Texas Data Repository Tutorial

ILSSI Dataset Submission

Last revised: 2017-08-02
Revised by: Anna Dabrowski

About the Texas Data Repository

The Texas Data Repository (TDR) is an online platform for Texas A&M University researchers to publish and archive datasets and data products. It is accessed as a website.

The TDR is built on Dataverse, an open-source software developed by Harvard University. It is hosted by the Texas Digital Library, a consortium of Texas academic libraries focused on long-term access and preservation of digital content. The TDR is committed to preserving and providing ongoing access to research data for at least 10 years after submission.

More information about the TDR is available here: http://data.tdl.org/about/

Datasets

In the Texas Data Repository, data files are uploaded to "datasets." Each dataset receives it's own webpage and a Digital Object Identifier (DOI).

A dataset may hold multiple files, these files can contain data or documentation about the data. Each file within a dataset may be **2GB** or less and should be saved in a **non-proprietary** file format, for example a comma-separated value file (CSV) instead of a Microsoft Excel Spreadsheet file (XLSX). The files uploaded to a dataset can also be ZIP or TAR files that contain multiple files within them.

Data file(s) uploaded to a dataset are stored along with additional information about the data, such as the title, creator, and date created. This "metadata" is structured, and describes the dataset to users and the Texas Data Repository system in a consistent way. It is visible on the dataset webpage and used by the Texas Data Repository software for search and filter features.

Dataverses

"Dataverse" is both the name of the software running the Texas Data Repository and the name of a container that holds datasets. The dataverse container is similar to a directory or folder in a computer file system.

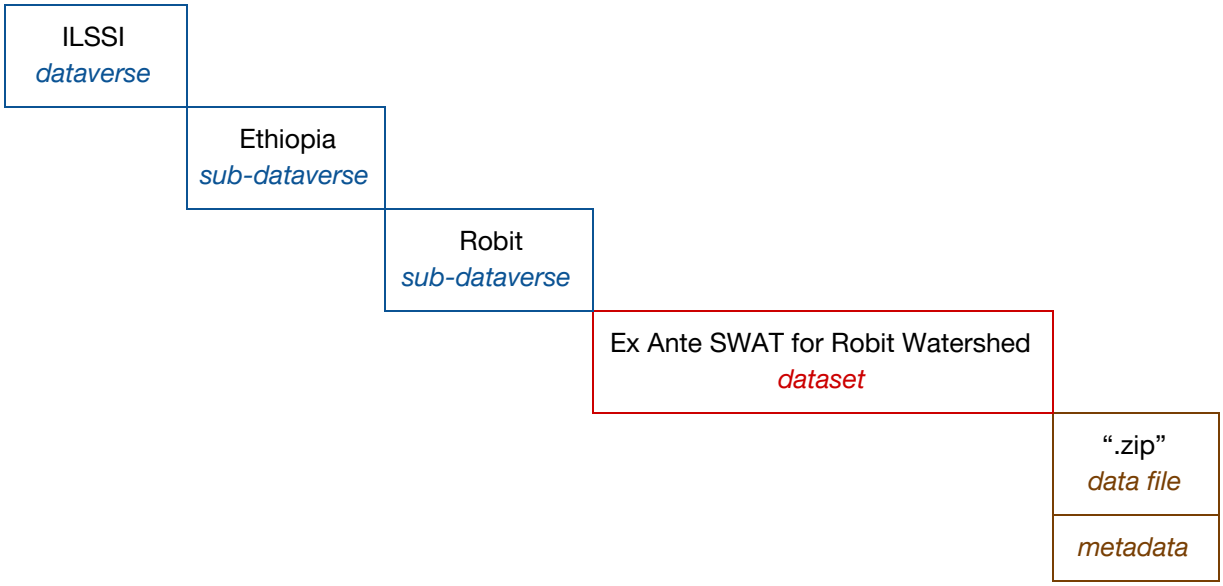
A dataverse may hold one or more datasets, it can be used to organize datasets into collections based on a project or other criteria. The dataverse container can also hold sub-dataverses, that is, sub-collections of datasets.

ILSSI Dataverse

The ILSSI dataverse is organized by country and region. It contains sub-dataverses named after countries, these country dataverses contain sub-dataverses named after regions, they may also hold datasets that apply to the country level.

Each regional sub-dataverse includes datasets relevant to that region. Each dataset may be based on a different model and analysis conducted in a particular region.

