
Diving into Data:

Implementing a Data Repository at the Texas Digital Library

TDL Dataverse Implementation Working Group



Panel Outline



The Introduction

(Kristi Park, Texas Digital Library)



The Demo

(Ryan Steans, Texas Digital Library)



The Need

(Bruce Herbert, Texas A&M)



The Future

(Elizabeth Quigley, Harvard IQSS)



The Design

(Santi Thompson, University of Houston)



Q&A



The Benefits

(Santi Thompson, University of Houston)



Part 1:

The Introduction



“Sound, reproducible scholarship rests upon *a foundation of robust, accessible data*. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, *data should be considered legitimate, citable products of research...*”



Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://www.force11.org/datacitation>.



TDL Mission Statement

The Texas Digital Library is a consortium of Texas higher education institutions that *builds capacity* for preserving, managing, and providing access to unique digital collections of enduring value.



Bruce
Herbert,
Texas
A&M
(Chair)



Maria Esteva (TACC); Colleen
Lyon (UT Austin); Jeremy
Donald (Trinity); Martha
Buckbee (UT Southwestern);
Christie Peters, Santi
Thompson (UH); Kristi Park,
Ryan Steans (TDL)

Step 1: Put our heads together

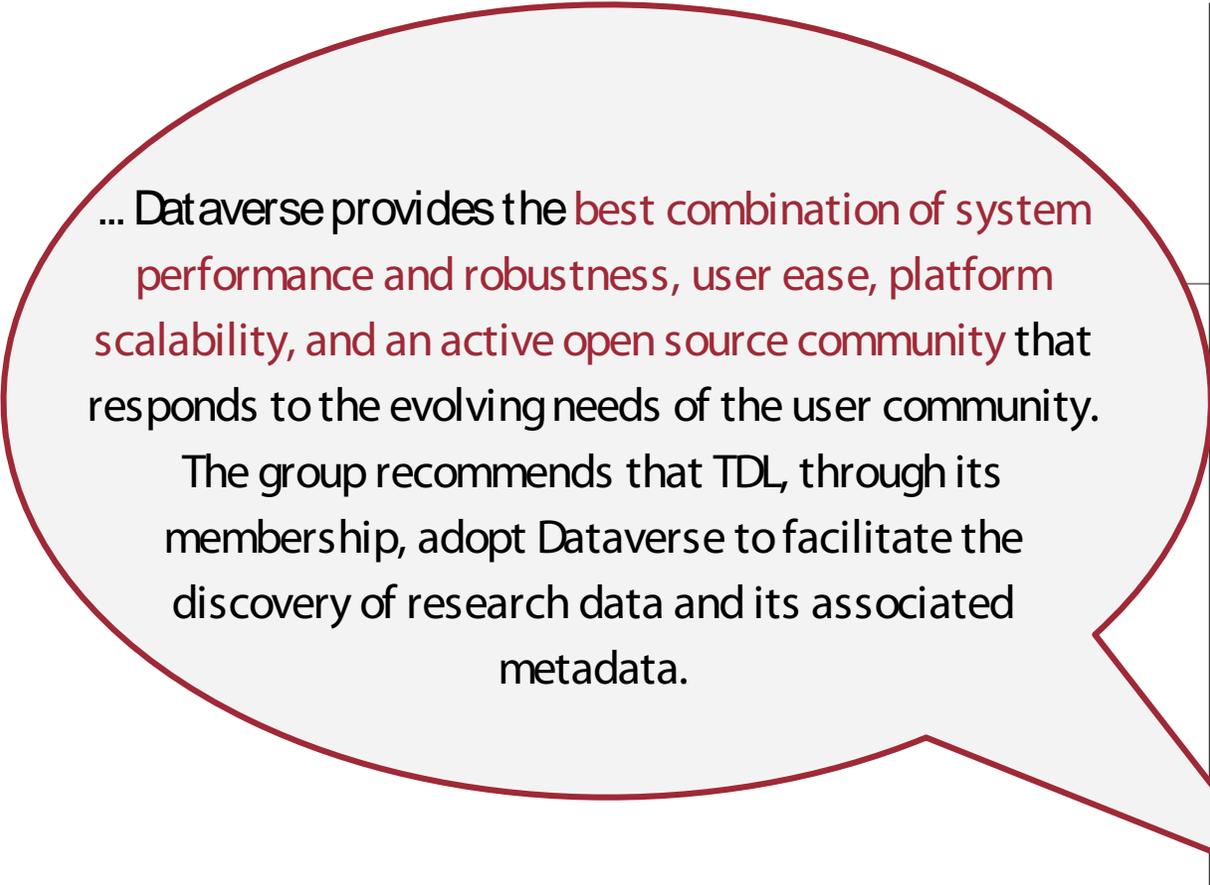
TDL Data Management Working Group

Charge: Help the TDL determine what kinds of data management services it could provide at a consortial level.

- Develop criteria
- Evaluate proposed projects
- Investigate issues
- Document findings
- Make recommendations for services



TDL Data Management Working Group Report



... Dataverse provides the best combination of system performance and robustness, user ease, platform scalability, and an active open source community that responds to the evolving needs of the user community.

The group recommends that TDL, through its membership, adopt Dataverse to facilitate the discovery of research data and its associated metadata.



TDL Data Management Working Group Report

Published August 28, 2015

Table of Contents

Introduction	1
Methodology	2
Evaluation of Dataverse	3
Recommendation	5
Next Steps	5
Appendices	7

Introduction

The need for Data Management services is one of two large-scale needs consistently expressed by Texas Digital Library (TDL) members, a need driven in part by the February 2013 mandate from the White House's Office of Science and Technology Policy to make the results of federally funded research publicly accessible.¹ For more information on how federal agencies plan to implement this policy, please see Appendix D.

The TDL Data Management Working Group convened in Fall 2013 to begin to address this gap, with a particular focus on finding solutions for making research data accessible and reusable.

The charge of the group was to help the Texas Digital Library determine what kinds of data management services it could provide at a consortial level.

Its objectives included:

- Articulating criteria for selecting pilot projects
- Evaluating proposed projects based on that criteria
- Selecting no more than three projects to implement
- Investigating issues related to storage and accessibility of data sets
- Documenting findings and recommendations for services

¹ The February 2013 OSTP directive, entitled "Increasing Access to the Results of Federally Funded Research" mandated that, each Federal agency with over \$100 million in annual research and development expenditures develop a plan to support increased public access to the results of research.



Step 2: Make it happen

TDL Dataverse Implementation Working Group

Charge: Pilot test, assess, and launch a consortial repository for research data archiving and management.

