

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- Historical documents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

(and Data Management and Sharing 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 (DMPs) 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

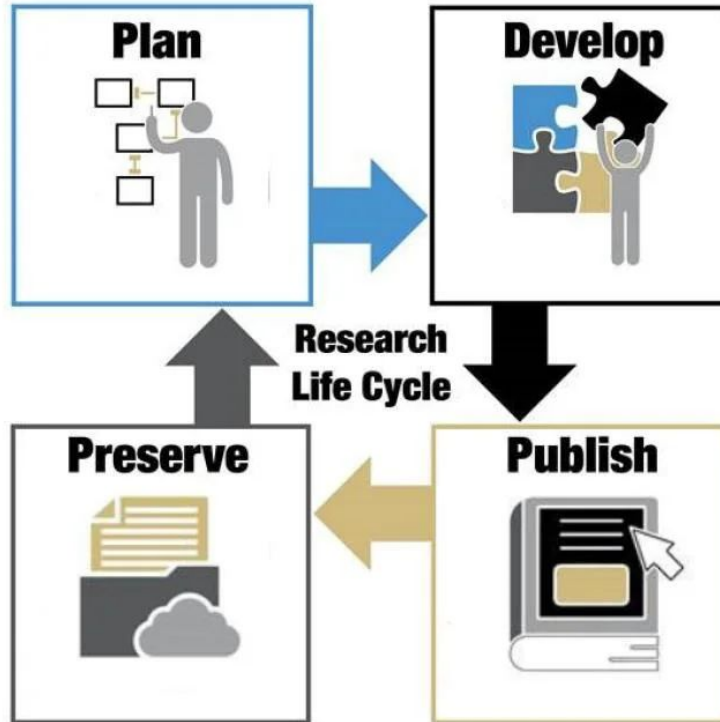
- Data management is what you *do*.
- Data Management Plans are where you write about what you're going to do. (Usually because a grant requires one.)

Why is data management important?

- Protects 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



Research Lifecycle



What Data Management isn't

- Data management is **not** data sharing
- You can manage your data without sharing it
- You can share your data without managing it
 - (please don't do this)

Data Storage & Access



Data Storage & Access

- Where will the data be stored?
 - When the data is being collected
 - When the data is being analyzed
 - After the project is over
- 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?

 r/cuboulder · 1y ago

I lost several years of DnD world-building with the google email change, looking for help

3 votes · 8 comments

 r/cuboulder · 2mo ago

Has anyone had any success trying to have their Google school account recovered? Or am I screwed?

2 votes · 6 comments

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
- Rogue AIs

All of your data has been deleted.

Pod 042

All of your data has been deleted.



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?

3-2-1 Rule

- Not always feasible
 - Big data (terabytes or petabytes of data)
 - Frequently changing data
 - Data collected in places with terrible internet
- Ask yourself:
 - How often do you need to backup your data?
 - If you lost a specific piece of data could your research project continue?

USBs are **Not** Data Storage

- You will lose them



Versioning

- Versioning is when you save specific versions of your files
- Some software (Git!) does this automatically
- Can be as basic as having “raw” and “cleaned” versions
- Be consistent in giving version numbers

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
- There are many different ways to do this

Why is Documentation 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

Examples of Documentation

- Data Dictionaries & Codebooks
 - Contain 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.

Codebooks

- Define the variables and their units
- Explain the formats for dates, time, geographic coordinates
- Define any coded values and missing values
- Allows others outside of your research group to understand the data

