

Data Storage 102:

Time-saving organizational strategies

Presenter: Llewellyn Fernandes

Date: October 15, 2025

Today we'll cover



Saving time with simple structures and efficient management



Building reproducible workflows that anyone can follow



Protecting your data and manage access securely



Staying organized at both the personal and project levels



Sharing and publishing your data confidentially

Why Organization Matters

- Data is getting bigger and more complex
- Vendors are increasing prices
- Can't keep everything in the same place forever anymore
- Smart organization saves **time, money, and frustration**

Data Storage is getting expensive



Benefits of Data Organization

- **Conserve storage space**
 - Identify unused data
- **Increase findability**
 - Standardizing structure
- **Promote collaboration**
 - Easier to share with colleagues



Recap: Storage Options at Northwestern

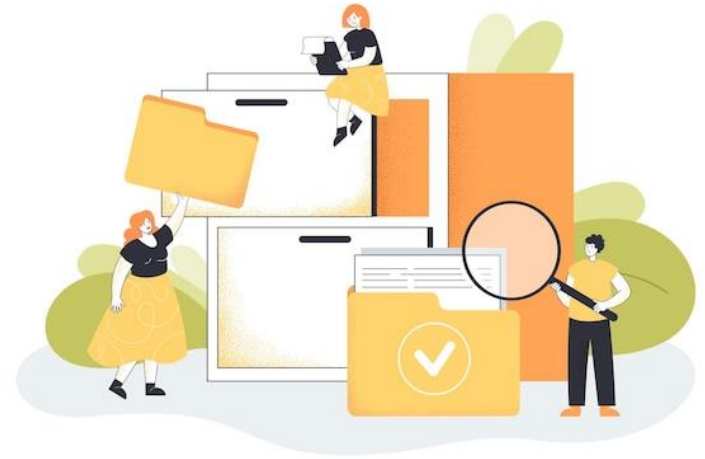
Storage	Who is it for	Good For	Key Feature
OneDrive	Individual/External	Storing working data that only you need to access	Auto-sync, version history
SharePoint	Group/External	Storing working data shared with a team	Controlled sharing
RDSS (resfiles, resfiles-aduit, fsmresfiles)	Group	Storing working data shared with a team, especially data with large individual file sizes	Reliable, centralized
Public Cloud (Google Cloud, AWS, Azure)	Group/External	Storing working data and archiving research data	Easy collaboration, pay-per-use
Quest (High-Performance Computing)	Group	Storing data being actively analyzed on Quest	Secure, large-scale storage

Principles of Data Organization

“Tidy” data is *findable, accessible, interoperable, reusable*

Would your "future self" thank you for how organized you are today?

- Good organization supports reproducibility and trust.
- Explore common formatting challenges.
- It saves time later.
- Documentation is part of the data.
- Think about and understand the types of metadata
- Think long-term.



Ask the right questions

- Can a new member in your project **easily find** and understand your data?
- If you stepped away for six months, could you **access** and make sense of it again?
- Can your files and formats **interact** smoothly with other systems or collaborators?
- Could someone else **reuse** your data, or even you, years from now?



