

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?

- Spreadsheets / CSV files
- Computer code & scripts
- Audio & Video files
- Transcripts
- GIS files
- Lab notebooks
- Documentation
- Questionnaires
- Physical artifacts & specimens
- 3D models
- Point clouds
- Methodologies & workflows
- Designs & blueprints
- Codebooks

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

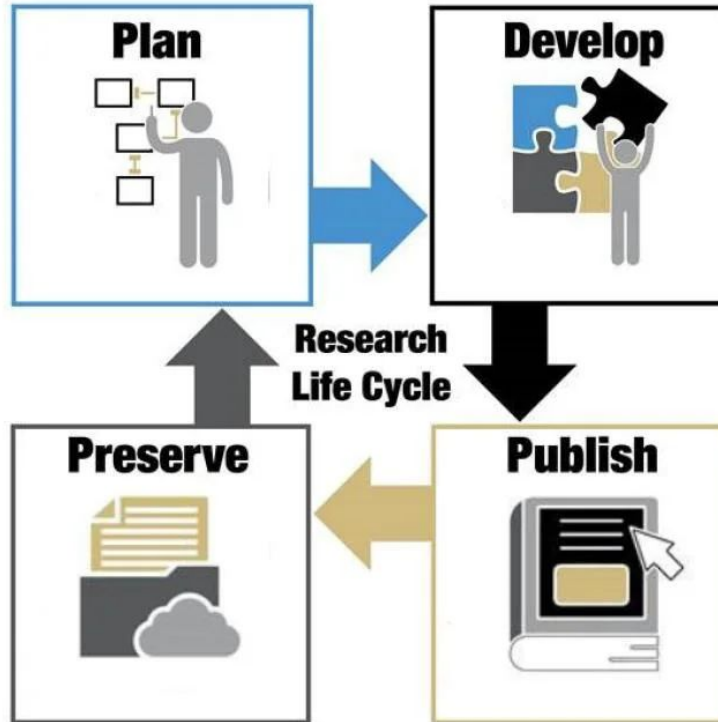
- Data management is what you *do*.
- Data Management Plans are where you write about what you're going to do. (Usually because a grant requires one.)

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



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.

