



# DATA INTEGRITY: ACQUISITION, MANAGEMENT, SHARING, AND OWNERSHIP

---

RESPONSIBLE CONDUCT OF RESEARCH

XUAN ZHOU, PHD  
DATA CURATION SPECIALIST  
UNIVERSITY LIBRARIES  
OCTOBER 11, 2023

## Warm Up Questions

1. Have you ever been asked to share your data with your team members or a journal publisher?
2. Have you ever submitted a data management plan to a funding agency?
3. Have you ever used a research data repository to publish or share your data?



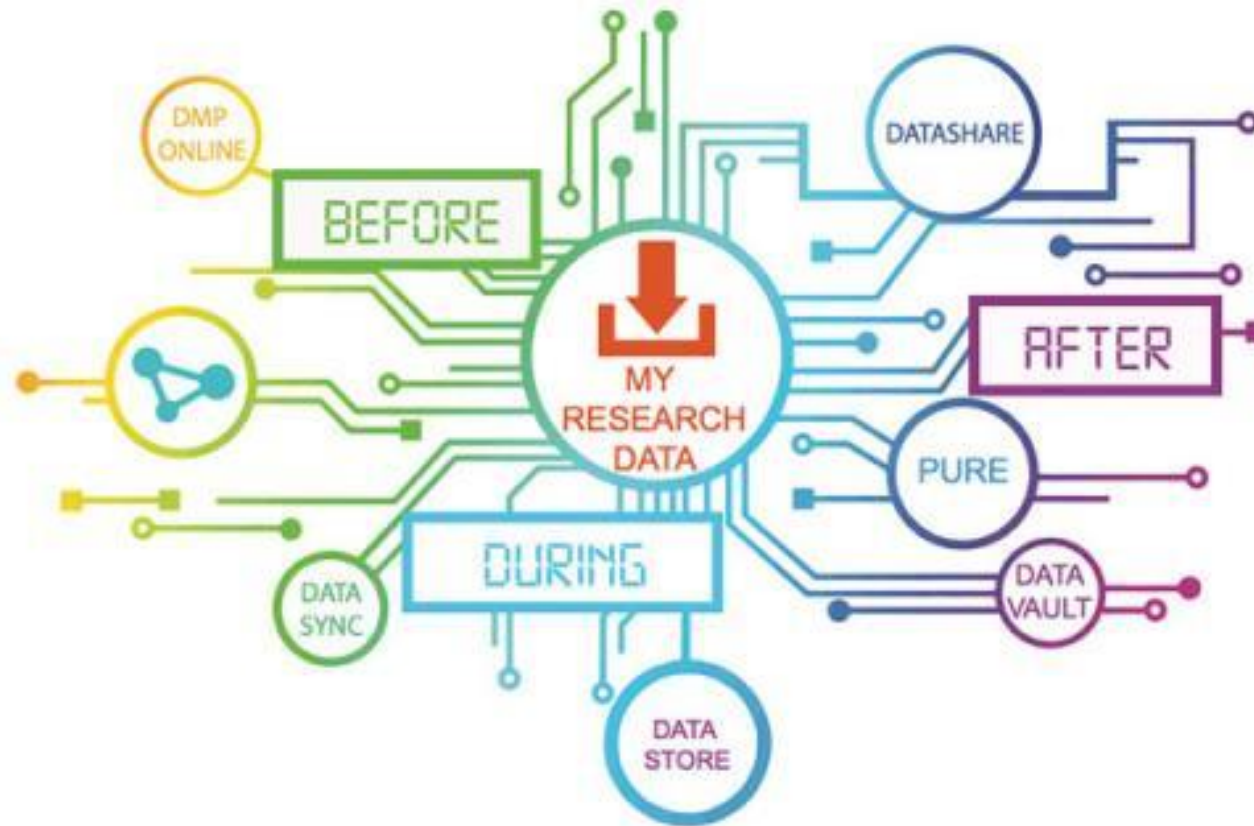
# Goals for Today

---

- ❖ Understand the research data life cycle
- ❖ Know DMPTool to create a quality DMP
- ❖ Recognize the importance of Research Data Management (RDM)
- ❖ Know good practices and available resources of RDM at TXST

# What do we mean when we talk about research data?

---



# What data do you use and create?

---

**Research Data** is recorded, factual material commonly accepted in the scientific community as necessary to validate research findings. (Awasthi & Tripathi, 2019)

Numeric data  
Spreadsheets  
Binary files  
Code

PDFs  
Image files  
Audio files  
Physical specimens  
Archival materials  
Geospatial data

Or something else



# Research Scenario

---

You published a study on a high-prestige journal. This study has been cited widely by others who have built upon your findings. However, three years later another researcher said, after several replications he could not get the similar results as found by your study and he has accused you of having falsified the data.

