Analytics

What do we mean by Analytics?

By examining a combination of Google Analytics data and data extracted from the Solr database (BTAA Geoportal backend), we can analyze trends in geoportal usage and attempt to construct a portrait of our users.

What Analytics are we focused on?

- We are concentrating on the analytics needed as outlined in the Analytics Working Group report
- The report identified these questions as high priority to be answered using analytics:
    - Where are the users geographically located?
    - What collection materials are of most interest to users?
    - What is in the collection?
    - What is the geographic coverage of the collection?
    - What collection materials are users downloading?
    - By user location, what geographies are of interest?

- Working with staff at the University of Minnesota Libraries, we developed a Tableau Dashboard of analytics data.

What did we find?

- Where are the users geographically located?

    - Top 20 Cities: Minneapolis, Ann Arbor, Madison, Chicago, State College, East Lansing, Bloomington, West Lafayette, Columbus, Iowa City, Saint Paul, College Park, New York, Austin, Corvallis, Boston, Washington, Champaign, Stanford, Amsterdam

- What collection materials are of most interest to users?

    - Users who selected a Genre from homepage: Geospatial Data (46%), Maps (34%), Aerial Imagery (20%)

    - Top text searches: water, minnesota, asia, parks, chicago, census, health, wisconsin, ann arbor

    - Top place facet values: Minnesota, Wisconsin, Michigan, Pennsylvania, Indiana, Maryland

    - Top subject facet values: Environment, Transportation, Society, Imagery and Base Maps, Boundaries, Fire Insurance Maps, Maps, Inland Waters, Planning and Cadastral, Geoscientific Information
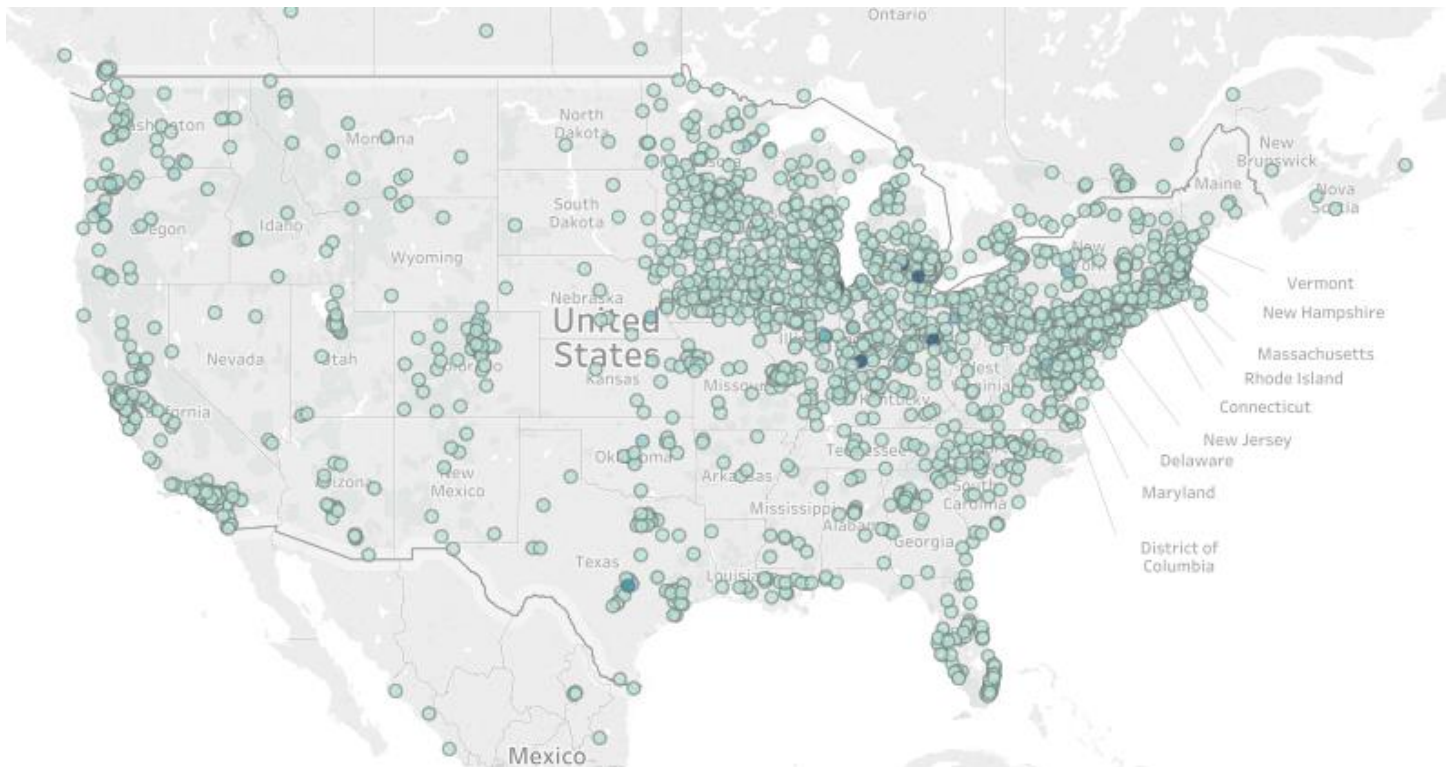
- What is in the collection?