Example of Structure



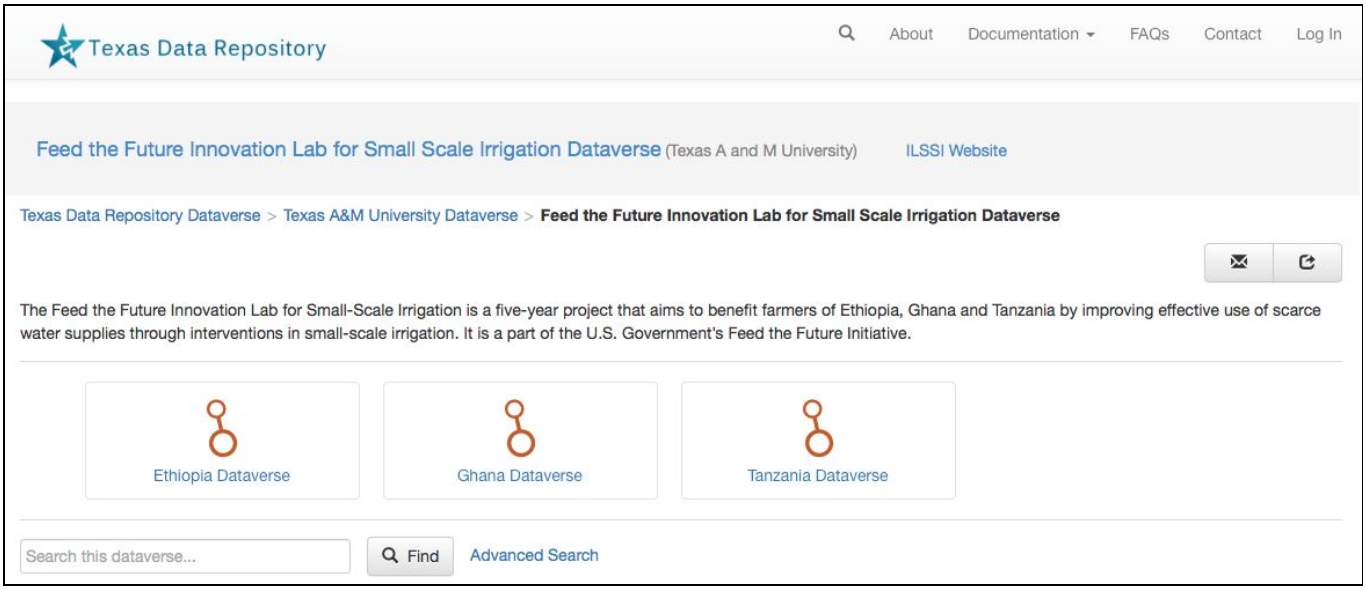
Permissions

Researchers are given permissions to create datasets within a particular dataverse by the ILSSI Program Coordinator, or alternatively, the Data Management Librarian.

Adding Datasets

Access the ILSSI Dataverse

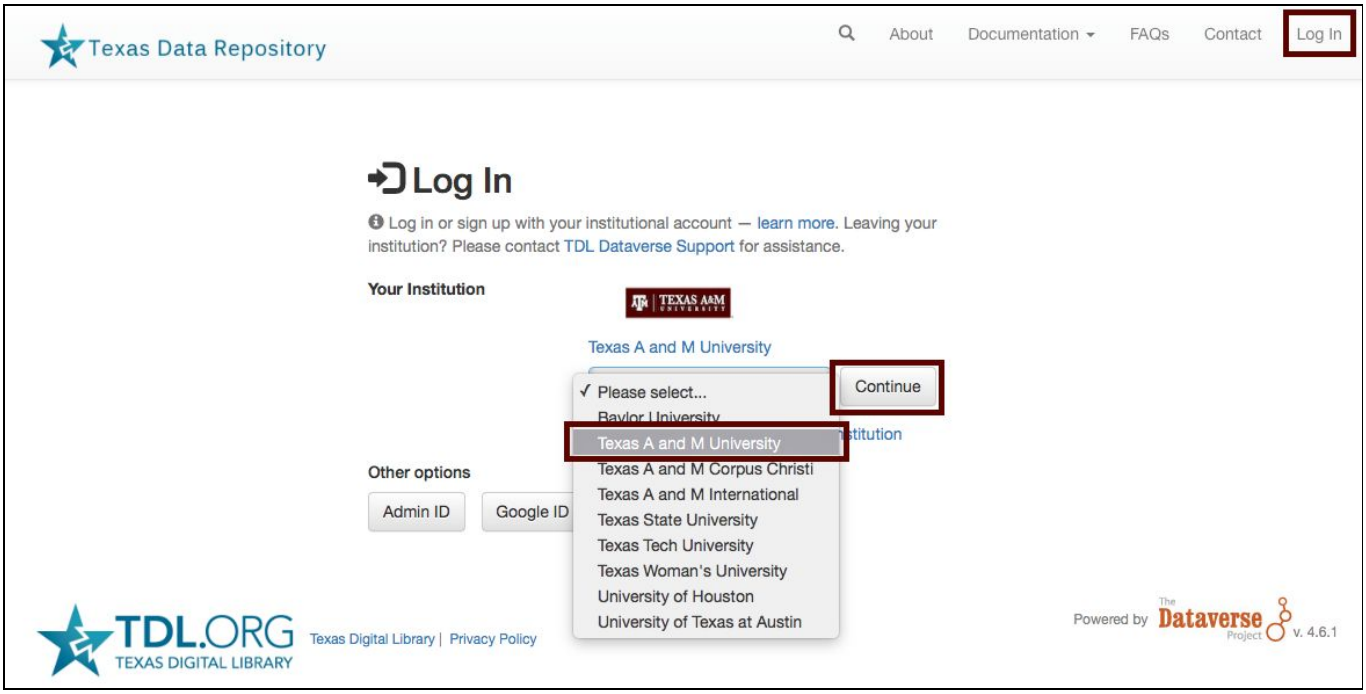
Go to: https://dataverse.tdl.org/dataverse/ilssi



Log In

Click “Log In” at the upper right.

