The Ins and Outs of Data Management

David Stone and Joanna Thielen
January 25, 2017

CETL Research Workshop Series

CC BY NC SA

Today's workshop

---

1. **Overview** of data management and its importance
2. Snapshot of **data management topics**
    a. Data Management Plans
    b. Documentation and Organization
    c. Storage and Security
    d. Preservation
3. **Key takeaways** about data management
4. **Introduction to ORCID** (Open Researcher and Contributor ID)

Overview of data management and its importance

Have you ever…

- Lost data? (loss/theft of laptop, hardware/software failure, etc.)
- Been unable to locate a specific file?
- Been unable to locate and/or re-use someone else’s data?

What is data?

Data is **anything you perform analysis on.**

Data can be physical or digital

What is research data management (RDM)?

Research data management is the **compilation of small practices** that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later.

In Chapter 1: The Data Problem in *Data Management for Researchers* by Kristin Briney

RDM includes...

- Data management planning
- Documenting your data
- Creating metadata about your data
- Organizing your data
- Improving analysis procedures
- Securing sensitive data properly
- Having adequate storage and backups
- Taking care of your data after a project
- Sharing data effectively
- Finding data for reuse in a new project

In Chapter 1: The Data Problem in *Data Management for Researchers* by Kristin Briney

RDM includes...

- **Data management planning**
- **Documenting your data**
- Creating metadata about your data
- **Organizing your data**
- Improving analysis procedures
- **Securing sensitive data properly**
- **Having adequate storage and backups**
- **Taking care of your data after a project**
- Sharing data effectively
- Finding data for reuse in a new project

In Chapter 1: The Data Problem in *Data Management for Researchers* by Kristin Briney

Why is RDM important?

You don't want to lose your data!!!

Saves valuable time and resources

Support open data initiatives

Comply with federal funding agency requirements

Federal funding mandates

Each of the federal funding agencies has their own data management policy

- At NSF, each directorate has their own specific policy
- NIH has a public access policy and a data sharing policy for grants in excess of $500,000

For a complete rundown, see: Wapole and Stone, “Data management and research integrity,” in Research Regulatory Compliance, (eds. Suckow and Yates), Academic Press, 2015

Snapshot of data management practices

Data Management Plans

What is a Data Management Plan (DMP)?

Two page addendum to a grant application that states your **data management practices during and after the proposed project**

Other names:

- Data Sharing Plan
- Data Products Plan
- Data Access Plan
- Data Plan
- Information Products Appendix

A DMP should answer these questions:

- What **type of data** will be produced? Will it be reproducible? What would happen if it got lost or became unusable later?
- **How much data** will it be, and at what growth rate? How often will it change?
- **Who will use it** now, and later?
- **Who controls it** (PI, student, lab, institution, funder)? Who in the research group will be responsible for data management?
- How long should it be **retained**? e.g. 3-5 years, 10-20 years, permanently
- Are there **tools or software** needed to create/process/visualize the data?
- How will the data be **shared**?
- Any **special privacy or security requirements**? e.g., personal data, high-security data
- How is the project and data being **documented**?
- What **directory and file naming convention** will be used?
- What **project and data identifiers** will be assigned?
- What **intellectual property issues** are there?
- What **file formats**? Are they long-lived?
- What is the **storage and backup strategy**?
- When will I **publish** it and where?
- Is there an **ontology or other community standard** for data sharing/integration?

