

A Multi-institutional Geoportal to Enhance Geospatial Data Discoverability and Increase Accessibility



Kelley J. O'Neal¹, Nicole Kong², Bria Parker¹, and Ryan Matkke³
¹University of Maryland; ²Purdue University; ³University of Minnesota

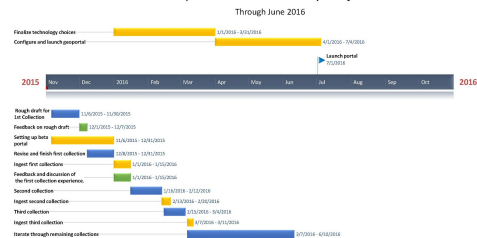
Corresponding Author:
 Kelley O'Neal kelleyo@umd.edu



Introduction

Geospatial data and methodologies are becoming increasingly popular in fields beyond Geography. The exponential increase in geospatial data sets available to users and varied platforms on which they are served can make data searches difficult and leave some data undiscovered, particularly for users with more limited knowledge of geospatial data sources. The Committee on Institutional Cooperation (CIC), which provides unique opportunities for and supports collaborative efforts among Big Ten member universities, initiated the CIC Geospatial Data Discovery Project to enhance geospatial data discoverability and access among member institutions.

CIC Geospatial Data Discovery Project



Project Summit November 2015



In November of 2015, 18 members of the project task force from nine CIC institutions convened for a two-day summit to discuss discovery software, institutional collection surveys, preferred metadata formats for GIS data and scanned maps, as well as metadata conversion tools and workflows.

Participating Institutions

Nine of the fifteen Big Ten institutions have come together to develop and populate a geoportal for use by all member institutions.



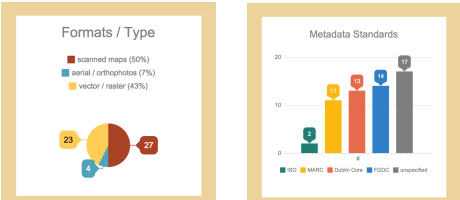
Collaborative Metadata Workflow

- 1. Submit Records**
Members from the 9 institutions on the task force obtain and submit metadata records representing local and regional geospatial resources.
- 2. Metadata Transition**
The Project Metadata Coordinator cleans, reorganizes, and adds template information to the records.
- 3. Edit Records**
The records are added to an online geospatial cataloging application, GeoNetwork. Task Force Members and the Project Metadata Coordinator can work together to edit the records manually or with a spreadsheet update process.
- 4. Publish Records**
The completed records are published in a multi-protocol search-friendly geospatial, GeoBlacklight.

Metadata

One challenge of this project was to arrive at a decision regarding metadata schema. While FGDC is the prevalent standard in the United States, it is being phased out in favor of ISO, the emerging standard. As much of the existing metadata accompanying submitted datasets is in FGDC, metadata may undergo multiple transformations in order to become ISO compliant. Metadata for scanned maps will generally not have FGDC or ISO metadata unless they are fully georeferenced, and will be submitted in or transformed to DublinCore.

Institutions submit metadata records to GitHub for evaluation and transformation, while also creating a tracking ticket with basic information which the Metadata Coordinator uses for tracking the transformation process.



Metadata Enhancement

The project is using GeoNetwork (an open source, ISO-centric geospatial metadata application) as a collaborative metadata editing tool. Once the submitted metadata records have been transformed into valid ISO, they can be batch-loaded into GeoNetwork and edited one record at a time, or as a batch process. This tool can be customized to accommodate various keyword thesauri to aid in metadata normalization. GeoNetwork also includes an implementation of OAI-PMH, which can simplify how metadata for scanned maps is imported and managed.

Discovery Platform

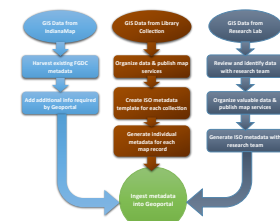
GeoBlacklight was identified as the best discovery portal and platform for the geoportal due to its open-source platform, strong development community and documentation, intuitive interface, and easy metadata integration. It does use its own GeoBlacklight metadata schema, but conversion tools already exist to convert from our validated ISO to GeoBlacklight.

University of Maryland's Data



University of Maryland chose to submit metadata from MD iMap, the state of Maryland's Mapping and GIS Data Portal. There were two ways of obtaining this metadata: downloading the shapefile (and generating metadata using tools like ArcGIS if a .shp.xml file is not included) or scraping metadata from a REST endpoint as a json file. The scraped data from the REST endpoint does not have geometric objects, attribute tables, or reference systems. We are currently working out how to combine these two types of files for the most complete metadata for discovery purposes.

Purdue University's Workflow



At Purdue, the institutional geodata portal has been built based OpenGeoportal project before joining the CIC collaboration project. Since both OpenGeoportal and GeoBlacklight share similar metadata requirements, Purdue University Libraries just need to share the existing metadata with the CIC project. The geospatial data included in Purdue geodata portal include three types: GIS data from Indiana State data clearing house (IndianaMap), GIS data collection at Purdue University Libraries, and research generated data at Purdue. For each of these datasets, we followed the following workflow to create/harvest metadata and ingest the information into geodata portal.

Conclusions and Future Work

The collaborative nature of this project serves to reduce workload among individual institutions, simply workflow, encourage communication between institutions, and increase geospatial data collections for each institution. The portal is slated to be launched in the summer of 2016.

Future work involves developing workflows and best practices for enhancing and ingesting metadata for scanned map collections. This includes how to ingest and represent coordinate information in Dublin Core, or in an extension to that schema. The task force also hopes that the work done so far will encourage other CIC institutions to participate.

Acknowledgements

Special thanks to project collaborator Karen Majewicz (U Minnesota) for assistance with figures represented in this poster. Also thanks to our project collaborators James Whitacre (U Illinois), Cathy Hodge (U Iowa), Rob Shepard (U Iowa), Mara Blake (U Michigan), Tim Utter (U Michigan), Kathleen Weesses (Michigan St), Amanda Tickner (Michigan St), Kevin Dyke (U Minnesota), Linda Ballinger (Penn St), Nathan Piekielek (Penn St), Paige Andrew (Penn St), Shirley Li (Purdue U), Jaime Martindale (U Wisconsin), and AJ Wortley (U Wisconsin).