Why is it Important?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	AB	AC	AD
1	79c7a2ce89	JC199149		03/25/201	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT TURNING ONE-WAY	4	STRAIGHT DRY			NO DEFEC ON SCENE INJURY ANY								OVER \$1,503/25/201 IMPROPEI DRIVING				121		
2	792b539dee	JB422857		09/05/201	30	NO CONF# NO CONF# CLEAR		DAYLIGHT ANGLE NOT DIVIE	2	STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,509/05/201 VISION OF FAILING				1712		
3	0115ade9a7	JB318029		07/15/202	30	UNKNOWN# UNKNOWN# CLEAR		DARKNESS# ANGLE NOT DIVIDED		STRAIGHT DRY			UNKNOWN# ON SCENE NO INJURY / DRIVE AWAY								OVER \$1,507/15/202 UNABLE T UNABLE				1123		
4	017040c619	JB324552		07/15/202	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT REAR ENC NOT DIVIDED		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,507/20/202 UNABLE T UNABLE				1811		
5	78eee027ec	JB291672		06/03/201	30	NO CONF# NO CONF# CLEAR		UNKNOWN# PARKED N ONE-WAY	1	STRAIGHT DRY			NO DEFEC ON SCENE NO INJURY / DRIVE AWAY							Y	OVER \$1,506/03/201 UNABLE T UNABLE				2534		
6	7943cacbae	JB330061		07/24/202	25	NO CONF# NO CONF# CLEAR		DAYLIGHT PARKED N ONE-WAY		STRAIGHT DRY			NO DEFEC ON SCENE NO INJURY / DRIVE AWAY								OVER \$1,507/24/202 NOT APPL NOT APP				823		
7	01aaa759c6	JB319819		07/15/202	40	NO CONF# NO CONF# CLOUDY	C	DAYLIGHT ANGLE DIVIDED - W/MEDIA		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,507/16/202 IMPROPEI NOT APP				2024		
8	7b1537e0a2	JA252488		05/05/201	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT SIDESWIPI NOT DIVIDED		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								\$501 - \$1,05/05/201 UNABLE T UNABLE				1731		
9	7b4700df1a	JB529385		11/26/201	30	TRAFFIC SI FUNCTION SNOW		DAYLIGHT ANGLE NOT DIVIE	2	STRAIGHT SNOW OR			NO DEFECTS NO INJURY								OVER \$1,506/26/201 WEATHER UNABLE				412		
10	7b3545fb91	JD298172		07/15/202	25	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT PEDALCYC NOT DIVIDED		STRAIGHT DRY			NO DEFEC ON SCENE INJURY ANY								\$500 OR L 07/15/202 DISREGAR DRIVING				1533		
11	79cac9c36c	JA441727		09/13/201	30	NO CONF# NO CONF# CLEAR		DAYLIGHT PARKED N PARKING LOT		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY							Y	OVER \$1,509/22/201 UNABLE T UNABLE				1823		
12	01d4adbecb	JB399117		09/16/202	35	NO CONF# NO CONF# CLEAR		DARKNESS# REAR TO FALLEY		STRAIGHT DRY			NO DEFEC NOT ON SI INJURY AND / OR TOW DUE TO Y							Y	OVER \$1,509/17/202 UNABLE T UNABLE				1832		
13	0362b3f8dd	JB318271		07/15/202	30	NO CONF# NO CONF# RAIN		DAYLIGHT REAR ENC DIVIDED - W/MEDIA		STRAIGHT WET			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,507/15/202 UNABLE T UNABLE				1634		
14	03d3679ef4	JB318956		07/15/202	30	STOP SIGN FUNCTION RAIN		DAYLIGHT PEDESTRI# FOUR WAY		STRAIGHT WET			NO DEFEC ON SCENE INJURY ANY							Y	\$500 OR L 07/15/202 FAILING T UNABLE				1235		
15	049099b67e	JB318460		07/15/202	35	TRAFFIC SI NOT FUNC RAIN		DAYLIGHT TURNING NOT DIVIDED		STRAIGHT WET			UNKNOWN# ON SCENE NO INJURY / DRIVE AWAY								OVER \$1,507/15/202 IMPROPEI DRIVING				815		
16	04ca492845	JB319187		07/15/202	30	NO CONF# NO CONF# RAIN		DAYLIGHT PARKED N NOT DIVIDED		STRAIGHT WET			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								\$500 OR L 07/16/202 NOT APPL NOT APP			Y	1103		
17	051542719f	JB319114		07/15/202	30	UNKNOWN# UNKNOWN# RAIN		DARKNESS# REAR TO F DIVIDED - W/MEDIA		STRAIGHT WET			UNKNOWN# ON SCENE INJURY AND / OR TOW DUE TO CRASH								OVER \$1,507/15/202 NOT APPL NOT APP				1834		
18	7b1c666d9d	JB163972		02/24/201	25	NO CONF# NO CONF# CLEAR		DAYLIGHT SIDESWIPI DIVIDED -	4	STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY							N	\$501 - \$1,02/24/201 UNABLE T UNABLE				123		
19	79704e1b74	JC315780		06/21/201	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT TURNING DIVIDED - W/MEDIA		STRAIGHT DRY			NO DEFEC ON SCENE INJURY AND / OR TOW DUE TO Y							Y	OVER \$1,506/21/201 FOLLOWIN# UNABLE				123		
