Fixing Your Files: 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

You can’t do everything

- I’m going to talk about a lot of different things
- Doing *anything* is good
- Pick one or two things and concentrate on doing them well

What is Research Data?

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

Things we can’t agree on

- How it’s pronounced
  - “day-ta” vs. “dah-ta”
- Whether it’s singular or plural
  - “data is” or “data are”
- Different disciplines will view different things as data

What is Research Data?

- Primary
  - Data you gather yourself
- Secondary
  - Data someone else gathered
- Qualitative data
  - Non-numeric data (e.g. Open-ended survey responses)
- Quantitative
  - Data represented as numbers that can be counted (measurements)

Quantitative Data

- Measurements
  - Distances
  - Temperatures
  - Amount
  - Length of time
- Location
  - Longitude/Latitude
  - Altitude
  - Planet
- Date / Time

What is Research Data?

- Experimental
  - Data gathered through experiments where researchers control variables
- Observational
  - Data gathered from observations
- Individual data
  - Data from/about an individual person or location
- Generalized data
  - Aggregated, anonymized data

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- Historical documents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 means people

Regardless of what the data is, humans were involved with it somehow.

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

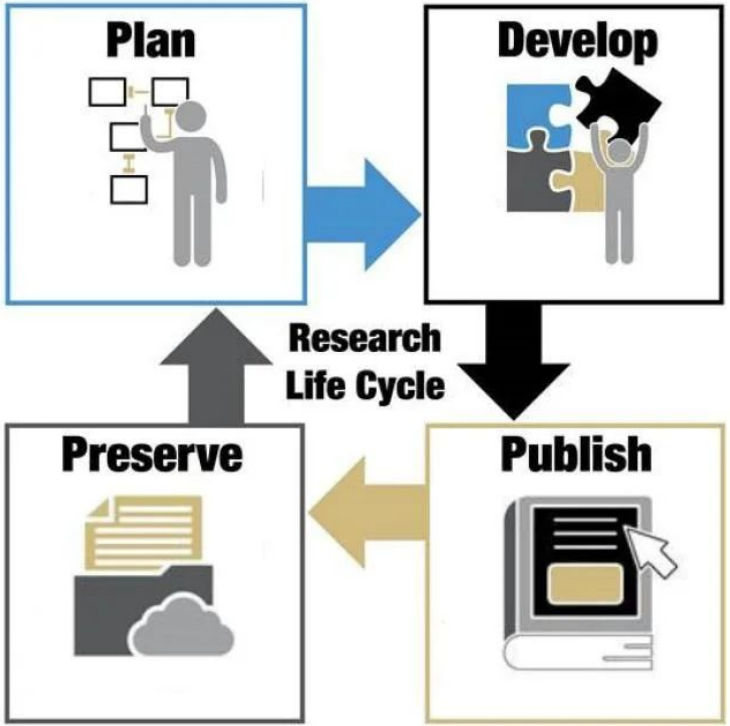
- Data management is what you ***do.***

- 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)

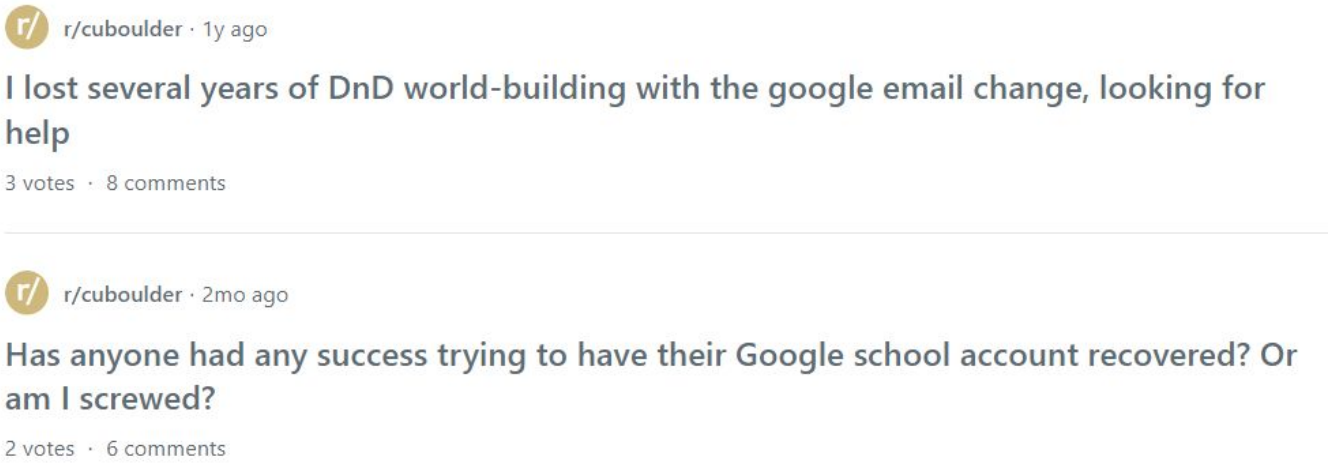
Data Storage & Access

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?



Content/Data requirements

Content/ Requirement	OneDrive	Teams	SharePoint*	UCB Files	PetaLibrary	Google**
User files that are stored or shared with others for collaboration	✓	✗	✗	✗	✗	✗
Data shared for unit or department use	✗	✓	✓	✓	✗	✗
Research data (less than a TB)	✓	✓	✓	✓	✓	✗

https://oit.colorado.edu/services/file-transfer-storage-infrastructure

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

Pod 042

All of your data has been deleted.

Why download & backup someone else's data?

CIA says it will cease publishing the World Factbook, a free online resource used by millions

By Andrew Thorpe | Internet

3h ago

How much government data was lost in a data center fire in South Korea?

Answer: 858 terabytes.