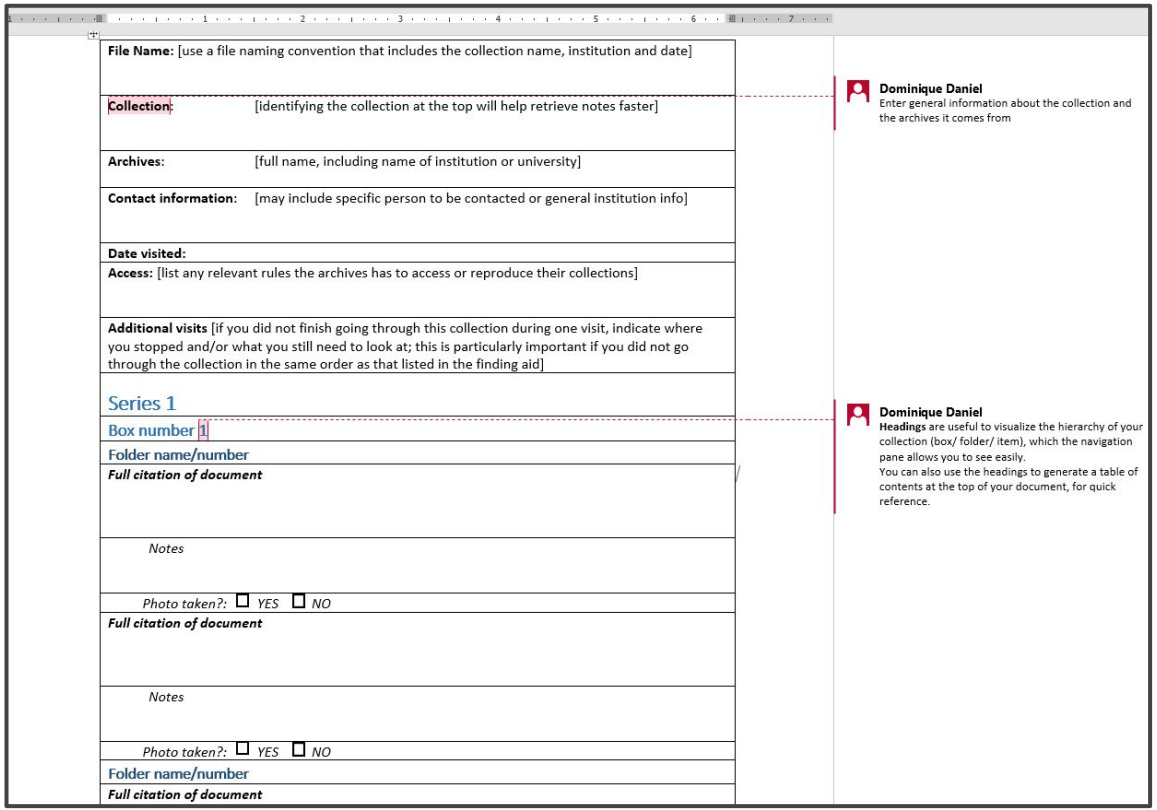
UCLA Library Data Management for the Sciences http://guides.library.ucla.edu/c.php?g=180539&p=1190233

Writing a high quality DMP sets you apart from other grant applications

“**I am and will continue to be** a trustworthy steward of your limited grant dollars”

Documentation and Organization

Using a template for your research notes/lab notebooks



File Name: [use a file naming convention that includes the collection name, institution and date]

Collection: [identifying the collection at the top will help retrieve notes faster]

Archives: [full name, including name of institution or university]

Contact information: [may include specific person to be contacted or general institution info]

Date visited:

Access: [list any relevant rules the archives has to access or reproduce their collections]

Additional visits [if you did not finish going through this collection during one visit, indicate where you stopped and/or what you still need to look at; this is particularly important if you did not go through the collection in the same order as that listed in the finding aid]

Series 1

Box number 1

Folder name/number

Full citation of document

Notes

Photo taken?: ☐ *YES* ☐ *NO*

Full citation of document

Notes

Photo taken?: ☐ *YES* ☐ *NO*

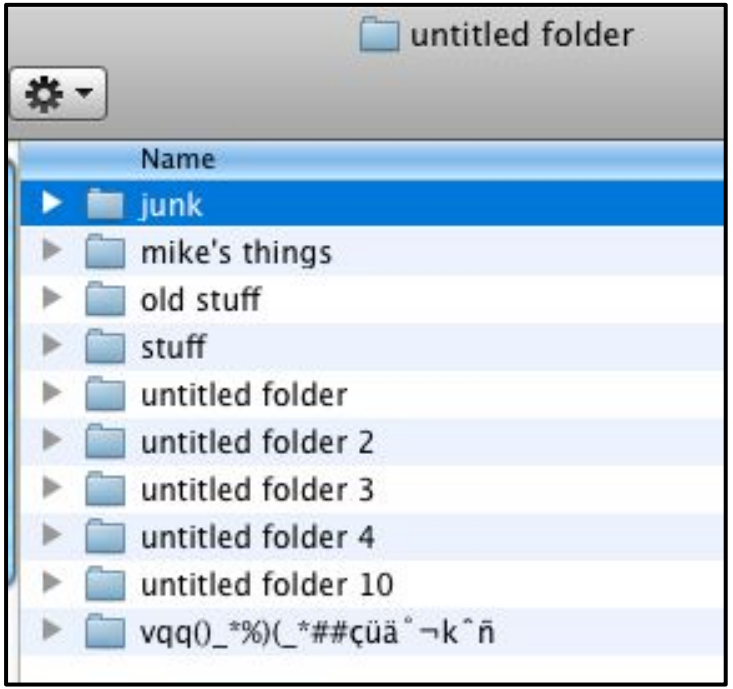
Folder name/number

Full citation of document

Dominique Daniel
Enter general information about the collection and the archives it comes from