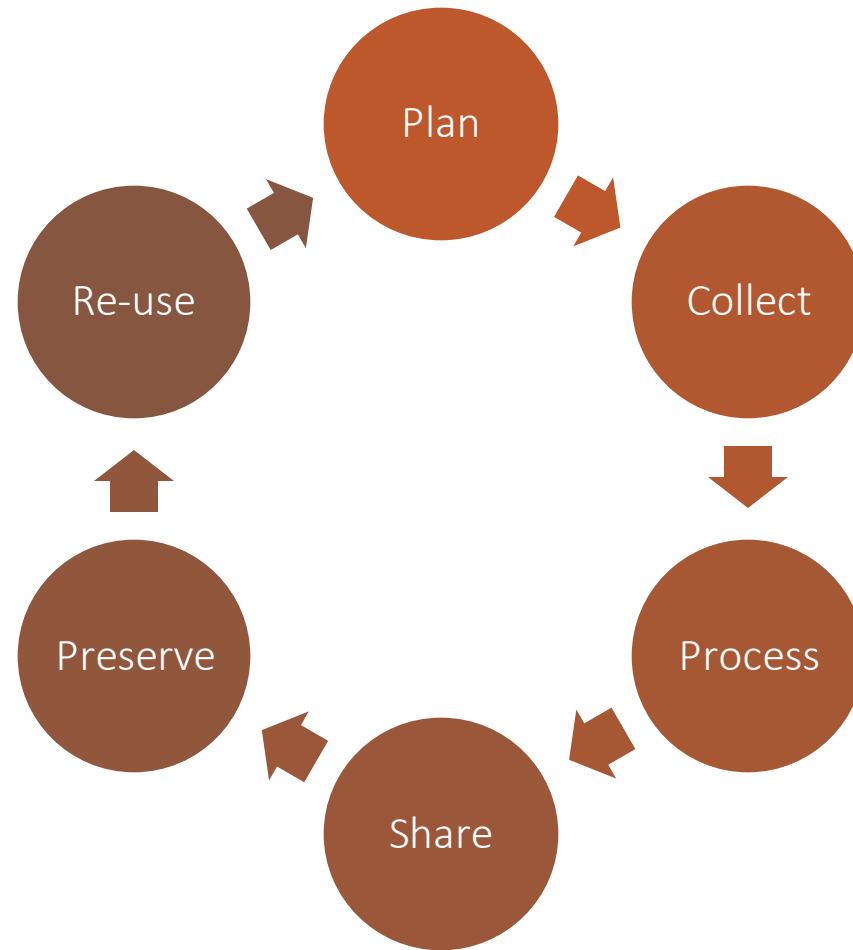
# What is research data management?

---

**Research Data Management (RDM)** is the organization, management, publication, and preservation of the products of research.

Mandate	Facilitate	Reuse	Impact
Meet requirements and expectations set by funding agencies, publishers and domain associations	Ensure that your data is complete, documented, and accessible to you and to future researchers	Encourage the discovery and reuse of your data to further discoveries in your field of research	Receive credit for your data and increase its impact and visibility

**Beneficial to you and your research in a long run!**



# Research Data Lifecycle



# Start With a Data Management Plan (DMP)

---

- Funder requirements
- Institutional policy
- Mitigate error and loss
- Avoiding unforeseen costs
- Be able to return the data
- Getting a handle on the complexity of data



National  
Science  
Foundation

# Basic Elements of a DMP

---

Simply a 1-2 page summary explaining how you are planning to manage the data gathered in the course of your research project.