October 16, 2025 • News Staff

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

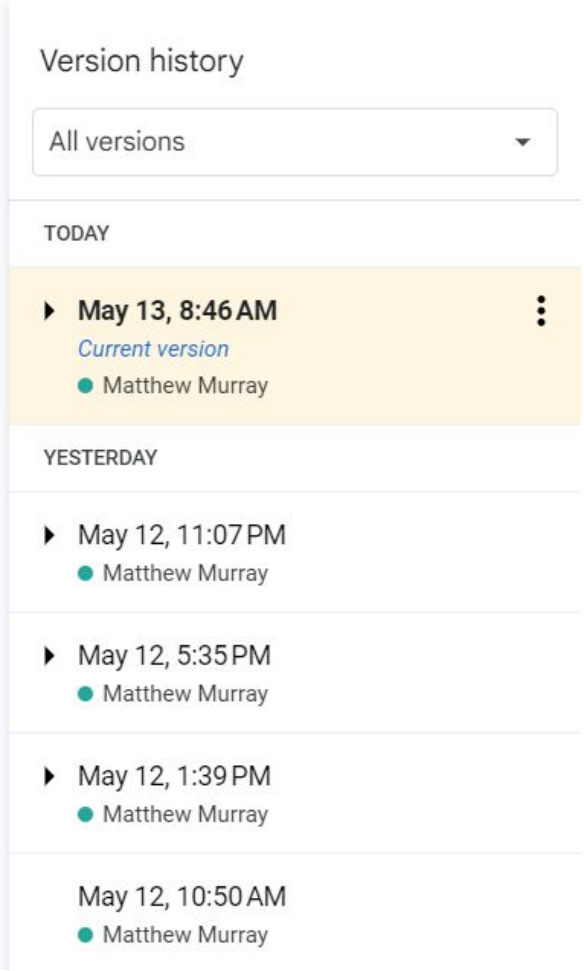
- Not always feasible
  - Big data (terabytes or petabytes of data)
  - Frequently changing data
  - Data collected in places with terrible internet
- Ask yourself:
  - How often do you need to backup your data?
  - If you lost a specific piece of data could your research project continue?

USBs are *Not* Data Storage

- You will lose them

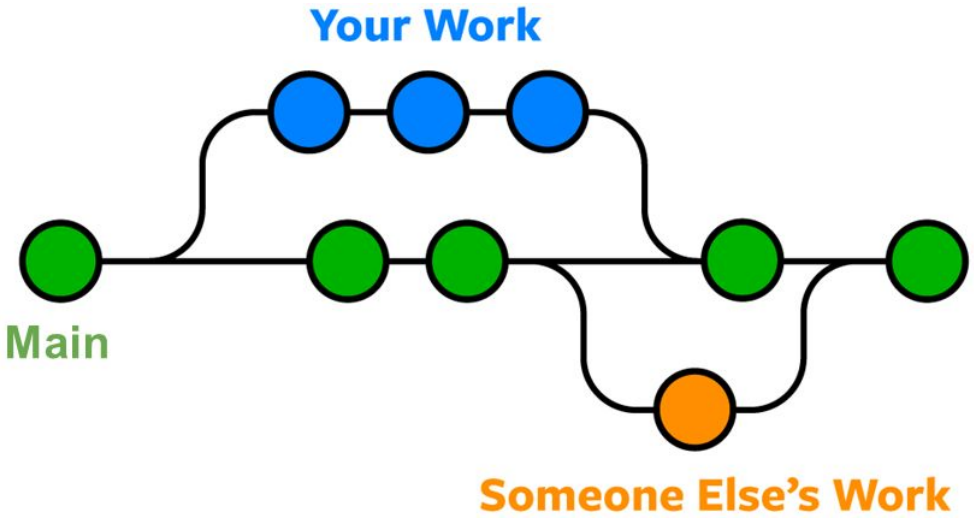
Versioning

- 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



Git/GitHub

- Can be used for anything (not just code!)
- Manage and view of different branches
- Collaboration through merging of branches
- Tracks history



Deleting Files

- Deleting files is also an important part of data management
- I have a monthly calendar reminder to spend some time and delete files I no longer need