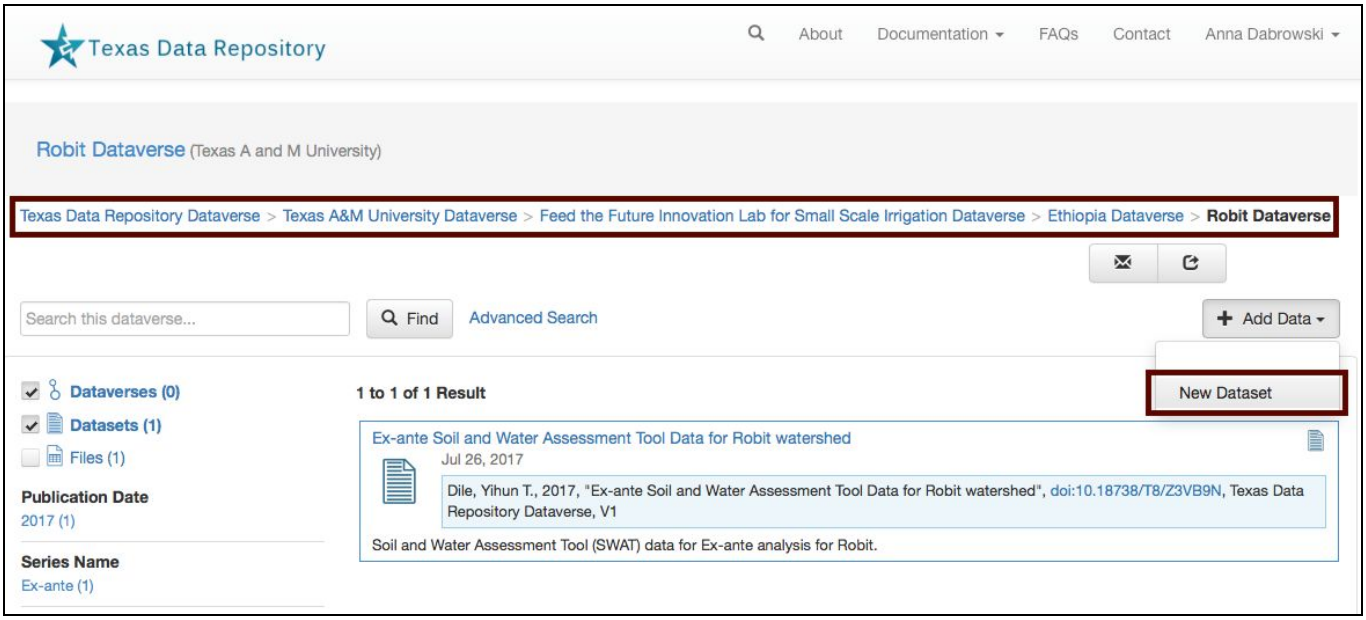
For researchers at Texas A&M University, choose “Texas A and M University.” Use your TAMU NetID and password to log in.



Create a Dataset

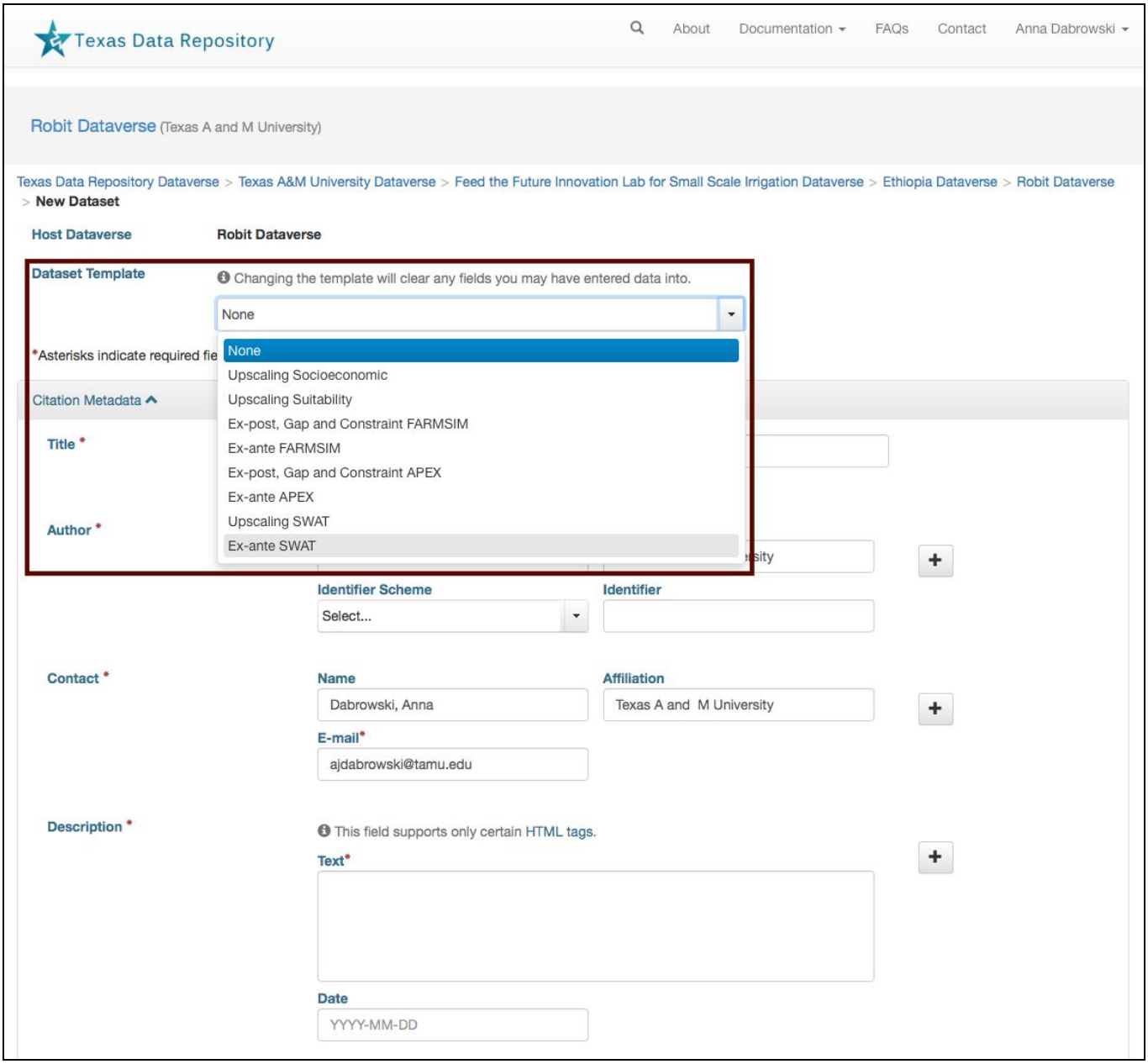
Go to the dataverse for the region where you would like to submit a data.

