Publishing Data in the Texas Data Repository

What to expect today

- Learn about FAIR Data
- Learn how to find data repositories
- Learn how to create a dataverse collection in the Texas Data Repository (TDR) test site
- Learn how to add a dataset in the TDR
- Assign metadata to your dataset

How to Communicate with each other

Ask us questions!

You can unmute your microphone or put a question in chat

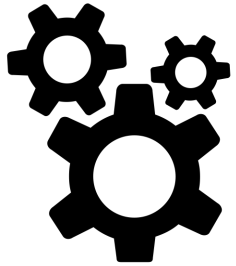
If we ask a question - please type a response in the chat, even if it is to say “no”, you don’t have any questions or “yes” I’m all set.

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

Data Repositories should maximize discovery and access.

- Does the repository provide DOIs?
- Does the repository provide help for metadata creation?
- Does the repository have fees for storage/hosting?
- Does the repository allow deposit of sensitive data?
- Does the repository require review of data before posting openly?

Data Repository Considerations

- Access
- Permanent Identifiers - a DOI or permanent URL
- Embargoes
- Licensing: Open Data Commons, Creative Commons
- Depositing Data: self-deposit and mediated deposit
- Curation
- Data disposal
- Accreditation or Certification (https://www.coretrustseal.org/)

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?

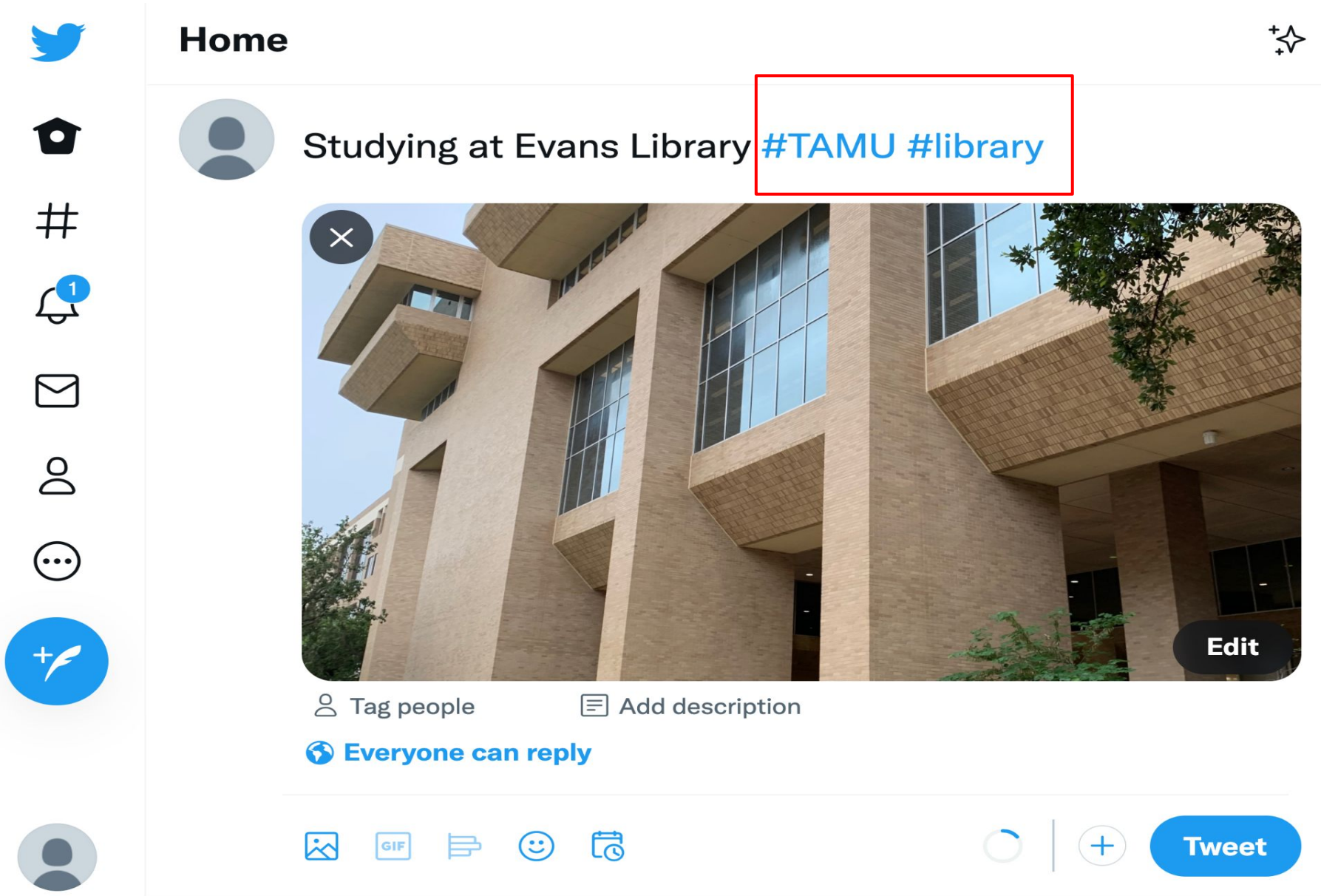
- Helps others and your future self understand and use your data
- Contextualizes the data
- Allows for machine-readability
- Makes it easier to retrieve, use or manage an information resource