Members; Denyse Rodgers (Baylor); Bruce Herbert, Sean Buckner, Wendi Kaspar, Cecilia Smith (TAMU); Ray Uzwyshyn, Todd Peters (Texas State); Christopher Starcher (Texas Tech); Jeremy Donald (Trinity); Kristi Park, Ryan Steans, Nick Lauland, Laura Waugh (TDL)

14

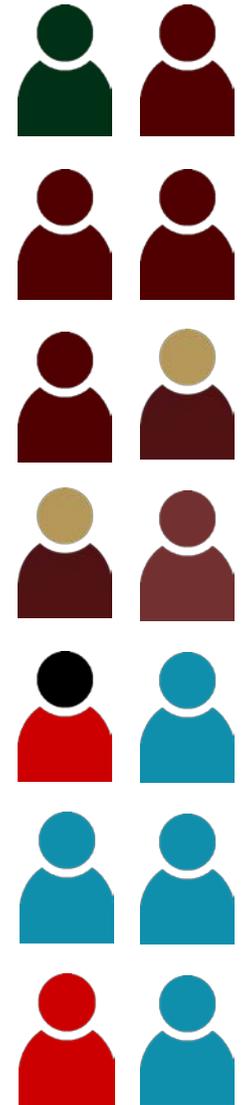
Working Group members

7

TDL member institutions

1

Santi Thompson
(Chair)



Part 2:

The Need

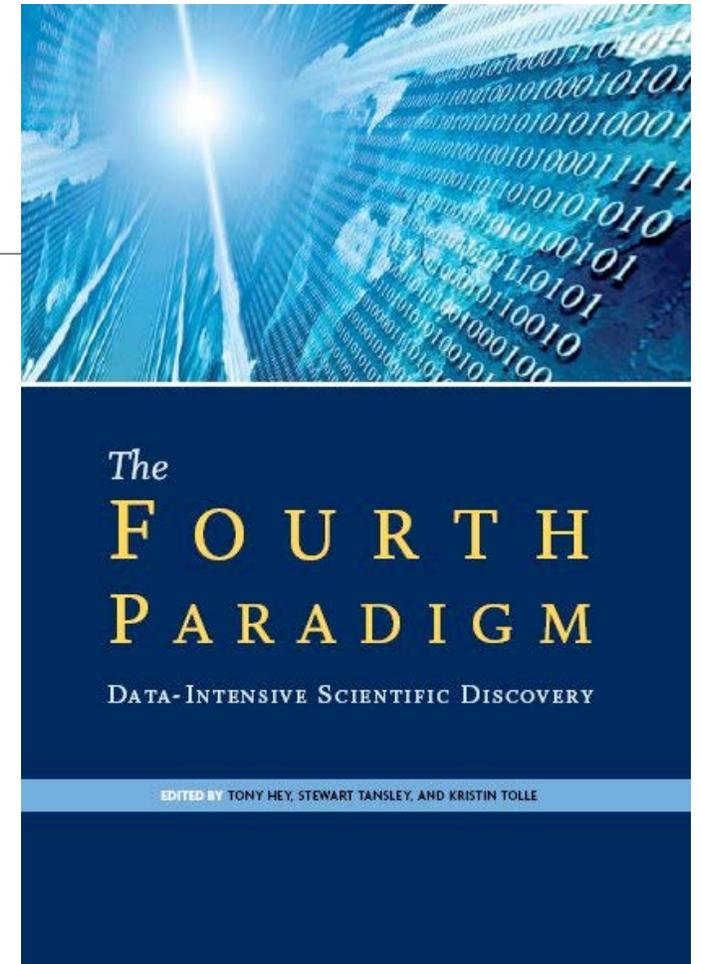
The Fourth Paradigm: Data-Intensive Scientific Discovery

“Jim Gray described his vision of the fourth paradigm of scientific research.

He outlined a two-part plea for the funding of tools for **data capture, curation**, and analysis, and for a **communication and publication infrastructure**.

He argued for the establishment of modern stores for data and documents that are on par with traditional libraries.”

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>



Use Case: Make Research Data Publicly Available

Primary Actors:

PIs of federally funded research

Researchers working on unfunded research or funded research with no retention requirements

Graduate students working on theses, dissertations, or other data-generating projects.

Federal Mandates For Public Access to Research

the WHITE HOUSE PRESIDENT BARACK OBAMA

Get Email Updates | Contact Us

BLOG | PHOTOS & VIDEO | BRIEFING ROOM | ISSUES | the ADMINISTRATION | the WHITE HOUSE | our GOVERNMENT

Home • The Administration • Office of Science and Technology Policy

Search WhiteHouse.gov

Office of Science and Technology Policy

About OSTP | Pressroom | OSTP Blog | Divisions | Initiatives | R&D Budgets | Resource Library | NSTC | PCAST | Contact Us

Expanding Public Access to the Results of Federally Funded Research

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST

E-Mail | Tweet | Share | +

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand of whom recently signed a *We the People* petition asking for expanded public access to the results of taxpayer-funded research.

To see the new policy memorandum, please visit: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

To see Dr. Holdren's response to the *We the People* petition, please visit: <https://petitions.whitehouse.gov/response/increasing-public-access-results-scientific-research>

Michael Stebbins is Assistant Director for Biotechnology at OSTP

GIVE FEEDBACK ABOUT THIS PAGE

YOUR FEDERAL TAXPAYER RECEIPT

Launch the Receipt

WHITE HOUSE BLOGS

The White House Blog
Middle Class Task Force
Council of Economic Advisers
Council on Environmental Quality
Council on Women and Girls

The Library Supports:

Publication repositories

Tools to create data management plans

TDL Data repository

Workflows, standards, & policies

<http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>



Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar , Roger S. Day, Douglas B. Fridsma

Published: March 21, 2007 • DOI: 10.1371/journal.pone.0000308 • Featured in PLOS Collections

Article	Authors	Metrics	Comments	Related Content

Abstract

- Introduction
- Results
- Discussion
- Materials and Methods
- Supporting Information
- Author Contributions
- References

- Reader Comments (6)
- Media Coverage (0)
- Figures

Abstract

Background

Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available.

Principal Findings

We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.

Significance

This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Figures



545 Saves	152 Citations
43,264 Views	107 Shares

[Download PDF](#)

[Print](#) [Share](#)



Included in the Following Collection

[Open Access Collection](#)

Subject Areas

- Microarrays
- Linear regression an...
- Internet
- Archives
- Gene expression
- Clinical trials
- Confidence intervals
- DNA sequence anal...

Open Data Data Citation

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>

ETDs in Institutional Repositories

TEXAS A&M SCHOLARSHIP
GLOBAL IMPACT

Since the 2007 launch of **OAKTrust**, Texas A&M's open access repository, its **>71,000** articles, reports, theses, dissertations and other digital objects have been downloaded *more than* **36 million times** by users in **nearly every country** on Earth.



Contribute Your Scholarship to the World
OAKTRUST.LIBRARY.TAMU.EDU

OPEN ACCESS
Open Access Week is an international event promoting open access as the new default for peer-reviewed scholarship and research.

ATM LIBRARIES
TEXAS A&M UNIVERSITY
The Office for Scholarly Communication spearheads campus-wide initiatives to open, share, and preserve scholarship.

OAKTrust
Open Access to Knowledge
OAKTrust is the University's open-access repository, sharing and preserving the scholarship and research of Texas A&M University's faculty and students.

