Smart Data Management

Strategies for Efficient Research Workflows

Instructor: Llewellyn Fernandes – Data Management Specialist

Date: February 11, 2026

Materials: https://github.com/nuitrcs/rdm-workshops

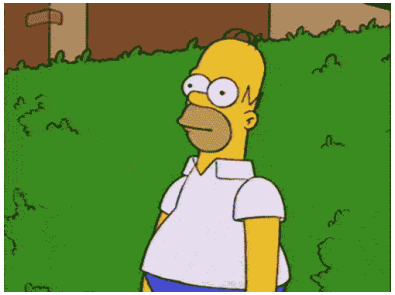
Northwestern
INFORMATION TECHNOLOGY

Let’s talk about your data

Why a workshop about data management?

Data Management is often an afterthought

- Data collection was **slow and limited in volume**
- Storage was relatively inexpensive
- Labs could realistically **keep everything forever**
- Data reproducibility demands were simpler



Managing Old Data

Collecting New Data

This strategy used to work,
BUT...

The storage landscape is changing

- 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**

Some common issues

- Version Chaos and File Naming Breakdown (*Final_v3_REAL_final.xlsx*)
- Data stored on personal laptops
- No one knows which copy is authoritative
- Students graduate → knowledge disappears
- Old data is taking up expensive storage
- Accidental overwrites or deletions
- Compliance / IRB / grant data retention risks

Today we'll cover how to...

- Take stock of your research data
- Note key characteristics that affect your storage decision
- Choose storage that fits your workflow
- Organize data for efficiency and reproducibility
- Manage, protect, and share data with confidence

Start with what you Actually Have

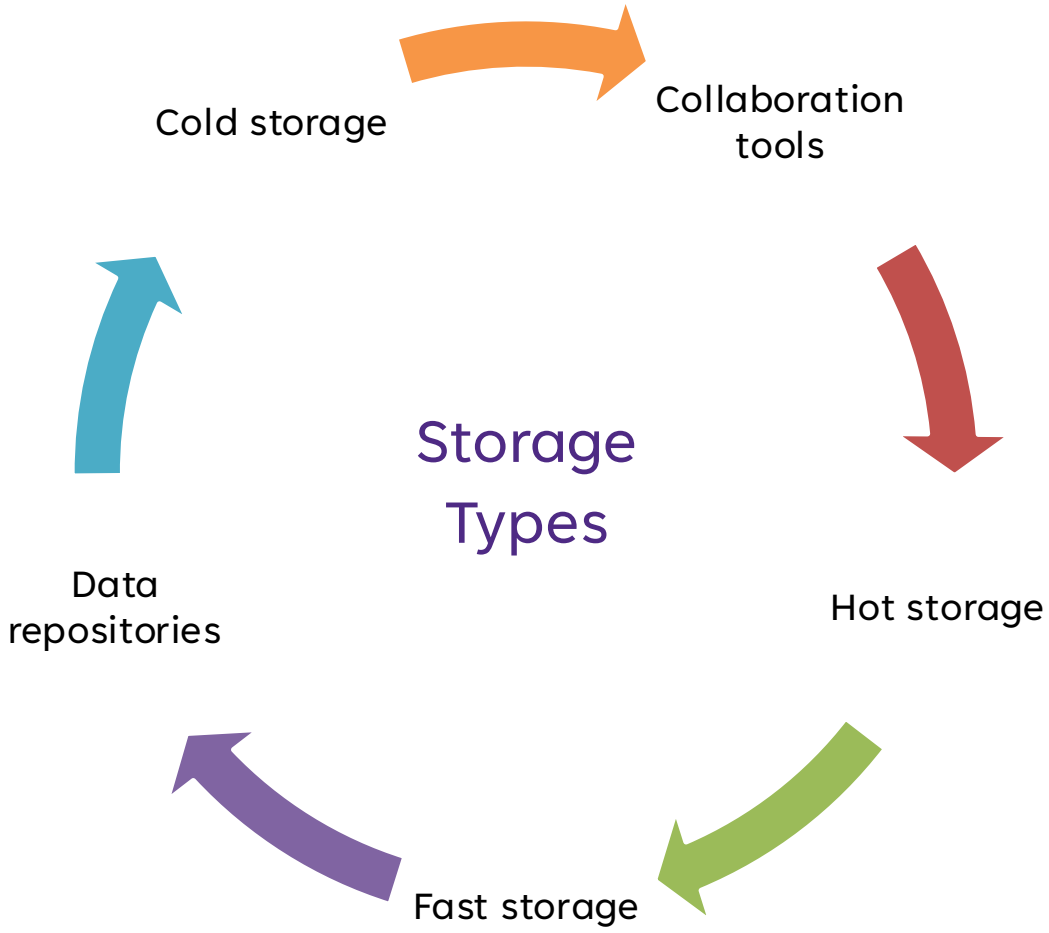
- What kind of data do you generate?
- How large are the datasets?
- How fast will the data grow?
- Who needs access?
- How long must it be retained?
- What data is still active vs. untouched?

