

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.



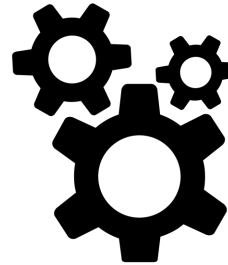
Data should be Findable, Accessible, Interoperable, Reusable (FAIR)



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



To be **Accessible**:

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



To be **Reusable**:

- Good provenance
- Clear data licenses

Created by Gregor Cresnar
from Noun Project



LIBRARIES
TEXAS A&M UNIVERSITY

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.



Home + ✨


 Studying at Evans Library #TAMU #library



Edit

 Tag people  Add description

 **Everyone can reply**

      |  **Tweet**

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

AGROVOC Multilingual Thesaurus

Content language

English ▾

corn

x

Search

Search options

By subvocabulary

By parent

Limit search

29 results for 'corn'

corn (maize) (en) → **maize (en)**

↶ cereals (en)

↷ dent maize (en), flint maize (en), popcorn (en), soft corn (en), soft maize (en), sweet corn (en), v... (7)

🗑 *corn (maize)*

🌐 ذرة صفراء (ar), 玉米 (zh), 苞谷 (zh), **kukuřičné zrno (cs)**, *zrno kukuřice (cs)*, **maïs (fr)**, *ბობოხე (ka)*, **Mais (de)**, **मक्का (hi)**, *अनाज (मक्का) (hi)*, **kukorica (hu)**, **Mais (it)**, *Granoturco (it)*, *トウモロコシ (ja)*, *コーン (ja)*, **옥수수 (ko)**, **ᩈᩬ᩵ (lo)**, **Jagung (ms)**, **ثرت (fa)**, **Kukurydza (ziarno) (pl)**, *Ziarno kukurydzy (pl)*, **milho (pt)**, **porumb (ro)**, **кукуруза (зерно) (ru)**, *зерно кукурузы (ru)*, **кукуруз (sr)**, **kukurica siata (sk)**, *zrno (kukurica) (sk)*, **Maíz (es)**, **ข้าวโพด (th)**, **misir (tr)**

http://aims.fao.org/aos/agrovoc/c_12332

corn (soft wheat) (en) → **soft wheat (en)**

↶ wheat (en)

🗑 *bread wheat, corn (soft wheat)*

🌐 قمح طري (ar), 软质小麦 (zh), 面包专用小麦粉 (zh), **měkká pšenice (cs)**, *potravinářská pšenice (cs)*,

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:

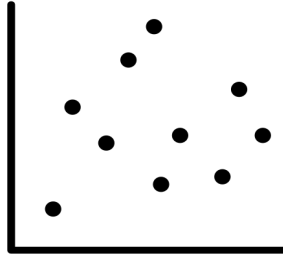
Small Data Projects - Datasets must be **under 10GB**

- 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

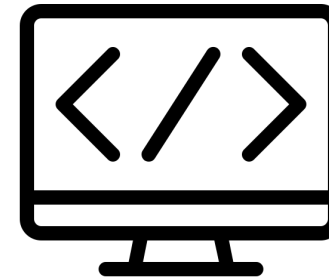


What to include?

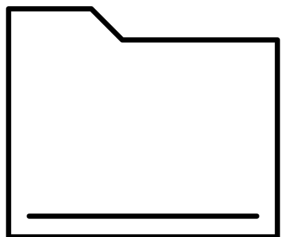
Data



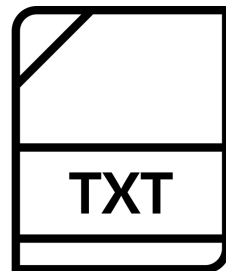
Code



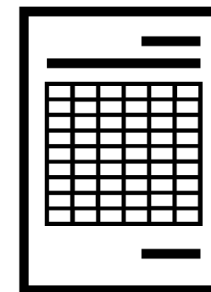
Data dictionary



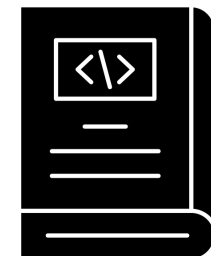
README text



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 Bats Data (Texas A and M University)

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

Texas Bat Habitat Data

Version 2.0



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

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

Access Dataset ▾
Contact Owner Share

Dataset Metrics ⓘ

255 Views ⓘ

45 Downloads ⓘ 

0 Citations ⓘ

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

Files Metadata Terms Versions

Search this dataset...

Filter by

File Type: All ▾ Access: All ▾

Sort ▾

1 to 5 of 5 Files



[Averages.csv](#)

Comma Separated Values - 9.3 KB - Jul 29, 2019 - 29 Downloads
MDS: 247da14eca48742122dbf95b376ffdde

Contains species, location, site type, average roost temperature (AveAMBIENT), average bat skin temperature (Ave BATTEMP), average substrate temperature (AveSUB), and average vapor pressure deficit (AveVPD)



Texas Data Repository Example 2

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

Texas Data Repository Dataverse > Texas A&M University Dataverse > **Feed the Future Innovation Lab for Small Scale Irrigation Dataverse**

[✉ 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.



Search this dataverse...

 Find

[Advanced Search](#)

 **Dataverses (15)**


 **Datasets (64)**

 **Files (563)**

Dataverse Category


Research Group (15)

1 to 10 of 79 Results

 Sort ▾

 [Ex Post, Gap and Constraints Farm simulation model data](#)

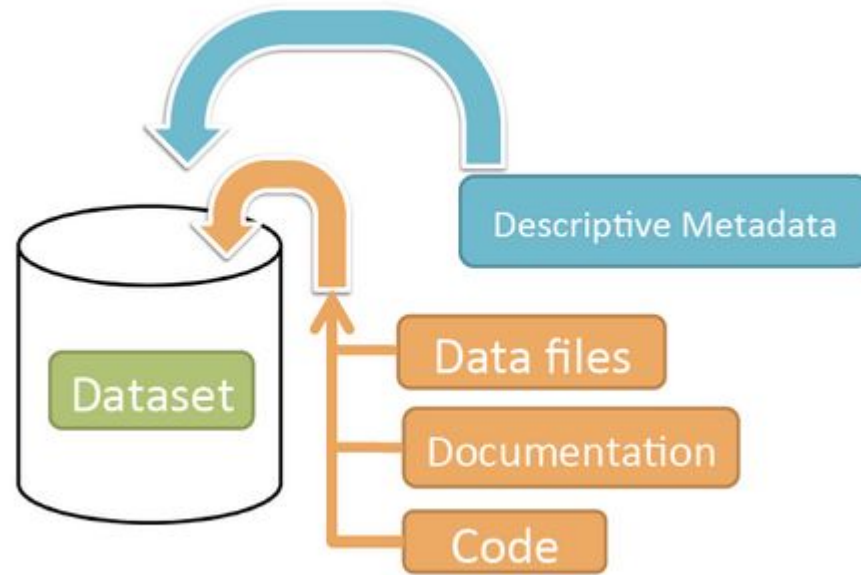
Feb 10, 2020 - Kilosa Dataverse

 Bizimana, Jean, 2020, "Ex Post, Gap and Constraints Farm simulation model data", <https://doi.org/10.18738/T8/KDETRK>, Texas Data Repository Dataverse, V1

Farm Simulation Model (FARMSIM) data for Ex-post, Gap and Constraint analysis

Texas Data Repository Demo

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

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=/216006691/289177628/TDR-Metadate-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>

