AnSRS4U

Sponsored Research Services

DIVISION OF RESEARCH
TEXAS A&M UNIVERSITY

Research Data Lifecycle

Ateanna Uriri (ateauriri@tamu.edu), Data Planning and Presentation Librarian

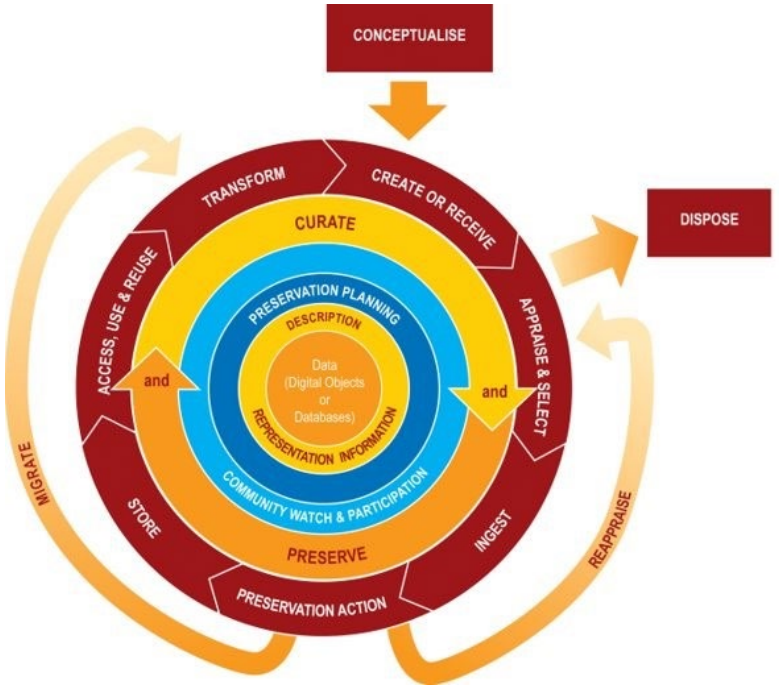
Overview

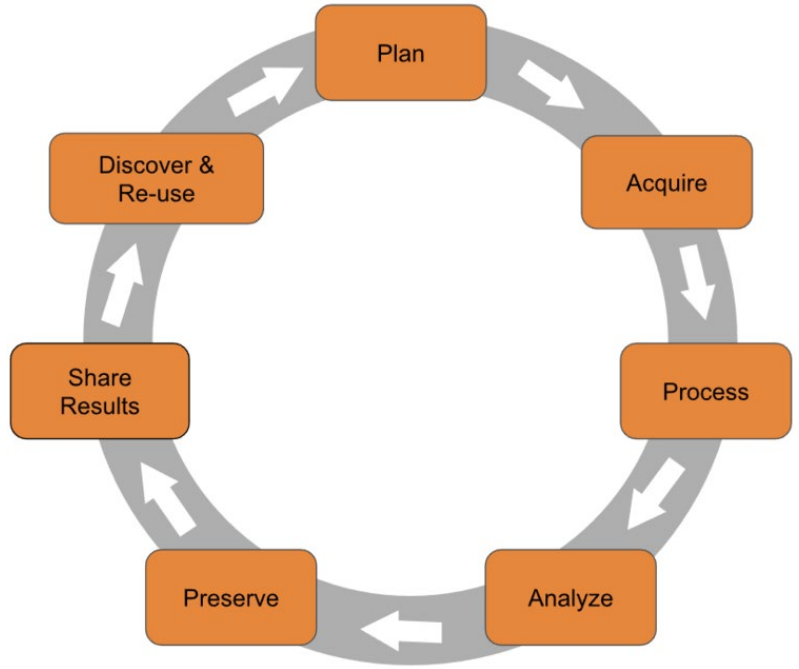
- What is the data lifecycle
- Why the lifecycle is helpful in research data management
- Stages of the lifecycle

What is the data lifecycle?

A **data lifecycle** illustrates how **data** (in all its various forms and derivatives, including data points, datasets, databases, data files, visualizations, and code) conceptually flows through its lifecycle of usefulness. While data lifecycles are helpful frameworks to discuss appropriate actions taken at different stages, it's important to remember that for most data the path is not linear, and some actions may not occur at all (University of Wisconsin-Madison).

Lifecycle Diagrams





From left to right: Research Data Lifecycle by LMA Research Data Management Working Group; Curation Lifecycle Model by Digital Curation Centre; Research Data Lifecycle by Princeton Research Data Service

The Lifecycle and Data Management

- Identify the resources and services that provide research support
- Find and understand data when needed
- Comply with funder and institutional requirements (TAMU SAP 15.99.03.M1.03)
- More easily validate results and write up research results for publication
- Reduce the risk of lost, stolen, or misused data
- Maintain project continuity through researcher or staff changes
- Share data, leading to collaboration and greater impact

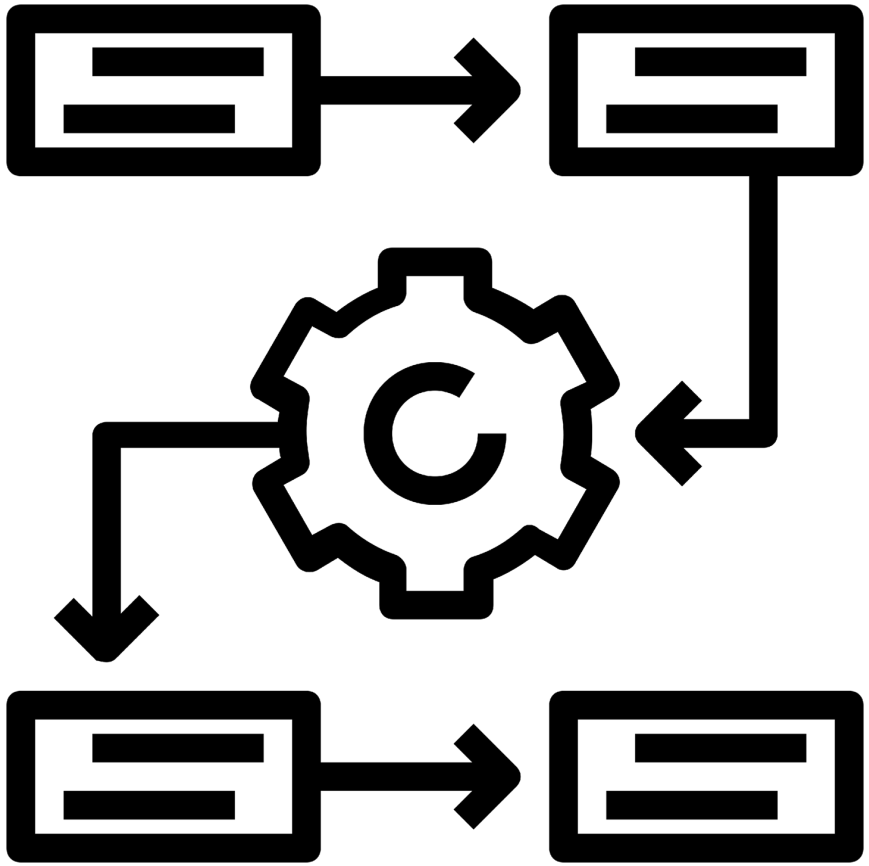
Planning

In the planning stage, researchers will work to create several plans that establish a foundation for the research project – plans for the overall project, who will be responsible for the different components of the project (roles and responsibilities), and how data will be organized and stored, for example.

Collecting

The active stages of the project start with the collecting stage, when data are gathered. To assist in useful data management, data collected must be properly organized, formatted, and documented. Clear and concise file names, stable data formats, the inclusion of metadata, and codebooks are examples of actions that are taken during the collecting stage.

Processing



Processing denotes actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.

Analyzing

Data analysis is the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making (Brown, 2014). Analysts and data scientists use different tools and strategies to conduct these analyses. Some of the more commonly used methods include statistical modeling, algorithms, artificial intelligence, data mining, and machine learning.

Sharing

Sharing often occurs during and after active stages of the project. The data may be distributed to others or published. In addition to publishing the results of your research, you'll also need to decide what will happen to the data themselves after the end of the project and how you can be sure that the data will be available in the future.

Questions to consider - Sharing

- Which repository am I going to store my data in? (TAMU Libraries – Data Repositories LibGuide)
- Am I going to publish my data and make it publicly accessible and citable?
- How will I make sure other researchers can find my data?
- Do I know which data should be stored and/or published?
- Do the data comply with any applicable confidentiality regulations or other restrictions? (TAMU Data Classification Guide; TAMU Data/Research Tools)
- Is there sufficient documentation/metadata to ensure that the data can be reused in the future?

Questions from Princeton Research Data Service Lifecycle Guide

Archiving

During the archiving step, data is deposited into a data archive or repository with the goal of long-term preservation. The deposit site may be the same as the site from the sharing/publishing stage or can be an archive/repository that is better suited for the long-term because of stability, data size, or discipline.

Other Stages

Transform – creating new data from the original

Reappraisal – periodically reviewing data and applicable policies to ensure that data continue to be accurate, secure, useful, and compliant

Migration – changing the format or storage location of data to maintain security, accessibility, and integrity

Disposal/Destruction – data that has not been selected for long-term preservation may either be transferred to another person's/institution's custody (e.g. cold storage) or securely destroyed

Summary

- Good practices will make it easier to identify your workflow and the decisions you've made during your research.
- While the data lifecycle – both in the description of stages and diagrams – are a useful tool in research data management, the process is not linear and there may be stages that are not applicable to your data management process.
- Reaching out to relevant parties – colleagues and college/departmental support, the Division of Research, Research Support and Data Services at the University Libraries, and IT – can help you navigate research data management and provide resources to support the process before, during, and after.

References

- *Biomedical Data Lifecycle.* (n.d.). Longwood Research Data Management. https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle
- Brown, M. S. (2014). Transforming Unstructured Data into Useful Information. In *Big Data Mining and Analytics* (pp. 227–246). https://doi.org/10.1201/b16666-14
- *Curation Lifecycle Model | DCC*. (n.d.). https://www.dcc.ac.uk/guidance/curation-lifecycle-model
- *Research Lifecycle Guide.* Princeton Research Data Service (n.d.). https://researchdata.princeton.edu/research-lifecycle-guide
- *Introduction to the Data Lifecycle*. (n.d.). University of Wisconsin-Madison. https://data.wisc.edu/data-literacy/lifecycle/

Data Management Plans & DMPTool

Dianna Morganti, MLIS, PMP®

Texas A&M University

Overview

- Why is data management important?
- What is a data management plan?
- Use the DMPTool

Overview

- **Why is data management important?**
- What is a data management plan?
- The DMPTool

Why is data management important?