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

We enter metadata manually whenever we describe or tag photos on social media.



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
Delta

Controlled Vocabulary

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

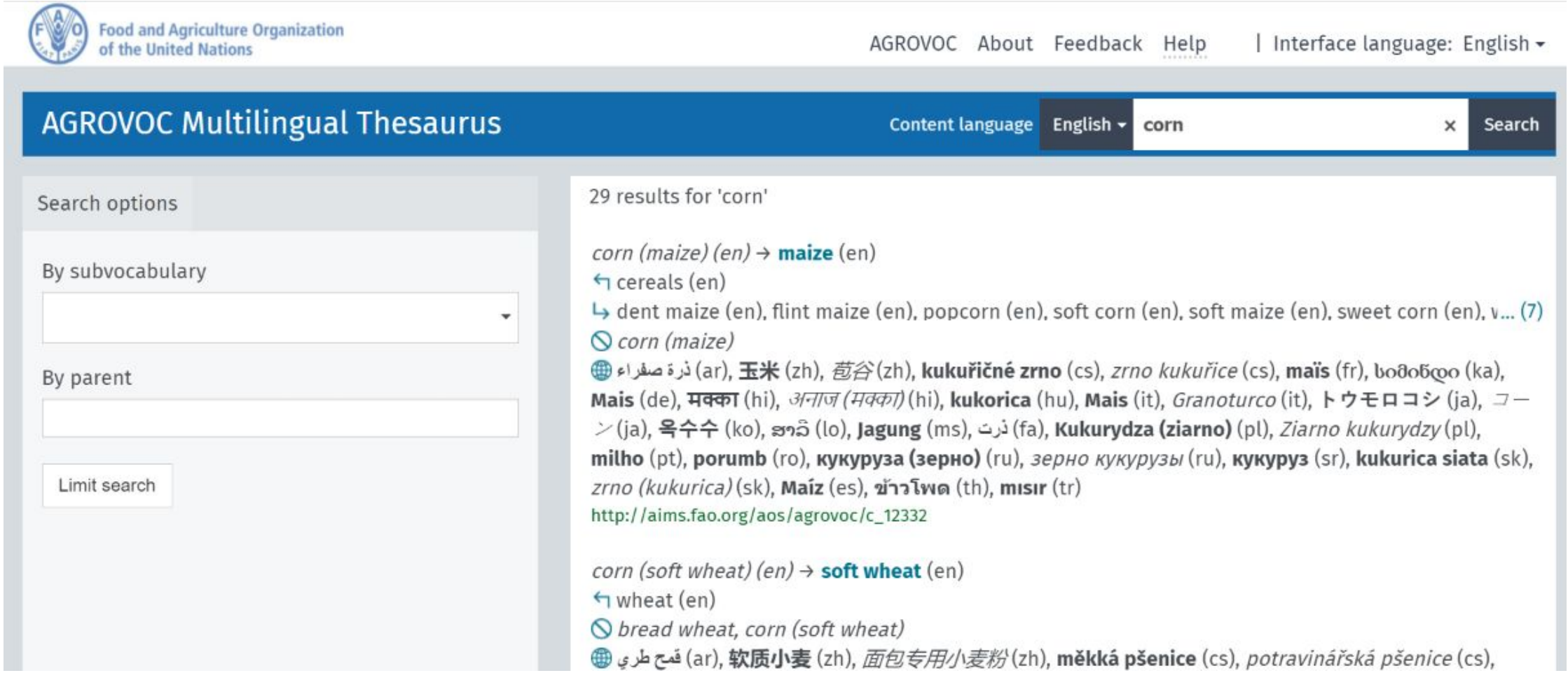
Ecological Metadata Language (EML) https://eml.ecoinformatics.org/