Choose “New Dataset.”



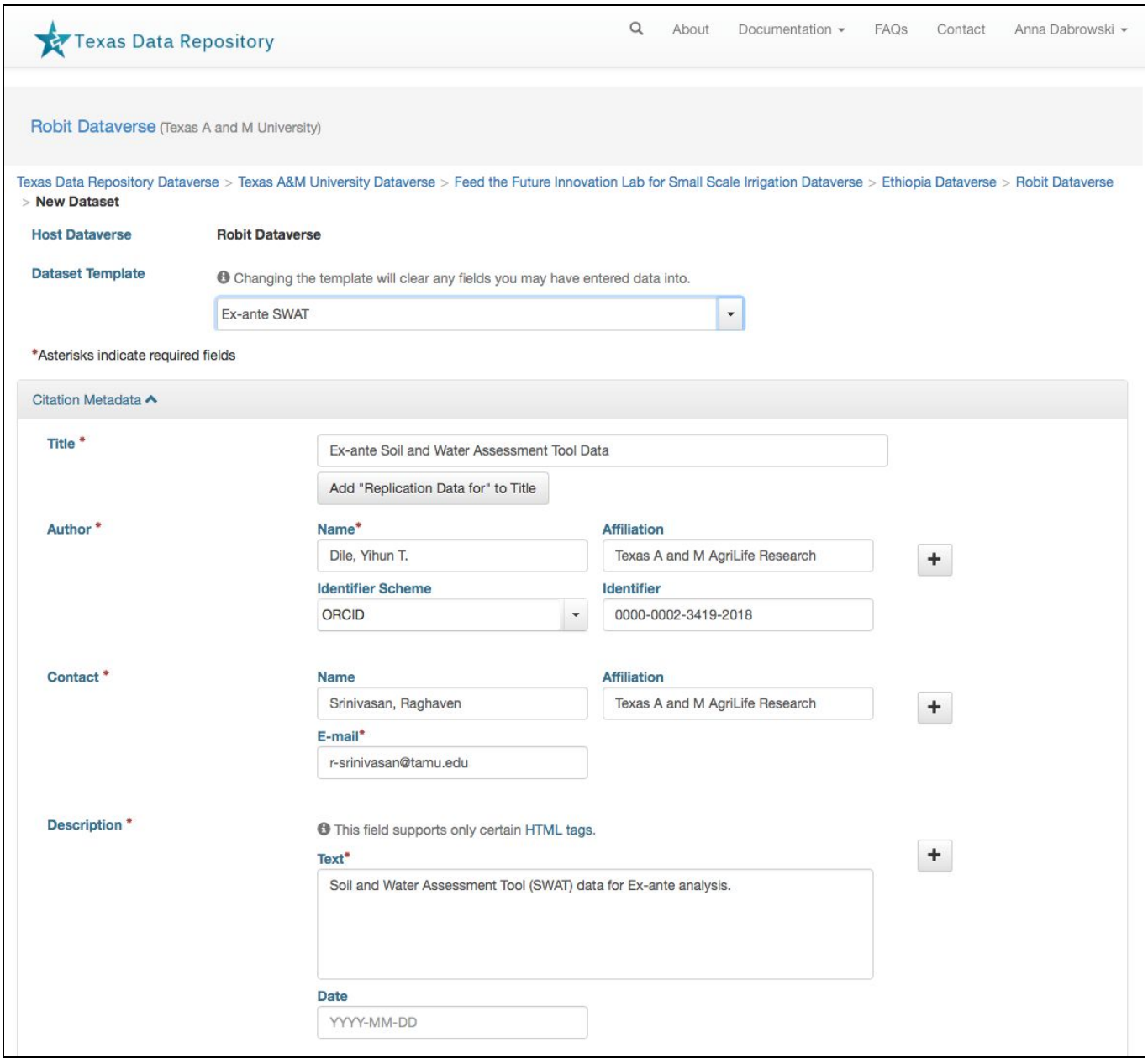
Add Metadata

Choose a “Dataset Template” named after the model and analysis used for your data.



N.B. The templates will automatically fill descriptive information for you.

If a template is incorrect, please contact the ILSSI Program Coordinator or Data Management Librarian to change the template.