1. What are you creating / generating?
2. How is it securely handled during the project?
3. How is access and data integrity maintained long term?
4. Additional Details: Roles and responsibilities, systems used, documentation, security

# NSF general DMP guidelines

---

- The **TYPES** of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- The **STANDARDS** to be used for data and metadata format and content;
- Policies for **ACCESS** and **SHARING** including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for **RE-USE**, re-distribution, and the production of derivatives; and
- Plans for **ARCHIVING** data, samples, and other research products, and for preservation of access to them.

Based on [NSF general DMP guidelines](#)

**Always consult the specific data management requirements for your funding agency to write your DMP.**

# Planning and Writing Resources

---

- Data Management Plan
- DMPTool
- Sample text & templates
- Login with TXST Email
- <https://dmptool.org/>



- 🌐 TXST Dataverse Repository
- 🌐 TXST Research Data Management Services
- 🌐 TXST Data Management Planning
- ✉ TXST Data Contact

# Templates for TXST Researchers

Template Text: Data Sharing and Access

Template Text: Metadata for Data Management

<https://guides.library.txstate.edu/research-data/DMP>

## University Libraries: Researcher Support

Search this G

Guidance on support and resources through the University Libraries for researchers at all stages of the research lifecycle process.

Home

Research Data Management

Data Management Plan  
Templates

Sharing Research Work

Digital Publishing Services

Teaching and Learning

Open Education

Technologies and Micro-  
Credentials

Special Collections and Exhibits

### Texas State University Template Language for DMPs

Applicants for Federal funding requiring Data Management Plans (DMP) may incorporate or adapt language in their plans if they intend to use the [Texas State University Research Data Repository](#) research data. Principal Investigators who intend to use the Data Repository can start the process with the Research Data Services Department at [UL-RDS@txstate.edu](mailto:UL-RDS@txstate.edu).

Texas State University provides Data Management Plan development support with the DMPTool.

### Data Sharing and Access: Template Text

The online Texas State University Research Data Repository (<https://dataverse.tdl.org/datavers>) to share datasets through the Texas Digital Library and managed by local Texas State University I Digital Library (TDL) is a consortium of academic libraries in Texas with a proven history of providing technology services to support secure, reliable access to digital collections of research and scholarly data. The Research Data Repository is a project of the TDL and its member institutions to develop a consortial state repository for researchers at Texas higher education institutions.

Data will be curated in the repository following accepted standards (NISO Framework Advisory Committee). A persistent identifier, a DOI, is created for each data set published. Datasets in the repository will be available for long term use.

The project team will work with Texas State University Research Data Service Department as needed in assigning appropriate metadata and in determining appropriate embargo periods for the individual datasets. For data sources that are embargoed for some period of time, the metadata records will be available for discovery of the resources. All project-related materials, such as technical reports, presentations, and other documents will be made accessible through the Texas State University Institutional Repository (digital.library.txstate.edu) linked with the dataset in TXST Research Data Repository accordingly.

### Metadata for Data Management: Template Text

Metadata records will be created to describe each of the project's digital resources. Metadata records will provide information on subject, provenance, authorship, methods and post-processing, and



FILE NAMING STRATEGIES



DATA STORAGE



DATA PRESERVING/  
PUBLISHING

Good Practices in RDM

# SAR\_090320.doc

---

## What does this mean?

- Survey Analysis Results?
- Survey of Agriculture Research?
- Sam A. Rodriguez, a researcher?
- September 03, 2020?
- March 09, 2020?
- March 20, 2009?

# File Naming

Two main criteria: **Context & Consistency**

## Good File Naming Practices

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent



# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

- ❖ Date (YYYY-MM-DD)
- ❖ Project name/Grant #
- ❖ Type of data
- ❖ Location/site/spatial coordinates
- ❖ Researcher info
- ❖ Version

# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

**SAR\_090320.doc ?**

**in YYYY-MM-DD format (2023-09-19)**

**Sort, with standard dates**

2023-03-16\_Code\_descriptions.docx

2023-05-24\_Code\_descriptions.docx

2023-11-03\_Code\_descriptions.docx

**Sort, without standard dates**

11-3-23\_Code\_descriptions.docx

3-16-23\_Code\_descriptions.docx

5-24-2023\_Code\_descriptions.docx

# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

## Sort, with a leading zero

Test01\_RDM assessment.xlsx  
Test02\_RDM assessment.xlsx  
Test03\_RDM assessment.xlsx  
...  
Test10\_RDM assessment.xlsx  
Test11\_RDM assessment.xlsx