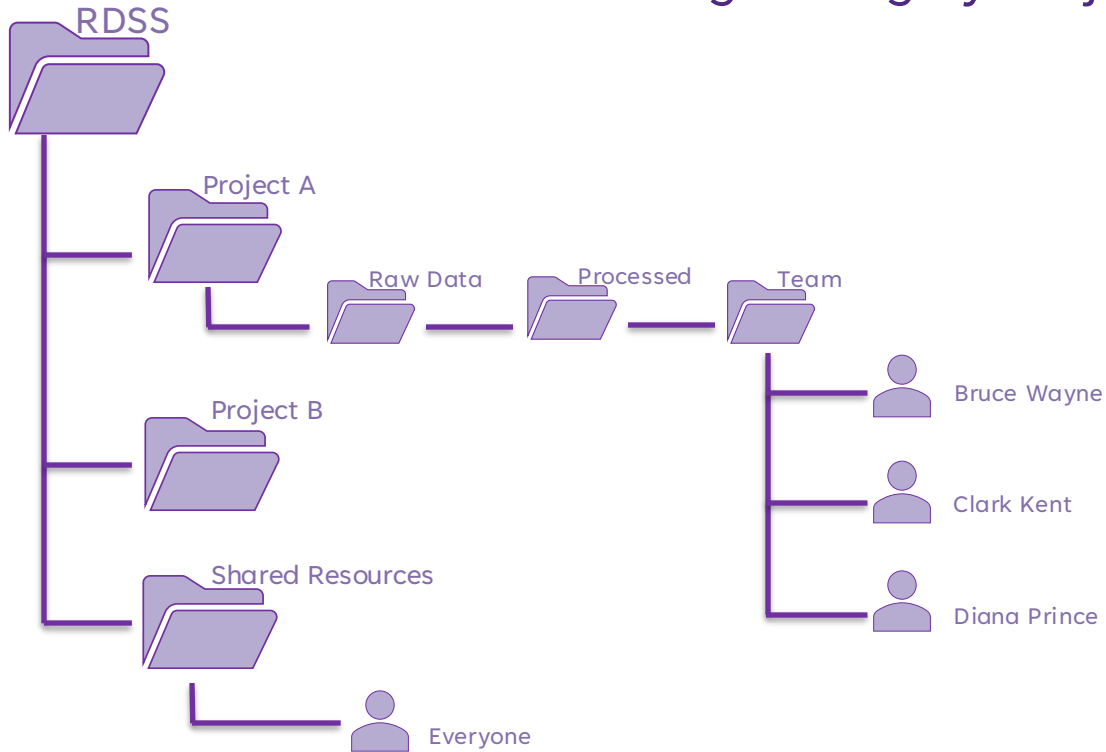
Organize your files

You can organize your files by...

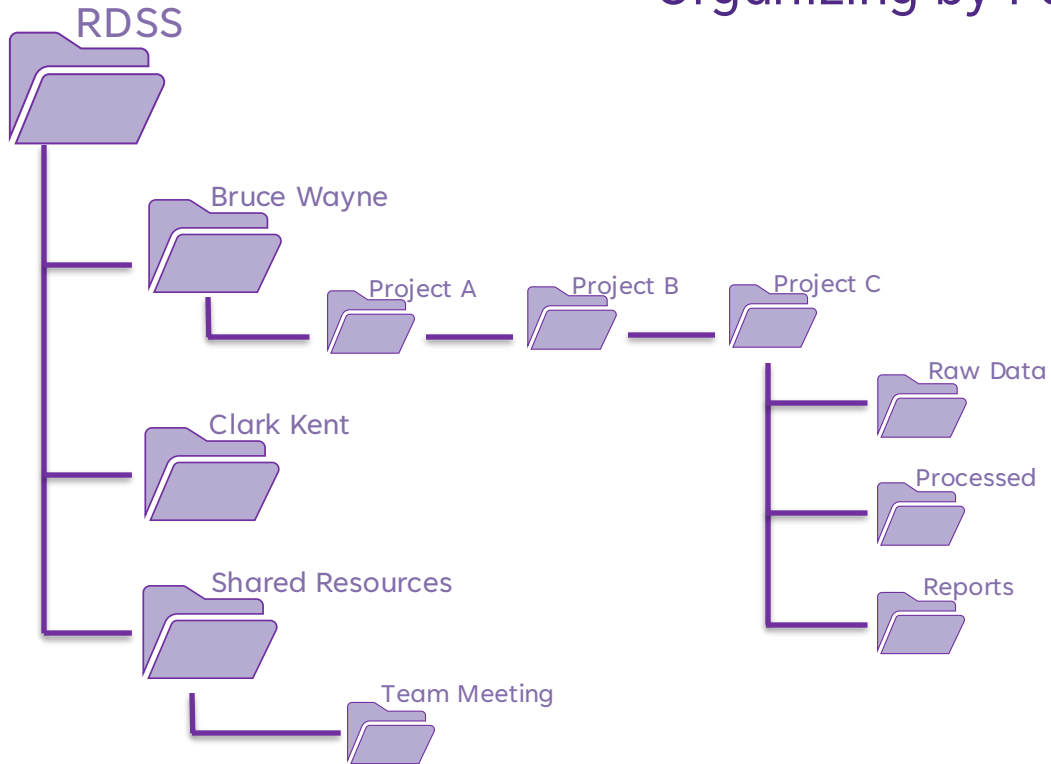
- Project or Individual Researcher
- Data Type (.csv, .fasta, .png, etc)
- Type of research activity (survey, assay)
- Subject characteristic (sex, species, etc.)
- Who needs access (Internal vs. External)
- Chronologically (Year 1, Year 2)



