

Research Data Management

A Guide to Good Practice



LIBRARIES
TEXAS A&M UNIVERSITY

Goals for Today

1. Conceptualize your personal data workflow
2. Create practices to streamline file management
3. Choose appropriate data storage solutions and strategies



What to Expect

1. This will be a broad overview of good practices.
2. Slides are shared after the workshop via email.
3. Ask questions via chat at any time.
4. Grad Aggie Certificate and RCR sign-in sheet shared at close.



Action Plan

3 - 3 - 3 Action Plan for better data management

- 3 goals to accomplish over a...
- 3 month timeline with...
- 3 tools or good practices

Keep this in mind throughout the workshop!



Definition

Data management is a set of practices across the research lifecycle to improve:

- Integrity of files and data
- Protection and security of data
- Efficiency and reliability of research
- Reuse of data (citations)

How Data Management Works for You

Data are citable products of research that can:

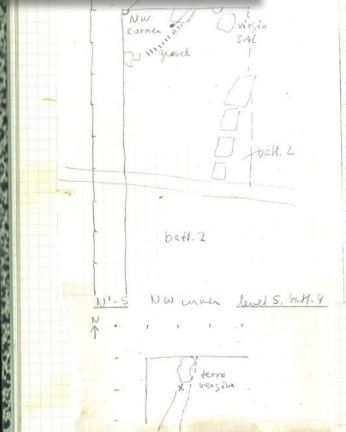
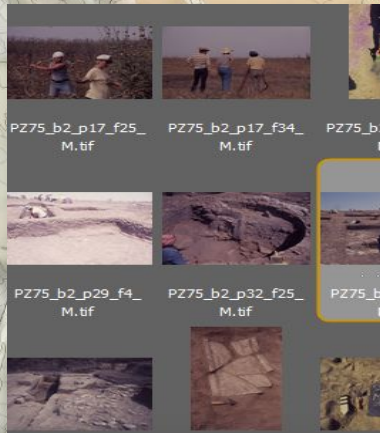
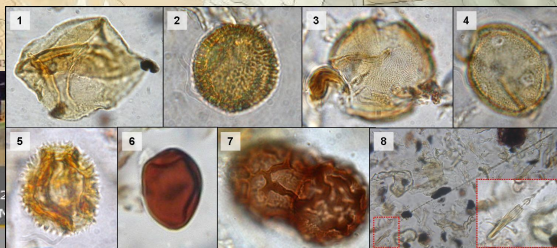
- Foster collaboration
- Enable reproducibility of findings
- Demonstrate impact of your work
- Contribute to your scholarly record



**What do we mean when we
talk about data?**

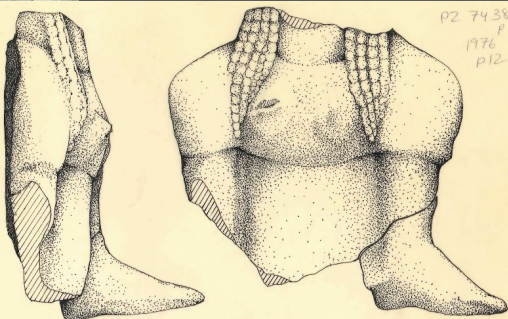


What are "data"?



PIZZICA, 1977 TVERDN, 5 JULY 1977

11-5, Level 5, bath 4
 purpose: to define extent of
 brown soil (which overlies
 in N.E. corner of bath 4) flood
 beach soil (N.W. corner)
 black heavy soil - lighter
 shade to indicate (irregularity
 in soil of bath 2)
 PZ 77-749 P, PZ 77-973 P
 PZ 77-749 P (bath 2), less shade, very and
 in bath 3
 PZ 77-749 NW corner, miniature vase
 PZ 77-749 complete female figurine
 black from bath 2 (20 m.
 (not from stake in N.W.))
 at the end of bath 4, there
 is a depression approx. in a
 N-S line with light soil
 on either side - this surface
 was cleaned and photo-
 graphed.
 level 5, bath 5
 again concentration of good
 soil cont. material at NW
 corner - mainly white ware,
 iron and traces of others.
 11-5 level 5, bath 3, 8 and
 9 of level 5, PZ 77-965 P
 Fract. bronze shill
 PZ 77-735 M
 PZ 77-966 P, PZ 77-967 P, PZ 77-968 P



art_prop	Year	Weathering
100	2005	Heavy
100	2007	Light
100	2007	Heavy
100	2007	Light
95	2007	Heavy
95	2005	Heavy
95	2007	Light
95	2007	Light
90	2007	Moderate
90	2007	Light
90	2007	Heavy
90	2007	Light
90	2006	Moderate
90	2007	Light
85	2005	Heavy

What do you use and create?

