Me: “I’m very organized”
Also me:

source

Organizing Files & Folders

Nick Rochlin, Data Science Librarian

Session Overview

Learning Objectives

- Introduce concepts and best practices of file naming
- Implement file naming best practices on example files
- Introduce concepts and best practices of directory organization
- Implement directory structure best practices on an example project

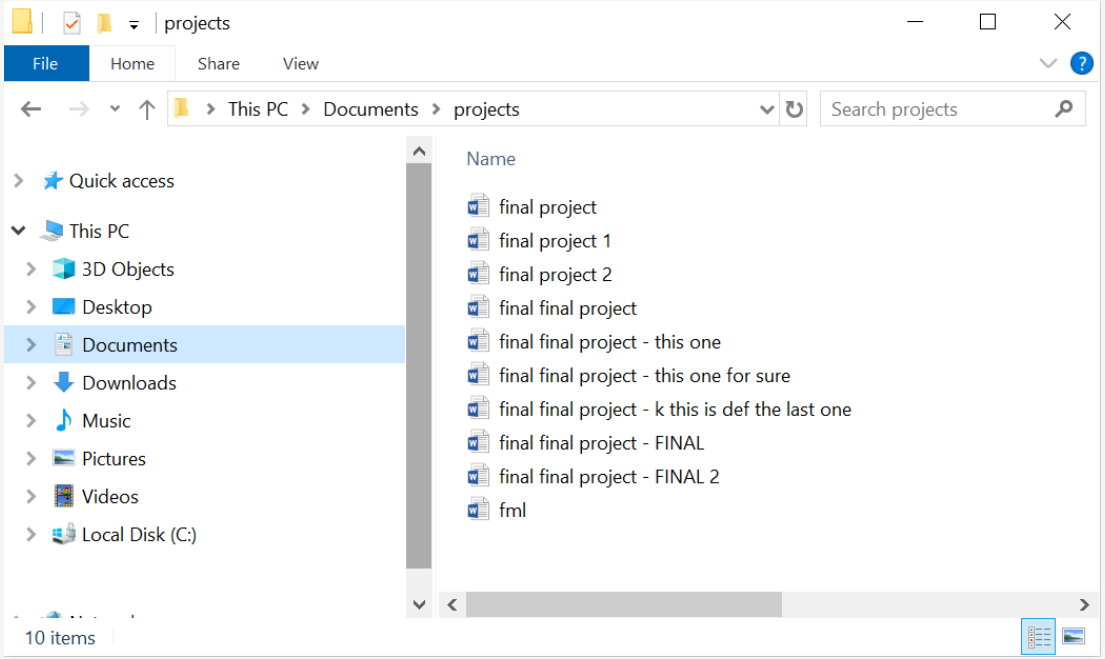
Session Overview

Learning Outcomes

By the end of this session, you'll be able to...

- Describe best practices for file naming and directory organization
- Begin conceptualizing how to apply these practices to your own files and directories

File Naming Best Practices



File Naming Best Practices

- Human readable
- Machine readable
- Be consisent!

Human Readable

- Can you look at a file name and know what it is? What about in a year from now?
- Will others be able to look at your files and know what they are?
- Will you/others be able to easily find a file you're/they're looking for?

Human Readable

- Short but complete names
- Ideally 3-5 conceptual elements
- Write down your naming conventions and document in a README file
- Define acronyms, abbreviations, codes, etc.

Human Readable

- Short but complete names
- Ideally 3-5 conceptual elements
- Write down your naming conventions and document in a README file
- Define acronyms, abbreviations, codes, etc.

Elements to consider in naming files:

- Date of creation/collection
- Short description
- Group/affiliation
- Activity
- Location
- Editor/creator
- Other relevant information

Machine Readable

- How will a computer sort your file names?
- If files move from one computer / application / operating system to another, will they remain interpretable in the same way?

Machine Readable

- How will a computer sort your file names?
- If files move from one computer / application / operating system to another, will they remain interpretable in the same way?

Machine Readable Qualities:

- Only contain letters in the English alphabet, numbers 0-9, dashes -, and underscores _
- Do not use spaces or special characters such as: !@#$%^&*()+={}[]|
- Separate naming elements with underscores and dashes
- Use date format:
  YYYYMMDD or **YYYY-MM-DD**

Examples: Data

- **Example:** lldr_mpp_20240723.csv
- **Documentation:**
  - *Convention*: project_location_collection-date.file-type
  - *lldr*: Leaf Litter Decomposition rate
  - *mpp*: Monk Provincial Park

Examples: Manuscript

- **Example:** lldr_manuscript_V01.docx
- **Documentation:**
  - *Convention*: project_content_version.file-type
  - *lldr*: Leaf Litter Decomposition rate

Examples: Feedback

- **Example:** lldr_manuscript_V01_NR.docx
- **Documentation:**
  - *Convention*: project_content_version_editor.file-type
  - lldr: Leaf Litter Decomposition rate

Likeness & Importance

source

Questions?

