Me: "I'm very organized"
Also me:
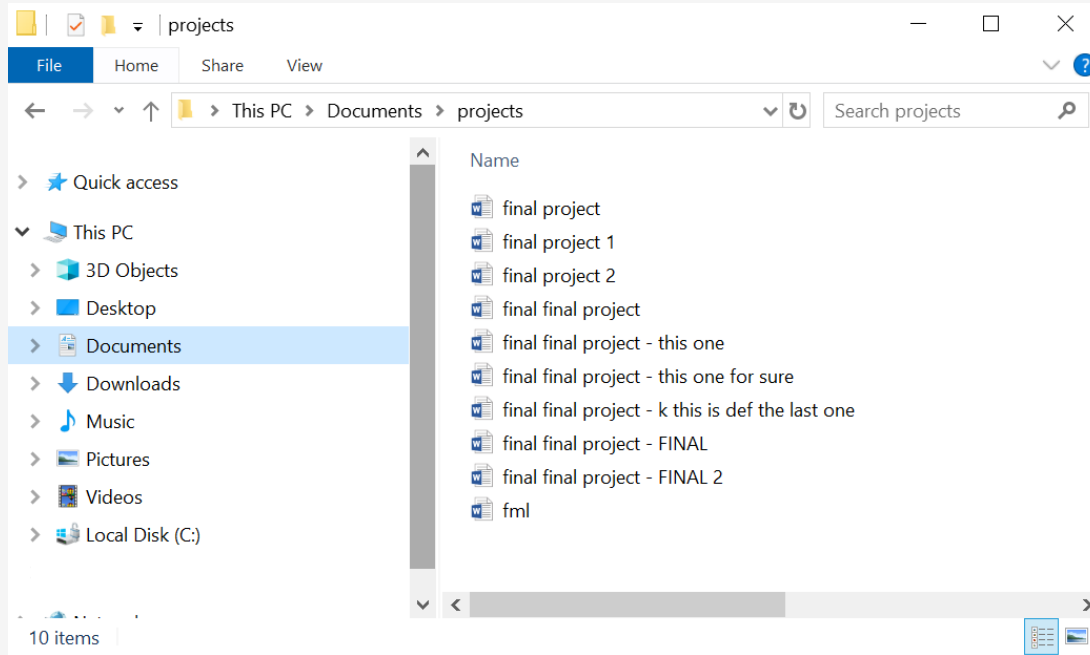

Organizing Files & Folders

Nick Rochlin, Data Science Librarian

source

Session Overview

- Introduce concepts and best practices in file naming
  - Human readability
  - Machine readability
- Best practices in directory (folder) structures
- Versioning files via file naming

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

Question from a student:

- If you have a folder dedicated to a project, isn't it unnecessary to have a project prefix in front of every file name within that folder?

Questions?

Managing Directories

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/
|   | _Data-dictionary.txt
|   | _README.txt
|   | -- Data/
|   | -- Funding/
|   | -- Manuscript/
|   | -- Scripts/
```

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

```
| -- Project/
|   | _Data-dictionary.txt
|   | _README.txt
|   | -- Data/
|   | -- Funding/
|   | -- Manuscript/
|   | -- Scripts/
```

A Deep(er) Structure

```
| -- Project/
|   | _Data-dictionary.txt
|   | _README.txt
|   | --Analysis
|   |   | -- Location-subsets/
|   |   | -- Scripts/
|   |   | -- Species-subsets/
|   | -- Data/
|   |   | -- Processed/
|   |   | -- Unprocessed/
|   | -- Funding/
|   | -- Manuscript/
```

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

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
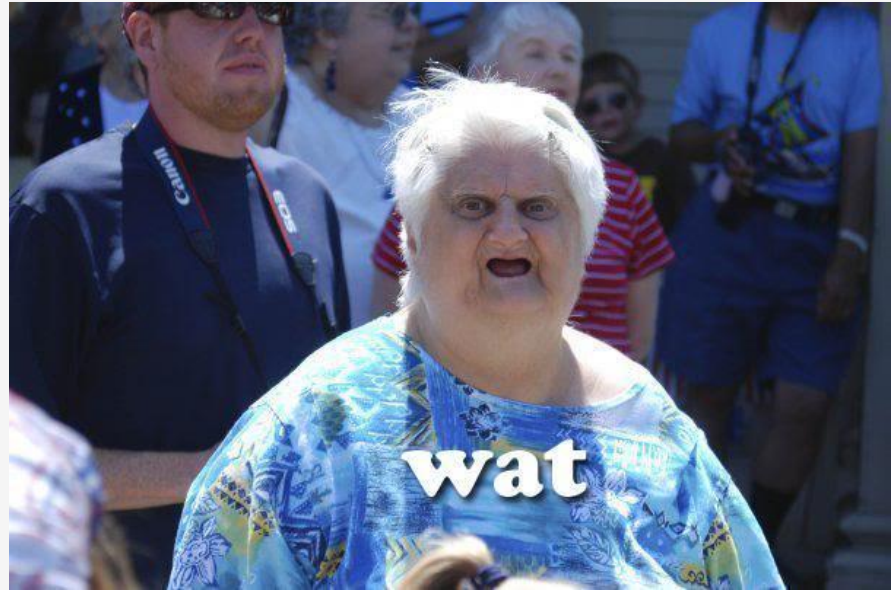- A good safeguard if things fall off the rails

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

Questions?