Spreadsheets

Databases

Instrument readings

Binary files

Code

PDFs

Image files

Audio files

Physical specimens

Archival materials

Geospatial data

Or something else!



Research Lifecycle



Think through your research lifecycle

- What are the steps you take in a research project?
- What do you create and use at each stage?

Let's look at some examples!



SOCIAL SCIENCE
EXAMPLE
LIFECYCLE

IDEA & LIT
SEARCH

PDFs
Lit notes
Mendeley Library

STUDY DESIGN

IRB
Docs for design
Experimental protocols
Grant & funding info

DATA COLLECTION

E prime scripts
Qualtrics surveys

DATA MERGING

R Scripts
Protocols
Combined datasets

DATA ANALYSIS

R Scripts
Data Visualizations
Subset Protocols

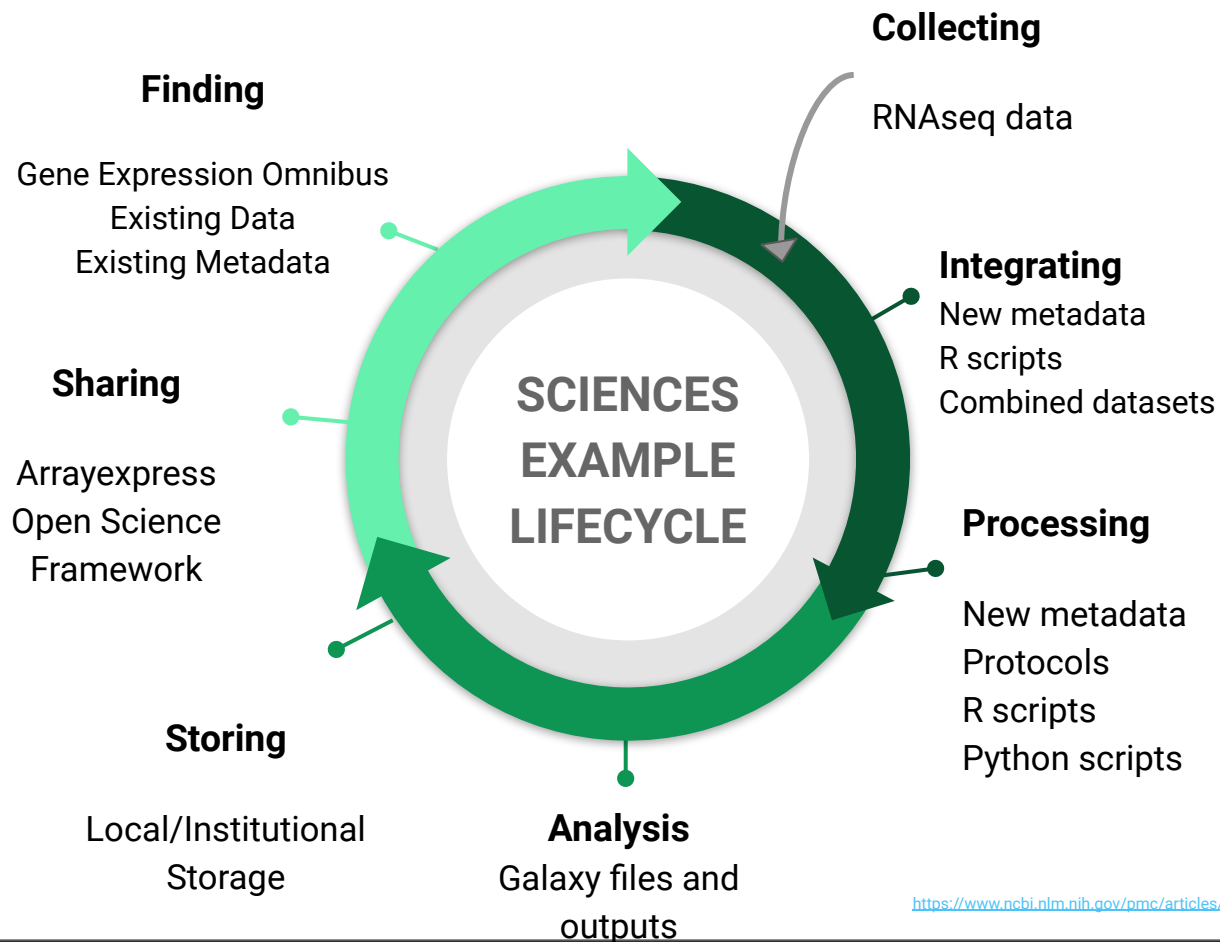
DISSEMINATION

Tables
Final figures
MS drafts
Mendeley Library

SHARING

Data dictionaries
De-identification scripts
Clean analysis scripts
Public use version of data





<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6069748/#:po=7.89474>



GATHER IDEAS & MATERIALS

Archival materials, interviews, media, books, articles (pdfs, images, video, audio, ebooks, papers and books)

READING, NOTE TAKING, ANNOTATION

Highlighting, marginalia, commentary

Books, hard copies, pdfs, folders, piles of stuff, notebooks, screenshots

FEEDBACK & REVIEW

Documents (MS Word, Google Docs), emails, folders

ANALYSIS, INTERPRETATION, CRITIQUE

Notebooks, docs, mind maps, outlines, spreadsheets

WRITING, DRAFTING

Docs created in various writing software, notebooks, slides

PREPARE FOR PUBLISHING

Editing, citations, indexing, proofreading, permissions

HUMANITIES EXAMPLE LIFECYCLE

Activity

Think about a specific current or future project and map out your research lifecycle.

- What are the steps you take in a research project?
- What materials and data do you create and use?