Data Documentation Initiative https://ddialliance.org/

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

AGROVOC Linked Open Data about agriculture http://www.fao.org/agrovoc/

Where to find Metadata



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

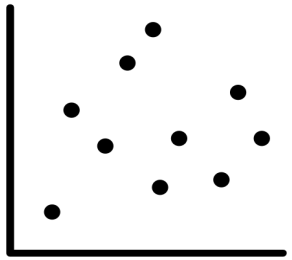
Appropriate for:

Small Data Projects - Datasets must be **under 10GB**

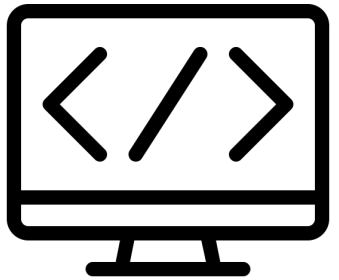
- Data in any file type
- Data from any field of research
- Static or evolving datasets
- Individual files up to 4GB (up to 2GB is best)
- Data **without** confidential or sensitive information

What to include?

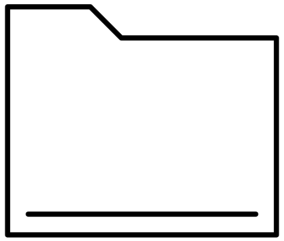
Data



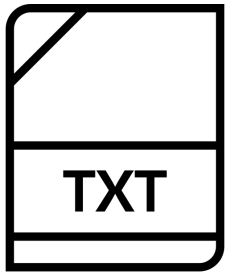
Code



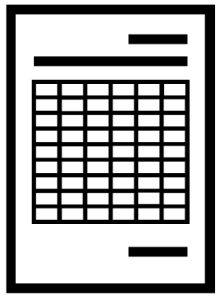
Data dictionary



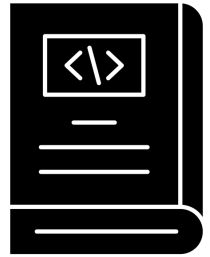
README text



Code books



Markdown



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

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

Texas Data Repository

Texas A&M Libraries:

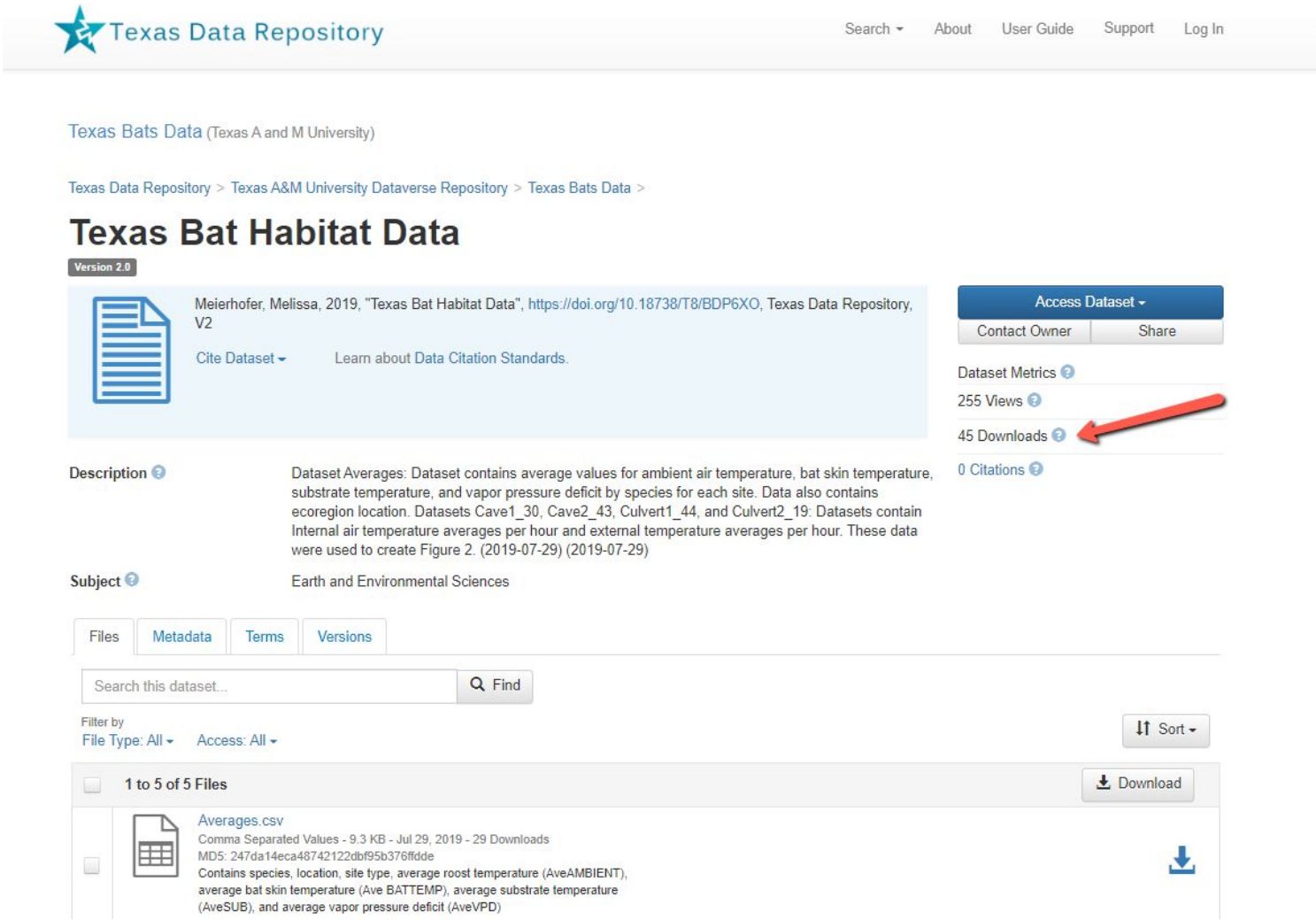
- Help with data management planning and with preparing data for deposit
- Offer advice on appropriate file formats, metadata, and licensing options
- Provide training materials or workshops for users to upload and manage their own data collections

