

Talking Research Data Management:

A Guide to *Good* Practice for Librarians

Jessica Trelogan

Research Data Services Coordinator



Objectives

Build confidence and skills

Identify gaps and challenges

Share experiences and resources

Discuss RDM service models

Build community

Roadmap

Defining things

Framing the conversation

Researcher pain points

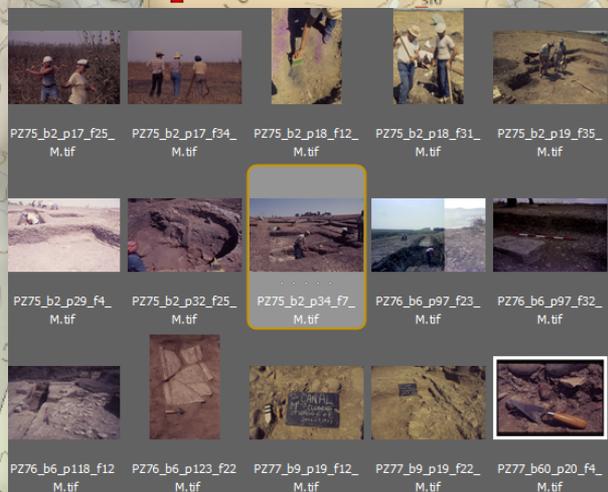
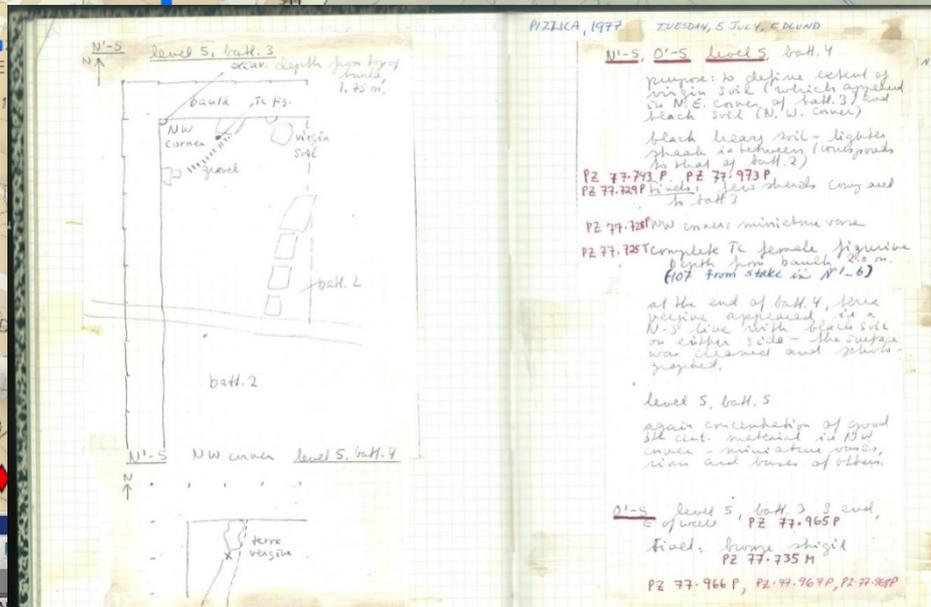
Moving the needle

Implementing services at UT



Defining things

What are "data"



loc_ID	LA							Year	Weathering
3	Uncultivated		None	30	75	Small cemetery	Heavy	100	2005 Heavy
16	Orchard		Normal plowing	10	70	Small cemetery	Light	100	2007 Light
42	Scrub		None	75	60	Small cemetery	Very light	100	2007 Heavy
25	Olive	Plowed	Harrowing	80	60	Small cemetery	Light	100	2007 Light
40	Scrub		None	50	20	Farmhouse	Light	95	2007 Heavy
3	Uncultivated		None	30	60	Other agrarian	Very light	95	2005 Heavy
26	Olive	Plowed	Harrowing	80	60	Small cemetery	Light	95	2007 Light
20	Olive	Plowed	Harrowing	65	65	Kiln/industrial	Moderate	95	2007 Light
96	Plowed		Normal plowing	50	0	Farmhouse	Light	90	2007 Moderate
23	Plowed		Harrowing	35	65	Small cemetery	Light	90	2007 Light
41	Scrub		None	75	25	Other agrarian	Light	90	2007 Heavy
1008	Plowed		Harrowing	20	40	Farmhouse	Very light	90	2007 Light
1002	Orchard		Harrowing	5	70	Other agrarian	Very light	90	2006 Moderate
1024	Plowed		Harrowing	35	65	Other agrarian	Light	90	2007 Light
959	Uncultivated		None	50	50	Other agrarian	Heavy	85	2005 Heavy