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

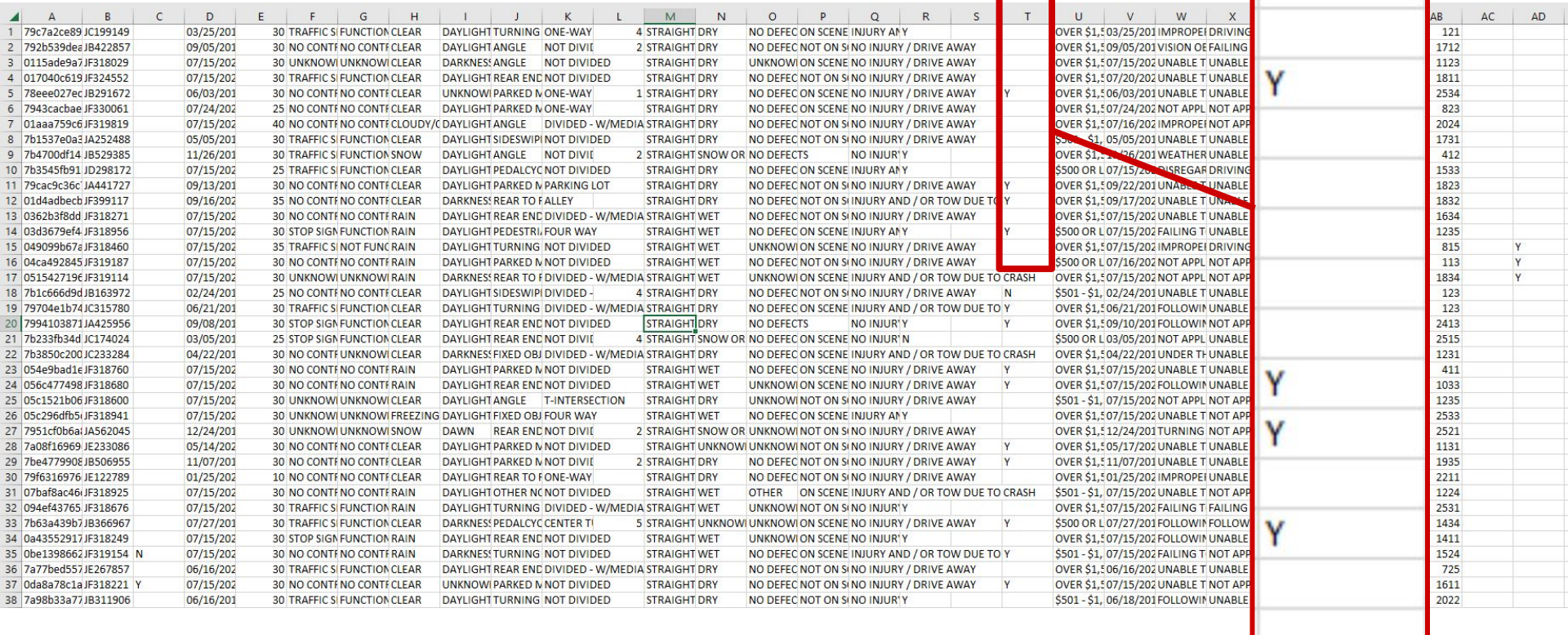
Examples of Documentation

- 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 S	FUNCTION	CLEAR	DAYLIGHT	TURNING	ONE-WAY	4	STRAIGHT	DRY	NO DEFEC	ON SCENE	INJURY AN	Y			OVER $1,5	03/25/201	IMPROPEI	DRIVING	121		
2	792b539dea	JB422857		09/05/201	30	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	ANGLE	NOT DIVI	2	STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	09/05/201	VISION OE	FAILING	1712		
3	0115ade9a7	JF318029		07/15/202	30	UNKNOWI	UNKNOWI	CLEAR	DARKNESS	ANGLE	NOT DIVIDED		STRAIGHT	DRY	UNKNOWI	ON SCENE	NO INJURY / DRIVE AWAY				OVER $1,5	07/15/202	UNABLE T	UNABLE	1123		
4	017040c619	JF324552		07/15/202	30	TRAFFIC S	FUNCTION	CLEAR	DAYLIGHT	REAR END	NOT DIVIDED		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	07/20/202	UNABLE T	UNABLE	1811		
5	78eee027ec	JB291672		06/03/201	30	NO CONTF	NO CONTF	CLEAR	UNKNOWI	PARKED M	ONE-WAY	1	STRAIGHT	DRY	NO DEFEC	ON SCENE	NO INJURY / DRIVE AWAY			Y	OVER $1,5	06/03/201	UNABLE T	UNABLE	2534		
6	7943cacbae	JF330061		07/24/202	25	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	PARKED M	ONE-WAY		STRAIGHT	DRY	NO DEFEC	ON SCENE	NO INJURY / DRIVE AWAY				OVER $1,5	07/24/202	NOT APPL	NOT APP	823		
7	01aaa759c6	JF319819		07/15/202	40	NO CONTF	NO CONTF	CLOUDY/C	DAYLIGHT	ANGLE	DIVIDED - W/MEDIA		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	07/16/202	IMPROPEI	NOT APP	2024		
8	7b1537e0a3	JA252488		05/05/201	30	TRAFFIC S	FUNCTION	CLEAR	DAYLIGHT	SIDESWIPI	NOT DIVIDED		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				$50[illegible]	$1, 05/05/201	UNABLE T	UNABLE	1731		
9	7b4700df14	JB529385		11/26/201	30	TRAFFIC S	FUNCTION	SNOW	DAYLIGHT	ANGLE	NOT DIVI	2	STRAIGHT	SNOW OR	NO DEFECTS		NO INJUR	Y			OVER $1,5	[illegible]/26/201	WEATHER	UNABLE	412		
10	7b3545fb91	JD298172		07/15/202	25	TRAFFIC S	FUNCTION	CLEAR	DAYLIGHT	PEDALCYC	NOT DIVIDED		STRAIGHT	DRY	NO DEFEC	ON SCENE	INJURY AN	Y			$500 OR L	07/15/20[illegible]	DISREGAR	DRIVING	1533		
11	79cac9c36c	JA441727		09/13/201	30	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	PARKED M	PARKING LOT		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY			Y	OVER $1,5	09/22/201	UNABLE T	UNABLE	1823		
12	01d4adbecb	JF399117		09/16/202	35	NO CONTF	NO CONTF	CLEAR	DARKNESS	REAR TO F	ALLEY		STRAIGHT	DRY	NO DEFEC	NOT ON SI	INJURY AND / OR TOW DUE TO			Y	OVER $1,5	09/17/202	UNABLE T	UNABLE	1832		
13	0362b3f8dd	JF318271		07/15/202	30	NO CONTF	NO CONTF	RAIN	DAYLIGHT	REAR END	DIVIDED - W/MEDIA		STRAIGHT	WET	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	07/15/202	UNABLE T	UNABLE	1634		
14	03d3679ef4	JF318956		07/15/202	30	STOP SIGN	FUNCTION	RAIN	DAYLIGHT	PEDESTRI	FOUR WAY		STRAIGHT	WET	NO DEFEC	ON SCENE	INJURY AN	Y		Y	$500 OR L	07/15/202	FAILING T	UNABLE	1235		
15	049099b67a	JF318460		07/15/202	35	TRAFFIC S	NOT FUNC	RAIN	DAYLIGHT	TURNING	NOT DIVIDED		STRAIGHT	WET	UNKNOWI	ON SCENE	NO INJURY / DRIVE AWAY				OVER $1,5	07/15/202	IMPROPEI	DRIVING	815		Y
16	04ca492845	JF319187		07/15/202	30	NO CONTF	NO CONTF	RAIN	DAYLIGHT	PARKED M	NOT DIVIDED		STRAIGHT	WET	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				$500 OR L	07/16/202	NOT APPL	NOT APP	113		Y
17	051542719e	JF319114		07/15/202	30	UNKNOWI	UNKNOWI	RAIN	DARKNESS	REAR TO F	DIVIDED - W/MEDIA		STRAIGHT	WET	UNKNOWI	ON SCENE	INJURY AND / OR TOW DUE TO	CRASH			OVER $1,5	07/15/202	NOT APPL	NOT APP	1834		Y
18	7b1c666d9d	JB163972		02/24/201	25	NO CONTF	NO CONTF	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 S	FUNCTION	CLEAR	DAYLIGHT	TURNING	DIVIDED - W/MEDIA		STRAIGHT	DRY	NO DEFEC	ON SCENE	INJURY AND / OR TOW DUE TO			Y	OVER $1,5	06/21/201	FOLLOWIN	UNABLE	123		
