Community-driven Collaboration:
Texas Data Repository
Updates & Activities

Laura Waugh, Texas State University Digital Collections Librarian & TDR Steering Committee Chair, with Courtney Mumma, MAS/MLIS, TDL Deputy Director
2019 Dataverse Community Meeting
tdl.org
The Texas Digital Library consortium builds capacity for preserving, managing, and providing access to unique digital collections of enduring value.
Joined Global Consortium

The Global Dataverse Community Consortium
Supporting Dataverse repositories around the world.

http://dataversecommunity.global/
*unofficial* Core Trust Seal

- Draft: Michael Nugent, Library Technical Associate from UT Austin, [Center for Transportation Research](https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/913014797/Core+Trust+Seal+documentation)
- [https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/913014797/Core+Trust+Seal+documentation](https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/913014797/Core+Trust+Seal+documentation)
Data Repositories Conformant with the DOT Public Access Plan

Researchers are encouraged to use publicly accessible repositories for the deposit of their data, where appropriate and available.

Conformant Repositories

The list is updated regularly.

- Dryad -- http://www.datadryad.org [Reviewed 2016-01]
- Inter-university Consortium for Political and Social Research (ICPSR) -- https://www.icpsr.umich.edu [Reviewed 2016-01]
- Iowa State University DataShare -- https://iastate.figshare.com [Reviewed 2019-05]
- Missouri University of Science and Technology Scholars’ Mine -- http://scholarsmine.mst.edu [Reviewed 2019-05]
- Odum Institute Data Archive -- https://dataverse.unc.edu
- Texas Digital Library Texas Data Repository -- https://data.tdl.org [Reviewed 2019-05]
- Zenodo -- http://zenodo.org [Reviewed 2016-01]

Locate Potential Data Repositories

Researchers may use the Registry of Research Data Repositories’ searchable listing of data repositories in the United States as a starting point to locate potential archiving options for their data.

Software Improvements

- Jim Myers, IQSS, contractor for TDL
- Highlights
  - DVUploader, a command line bulk uploader for dataverse
  - Automated assignment to TDR liaisons of administrative permissions for any new Dataverse within their institutional Dataverse
  - Move to S3
- All work contributed back to the core Dataverse
Assessment Group

● Monthly Metrics Reporting
  ○ Goal: Evaluate institutional monthly metrics
  ○ Considerations: Google Analytics and the Making Data Count requirements
  ○ Poster presentation here

● Usage Analysis
  ○ Goal: Establish yearly usage updates
  ○ Poster presentation here
# Monthly Metrics Reporting

## Considerations
- TDR Committee **Survey**
- Make Data Count **Code**
- **Google** Analytics
- Texas ScholarWorks (TSW)

## Metrics Categories
- Datasets and collections
- Users
- Views and access
- Sessions

<table>
<thead>
<tr>
<th>Metrics about datasets and collections</th>
<th>Survey</th>
<th>Code</th>
<th>Google</th>
<th>TSW</th>
<th>Source</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset download counts</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File, Google</td>
<td>yes</td>
</tr>
<tr>
<td>Size of datasets (MB)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td>Number of files within datasets</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>maybe</td>
</tr>
<tr>
<td>List of dataset DOIs</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td>Size of collections (MB)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td>Hierarchy collections and datasets</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>maybe</td>
</tr>
<tr>
<td>List of links to all collections and datasets</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Metrics about users</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institutional researchers using the TDR</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes*</td>
</tr>
<tr>
<td>Unique page visitors</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Google</td>
<td>yes</td>
</tr>
<tr>
<td>Percent new users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Percent users from referring site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Percent users directed from search engines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>maybe</td>
</tr>
<tr>
<td><strong>Metrics about views and access</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique Dataset Investigations</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td>Total Dataset Investigations</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File</td>
<td>yes</td>
</tr>
<tr>
<td>Number of pageviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>maybe</td>
</tr>
<tr>
<td>Mode of Access</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Log File, Google</td>
<td>maybe</td>
</tr>
<tr>
<td><strong>Metrics about sessions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sessions</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Google</td>
<td>yes</td>
</tr>
<tr>
<td>Average session duration</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Google</td>
<td>yes</td>
</tr>
<tr>
<td>Path Through the Site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Referral Traffic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Site Content Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Bounce Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
<tr>
<td>Search Terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google</td>
<td>no</td>
</tr>
</tbody>
</table>
Monthly Metrics Reporting

Continue to Include
- Dataset download counts
- Size of datasets
- List of dataset DOIs
- List of links to all dataverses and datasets

Recommended Additions
- Size of dataverses
- Unique page visitors
- Unique dataset views
- Total dataset views
- Average session duration
Usage Statistics

Key Findings
○ Increase in registered users vs. dataset and dataverse creators
○ More datasets created than released (published)
○ Majority of datasets are under 10GB (per requirements)

Considerations
○ Available staff time and resources
○ Researcher practices and outputs vary
○ Differing policies/practices at member institutions
Webinars

Topics in Research Data Management

- Participants - THANK YOU!
  - Odum Institute
  - UVA Libraries
  - MIT Libraries
  - University of Toronto Libraries
  - Carnegie Mellon University Libraries
  - Harvard Dataverse Project (upcoming, June 25th)

- All slides and closed captioned video available online:
  - Slides: https://tdl-ir.tdl.org/handle/2249.1/76164
  - Videos: https://www.youtube.com/user/texasdigitallibrary/videos
Coming Soon

- Official Core Trust Seal - [https://www.coretrustseal.org/](https://www.coretrustseal.org/)
- Chronopolis Distributed Digital Preservation- [https://libraries.ucsd.edu/chronopolis/](https://libraries.ucsd.edu/chronopolis/)
- Bigger Data Support
- Roadmap
- Fun new logo!
- Confluence documentation and resource sharing enhancements
Thank you and Questions?

Courtney Mumma, c.mumma@austin.utexas.edu
Laura Waugh, lwaugh@txstate.edu