Enhance the Impact of our ETDs:

Co-publish data sets

Vireo ingestion

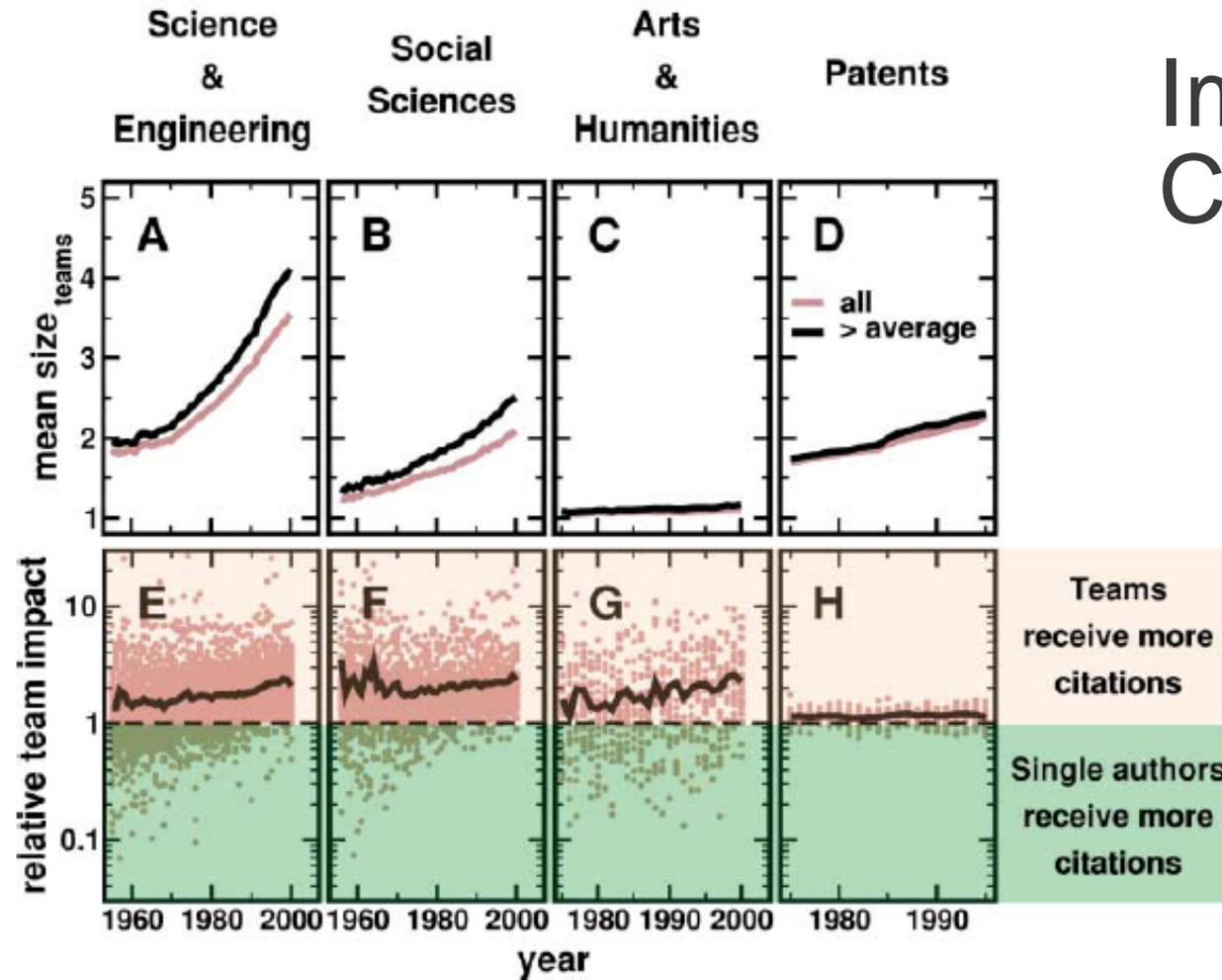
Links in metadata

Workflows, standards, & policies

Use Case: Share Data within a Trusted, Collaborative Network

Primary Actors:

Researchers involved in collaborative teams or networks



Impact of Team Collaboration

Fig. 2. The relative impact of teams. (A to D) Mean team size comparing all papers and patents with those that received more citations than average in the relevant subfield. (E to H) The RTI, which is the mean number of citations received by team-authored work divided by the mean number of citations received by solo-authored work. A ratio of 1 indicates that team- and solo-authored work have equivalent impact on average. Each point represents the RTI for a given subfield and year, whereas the black lines present the arithmetic average in a given year.

Wuchty et al. (2007). *Science* 316(5827): 1036-1039.

Collaboration Across Institutions

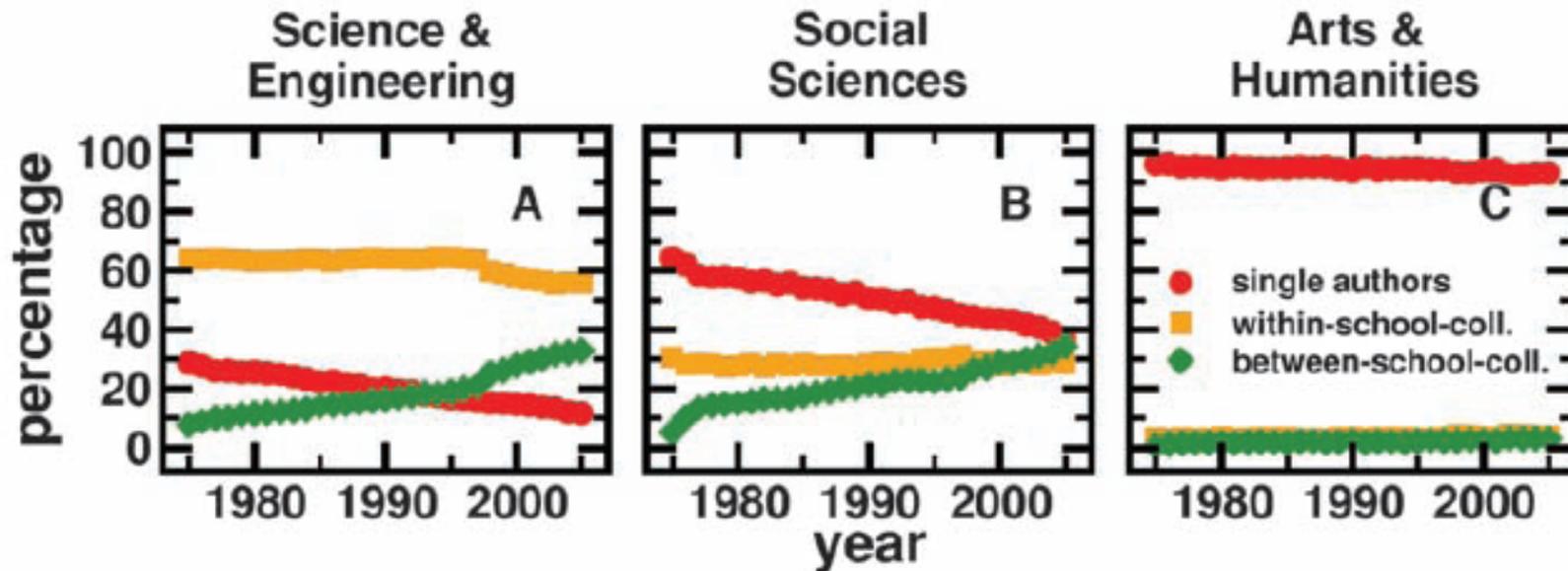


Fig. 1. The rise in multi-university collaboration. By comparing the incidence of papers produced by different authorship structures, we see that the share of multi-university collaborations strongly increases from 1975 to 2005. This rise is especially strong in SE (A) and SS (B), whereas it appears weakly in AH (C), in which collaboration of any kind is rare. The share of single-university collaborations remains roughly constant with time, whereas the share of solo-authored papers strongly declines in SE and SS.

Use Case: Seek Data to (Re)Use

Primary Actors:

Researcher is interested in conducting a meta study reusing data developed in earlier studies

Public using data for personal needs

Organizations seeking data for their needs.