20	7994103871	JA425956		09/08/201	30	STOP SIGN	FUNCTION	CLEAR	DAYLIGHT	REAR END	NOT DIVIDED		STRAIGHT	DRY	NO DEFECTS		NO INJUR	Y		Y	OVER $1,5	09/10/201	FOLLOWIN	NOT APP	2413		
21	7b233fb34d	JC174024		03/05/201	25	STOP SIGN	FUNCTION	CLEAR	DAYLIGHT	REAR END	NOT DIVI	4	STRAIGHT	SNOW OR	NO DEFEC	ON SCENE	NO INJUR	N			$500 OR L	03/05/201	NOT APPL	UNABLE	2515		
22	7b3850c200	JC233284		04/22/201	30	NO CONTF	UNKNOWI	CLEAR	DARKNESS	FIXED OBJ	DIVIDED - W/MEDIA		STRAIGHT	DRY	NO DEFEC	ON SCENE	INJURY AND / OR TOW DUE TO	CRASH			OVER $1,5	04/22/201	UNDER TH	UNABLE	1231		
23	054e9bad1e	JF318760		07/15/202	30	NO CONTF	NO CONTF	RAIN	DAYLIGHT	PARKED M	NOT DIVIDED		STRAIGHT	WET	NO DEFEC	ON SCENE	NO INJURY / DRIVE AWAY			Y	OVER $1,5	07/15/202	UNABLE T	UNABLE	411		
24	056c477498	JF318680		07/15/202	30	NO CONTF	NO CONTF	RAIN	DAYLIGHT	REAR END	NOT DIVIDED		STRAIGHT	WET	UNKNOWI	ON SCENE	NO INJURY / DRIVE AWAY			Y	OVER $1,5	07/15/202	FOLLOWIN	UNABLE	1033		
25	05c1521b06	JF318600		07/15/202	30	UNKNOWI	UNKNOWI	CLEAR	DAYLIGHT	ANGLE	T-INTERSECTION		STRAIGHT	DRY	UNKNOWI	NOT ON SI	NO INJURY / DRIVE AWAY				$501 - $1,	07/15/202	NOT APPL	NOT APP	1235		
26	05c296dfb5	JF318941		07/15/202	30	UNKNOWI	UNKNOWI	FREEZING	DAYLIGHT	FIXED OBJ	FOUR WAY		STRAIGHT	WET	NO DEFEC	ON SCENE	INJURY AN	Y			OVER $1,5	07/15/202	UNABLE T	NOT APP	2533		
27	7951cf0b6a	JA562045		12/24/201	30	UNKNOWI	UNKNOWI	SNOW	DAWN	REAR END	NOT DIVI	2	STRAIGHT	SNOW OR	UNKNOWI	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	12/24/201	TURNING	NOT APP	2521		
28	7a08f16969	JE233086		05/14/202	30	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	PARKED M	NOT DIVIDED		STRAIGHT	UNKNOWI	UNKNOWI	NOT ON SI	NO INJURY / DRIVE AWAY			Y	OVER $1,5	05/17/202	UNABLE T	UNABLE	1131		
29	7be477990[illegible]	JB506955		11/07/201	30	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	PARKED M	NOT DIVI	2	STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY			Y	OVER $1,5	11/07/201	UNABLE T	UNABLE	1935		
30	79f6316976	JE122789		01/25/202	10	NO CONTF	NO CONTF	CLEAR	DAYLIGHT	REAR TO F	ONE-WAY		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	01/25/202	IMPROPEI	UNABLE	2211		
31	07baf8ac46	JF318925		07/15/202	30	NO CONTF	NO CONTF	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	JF318676		07/15/202	30	TRAFFIC S	FUNCTION	RAIN	DAYLIGHT	TURNING	DIVIDED - W/MEDIA		STRAIGHT	WET	UNKNOWI	NOT ON SI	NO INJUR	Y			OVER $1,5	07/15/202	FAILING T	FAILING	2531		
33	7b63a439b7	JB366967		07/27/201	30	TRAFFIC S	FUNCTION	CLEAR	DARKNESS	PEDALCYC	CENTER TI	5	STRAIGHT	UNKNOWI	UNKNOWI	ON SCENE	NO INJURY / DRIVE AWAY			Y	$500 OR L	07/27/201	FOLLOWIN	FOLLOW	1434		
34	0a43552917	JF318249		07/15/202	30	STOP SIGN	FUNCTION	RAIN	DAYLIGHT	REAR END	NOT DIVIDED		STRAIGHT	WET	UNKNOWI	ON SCENE	NO INJUR	Y			OVER $1,5	07/15/202	FOLLOWIN	UNABLE	1411		
35	0be1398662	JF319154	N	07/15/202	30	NO CONTF	NO CONTF	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	7a77bed557	JE267857		06/16/202	30	TRAFFIC S	FUNCTION	CLEAR	DAYLIGHT	REAR END	DIVIDED - W/MEDIA		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY				OVER $1,5	06/16/202	UNABLE T	UNABLE	725		
37	0da8a78c1a	JF318221	Y	07/15/202	30	NO CONTF	NO CONTF	CLEAR	UNKNOWI	PARKED M	NOT DIVIDED		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJURY / DRIVE AWAY			Y	OVER $1,5	07/15/202	UNABLE T	NOT APP	1611		
38	7a98b33a77	JB311906		06/16/201	30	TRAFFIC S	FUNCTION	CLEAR	DAYLIGHT	TURNING	NOT DIVIDED		STRAIGHT	DRY	NO DEFEC	NOT ON SI	NO INJUR	Y			$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]

