

Galter Health Sciences Library & Learning
Center
Northwestern University Feinberg School of
Medicine

10 Simple Rules for File Organization

Sara Gonzales, MLIS
Data Librarian





Hello!

Thank you for coming to this Galter DataLab class.

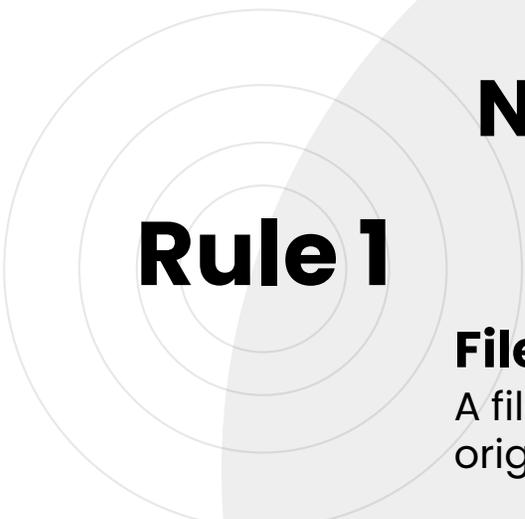
You can find more information on data management topics on Galter Library's website at:
galter.northwestern.edu/about/datalab.

You can find me at sara.gonzales2@northwestern.edu

The background features a large white circle centered on a black background. To the left of the white circle, there are two overlapping circles: a larger, semi-transparent grey one and a smaller, semi-transparent white one. To the right of the white circle, there are several concentric white circles of varying diameters.

Best Practices for File Naming

An essential first step to keeping data
files under control



Rule 1

Name files with a view to their future lives

File names should not be dependent on location

A file should still be identifiable even if it is removed from its original location

File names should be understandable by someone other than their original creator

New team members, or even you in five years, should be able to interpret a file's meaning from its name

Think about scalability

Do your naming conventions allow for files in a series increasing to the hundreds, or beyond?

Rule 2

Things to avoid in file names

- Special characters
 - Avoid: ~ ! @ # \$ % ^ & * () ` ; < > ? , [] { } ' " |
- Spaces in file names. Instead try:
 - Underscores _ or -hyphens
 - CamelCase
 - Run every element together with no spaces
- File names that are overly long
 - Names can be up to 255 characters, but something around 25 is preferable
- Short article words
 - "A", "and", "the"



Rule 3

Things to include in file names

20180817_StrsHlth_Survey1_KGH_verbal_23846_v01.pdf

Dates

Use the format YYYYMMDD to make files line up by date in folders

Location/spatial coordinates

Location in which data collection took place

Project Name

An abbreviation or acronym for the project title

Conditions of Data Collection

Note important details about things that had an effect on data collection

Researcher's Initials

If warranted, for instance if you need to record the name of the person who recorded data

Version of File

Using leading zeroes helps files line up correctly in folders even into hundreds of versions



Standard Operating Procedures

Rule 4

SOP Title: Naming files per the Systematic Review (SR) naming convention

Date: 2019-02-01

Purpose: To consistently name all SR files according to the naming convention

Responsibility: Head, Research and Information Services

Definitions:

1) SR Naming Convention: the format for naming files agreed upon by the SR workgroup. This format consists of a multi-part name separated by underscores. The first three parts of any folder or filename are always: Date (YYYYMMDD)_InvestigatorSurname_(Subject)SR

Procedure:

1. Make sure that you are storing your file in the appropriate folder (see folder creation SOP).
2. Make a multi-part filename, with the parts separated by underscores, in the following format: Date (YYYYMMDD)_InvestigatorSurname_(Subject)SR_SearchStrategy.docx
3. The fourth part in the multi-part filename should always refer to a term from the Systematic Review process. If more detail is needed to describe the file, include this detail in a fifth part.
4. If there is more than one word in one of the parts, demarcate a new word with capital letters (see "SearchStrategy" in the example in step 2).
5. Do not give two files exactly the same name. Differentiate them with "v02" for version 2 at the end of the filename, or another appropriate designation.