Data management is a set of practices across the research lifecycle that ensures:

- Compliance for grant requirements
- Research integrity and reproducibility
- Research efficiency and accuracy

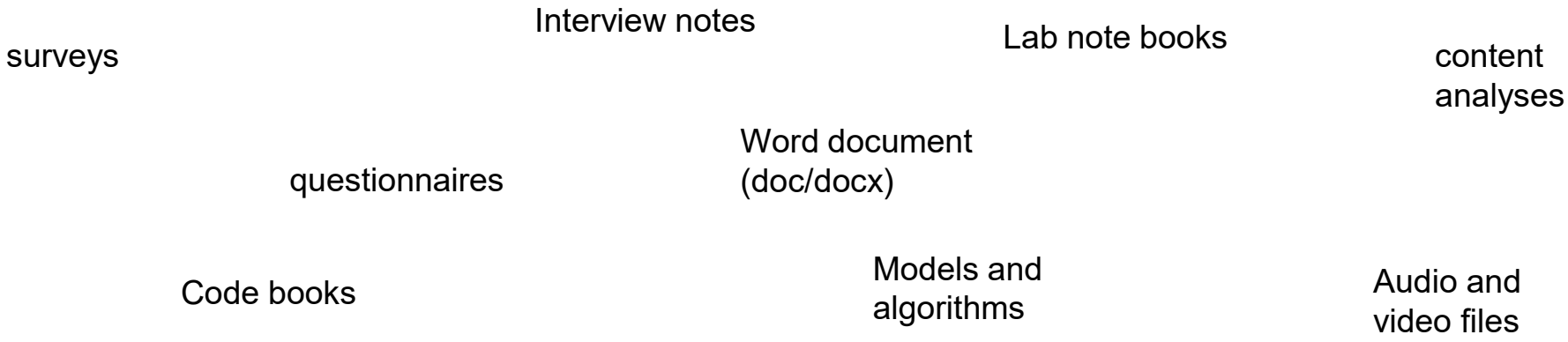
Overview

- Why is data management important?
- **What is a data management plan?**
- The DMPTool

DMPTool
Build your Data Management Plan

What is research data?

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

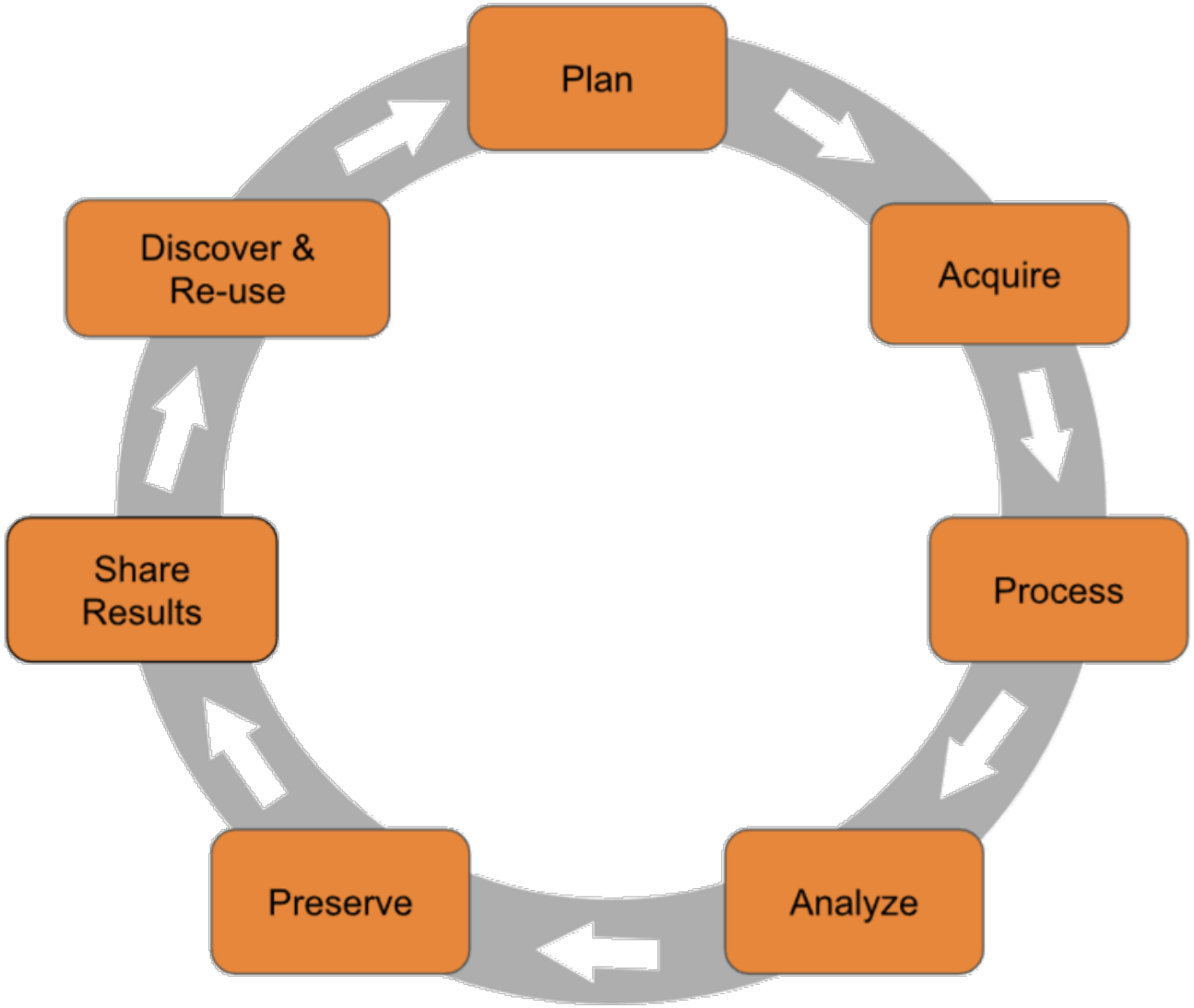


DMPTool
Build your Data Management Plan

What is research data management?

Research data management (RDM) is about handling research data effectively and appropriately throughout **the life of a research project and beyond**. Research data management refers to all aspects of creating, storing, sharing and archiving data and is an essential aspect of conducting responsible research.

Research Data Life Cycle



Research Data Lifecycle by Princeton Research Data Service

DMPTool
Build your Data Management Plan

Components of a Basic DMP

- Describing the Research Data
- Data Standards and Metadata
- Data Storage and Access
- Intellectual Property and Re-use
- Publishing Data and Preservation

DMPTool
Build your Data Management Plan

Components of a Basic DMP

- **Describing the Research Data**
- Data Standards and Metadata
- Data Storage and Access
- Intellectual Property and Re-use
- Publishing Data and Preservation

DMPTool
Build your Data Management Plan

Describing the Research Data

Give a brief description of the data. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

Data Type	Description	Potential Size of Data Files	File Format	Other Necessary Information
Powerpoint file				
Audio file from workshop				
Survey instrument				
Survey responses – alphanumeric.				

Data Type	Description	Potential Size of Data Files	File Format	Other Necessary Information
Powerpoint file	PPT file used for instruction, which was originally created by Zhihong Xu	Exactly 230mb	.pdf	Not available for re-use, just provided for information – see (link to the original PPT) for licensing information
Audio file from workshop				
Survey instrument				
Survey responses – alphanumeric.				

Components of a Basic DMP

- Describing the Research Data
- **Data Standards and Metadata**
- Data Storage and Access
- Intellectual Property and Re-use
- Publishing Data and Preservation

DMPTool
Build your Data Management Plan

Data Standards and Metadata

Describe the types of documentation that will accompany the data to help other users to understand and reuse it. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.

Data Standards and Metadata

Data Type			Standards	Documentation and Metadata
Powerpoint file			Will follow WCAG 2.1 AA web content accessibility standards.	
Attendance information			Will follow FERPA guidelines, using Zoom's "attendance" feature, counting 80%	Will include pseudonym, time spent in session, and score on follow-up assessment (e.g. survey)

Components of a Basic DMP

- Describing the Research Data
- Data Standards and Metadata
- **Data Storage and Access**
- Intellectual Property and Re-use
- Publishing Data and Preservation

DMPTool
Build your Data Management Plan

Data Storage and Access

Consider where data will be stored and how you will control access to the data. Consider where, how, and to whom data with acknowledged long-term value should be made available.

Data Storage and Access

Consider where data will be stored and how you will control access to the data. Consider where, how, and to whom data with acknowledged long-term value should be made available.

During the project, data will be secured and stored on _____, following TAMU data security policy TAMUS 13.1.

After project completion, all instructional material, the audio recordings, and the pseudonymized student data (listed in Table 2 of this document) will be shared in the Texas Data Repository within 6 months of project completion.

DMPTool
Build your Data Management Plan

Components of a Basic DMP

- Describing the Research Data
- Data Standards and Metadata
- Data Storage and Access
- **Intellectual Property and Re-use**
- Publishing Data and Preservation

DMPTool
Build your Data Management Plan

Intellectual Property and Re-Use

Consider how the data may be reused. Decide which data to keep and for how long. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.

Intellectual Property and Re-Use

Consider how the data may be reused. Decide which data to keep and for how long. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.

Keep in mind: Not all data can be shared, and not all data should be shared!

More info: Sensitive Data Workshop from TAMU Libraries!

DMPTool
Build your Data Management Plan

Components of a Basic DMP

- Describing the Research Data
- Data Standards and Metadata
- Data Storage and Access
- Intellectual Property and Re-use
- **Publishing Data and Preservation**

DMPTool
Build your Data Management Plan

Publishing Data and Preservation

Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving.

DMPTool
Build your Data Management Plan

Publishing Data and Preservation

Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving.

DMPTool
Build your Data Management Plan

The data collected on employee motivation will be published at the Texas Data Repository which guarantees preservation for a minimum of ten years through the Chronopolis digital preservation network, in accordance with the organization's Data Preservation Policy (DPP-2022).

During this period, the data will be curated through Chronopolis to ensure long-term accessibility and integrity, including periodic format migration, replication across geographically distributed storage nodes, and documentation updates to maintain compliance with evolving data standards and ensure usability for future research.