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	Y	Z	AA	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 S	536	W	RANDOLP	121			
2	792b539de4	JB422857		09/05/201	30	NO CONF	NO CONF	CLEAR	DAYLIGHT ANGLE	NOT DIVI	2	STRAIGHT DRY	NO DEFEC NOT ON SI	NO INJURY / DRIVE AWAY							OVER \$1,509/05/201	VISION OE	FAILING T	4821	N	ELSTON A	1712			
3	0115ade97f	JF318029		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 T	300	N	CENTRAL I	1123			
4	017040c619	JF324552		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 T	2359	N	ASHLAND	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 T	1630	N	KEELER A	2534			
6	7943cacbae	JF330061		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 APPL	2817	W	61ST ST	823			
7	01aaa759c6	JF319819		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 APPL	5100	N	LAKE SHO	2024			
8	7b1537e0a2	JA252488		05/05/201	30	TRAFFIC SI	FUNCTION CLEAR	DAYLIGHT SIDES	WIPI	NOT DIVIDED		STRAIGHT DRY	NO DEFEC NOT ON SI	NO INJURY / DRIVE AWAY							\$501 - \$1,05/05/201	UNABLE T	UNABLE T	4498	W	ADDISON	1731			
9	7b4700fd14	JB529385		11/26/201	30	TRAFFIC SI	FUNCTION SNOW	DAYLIGHT ANGLE	NOT DIVI	2	STRAIGHT SNOW OR	NO DEFECTS	NO INJURY								OVER \$1,511/26/201	WEATHER	UNABLE T	8300	S	STONY ISL	412			
10	7b3545fb91	JD298172		07/15/202	25	TRAFFIC SI	FUNCTION CLEAR	DAYLIGHT PEDAL	CYC	NOT DIVIDED		STRAIGHT DRY	NO DEFEC ON SCENE	INJURY ANY							\$500 OR L07/15/202	DISREGAR	DRIVING S	298	S	CICERO A	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 T	225	W	DIVISION	1823			
12	01d4adbcbf	JF399117		09/16/202	35	NO CONF	NO CONF	CLEAR	DARKNESS	REAR TO F	ALLEY		STRAIGHT DRY	NO DEFEC NOT ON SI	INJURY AND / OR TOW DUE TO Y						OVER \$1,509/17/202	UNABLE T	UNABLE T	609	N	CLARK ST	1832			
13	0362b3f8dd	JF318271		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 T	3458	N	CICERO A	1634			
14	03d3679ef4	JF318956		07/15/202	30	STOP SIGN	FUNCTION RAIN	DAYLIGHT PEDESTRI	FOUR WAY		STRAIGHT WET	NO DEFEC ON SCENE	INJURY ANY						Y		\$500 OR L07/15/202	FAILING T	UNABLE T	1900	S	LAFLIN ST	1235			
15	049099b67e	JF318460		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 S	4400	W	47TH ST	815			Y
16	04ca492845	JF319187		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 L07/16/202	NOT APPL	NOT APPL	318	S	FEDERAL S	113			Y
17	051542719f	JF319114		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 APPL	401	N	WABASH	1834			Y
18	7b1c666d9d	JB163972		02/24/201	25	NO CONF	NO CONF	CLEAR	DAYLIGHT SIDES	WIPI	DIVIDED -	4	STRAIGHT DRY	NO DEFEC NOT ON SI	NO INJURY / DRIVE AWAY				N		\$501 - \$1,02/24/201	UNABLE T	UNABLE T	520	S	MICHIGAN	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									OVER \$1,506/21/201	FOLLOWIN	UNABLE T	1198	S	STATE ST	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 APPL	3003	W	GRANVILL	2413			
21	7b233fb34d	JC174024		03/05/201	25	STOP SIGN	FUNCTION CLEAR	DAYLIGHT REAR ENC	NOT DIVI	4	STRAIGHT SNOW OR	NO DEFEC ON SCENE	NO INJURY	N							\$500 OR L03/05/201	NOT APPL	UNABLE T	5331	W	FULLERTO	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 T	804	S	ASHLAND	1231			
23	054e9bad1e	JF318760		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 T	7804	S	SHENWOOD	411			
24	056c477498	JF318680		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 T	3047	W	CERMAK F	1033			
25	05c1521b06	JF318600		07/15/202	30	UNKNOWN	UNKNOWN	CLEAR	DAYLIGHT ANGLE	T-INTERSECTION		STRAIGHT DRY	UNKNOWN ON SCENE	NO SI	NO INJURY / DRIVE AWAY						\$501 - \$1,07/15/202	NOT APPL	NOT APPL	939	W	PULASKI R	1235			
26	05c296dfb5	JF318941		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 APPL	1502	N	CICERO A	2533			
27	7951cf0b6a	JA562045		12/24/201	30	UNKNOWN	UNKNOWN	SNOW	DAWN	REAR ENC	NOT DIVI	2	STRAIGHT SNOW OR	UNKNOWN	NOT ON SI	NO INJURY / DRIVE AWAY					OVER \$1,512/24/201	TURNING	NOT APPL	2797	N	LARAMIE	2521			
28	7a08f16969	JE233086		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 T	566	S	KILPATRIC	1131			
29	7be477990c	JB506955		11/07/201	30	NO CONF	NO CONF	CLEAR	DAYLIGHT PARKED N	NOT DIVI	2	STRAIGHT DRY	NO DEFEC NOT ON SI	NO INJURY / DRIVE AWAY					Y		OVER \$1,511/07/201	UNABLE T	UNABLE T	401	W	WRIGHTW	1935			
30	79f631697e	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 T	10343	S	PULASKI R	2211			
31	07baf8ac46	JF318925		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 APPL	1676	W	OGDEN A	1224			
32	094ef43765	JF318676		07/15/202	30	TRAFFIC SI	FUNCTION RAIN	DAYLIGHT TURNING	DIVIDED - W/MEDIA	STRAIGHT WET	UNKNOWN	NOT ON SI	NO INJURY	Y							OVER \$1,507/15/202	FAILING T	FAILING T	5600	W	DIVISION	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 L07/27/201	FOLLOWIN	FOLLOWIN	1941	N	WESTERN	1434			
34	04a3552917	JF318249		07/15/202	30	STOP SIGN	FUNCTION RAIN	DAYLIGHT REAR ENC	NOT DIVIDED		STRAIGHT WET	UNKNOWN	ON SCENE	NO INJURY	Y						OVER \$1,507/15/202	FOLLOWIN	UNABLE T	2857	N	KEDZIE A	1411			
35	0be1398662	JF319154	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 APPL	5335	W	CHICAGO	1524			
36	7a77bed557	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 T	1554	W	69TH ST	725			
37	0da8a78c1a	JF318221	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 APPL	5925	N	NORTHWI	1611			
38	79a8b33a77	JB311906		06/16/201	30	TRAFFIC SI	FUNCTION CLEAR	DAYLIGHT TURNING	NOT DIVIDED		STRAIGHT DRY	NO DEFEC NOT ON SI	NO INJURY	Y							\$501 - \$1,06/18/201	FOLLOWIN	UNABLE T	5700	N	KENMORE	2022			

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 CHAM © 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/>

PROTIP: 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 inyourfilenames.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

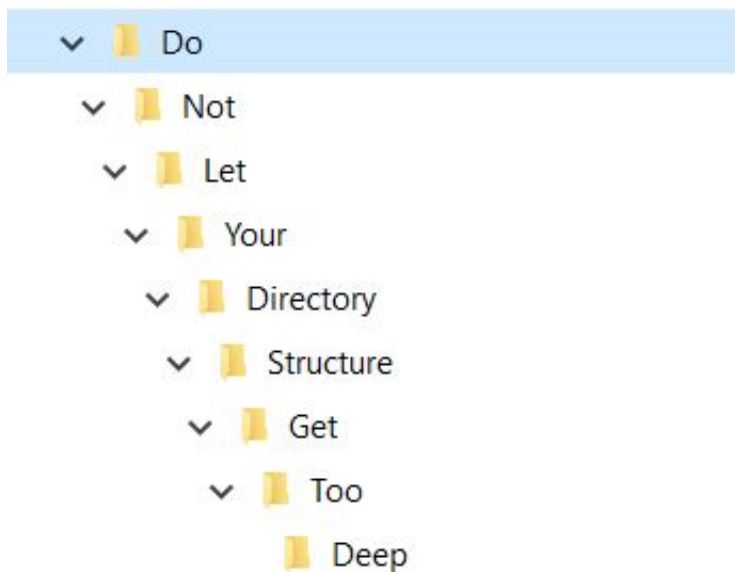
Dates

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



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: rdds@colorado.edu