| Genre	Records	% of total
Geospatial data	13,653	62%
Maps	7,561	35%
Aerial Imagery	668	3%

Total number of records contributed
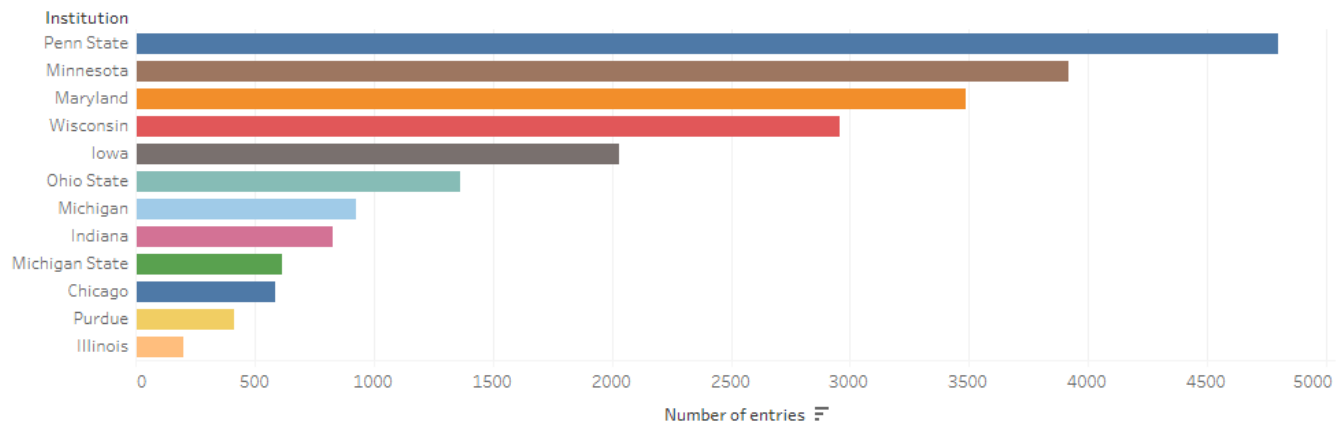
- What is the geographic coverage of the collection?

  - (See examples below)

- What collection materials are users downloading?

  - Top Five Data Sets:

    - 100 Year all Flood: Lycoming County, Pennsylvania

    - Topography 2 Foot Contours: Washington, D.C.

    - Census Tracts: Hennepin County, Minnesota, 2010

    - Parcels: Washtenaw County, Michigan, 2015

    - County Boundaries: Minnesota

  - Highest number of unique downloads for a single dataset: 14

  - Number of different datasets downloaded: 1,933

- By user location, what geographies are of interest?

  - [We are still working on tweaking this dashboard]

What conclusions can we draw?

- Where are the users geographically located?