Overview

- Why is data management important?
- What is a data management plan?
- **The DMPTool**

DMPTool
Build your Data Management Plan

What is DMPTool

The DMPTool is an online platform guiding DMP development according to the requirements of specific funding agencies.

Texas A&M University researchers log in with their NetID and passwords.

DMPTool
Build your Data Management Plan

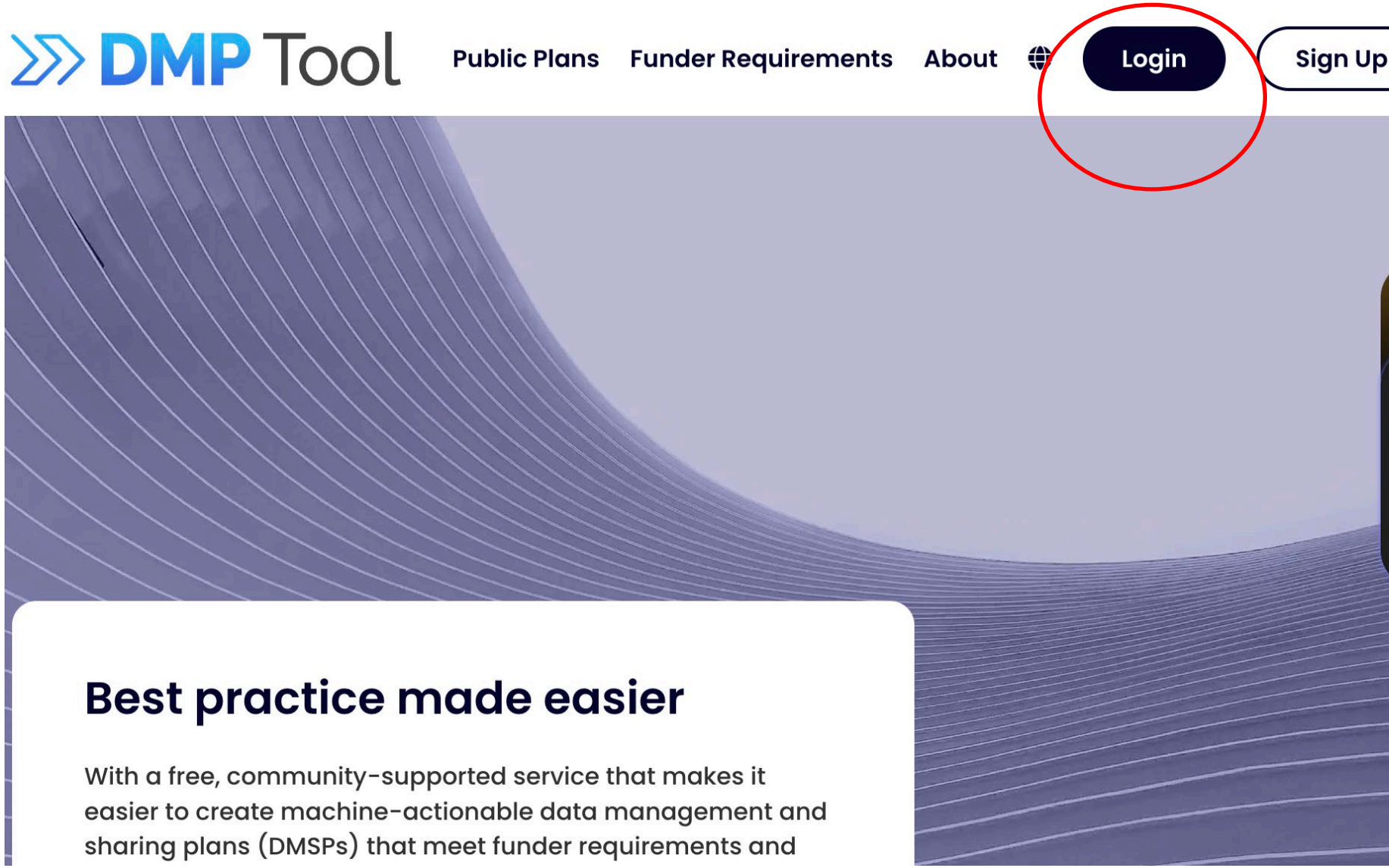
Machine Actionable DMPs

- DMPs now **interface** with funders and universities, often allowing online submission
- This allows for seamless **updating** when results are published (benefit for university and researcher)
- It also allows for **auditing** to be sure researchers fulfill their obligations for open publishing (benefit for funder)

DMPTool
Build your Data Management Plan

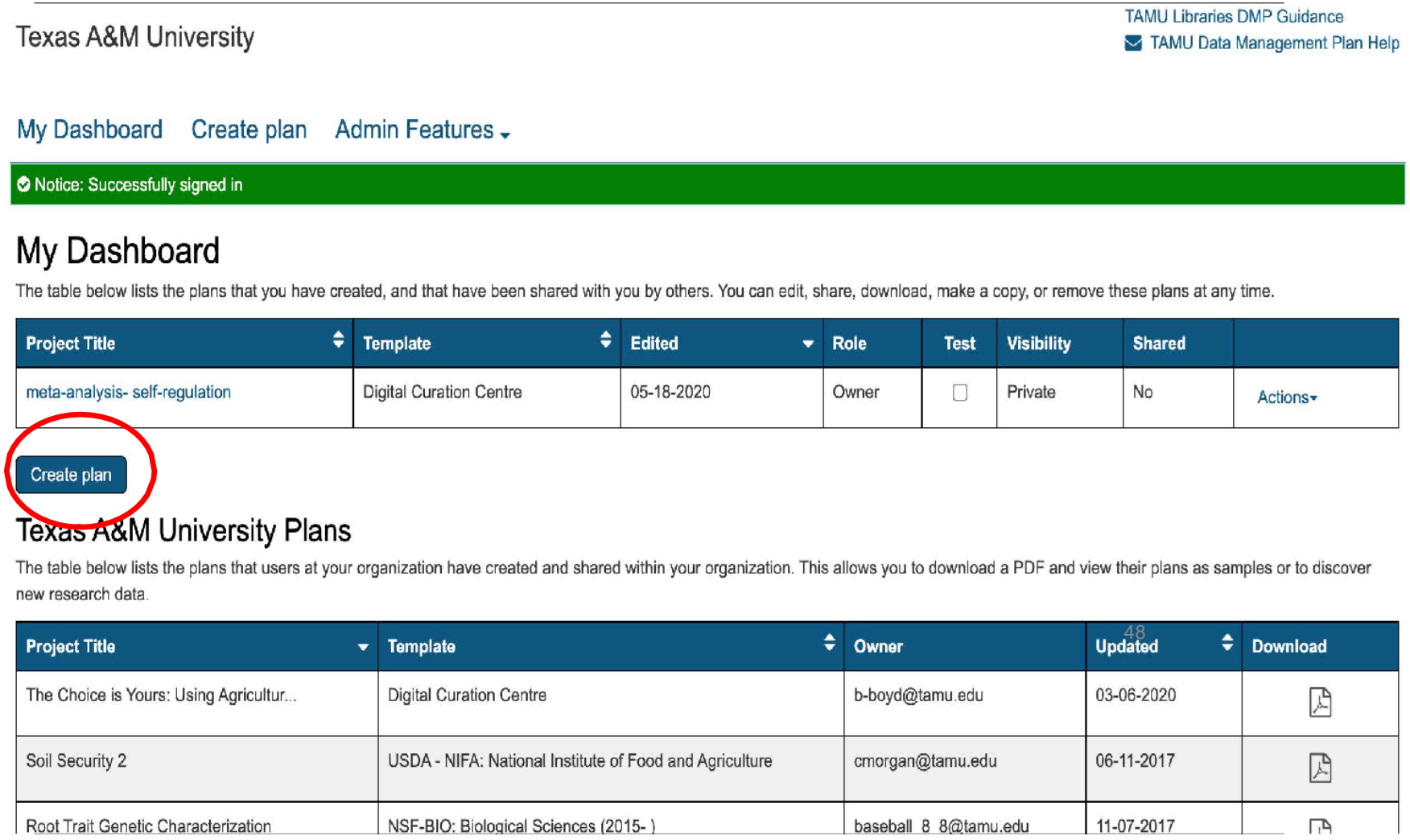
Using DMPTool

Go to https://dmptool.org/ and sign in through ‘Your Institution’.