## Sort, without a leading zero

Test1\_RDM assessment.xlsx  
Test10\_RDM assessment.xlsx  
Test11\_RDM assessment.xlsx  
...  
Test2\_RDM assessment.xlsx  
Test3\_RDM assessment.xlsx

# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

## ❖ Use \_ as delimiters

❖ Data\_v01

## ❖ Avoid Special Characters

Name

- 📁 Data
- 📁 Data 'in progress'
- 📁 Grad Shop Talk~
- 📁 Let's work on RDM!
- 📁 RDM workshop#1

# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

❖ Using consecutive numbering for major version changes

Code\_descriptions\_**20230919\_v01**.docx

❖ Using decimals for minor changes

Code\_descriptions\_**20230919\_v01.1**.docx

# Good File Naming Practices

---

- Use descriptive file names
- Use a standard date system
- Use leading zeros
- Use basic characters and avoid (/ , # ?)
- Version files
- Be consistent

## ❖ Consistency with Spaces

Data\_projectname**v03**.docx

Data\_**project name v01**.docx

Data\_**projectname\_v02**.docx

# Data Documentation: README File

---

**README files** are plain text documents that sit at the top level of project folders and describe the purpose of the project, contact details, and organization of files.



A standard document detailing information about the documents:

- Title of dataset
- Name/institution/contact information for
- Principal Investigator (or person responsible for collecting the data)
- File name structure and the description of the attributes used to name the files.
- Descriptions of every folder, file, format, data collection method, instruments, etc.
- Codes: Provide a complete list of any codes/abbreviations used.
- Dates/Locations of data collection
- Funding information
- People involved

# Data Storage and Backup **3-2-1 Rule**

---

3

Create at least  
three copies of  
your data



2

Store the copies  
on two different  
storage media



1

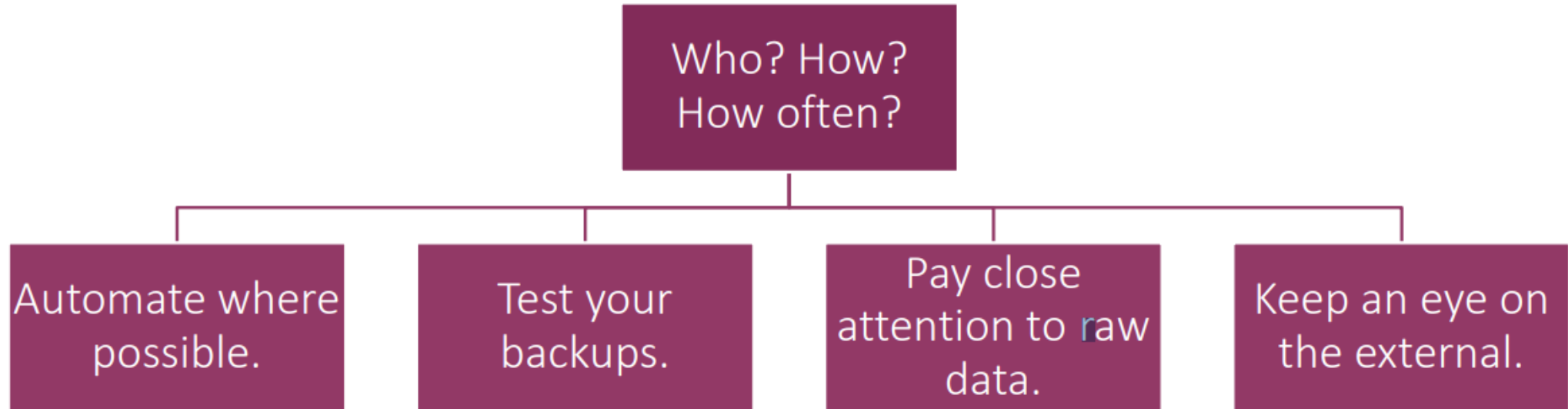
Store one copy  
on an offsite  
storage