(1 out of 108 Selected)
sites 1981-2003 sites 2005 +

Where do YOU fit in?



Liaison skills and mindsets translate to RDM support!



What are “data”?

Natural/Physical Sciences

Observational

Experimental

Simulation

Compiled

Social Sciences

Qualitative

Quantitative

Humanities

Raw

Primary

Interpretive/Derived

National Science Foundation

“...determined by the community of interest through the process of peer review and program management. This may include, but is not limited to: **data**, publications, samples, **physical collections**, software and models.”



National Institutes of Health

“Recorded factual material commonly **accepted in the scientific community** as necessary to validate research findings....”



What is Research Data Management?



A collection of tasks practiced throughout the lifecycle of research that make it easier to find, understand, navigate, and use your data.



Framing the conversation

Why bother?



Save time and money

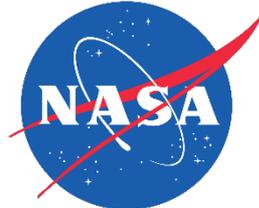
Maximize your impact

Allow for reuse

Do better research



Why else?



NATIONAL ENDOWMENT FOR THE
Humanities



GORDON AND BETTY
MOORE
FOUNDATION



INSTITUTE of
Museum and Library
SERVICES



wellcometrust

It's required.



Data Management Plans

A data management plan (DMP) is a formal but **living document** that describes the **nature** and **structure** of the data you will likely use or produce in the course of research, along with your **strategies** for dealing with it **throughout and after** a project.



Common Elements of a DMP

1. Data description
2. Data documentation
3. Access, sharing, re-use
4. Storage and backups
5. Preservation and archiving
6. Resources and responsibilities





Researcher pain points

Creating and Collecting



Photo by Jessica Trelogan

Test **everything!**

Automate where possible, but be mindful.

Create data snapshots

Be selective **before** collecting

Ensure compliance



Finding and Re-using

Find the right data

- **Talk to your liaison librarian!**
- Research data repositories: re3data.org

Integrate and tidy

- [OpenRefine](http://OpenRefine.org) for clean up
- R and the [tidyverse](http://tidyverse.org)
- Python (pandas)