DMPTool
Build your Data Management Plan

Create a plan



DMPTool
Build your Data Management Plan

LIBRARIES
TEXAS A&M UNIVERSITY

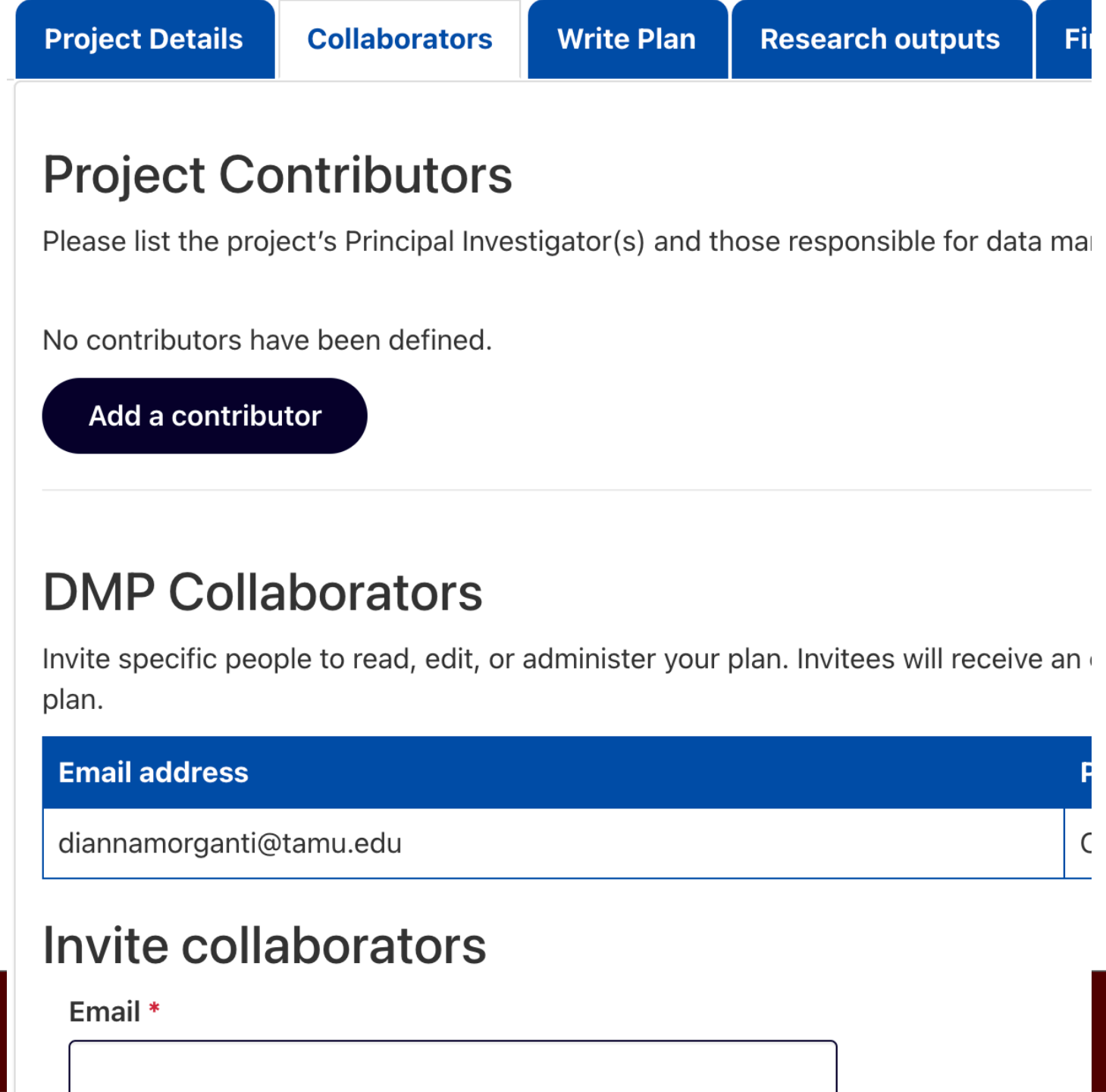
Collaborators & Contributors

Contributors will be acknowledged by the DMP output as copauthors.

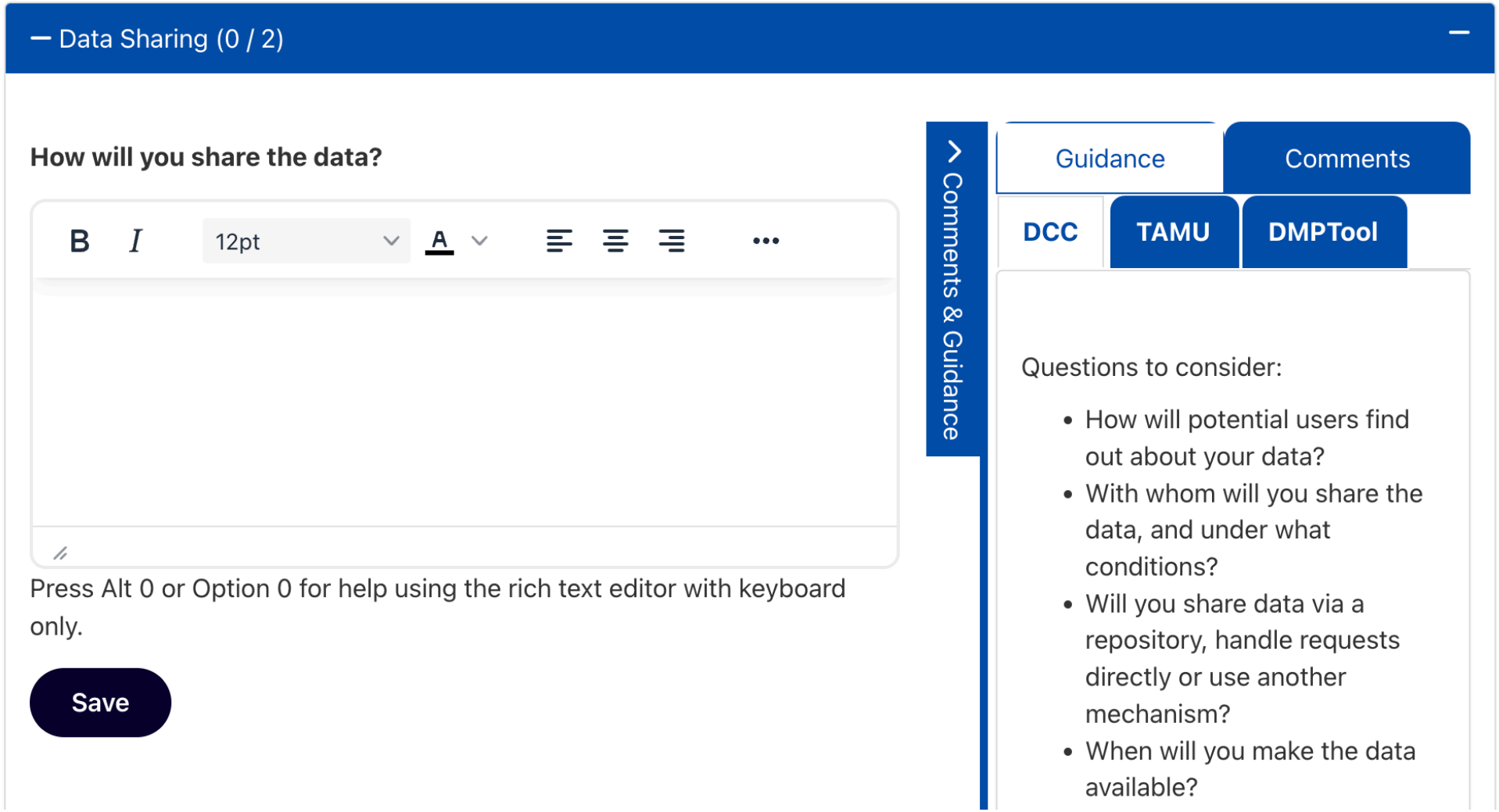
Collaborators will actually be invited to edit the plan.

You can have contributors who are collaborators, or not, and vice versa.

DMPTool
Build your Data Management Plan



Write Plan & Use Guidance

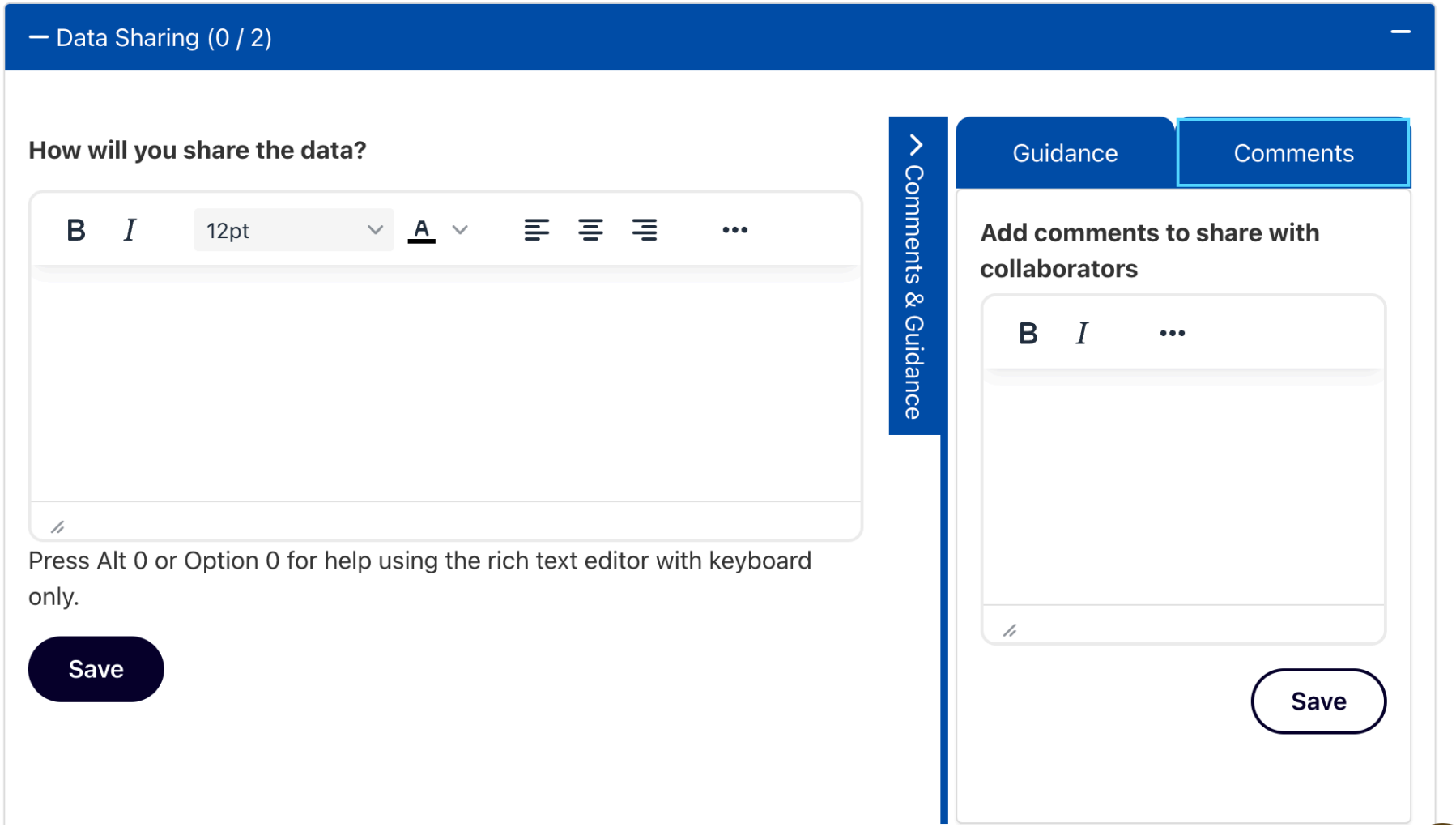


DMPTool
Build your Data Management Plan

LIBRARIES
TEXAS A&M UNIVERSITY

Access Comments

In the second tab beside Guidance, your collaborators can provide you comments about your DMP.