20	7994103871	JA425956		09/08/201	30	STOP SIGN FUNCTION CLEAR		DAYLIGHT REAR ENC NOT DIVIDED		STRAIGHT DRY			NO DEFECTS NO INJURY							Y	OVER \$1,509/10/201 FOLLOWIN# NOT APP				2413		
21	7b233fb34d	JC174024		03/05/201	25	STOP SIGN FUNCTION CLEAR		DAYLIGHT REAR ENC NOT DIVIE	4	STRAIGHT SNOW OR			NO DEFEC ON SCENE NO INJURY N								\$500 OR L 03/05/201 NOT APPL UNABLE				2515		
22	7b3850c20c	JC233284		04/22/201	30	NO CONF# UNKNOWN# RAIN		DARKNESS# FIXED OBJ DIVIDED - W/MEDIA		STRAIGHT DRY			NO DEFEC ON SCENE INJURY AND / OR TOW DUE TO CRASH								OVER \$1,504/22/201 UNDER TH UNABLE				1231		
23	054e9bad1e	JB318760		07/15/202	30	NO CONF# NO CONF# RAIN		DAYLIGHT PARKED N NOT DIVIDED		STRAIGHT WET			NO DEFEC ON SCENE NO INJURY / DRIVE AWAY							Y	OVER \$1,507/15/202 UNABLE T UNABLE				411		
24	056c477498	JB318680		07/15/202	30	NO CONF# NO CONF# RAIN		DAYLIGHT REAR ENC NOT DIVIDED		STRAIGHT WET			UNKNOWN# ON SCENE NO INJURY / DRIVE AWAY							Y	OVER \$1,507/15/202 FOLLOWIN# UNABLE				1033		
25	05c1521b06	JB318600		07/15/202	30	UNKNOWN# UNKNOWN# CLEAR		DAYLIGHT ANGLE T-INTERSECTION		STRAIGHT DRY			UNKNOWN# NOT ON SI NO INJURY / DRIVE AWAY								\$501 - \$1,07/15/202 NOT APPL NOT APP				1235		
26	05c296dfb5	JB318941		07/15/202	30	UNKNOWN# UNKNOWN# FREEZING		DAYLIGHT FIXED OBJ FOUR WAY		STRAIGHT WET			NO DEFEC ON SCENE INJURY ANY								OVER \$1,507/15/202 UNABLE T NOT APP				2533		
27	7951cf0b6a	JA562045		12/24/201	30	UNKNOWN# UNKNOWN# SNOW		DAWN REAR ENC NOT DIVIE	2	STRAIGHT SNOW OR			UNKNOWN# NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,512/24/201 TURNING NOT APP				2521		
28	7a08f16969	JB233086		05/14/202	30	NO CONF# NO CONF# CLEAR		DAYLIGHT PARKED N NOT DIVIDED		STRAIGHT UNKNOWN#			UNKNOWN# NOT ON SI NO INJURY / DRIVE AWAY							Y	OVER \$1,505/17/202 UNABLE T UNABLE				1131		
29	7be477990c	JB506955		11/07/201	30	NO CONF# NO CONF# CLEAR		DAYLIGHT PARKED N NOT DIVIE	2	STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY							Y	OVER \$1,511/07/201 UNABLE T UNABLE				1935		
30	79f6316976	JE122789		01/25/202	10	NO CONF# NO CONF# CLEAR		DAYLIGHT REAR TO F ONE-WAY		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,501/25/202 IMPROPEI UNABLE				2211		
31	07baf8ac46	JB318925		07/15/202	30	NO CONF# NO CONF# RAIN		DAYLIGHT OTHER NC NOT DIVIDED		STRAIGHT WET			OTHER ON SCENE INJURY AND / OR TOW DUE TO CRASH								\$501 - \$1,07/15/202 UNABLE T NOT APP				1224		
32	094ef43765	JB318676		07/15/202	30	TRAFFIC SI FUNCTION RAIN		DAYLIGHT TURNING DIVIDED - W/MEDIA		STRAIGHT WET			UNKNOWN# NOT ON SI NO INJURY								OVER \$1,507/15/202 FAILING T FAILING				2531		
33	7b63a439b7	JB366967		07/27/201	30	TRAFFIC SI FUNCTION CLEAR		DARKNESS# PEDALCYC CENTER TI	5	STRAIGHT UNKNOWN#			UNKNOWN# ON SCENE NO INJURY / DRIVE AWAY							Y	\$500 OR L 07/27/201 FOLLOWIN# FOLLOW				1434		
34	0a43552917	JB318249		07/15/202	30	STOP SIGN FUNCTION RAIN		DAYLIGHT REAR ENC NOT DIVIDED		STRAIGHT WET			UNKNOWN# ON SCENE NO INJURY								OVER \$1,507/15/202 FOLLOWIN# UNABLE				1411		
35	0be1398662	JB319154	N	07/15/202	30	NO CONF# NO CONF# RAIN		DARKNESS# TURNING NOT DIVIDED		STRAIGHT WET			NO DEFEC ON SCENE INJURY AND / OR TOW DUE TO Y								\$501 - \$1,07/15/202 FAILING T NOT APP				1524		
36	7a77bed55c	JE267857		06/16/202	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT REAR ENC DIVIDED - W/MEDIA		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY								OVER \$1,506/16/202 UNABLE T UNABLE				725		
37	0da8a78c1a	JB318221	Y	07/15/202	30	NO CONF# NO CONF# CLEAR		UNKNOWN# PARKED N NOT DIVIDED		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY / DRIVE AWAY							Y	OVER \$1,507/15/202 UNABLE T NOT APP				1611		
38	7a98b33a77	JB311906		06/16/201	30	TRAFFIC SI FUNCTION CLEAR		DAYLIGHT TURNING NOT DIVIDED		STRAIGHT DRY			NO DEFEC NOT ON SI NO INJURY								\$501 - \$1,06/18/201 FOLLOWIN# UNABLE				2022		

