

Data Management Best Practices: File Organization

File Naming Best Practices

There are two general rules for file organization: Be consistent and Be descriptive. You want to make sure you and your colleagues can find anything you and they are looking for quickly. You'll need to figure out which specifics make the most sense to you and document your convention in a place everyone in your research group can follow. Here are some guidelines to include in your convention:

- Choose 2-3 descriptors to identify the project or collection the item belongs to and what the specific item is. Have a standard for your research group so things can easily be found and shared.
- Use capitals (camel case) or underscores instead of periods or spaces.
Example: surveyResponseData.csv or survey_response_data.csv
- Use no more than 30 characters whenever possible
- Use date format ISO 8601: YYYY-MM-DD
The year first format makes it easy to find newest/oldest files. Wikipedia has a nice page about about additional ISO 8601 date and time standards.
- Avoid special characters (See this page for more info) in a file name. Common things to avoid are using spaces or ampersands (&).
So **don't** name a file WBS+--+Final.docx or Research Data(finsished & final)!!!.csv
- Document your naming convention so you remember what it is and your project collaborators know what it is.

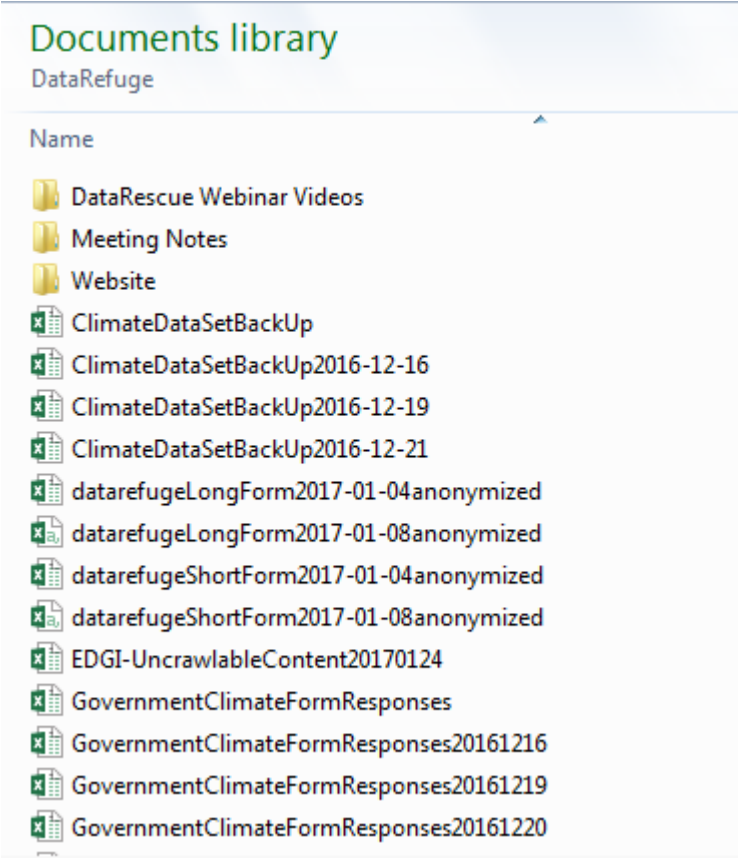
Version Control

How and when to save a new version is largely up to you and your research partners. Even if you're tracking changes with the software you're using, you should **always** keep a copy of the original unedited data available and save a new version when substantial changes are made

Most file naming conventions show new version by including a version number at the end of the file name, as in the example below, although there are other methods. Best practice is to be consistent.

- **Example:** dataMgmtNotesv5.txt [instead of, say, dataMgmtNotesFinalAgainReally2.txt]

Example



This is one example of file naming and organization structure. There are many ways to organize your files - think about what makes the most sense for your work. Notice in this example, though, some good practices and a few

Tools for File Renaming

- Bulk Rename Utility
- PSRenamer
- WildRename

Tools for Version Control

- Bazaar
- Git and GitHub

Guidance from Other Universities

- Oregon State University
- CalTech
- Temple University

Email Us

Also search Research DataQ for answers

things that could be improved.

Good:

- Most of the supplemental files are in their own descriptively named folders.
- New versions of files are dated using ISO 8601 format.
- No special characters or spaces are used in the file names.
- The file names are generally descriptive enough for the user to know what each contains.

Room for Improvement:

- The ISO 8601 dates should consistently use dashes or not use dashes.
- One file (EDGI-UncrawlableContent20170124) seems out of place. This might be fine, but most of the files in this folder have multiple versions so it's possible this one is just misnamed.
- A README file would help users recognize which files belong in this folder and provide information about the content of the files. See the Documentation tab for more information.

[Penn Libraries](#) [Hours](#) [Library locations](#) [Spaces](#) [Workshops](#)

QUESTIONS OR COMMENTS



Report accessibility issues and get help

3420 Walnut St., Philadelphia, PA 19104-6206

215-898-7555 | [CAMPUS LIBRARIES MAP](#)

©University of Pennsylvania