Get an Inventory Done

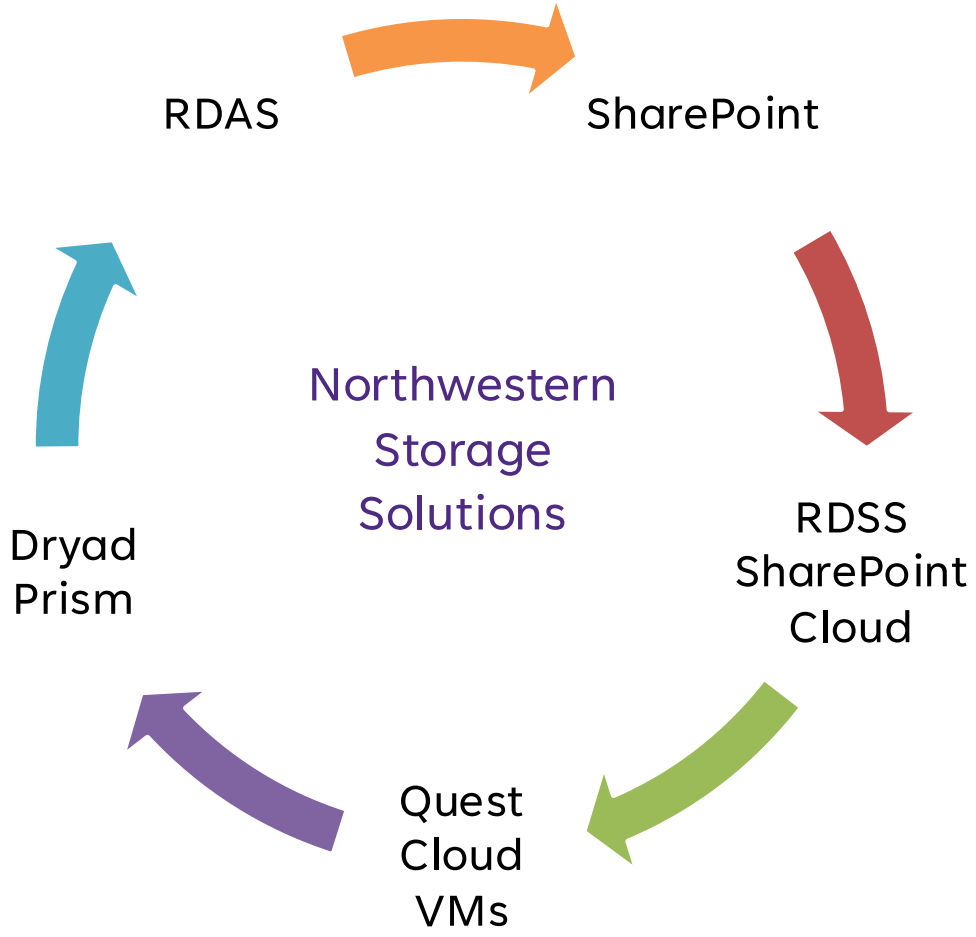
Research Workflow

Not All Data Should Live in the Same Place

Characteristic	Why It Matters
Size & growth rate	Determines scalability needs
Frequency of access	Active vs. archival placement
Sensitivity	Security and compliance requirements
Collaboration level	Sharing platform choice
Longevity	Retention vs working storage
Compute needs	Proximity to analysis environments