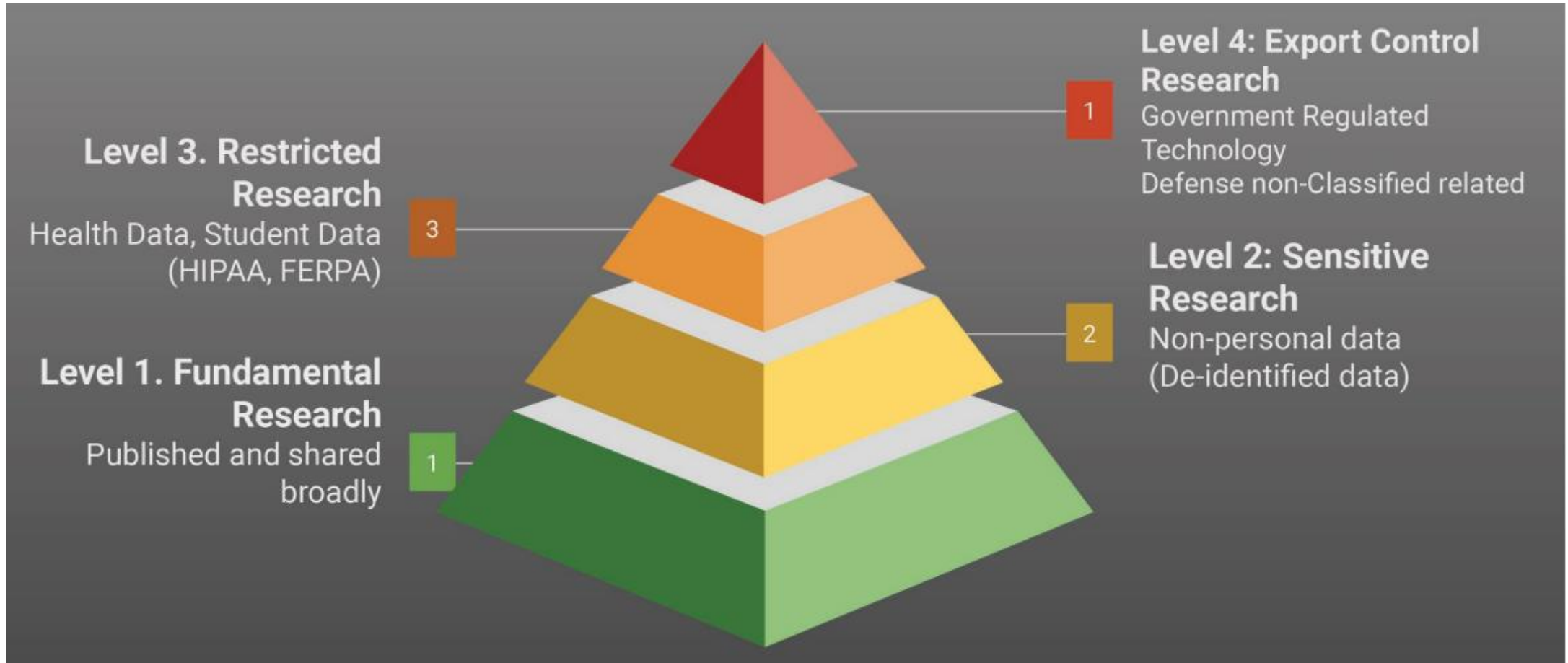


# Consideration for back-up

---



# Security Measures: Know Your Data



# TXST Data Classifications

	Confidential Information	Sensitive Information	Public Information
Level of Sensitivity	High	Moderate	Low
Legal Requirements	Protection of data is required by law (e.g., TPIA, FERPA, and HIPAA data) or contractual agreements.	Often considered “public” in the sense it is releasable under the Texas Public Information Act, some assurance is required so release of information is both controlled and lawful.	Public information by its very nature is designed to be shared broadly, without restriction, at the complete discretion of the owner.
Disclosure Risk	Confidential information presents the most serious risk of harm if improperly disclosed.	Unauthorized disclosure of Sensitive information could adversely impact the University, individuals or affiliates.	From the perspective of confidentiality, public information may be disclosed or published by any person at any time.
Examples of Information	<ul style="list-style-type: none"><li>• Social Security numbers</li><li>• Credit card info</li><li>• Personal health info</li><li>• Student records</li><li>• Crime victim info</li><li>• Library transactions</li><li>• Court sealed records</li><li>• Access control credentials</li></ul>	<ul style="list-style-type: none"><li>• Performance appraisals</li><li>• Employee DOB</li><li>• Employee email addresses</li><li>• Donor information</li><li>• Voicemail records</li><li>• Email contents</li><li>• Unpublished research</li></ul>	<ul style="list-style-type: none"><li>• Job posting</li><li>• Service offerings</li><li>• Published research</li><li>• Directory information</li><li>• Degree programs</li><li>• General information about university products and services</li></ul>

---

# Data Archiving/Publishing

---

# TXST Dataverse Repository

---



A research data management system



Add, share, publish, and manage your data



Find datasets from across Texas institutional Dataverse collections.

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



# TXST Dataverse Repository is Appropriate for:



**Data in any file type**



**Data from any field of research**



**Static or evolving datasets**



**Data without confidential or sensitive information.**



**Individual file up to 4GB (Small- Medium size)**



**Large file: consult RDM/ IT service department**

# Why TXST Dataverse Repository?

---



Provides a platform for archiving and publishing the data developed or used in support of research at Texas State University



An open access data repository for researchers affiliated with TXST



