

Two libraries using one Texas Data Repository

Anna J Dabrowski



LIBRARIES
TEXAS A&M UNIVERSITY

Jessica Trelogan



The University of Texas at Austin
University of Texas Libraries

The Texas Data Repository (TDR)



Consortial data repository using the Dataverse platform.

Launched in March 2017.

Hosted by the Texas Digital Library.

9 participating institutions in Texas.

Steering committee composed of institutional liaisons.

Research data management services

Texas A&M University (TAMU)

Data services since 2017.

Dabrowski is Data Management Librarian since March 2017.

Pilot project with Feed the Future Innovation Lab for Small Scale Irrigation (ILSSI).

University of Texas (UT)

Data services since 2011.

Trelogan is Data Management Coordinator since February 2016.

Pilot project with High-Resolution X-ray Computed Tomography Facility (UTCT).

Pilot projects

ILSSI

Five-year (2013–2018) agreement with the United States Agency for International Development (USAID).

Examines mechanisms for improving access to agricultural irrigation technologies internationally.

UTCT

Designated as a national multi-user facility by the National Science Foundation's (NSF) Earth Sciences directorate.

Tomographic scanning and digital image-analysis laboratory, serving the biological, earth, and engineering sciences.

Expressed data needs

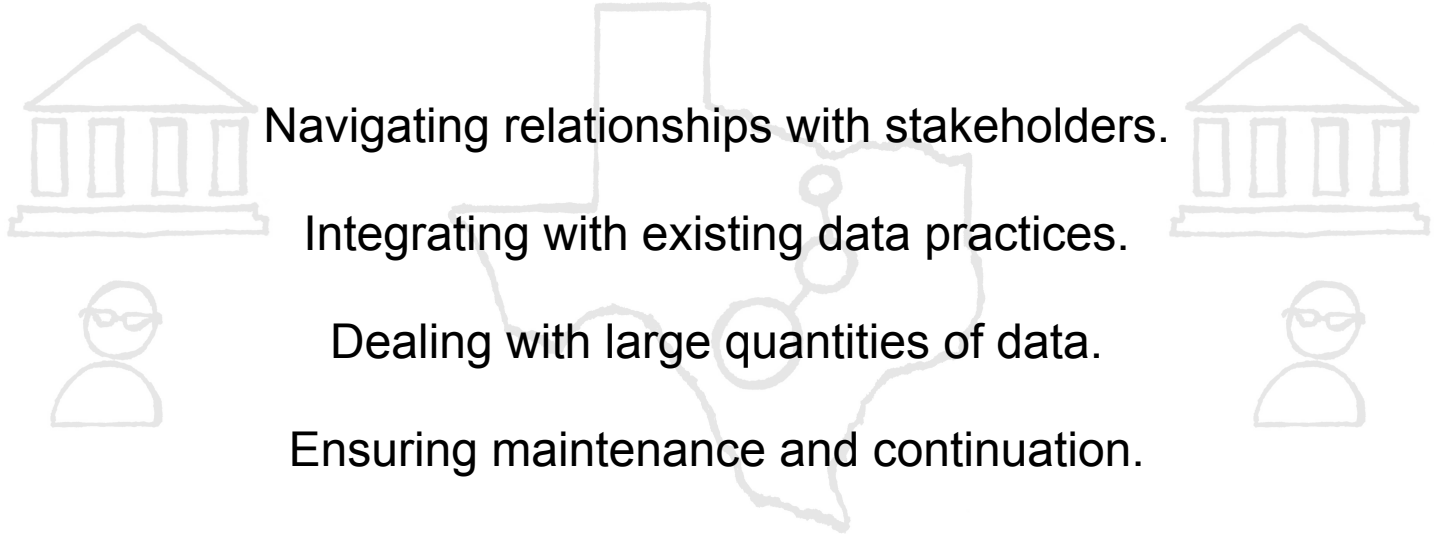
ILSSI

1. Satisfy USAID requirements for data sharing.
2. Provide researchers with a quick dataset submission workflow.
3. Give the Program Manager administrative control over dataset publication.
4. Organize datasets into navigable and extensible collections.
5. Track dataset submission.

UTCT

1. Publish full datasets, including raw data.
2. Increase discoverability of data and visibility of lab.
3. Provide citable, persistent access to data.
4. Maintain existing data documentation and archiving workflows.
5. Develop a long-term data preservation plan.

Shared challenges



Navigating relationships with stakeholders.

Integrating with existing data practices.

Dealing with large quantities of data.

Ensuring maintenance and continuation.

Navigating relationships with stakeholders

ILSSI

UTCT

USAID

University of Texas

ILSSI

UTCT

4 globally distributed partner institutions

Clients and research partners



Integrating with existing data practices

ILSSI

Data distributed across collaborators, including local and external researchers.

Data organization needs to map into a collection hierarchy.

Researchers need to self-report appropriate metadata for templates.

New workflows need to be established.

UTCT

Technical team manages data curation.

Host existing website where select data are shared.

Descriptive metadata generated in text files needs appropriate fields.

Legacy collection and established data capture workflows need to be transferred.

Dealing with large quantities of data

ILSSI

USAID deadline.

Appropriate datasets need to be identified and tracked.