Organizing by Project



Organizing by Person



Deep Folder structures

No technical limit to how deep a folder structure can go, but...

Harder to Navigate

- People get lost in long, nested paths.

File Path Length Limits

- Some systems (especially Windows, Globus) **reject long paths**, causing errors in saving, syncing, or backups.

Redundant Structure

- Too many levels often duplicate information already in filenames

Collaboration Friction

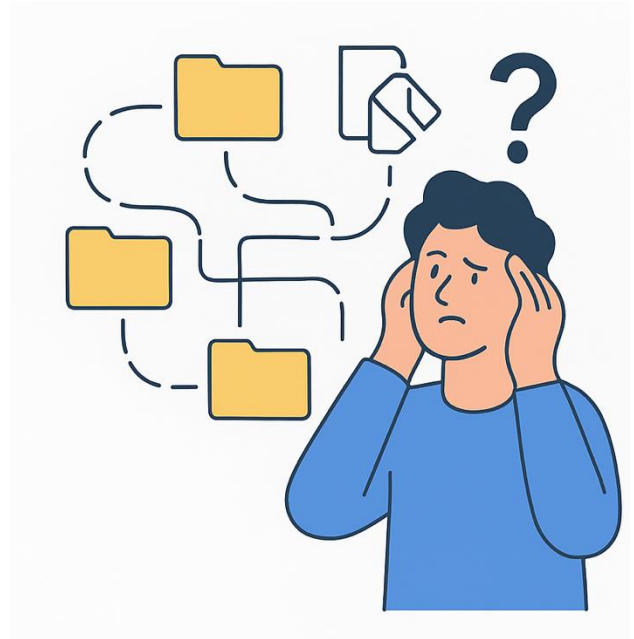
- Teammates may not know *where* to store or find files.

Transfer and Archival Issues

- Long paths and nested folders can cause **missed files** during automated backup or archiving processes.