Know your sources

- Restrictions
- Copyright v. ownership
- Data citation
datacite.org



Organizing

Pick a system that works for **you**

Write it down as convention

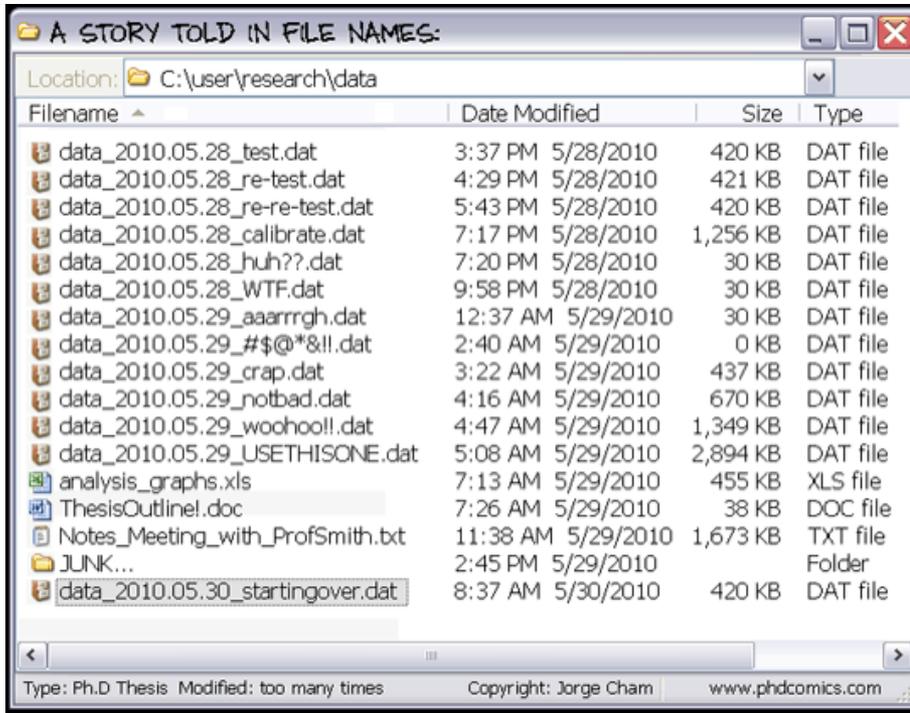
Make using it a habit



condensedconcepts.blogspot.com/2009_09_01_archive.html



File Names and Hierarchies



Be descriptive, not generic

Include dates

Avoid special characters

Describe your convention

Use batch tools or scripts

<http://www.phdcomics.com/comics/archive.php?comid=1323>



File Formats



Non-proprietary, open standards



Used commonly in your domain



Encoded with standard characters



Uncompressed (if you can afford it)



Metadata



Photo courtesy Institute of Classical Archaeology

Document! Document! Document!

Data are only useful if understandable

Strive for **FAIR** but ...

Minimum effort is better than none!

README.txt

Dublin Core 15



FAIR Data Principles

Findable: Both data and metadata should be easy to find by both humans and computers.

Accessible: Once the user finds the required data, she/he needs to know how can they be accessed.

Interoperable: Data usually need to be integrated with other data; they also need to work with applications or workflows for analysis, storage, and processing.

Reusable: The ultimate goal of **FAIR** is to optimize data reuse. Metadata and data should be well-described so that results can be replicated and/or combined in different settings.





Moving the needle

Why share?



Fosters public **trust**

Maximizes funder investment

Supports **transparency** and reduces misconduct

Encourages new research and collaboration

Increases visibility, citation rates, and **impact**

Teaching and learning



Why a repository?