Storage Types
Collaboration tools
Hot storage
Fast storage
Data repositories
Cold storage

1. Align your Storage with the Research Lifecycle


RDAS
SharePoint
Northwestern
Storage
Solutions
RDSS
SharePoint
Cloud
Quest
Cloud
VMs
Dryad
Prism

Good Organization is a Time-Saving Tool

- **Clear folder structure** beats complex systems
- **Consistent naming** prevents duplication
- **Reduce "search time"** and rework
- Store in an **accessible** location
- Keep **multiple copies** in case one gets corrupted
- **Limit access** to those who need it

2. Build Reproducible Workflows Others Can Follow

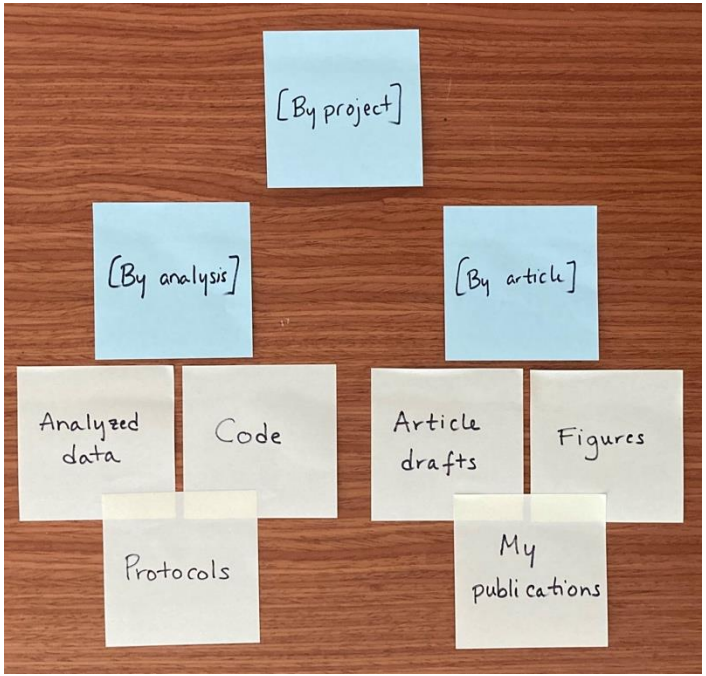
Make Your Research Understandable Without You

- **DOCUMENT** structure, not just results
- Store **metadata** alongside data
- Avoid personal-device silos
- Enable new students to **onboard quickly**
- Explore common formatting challenges.
- **Think long-term.**

How to organize?

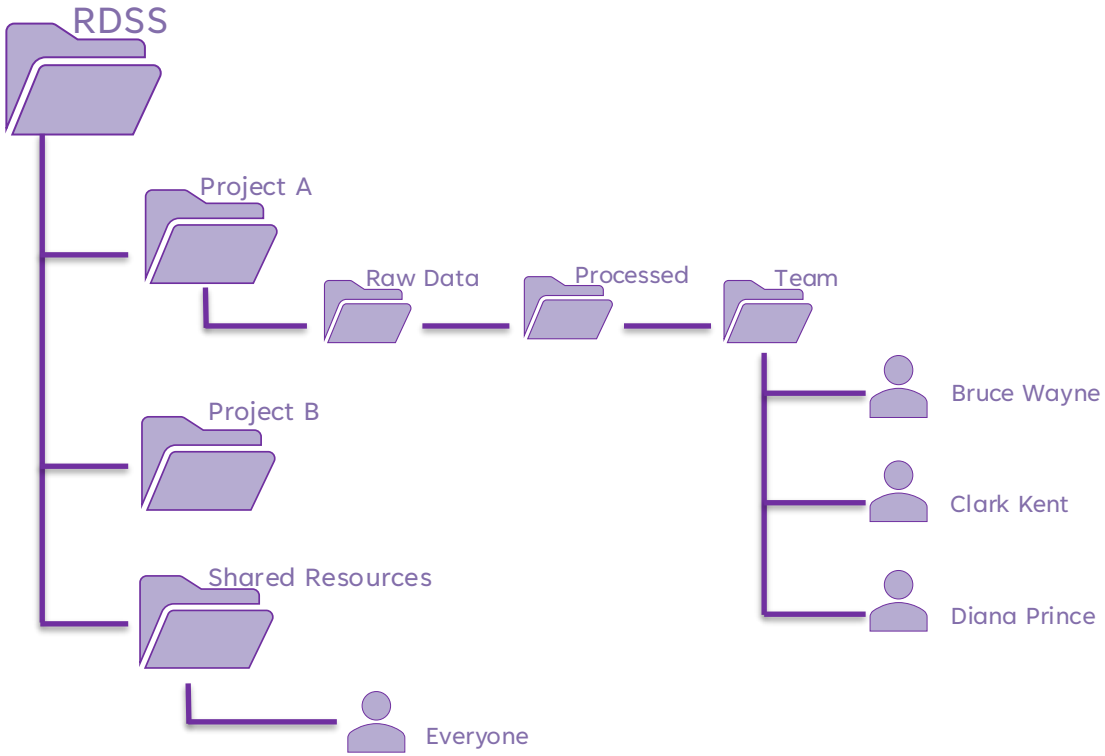
You can organize your files by...