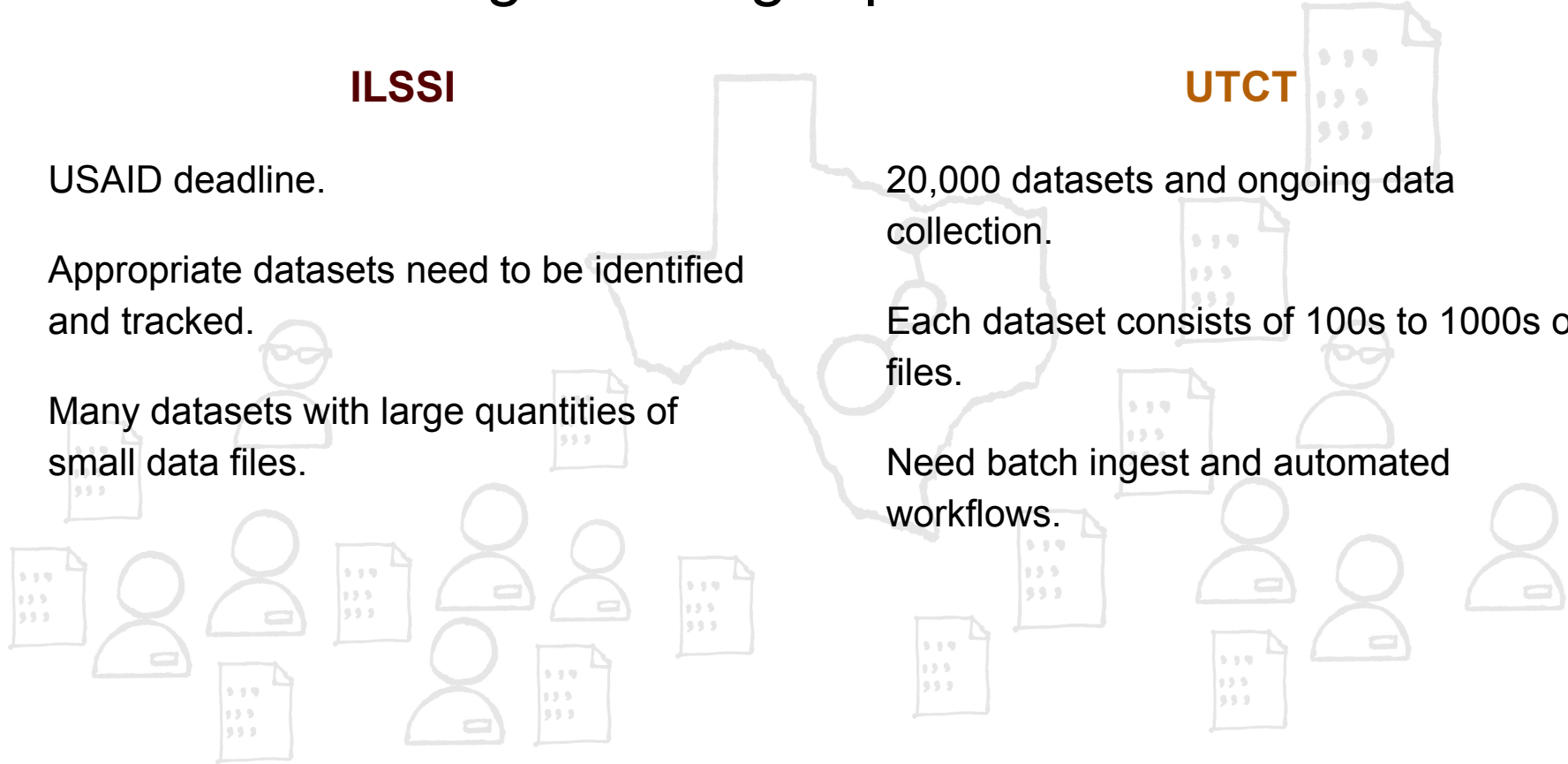
Many datasets with large quantities of small data files.

UTCT

20,000 datasets and ongoing data collection.

Each dataset consists of 100s to 1000s of files.

Need batch ingest and automated workflows.



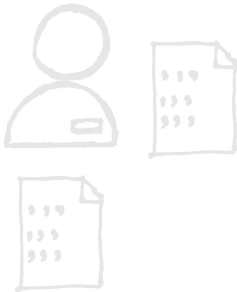
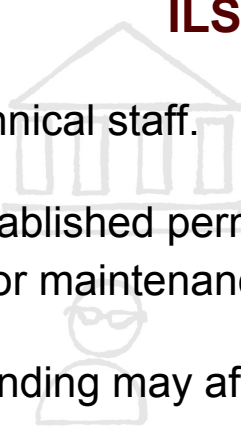
Ensuring maintenance and continuation

ILSSI

Lack technical staff.

Need established permission controls and training for maintenance.

Future funding may affect continuation.

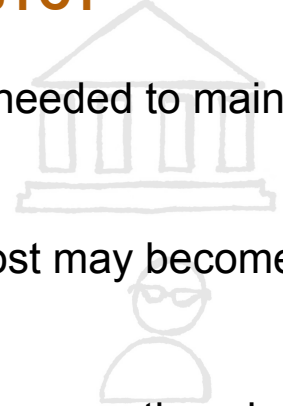


UTCT

Personnel may be needed to maintain collections.

Ongoing storage cost may become a barrier.

Need a long-term preservation plan.



Lessons learned so far

Labs face complicated social, policy, and technical challenges.

Our work surfaces areas for further software and policy development.

The consortial model allows us to share solutions.

Scaling local services requires additional engagement with liaison librarians.

Future directions

Preservation policies and storage cost models.

Continuing to work with labs to identify challenges.

Working with local liaisons to scale at the institutional level

Contributing back to Dataverse community