Open Data

Open Sharing of the Paper and the Data

PeerJ PeerJ Computer Science ARTICLES PREPRINTS More ▾  SUBMIT ARTICLE Login

2-year citation median PeerJ articles **4**

[Subscribe](#)  [Download as !\[\]\(03d63843f61b0093ef0dc2a37c30b212_img.jpg\)](#)

Introduction
Methods
Results
Description
Discussion
Conclusions
Supplemental Information
Additional Information and Declarations

 Peer Review history
 Articles citing this paper **3**
 Questions **2**
 Links **5**

Subject areas
Evolutionary Studies
Paleontology
Zoology

23,954 Visitors 29,341 Views 1,368 Downloads
[View all metrics + mentions on the Web](#)

Ontogeny in the tube-crested dinosaur *Parasaurolophus* (Hadrosauridae) and heterochrony in hadrosaurids

[Andrew A. Farke](#)¹, [Derek J. Chok](#)², [Annisa Herrero](#)²,
[Brandon Scolieri](#)², [Sarah Werning](#)³

Published October 22, 2013
PubMed [24167777](#)

 Part of the PeerJ [PeerJ Picks 2014 Collection](#)

 Part of the PeerJ [Top Paleontology Papers - October 2014](#)

July 1, 2014: **(Minor Correction):** "FMNH" was inadvertently omitted from the list of institutional abbreviations. The abbreviation list should include: FMNH, Field Museum of Natural History, Chicago, Illinois, USA.

“ Also see the associated PeerJ [guest blog post](#) by author [Andrew Farke](#) on this paper as well as the ["Dinosaur Joe"](#) website built specifically for this new find.

➤ Author and article information

<https://peerj.com/articles/182/#supplemental-information>



JOE'S DISCOVERY

JOE'S BONES

JOE'S LIFE

JOE'S HOME

JOE'S RECONSTRUCTED SKULL



This reconstruction shows how the skull of "Joe" the baby *Parasaurolophus* might have looked when complete. The keratinous beak has been included here; its attachment to the bone is shown by a subtle line around the upper beak. The model was based upon CT scans of the fossil skull, with missing parts filled in from related dinosaurs.

Reconstruction copyright [Ville Sinkkonen](#), used with permission.

Having trouble viewing the model? This 3D viewer works best on Firefox, Chrome, and Safari (no Internet Explorer, sorry!). Many of the files are viewable as 3D PDFs (via Adobe Acrobat) for download from the [journal article at PeerJ](#), including a [3D pdf of the skull](#). A table with links to all of the raw data hosted at [Figshare](#) (including printable STL files) is [available at PeerJ](#).

Media
Thanks

Copyright © 2015
Raymond Alf Museum.
All Rights Reserved.

Open Data Open Sharing of the Paper and the Data

<http://dinosaurjoe.org/joes-bones/digital-joe/joes-skull-reconstruction/>



Part 3:

The Design



Researcher Use Cases

Title: Researcher needs to make their research data publicly available

Primary Actor



Primary actors may include PIs of federally funded research, researchers working on unfunded research or funded research with no retention requirements, and graduate students working on theses, dissertations, or other data-generating projects.

Title: Researcher needs a virtual research environment to share active data, which may or may not be publicly accessible, within a prescribed collaborative network



Primary Actor

Researchers involved in collaborative networks.

Title: Researcher seeks data to (re)use



Primary Actor

Researcher is interested in conducting a meta study reusing data developed in earlier studies.

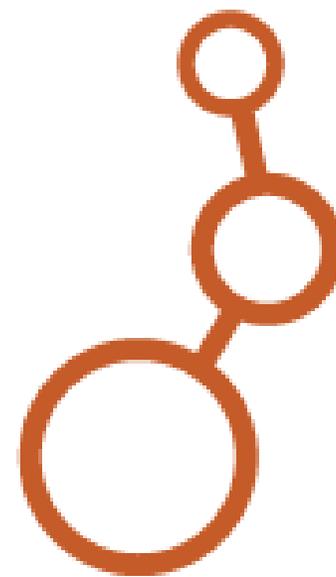


ID	Function	Use Case #	Evaluation Factor	How important is this feature? (Average Score 0-3)	How well does the system perform this function? (Average Score 0-3)	Summary
I.1	Ingest	1	Upload -- the system offers a simple ingest option for user	3	2.5	Platform offers user the ability to drag and drop files from their desktop. Unclear how it would interact with other file destinations (including drop box).
I.2	Ingest	1	Controlled vocabulary -- the system provides users with standardized lists of terms to describe their data (using drop down menus or other interfaces)	2.33	1.4	Controlled vocabulary terms are offered only as a broad list at the subject level.
I.3	Ingest	1	Copyright Permissions Verification/Notification -- the system requires the user to agree to a series of statements regarding	3	0	The system does not alert user of copyright issues or policies prior to the ingest

[View Full Evaluation Matrix](#)



The
Dataverse
Project



TDL Dataverse Implementation Working Group

Policy and Governance

- Sean Buckner
- Santi Thompson
- Ray Uzwysyn

Workflows and Outreach

- Jeremy Donald
- Wendi Kaspar
- Cecilia Smith
- Chris Starcher

Budget/Business Model

- Bruce Herbert
- Kristi Park
- Ryan Steans
- Santi Thompson

Technical Configuration

- Nick Lauland
- Todd Peters
- Denyse Rodgers
- Ryan Steans

Marketing and Coordinator Extraordinaire: Laura Waugh



Policy and Governance

Internal and external policies
creation





Workflows and Outreach

Develop and test workflows for researchers and librarians





Budget and Business Model

Assess costs and identify potential funding models





Technical Configuration

Setup, configure, and test system
and its features





TEXAS RESEARCH DATA REPOSITORY



User Guide

Texas Research Data Repository Pilot Project

Participants in the Texas Research Data Repository Pilot Project are asked to complete the Required Tasks (steps 1 – 8) below. Participants are also welcome to spend as much time as they like in the repository and complete the Optional Tasks (steps 9 – 15).

After completing the Required Tasks (and Optional Tasks, if applicable), participants are asked to complete the follow-up survey and provide feedback and observations about using Dataverse for the Texas Research Data Repository: [SURVEY LINK](#)

Required Tasks:

1. Create a user account
2. Prepare data, code, and additional documentation files
3. Create metadata
4. Create a Dataverse (i.e., collection)
5. Upload information about the dataset
6. Share dataset
7. Publish dataset
8. Download dataset

Optional Tasks:

9. Use mapping and statistical analysis tool
10. Alter default Terms of Use
11. Make your dataset restricted
12. Create multiple versions of a dataset
13. Deaccession a dataset
14. Turn on the Guestbook feature
15. Add a logo to your Dataverse

Resources

- [About the Repository](#)
- [User Guide](#)
- [Licensing](#)
- [Policies](#)
- [Frequently Asked Questions](#)
- [Texas Research Data Repository Metadata Guidelines](#)
- [Data Management Working Group Report](#)
- [Pilot Project Promotional Flyer](#)



Pilot Survey Demographics

Type of Respondent	Participation Percentage
Researchers	31%
Librarians	69%
Overall Rate of:	%
Response	59%
Completion	89%

Completing Required Tasks

Answer	%
Create a user account	100%
Create a Dataverse (i.e., collection)	88%
Upload at least one dataset	100%
Provide metadata information for dataset(s)	100%
Publish dataset(s)	94%
Download a dataset	100%



Completing Optional Tasks

Answer	%
Utilize the mapping analysis tool	17%
Utilize the statistical analysis tool	33%
Request access to a restricted dataset	17%
Utilize versioning of data	17%
Turn on the Guestbook feature in Dataverse	50%
Add a logo to the Dataverse instance that you created	17%



Meet Disciplinary Data Needs?