- 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)

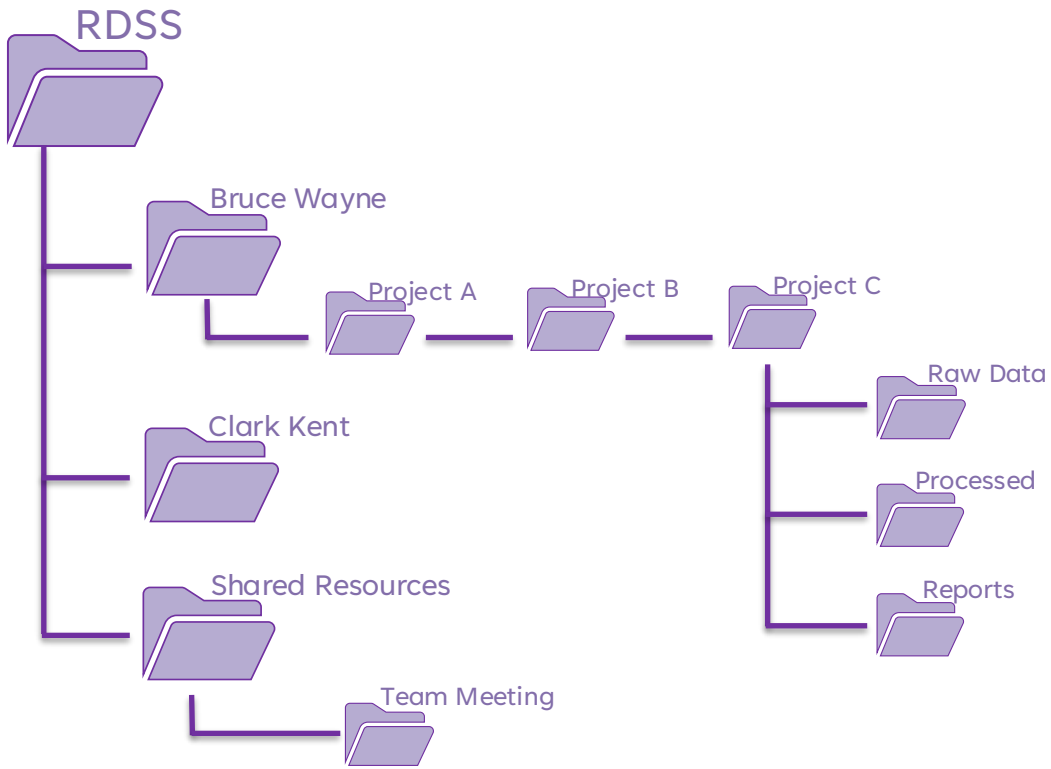


The Research Data Management Workbook - Kristin Briney

Organizing by Project



Organizing by Person

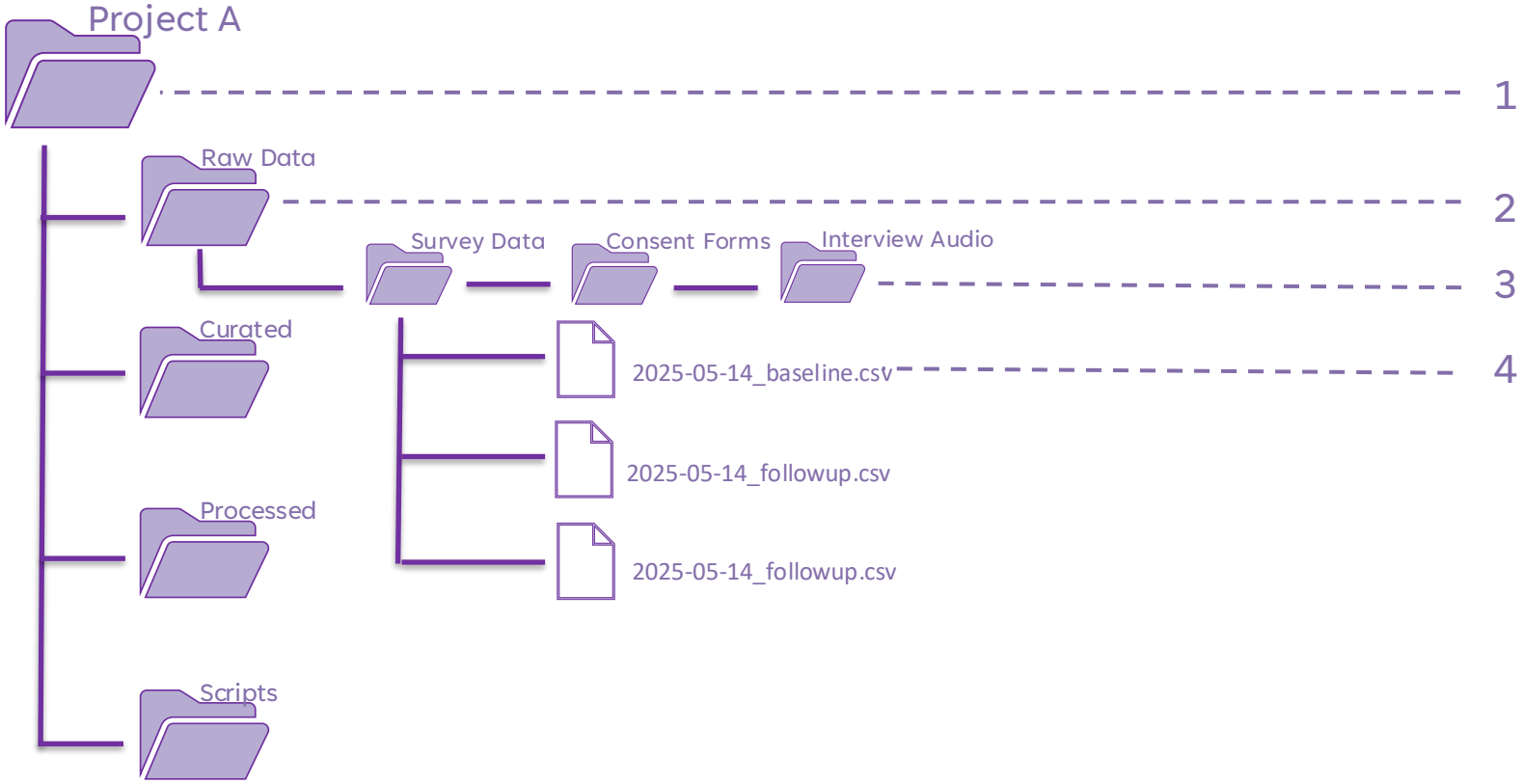


Data Management Is Ongoing, Not a Grant Requirement

- People get lost in long, nested paths
- Long paths can break saving, syncing, or backups
- Multiple levels often duplicate info already in filenames
- Teammates struggle to find or store files
- Deep folders can cause missed files
- Hard to mentally map the structure

Follow the 4-level Rule...