Documentation matters

- HIT_AND_RUN_I
- INTERSECTION_RELATED_I
- STATEMENTS_TAKEN_I

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
- List of software (with version numbers) and instruments

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!



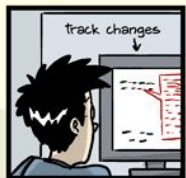
FINAL_rev.2.doc



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



PHD Comics:

NotFinal.Doc. Jorge

Cham, 2012.

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

Untitled 158.docx
Untitled 241.doc
Untitled 138 copy.docx
Untitled 138 copy 2.docx
Untitled 139.docx
Untitled 40 MOM ADDRESS.jpg
Untitled 242.doc
Untitled 243.doc
Untitled 243 IMPORTANT.doc
Untitled 41.doc



xkcd: Documents.

Randall Munroe.

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

PROTIP: NEVER LOOK IN SOMEONE
ELSE'S DOCUMENTS FOLDER.

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

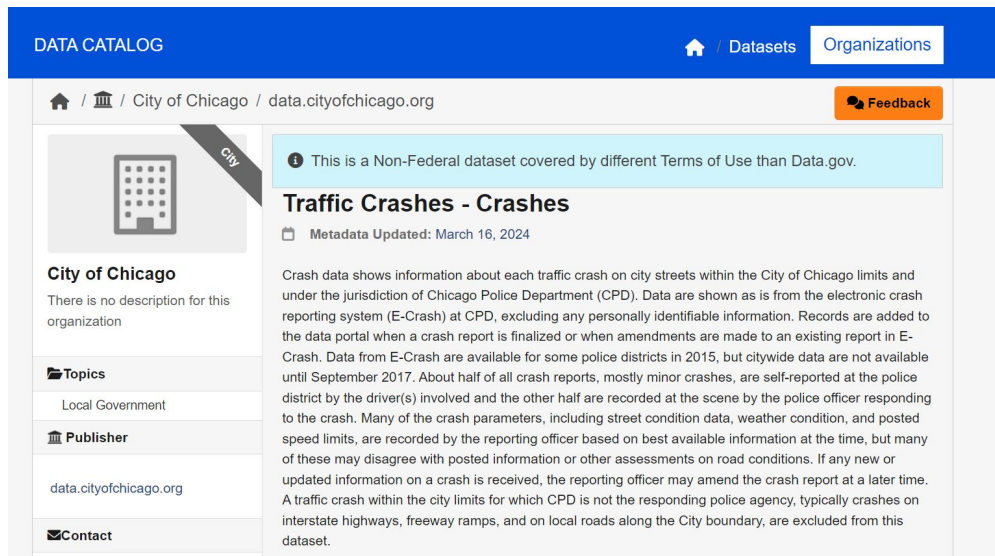
Abbreviations

- Use meaningful abbreviations
- Document your decisions
 - Don't assume you'll remember what an abbreviation means
- Have group & individual identifiers
- Use version numbers

File Naming Best Practices

- Use CamelCase (not all systems preserve case)
 - Do: FileName-2023.pdf
 - Don't: use spaces in your file names.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
 - - and _ are okay!