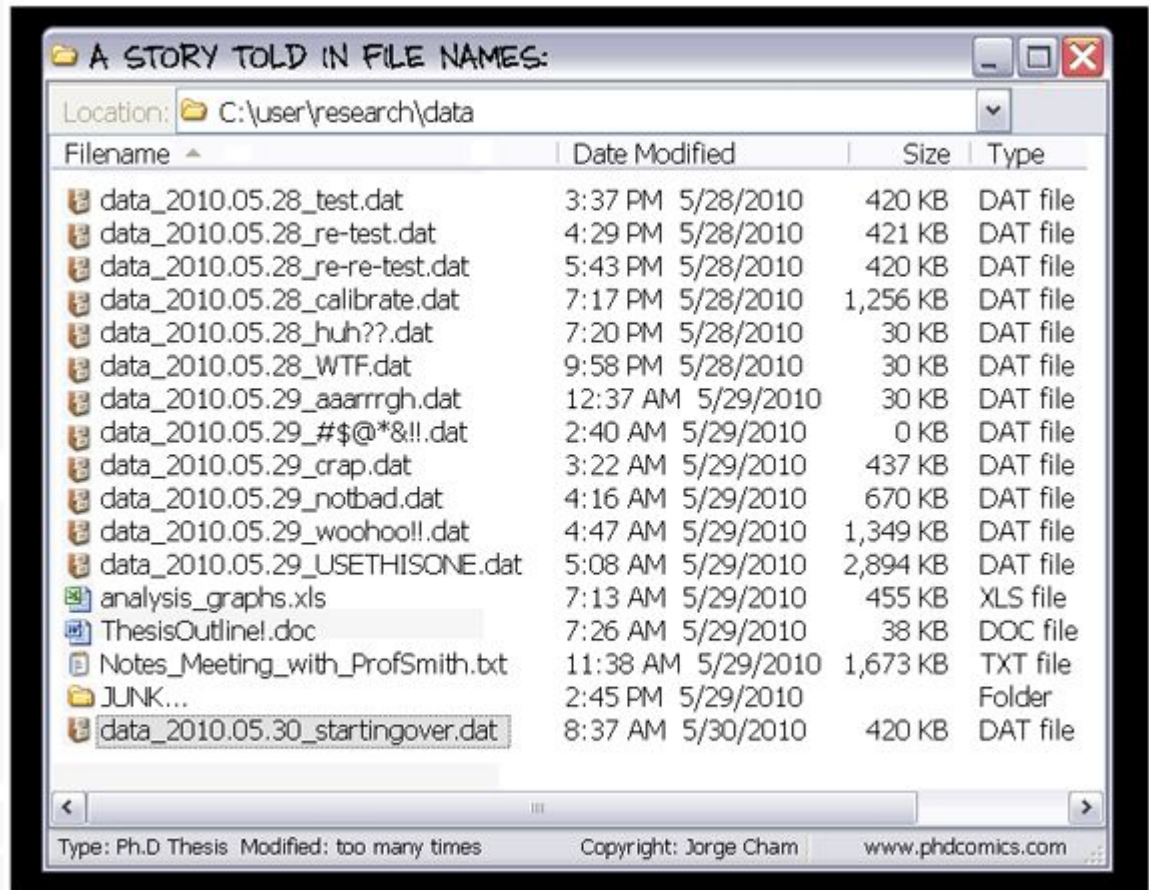


File Management



File Management

You (and others) should be able to understand what is in the file without opening it.



https://phdcomics.com/comics/archive_print.php?comicid=1323



LIBRARIES
TEXAS A&M UNIVERSITY

sam_1262020.tif

What does it represent?

- Scanning Acoustic Microscope?
- Survey of Adolescent Mothers?
- Sam the Grad Student?
- 12, June 2020?
- December 6, 2020?
- Jan 26, 2020?



What is in a file name?

Three main criteria to consider for labelling of data files:

- Context - describe the content, origin, or time period
- Consistent - create a convention used by everyone
- Unique - distinguishes a file from files with similar subjects as well as different versions of the file

[\(Mantra Research Data Management Training\)](#)



Context and Uniqueness

Elements to consider for your file names:

- Date
- Topic/Subject
- Type of data
- Location/site/spatial coordinates
- Researcher info
- Version



Choose a name that makes sense for you

Order by date:

2013-04-12_interview-recording_THD.mp3
2013-04-12_interview-transcript_THD.docx
2012-12-15_interview-recording_MBD.mp3
2012-12-15_interview-transcript_MBD.docx

Order by type:

Interview-recording_MBD_2012-12-15.mp3
Interview-recording_THD_2013-04-12.mp3
Interview-transcript_MBD_2012-12-15.docx
Interview-transcript_THD_2013-04-12.docx

Order by subject:

MBD_interview-recording_2012-12-15.mp3
MBD_interview-transcript_2012-12-15.docx
THD_interview-recording_2013-04-12.mp3
THD_interview-transcript_2013-04-12.docx

Forced order with numbering:

01_THD_interview-recording_2013-04-12.mp3
02_THD_interview-transcript_2013-04-12.docx
03_MBD_interview-recording_2012-12-15.mp3
04_MBD_interview-transcript_2012-12-15.docx



Use A Standard Date System

YYYY-MM-DD

International Organization for Standardization (ISO)

This ISO standard helps remove doubts that can result from the various day–date conventions.

International Organization for Standardization
<https://www.iso.org/standard/70908.html>



It Also Helps You Stay Organized

Sort, *without* standard dates

12-8-22_code_descriptions.docx

2-14-2022_code_descriptions.docx

8-1-2022_code_descriptions.docx

Sort, *with* standard dates

2022-02-14_code_descriptions.docx

2022-08-01_code_descriptions.docx

2022-12-08_code_descriptions.docx



Use Leading Zeros

Sort,
without a
leading zero

X Day1_test results.xlsx

X Day10_test results.xlsx

X Day11_test results.xlsx

X Day2_test results.xlsx

X Day3_test results.xlsx

X Day4_test results.xlsx

X Day5_test results.xlsx

X Day6_test results.xlsx

X Day7_test results.xlsx

X Day8_test results.xlsx

X Day9_test results.xlsx

Sort,
with a
leading
zero

X Day01_test results.xlsx

X Day02_test results.xlsx

X Day03_test results.xlsx

X Day04_test results.xlsx

X Day05_test results.xlsx

X Day06_test results.xlsx

X Day07_test results.xlsx

X Day08_test results.xlsx

X Day09_test results.xlsx

X Day10_test results.xlsx

X Day11_test results.xlsx



Human and Machine Readable

- Avoid spaces and special characters
- “_” underscores to delimit units in names
- “-” hyphens to delimit names for readability