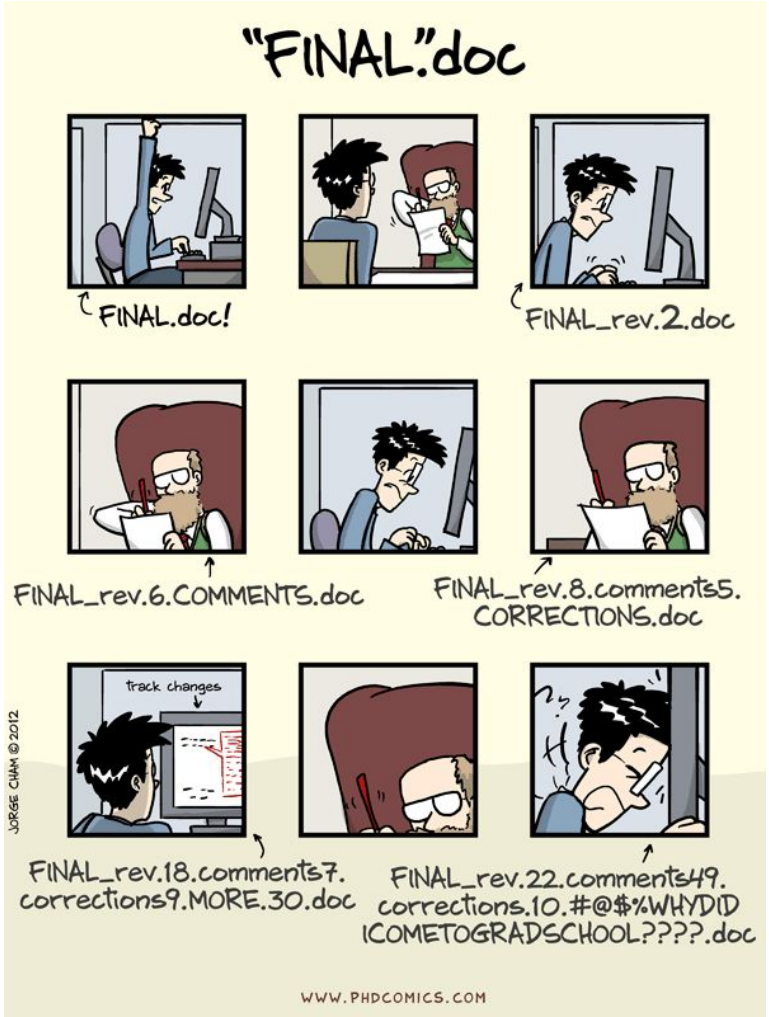
License

Creative Commons BY Attribution 4.0 International

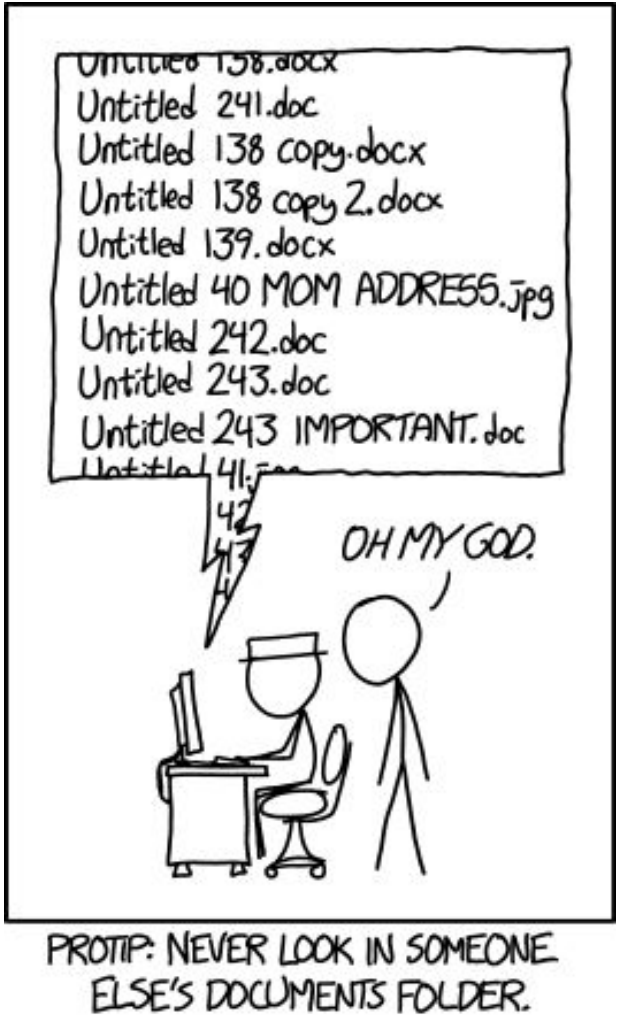
What to document for other people's data?

- When you downloaded the data
- Where you downloaded it from
- Who created it
- Any associated documentation related to the dataset that is not in the dataset itself
- Any changes you make to the dataset

File Names & Structures



PHD Comics: NotFinal.Doc. Jorge Cham, 2012. https://phdcomics.com/comics.php?f=1531 .



xkcd: Documents. Randall Munroe. https://xkcd.com/1459/

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 (some) 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!

OLS, OntoPortal, Skosmos - API mapping specifications for API Gateway and Termiology Service Suite - entity info ✅.csv

md5:b22bf8780e38a5c880895aa2986a7f8a

62 Bytes

Preview Download

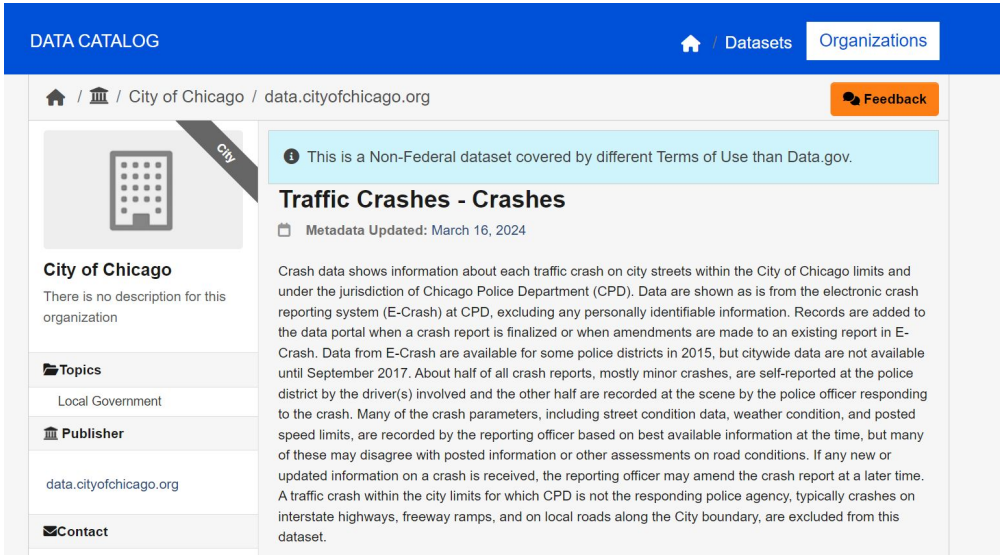
OLS, OntoPortal, Skosmos - API mapping specifications for API Gateway and Termiology Service Suite - entity relations ❌.csv

md5:caf5a911e96a28f6811eba834a37ce5b

683 Bytes

Preview Download

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

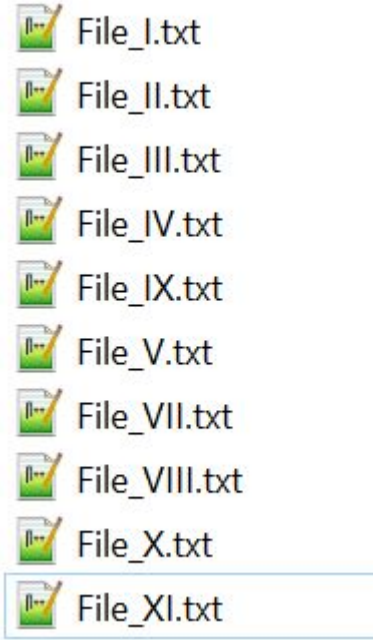
- Don’t use terms like “Quarter 1” or “spring”
- Use ISO 8601
  - 2023-10-31 or 20231031 (YYYY-MM-DD)
- Remember that historically dates were not consistent
  - 1712-02-30 (February 30th, 1712) was a real date in Sweden

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 ﷺ and Mara L. Thacker

Additional file 4: of The gut of the finch: uniqueness of the gut microbiome of the GalĂĄpagos 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 Oâ□□Neill, 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