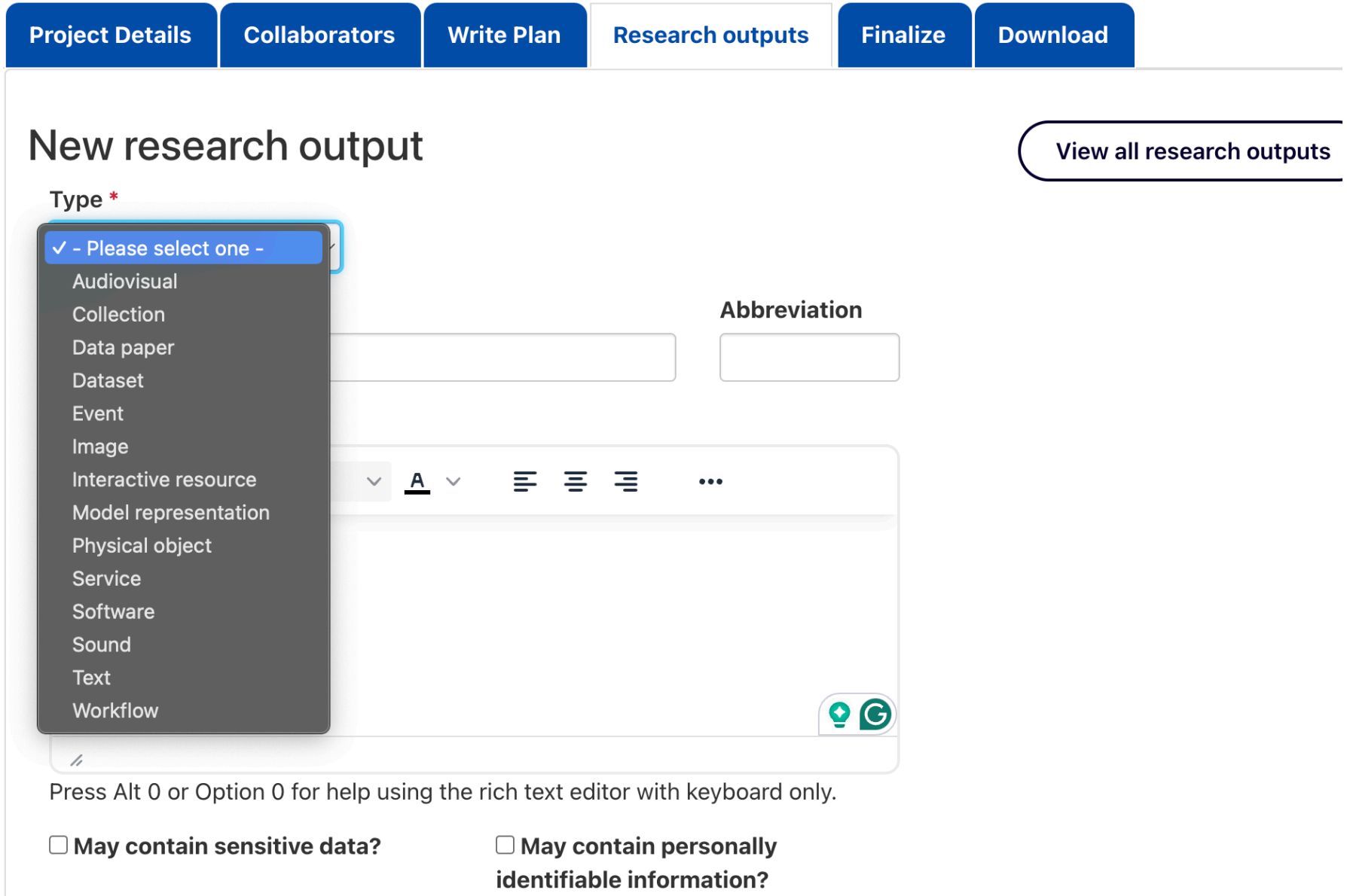


DMPTool
Build your Data Management Plan

LIBRARIES
TEXAS A&M UNIVERSITY

Research Outputs

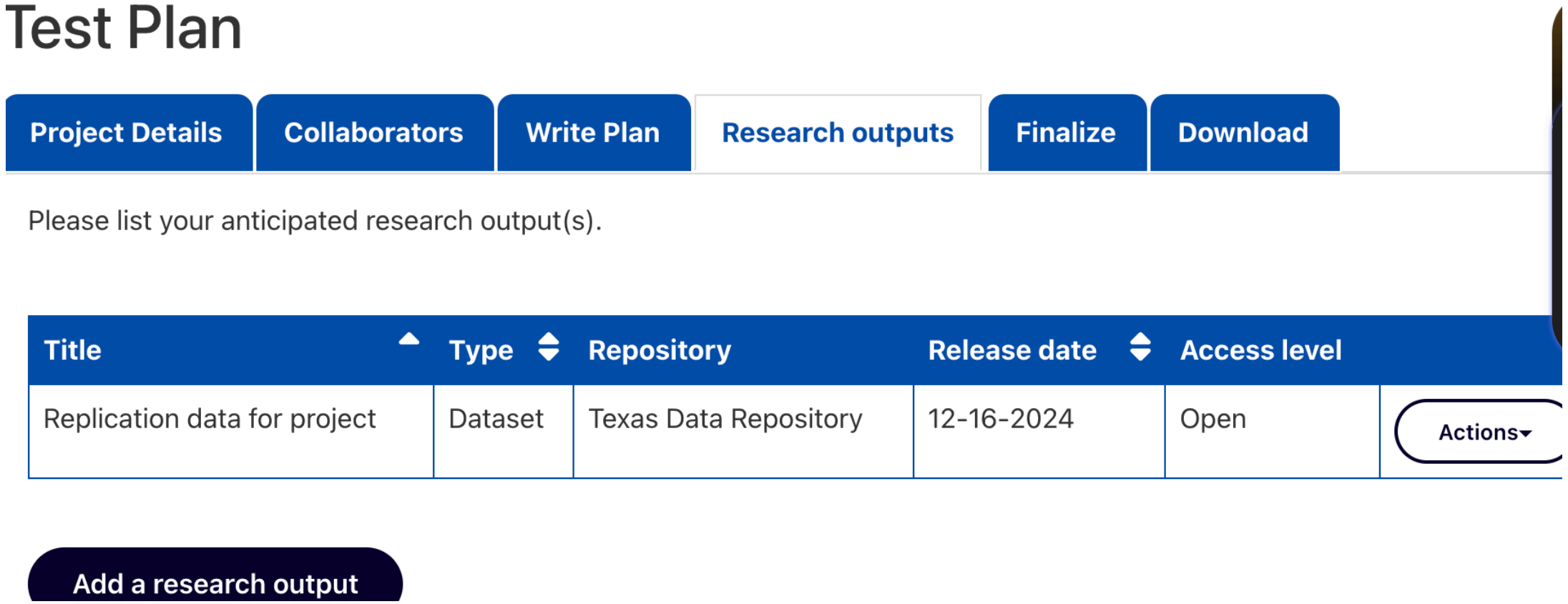
On this page, you will be able to list your anticipated research outputs in your DMP.



DMPTool
Build your Data Management Plan

LIBRARIES
TEXAS A&M UNIVERSITY

Research Outputs



Requesting Feedback

Request expert feedback

Click below to give data management staff at Texas A&M University , the Plan Owner's org, access to read and comment on your plan.

A data librarian from Texas A&M University will respond to your request within two business days. When requesting feedback, please p
Additional resources about using the DMPTool can be found in the "Create a Data Management Plan" guide. If you have questions pert
through the AskUs form or directly via email at ateauriri@tamu.edu.

You can continue to edit and download the plan in the interim.

Request feedback

Finalize: Publishing & Sharing Plans

Public plans:

- Visible to anyone on DMPTool.org
- Not recommended

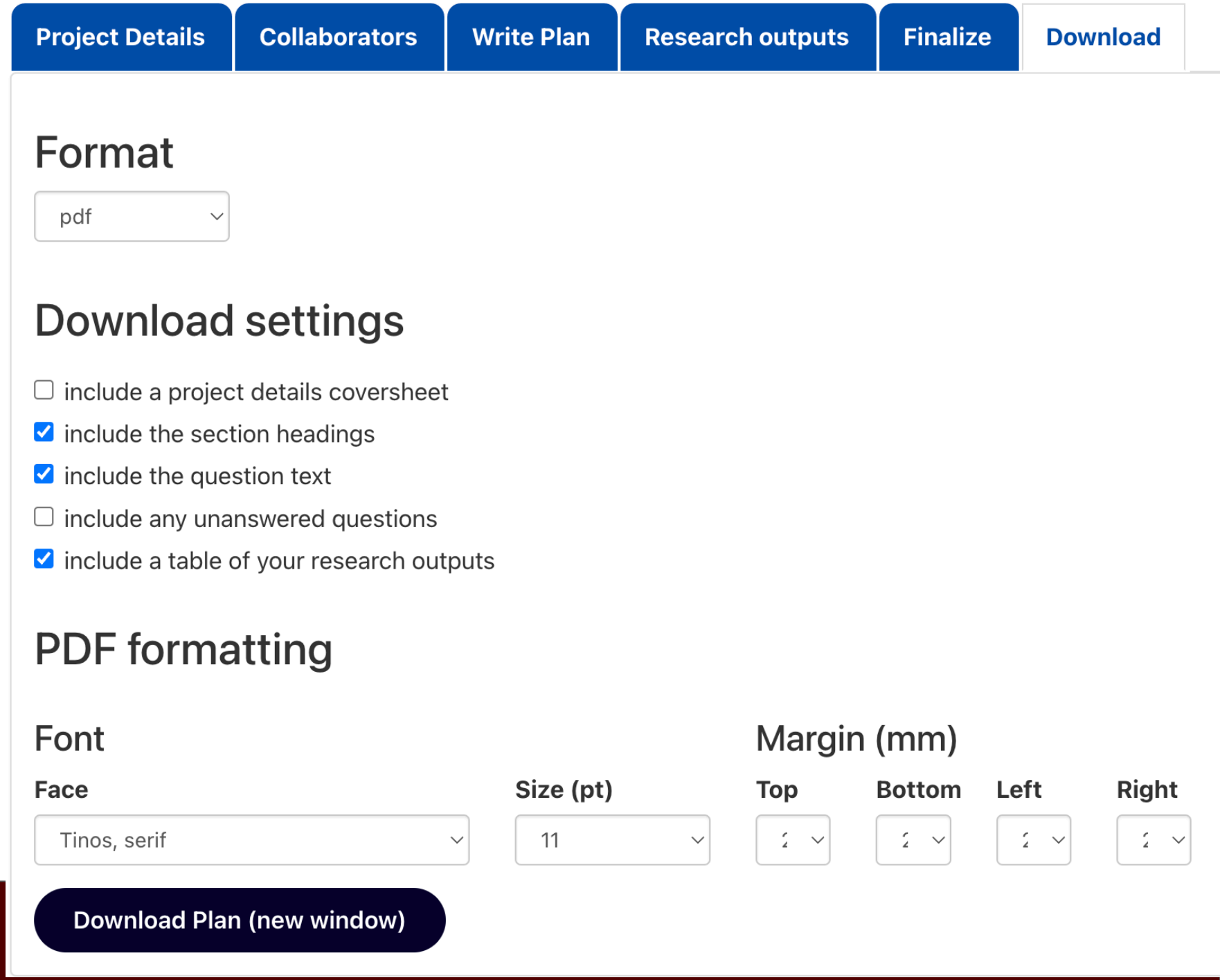
Registered plans:

- Receive a DMPID (may be required by grants),
- and assigned to your ORCID (which helps with machine actionability)

DMPTool
Build your Data Management Plan

Download

Download with all requirements!