Rule 5

Batch re-naming: when everything needs to be fixed



- Popular batch renaming tools:**
- [Bulk Rename Utility](#) (Windows)
 - [File Renamer Basic](#) (Windows)
 - [Wild Rename](#) (Windows)
 - [Renamer](#) (Mac)
 - [Name Changer](#) (Mac)
 - [Name Munger](#) (Mac)
 - [Gnome Commander](#) (Linux)
 - [PSRenamer](#) (Windows, Mac or Linux)
 - [Adobe Bridge](#) (via Creative Cloud – Windows or Mac)



**Best
Practices for
File
Organization**

Rule 6

Choose a folder organizational style: hierarchical or tag-based

	Hierarchical	Tag-based
Structure of Information	Good at representing structure of information	Less good at representing structure; not how operating systems store files
Display of individual files	Similar items are stored together	Related items may be split up in multiple places
Difficulty of Setup	Can be challenging to set up; needs much forethought and balance of breadth vs. depth	Easy to set up
Difficulty of change	Can be challenging to reorganize when changes occur	Can adapt to changes more readily, especially with batch changes

Rule 6

Tag-based file organization

Windows:

<https://www.maketech-easier.com/tag-files-in-windows/>)

- Click on one or more files, right-click, and select Properties
- Go to Details tab, click on Tags, and add as many as needed
- Can do same from View – Details Pane in Windows 10

Mac

<https://www.imore.com/how-set-and-start-using-finder-tags-macos>

- Open a new Finder window
- Click Finder in the top menu bar, then Preferences, then Tags
- Use the color tags or add more, create your own names



Rule 7

How to set up hierarchical folder organization

Try not to repeat the same categories in multiple places, or if categories do repeat, there should be a business-related reason

Potentially overlapping categories

Name

- 📁 Interview Data
- 📁 Interview Data For Publication
- 📁 Processed Interview Data

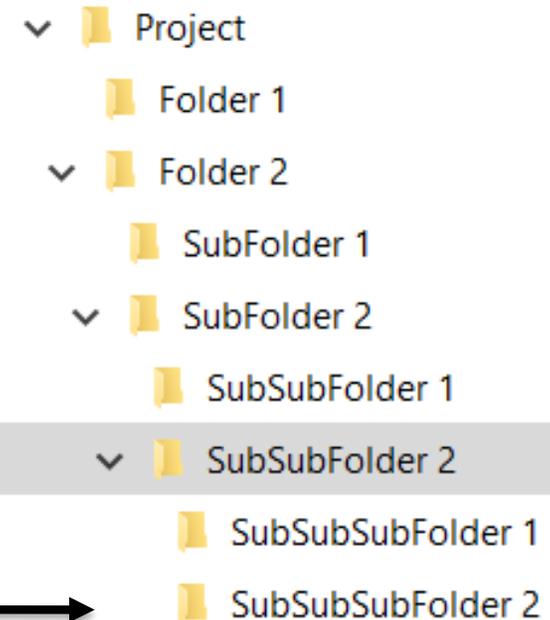
More clearly defined categories

- ▼ 📁 Data
 - 📁 Processed Data
 - 📁 Raw Data
- ▼ 📁 Manuscript
 - 📁 Processed Data for Publication
- ▼ 📁 Reports
 - 📁 Processed Data for Annual Report
 - 📁 Processed Data for Funder Report

Rule 7

How to set up hierarchical folder organization, part 2

- Strive for balance between amount of files per folder and number of sub-folders
- Don't allow folders to get overly-full
- Don't allow a nested folder structure to get too deep
- The deepest folders can be a pain to navigate to, and their relation to the top levels can be hard to see after enough subfolders are made



Rule 8

Keep the hierarchy organized and easy for others to understand

File everything right away

- Stray files might have simply missed filing in the rush of work
 - Try to file immediately, even as you name a file
- Stray files can also mean the older hierarchical constructs are no longer relevant

