A Consortial Approach to Research Data Repository Services

Laura Waugh, Texas Digital Library
On behalf of the TDL Dataverse Implementation Working Group
Overview

- Background
- Deciding on Dataverse
- About Dataverse
- TDL Dataverse Implementation
- Next Steps
- Q&A
BACKGROUND
Background
Background

**Texas Digital Library:** A consortium of Texas higher education institutions that builds capacity for preserving, managing, and providing access to unique digital collections of enduring value.
DECIDING ON DATAVERSE
TDL Data Management Working Group

Fall 2013 - 2015

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

Bruce Herbert (Chair)
Texas A&M

Santi Thompson, University of Houston Libraries
Maria Esteva, Texas Advanced Computing Center
Colleen Lyon, UT Austin Libraries
Jeremy Donald, Trinity University Libraries
Martha Buckbee, UT Southwestern Libraries
Mark Phillips, UNT Libraries
Kristi Park, Texas Digital Library
Ryan Steans, Texas Digital Library
Title: **Researcher needs to make their data publicly available**

Primary Actor: Primary actors may include PIs of federally-funded 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.
<table>
<thead>
<tr>
<th>ID</th>
<th>Function</th>
<th>Use Case #</th>
<th>Evaluation Factor</th>
<th>How important is this feature? (Average Score 0-3)</th>
<th>How well does the system perform this function? (Average Score 0-3)</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.1</td>
<td>Ingest</td>
<td>1</td>
<td>Upload -- the system offers a simple ingest option for user</td>
<td>3</td>
<td>2.5</td>
<td>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).</td>
</tr>
<tr>
<td>I.2</td>
<td>Ingest</td>
<td>1</td>
<td>Controlled vocabulary -- the system provides users with standardized lists of terms to describe their data (using drop down menus or other interfaces)</td>
<td>2.33</td>
<td>1.4</td>
<td>Controlled vocabulary terms are offered only as a broad list at the subject level.</td>
</tr>
<tr>
<td>I.3</td>
<td>Ingest</td>
<td>1</td>
<td>Copyright Permissions Verification/Notification -- the system requires the user to agree to a series of statements regarding</td>
<td>3</td>
<td>0</td>
<td>The system does not alert user of copyright issues or policies prior to the ingest</td>
</tr>
</tbody>
</table>
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.

Full Report: http://hdl.handle.net/2249.1/68438
Why Dataverse?

• (Relatively) simple **self-service** ingest workflow for researchers
• Ability to **share with trusted group of researchers** prior to publication
• Creation of **DOIs** for data citation
• Helps researchers fulfill **Data Management Plan** requirements
• Comes with an **active open source community** led by Harvard IQSS
ABOUT DATAVERSE
About Dataverse

• Open Source Software
• Developed and used by Harvard
• Collaboration of the Institute for Quantitative Social Sciences (IQSS) and Harvard University Libraries
• 17 installations worldwide
About Dataverse

Schematic Diagram of a Dataset in Dataverse 4.0

Container for your data, documentation, and code.
About Dataverse

Schematic Diagram of a Dataverse in Dataverse 4.0

Container for your Datasets and/or Dataverses*

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)
**TDL Dataverse Implementation Working Group**

**Fall 2015 - Present**

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

- TDL Dataverse Installation
- Pilot test and Assessment
- Create Documentation
- Prepare for Launch
- Name the Repository!

**Santi Thompson, UH Libraries (Chair)**

- Denyse Rodgers, Baylor Libraries
- Bruce Herbert, Texas A&M Libraries
- Sean Buckner, Texas A&M Libraries
- Wendi Kaspar, Texas A&M Libraries
- Cecilia Smith, Texas A&M Libraries
- Ray Uzwyshyn, Texas State University Libraries
- Todd Peters, Texas State University Libraries
- Christopher Starcher, Texas Tech University Libraries
- Jeremy Donald, Trinity University Libraries
- Kristi Park, Texas Digital Library
- Ryan Steans, Texas Digital Library
- Nick Lauland, Texas Digital Library
- Laura Waugh, Texas Digital Library

**Texas Digital Library**
Sidenote

Name the TDL Data Repository Contest!

• Submit ideas through Friday, August 5th (today)!
• Contributor of the winning name will win a $50 Amazon gift card

http://goo.gl/forms/al04IGQHZgsa3Y9p2
Pilot Project

TDL Dataverse Implementation Working Group

• Pilot test March - April 2016

<table>
<thead>
<tr>
<th>Overall Rate of:</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>59%</td>
</tr>
<tr>
<td>Completion</td>
<td>89%</td>
</tr>
</tbody>
</table>
Pilot Project

Required Tasks Completed

- Download a dataset
- Upload at least one dataset
- Provide metadata information for dataset(s)
- Upload at least one dataset
- Create a Dataverse (i.e., collection)
- Create a user account

Texas Digital Library
Pilot Project

Optional Tasks Completed

Utilize the mapping analysis tool
Utilize the statistical analysis tool
Request access to a restricted dataset
Utilize versioning of data
Turn on the Guestbook feature in...
Add a logo to the Dataverse instance
Other

Other:
- Edit Terms of Use
- Deaccession data
- Make a dataset restricted
Pilot Project

Disciplines:
- Physics
- Geography
- Ecosystem Science & Mgmt.
- Nutritional Sciences
- Hospitality & Retail Mgmt.
- Family & Consumer Sciences
- Geology
- Behavioral Ecology
- Sociology

Meeting Data Needs for Your Discipline
TDL Dataverse

TDL Dataverse Implementation Working Group Documentation

- User Guides
- Policies and Terms of Use
- Promotional Materials/Talking Points
- Memorandum of Understanding
- Site Branding (e.g., Homepage)
- API Guide

http://data.tdl.org/
TDL Dataverse

Institutional dataverses
TDL
- Technology Hosting
- Helpdesk (tech support)
- Training
- Coordination & Governance

TDL Dataverse Repository

Member Libraries
- Data Curation Support, Mediated Deposit, etc.
- Campus Outreach
- Active Management of Institutional “dataverse”

Researchers
- Deposit and publish data
- Share data with research group
TDL Dataverse

Additional Integrations:

• EZID accounts for generating DOIs
• TwoRavens: Harvard-built package for statistical data exploration and analysis
• Geoconnect for WorldMap: Connects Dataverse files to Harvard’s WorldMap for GIS data exploration
• Shibboleth authentication system

http://data.tdl.org/
NEXT STEPS
Next Steps

Implementation and Outreach

• Finalizing documentation through August 2016
• Soft launch September 2016
• Data Symposium November 15th and 16th 2016 (Save the Date!)
• Additional training for library liaisons throughout fall and spring (2016 - 2017)
Thank you!

- Laura Waugh, l.waugh@austin.utexas.edu
- TDL Information, info@tdl.org
- TDL Data Repository Homepage: http://data.tdl.org/
- Name the Repository Contest: http://goo.gl/forms/al04IGQHZgsa3Y9p2