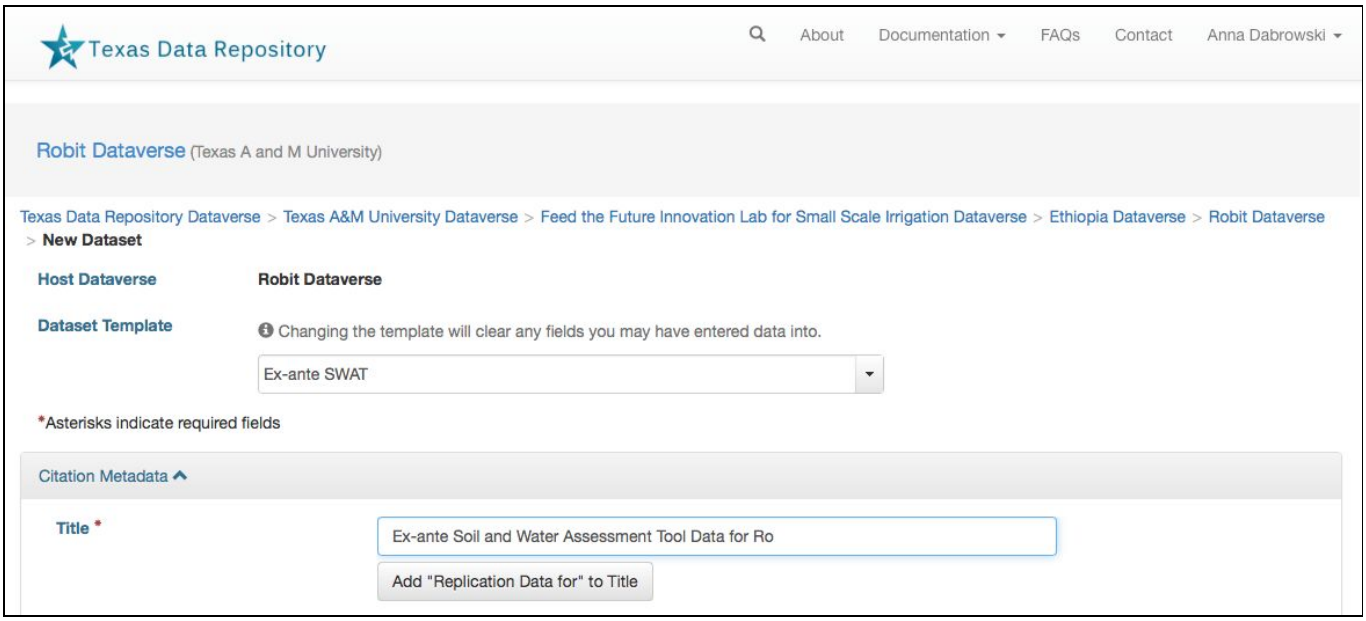


You will need to add some information:

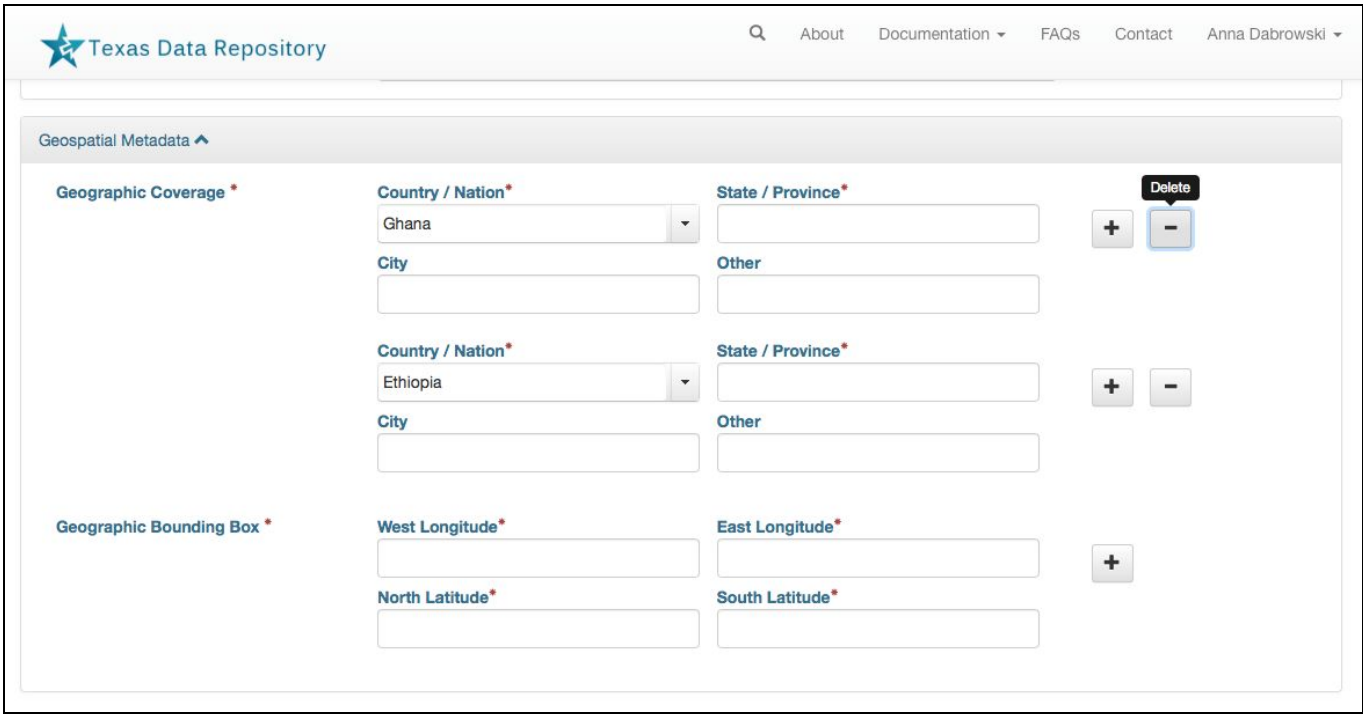
1. Change the title to include the location
2. Choose the correct country
3. Add geographic state/region

N.B. Be consistent in capitalization, spelling, terminology, and the format of this information.