Include a README file in the top-level folder of the hierarchy

- Explains the folder organizational structure for newcomers to the project
- Can be updated as things change
- Can include a glossary of terms



Rule 9

Maintain your folder hierarchy

1 Create an archive folder

- Structure can mirror the main hierarchy
- Archive can house all earlier file versions you might want for reference

2 Cull files regularly

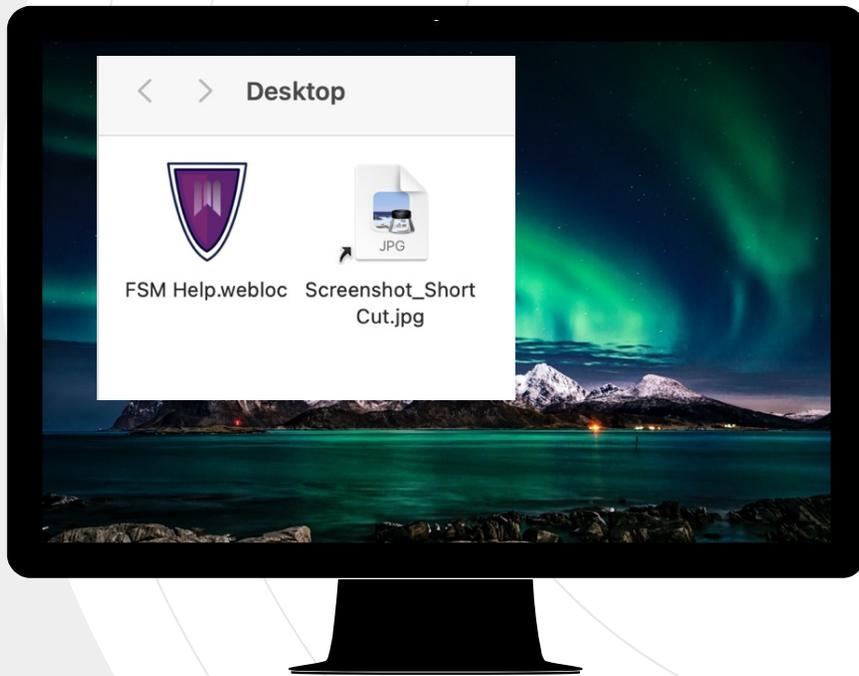
- Records management is part of good clinical practice
- Dispose of what you are not required to keep
- Save records you need for retention purposes in the archive folder(s)

3 Schedule a regular folder maintenance day

- Partners at multiple sites or different parts of an organization can clean and organize folders on the same day
- Monthly reminders can help ensure compliance

Rule 10

Need to frequently access files deep in a hierarchy? Create shortcuts



- Windows shortcuts: Right click, select "Create shortcut." Appears in the same folder, but can be moved
- Mac shortcuts: Select file name, then "Make Alias." Appears in the same folder, but can be moved
- Google Docs: Highlight doc URL and drag to desktop. Works best in Chrome.



Thanks!

Any questions?

You can find me at

- @saragon02 (GitHub)
- sara.gonzales2@northwestern.edu

Credits

- Malinowski, Christine. Data Management: File Organization. MIT Libraries, 2016. https://libraries.mit.edu/data-management/files/2014/05/FileOrgSlides_20170118sm.pdf
- MANTRA Research Data Management Training, The University of Edinburgh <https://mantra.edina.ac.uk/organisingdata/>
- McNeill, Katherine and Bailey, Helen. Research Data Management: File Organization. MIT Libraries, 2014. <https://libraries.mit.edu/data-management/files/2014/05/file-organization-july2014.pdf>
- Microsoft: Windows Dev Center. Naming Files, Paths, and Namespaces, 2018. <https://docs.microsoft.com/en-us/windows/win32/fileio/naming-a-file>
- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)

This slideshow template is free to use under [Creative Commons Attribution license](#)

Developed resources reported in this presentation are supported by the National Library of Medicine (NLM), National Institutes of Health (NIH) under cooperative agreement number 1UG4LM012346. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.