Managing Directories

source

Managing Directories

- Directories, AKA folders, are a way of keeping your files organized and easy to find
- Developing a directory structure before you begin a project can help with managing all the files that will be collected / generated
- The same principles for file naming apply to directories
- Directories are denoted by a “/” at the end in diagrams

Directory Structures

Directory structures typically have:

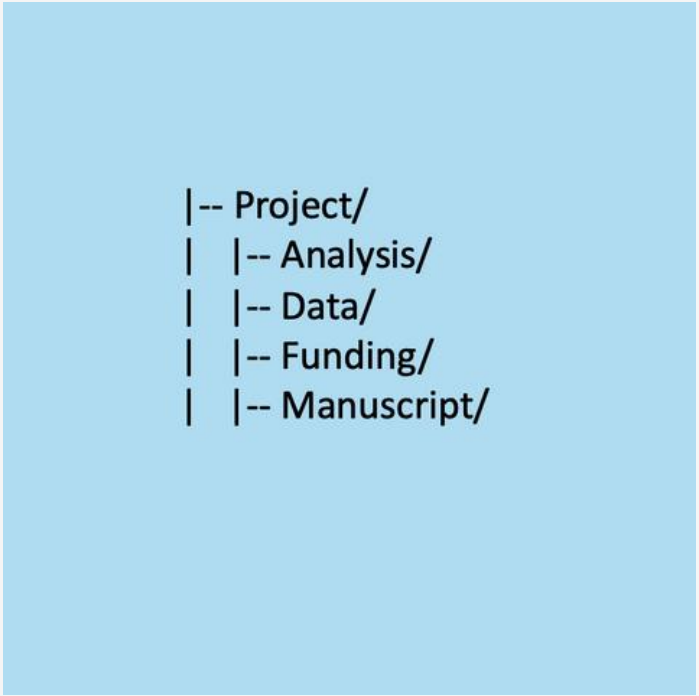
- A root directory (top-level folder)
- Subdirectories (subfolders)
- Relevant files

```
|-- Project/
|   |-- Analysis/
|   |   | -- scripts
|   |   | -- processed-data
|   |-- Data/
|   |   | -- data-collection-tools
|   |   | -- unprocessed-data
|   |-- Funding/
|   |   | -- application
|   |   | -- wages
|   |-- Manuscript/
|   |   | -- manuscript
|   |   | -- conferences
```

Hierarchy Depth

- A “shallow” directory structure has minimal nesting of folders
- A “deep” structure contains (potentially many) subdirectories
- Choosing which type of structure you want will depend on:
  - How many files the project has
  - The types of files the project has
  - The size/nature of your research team
  - Personal preference

A Shallow Structure



A Deep(er) Structure

```
|-- project/
|   |-- analysis/
|   |   |-- location-subsets/
|   |   |-- scripts/
|   |   |-- species-subsets/
|   |-- data/
|   |   |-- processed/
|   |   |-- unprocessed/
|   |-- funding/
|   |-- manuscripts/
```

Whiteboard a Plan

- Before jumping into a project, it's very useful to whiteboard a plan of directories

- It can be helpful to think about natural and distinct groups of data and files that you'll be working with

- To start whiteboard, you can think about:
  - Directory names
  - Directory contents
  - Access permissions
  - Other relevant aspects of the project

Final Thoughts

source

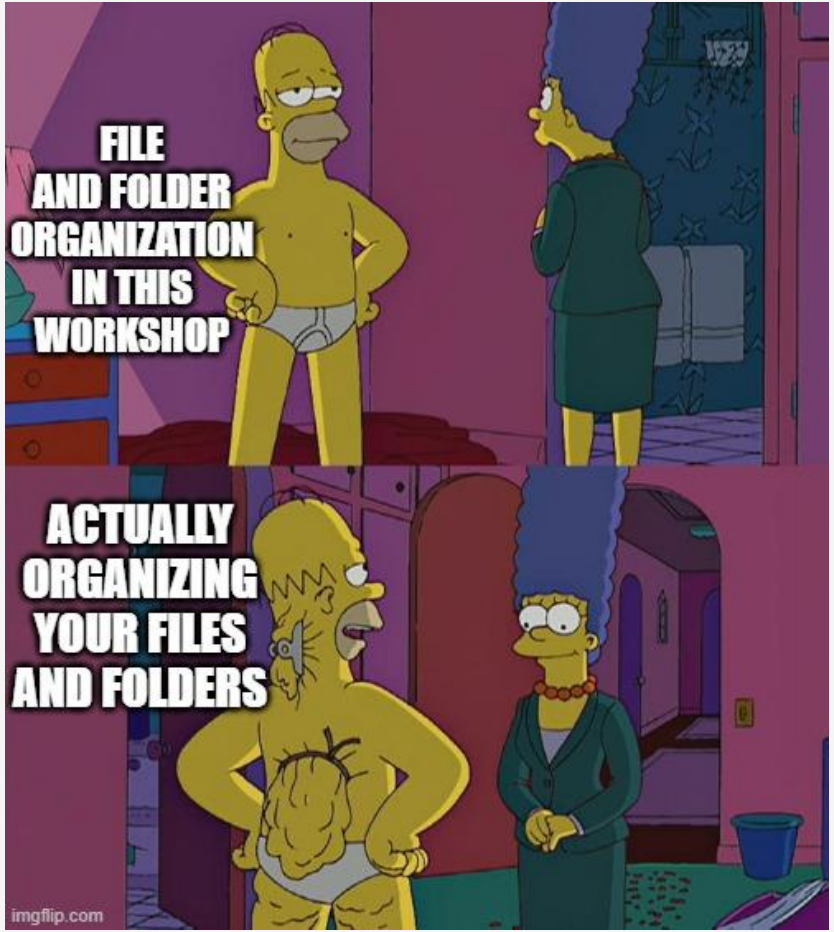
Versioning Data

- Data files can change and evolve over the course of a project
- Versioning is a way to keep “snapshots” of this progress
- Promotes transparency
- A good safeguard if things fall off the rails

Versioning Data – Best Practices

- Keep an original copy of your data files, and DON’T TOUCH!

- Consider various stages of data cleaning/analysis, and what versions might be valuable, and if these are best captured by naming or by directories:
  - Cleaning/cleaned
  - Subsets
  - Analysis/analyzed
  - Final


FILE
AND FOLDER
ORGANIZATION
IN THIS
WORKSHOP
ACTUALLY
ORGANIZING
YOUR FILES
AND FOLDERS
imgflip.com

source

Questions?

source