(Briney, 2015)



Avoid Special Characters

Google Drive

My Drive > Shared Folders > Folders with Special Characters

Name ↑	Owner
Data	me
Data_Analysis	me
What is even in this folder?	me
Work "in progress"	me

On a Mac

Folders with Special Characters

Name	Date M
Data	Today
Data_Analysis	Today
What is even in this folder?	Today
Work "in progress"	Today

On a PC

This PC > Desktop > Special-Characters

Name
Data
Data_Analysis
What is even in this folder
Work in progress

Add version numbers to file names

Using consecutive numbering for major version changes








Code_descriptions_20120815_v01.docx

Using decimals for minor changes

Code_descriptions_20120815_v01.1.docx



Inconsistency Strikes Again!

 2021_Week1	me	Jul 16, 2021
 2021_Week2	me	Jul 11, 2021
 2021_Week3	me	Jul 16, 2021
 2021_Week4	me	Jul 26, 2021
 2021_Week5	me	Aug 1, 2021
 2021_Week7	me	10:01 AM
 2021-Week6	me	Aug 1, 2021



What to include?

Not Proprietary Formats

- Excel (.xls, .xlsx)
- Word (.doc, .docx)
- PowerPoint (.ppt, .pptx)
- Photoshop (.psd)
- Quicktime (.mov)
- MPEG 4 Protected Audio (.m4p)

Open Format Equivalents

- Comma Separated Value (.csv)
- Plain Text (.txt)
- PDF/A (.pdf)
- TIFF (.tif, .tiff) or PNG (.png)
- MPEG-4 (.mp4)
- MP3 (.mp3)

Open formats also help preserve documents for the long term



Folder Management Strategy

Possible strategies:

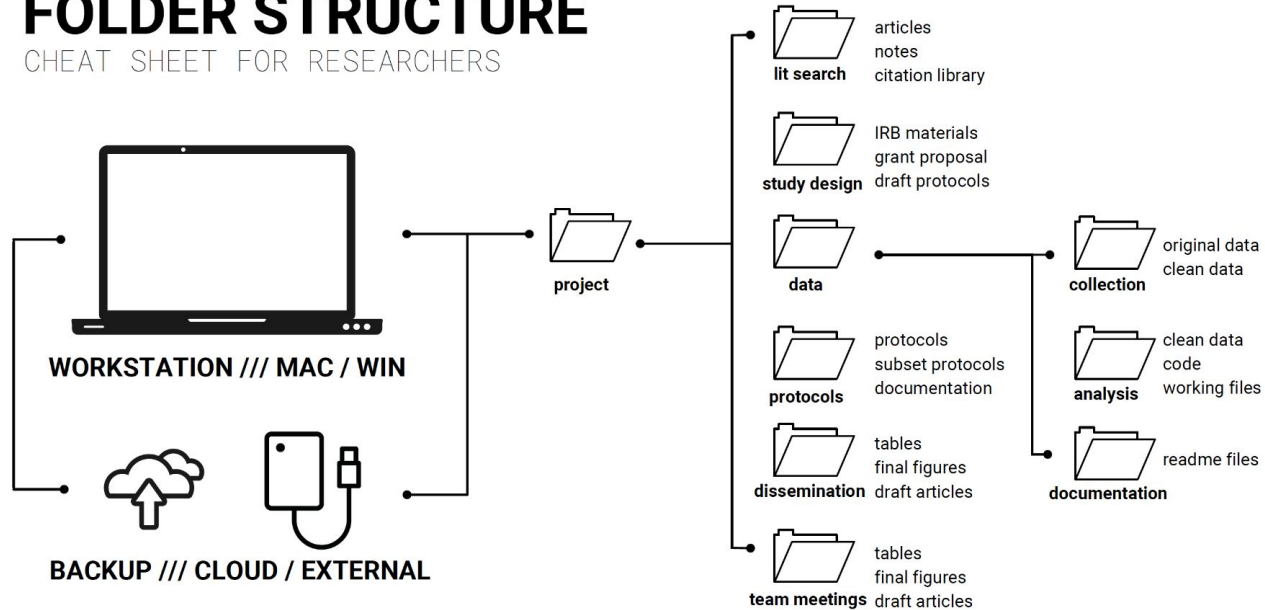
- **By stage:** protocol, data collection, data analysis, manuscript, etc.
 - **By research activities:** interviews, surveys, experiments, etc.
 - **By data/material type:** databases, text, images, models, field notes, working and posted files, etc.