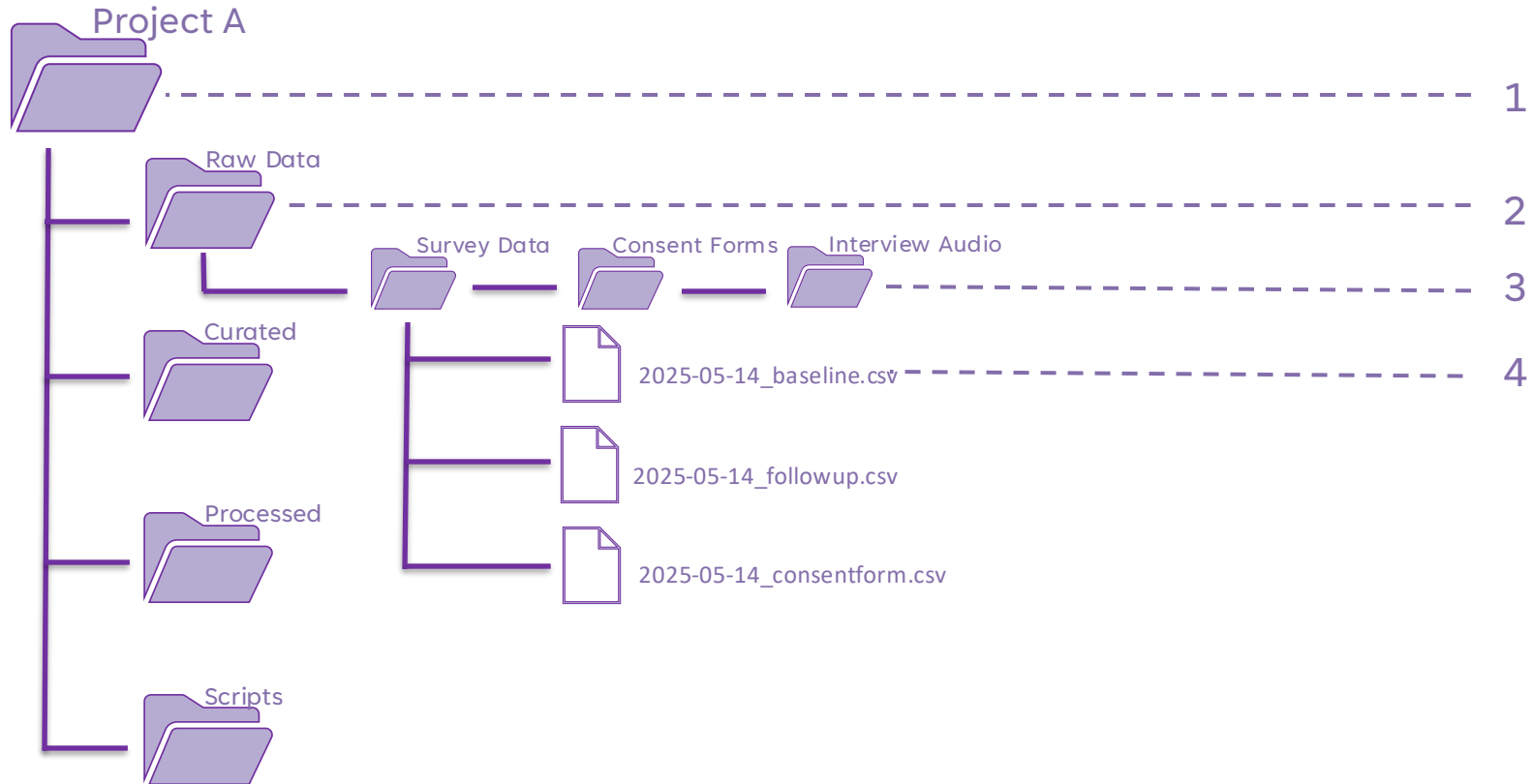
Cognitive Load

- The more levels, the harder it is to mentally map the structure



Follow the 4-level Rule...

The 4 Level Rule



Organizing data on Northwestern Servers

RDSS

- Self-Service: *You build the structure from scratch.*
- ResFiles & ResFiles-Audit users gain access to the entire share and all its files within. Granular permissions cannot be configured.
- FSMResFiles users can request granular permissions at a folder level.

Quest

- Home Directory: Code Files
- Project: Raw Data
- Scratch Space: Raw Data & Results

Organizing data Online

Public Cloud

- Self-Service: You build the structure from scratch.
- Granular permissions can be set.

SharePoint

- You can create Document Libraries to organize your project files and set granular permissions.
- You have User Groups to manage permissions.

OneDrive

- Meant for personal files and not research data, but you can share folders/files with others.

Good Organization Practices

There is no one right answer.

Make a plan.

Be consistent



Portability: Put all project files in one folder



Findability: Use descriptive file names



Navigability: Flatten your subfolder structure



Integrity: Use controlled access and backups to protect your data

File Naming Conventions

The file name should tell you what's in the file

Be kind to yourself

Name today, understand tomorrow

- If your file name looks like a secret code, six months from now, you won't crack it either.

