

A Spatial Data Infrastructure (SDI) for the BTAA

Description and Current BTAA Capabilities

Abstract: This document provides a description of the technology behind a Spatial Data Infrastructure (SDI). It also gives a background for the BTAA Geospatial Data Project (GDP), and some of the current challenges this project faces that could be remedied by a shared, or common, SDI.

Existing collaborations across the BTAA for geospatial data

In 2012, three BTAA map librarians issued a white paper, *A Collaborative Vision for Spatial Scholarship across the CIC*¹, that described geospatial data as a long term, interdisciplinary need and proposed several collaborative possibilities for managing it. This paper was a major impetus for the creation of the BTAA Geospatial Data Project (GDP) three years later. The BTAA GDP has accomplished several of the recommendations laid out in the paper, including creating a collective specialized tool for discovering geospatial resources, combining scanned map collections, and encouraging robust communication across the BTAA geospatial community.

BTAA GDP project quick facts²

- Number of participating institutions: 12
- Number of project Task Force members: 22
- Year the project began: 2015
- Month the BTAA Geoportal went online: August 2016
- Number of resources in the geoportal at launch: 2,300
- Number of resources as of January 2019: 18,195
 - Number of scanned maps: 6,974
 - Number of geospatial datasets: 10,793
 - Number of geospatial web services: 5,298

The next phase of the BTAA GDP proposes to address another of the paper's main goals, that of building a "common infrastructure for storing, disseminating, and archiving geospatial data." (p. 2). A common infrastructure would address two major challenges that BTAA libraries continue to face: managing geospatial data created or licensed at individual libraries and dealing with the ephemeral nature of publicly available open data.

¹ Bidney, M., Mattke, R. & Weessies, K. (2012). *A Collaborative Vision for Spatial Scholarship across the CIC*, 2012, white paper. [Link](#)

² Internal document: [BTAA Continuation Proposal 2019-2021](#)

Ongoing challenges across the BTAA for geospatial data

The challenge of managing licensed geospatial data in libraries

Many of the BTAA map librarians are still storing licensed geospatial data purchased from vendors on physical media or local hard drives, or using a cloud service such as Google Drive or Box. As described in the 2012 white paper³, this leaves the libraries in “reactive mode,” where they must develop piecemeal solutions for delivering geospatial data to patrons. As evidenced by the Summary tab of the [BTAA Spatial Data Infrastructure Details](#)⁴, there is inconsistency across the BTAA in the institutional infrastructure for this type of geospatial data. Combining our resources and expertise into a shared SDI would fill in the gaps across the chart.

The challenge of ephemeral open data

Academic libraries with physical map collections have well-established programs to collect and archive these important information resources for the benefit of their local researcher populations. Government agencies now issue most geospatial information as digital data instead of as physical maps. However, many academic libraries have not yet expanded their collection scopes to include publicly available digital data, and are therefore no longer systematically capturing the changing geospatial landscape for future researchers.

The value of this data is high, as researchers routinely use it to form the base layers for web maps or geographic analysis. However, the lack of standard policies at this level of government means that this data can be considered **ephemeral**, as providers regularly migrate, update, delete, and re-publish data without saving previous versions and without notification to the public.⁵

The BTAA Geoportal fills a gap in the geospatial data ecosystem by cataloging metadata records for current publicly available state, county, and municipal geospatial data. Because the BTAA Geoportal indexes geospatial data on a monthly basis, we have discovered that the rate at which this ephemeral data disappears is rapid.⁶ This continual turnover creates a difficult environment for researchers to properly source data and replicate results, and it prevents researchers from conducting longitudinal studies that examine change over time for a region. For the geoportal, it requires a great deal of dedicated labor to maintain the correct access links.⁷ As the geoportal’s collection grows, the labor required to maintain it grows as well.

The barriers that keep libraries from collecting copies of open data, rather than tracking its metadata as the BTAA Geoportal does, include lacking local technology and staff expertise to effectively curate and archive the resources. A shared SDI that utilizes a combination of existing skill sets across the BTAA and new shared technology development would allow libraries to expand their scope to include these types of important research resources.

³ Bidney, M., Mattke, R. & Weessies, K. (2012). A Collaborative Vision for Spatial Scholarship across the CIC, 2012, white paper.

