

Research Data Management

Matthew Murray

Data Librarian

matthew.murray-2@colorado.edu



What we'll talk about

- What is research data?
- Data Management vs. Data Management Plans
- Data Storage & Access
- Backups & Versioning
- Documentation
- File Names & Structures
- File Formats & Units of Measurement
- Data Management Plans
- Tools

What is Research Data?



**No two people on any
campus will agree on
what “Data” actually is.**



What is Research Data?

- Primary / Secondary
- Qualitative / Quantitative
- Experimental / Observational

What is Research Data?

- 3D models and printable files
- Accreditation reports
- Archival university papers
- Artistry and performance materials
- Audio
- Books
- Computer code & scripts
- Conference proceedings
- Course catalogs
- Datasets
- Designs & blueprints
- Digital journals
- Dissertations
- Documentation
- GIS files
- Grant proposals
- “Grey” literature
- Historical documents
- Images
- Interviews
- Journals
- Lab notebooks
- Learning materials
- Lecture transcriptions
- Maps
- Methodologies & Workflows
- MOOCs
- Newsletters
- Oral History
- Physical artifacts and specimens
- Point clouds
- Posters
- Presentations
- Seismic recordings
- Software
- Spreadsheets / CSV files
- Surveys
- Technical reports
- Teaching tools designed by faculty
- Theses
- Transcripts
- Video
- Visualizations
- Websites
- White papers

What is Research Data?



Data Management

vs.

Data Management Plans



Data Management vs DMPs

- Data management refers to the things researchers do as they create, collect, describe, store, and work with data generated by their research.
- Data Management Plans (DMP) are a written description of what data is generated or otherwise acquired for a particular research project, and how the data will be utilized and stored during and after a project.

Data Management vs DMPs

- Data management is what you *do*.
- Data Management Plans are where you write about what you're going to do. (Usually because a grant requires one.)

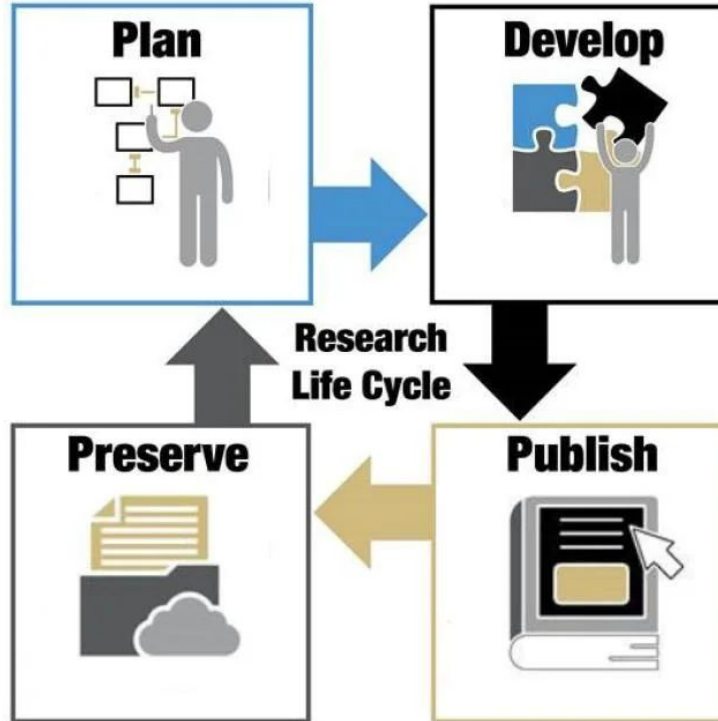
Why is data management important?

- Protect data from loss
- Saves you time
- You can find your data when you need it
- Helps new members of research teams understand processes faster
- Facilitates reproducibility
- Improves the quality of published data
- Keep sensitive data secure



afrengel

Research Lifecycle



Data Storage & Access



Data Storage & Access

- Where will the data be stored?
 - Data being collected
 - Data being analyzed
- Who can access the data?
 - How can they access it?
 - Is there sensitive or confidential data?
 - What security measures are in place to protect the data?

Things to think about

- Who's paying for data storage?
- How long will you have access to this storage?
 - What happens when you graduate?

Big Data

- Larger data sets means increased complexity!
- It's harder to store and provide access to data when you're working with terabytes or petabytes of data

Backups & Versioning



Why backup your data?

- Technology failure
- Natural disasters
- Theft
- Human error

3-2-1 Rule

- Three copies
 - One primary and two backups
- Two formats/media
 - e.g. External hard drive & cloud storage
- One off site
 - Where is your cloud storage located?