Don't use spaces in your file names



Traffic_Crashes_-_Crashes.csv

- This also breaks certain software (or makes it harder to use)

<https://catalog.data.gov/dataset/traffic-crashes-crashes>

Dates

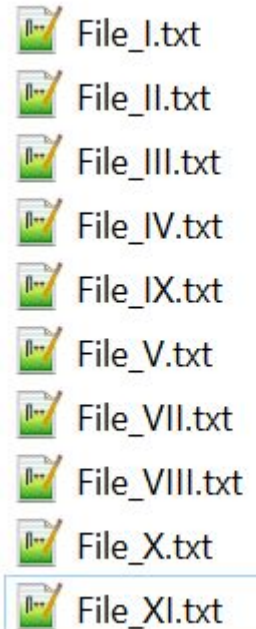
- Don't use terms like “Quarter 1” or “spring”
- ISO 8601
- 2023-10-31 or 20231031 (YYYY-MM-DD)
- Remember that historically dates were not consistent
 - 1712-02-30 (February 30th, 1712) was a real date in Sweden

Use the Latin alphabet

- This sucks
- I'm sorry
- If your data features non-latin characters reach out to us and we can help make sure it will work
- When you use them, make sure they're encoded properly

Don't use Latin numerals

- Don't use Latin numerals to name your files (they won't sort properly)
- You can use them in documentation (sometimes)



235

Chapter 17: What About Crowdfunded Comics?

Matthew Murray 🦋 and Mara L. Thacker

Part 4. How Can Comics Be Used as Primary and Secondary Source Material by Students and Faculty?

Chapter 17: What About Crowdfunded Comics?

Matthew Murray 🦋 and Mara L. Thacker

Additional file 4: of The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch

[Cite](#)[Download \(132.15 kB\)](#)[Share](#)[Embed](#)[+ Collect](#)

Journal Contribution posted on 2018-09-18, 23:00 authored by Alice Michel, Lewis Ward, Shana Goffredi, Katherine Dawson, Daniel Baldassarre, Alec Brenner, Kiyoko Gotanda, John McCormack, Sean Mullin, Ariel O'Neil, Gabrielle Tender, J. Uy, Kristie Yu, Victoria Orphan, Jaime Chaves

Additional information regarding the observations of finch feeding behavior, core microbiome results, and microbiome differences between vampire finches on Darwin and Wolf islands. (DOCX 20 kb)

FUNDING

Gordon and Betty Moore Foundation (US)

HISTORY

- 2018-09-18 - Posted date
- 2018-09-19 - First online date

RELATED MATERIALS

- DOI - Is supplement to [The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch](#)

 [View PDF](#)

USAGE METRICS

118
views

10
downloads

0
citations

Is supplement to

[The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch](#)

SPRINGER NATURE

CATEGORIES

- Evolutionary Biology
- Environmental Sciences not elsewhere classified
- Ecology
- Biological Sciences not elsewhere classified