Short, but not cryptic

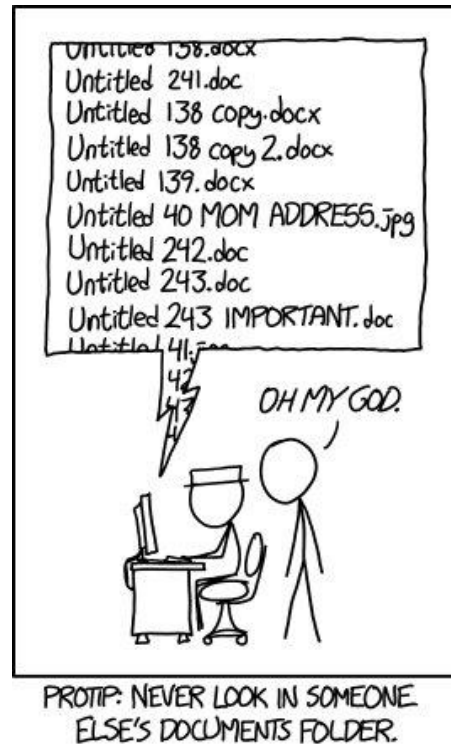
- Enough detail to understand *what and when*, but no essays.

Consistency beats creativity

- You don't need clever, you need predictable.
- Every file should follow the same naming format.

Create a Team Sheet

- Agree on what prefixes and terms mean (raw_, processed_, draft_).



Be kind to your computer

Spaces are troublemakers

- Use underscores '_' or hyphens '-' instead

Skip the Drama

- No more "final_final_reallyfinal_thisone.xlsx."
- Use simple version numbers (v1, v2) instead.

Use standardized dates

- ISO format (YYYY-MM-DD) keeps your files in perfect order.

Use sorting to your advantage

- Your computer is good at sorting things
- Name your files so that your computer's sorting is useful to you

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS **THE** CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013.II.27. 27½-13 2013.158904109
MMXIII-II-XXVII MMXIII ^{LVII}/_{CCLXV} 1330300800
((3+3)×(111+1)-1)×3/3-1/3³ 2013 Miss 2-2-13
10/11011/1101 02/27/20/13 0 1 2 3 4 5 6 7 8

<https://xkcd.com/1179/>

Use Descriptive File Naming

Trace the timeline

data.xlsx → data_final.xlsx → data_FINAL_really.xlsx → 2025-02-14_survey_data.csv

Create a cheat sheet

Topic	PREFIX	Content Type	Example
Social Science	raw	Raw survey data	2025-02-14_raw_survey.csv
Genomics	norm	Sequencing data	2025-03-02_norm_seq_sampleA.fastq
Engineering	Sensor	Sensor Readings	2025-03-10_sensor_bridgeA.csv

Finding the right balance

Too Vague	Too Verbose	Just Right!
data.csv	experiment_42_temperature_data_collected_from_labA_day3.csv	2025-02-14_exp42_temp_labA.csv

“When every file is called ‘final’, none of them are.”

Document your Approach

Use a ReadMe file to explain...

- Where the data comes from (instrument, survey, external dataset)
- How your folders are structured and what they should contain
- Your file naming convention
- Abbreviations or variable names you use
- How to store data for a new experiment (provide a template!)
 - Instructions on how to run or replicate analysis.
- Create metadata at each level

Make it your own!

Naming Best Practices



Consistency is key: Use naming conventions that help the future you



Standardize: Add in standards that help your system find/query your data efficiently



Descriptive yet meaningful: Find the right balance between informative without overload



Reproducibility: Document your approach, work with your team

Moving or Migrating Your Data

Consider this Scenario

Situation

The lab's premium storage subscription is being canceled due to reduced funding. All project data must be moved to a lower-cost storage solution before the current subscription expires.