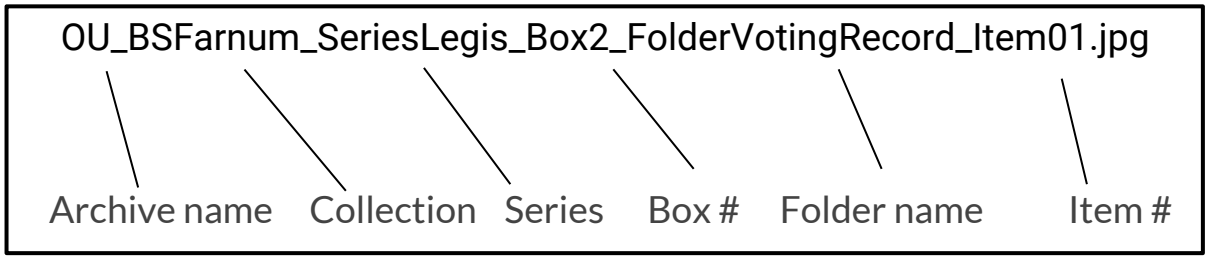
Dominique Daniel
Headings are useful to visualize the hierarchy of your collection (box/ folder/ item), which the navigation pane allows you to see easily.
You can also use the headings to generate a table of contents at the top of your document, for quick reference.

Do your digital files look like this?

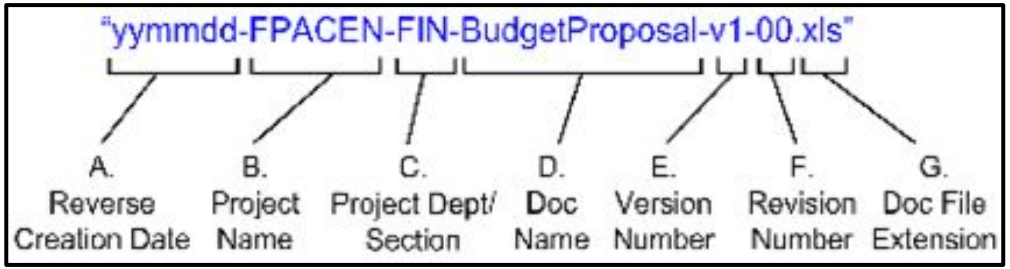


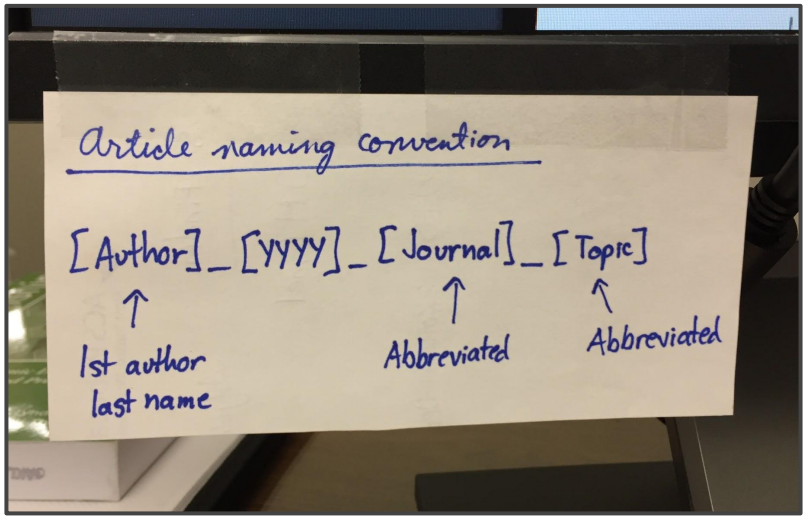
Use a file naming convention

- **Standardized naming system that gives each file a unique name**
  - Name describes:
    - Contents of file
    - Relation to other files
- **FNC makes your files easy to understand and search**