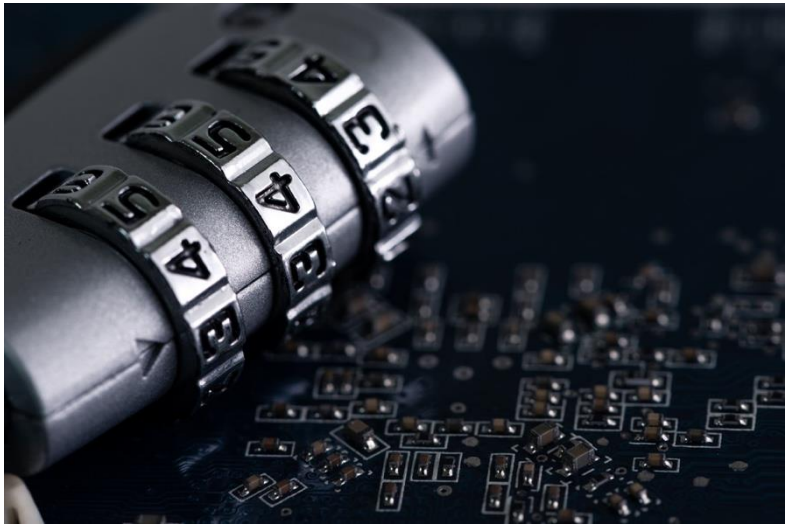
The 4 Level Rule



3. Protect Your Data and Manage Access Securely

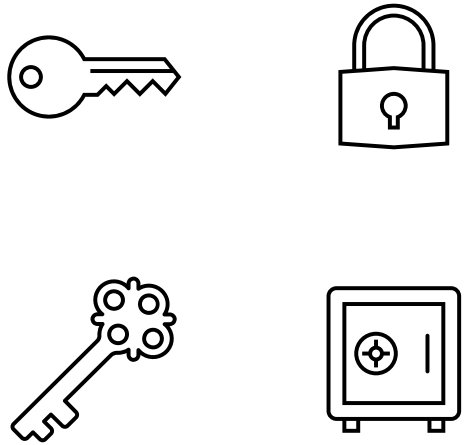
Security and Access are part of Data Management

- Avoid uncontrolled sharing
- Use role-based permissions
- Keep sensitive data in approved environments
- Reduce risk of accidental loss or exposure



Who needs access to the data and when?

- People from your research group
- People at Northwestern
- Collaborators external to Northwestern
- The Public



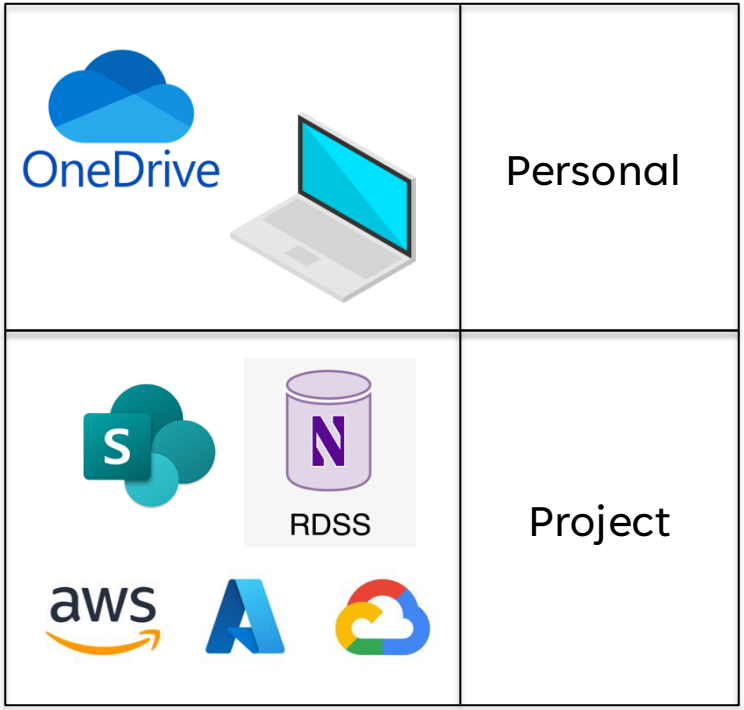
Know what requirements your data is subject to