Served by RDM service team: help with DMP and preparing data to deposit




University libraries offer advice on appropriate file formats, metadata, and licensing options




Provide consultation services or training workshops for users to upload and manage their own data collections

# Find Open Data

 **Texas Data Repository**

Search ▾ About User Guide Support Xuan Zhou ▾

**Texas Data Repository** [A statewide collaboration of Texas higher education institutions](#)


 Metrics 1,178,190 Downloads [Contact](#) [Share](#)


## Share, publish, and manage your data. Find and cite data across all research fields.


Welcome to the Texas Data Repository, a research data management system for Texas Digital Library (TDL) member institutions. To add, share, and publish your data or work on a project, select your local institutional repository from the institutions below. To find datasets from across Texas institutional Dataverse collections, start [here](#).


**LEARN MORE**

- [Video Tutorials](#)
- [Go to the user guide.](#)
- [Contact a local university liaison librarian for help.](#)

  
SMU Dataverse Repository


  
Texas A&M University Dataverse Repository

  
TAMU Galveston Dataverse

  
Texas A&M International University Dataverse



# Facilitate Collaboration

 **Texas Data Repository**

Search ▾ About User Guide Support Log In


**Feed the Future Innovation Lab for Small Scale Irrigation Dataverse** (Texas A and M University) [ILSSI Website](#)


Texas Data Repository > Texas A&M University Dataverse Repository >


Contact Share


The Feed the Future Innovation Lab for Small-Scale Irrigation is a five-year project that aims to benefit farmers of Ethiopia, Ghana and Tanzania by improving effective use of scarce water supplies through interventions in small-scale irrigation. It is a part of the U.S. Government's Feed the Future Initiative.

<


  
Ethiopia Dataverse


  
Ghana Dataverse


  
Tanzania Dataverse


  
SIPSN

>

Search this dataverse...  [Advanced Search](#)


☒  **Dataverses (17)**

☒  **Datasets (75)**


☐  **Files (588)**

**Dataverse Category**  
[Research Group \(15\)](#)  
[Research Project \(2\)](#)

1 to 10 of 92 Results



**Karnali Watershed - SWAT simulated scenarios**  
Mar 24, 2022 - Nepal  
Risal, Avay, 2022, "Karnali Watershed - SWAT simulated scenarios", <https://doi.org/10.18738/T8/UI8Y4C>, Texas Data Repository, V1  
Baseline scenario Fully irrigated rice-wheat scenario Rainfed rice-wheat and rice-lentil scenario Rice-vegetable- spring rice scenario & Rice-irrigated maize scenario year of simulation : 1985-2020 (using meteorologic data), 2021- 2050 (climate data) (2022-03-23)