Answer	%
Extremely well	13%
Very well	56%
Moderately well	6%
Slightly well	25%
Not well at all	0%



Future Repository Services?

Answer	%
Assistance describing data	40%
Assistance setting up a location in the repository for research projects	20%
Assistance finding data in the repository for reuse	20%
Assistance managing data prior to submitting it to the repository	47%
Assistance applying digital preservation best practices with research data	53%



Future Repository Features?

Answer	%
Linking research data with an existing publication	100%
Linking supplemental data with an electronic theses or dissertation	50%
Management of collaborative teams within the data repository	13%
Customizable submission screen with instructions	6%
Development and growth of interdisciplinary research data related to Texas geographic regions and topics	13%



Most Important Benefits?

Answer	%
Fulfill federal mandates for sharing publications and research data	56%
Make your research data more widely available	50%
See statistics on downloads and citations of my data	31%
Make my data citeable through the assignment of a DOI (digital object identifier)	44%
Save versions of your dataset	31%
Collecting all my data in one place	63%

Part 4:

The Benefits



Metrics Shares Coming soon...



Share, publish, and archive your data. Find and cite data across all research fields.

Welcome to the Texas Digital Library Test Dataverse!

IMPORTANT: This Dataverse server does NOT include the [TwoRavens](#) add-on.

Because of this, you may receive errors when ingesting certain datasets and the "explore" button will not work.

<

TRINITY UNIVERSITY
 Trinity University Dataverse

utmb Health
Working together to work wonders™
 UT Medical Branch Dataverse

TEXAS
 University of Texas Dataverse

TEXAS STATE UNIVERSITY
SAN MARCOS
 Texas State University Dataverse

>

Search this dataverse... Find [Advanced Search](#)

- Dataverses (29)**
 - Datasets (20)**
 - Files (46)**
- Dataverse Category**
- Organization or Institution (19)
 - Researcher (5)
 - Research Project (2)
- Publication Date**

1 to 10 of 49 Results Sort << < Previous 1 2 3 4 5 Next > >>

size test 6.1 GB 📄

Jan 15, 2016 - Todd Testverse Dataverse

Peters, Todd, 2016, "size test 6.1 GB", <http://dx.doi.org/10.5072/FK2/8FSKOE>, Texas Digital Library Prod/Test Dataverse, V1

upload test 6.1 GB mp4

New file for multi-gig files 📄

Jan 13, 2016 - nicktest Dataverse

Add Data
Share, publish, and archive

Find Data
Search across disciplines

Get Recognition
Obtain a citation and unique identifier

Part 5:

Texas Research Data Repository Demonstration

What we'll cover

1. What is a Dataverse versus a Dataset?
2. Uploading Data
 - a. Creating a Dataverse
 - b. Creating a Dataset
3. Visualizations
4. Re-Using and Sharing your data

Dataverse and Datasets

- A **Dataverse** is a home for your research project, your community, etc...
 - You can easily build a Dataverse **within** a Dataverse (Ex: University Dataverse > Archaeology Dataverse)
 - You can **stack Dataverses within one another** (Ex: University Dataverse > Archaeology Dataverse > Roman Archaeology Dataverse) to create sub-sub Dataverses
- **Datasets live within a Dataverse** - you can associate multiple datasets within a Dataverse (Ex: Different dig site **datasets** from different locations may all be under one **Dataverse** for a single researcher archaeologist)

A Dataverse Demo (Uploading Data)

The screenshot displays the Dataverse website interface. At the top, the Dataverse logo is on the left, and navigation links for 'About', 'Guides', 'Support', 'Sign Up', and 'Log In' are on the right. Below the navigation, the 'Texas Research Data Repository Dataverse' is identified as part of the 'Texas Digital Library'. A tagline states: 'A statewide collaboration of higher education institutions in Texas'. A 'Metrics' bar shows '234 Downloads'. Below this, a message encourages users to 'Share, publish, and archive your data. Find and cite data across all research fields.' A welcome message follows: 'Welcome to the Texas Research Data Repository, a repository for research data collected by researchers at the Texas Digital Library (TDL) member institutions. Select your TDL member institution below.' A carousel of logos for member institutions is shown: Trinity University, UT Medical Branch, University of Texas, and Texas State University. Below the carousel is a search bar with a 'Find' button and an 'Advanced Search' link, along with an 'Add Data' button. The main content area shows search results for '1 to 10 of 69 Results'. On the left, there are filters for 'Dataverses (47)', 'Datasets (22)', and 'Files (115)'. Under 'Dataverse Category', there are options for 'Organization or Institution (23)', 'Researcher (12)', and 'Research Project (9)'. Under 'Publication Date', there are options for '2016 (43)' and '2015 (26)'. Under 'Author Name', there are options for 'Admin, Dataverse (4)' and 'Institute of Classical Archaeology (2)'. The search results list two datasets: 'Spatial extents of ICA projects' and 'Data on nestbox use', each with a brief description and a DOI link.

Dataverse Search About Guides Support Sign Up Log In

Texas Research Data Repository Dataverse (Texas Digital Library) A statewide collaboration of higher education institutions in Texas

Metrics 234 Downloads Share Copy

Share, publish, and archive your data. Find and cite data across all research fields.

Welcome to the Texas Research Data Repository, a repository for research data collected by researchers at the Texas Digital Library (TDL) member institutions.
Select your TDL member institution below.

Trinity University Dataverse **UT Medical Branch Dataverse** **University of Texas Dataverse** **Texas State University Dataverse**

Find Advanced Search + Add Data

Dataverses (47)
 Datasets (22)
 Files (115)

Dataverse Category
Organization or Institution (23)
Researcher (12)
Research Project (9)

Publication Date
2016 (43)
2015 (26)

Author Name
Admin, Dataverse (4)
Institute of Classical Archaeology (2)

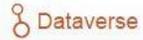
1 to 10 of 69 Results Sort

Spatial extents of ICA projects
May 9, 2016 - Institute of Classical Archaeology Dataverse
Institute of Classical Archaeology, 2016, "Spatial extents of ICA projects", <http://dx.doi.org/10.18738/T8/ZGS0J3>, Texas Research Data Repository Dataverse, V1
This dataset represents the rough location and spatial extents of ICA's field research projects carried out between 1974 and 2013 in the agricultural territories of Metaponto and Croton in southern Italy and Chersonesos in Crimea.

Data on nestbox use
May 6, 2016 - Nestbox use data Dataverse
Murphy, Troy, 2016, "Data on nestbox use", <http://dx.doi.org/10.18738/T8/KVPV4L>, Texas Research Data Repository Dataverse, V1 [UNF:6:E7+zWwvNhu7bMGFj8SJAVQ==]
Data on location of nest boxes and whether they were occupied or not by black-crested titmice.

Nestbox use data Dataverse (Trinity University)

Add a New Dataverse (a home for your datasets)



Texas Research Data Repository Dataverse (Texas Digital Library)

A statewide collaboration of higher education institutions in Texas

Metrics 234 Downloads



Share, publish, and archive your data. Find and cite data across all research fields.

Welcome to the Texas Research Data Repository, a repository for research data collected by researchers at the Texas Digital Library (TDL) member institutions.

Select your TDL member institution below.

Navigation bar with logos for Trinity University, UT Medical Branch, University of Texas, and Texas State University.