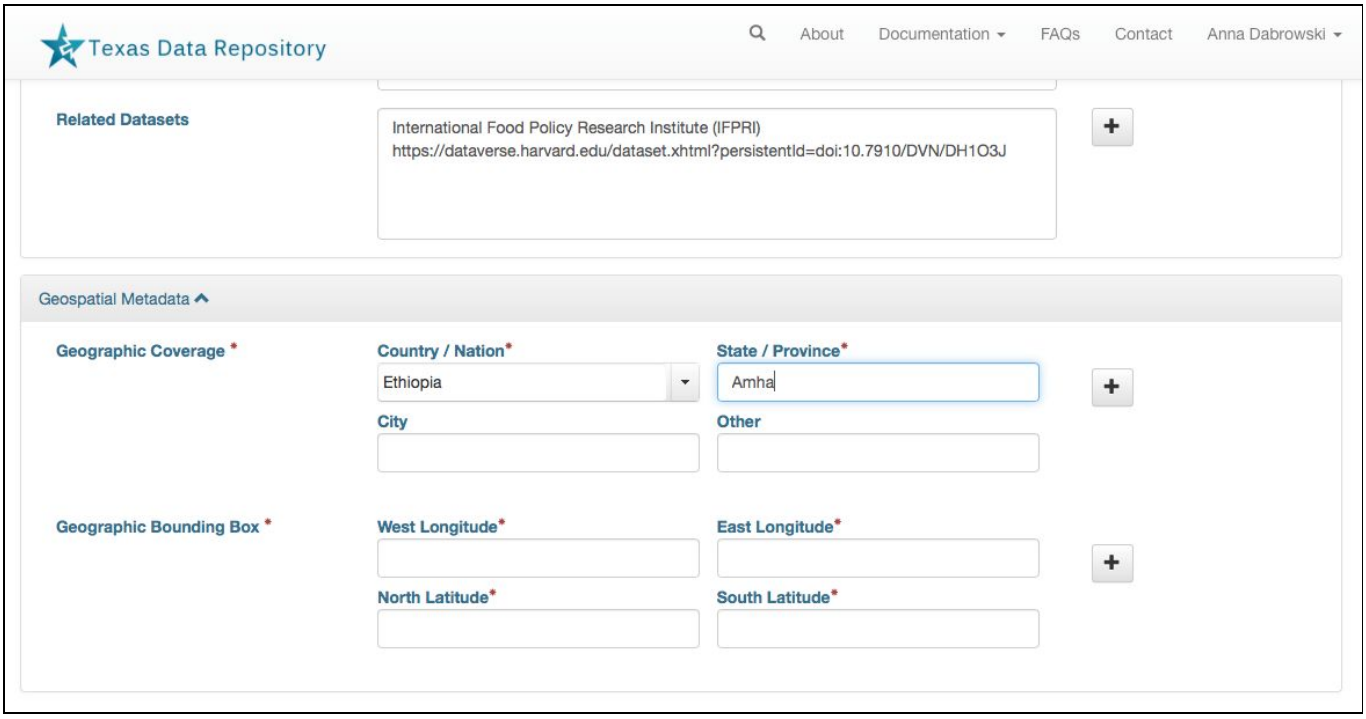
1. Change the title to include the location:



2. Choose the correct country:

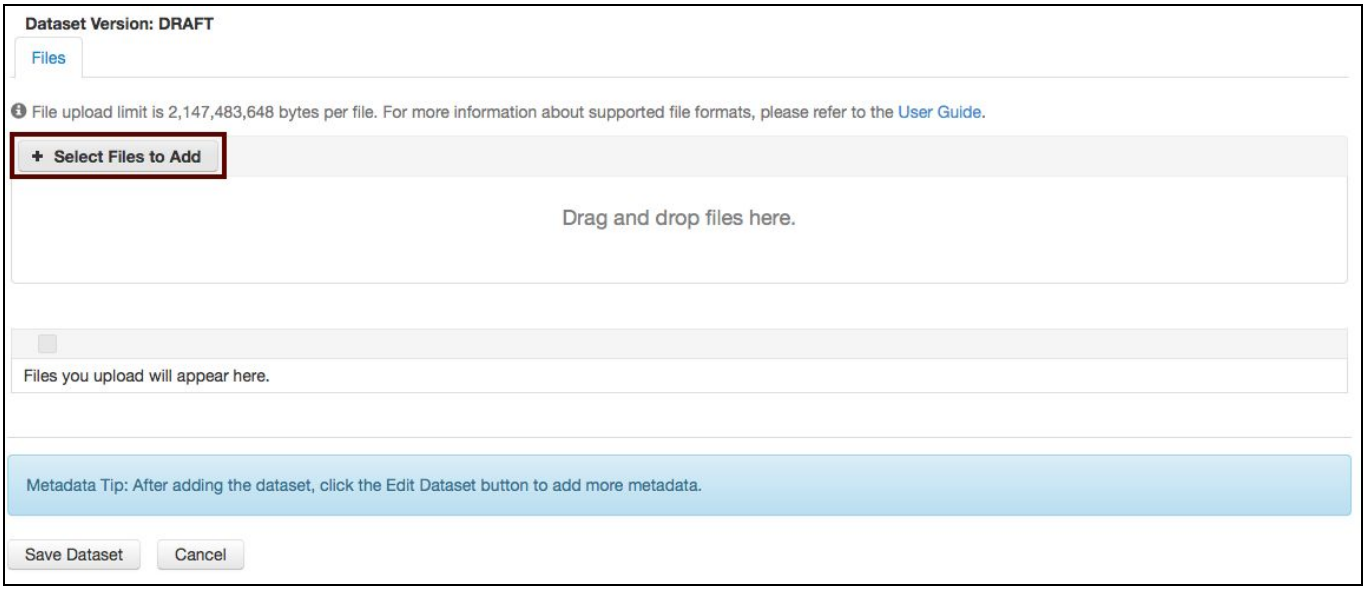


3. Add geographic state/region



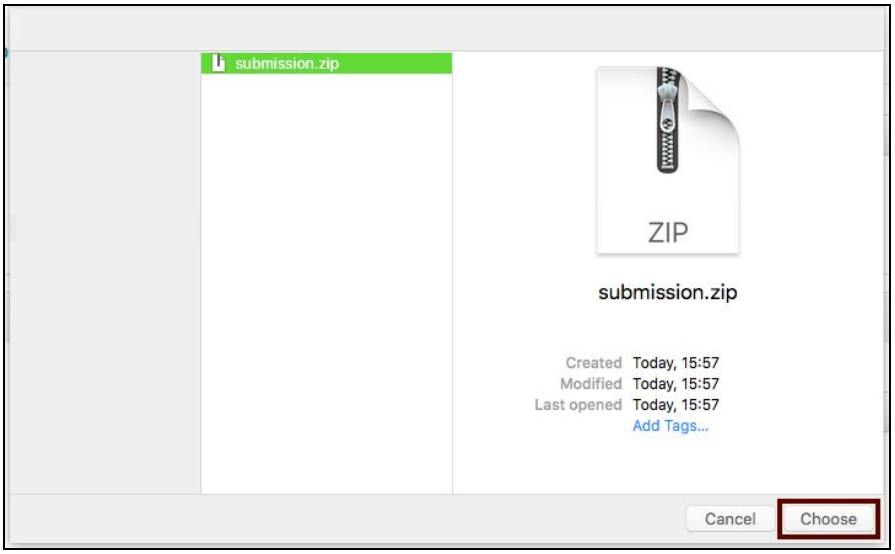
Upload Data Files

Make sure that each ZIP file is less than 2GB, split larger files into multiple ZIP files. Click “Select Files to Add”