1. 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

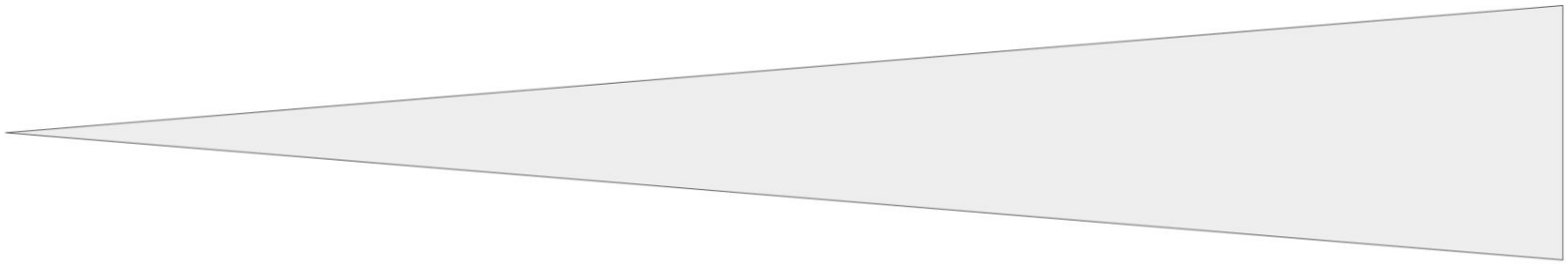
GalĂĄpagos Islands | Darwinâ□□s finches | Blood-feeding | Microbiome | Geospiza