Search this dataverse... Find Advanced Search + Add Data

Main content area showing search results for 'Spatial extents of ICA projects' and 'Data on nestbox use'.

Close-up of the '+ Add Data' dropdown menu, highlighting the 'New Dataverse' option with a yellow arrow.

Add Metadata and Settings to your Dataverse

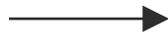
Basic Information



Metadata Fields



Search Facets



Now - add a dataset to your Dataverse

The screenshot shows the Dataverse web interface. At the top, the Dataverse logo and navigation links (About, Guides, Support) are visible. The user is logged in as 'Dataverse Admin'. The main content area displays a success message: 'Success! - You have successfully created your dataverse! To learn more about what you can do with your dataverse, check out the User Guide.' Below this, there are buttons for 'Publish' and 'Edit'. A search bar is present with a 'Find' button and a link to 'Advanced Search'. A prominent '+ Add Data' button is highlighted with a yellow arrow pointing to a callout box. The callout box contains the text 'New Dataverse' and 'New Dataset'. The footer includes copyright information and the Dataverse Project logo.

Dataverse

TCDL2016_Test_Dataverse Dataverse (TDL.org) Unpublished

Texas Digital Library Test Dataverse > TCDL2016_Test_Dataverse Dataverse

Success! - You have successfully created your dataverse! To learn more about what you can do with your dataverse, check out the [User Guide](#).

✉ 🔗 🔄 Publish Edit

This is a test for a live demo dataverse

Search this dataverse... Find Advanced Search + Add Data

Dataverses (0)
 Datasets (0)
 Files (0)

This dataverse currently has no dataverses, datasets, or files. You can add to it by using the Add Data button on this page.

Data Science at The Institute for Quantitative Social Science | Dataverse Project on [Twitter](#) | Code available at [GitHub](#)
Copyright © 2015, The President & Fellows of Harvard College | [Privacy Policy](#)

Powered by **The Dataverse Project** v. 4.2.2 build 54-9883aaa

+ Add Data

New Dataverse
New Dataset

First, add Metadata for each dataset

 Dataverse Search About Guides Support Dataverse Admin 7

TCDL2016_Test_Dataverse Dataverse (TDL.org) Unpublished

Texas Digital Library Test Dataverse > TCDL2016_Test_Dataverse Dataverse > **New Dataset**

Host Dataverse TCDL2016_Test_Dataverse Dataverse

*Asterisks indicate required fields

Citation Metadata ^

Title *

Author *

Name* <input type="text" value="Admin, Dataverse"/>	Affiliation* <input type="text" value="Dataverse.org"/>	<input type="button" value="+"/>
Identifier Scheme Select...	Identifier <input type="text"/>	

Contact *

Name <input type="text" value="Admin, Dataverse"/>	Affiliation <input type="text" value="Dataverse.org"/>	<input type="button" value="+"/>
E-mail* <input type="text" value="Support@tdl.org"/>		

Description *

ⓘ This field supports only certain HTML tags.

Text*

Date

Adding data files requires only a simple upload

Files

i File upload limit is 2GB per file. For more information about supported file formats, please refer to the [User Guide](#).

+ Select Files to Add 

Drag and drop files here.

Files you upload will appear here.

Your files will appear inside the dataset

Metrics 0 Downloads

✉️ 🔗 🔄 Publish Edit

TCDL 2016 Test Dataset Draft Unpublished

Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", <http://dx.doi.org/10.5072/FK2/QL4Q3T>, Texas Digital Library Test Dataverse, DRAFT VERSION [UNF:6:0fwmGJPw+0VwSEggG+fx/g==] [🔗](#) Download Citation

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description This is a test Dataset for TCDL 2016 Demo

Subject Earth and Environmental Sciences; Other

Keyword Earth Science Stuff

Files Metadata Terms Versions

Search this dataset... Find

3 Files + Upload Files Edit Files Download

<input type="checkbox"/>	 catdata.tab Tabular Data - 21.5 KB - May 9, 2016 - 0 Downloads 3 Variables, 1000 Observations - UNF:6:0fwmGJPw+0VwSEggG+fx/g== Explore Download
<input type="checkbox"/>	 climate_data.xls MS Excel - 183.0 KB - May 9, 2016 - 0 Downloads MD5: d7ff4f3a6fa5197df4de8e85a1501d1; Download
<input type="checkbox"/>	 german.data-numeric.txt Plain Text - 99.6 KB - May 9, 2016 - 0 Downloads MD5: dd6d1bd8f4bcc4555b900b9c00955c33; Download

Users can manage Metadata and Terms



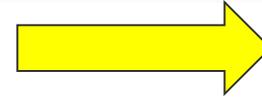
Add to and Edit the full Metadata record



Metadata

Terms

Versions



 Add + Edit Metadata

Citation Metadata 	
Dataset Persistent ID	doi:10.5072/FK2/QL4Q3T
Title	TCDL 2016 Test Dataset
Author	Admin, Dataverse (TDL.org)
Contact	 Use email button above to contact. Admin, Dataverse (TDL.org)
Description	This is a test Dataset for TCDL 2016 Demo (2016-03-03)
Subject	Earth and Environmental Sciences; Other
Keyword	Earth Science Stuff
Producer	TDL Science Labs (Texas Digital Library) http://tdl.org
Depositor	Admin, Dataverse
Deposit Date	2016-05-09

Specialized metadata can be expanded

Geospatial Metadata 

Geographic Coverage *

Country / Nation *	<input type="text" value="United States"/>	State / Province *	<input type="text" value="Texas"/>	<input type="button" value="+"/>
City *	<input type="text" value="Austin"/>	Other	<input type="text"/>	

Geographic Unit

<input type="text"/>	<input type="button" value="+"/>
----------------------	----------------------------------

Geographic Bounding Box

West Longitude	<input type="text"/>	East Longitude	<input type="text"/>	<input type="button" value="+"/>
North Latitude	<input type="text"/>	South Latitude	<input type="text"/>	

Users can manage Terms of Use

The screenshot displays a dataset page for 'TCDL 2016 Test Dataset'. At the top, there are navigation elements including 'Metrics', '0 Downloads', and action buttons for 'Publish' and 'Edit'. The dataset title is followed by 'Draft' and 'Unpublished' status tags. Below this is a citation string: 'Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", http://dx.doi.org/10.5072/FK2/QL4Q3T, Texas Digital Library Test Dataverse, DRAFT VERSION [UNF:6:0fwmGJPw+0VwSEggG+fx/g=]'. A 'Download Citation' button is present. A note states: 'If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.' Below this is a metadata table:

Description	This is a test Dataset for TCDL 2016 Demo
Subject	Earth and Environmental Sciences; Other
Keyword	Earth Science Stuff

Below the metadata table are three tabs: 'Files', 'Terms', and 'Versions'. A yellow arrow points to the 'Terms' tab. To the right, a button labeled 'Edit Terms Requirements' is also highlighted with a yellow arrow. The 'Terms of Use' section is expanded, showing a 'Waiver' section with the text: 'Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.' Below the waiver text is the text 'CC0 - "Public Domain Dedication"' and a 'PUBLIC DOMAIN' logo. The 'Guestbook' section is also expanded, showing a 'Guestbook' entry with the text: 'There are no guestbooks available in TCDL2016_Test_Dataverse Dataverse to assign to this dataset.'



Users may alter their license and terms of access

Terms

Terms of Use ▾

Waiver

Datasets will default to a [CC0 public domain dedication](#). CC0 facilitates reuse and extensibility of research data. Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. If you are unable to give datasets a CC0 waiver you may enter custom Terms of Use for datasets.