Example:

FOLDER STRUCTURE

CHEAT SHEET FOR RESEARCHERS



FILE TYPES

.txt / .csv / .pdf / .rtf



IMAGE TYPES

.tif / .jpg / .png



FILE NAME RULES

- My File 11-06-2021
 - 2021-11-06_myfile
- Use lowercase
-Use underscores or hyphens
-Use standard calendar dates
-Put dates at the beginning of name
-Avoid spaces and special characters



Example:



Grant



IRB



Recruitment



Intervention
Materials



Tracking Logs



Data



Evaluation



Protocols



Team Meetings



Communication



Dissemination



Project Close Out

Folder Management Pro Tips

1. Choose a mnemonic for each project (e.g.: RDM=Research Data Management) to use in associated directories and file names
2. Create separate folders for “working” and “posted” files
3. Document your strategies in a README file



Documentation: README File

A standard document detailing information about other documents:

- Title of dataset
- Name/institution/contact information for
- Principal Investigator (or person responsible for collecting the data)
- File name structure
- Attributes: Describe the attributes used to name the files.
- Codes: Provide a complete list of any codes/abbreviations used.
- File formats
- Calculations
- Versioning



Batch Renaming Files

Sometimes you may have specific requirements for renaming large numbers of files, for example:

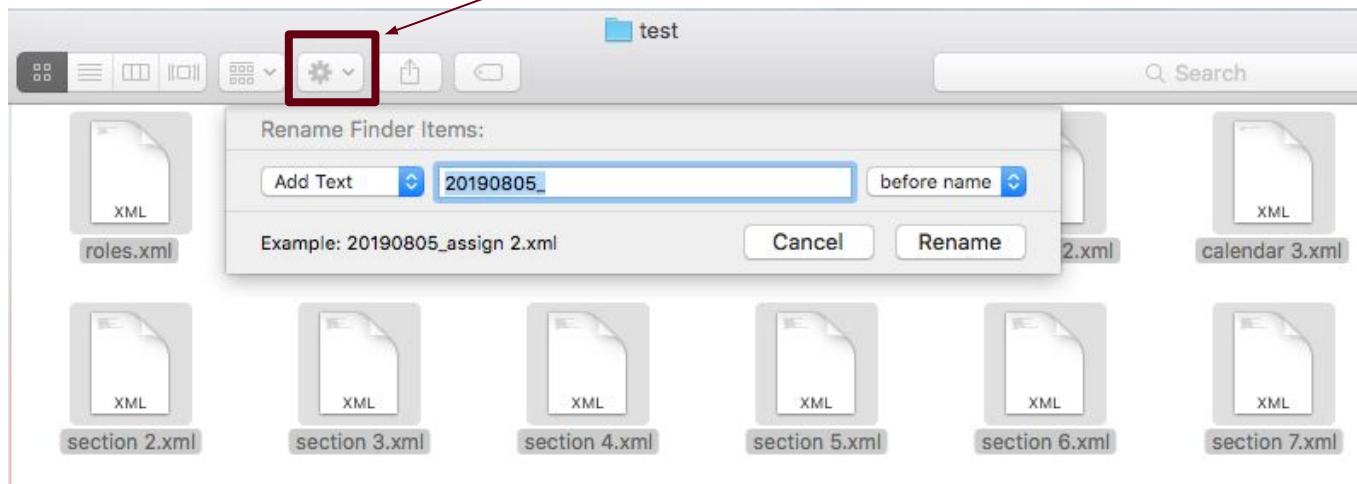
- Adding a date
- Adding a code
- Adding, replacing, or removing individual characters
- Shortening the file name by X # of characters



Batch Renaming Files

Macs:

Select all files, then select "Rename files" from settings menu



Batch Renaming Files

Windows: For advanced renaming operations, you will need to use the command line or an external program, such as [Bulk Rename Utility](#), which provides a range of options through a graphical user interface



Data Storage



Good Practice: 3-2-1 Rule

3 copies of your work

1 working copy, 2 backups

On **2** different kinds of storage

At least **1** copy off site



What are some risks to your data?