Consider this Scenario

Problem

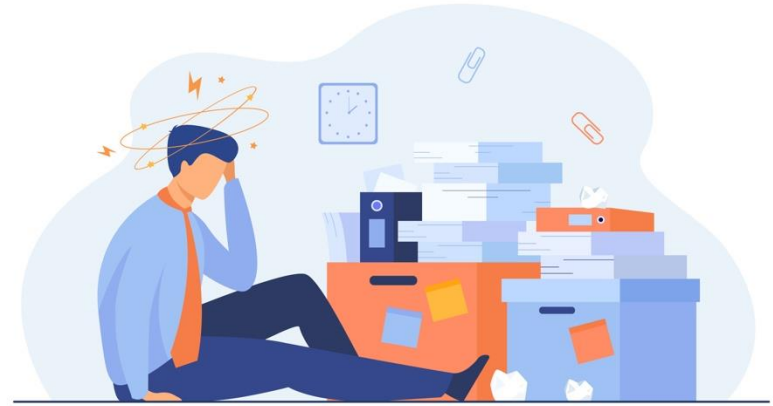
- Files are scattered across multiple drives and team members' personal accounts.
- Folder structures are inconsistent, and many files are poorly named.
- Documentation of datasets, metadata, and version history is incomplete.
- Scripts and pipelines reference paths in the premium storage, which will no longer exist.



Consider this Scenario

Impact

- Urgent migration is stressful and time-consuming.
- Risk of losing or overwriting critical data increases.
- Collaboration is disrupted because files are not centralized or clearly organized.
- Team may spend days reorganizing and relinking data instead of progressing on research.



Organization can help with migration

An organized data structure will help you...

- **Bundle** files into a file archive (.tar, .zip, etc)
- **Compress** files to save space (and \$\$\$) on the destination
- **Create** optimal file size for cloud storage and data transfer
- **Verify** all data arrived safely using checksums, manifests, and structured folders.
- **Resume** work quickly after migration with clearly organized files.
- **Locate** and **access** needed data effortlessly, reducing frustration for every team member.



Need Data Transferred?

Globus is the preferred method of data transfer

- Connects to RDSS/FSMResFiles, Quest, SharePoint, Cloud Storage, your computer
- Web or command line interfaces
- Retries if you get disconnected
- Checksum verification



Automations and Workflows

Automate the Boring (but Important) Stuff



File Synchronization: Automatically sync project folders across devices and collaborators.



Backup & Snapshots: Schedule automated backups to institutional storage or cloud servers.



Structure & Naming Enforcement: Use scripts (Python, R, PowerShell) to validate structure, naming, and required metadata.



Workflow Automation: Automate multi-step research workflows (e.g., data cleaning → analysis → output).



Monitoring & Reporting: Track file changes, generate daily logs, or send alerts when storage quotas are near limits.

Finally

If you plan, you're ahead of the game

1. Structure with purpose

Organize around projects, not people. Keep folder depth manageable (≤ 4 levels).

2. Design File Names with Intent

Use clear, consistent, and compact naming conventions, your future self will thank you.

3. Document everything

A simple README beats hours of detective work later. Explain structure, naming, and abbreviations.

4. Automate where possible

Let synchronization, backup, and validation tools do the heavy lifting.

5. Plan for portability

Well-organized data moves easily — whether you're changing platforms, collaborators, or roles.





FIND WHAT YOU NEED



PLANNING

- [Writing a Data Management Plan](#)
- [Protecting the Sensitive Information in My Data](#)



DATA COLLECTION AND STORAGE

- [Choosing Appropriate Storage](#)
- [Documenting Your Research](#)
- [Transferring Data to or from Northwestern](#)
- [Sharing Data with an External Collaborator](#)



DATA SHARING AND ARCHIVING

- [Making Your Data Reusable](#)
- [Sharing Data Publicly](#)
- [Archiving Data When a Project is Done](#)



SUPPORT AND RESOURCES

- [Talk to a Data Management Expert](#)
- [Northwestern Research Data Management Resources](#)
- [External Research Data Management Resources](#)

Reach out!

Visit the: [Research Data Management Website](#)

Email: researchdata@northwestern.edu

[RCDS Consult Form](#)

[RCDS Cloud Consult form](#)

[Galter Data Lab Consult form](#)

[Information Security: Protect your research](#)

Office Hours: Every Monday
3 p.m. – 4 p.m.
Mudd Library
Rooms 2202-2205
(2nd Floor across
the bridge to Tech)



Questions?