USBs are *Not* Data Storage

- You will lose them



Versioning

- Versioning is when you save specific versions of your files
- Some software does this automatically
- Can be as basic as having “raw” and “cleaned” versions

Version history

All versions ▾

TODAY

▶ May 13, 8:46 AM ⋮
Current version
● Matthew Murray

YESTERDAY

▶ May 12, 11:07 PM
● Matthew Murray

▶ May 12, 5:35 PM
● Matthew Murray

▶ May 12, 1:39 PM
● Matthew Murray

May 12, 10:50 AM
● Matthew Murray

Documenting Data



What is Documentation?

- Documentation is capturing your research process (from data collection through analysis)
- Includes the Five Ws (and One H)
 - What was done, Who did it, When it happened, Where it happened, Why it was done, and How it was done
- Many different ways to do this

Why is it Important?

- You can keep track of what still needs to be done.
- Your future self will thank you for writing down what you did.
- Allows others to understand your process and replicate your work.

Documentation matters

- HIT_AND_RUN_I
- STATEMENTS_TAKEN_I

Examples of Documentation

- Data Dictionaries
 - Contains a description of elements in a dataset, including names, definitions, acronyms, and other relevant information.
- README files
 - A plain text file that provides information and instructions about a project, including its purpose, usage instructions, known issues, and contact information for support or collaboration.

README Files

- Title of Dataset
- Authors
- Contact information
- Date of data collection
- Licenses/restrictions placed on the data
- Links to publications that cite or use the data
- Recommended citation for the data
- Structure and organization of the data files
- Descriptions of variables
- List of software and version numbers

Metadata

- Data about data!
- Descriptions that help you find and understand data
- Different fields/disciplines use different metadata standards

Creator

Gifford, Lauren
Nacu-Schmidt, Ami
Osborne-Gowey, Jeremiah
Boykoff, Max

Date Issued

2023-04

Academic Affiliation

Cooperative Institute for Research in Environmental Sciences

Last Modified

2023-05-02


Resource Type

Data Set

Rights Statement

In Copyright 

DOI

 <https://doi.org/10.25810/c862-0e81.60>

Language

English [eng] 

License

Creative Commons BY Attribution 4.0 International 

File Names & Structures



"FINAL".doc



FINAL.doc!



FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc

JORGE CHAN © 2012

WWW.PHDCOMICS.COM

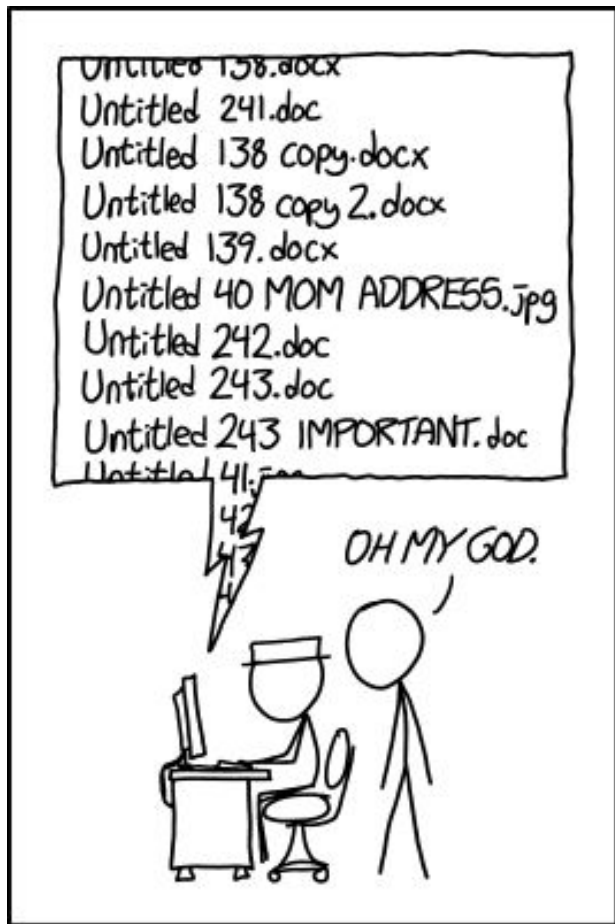
PHD Comics:

NotFinal.Doc. Jorge Cham, 2012.

<https://phdcomics.com/comics.php?f=1531> .

xkcd: Documents.
Randall Munroe.

<https://xkcd.com/1459/>



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Bad File Names

- Traffic_Crashes_-_Crashes.csv
- “Traffic Crashes - Crashes.”
data.cityofchicago.org, 2023-05-04.
<https://catalog.data.gov/dataset/traffic-crashes-crashes>.