KEYWORDS

Galápagos Islands

Darwin's finches

Blood-feeding

Microbiome

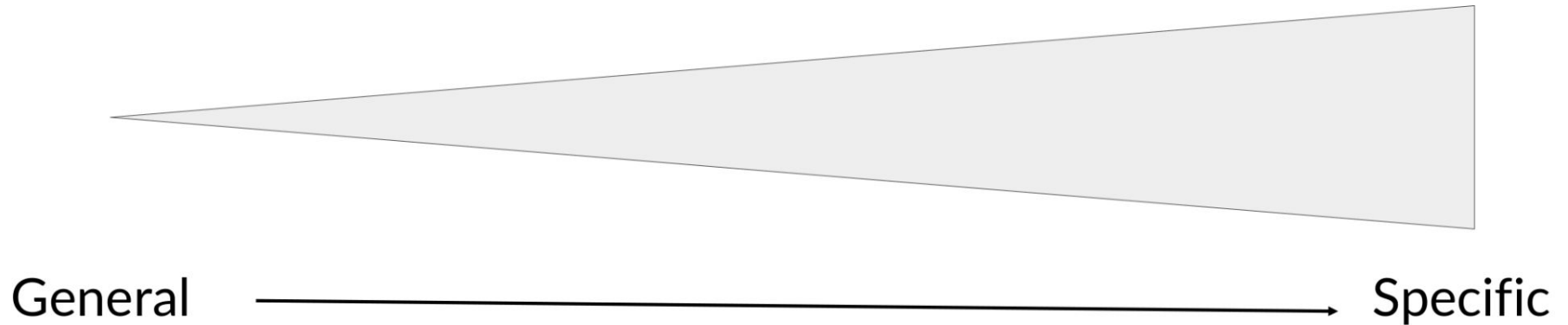
Geospiza

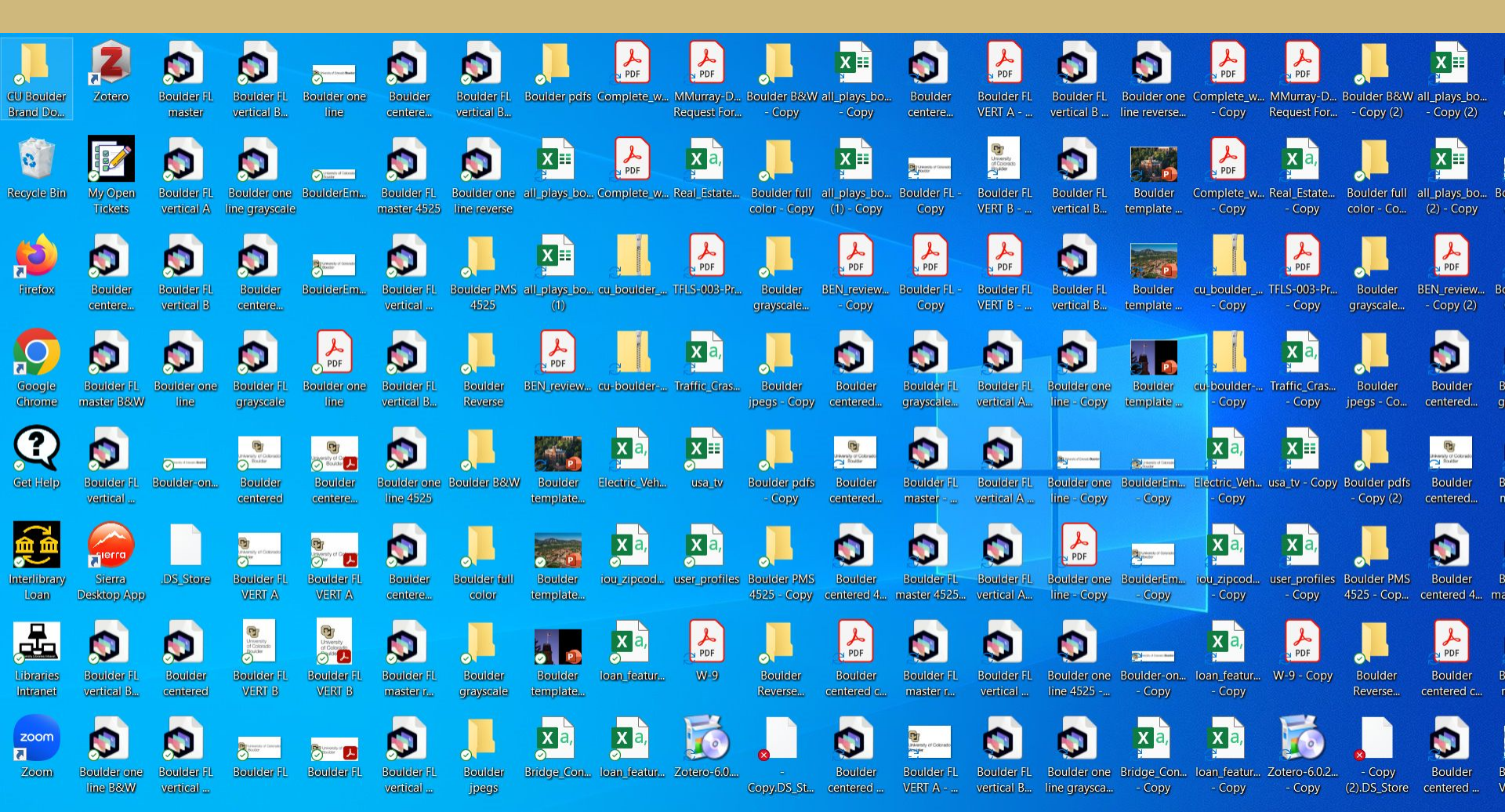
LICENCE

CC BY + CC0

20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx

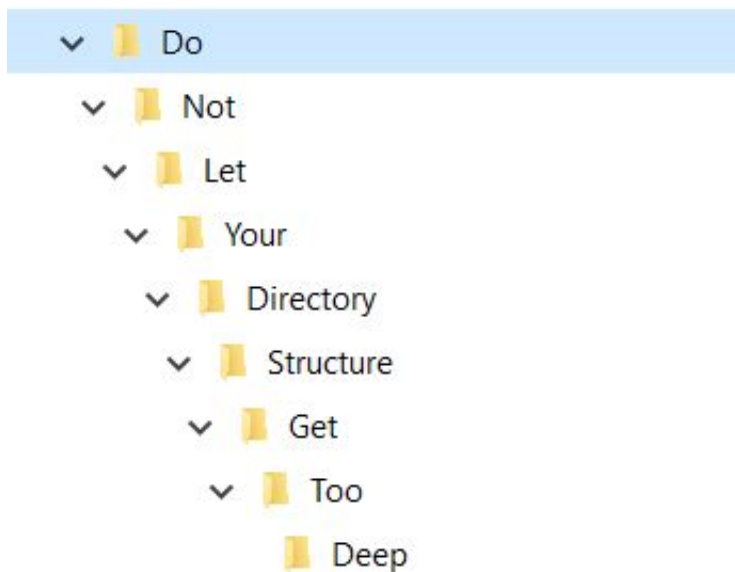
Date Project Experiment Type ID Version






File Structuring

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



Project directory structure

Project_1

- methods
- raw_data 
 - readme
- analysis
 - analysis_method_1
 - 2017
 - 2018
 - analysis_method_2
- scripts
- manuscript
 - text
 - version_1
- readme and/or ELN link

Always keep
your raw
data!

(Raw data should
be “read-only” if
possible.)

File Archiving and Compression



- Usually called “zipping”
- Means collecting different files and directories into one file
- Many programs/methods of doing this
 - .zip, .rar., .7z, .tar, .gz, etc.
- Done to minimize size of files (compression)