LICENCE

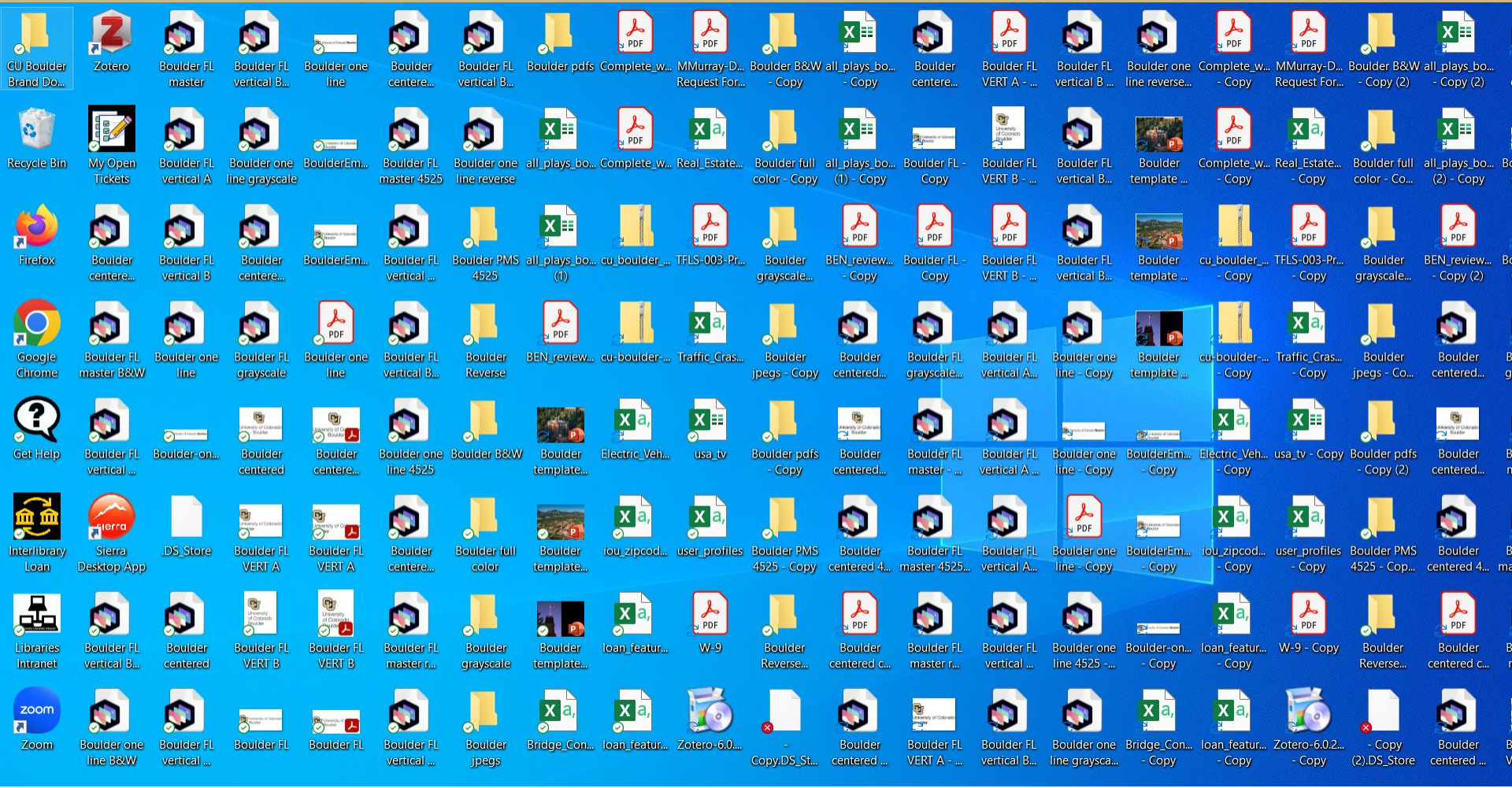
CC BY + CC0

20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx

Date **Project** Experiment **Type** ID **Version**

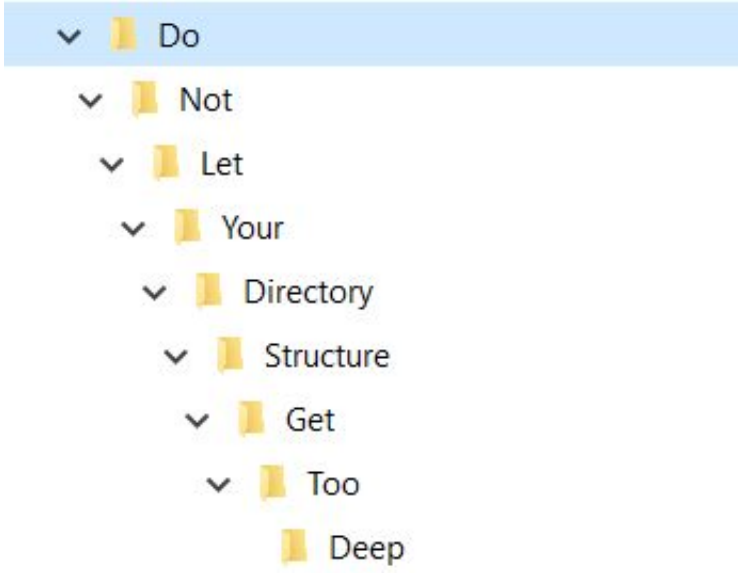


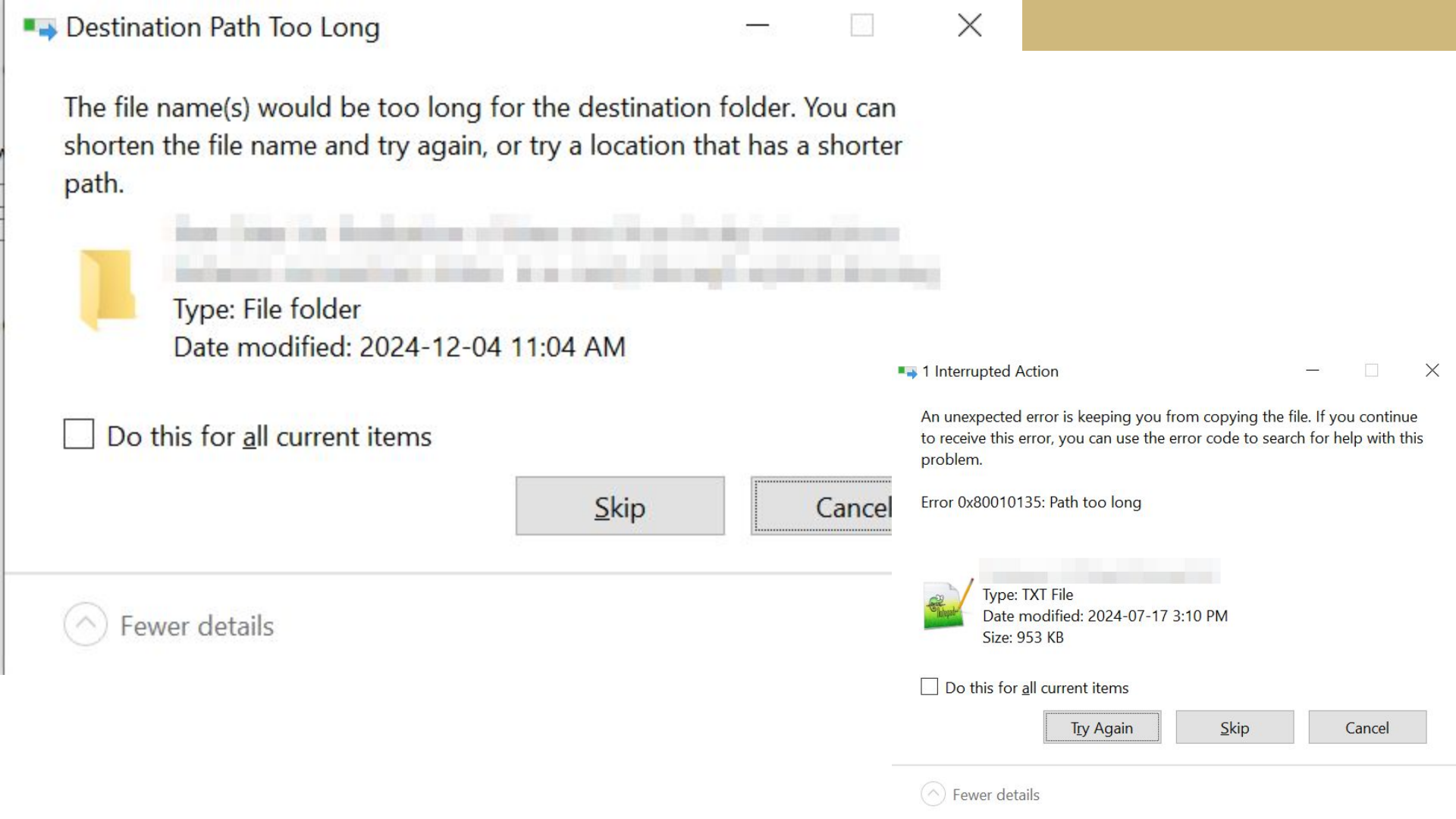
General → Specific


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 - Copy
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 - 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...
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 graysca...
Bridge_Con... - Copy
loan_featur... - Copy
Zotero-6.0.2... - Copy
- Copy (2).DS_Store
Boulder centered ...

File Structuring

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




Destination Path Too Long
The file name(s) would be too long for the destination folder. You can shorten the file name and try again, or try a location that has a shorter path.
Type: File folder
Date modified: 2024-12-04 11:04 AM
Do this for all current items
Skip
Cancel
Fewer details
1 Interrupted Action
An unexpected error is keeping you from copying the file. If you continue to receive this error, you can use the error code to search for help with this problem.
Error 0x80010135: Path too long
Type: TXT File
Date modified: 2024-07-17 3:10 PM
Size: 953 KB
Do this for all current items
Try Again
Skip
Cancel
Fewer details

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

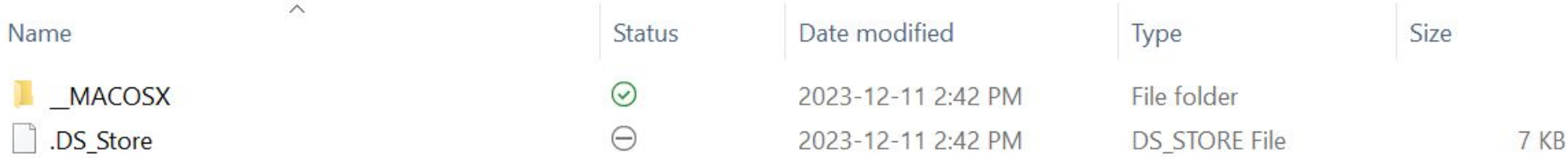
- Usually called “zipping”
- Means collecting different files and directories into one file
- Many programs/methods of doing this
  - .zip, .rar., .7z, .tar, .gz, etc.
- Done to minimize size of files (compression)
- Done to collect many files and directories into one file (archiving)
- Makes it easier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