DMPTool
Build your Data Management Plan

Getting Help

- Guidance section
- Feedback Button
- AskUs at the Library
- Funder requirements
- Public Plans (take with a grain of salt – you've seen how easy it is to publish a plan!)
  - DMPTool marks some as "Recommended" public plans!

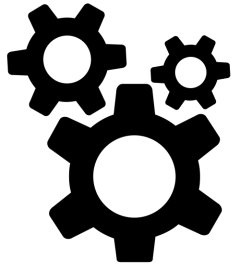
Data Repositories

Data should be Findable, Accessible, Interoperable, Reusable (FAIR)

Created by Diego Naive from Noun Project

To be **Findable**:

- Persistent ID
- Described with good metadata



Created by Daniel Shoreman from Noun Project

To be **Interoperable**:

- References to other metadata
- FAIR vocabularies
- Use discipline ontologies

Created by Gregor Cresnar from Noun Project

To be **Accessible**:

- Open standard protocol
- Easily Downloadable
- Open metadata readable by humans and machines

To be **Reusable**:

- Good provenance
- Clear data licenses

Finding a Data Repository

- Registry of Research Data Repositories: re3data.org
- The Directory of Open Access Repositories: http://opendoar.org
- FAIRSharing: https://fairsharing.org/databases/
- Nature *Scientific Data* recommended repositories: https://www.nature.com/sdata/policies/repositories

Data Repositories: Find out First

- Does the repository provide permanent identifiers or DOIs?
- Does the repository have self-deposit or mediated deposit?
- Does the repository have fees for storage/hosting?
- Does the repository allow deposit of sensitive data?
- Does the repository require review (or curation) of data before posting openly?
- How long are data held?
- File size limitations? File count limitations?
- Usage monitoring
- How is data backed up and archived?
- Access protocols?

What is Metadata

Metadata

What is it?

- Data about data or information about information
- Summary of the basic information about data
- Structured information

Why is it important?

- Helps others and your future self understand and use your data
- Contextualizes the data
- Allows for machine-readability
- Makes it easier to retrieve, use or manage an information resource

Metadata in daily life

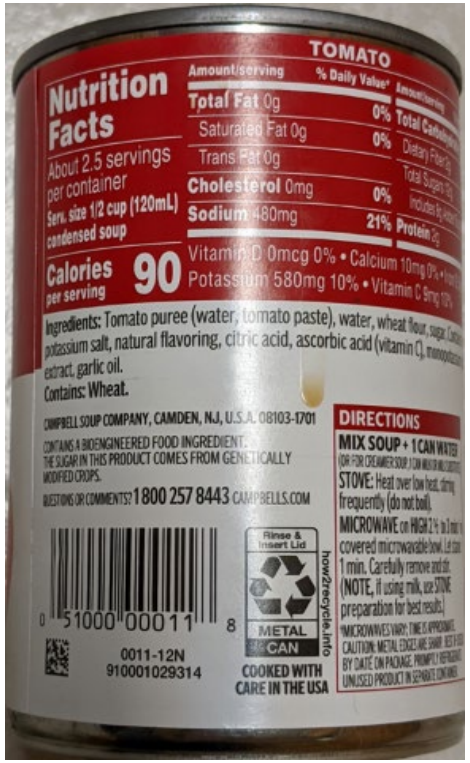
Metadata is automatically captured by phones and embedded in photos:

- Location
- Date
- Time
- File size
- File format
- Technical specifications about the camera

How Metadata helps others

POOR

BETTER



BEST

Keywords

Many words for the same thing

Homeless
Destitute
Unhoused
Street People

Coronavirus
COVID-19
SARS-CoV-2

Controlled Vocabulary

For Subject terms, or keywords, use your discipline's ontology - Controlled Vocabularies ensure interoperability of your data.

Ecological Metadata Language (EML) https://eml.ecoinformatics.org/

Data Documentation Initiative https://ddialliance.org/

Darwin Core Standard (biology) - https://github.com/tdwg/dwc

FAIR Sharing.org - registry of terminology artefacts https://fairsharing.org/search?fairsharingRegistry=Standard

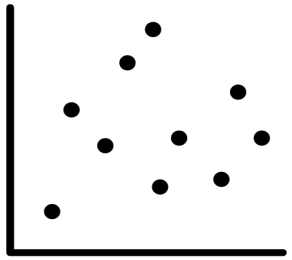
Documentation

Project-level documentation - or collection level documentation, is information about the whole research project including an abstract; keywords; methodology; the who, what, when, and where about the data collection, and how the data can be accessed.

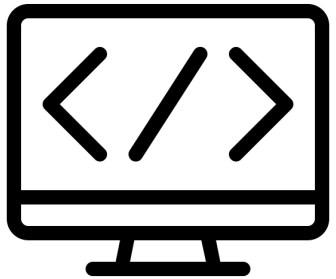
Data-level documentation - or object-level documentation, is information about the individual files and information within the files such as spreadsheets and transcripts, variables, acronyms, and labels.

What to include?

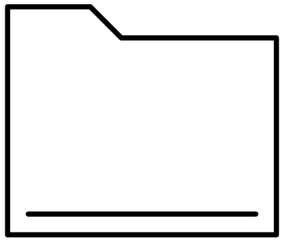
Data



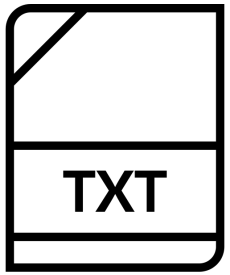
Code



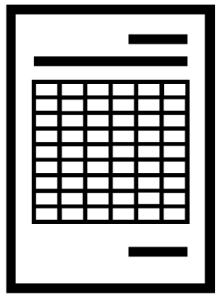
Data dictionary



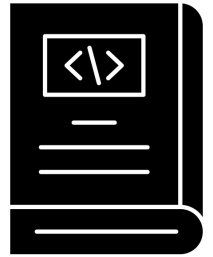
README text



Code books



Markdown



Documentation: README File

A standard document detailing information about other documents:

- Title of dataset
- Name/institution/contact information of person responsible for 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 (major and minor revisions explained)
- How data was collected, methodology, populations

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

Texas Data Repository - https://data.tdl.org/

Purpose: to provide a platform for archiving and publishing the data developed or used in support of research at Texas A&M University, to promote the reproducibility of that research, and to facilitate sharing and collaboration

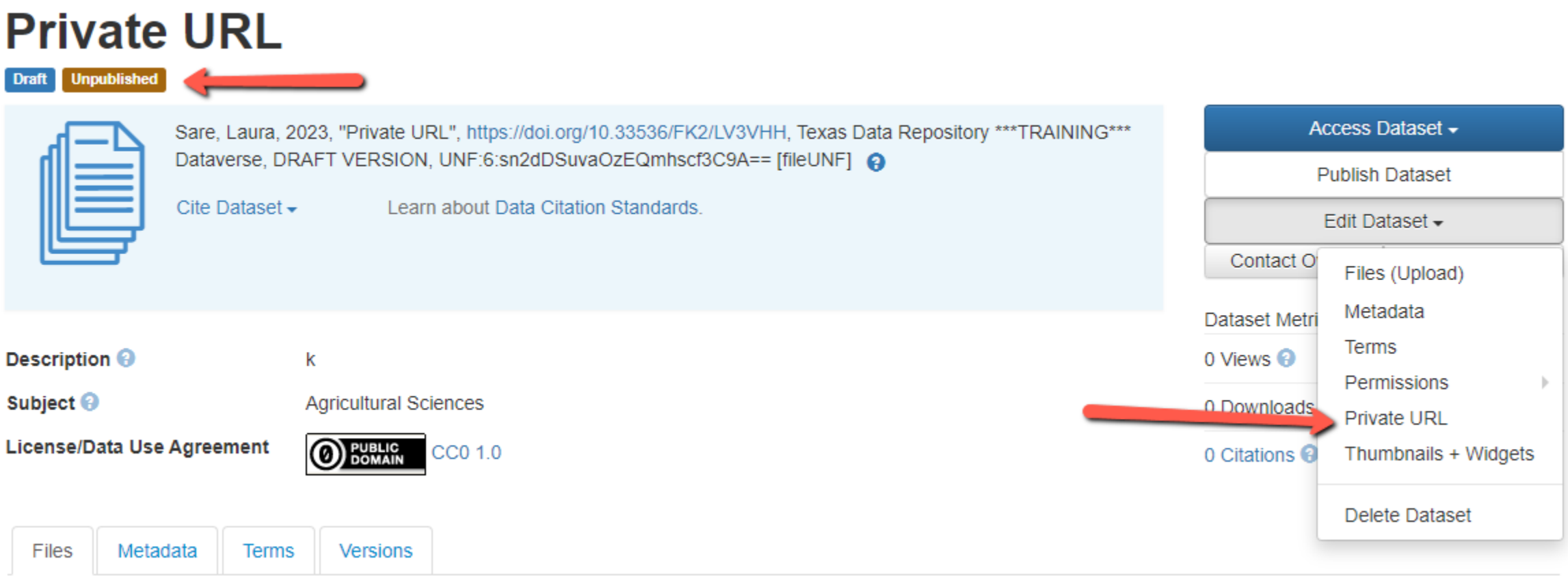
(works best in Chrome browser)

Texas Data Repository

Small Data Projects - Datasets **under 50GB**

- Data in any file type
- Data from any field of research
- Static or evolving datasets
- Individual files up to **4GB** (up to 2GB is best)
- Data **without** confidential, proprietary, or sensitive information
- Restrict Files - define terms on how files can be used
- Embargo date - blocks access to files for limited time period
- Private URL
- Upload folders of files
- Customizable metadata tags and fields