Examples of file naming conventions





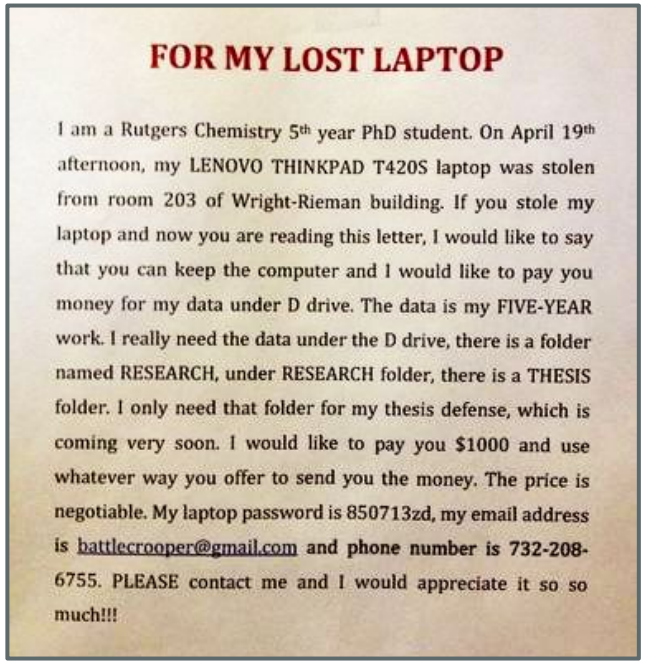
CONSISTENCY IS KEY!

Why? Because **YOU** are the likely beneficiary of good documentation & organization practices!

Storage and Security

Why is data storage important?

Natural disaster



FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you $1000 and use whatever way you offer to send you the money. The price is negotiable. My laptop password is 850713zd, my email address is battlecrooper@gmail.com and phone number is 732-208-6755. PLEASE contact me and I would appreciate it so so much!!!

Loss or theft

Practice the 3-2-1 rule

Example storage plan:

I will keep my data on my personal laptop and back it up on my personal external hard drive and OakShare (files.oakland.edu). My laptop and external hard drive will be stored at my house whereas OakShare is stored on servers at OU.

Storage media	Recommended?	Disadvantages
Personal computer	YES	Prone to theft or loss
External hard drive	YES	Subject to degradation; lifetime is ~ 5 years
CD/DVD	YES	Subject to degradation due to mishandling; can be laborious to use
USB flash drive	NO	Easy to lose; very fallible
Cloud service (Dropbox, Box, etc.)	YES	Don’t use for confidential data; by agreeing to their Terms of Service, you may be giving the company a license to use your files (including data files)

Confidential research data

OU Policy 860 on Information Security:
Confidential research data includes “information related to a forthcoming or pending patent application, grant applications and proposals, information related to human subjects”

Securely storing confidential, digital research data

- Password protect files
- Don’t put confidential data in email or on the internet (e.g. DropBox)
  - Use OakShare (files.oakland.edu)
- Encrypt all files and devices that access confidential data

Preservation

Why preserve research data?

So you (or others) can **continue to access and re-use** your data for as long as necessary.

Unlike physical data, **digital data needs to be carefully curated** to mitigate the risk of software obsolescence, hardware failure, etc.

How to preserve your research data

- Convert to non-proprietary formats (CSV, PDF, TXT, etc.)
- Consider using a data repository

Examples: 

Lists of data repositories:

- Registry of Research Data Repositories (re3data.org)
- Simmons College list of repositories by discipline (bit.ly/data_repositories)

Takeaways about data management

Implementing RDM practices can seem arduous, time-consuming, a bottomless pit of busy work…

BUT

For every **one minute spent** on implementing RDM practices, you could **save ten minutes** of headache later

Thinking back to last week's CETL workshop on Research Positioning…

Practicing RDM can also help to position yourself as a researcher and grant writer

Start small with RDM – don't be overwhelmed!

Pick one RDM practice to implement… Master it before moving on

1. Storage - start regular back ups!
2. File naming convention
3. More robust documentation
4. …

Consistency is more important than complexity!

Resources at the OU Libraries

- Data Management online guide (bit.ly/OUDataMgmt)
- Data Management for Researchers by Kristin Briney (eBook)

Introduction to ORCID

What is ORCID (Open Researcher and Contributor ID)?