- **Northwestern polices**: Use approved storage systems
- **Grant and contract terms**: Controlled Unclassified Information (CUI)
- **Data use agreements (DUAs)**: Specific controls (e.g., encryption) or security standards (NIST 800-171, HIPAA security rule)
- **Federal and state regulations**: HIPAA, BIPA, FERPA

4. Stay Organized at Both the Personal and Project Levels

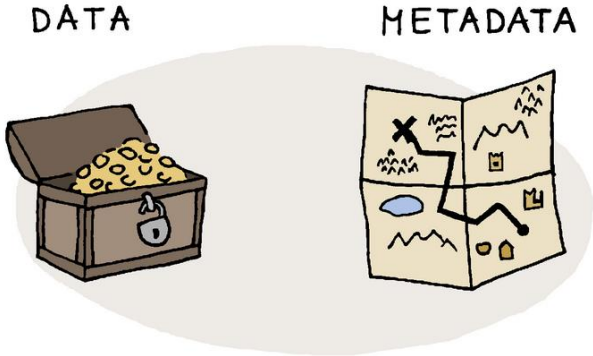
Separate Personal Workspace From Research Repository

- Mixing personal drafts and official project data leads to duplication, lost files, and confusion
- Personal folders often lack consistent naming, metadata, or access control
- Institutional storage must remain the ***single source of truth***



Who needs access to the data and when?

- If you're storing research data in OneDrive, move finalized work to shared storage regularly
- Use a consistent folder naming convention when transitioning files
- Keep metadata and documentation alongside files, not just in personal notes

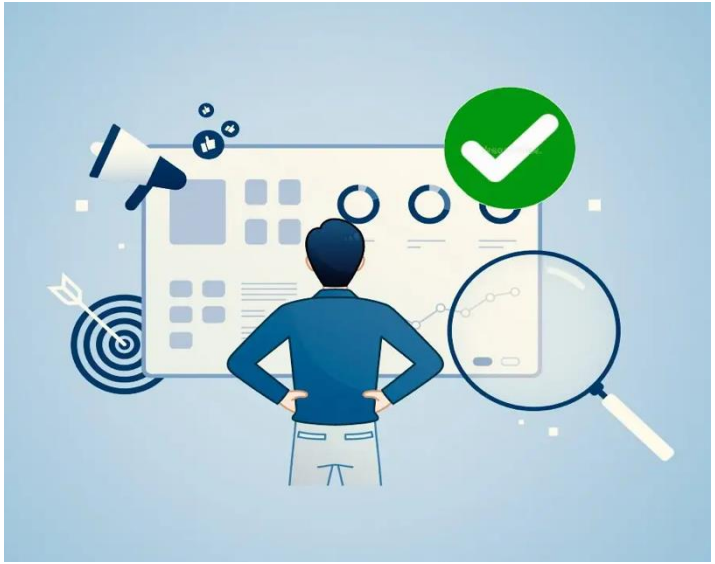


Source: https://dataedo.com/blog/data-vs-metadata

5. Share and Publish Your Data With Confidence

Well-Managed Data Is Easier to Share, Validate, and Archive

- Cleaner handoffs to collaborators
- Easier compliance with funders and journals
- Clear archival pathway after publication
- Data remains usable years later



Storage ≠ Backup & Backup ≠ Archive

Function	Purpose	What It Protects Against	What It Is Not
Active Storage	Where you work on data	Daily access needs	Not a safety net
Backup	Restore after failure or mistakes	Deletion, corruption, hardware loss	Not long-term curation
Archive	Maintain a stable, final record	Time, turnover, compliance needs	Not for editing

Research Data Archival Service

Finally

Key Takeaways

Align your storage with the research lifecycle

- Keep active work, collaboration, and archives in the right environments.

Build reproducible workflows others can follow

- Structure your data so it remains understandable beyond any one researcher.

Protect your data and manage access securely

- Safeguard integrity while enabling the right level of collaboration.

Stay organized at both the personal and project levels

- Separate individual workspaces from the shared source of truth.

Share and publish your data with confidence

- Well-managed data is easier to validate, archive, and reuse.

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