Texas Data Repository

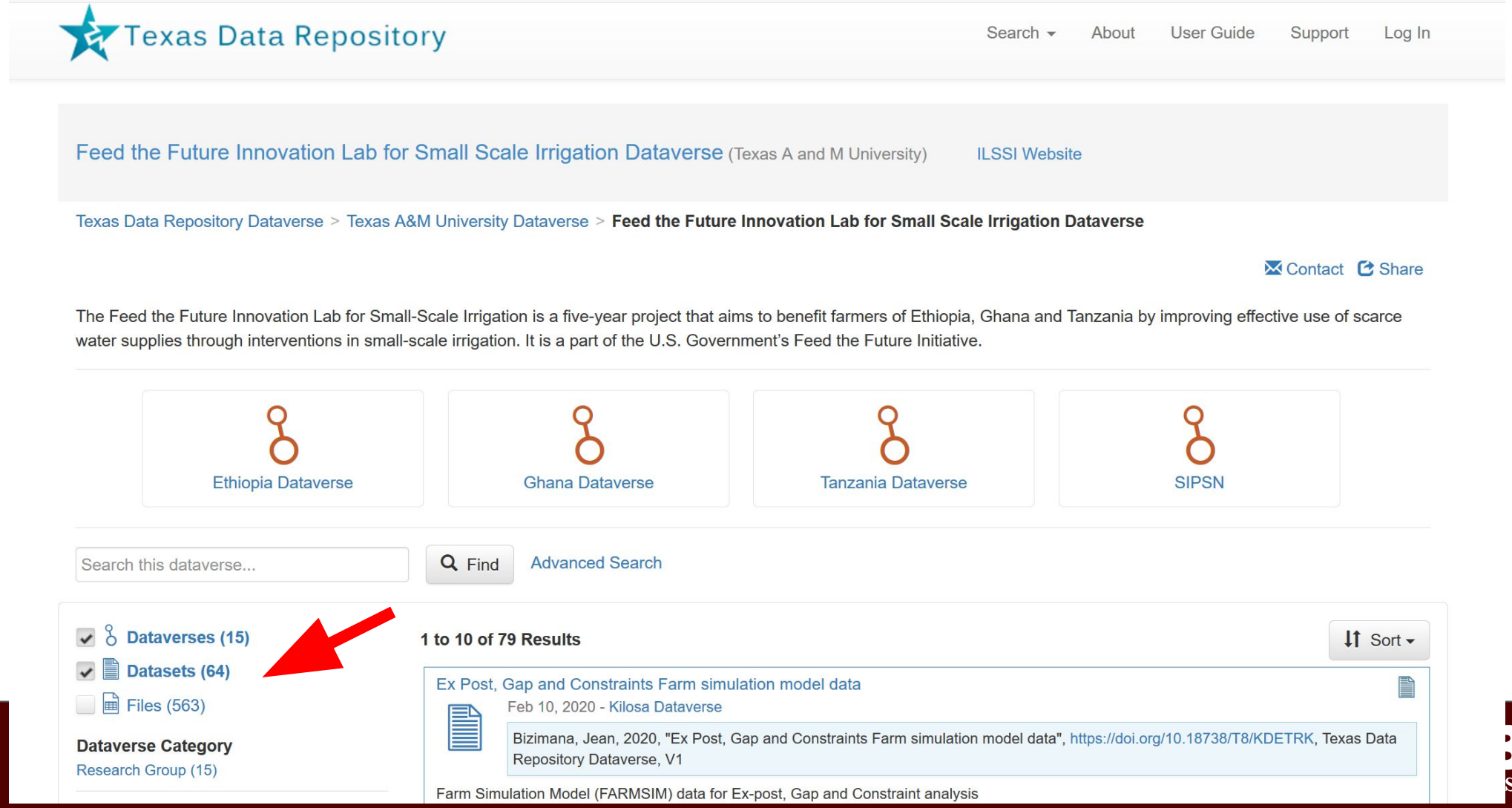
Benefits:

- 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



Texas Data Repository Demo

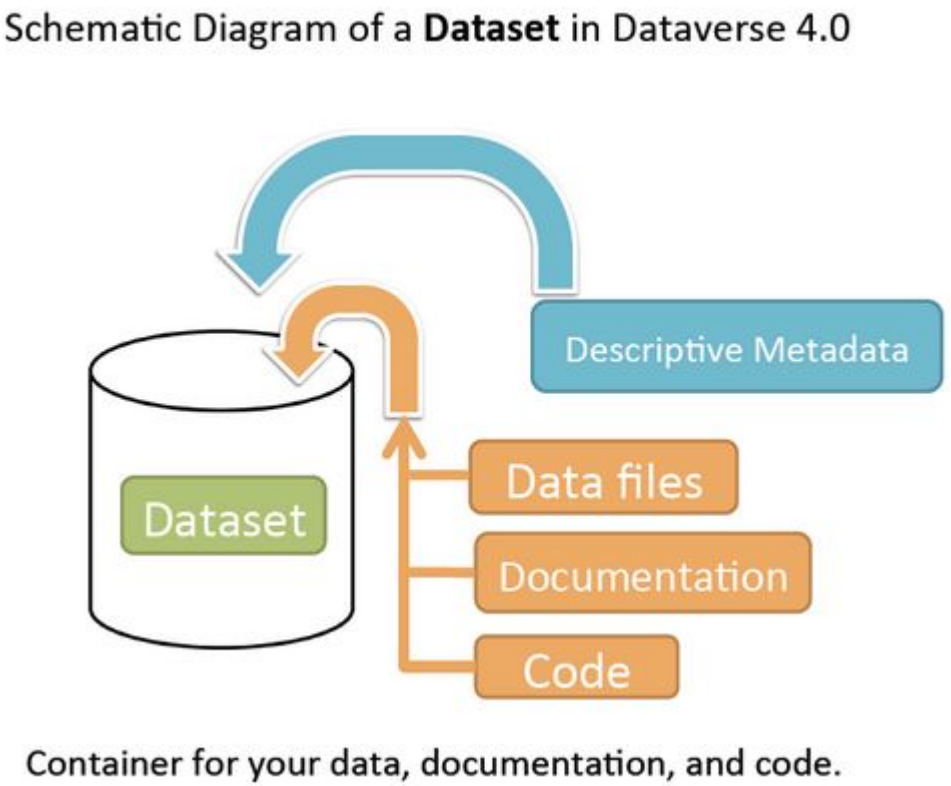


Image from: Dataverse http://guides.dataverse.org/en/latest/user/dataset-management.html

Texas Data Repository Resources

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=/216006691/2891177628/TDR-Metadata-Dictionary.pdf

TAMU TDR Cheatsheets and copy of this presentation - https://osf.io/t6a9e/

FAIR Resources

Force 11 FAIR Data Principles

https://www.force11.org/group/fairgroup/fairprinciples

The FAIR Guiding Principles for scientific data management and stewardship by Mark D. Wilkinson et al.#

https://www.nature.com/articles/sdata201618

Go FAIR

https://www.go-fair.org/fair-principles/

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