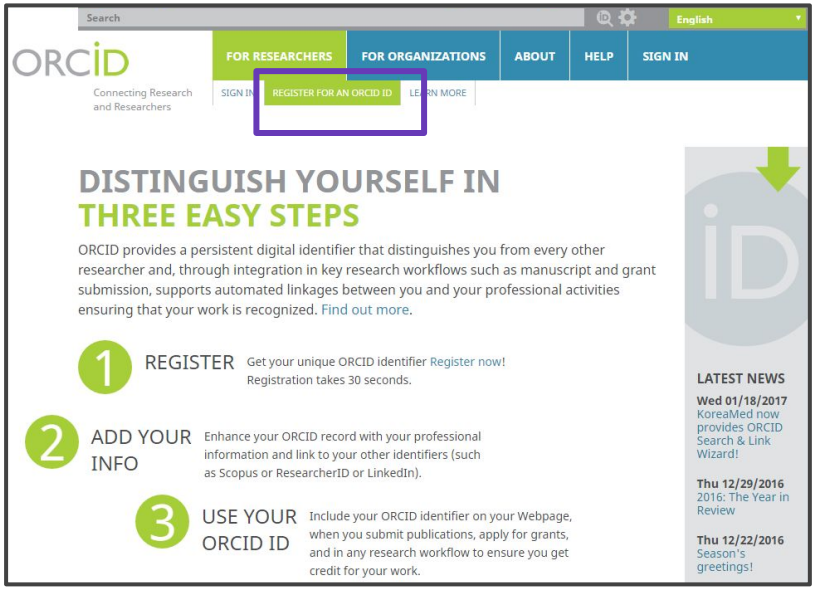
- Unique, persistent numerical identifier
  - Joanna's ORCID:0000-0002-2983-5402
- Aim to solve the name ambiguity problem
- Profile: digital version of CV
- ORCID required for manuscript submission in many journals (PLOS, IEEE, Hindawi, Science journals, Wiley, etc.)

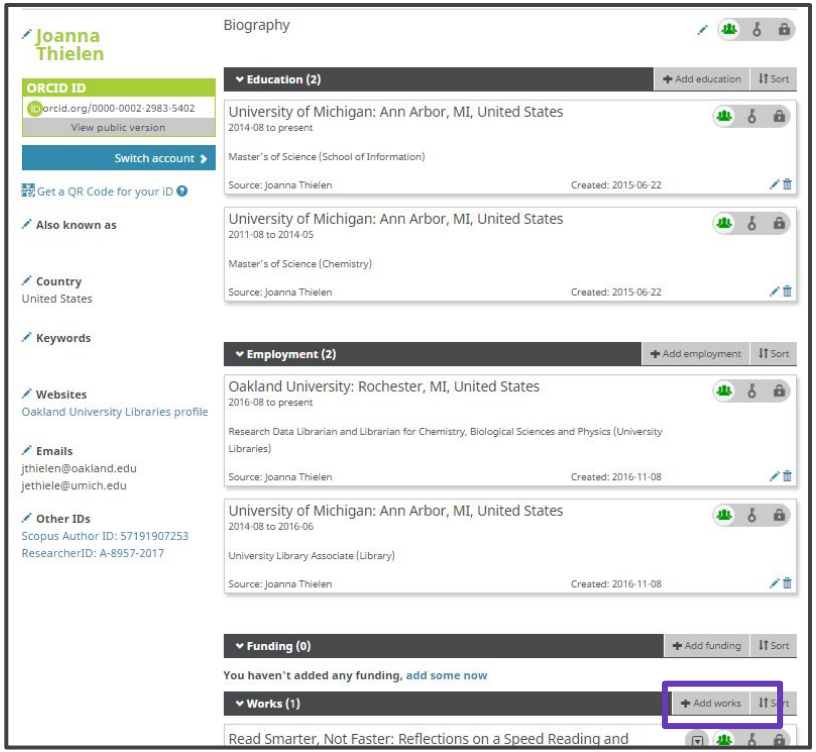
ORCID

Benefits of ORCID

- Profile links to your publications
- Differentiates you from other researchers with same name
- Static identifier regardless of job change, name change, etc.
- Positions you as a scholar in your field

How to Create an ORCID account (orcid.org)





Other Online Profiles

Research Author Profiles

- ORCID
- ResearcherID - Thomson Reuters product
- Scopus Author ID - Elsevier product

Scholarly Network Profiles

- Google Scholar
- Academia.edu
- ResearchGate
- Mendeley

Thank you for attending!

Please fill out the online CETL survey!

David Stone
Chief Research Officer
dstone@oakland.edu

Joanna Thielen
Research Data Librarian
jthielen@oakland.edu