Private URL


Private URL
Draft
Unpublished
Sare, Laura, 2023, "Private URL", https://doi.org/10.33536/FK2/LV3VHH, Texas Data Repository ***TRAINING*** Dataverse, DRAFT VERSION, UNF:6:sn2dDSuvaOzEQmhscf3C9A== [fileUNF]
Cite Dataset
Learn about Data Citation Standards.
Access Dataset
Publish Dataset
Edit Dataset
Contact O
Files (Upload)
Metadata
Terms
Permissions
Private URL
Thumbnails + Widgets
Delete Dataset
Dataset Metri
0 Views
0 Downloads
0 Citations
Description
k
Subject
Agricultural Sciences
License/Data Use Agreement
PUBLIC DOMAIN
CC0 1.0
Files
Metadata
Terms
Versions

Texas Data Repository

Texas A&M Libraries:

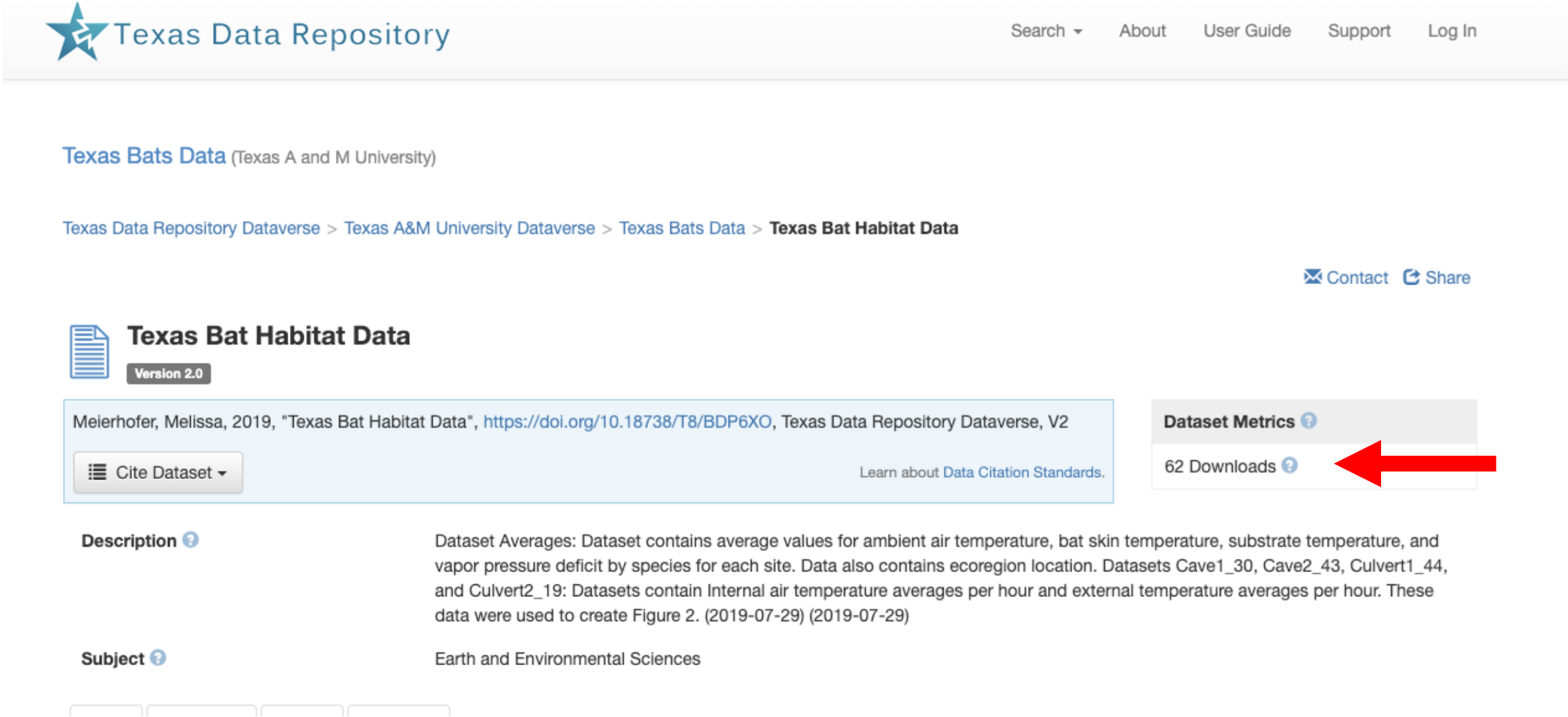
- Help with data management planning and with preparing data for deposit
- Offer advice on appropriate file formats, metadata, and licensing options
- Provide training materials or workshops for users to upload and manage their own data collections

Texas Data Repository

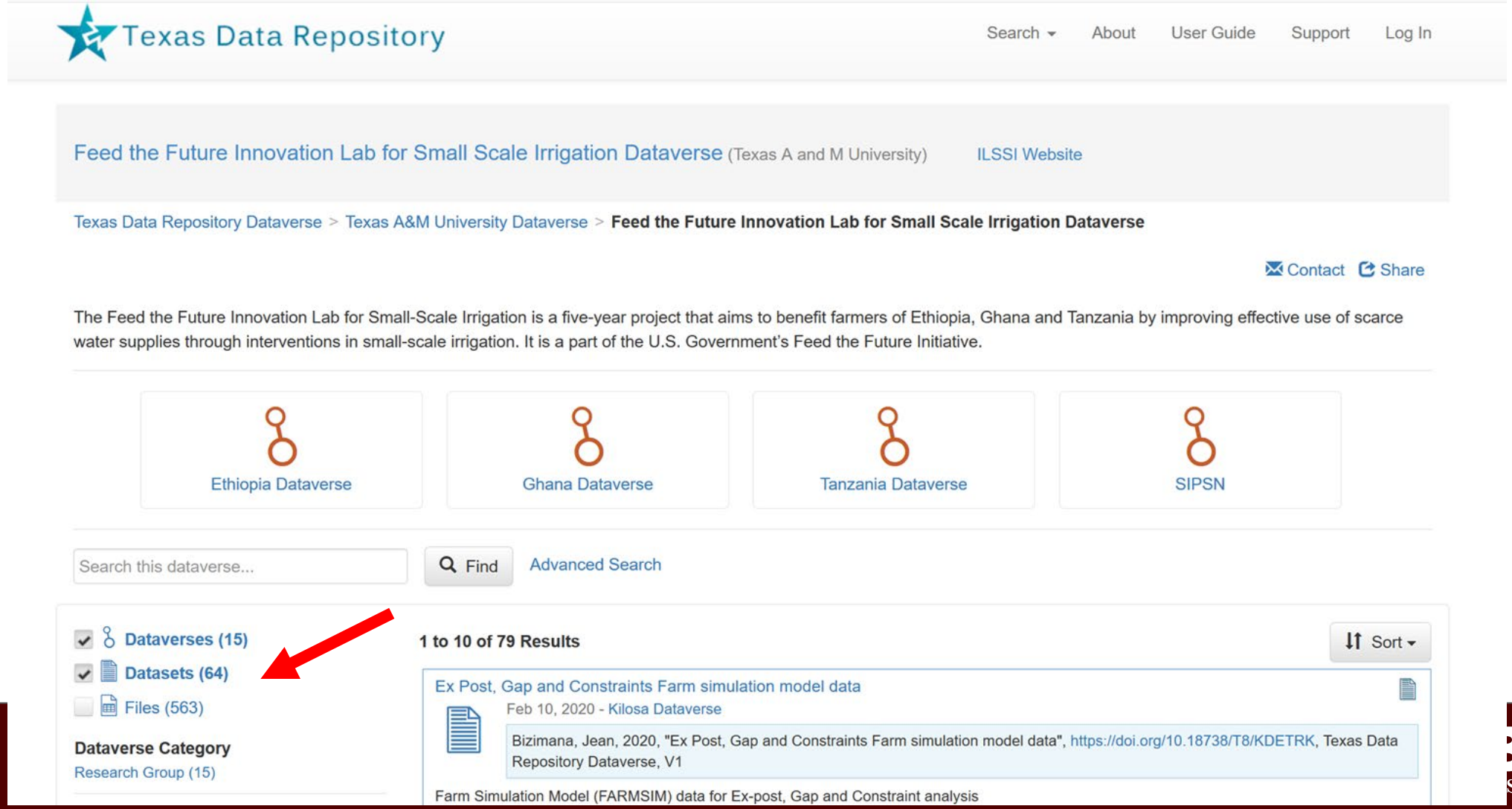
Benefits:

- Version control helps track progress and keep collaborators up-to-date
- Published datasets are assigned a DOI (persistent identifier) to allow citation
- Flexible access controls let users decide when, with whom, and how much data to share
- Customizable metadata fields, permissions
- Long term preservation of uploaded data

