes or terabytes of data you need to store is usually sufficient.
 - Storing **text (characters or numbers)** takes very little space -- just one byte per character. The text of an entire 500 page book could be stored in 1-2 megabytes (MB), while a single image might easily take up 4-5 MB, depending on its size and resolution. If your content is all textual, estimating the archive size will be less important, as almost all of the memory you use will be for the program/tool and not the data in it.
 - If you will be including **images, video, or other media files**, look at the size of a few representative files to get a sense of how big they are. Remember that scanned pages of books are usually stored as images, even though they are texts. Once you have a good sense of your averages for each of these, you can multiply them times the number of records you will have for a good estimate. A useful tip: google understands units like kb and mb, so if you have 1500MB of images and 2 TB of video, you can just type “1500 MB + 2 TB” into the Google search bar and it will put the total in the correct units for you.

- While it is possible to compress your data some, this always results in some loss of quality or slower access times, so you want to avoid this for most archival projects. You may start out with a small number of records initially, but **if you expect to add content over time it is worth trying to estimate how much you will be adding and how quickly**. Digital storage space has gotten cheaper over time, but you may find yourself with a huge increase in storage costs if your project's growth takes you over the limit for a tool's free version.
- After you've thought about all this, you should have a draft that includes a list of the data that you will include and its formats; an estimate of how many records or entries you will have (at least initially); and an estimate of how big your content will be initially and how much you expect it to grow.
- The answers to this question will be important for building a budget, choosing tools or programs, developing a team, structuring your workflow, and estimating a timeline, so **it is crucial that you spend some time being very explicit about your content**.

Is the archive my end goal, or do I expect to create a different project using the digitized material in my archive?

- **In some cases, the archive is the first step toward** creating a database-driven website, setting up an online exhibition, uploading content to a mapping program, or otherwise **using structured content to create a new interpretive output**. If you have a definite end use in mind, you will want to use the parameters of that later project to structure your archive. For example, in a mapping project, you might need to encode locations as latitude/longitude pairs rather than simply as city names, while for a virtual exhibit you might need to link to high resolution tiffs or video files and you would want to use a structure that supported those file types.
- **If the archive itself is the end goal, think about how it will be used**. Decide on your priorities; if you try to make everything equally important you will have no way to choose tools or structure your content. What are some typical tasks you want to perform using the archive? What information in it is most important to its final audience?

- At the beginning, your goal may feel like “collect all the information available about topic x,” but this is not specific enough to help you plan. In the last question, you assessed what information you wanted to include. Now you want to **think about why that data is important.**

For example, imagine you are making a database of historic photos:

1. Are the locations in the photos significant? If so, you might want to record their latitude and longitude in addition to their names, or you might want to include keywords or tags about the types of places to make them easier to search.
 2. Are the people in the photos important? If so, you might want to tag them, but you also might want to connect them to other biographical information.
 3. Are the photos part of a larger event or experience? If so, you might want to include more specific dates or times, or put them into a particular order.
 4. Do I need a certain number of entries before the database becomes usable? Consider limiting your initial archival project to a single collection, or to materials you already own, or to a certain place -- whatever makes sense given your topic.
- If you try to do everything that is possible or interesting, your archive will take a very long time to create! While you want to set up an archival project to have some flexibility, **setting your priorities helps you keep costs reasonable and make your archive usable in a shorter time.** You can also use this priority assessment to think about working on your project in phases and expand its purpose or function over time.
 - **Once you’ve completed this step, you should have a good sense of how you or others will use the archive:** what information is most important and what kinds of tasks people may perform with it, or what parts of it need to interact with another program and how. This will help you choose programs or tools for both storing and interacting with the content, and help you identify what formats you might use for different types of content.

Who needs to access or interact with the content in my archive? What tasks do they need to perform with the archive?

In some cases, one person will build the archive and be its primary user. However, it is common to have a larger team involved on adding material to the archive or using the final archive.

1. Who needs to *create* new records or add completely new content to the archive?
2. Who needs to *edit* records or fill in missing metadata?
3. Who needs to *use* the archive? This is the final audience, although if your archive is intended as a first step toward a larger project this final “user” may be a different program or tool.

For each category above, think about the needs and abilities of the people involved.

- *How many people are in each category?* If you have multiple people adding or editing records, you will need some method of version control.
- *Do they need to access the archive remotely or will they be in the same place (even using the same computer) as the main owner of the archive?*
- *Do they need to restructure content in any way?* This might include changing file types for people in the second category, or filtering/sorting for the final audience. It also might include features like downloading content in individual records or exporting content in a structured form for use in another program.
- *Do the users typically have existing experience with any archives or tools used for archives?* Will you have the ability to provide training or instruction? You don’t want to make an archive that other people on your team or in your final audience are unable to use!

Do I need to get permission to use this source material?

- **You need to determine whether any of your material is protected by copyright, and if so, who owns it.** In many cases, you will be able to make an

archive of copyrighted materials for personal research or for classroom use, but you are much more limited if you will be making the archive available to other people.

- **Check out the [guide to copyright and permissions for digital projects](#)** for an overview of the issues that might come up in the creation of digital archives.
- It is always a good idea to **talk to an expert about this issue**. Try a librarian at your home institution, as they can usually connect you with the people at your institution who specialize in this area.
- **Another good resource is the US Copyright Office's [series of circulars on copyright topics](#)**, which are written for laypeople. Although these are useful, they are very detailed and can be a little overwhelming when you're just getting started thinking about copyright. These are most helpful if you are thinking about how you might claim copyright on your archive.
- **If you need to get permission or to pay for licensing, you want to know this at the start of your project!** This can affect your budget, of course, but it can also take a lot of time, change your workflow, or change how you make your archive accessible to others.

Continue reading: [Prototyping & Wireframing an Archival Project](#)

Footnotes

1. **Record:** a set of primary objects, people, or ideas stored with related information about them. [↪](#)
2. This information is often termed “metadata.” [↪](#)