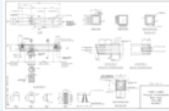


# Increase scholar impact

Texas Data Repository > Texas A&M University Dataverse Repository > TxDOT 0-6863: Pretensioned Concrete Bent Caps >

## TxDOT 0-6863: Pretensioned Concrete Bent Caps Phase 2 Experimental Data

Version 1.0



McKee, Codi D.; Lee, Ju Dong; Birely, Anna C.; Mander, John B., 2018, "TxDOT 0-6863: Pretensioned Concrete Bent Caps Phase 2 Experimental Data", <https://doi.org/10.18738/T8/CPNVA5>, Texas Data Repository, V1, UNF:6:rhD7jjUKeHH3Lu7bKo2DKA== [fileUNF]

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

[Access Dataset](#) [Link Dataset](#) [Contact Owner](#) [Share](#)

**Dataset Metrics**

- 1,271 Views
- 158 Downloads
- 1 Citation

**Description**

This dataset contains metadata and data collected during TxDOT Project 0-6863 on development of standards for precast, pretensioned concrete bent caps.

Phase 2 tests are contained in this dataset. The Phase 2 setup consisted of a longer overhang and interior span than Phase 1, allowing for application of larger moment demands. Phase 2 consisted of two specimens, both with 28 longitudinal prestressing strands, 12" spacing of shear reinforcement, and internal voids for weight reduction. Void location and details varied between the two specimens (**PSV-28A** and **PSV-28B**).

Data provided includes specimen as-built drawings, construction timeline, measured material properties, test setup details, load patterns/sequence, applied loads at key points during test, and crack data (location and width).

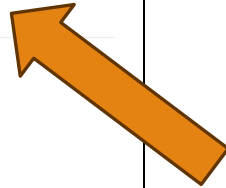
**Subject**

Engineering

**Related Publication**

Birely, A.C., Mander, J.B., Lee, J.D., McKee, C.D., Yole, K.J., and Barooah, U.R. (2018). "Precast, Prestressed Concrete Bent Caps: Volume 1 Preliminary Design Considerations and Experimental Test Program." Rep. No. FHWA/TX-18/0-6863-1-Vol1, Texas Department of Transportation and Texas A&M Transportation Institute.

**License/Data Use Agreement** [Custom Dataset Terms](#)





# Final Tips and Reminders

---

- Know your data
- Decide which data you want to share
- Choose file formats that last
- Remember the documentation
- Consider ownership and privacy
- **Make a Plan!**

# Thank you!

**Xuan Zhou, PhD**

Data Curation Specialist

[x\\_zhou@txstate.edu](mailto:x_zhou@txstate.edu)

Research Data Services

Texas State University Libraries

<https://www.library.txstate.edu/>

# Love Data Week (February 2024)

---

## The Carpentries Workshop

### Software Carpentry (R for Reproducible Scientific Analysis)

Our more introductory R lesson. In addition to our [standard content](#), this workshop covers data analysis and visualisation in R, focusing on working with tabular data and other core data structures, using conditionals and loops, writing custom functions, and creating publication-quality graphics. As our more introductory R offering, this workshop also introduces learners to RStudio and strategies for getting help. This workshop is appropriate for learners with no previous programming experience. For audiences with some experience with R or other programming languages, we recommend our [Programming with R](#) lesson.

### Software Carpentry (Programming with R)

Our more advanced R lesson. In addition to our [standard content](#), this workshop covers data analysis and visualisation in R focusing on working with core data structures, using conditionals and loops, writing custom functions, and running R programs from the command line. This is the more advanced of our two R offerings for Software Carpentry and is appropriate for learners with some previous programming experience, in R or other languages. For audiences with no previous programming experience, we recommend our [R for Reproducible Scientific Analysis](#) lesson.



THE  
CARPENTRIES

# SUPPLEMENTAL INFO & ACTIVITIES

---

Who owns the data?

It may depend on who sponsors the research.

As employees of the university, you are working for hire for the university, which, in most cases, owns the rights to the data. In federally sponsored research, the university owns the data but allows the principal investigator on the grant to be the steward of the data. The PI takes responsibility for the collection, recording, storage, retention, and disposal of data.