Texas Data Repository Example 1

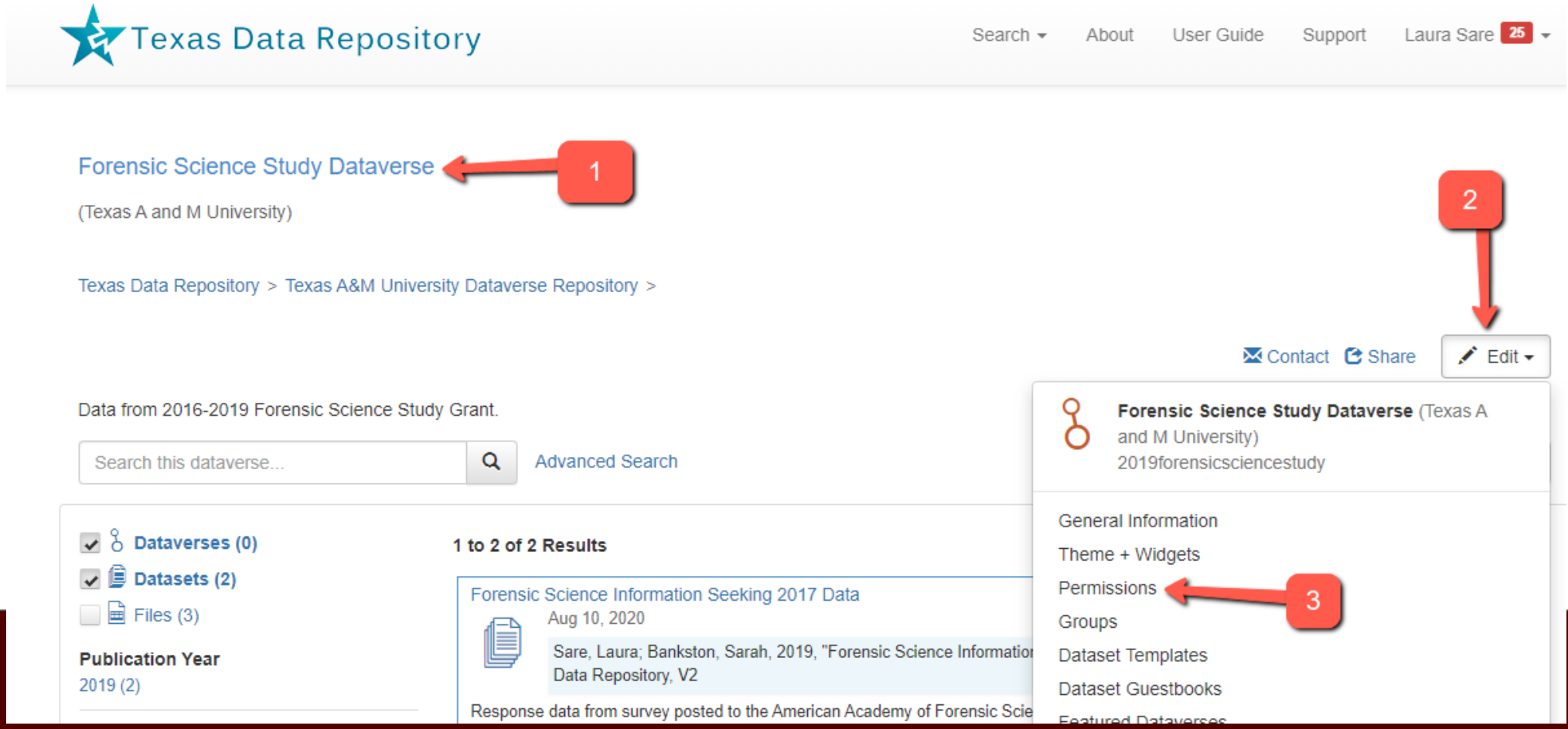


Texas Data Repository Example 2



Note to Graduate Students!

BEFORE you leave Texas A&M University, to have access to your data at a later time, you will need to create a new user for your dataverse and use a gmail email address.



Final Tips and Reminders

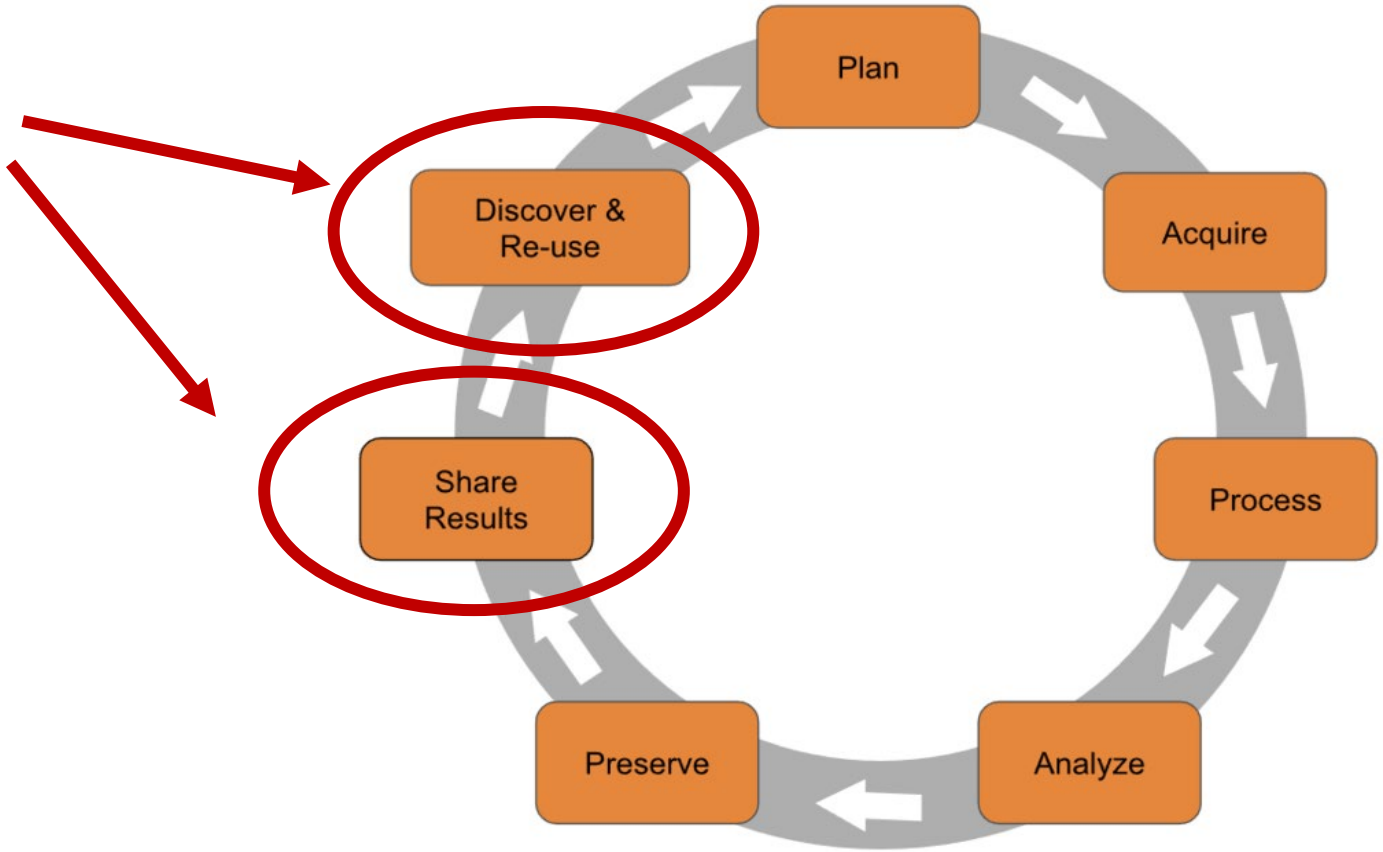
- Decide which data you want to share
- Choose file formats that last
- Remember the documentation
- Consider ownership and privacy
- Follow metadata standards, look at repository metadata fields to use during documentation

http://tamu.libguides.com/research-data-management/repositories

Data Publishing & Impact

Jayden Matus, Research & Scholarly Initiatives Librarian
Email: jreider@tamu.edu

YOU ARE HERE



Research Data Lifecycle by
Princeton Research Data Service

What is Data Impact?

The influence and value which your published data holds and affects the research ecosystem at large.

Ways to measure Data Impact:

- Data views
- Data downloads
- Data citations
- Author citations
- Institution recognition

Publish It!

Why Publish your Data?

- Increase your data's discoverability
- Show academic transparency
- Preserve your data/establish a home for it
- Your funder probably requires you to do so
- Increase your impact!

Impact from Publishing Data

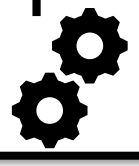
- Contribute to your field’s advancements, research, and discussions
- Can shape a country's policy and laws
- Improving medical and clinical practices
- Increased citations and potential collaborations with others in your field
- Encourages open science initiatives and importance of discoverability
- Demonstrates your commitment as a researcher to transparent data sharing practices

What is FAIR Data?

Findable

Accessible

Interoperable



Reusable

FAIR Data Practices and Publishing

- Unique Identifier (UID) for your dataset (usually generated by the repository/journal)
- Data is described with rich and precise metadata
- Publishing in a trusted repository
- Metadata should be accessible even when the original data is not
- Data should meet the domain-relevant standards for publishing

Sourced from GOFAIR FAIR Principles:
FAIR Principles - GO FAIR

Open Access Publishing

Inspired by Iowa State University: For Fun: Open Access Memes · OpenISU

Why Publish Open Access?

- Data is more accessible
- Data reaches a wider audience/more exposure (and more impact!)
- Higher chance of citation and reference in similar research
- Supports fair play in the access to knowledge
- Your funder may require you to publish in OA locations
- Discounted/waived Article Processing Charge programs

Publishing Options

- Data repositories
  - Institutional Repositories
  - Specialist repositories
  - Generalist repositories
- Data journals (for data articles)
  - TAMU Libraries OA publishing options: https://library.tamu.edu/services/scholarly_communication/Open_Access/opportunities.php

Establish your Scholarly Profiles

- Open Researcher and Contributor ID (ORCID) - ORCID
- Scholars@TAMU - Scholars@TAMU (for *TAMU Faculty*)
- Open Science Framework (OSF) - OSF | Home
- Research Gate - Login to ResearchGate

Help other researchers by providing contact information and increase your reach/collaboration!

Further Resources

TAMU Data Management

TAMU Library Event Calendar – Research Data Management

TAMU Libraries Research Data Management LibGuide

TAMU The Responsible Stewardship of Research Data Standard Administrative Procedure – 15.99.03.M1.03

TAMU Technology Services Data Classification Guide

TAMU Technology Services Data Tools

TAMU Division of Research

Research Data Management Services

Further Resources

Texas Data Repository

TDR User Guide - https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/287965260/User+Guide

TDR Video series -

https://www.youtube.com/playlist?list=PLHTJEapaCC32q5JxYH980SZwz1hMiGBKK

TDR Metadata dictionary - https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/723550229/Texas+Data+Repository+Educational+Resources?preview=/216006691/289177628/TDR-Metadata-Dictionary.pdf

Further Resources

TAMU Publishing and Impact

Scholarly Identity, Impact, and Reputation

Open Access Publishing Opportunities

ORCID Guide - https://tamu.libguides.com/orcid

Further Publishing Resources

Data - Publishing strategy guide - Library guides at UNSW Library

Galter Health Sciences Library & Learning Center | News Understanding the Impact of Data

Why publish and preserve data? - Research - University of St Andrews

Thank you!

For further assistance, please contact us!

Ateanna Uriri – ateauriri@tamu.edu
Dianna Morganti – diannamorganti@tamu.edu
Laura Sare – lsare@tamu.edu
Jayden Matus – jreider@tamu.edu

University Libraries
https://library.tamu.edu/

University Libraries – Ask Us
https://askus.library.tamu.edu/