File Naming Best Practices

- Be consistent
 - Do: Use the same format for all files
 - Don't: Keep changing file names
- Be descriptive
 - Do: Avoid generic terms
 - Don't: Use “Final” in your file name
- Limit file name length
 - Do: Use abbreviations
 - Don't: write-out-every-word-in-your-data-file.xlsx

File Naming Best Practices

- Use CamelCase and dashes
 - Do: FileName-2023.pdf
 - Don't: use spaces in your filenames.doc
- Use standardized numbers and dates
 - Do: Use leading zeros (001.png)
 - Don't: Have files named 9-12-11.csv
- Use the Latin alphabet
 - Don't: Use punctuation or special characters

Dates

- ISO 8601
- 2023-05-13
- Remember that historically dates were not consistent
 - 1712-02-30 (February 30th, 1712) was a real date in Sweden



CU Boulder Brand Do...

Zotero

Boulder FL master

Boulder FL vertical B...

Boulder one line

Boulder centere...

Boulder FL vertical B...

Boulder pdfs

Complete_w...

MMurray-D... Request For...

Boulder B&W all_plays_bo... - Copy

Boulder centere...

Boulder FL VERT A - ...

Boulder FL vertical B ...

Boulder one line reverse...

Complete_w... - Copy

MMurray-D... Request For...

Boulder B&W all_plays_bo... - Copy (2)

all_plays_bo... - Copy (2)

Recycle Bin

My Open Tickets

Boulder FL vertical A

Boulder one line grayscale

BoulderEm...

Boulder FL master 4525

Boulder one line reverse

all_plays_bo...

Complete_w...

Real_Estate...

Boulder full color - Copy

all_plays_bo... (1) - Copy

Boulder FL - Copy

Boulder FL VERT B - ...

Boulder FL vertical B...

Boulder template ...

Complete_w... - Copy

Real_Estate... - Copy

Boulder full color - Co...

all_plays_bo... (2) - Copy

Firefox

Boulder centere...

Boulder FL vertical B

Boulder centere...

BoulderEm...

Boulder FL vertical ...

Boulder PMS 4525

all_plays_bo... (1)

cu_boulder_...

TFLS-003-Pr...

Boulder grayscale...

BEN_review... - Copy

Boulder FL - Copy

Boulder FL VERT B - ...

Boulder FL vertical B...

Boulder template ...

cu_boulder_... - Copy

TFLS-003-Pr... - Copy

Boulder grayscale...

BEN_review... - Copy (2)

Google Chrome

Boulder FL master B&W

Boulder one line

Boulder FL grayscale

Boulder one line

Boulder FL vertical B...

Boulder Reverse

BEN_review...

cu-boulder_...

Traffic_Cras...

Boulder jpegs - Copy

Boulder centered...

Boulder FL grayscale...

Boulder FL vertical A...

Boulder one line - Copy

Boulder template ...

cu_boulder_... - Copy

Traffic_Cras... - Copy

Boulder jpegs - Co...

Boulder centered...

Get Help

Boulder FL vertical ...

Boulder-on...

Boulder centered

Boulder centere...

Boulder one line 4525

Boulder B&W Boulder template...

Boulder template...

Electric_Veh...

usa_tv

Boulder pdfs - Copy

Boulder centered...

Boulder FL master - ...

Boulder FL vertical A ...

Boulder one line - Copy

BoulderEm... - Copy

Electric_Veh... - Copy

usa_tv - Copy

Boulder pdfs - Copy (2)

Boulder centered...

Interlibrary Loan

Sierra Desktop App

.DS_Store

Boulder FL VERT A

Boulder FL VERT A

Boulder centere...

Boulder full color

Boulder template...

iou_zipcod...

user_profiles

Boulder PMS 4525 - Copy

Boulder centered 4...

Boulder FL master 4525...

Boulder FL vertical A...

Boulder one line - Copy

BoulderEm... - Copy

iou_zipcod... - Copy

user_profiles - Copy

Boulder PMS 4525 - Cop...

Boulder centered 4...

Libraries Intranet

Boulder FL vertical B...

Boulder centered

Boulder FL VERT B

Boulder FL VERT B

Boulder FL master r...

Boulder grayscale

Boulder template...

loan_featur...

W-9

Boulder Reverse...

Boulder centered c...

Boulder FL master r...

Boulder FL vertical ...

Boulder one line 4525 ...

Boulder-on... - Copy

loan_featur... - Copy

W-9 - Copy

Boulder Reverse...