Yes, apply CC0 - "Public Domain Dedication" No, do not apply CC0 - "Public Domain Dedication"

Terms of Use

If you are unable to use CC0 for datasets you are able to set custom terms of use. Here is an example of a [Data Usage Agreement](#) for datasets that have de-identified human subject data.

Additional Information [+]

Restricted Files + Terms of Access ▲

Terms of Access

Additional Information [+]

Request Access

Enable access request

Additional Information [+]



Publish Data with one button

The screenshot displays a data repository interface for a dataset titled "TCDL 2016 Test Dataset". The dataset is currently in "Draft" status, as indicated by the orange "Unpublished" tag. The interface includes a "Metrics" section showing "0 Downloads", a "Description" field with the text "This is a test Dataset for TCDL 2016 Demo", and a "Subject" field with the text "Earth and Environmental Sciences; Other". A "Keyword" field contains "Earth Science Stuff". Below the description, there are tabs for "Files", "Metadata", "Terms", and "Versions". A search bar is present with the text "Search this dataset..." and a "Find" button. The "Files" section shows three files: "catdata.tab" (Tabular Data - 21.5 KB - May 9, 2016 - 0 Downloads), "climate_data.xls" (MS Excel - 183.0 KB - May 9, 2016 - 0 Downloads), and "german.data-numeric.txt" (Plain Text - 99.6 KB - May 9, 2016 - 0 Downloads). A callout box highlights the "Publish" button in the top right corner of the dataset view, with an arrow pointing to a larger, more detailed view of the "Publish" button in the bottom right corner. The callout box also shows a "Download Citation" button and a "Test Dataverse," label.

Metrics 0 Downloads

TCDL 2016 Test Dataset **Draft** Unpublished

Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", <http://dx.doi.org/10.5072/FK2/QL4Q3T>, Texas Digital Library Test Dataverse, DRAFT VERSION [UNF:6:0fwmGJPw+0VwSEggG+fx/g==]

If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description This is a test Dataset for TCDL 2016 Demo

Subject Earth and Environmental Sciences; Other

Keyword Earth Science Stuff

Files Metadata Terms Versions

Search this dataset... Find

3 Files + Upload Files Edit Files

File Name	Format	Size	Date	Downloads	MD5	Actions
catdata.tab	Tabular Data	21.5 KB	May 9, 2016	0	UNF:6:0fwmGJPw+0VwSEggG+fx/g==	Explore
climate_data.xls	MS Excel	183.0 KB	May 9, 2016	0	d7#4f3a6fa5197df4de8e85a1501d1;	Download
german.data-numeric.txt	Plain Text	99.6 KB	May 9, 2016	0	MD5: dd6d1bd8f4bcc4555b900b9c00955c33;	Download

Test Dataverse, Download Citation

You have now published your data!

 **Dataverse** 🔍 About Guides ▾ Support  Dataverse Admin 10 ▾

 **Texas Digital Library Test Dataverse** A statewide collaboration of higher education institutions in Texas

 **22 Downloads**    Edit ▾

Share, publish, and archive your data. Find and cite data across all research fields.

Welcome to the Texas Digital Library Test Dataverse!
IMPORTANT: This Dataverse server is for TESTING only.


TRINITY UNIVERSITY
Trinity University Dataverse


utmb Health
Working together to work wonders®
UT Medical Branch Dataverse


TEXAS
University of Texas Dataverse


TEXAS STATE UNIVERSITY
Texas State University Dataverse

Search this dataverse... 🔍 Find Advanced Search + Add Data ▾

 **Dataverses (34)**

 **Datasets (25)**

 **Files (17)**

Publication Status

Published (37)

Deaccessioned (13)

Draft (7)

Unpublished (6)

Dataverse Category

Organization or Institution (21)

Researcher (8)

1 to 10 of 59 Results ⇅ Sort ▾

 **TCDL 2016 Test Dataset** 

May 9, 2016 - TCDL2016_Test_Dataverse Dataverse

Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", <http://dx.doi.org/10.5072/FK2/QL4Q3T>, Texas Digital Library Test Dataverse, V1 [UNF:6:0fwmGJPw+0VwSEggG+fx/g=]

This is a test Dataset for TCDL 2016 Demo

 **TCDL2016_Test_Dataverse Dataverse (TDL.org)** 

May 9, 2016

This is a test for a live demo dataverse

Visualization of Data

TCDL 2016 Test Dataset

Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", <http://dx.doi.org/10.5072/FK2/QL4Q3T>, Texas Digital Library Test Dataverse, V1 [UNF:6:0fwmGJPw+0VwSEggG+fx/g==]

Download Citation

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description This is a test Dataset for TCDL 2016 Demo

Subject Earth and Environmental Sciences; Other

Keyword Earth Science Stuff

Files Metadata Terms Versions

Search this dataset...

Find

Upload Files

Edit Files

Download

3 Files

<input type="checkbox"/>	 catdata.tab Tabular Data - 21.5 KB - May 9, 2016 - 0 Downloads 3 Variables, 1000 Observations - UNF:6:0fwmGJPw+0VwSEggG+fx/g==		Explore	Download
<input type="checkbox"/>	 climate_data.xls MS Excel - 183.0 KB - May 9, 2016 - 0 Downloads MD5: d7ff4f3a6fa5197dff4de8e85a1501d1;			Download
<input type="checkbox"/>	 german.data-numeric.txt Plain Text - 99.6 KB - May 9, 2016 - 0 Downloads MD5: dd6d1bd8f4bcc4555b900b9c00955c33;			Download

TwoRavens allows users to dynamically view data

The screenshot displays the TwoRavens software interface for a dataset named 'catdata'. The interface is divided into several sections:

- Data Selection:** Located on the left, it features two tabs: 'Variables' and 'Subset'. Under 'Variables', three variables are listed: 'color', 'age', and 'favorite-icecream', each with a corresponding icon (bar chart, line graph, and pie chart respectively).
- Model Selection:** Located on the right, it has three tabs: 'Models', 'Set Covar.', and 'Results'. The 'Models' tab is active, showing a list of statistical models: ls, logit, probit, poisson, normal, gamma, negbinom, exp, lognorm, tobit, quantile, logitgee, probitgee, zgammagee, znormalgee, and poissongee.
- Causal Diagram:** In the center, a directed acyclic graph (DAG) illustrates the relationships between variables. 'age' is a parent of 'color', and 'color' is a parent of 'favorite-icecream'.
- Buttons:** At the top right, there are buttons for 'Variable transformation', a refresh icon, and 'Estimate'. Below these are two smaller icons: a magnifying glass and a 'u' symbol.

Maps and Shape Data may work with WorldView

Testing geoconnect to worldmap on this server...

Admin, Dataverse, 2016, "Testing geoconnect to worldmap on this server...", <http://dx.doi.org/10.5072/FK2/3YFPD3>, Texas Research Data Repository Dataverse, V1

Download Citation

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description

A connector between Dataverse and WorldMap. Contribute to geoconnect development by creating an account on GitHub

Subject

Agricultural Sciences

Files

Metadata

Terms

Versions

Search this dataset...

