File Organization

updated: 2021-01-22

File Organization - Electronic

There aren't really "rules" about how to organize your files, but I think the best way to organize is to think in terms of "projects". Of course, how you define a "project" is sometimes tough…is the PhD the 'Project' or is each chapter of the dissertation a 'Project'? You'll have to give this some thought and think of a system that works for you. That said…

I like 'Project' based organization for two main reasons. **First**, it pushes you to think as 'projects' leading to of 'products' (e.g., paper, report, grant proposal), and therefore how and where data you are collecting fit in. If trhe data aren't for a new project, and it';'s not clear to you if the data fit into an existing project, then are you sure you need to spend time, effort, and money collecting it? **Second**, project-based organization focuses on the individual steps and components that take a project from start-to-finish: planning your data collection, data gathering and clean-up, analysis, presentation, and data archiving.

Example: Project Based Organization

A project-based organization scheme for a master's thesis might look something like this:

1. Top Folder: emb_masters_thesis
2. Sub-folder 1: `Chapter1_lit_review`
3. Nested Folders and Subfolders for:
    - Plans for Data Collection & Management
        - DMP
        - plans for storage and backup
    - data
        - raw data
        - clean data
        - metadata
    - analyses
        - intermediate outputs from analyses
        - final outputs from analyses
    - presentation (e.g., `masters_ch1_text` or `ch1_slides`)
    - data archiving

You can have one of these for each thesis chapter. Note that some (DMP, plans for storage and backup) might be better placed one folder up in the `emb_master_thesis` folder

But what if I will be using the same data for multiple chapters or papers? This is a great questions, and there are several options. One would be to include both chapters in the same folder, each with its own subfolder (manuscript 1, manuscript 2). The other would be to have each as a separate project, and pull the data in from the online archive where you deposited the data (or folder where the data are stored). As you can probably guess, I think #2 is better because I think of each paper as a different project.

File Organization - Physical

1. Whenever possible I try to organize in the same way. File folders, boxes, cabinets, whatever. Try to do things so that you go to one place for everything related to a project. Of course, you can't always do that because specimens go to museums or stay in an archive…but you can keep the information on how to *find* the originals there.
2. You can keep notes - electronic or paper - in the relevant folder. But one thing that has gotten complicated is email correspondence regarding projects, manuscripts, etc. Printing seems like a huge waste of paper, and yet if critical it might be worth doing. Other wise keep an email folder for each project as well - with the same name as the electronic version so you can easily find email correspondence related to each project.
3. Don't forget to include copies of the survey, data sheets, etc., as well as maps or other such resources.

Documentation

1. Prepare a file documenting your data organization structure so you can share it with team members (or look at it years later when looking for something). These can be `'README.txt'` files as needed for each folder. The `'README.txt'` file might include such things as explanations of naming conventions and how the structure of the directory relates to the structure of the project.
2. Use appropriate tools, such as version control tools, to keep track of the history of the data files (more on this later in the semester).