Data are **citable**

Work is more **discoverable**

Persistent IDs for deposits

Some may help with **curation** and ingest

Preservation and **maintenance**

Funding agencies expect it



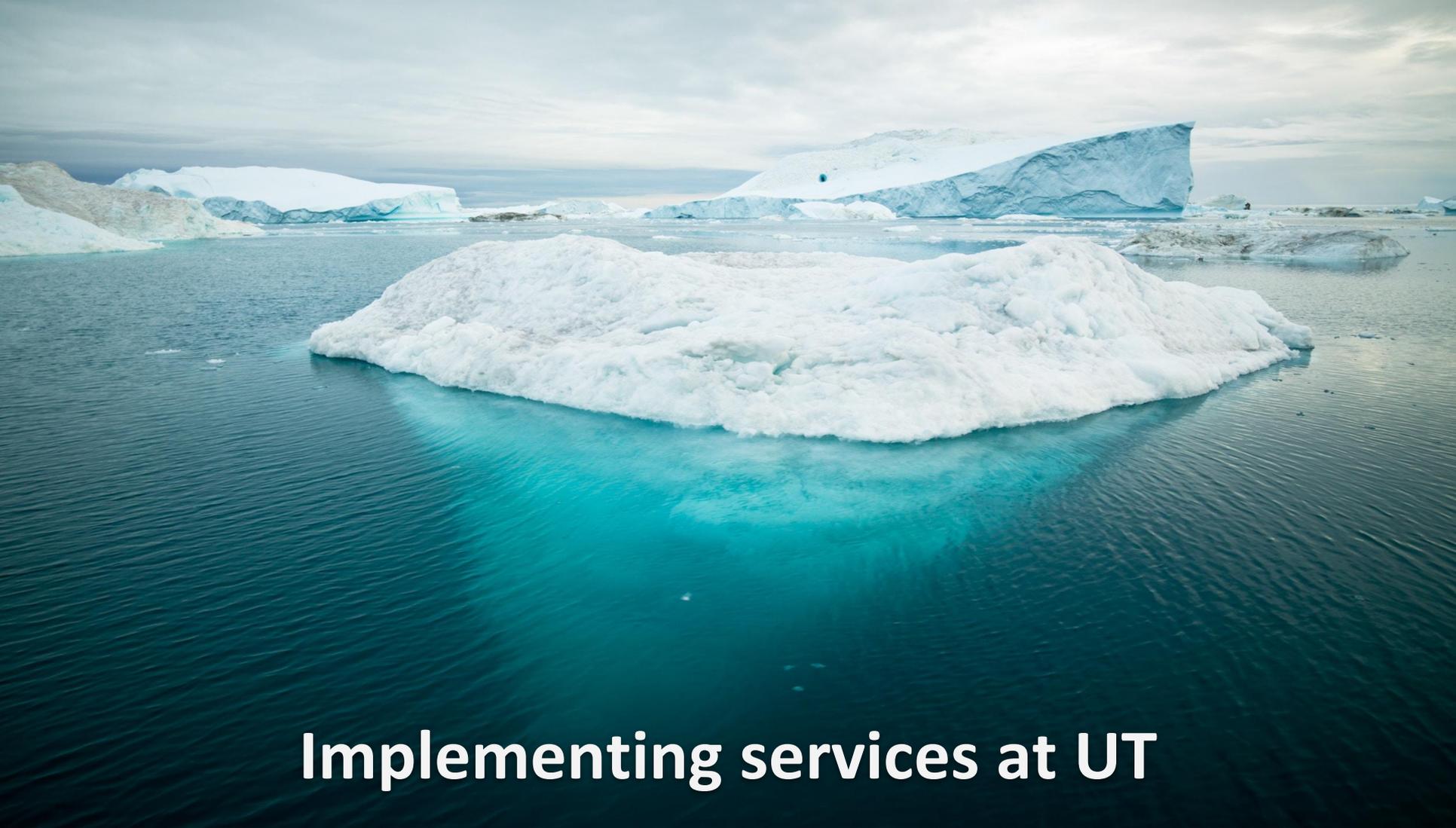
Types of repositories

Institutional repositories (e.g. [Texas ScholarWorks](#))

Disciplinary repositories (e.g. [ICPSR](#), [tDAR](#))

Open repositories (e.g. [Dataverse](#), [figshare](#))





Implementing services at UT

Building Research Data Services

Data management since 2011

Dedicated person since 2016

Added GIS person in 2018



Partnerships

Research Computing

Texas Advanced Computing Center
Central ITS
GIS Administrators

Grants

Office of Sponsored Projects
Vice President for Research
Grants administrators

Students: ~51,300
Teaching faculty: ~3,000
Libraries staff: ~230
Librarians: ~70
RDM staff: 2.5



Texas Data Repository



Consortial data repository using [Dataverse](#)

Launched in 2017

Hosted by the [Texas Digital Library](#)