⁴ Internal Document: [Geospatial Data and Repositories across the BTAA](#)

⁵ Future Proofing Civic Data: Exploring the Challenges of Preserving Open Civic Data for the Long Term (2017). Temple University Libraries. <https://drive.google.com/file/d/0B3MMB2pFOd4MXIkTUhLOGNvdXIIeGVtOURmZGO4S2gxY1I1N/view>

⁶ Internal document: [BTAA GDP ArcGIS Data Portals Rate of Change 2017-08 to 2018-12](#)

⁷ Internal document: [Comparing Staff Time for Maintaining Access to Ephemeral Resources vs Archived Data](#)

Introducing a Spatial Data Infrastructure (SDI)

A Spatial Data Infrastructure (SDI) is defined in a GIS dictionary as “A framework of technologies, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve the use of geospatial data across multiple public and private organizations.”⁸ The benefits of a shared SDI include reduced costs through shared technology, agreed upon standards, and improved data discovery and accessibility. The technology behind a spatial data infrastructure can be broken down into four parts:⁹

1. Digital Object Preservation and Access

About: The foundation of an SDI is the repository that stores the digital objects. These objects may be reproduced in multiple formats to ensure long term access, and the supporting technology may include preservation services, such as fixity checks and versioning. *non-BTAA example:* [Hydra/Samvera at Stanford University](#)

Across the BTAA: All of the surveyed BTAA institutions host a general data repository that is restricted to content created by university researchers. These repositories are typically used for articles, theses, and supporting data. Although geospatial file formats are accepted to these repositories, they represent a small percentage of the collections.

Most BTAA libraries offer purchased, licensed geospatial data to researchers. Often, this data must be either viewed on site or obtained by contacting a librarian directly. In some cases, university affiliates may be able to access the data from file servers.

For publicly available state and local data, Indiana University has a dedicated geospatial data repository, and the University of Wisconsin is systematically saving files on hard drives.

2. Geospatial Metadata

About: Geospatial data in an SDI should be fully documented with valid standards-compliant metadata. The metadata can be stored with the actual data files, but is sometimes also stored in a separate catalog. Like other digital archives, authoring metadata is typically the most labor intensive aspect of creating and maintaining an SDI. *BTAA example:* [GeoNetwork Metadata Catalog using ISO 19139 metadata standard for the BTAA GDP](#)

Across the BTAA: Three BTAA institutions have geospatial metadata experts on staff, who regularly catalog geospatial data (Purdue, Minnesota, Wisconsin) in a geospatial standard (ISO or FGDC). The remaining libraries employ staff who regularly catalog maps using MARC and transfer this knowledge to cataloging datasets as needed.

⁸ Wade, T., & Sommer, S. (2006). A to Z GIS, An illustrated dictionary of geographic information systems. Esri Press.

⁹ For a more technical explanation, see Durante, K. & Hardy, D. (2015). Discovery, Management, and Preservation of Geospatial Data Using Hydra, Journal of Map & Geography Libraries, 11:2, 123-154, DOI: [10.1080/15420353.2015.1041630](https://doi.org/10.1080/15420353.2015.1041630)

3. Geospatial Web Services

About: Geospatial web services provide an online connection to a dataset hosted on a server. These services allow users to stream and visualize geospatial data in a web browser or mapping application. Users can also load these services through a Desktop GIS without actually downloading the dataset. Geospatial web services are widely used by researchers to build online maps quickly without needing to rehost data.¹⁰

In order to provide these services, an organization needs to set up a server with a geospatial web service application and load access copies of geospatial datasets into it. *non-BTAA example:* [Princeton University's GeoServer](#) (click on Layer Preview, choose OpenLayers option)

Across the BTAA: There are a few instances of locally hosted geospatial web services at BTAA institutions, and they are used for special library projects or initiatives. At most BTAA institutions, campus affiliates can log in to an organizational account for ArcGIS Online to publish their data as ArcGIS REST Services. However, this proprietary platform is generally a temporary host, and users are not always able to port the data to custom applications. Geospatial web services are not currently integrated with any of the BTAA institutional digital repositories or discovery applications.

4. Discovery Applications

About: Discovery applications typically take the form of a web portal that allows users to search and filter through a normalized subsection of the full standards metadata. In the case of a geoportal, a graphical interface for searching on a map is also provided. If geospatial web services have been developed, a user can preview and query the dataset to analyze it without the need for specialized software. The BTAA Geoportal is an example of a standalone Discovery Application that relies upon repositories and geospatial web services from external data providers. *non-BTAA example:* [GeoData@Tufts built with OpenGeoPortal](#).

Across the BTAA: Indiana University makes their geospatial data available via a custom interface that allows users to select data from nested lists. Two universities (Wisconsin and Purdue) have implemented geoportal applications that also allow users to search by text and via a map interface.

Other universities (Minnesota and Penn State) contribute university created geospatial data to clearinghouses maintained by state agencies. Penn State provides IT assistance and web servers for the statewide clearinghouse.

See this spreadsheet for more information: [BTAA Spatial Data Infrastructure Details](#)

¹⁰ For those more familiar with image repository applications, a geospatial web service can be thought of as analogous to the [International Image Interoperability Framework](#) (IIIF) that enables image sharing across platforms.