Diving into Data: Implementing a Data Repository at the Texas Digital Library

Abstract
This poster documents the progress of two Texas Digital Library (TDL) working groups focused on the storage and accessibility of research data. The first group, the TDL Data Management Working Group, selected a platform to act as the statewide repository. To do this, the group created a set of use cases to formulate evaluation criteria for testing potential systems. After testing Dataverse, the team made a recommendation to move forward with its implementation. A second group, the TDL Dataverse Implementation Working Group, is currently charged with launching an instance of the Dataverse open source platform as the statewide data repository for Texas.

As more institutions consider developing a repository for research data, our poster presentation offers a case study on the assessment of data repositories, including potential methods and criteria for evaluating systems, as well as the challenges and benefits to building a collaborative, consortial data repository.

DESIGN CHARACTERISTICS

<table>
<thead>
<tr>
<th>Function Type</th>
<th>Selected Evaluation Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingest</td>
<td>System Notifications, Licensing, DOI, File Size</td>
</tr>
<tr>
<td>Processing</td>
<td>GIS data, Static Databases, Version Control</td>
</tr>
<tr>
<td>Curating</td>
<td>Access Levels for Teams</td>
</tr>
<tr>
<td>Access</td>
<td>Data and Metadata Reuse, OAI Harvest</td>
</tr>
</tbody>
</table>

WORKFLOW

Project Charter
- Needs
- Project Goals

Use Cases
- Federal Mandates
- Data & ETDs
- Data & Papers
- Collaborative Teams
- Discovery

Technology Assessment
- Assessment Matrix
- Testing

Implement
- Business Model
- Technology Configuration
- Governance
- Workflows & Outreach

Pilot
- Researcher Ingestion
- Surveys
- Focus Groups

TIMELINE

Investigate → Report → Pilot Project → Assess → Launch → Outreach

USE CASES

1. Researcher needs to make their research data publicly available
2. Researcher needs a virtual environment to share active data, which may or may not be publicly accessible, within a prescribed network
3. Researcher seeks data to (re)use