Find

Upload Files

Edit Files

Download

2 Files



t1_2014_17179_roads.zip

Shapefile as ZIP Archive - 1.9 MB - Jan 5, 2016 - 7 Downloads
MD5: d45fe80cd48b96e0c108fa6d6658eceb;

Preview

Explore

Map Data

Download



us_eco_l4.zip

Shapefile as ZIP Archive - 66.4 MB - Jan 5, 2016 - 1 Download
MD5: a674f7c308fbd20d80adac505e9b3fte;

Preview

Explore

Map Data

Download

World Map

The screenshot displays the WorldMap application interface. At the top, the logo "WorldMap" is on the left, and navigation links "Sign in | Create Map | View Map | Help" are on the right. Below the header, a "New Map" title is centered. A secondary navigation bar includes "Add Layers", "Save", "Identify", "Link", "Print", "Gazetteer", "About", "Notes", "Google Earth", "Street View", and "Share Map".

The left sidebar is divided into two sections:

- Overlays:** A sub-section "General" contains a checked item "tl_2014_17179_roads.zip" with a red line icon.
- Base Maps:** A list of map styles with radio buttons, including Google Roadmap, Google Hybrid, Google Terrain (selected), Google Satellite, ESRI Light Gray Reference, ESRI World Imagery, ESRI World Street Map, Stamen Toner, Stamen Watercolor, Bing Aerial With Labels, MapQuest Imagery, MapQuest OpenStreetMap, OpenStreetMap, and No background.

The main map area shows a detailed road network in red over a terrain background. The map is centered on Peoria, Illinois, with labels for various cities and towns such as Daninda, Williamsfield, Laura, Chillicothe, Rome, Washburn, Low Point, Benson, Peoria Heights, Peoria, Bartonville, Glasford, Manito, Forest City, Topeka, Easton, Mason City, New Holland, Lincoln, and Beason. Major roads like I-74, I-55, and I-150 are also visible.

At the bottom of the map, there is a search bar with "Enter search...", "Search", and "Reset" buttons. To the right of the search bar is a scale bar showing 10 km and 5 mi, and a scale indicator "1 : 545979". The Google logo and "Map data ©2016 Google" are visible in the bottom left corner. In the bottom right corner, there is a link for "Center for Geographic Analysis" and a "Report a map error" button.

Reuse of Data via Download

Files Metadata Terms Versions

Search this dataset... Find

3 Files + Upload Files Edit Files ▾ Download

<input type="checkbox"/>	 catdata.tab Tabular Data - 21.5 KB - May 9, 2016 - 0 Downloads 3 Variables, 1000 Observations - UNF:6:0fwmGJPw+0VwSEggG+fx/g==	Explore Download ▾
<input type="checkbox"/>	 climate_data.xls MS Excel - 183.0 KB - May 9, 2016 - 0 Downloads MD5: d7ff4f3a6fa5197dff4de8e85a1501d1;	 Download
<input type="checkbox"/>	 german.data-numeric.txt Plain Text - 99.6 KB - May 9, 2016 - 0 Downloads MD5: dd6d1bd8f4bcc4555b900b9c00955c33;:	Download

Some files have multiple download options

The image shows a screenshot of a data portal interface. At the top right, there is a "Download" button. Below it, a search bar and navigation tabs are visible. A list of three files is shown:

File Name	Format	Size	Date	Downloads	Actions
catdata.tab	Tabular Data	21.5 KB	May 9, 2016	1	Explore, Download
climate_data.xls	MS Excel	183.0 KB	May 9, 2016	0	Download
german.data-numeric.txt	Plain Text	99.6 KB	May 9, 2016	0	Download

A yellow arrow points to the "Download" button for the first file, "catdata.tab". A black arrow points from this button to a dropdown menu that lists the following options:

- Original File Format (Comma Separated Values)
- Tab-Delimited
- RData Format
- Variable Metadata
- Data Subset
- Data File Citation

Easy to share on social media

The image shows a screenshot of the Dataverse website interface. At the top left is the Dataverse logo. The top right navigation bar includes a search icon, 'About', 'Guides', 'Support', 'Sign Up', and 'Log In'. The main content area displays the dataset title 'TCDL2016_Test_Dataverse Dataverse (TDL.org)' and a breadcrumb trail 'Texas Digital Library Test Dataverse > TCDL2016_Test_Dataverse Dataverse'. Below this is a description: 'This is a test for a live demo dataverse'. A search bar with the placeholder 'Search this dataverse...' and buttons for 'Find' and 'Advanced Search' is present. On the left sidebar, there are filters for 'Dataverses (0)', 'Datasets (1)', and 'Files (3)', along with sections for 'Publication Date' (2016 (1)), 'Author Name' (Admin, Dataverse (1)), and 'Subject'. The main results area shows '1 to 1 of 1 Result' for the 'TCDL 2016 Test Dataset' (May 9, 2016) with a description: 'Admin, Dataverse, 2016, "TCDL 2016 Test Da [UNF:6:0fwmGJPw+0VwSEggG+fx/g=]' and a note 'This is a test Dataset for TCDL 2016 Demo'. A 'Sort' dropdown is visible. A yellow arrow points to a share icon (two overlapping documents) in the top right of the dataset entry. A black arrow points from this icon to a 'Share Dataverse' modal dialog box. The dialog box has a title bar 'Share Dataverse' and a close button. It contains the text 'Share this dataverse on your favorite social media networks.' and three social media icons: Facebook (f), Twitter (bird), and Google+ (g+). A 'Close' button is at the bottom left of the dialog.

Citation Options

TCDL2016_Test_Dataverse Dataverse (TDL.org)

Texas Digital Library Test Dataverse > TCDL2016_Test_Dataverse Dataverse > TCDL 2016 Test Dataset

Metrics 2 Downloads

TCDL 2016 Test Dataset

Admin, Dataverse, 2016, "TCDL 2016 Test Dataset", <http://dx.doi.org/10.5072/FK2/QL4Q3T>, Texas Digital Library Test Dataverse, V1 [UNF:6:0fwmGJPw+0VwSEggG+fx/g==]

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description	This is a test Dataset for TCDL 2016 Demo
Subject	Earth and Environmental Sciences; Other
Keyword	Earth Science Stuff

A screenshot of the 'Download Citation' button on the web page. A yellow arrow points to the button. A black arrow points from the button to a dropdown menu that is open, showing two options: 'EndNote XML' and 'RIS Format'.

A screenshot of a Notepad window titled 'DOI-10.5072-FK2-QL4Q3T (1).xml - Notepad'. The window shows the following XML content:

```
<?xml version='1.0' encoding='UTF-8'?><xml><records><record><ref-type name="online Database">45</ref-type><con >
```

Part 6:

Upcoming Dataverse Developments



Community & Future Work

Elizabeth Quigley
User Experience Lead

The Dataverse Community

Dataverse
Community
Meeting

Universities
Using
Harvard
Dataverse

Community
& Working
Groups

Code & Issue
Contributors

Dataverse
Installations

Dataverse
Advisory
Board

Biweekly
Community
Calls

Grant Funded
Collaborations

Dataverse Installations



Dataverse Installations



Future Work

Upcoming Releases:

- 4.4: Widgets Updates & Remote Authentication
- 4.5: Metadata Harvesting & Exporting, Private URL

File Level Metadata:

- Provenance
- Richer support for file level metadata

Sensitive Data Support:

- Secure Data Storage for Harvard Dataverse (hosting set up)
- DataTags compliant version of Dataverse

Large Data Support:

- Streaming social sciences data, e.g. - billions of GeoTweets
- Biomedical large scale data (SBGrid Repository)

Funding Agencies





Welcome to the Dataverse Community!



Questions and Discussion