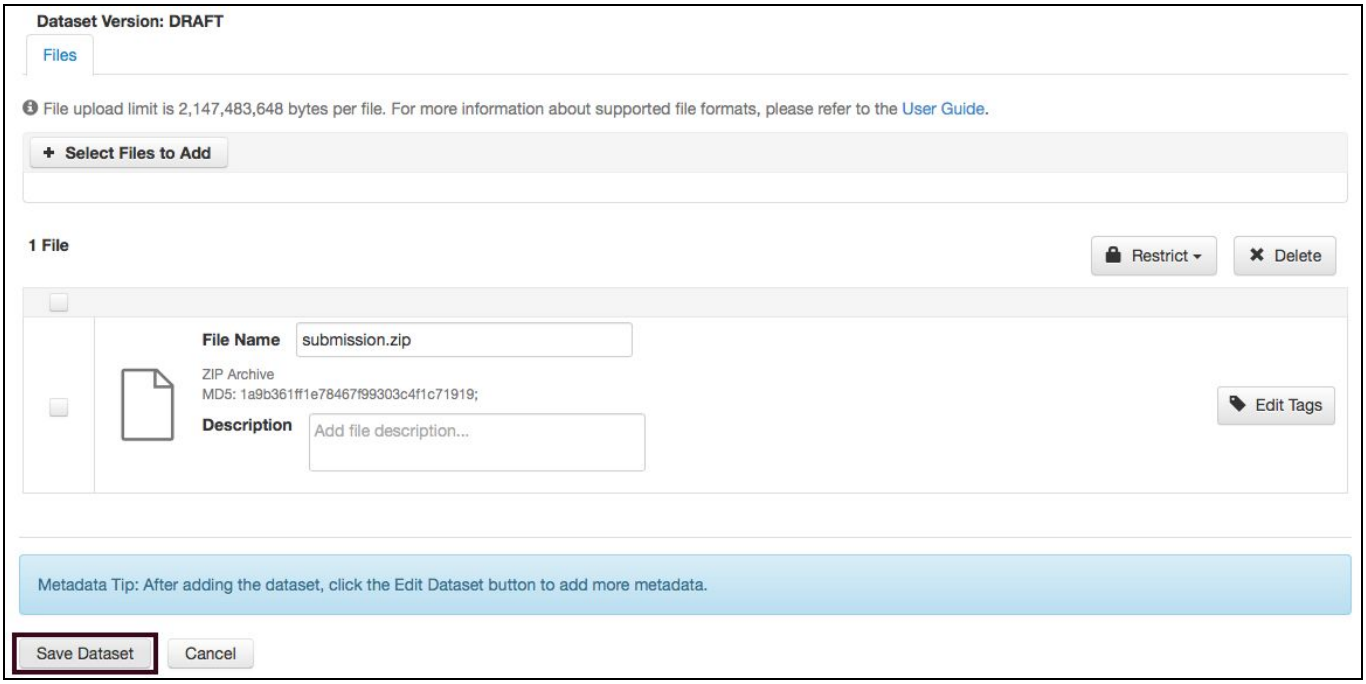


N.B. If your ZIP file contains hundreds of files, the system will be unable to extract them. Dataverse will provide a warning message.

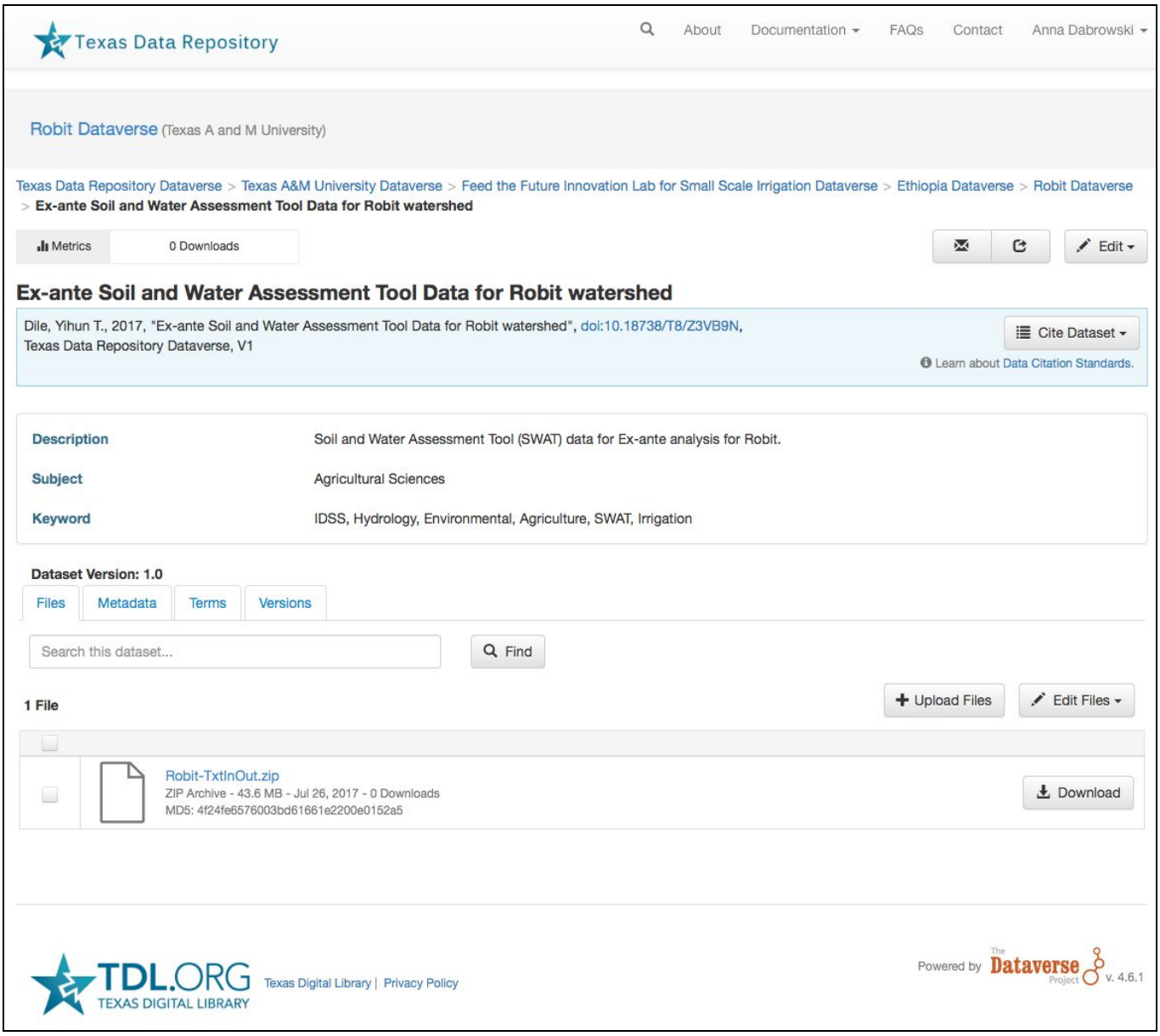
You may continue, this means that the repository will store your ZIP file, without un-zipping the contents.



Once your files have uploaded, click “Save Dataset” to submit the dataset for review.



Your dataset will be “in review” and “unpublished” until an administrator reviews and publishes it. Once it is published, the dataset will be publically visible and the files will be openly accessible for download.



Office of Scholarly Communications contact information

Contact us for information, consultations, and training.

Anna J Dabrowski
Data Management Librarian
ajdabrowski@library.tamu.edu
979.845.8847

Dr. Bruce Herbert
Director, Office of Scholarly Communications
beherbert@library.tamu.edu
979.845.1083