Boulder centered c...

Zoom

Boulder one line B&W

Boulder FL vertical ...

Boulder FL

Boulder FL

Boulder FL vertical ...

Boulder jpegs

Bridge_Con...

loan_featur...

Zotero-6.0...

- Copy.DS_St...

Boulder centered ...

Boulder FL VERT A - ...

Boulder FL vertical B...

Boulder one line grayca...

Bridge_Con... - Copy

loan_featur... - Copy

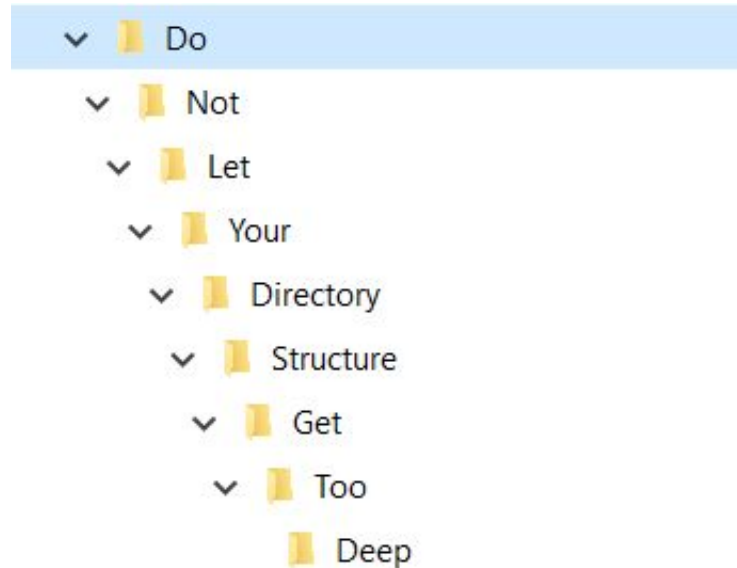
Zotero-6.0.2... - Copy

(2).DS_Store

Boulder centered ...

File Structuring

- Don't save everything to the desktop
- Don't let your structure get too deep



File Formats & Units of Measurement



File Formats

- Non-proprietary/Open
 - Can still be used even if original software is inaccessible
 - e.g. Use CSV files instead of .vc
- Unencrypted
- Uncompressed/lossless

Units of Measurement

- Be consistent
- Use standardized measurements
- Don't keep changing between them
- How big is a “cup”?
 - US Cup: 8 oz (236.6 ml)
 - Metric cup: 250 ml
 - Japanese rice cup: 180 ml

Data Preservation



What is Data Preservation?

- What you do with your data after the research project is over
- Not all data has to be preserved!
 - What data should be preserved?
 - What data can be gotten rid of? When do you get rid of it?
- All the stuff we talked about before (storage, access, metadata, etc.)

Data Management Plans



What is a DMP?

- A framework for how you'll manage your data
- Describe your plans for collecting, organizing, storing, and sharing your data
- About two pages long

When should you make a DMP?

- When you have to for grant requirements
 - Different grants have different requirements
- When you don't have to, but want to, keep track of your data

Two types of DMPs

- The “final” version you’ll submit with a grant proposal
- The “living” version that you’ll continue to update as your research project progresses

What goes in a DMP?

- Data generated in this project
- Software and file formats that will be used
- Where data will be stored and who can access it
- Any privacy, legal, or ethical constraints
- Metadata standards
- How the data will be preserved/shared
- How the data can be reused

DMP Tool



DMPTool

- Has pre-formatted DMP templates from various funding agencies
- Walks you through the process of completing the DMP

DMPTool Activity

- Go to <https://dmptool.org/>
- When signing in, indicate that you're from CU Boulder
- Select a DMP Template relevant for your field (if you're having trouble choosing, pick the generic NSF template)
- Look through the prompts
- Attempt to answer the prompts in the context of your proposed project

DMPTool Activity

- Were any of the prompts challenging or confusing?
- Is there anything you would like clarification on?
- We're happy to read drafts and provide feedback on your draft DMPs, so please send them our way!

Other Useful Tools



Data Management Tools

- Git and GitHub
 - Track changes to data and code
- Open Science Framework
 - One-stop shop for project management
- Open Refine
 - Clean data
- File Renamers (various)
 - Rename files so they're consistent
- Zotero
 - Reference management software

Consultations



How CRDDS can assist with Data Management

- One-on-one or small group consults
- Review draft DMPs and README files
- Help navigate data policies
- Find data repositories
- Advice on file formats, etc.
- Email us: crdds@colorado.edu