Steering committee of institutional liaisons

[11 participating institutions](#) in Texas

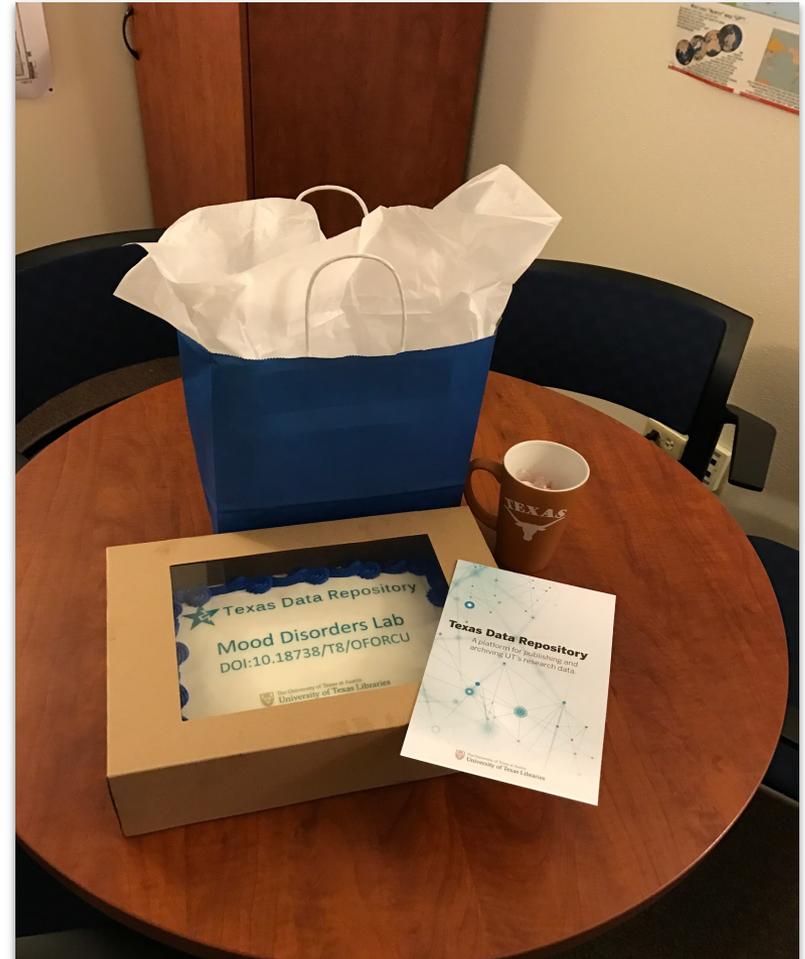


Getting the word out

Shout loudly, scale later

Training (in and out)

Team effort



Training

Piloting and assessing

Building community

Partnering

Data & Donuts

Want to learn how to take better care of your research data? Love donuts? Research Data Services is hosting a special Workshops @ UT Libraries series aimed at promoting good data management practices throughout the research lifecycle.

Join us at the The

Managing Research: A Guide to Good Practice
Jessica Trelogan, Research Data Services, UT Libraries
September 14, 2018

Metadata Basics
Melanie Cofield, Head of Research Data Services, UT Libraries
September 21, 2018

Designing Surveys and Assessments
Krystal Wyatt-Baxter, Head of Research Data Services, UT Libraries
September 28, 2018

For more information contact Jessica Trelogan at jtrelogan@atstin.utexas.edu | 512-495-1234

Data & Donuts Geospatial Edition

Want to learn how to realize the full potential of your geospatial data? Love donuts? Research Data Services is hosting a specially-themed Data & Donuts workshop series that is aimed at helping you develop the essential software skills to get the most out of your data.

Drop in at 3pm

Intro to GIS and Geospatial Data
Michael Shensky
February 1, 2019
PCL Learning Lab 1

Finding Geospatial Data Online
Michael Shensky
February 8, 2019
PCL Learning Lab 1

Managing Data with GIS Software
Michael Shensky & Jessica Trelogan
February 15, 2019
PCL Learning Lab 2

R for Geospatial Analysis
Emily Beagle
February 22, 2019
PCL Learning Lab 1

Designing ArcGIS Online Visualizations
Michael Shensky & Albert Palacios
March 1, 2019
PCL Learning Lab 1

Map Making with GIS Software
Michael Shensky
March 8, 2019
PCL Learning Lab 1

Jessica Trelogan, Research Data Services, UT Libraries
jtrelogan@atstin.utexas.edu

Data Carpentry Workshop

Want to learn how to realize the full potential of your research data? Interested in developing new skills with Python, SQL, and Open Refine? Save the date for the August 2019 Data Carpentry workshop co-hosted by the UT Libraries and Texas Advanced Computing Center (TACC).

Thu. (8/22) & Fri. (8/23)
8:30am - 4:00pm

Perry-Castañeda Library (PCL) Learning Lab 4

For more information about workshop content & registration:
<http://bit.ly/utdc819>

 The University of Texas at Austin
University of Texas Libraries

 TACC
TEXAS ADVANCED COMPUTING CENTER

 DATA CARPENTRY





Questions?