Data Risks

- Accidental sharing
- Software fail
- Natural disasters
- Pandemics
- Building collapse
- Corporate malfeasance
- Submit wrong file
- Rogue colleagues
- Theft
- Mischievous pets
- Misplaced flash drive
- Hackers



Consider

Thinking of the workflows we discussed earlier...

- What are the risks?
- What are some strategies for mitigating data loss?



Data Archiving/Publishing



Published Data = Scholarly Impact



PEOPLE

PROGRAMS

NEWS

BLOG

LEARNING

PUBLICATIONS

EVENTS

CONTACT



[Home](#) / [People](#) / [Research](#)

Melissa Meierhofer



Melissa Meierhofer

Research Associate

✉ melissa.meierhofer@ag.tamu.edu ☎ [\(979\) 862-7805](tel:979-862-7805) 📄 [Curriculum Vitae](#)

Melissa Meierhofer joined the Texas A&M Natural Resources Institute in 2015 as a research associate. Her research is focused on understanding the susceptibility of winter-roosting bats to white-nose syndrome, a cold-adapted fungus that has caused the deaths of millions of hibernating bats. She leads the bat research team in studying bat winter activity, composition, abundance and environmental conditions within hibernacula across Texas.

Prior to joining the institute, Melissa volunteered her time in the mammal and bird divisions at the Field Museum of Natural History.



LIBRARIES
TEXAS A&M UNIVERSITY

Published Data = Scholarly Impact

Texas Bats Data (Texas A and M University)

Texas Data Repository Dataverse > Texas A&M University Dataverse > Texas Bats Data > Texas Bat Habitat Data


 Contact  Share



Texas Bat Habitat Data

Version 2.0

Meierhofer, Melissa, 2019, "Texas Bat Habitat Data", <https://doi.org/10.18738/T8/BDP6XO>, Texas Data Repository Dataverse, V2

 Cite Dataset ▾

[Learn about Data Citation Standards.](#)

Dataset Metrics

62 Downloads 



Description

Dataset Averages: Dataset contains average values for ambient air temperature, bat skin temperature, substrate temperature, and vapor pressure deficit by species for each site. Data also contains ecoregion location. Datasets Cave1_30, Cave2_43, Culvert1_44, and Culvert2_19: Datasets contain Internal air temperature averages per hour and external temperature averages per hour. These data were used to create Figure 2. (2019-07-29) (2019-07-29)

Subject

Earth and Environmental Sciences



Texas Data Repository

Texas A&M University

<https://data.tdl.org/>

- Create an account with your NetID
- Share with your team and set roles
- Simplified versioning
- Provides a persistent identifier (DOI)

Possible Action Steps

- Inventory your files
- Analyze workflow, make sure it's replicable
- Start a data management plan with file naming and organization details
- Set up a meeting with your group to decide on file organization plan



Make a Plan!

3 - 3 - 3 Action Plan for better data management

- 3 goals to accomplish over a...
- 3 month timeline with...
- 3 tools or good practices



References

Briney, K. (2015) *Research data management for researchers: Organize, maintain, and share your data for research success*. Pelagic Publishing, Exeter: UK.

Harris, R. (2017). *Rigor mortis: How sloppy science creates worthless cures, crushes hope, and wastes billions*. Hachet Book Group: New York, NY.

University of Minnesota Libraries Research Data Services. (2019). *University of Minnesota Research Data Management Boot Camp Slide Deck*. Available at: <https://z.umn.edu/exampleUMNRDMBootCamp>

The University of Edinburgh. (2020, September). *MANTRA: Research Data Management Training*. Available at: <https://mantra.edina.ac.uk/index.html>

