The Texas GeoData Portal

A New System for Enhancing Access to Geospatial Data

Michael Shensky
GIS & Geospatial Data Coordinator
m.shensky@austin.utexas.edu
The University of Texas Libraries has for many years maintained thousands of unique geospatial datasets, but lacked a way to share them effectively with the research community for use in geospatial research.
Project Context: Peer Institutions

A study of the strategies employed by peer institutions led to the conclusion that a geospatial data portal was the ideal solution for sharing GIS compatible data with the UT campus community.

Stanford
https://earthworks.stanford.edu/

Princeton
https://maps.princeton.edu/

NYU
https://geo.nyu.edu/
Project Context: Building Momentum

Building demand on campus for geographic information system (GIS) services led to the creation of 2 GIS oriented positions within the UT Libraries in 2017.

The year 2017 also saw the successful launch of the Texas Data Repository (TDR) which served as an example for data portal development.
In late 2017 and early 2018, a GIS needs assessment and GIS portal pilot project were initiated to determine how to best build out a suite of GIS services and facilitate sharing of the UT Libraries’ geospatial data.
Development Considerations & Challenges

No existing metadata in standard GIS formats and portal search functionality completely dependent on metadata.

Over 60,000 diverse datasets to process so data management had to be: scalable, sustainable, and adaptable.

Digital collection managers and IT staff were new to working with GIS software.

And more!
Project Starting Point

Test Collections
- Relaciones Geograficas
- Buildings of Texas
- PCL Map Collection

Stakeholders
- Test collection content owners
- Librarian liaisons to academic depts.
- Digitization personnel
- Cartographic cataloger
- Research Support & Digital Initiatives
- Research data services

Libraries IT Developers
- Project management
- Systems administration
- Software development
- UI design
Development Coordination

Agile Development Principles

- Minimum Viable Product
- 2 week sprints
The project was carried out in two phases. Phase one focused on development of the Texas GeoData portal backend system and was completed in fall 2018.

The front end of the portal was successfully completed during the summer of 2019, and the portal officially launched on 11/13/2019.

https://geodata.lib.utexas.edu/
**System Architecture Overview**

- **Linux**: Open source OS used on system virtual machines (VMs)
- **PostgreSQL**: Open source relational database management system
- **Python**: Open source interpreted scripting language
- **ArcGIS Server**: Proprietary GIS server software
- **GeoBlacklight**: Open source Ruby on Rails project
- **Solr**: Open source search platform built on Apache Lucene
- **Sierra / Ex Libris Alma**: Library services platforms
- **Docker**: Open source OS-level virtualization software
GeoData Portal Content

UT Libraries Content
- Benson Latin America Collection
- PCL Map Collection
- Alexander Architectural Archives

Content from Other Universities
Over 30,000 open datasets from 23 additional institutions that participate in OpenGeoMetadata
GeoData Portal Functionality

- Federated Searching
- Browsing
- Faceting
- Featured Collections
- Data Downloads
- Metadata Downloads
- Detailed Help Page
Belize E, Central America (AMS E501), 1:250000

Attribute: Value

- Description: Scale: 1:250000; Projection: transverse mercator; Map Type: Topographic maps; Other Descriptive Notes about Map Series: E501: Transverse Mercator projection. Compiled from various sources by Military Intelligence Division, General Staff, U.S. Army. Cities are symbolized for size and importance. Index map on separate sheet 24 x 26 cm.
- Collection: Perry-Castañeda Library Map Collection and Army Map Service collection
- Place(s): Petén, Acanceh, Hondo, Belmopán, Corozal, Cayo, Campeche, Maya Mountains, Quintana Roo, El Cayo, Guatemala, Stann Creek, San Pedro, Central America, Orange Walk, Belize City, Belize, and Yucatán Peninsula
- Subject(s): Scanned maps

Buildings of Texas (Points)

Attribute: Value

- Author(s): Josh Conrad, Grace Hansen, Kate Pierce Meyer, Michael Shensky, Jessica Trelogan
- Description: Dataset produced from the spreadsheet donated by Gerald R. Moorhead and created by researchers collecting data for the two-volume publication Buildings of Texas. Architects and other contributor names have been normalized and building types have been added to portions of the dataset.
- Publisher: University of Texas Libraries
- Place(s): Texas
- Subject(s): architecture, architectural history, buildings, built works, and history
Utility for Campus Researchers

- Analyzing and mapping vector data
- Using georeferenced map scans as historical basemaps
- Creating new vector data from georeferenced map scans

Data is compatible with all major GIS software including:
- ArcGIS Pro
- ArcMap
- QGIS
- and more…
Technical Overview
Guide for Metadata Authors

Manual Metadata Entry

Scripted Metadata Export
Metadata Processing for Raster Datasets
Metadata Schema Mapping

**MARC Schema**

```json
{
  "id": "3886668",
  "varFields": [  
    {  
      "fieldTag": "a",
      "marcTag": "110",
      "ind1": "1",
      "ind2": " ",
      "subfields": [  
        {  
          "tag": "a",
          "content": "United States"
        },
        {  
          "tag": "d",
          "content": "Army Map Service"
        }
      ]
    }
  ]
}
```

**ISO 19139 Schema**

```xml
<characterstring gco:characterString>Topographic maps</characterstring>
```

**GeoBlacklight Schema**

```json
{
  "dc_title": "US Army Map Service (AMS) 2020",
  "dc_identifier_s": "100418_5204_5656_L"
}
```

**PHASE 1**

```
import json

mrc_data = json.dumps(marc_data)
```

**PHASE 2**

```
geojson_data = json.dumps(geojson_data)
```
Metadata Based Publishing
OpenGeoMetadata Harvesting
Future Improvements

676 datasets shared...
60,000+ datasets to go
Lessons Learned

☑ Data discoverability highly dependent on metadata quality

☑ Scripted processes can greatly enhance efficiency

☑ Scalability, sustainability, and adaptability are critical

☑ Effective team communication strategies are essential

☑ Important factors to consider when choosing software: open source, use by peer institutions, consider existing expertise

Acknowledgements

GIS Project Stakeholders past and present
Jessica Trelogan, Jenifer Flaxbart, Albert Palacios,
Katie Pierce Meyer, Katherine Strickland, Benn Chang,
Elle Covington, Dennis Trombatore, Anna Lamphear,
Loretta Wallace, Beth Dodd

UT Libraries IT
Dustin Slater, Larry Yang, Heather Langley,
Noah King, Dave Ronn, Brandon Stennett,
Crystal Arnspiger, Perry Thompson
Questions or Comments?
Contact Michael Shensky, UT Libraries GIS & Geospatial Data Coordinator at m.shensky@austin.utexas.edu

Interested in Learning More?
Explore the Texas GeoData Portal at https://geodata.lib.utexas.edu