- Done to collect many files and directories into one file (archiving)
- Makes it easier to share many files

File Archiving and Compression

- Consider how the files will look once they're extracted (unzipped).
- Create a directory structure that will make it easy for people to understand and use your files.
- Ensure all file paths will allow associated code to work.

MacOS

- __MACOSX, .DS_Store, & ._ files are created automatically in MacOS
- They're invisible on Mac, but visible on Windows and Linux systems
- They will be included in .zip files

Name	Status	Date modified	Type	Size
 __MACOSX		2023-12-11 2:42 PM	File folder	
 .DS_Store		2023-12-11 2:42 PM	DS_STORE File	7 KB

1,018,676 items | 509,338 items selected



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

File Formats

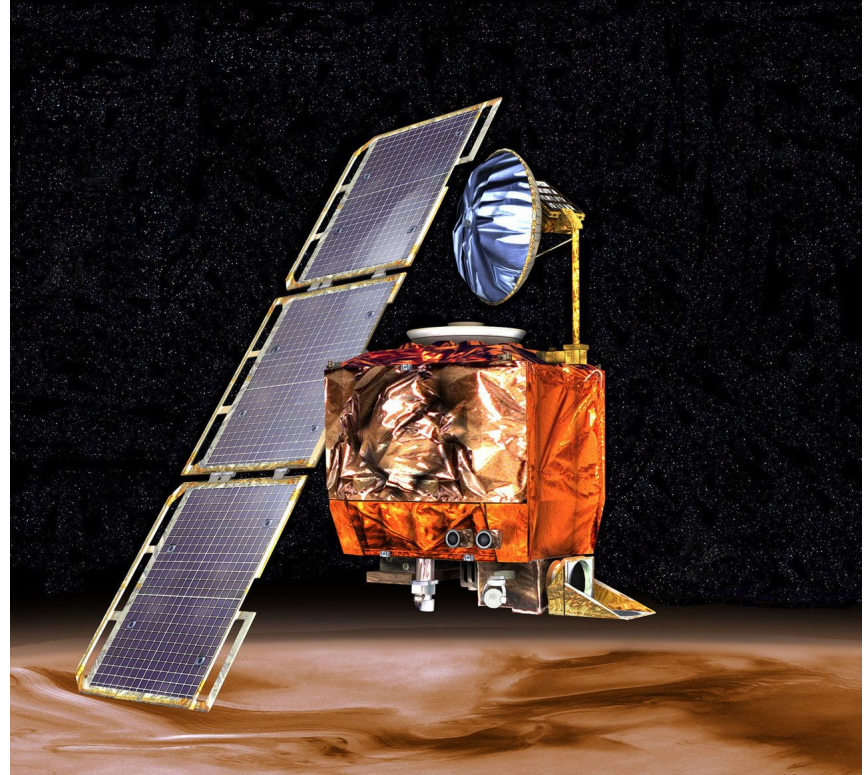
- Use the default file extension
 - Do: Save Excel files as .xlsx
 - Don't: Save Excel files as .spreadsheet
- Yes, I have seen people use non-standard file extensions.