File Archiving and Compression

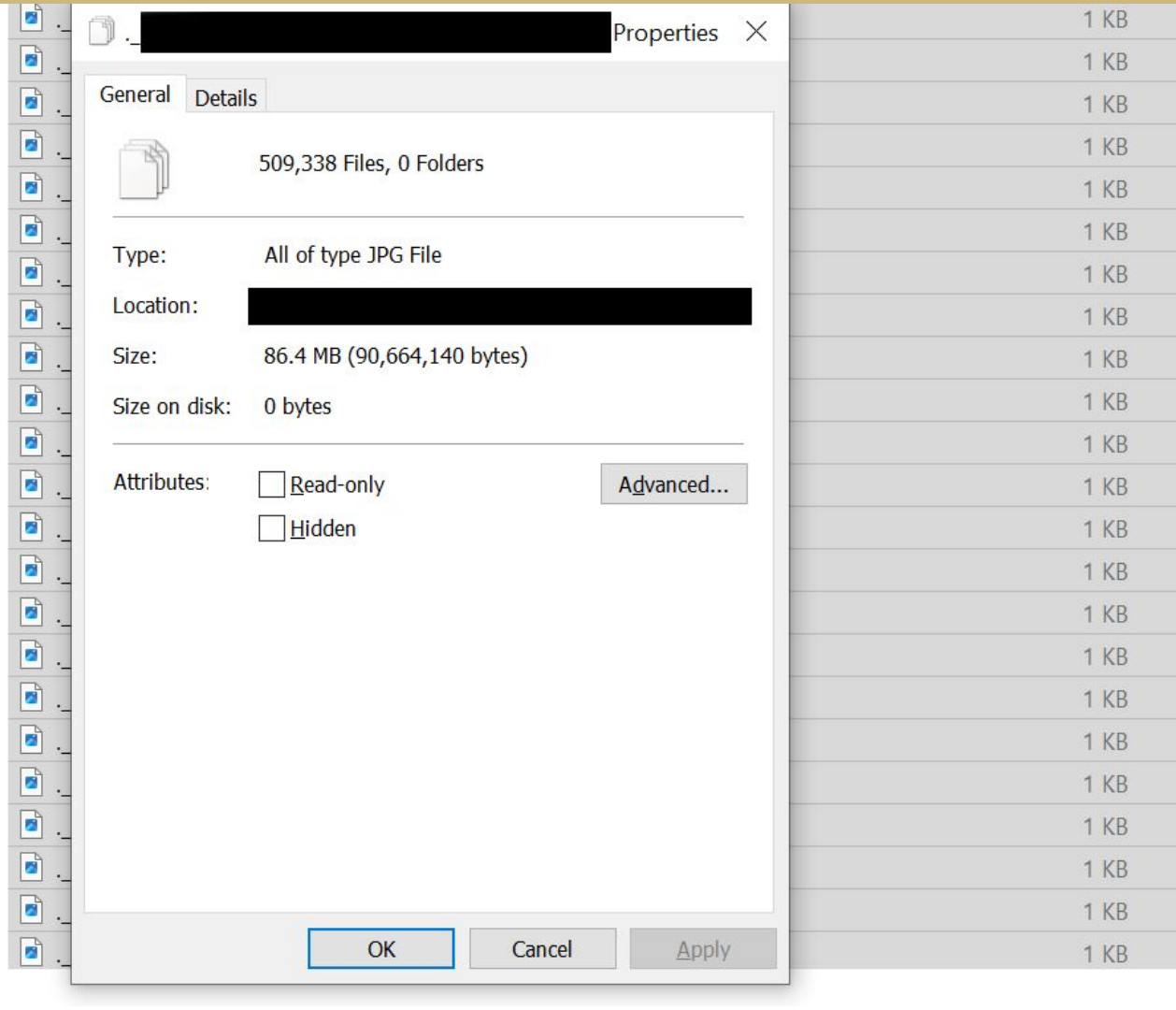
- Consider how the files will look once they’re extracted (unzipped)
- Create a directory structure that will make it easy for people to understand and use your files
- Ensure all file paths will allow associated code to work

CHAFF

- CHAFF: Concealed, Hidden, and Forgotten Files
- __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
- Other types of CHAFF exist



This .zip file contained over 500,000 ._ files



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

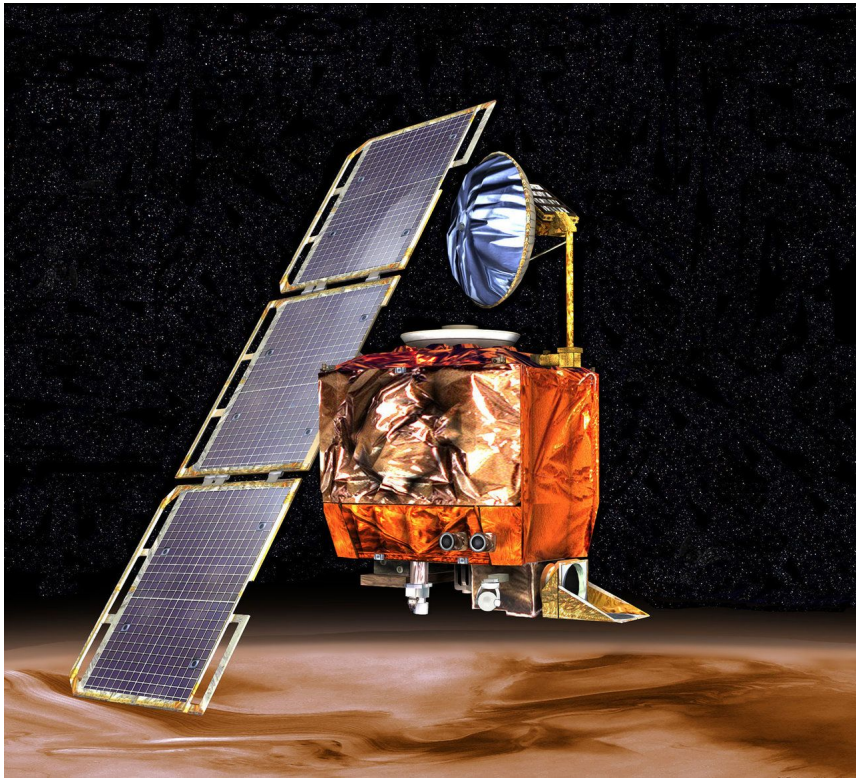
Units of Measurement

- 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

- Spacecraft that was probably destroyed in the Martian atmosphere
- The failure was due to a measurement mismatch between SI units (metric) and US customary units causing numbers to be incorrect by a factor of 4.45



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 Open

openforcefield/openff-qcsubmit: 0.50.0

JoshHorton; Josh Horton[1]; SimonBoothroyd; trevorgokey; David L. Dotson[2]; Josh A. Mitchell[3]; Brent Westbrook[3]; Matt Thompson[4]; Jeff Wagner[3]; Chapin Cavender[5]; John Chodera[6]

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.yaml 616 Bytes
  - .gitignore 1.2 kB
  - .lgtm.yml 332 Bytes

Files (325.3 kB)

Name	Size	Download all
openforcefield/openff-qcsubmit-0.50.0.zip md5:f37072b3cfddcb55dc6741df8fe001c2	325.3 kB	Preview Download

Versions

Version 0.50.0 Nov 1, 2023
10.5281/zenodo.10060049

Version 0.5.0 Aug 16, 2023
10.5281/zenodo.8253341

Version 0.4.1 Jan 6, 2023
10.5281/zenodo.7510744

Version 0.4.0 Nov 15, 2022
10.5281/zenodo.7320745

Version 0.3.2 Nov 8, 2022
10.5281/zenodo.7305244

View all 9 versions

Cite all versions? You can cite all versions by using the DOI 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

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

- Open Science Framework
  - One-stop shop for project management
- Open Refine
  - Clean data
- File Renamers (various)
  - Rename files so they’re consistent
  - You can also batch rename files in the command line
- 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