Units of Measurement

- Be consistent
- Use standardized measurements
- Don't keep changing between them
- Ensure everyone is using the same system
- Define them in your documentation

Height
3
6
1
4
6
7
4
10
12
1
12
2
3

Learn from the Mars Climate Orbiter



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?
- Consider all the stuff we talked about before (storage, access, documentation, etc.)

Computer Code

- GitHub is a collaboration tool (not a repository)
- To preserve code or software put it in a repository
- Connect your GitHub account to Zenodo

Published November 1, 2023 | Version 0.50.0

Software 

openforcefield/openff-qcsubmit: 0.50.0

JoshHorton; Josh Horton¹; SimonBoothroyd; trevorgokey; David L. Dotson²; Josh A. Mitchell³; Brent Westbrook³; Matt Thompson⁴; Jeff Wagner³; Chapin Cavender⁵; John Chodera⁶

Show affiliations

This release adds support for QCFractal's refactor and 0.50.0 release.

What's Changed

- Update versioneer by @Yoshanuikabundi in <https://github.com/openforcefield/openff-qcsubmit/pull/235>
- Changes to [openff-qcsubmit](#) to use [next](#) QCFractal by @dotsdl in <https://github.com/openforcefield/openff-qcsubmit/pull/195>



New Contributors

- @pre-commit-ci made their first contribution in <https://github.com/openforcefield/openff-qcsubmit/pull/234>


Full Changelog: <https://github.com/openforcefield/openff-qcsubmit/compare/0.5.0...0.50.0>

Files


openforcefield/openff-qcsubmit-0.50.0.zip ▾


 openforcefield/openff-qcsubmit-0.50.0.zip 

openforcefield-openff-qcsubmit-a968d44


 .gitattributes

40 Bytes


 github


 CONTRIBUTING.md

2.2 kB


 PULL_REQUEST_TEMPLATE.md

224 Bytes


 workflows

 CI.yml


1.8 kB

 Lint.yml

616 Bytes





 .gitignore

1.2 kB

 .lgtn.yml

332 Bytes

Files (325.3 kB) ▾

Name	Size	 Download all
openforcefield/openff-qcsubmit-0.50.0.zip md5:f37072b3cdcd55dc5741d89e01c2 	325.3 kB	 Preview  Download

129
👁 VIEWS

7
📄 DOWNLOADS

▸ Show more details

Versions

Version 0.50.0 10.5281/zenodo.10060049	Nov 1, 2023
Version 0.5.0 10.5281/zenodo.8253341	Aug 16, 2023
Version 0.4.1 10.5281/zenodo.7510744	Jan 6, 2023
Version 0.4.0 10.5281/zenodo.7320745	Nov 15, 2022
Version 0.3.2 10.5281/zenodo.7305244	Nov 8, 2022

[View all 9 versions](#)


Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.5503441](https://doi.org/10.5281/zenodo.5503441). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

External resources

Available in

 [openforcefield/openff-qcsubmit](#)
Release: 0.50.0

Indexed in

 [OpenAIRE](#)

Details

DOI
[DOI 10.5281/zenodo.10060049](https://doi.org/10.5281/zenodo.10060049)

Resource type
Software



Before you share your code...

- Remove irrelevant comments
- Ensure you're using relative paths and not absolute paths
 - */MMurray/i_hate_this_project/script.sh*

Data Management Plans & Data Management and Sharing Plans



What is a DMP/DMSP?

- 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?

- Types of data generated in this project
- Estimated size of data
- Software and file formats that will be used
- Where data will be stored, who can access it, and any security considerations
- Any privacy, legal, or ethical constraints
- Metadata standards
- How the data will be preserved/shared
- How the data can be reused
- A description of roles and responsibilities

DMPTool



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