  - Heaviest use is from the cities where the participating institutions are located, but there is use from across the country
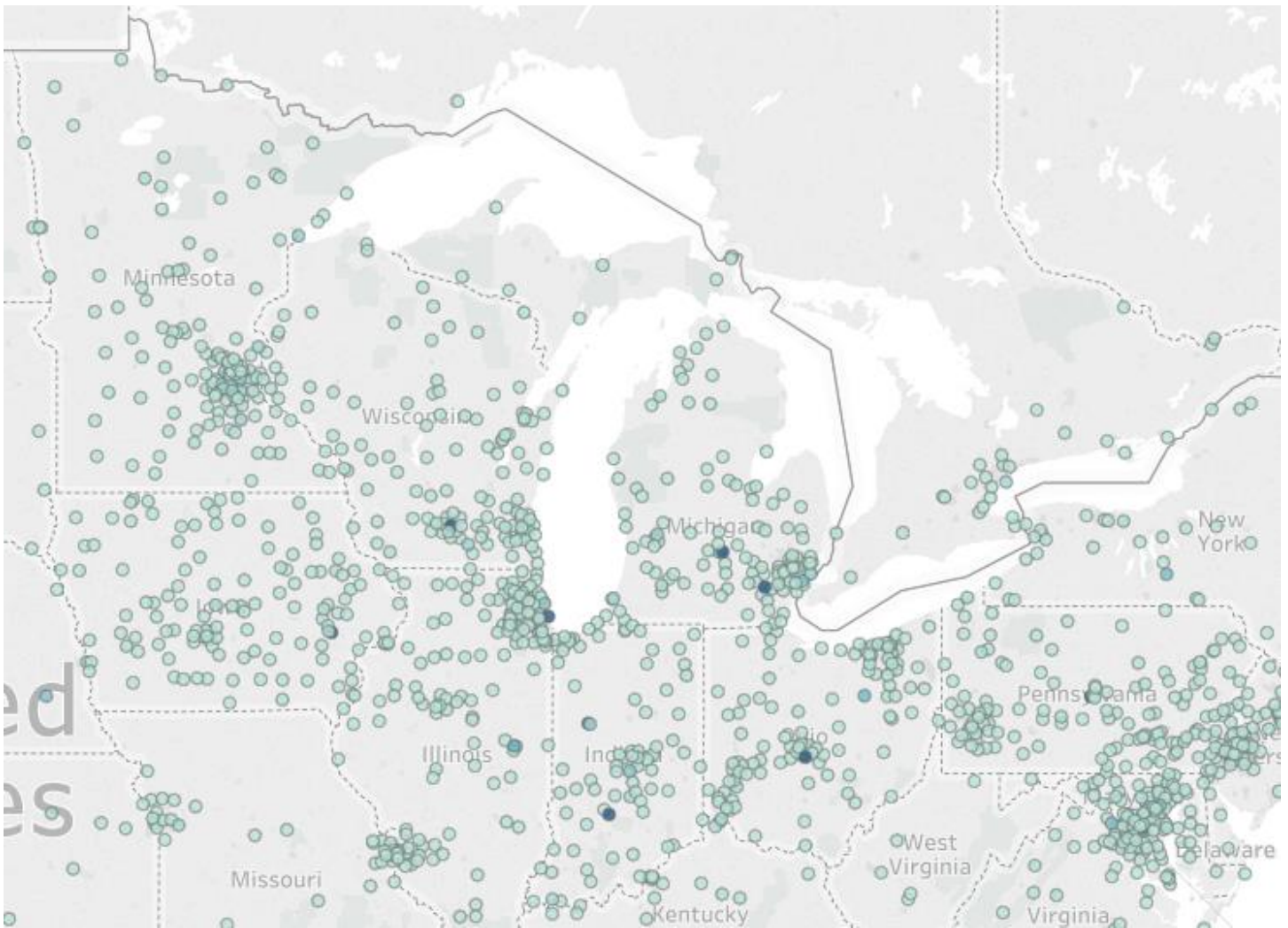
- What collection materials are of most interest to users? What's in the collection?
  - Some interesting data here comparing what genre type users click on vs. what is in the collection

- What is the geographic coverage of the collection?

   Map showing coverage of "smaller countries" scale datasets and maps
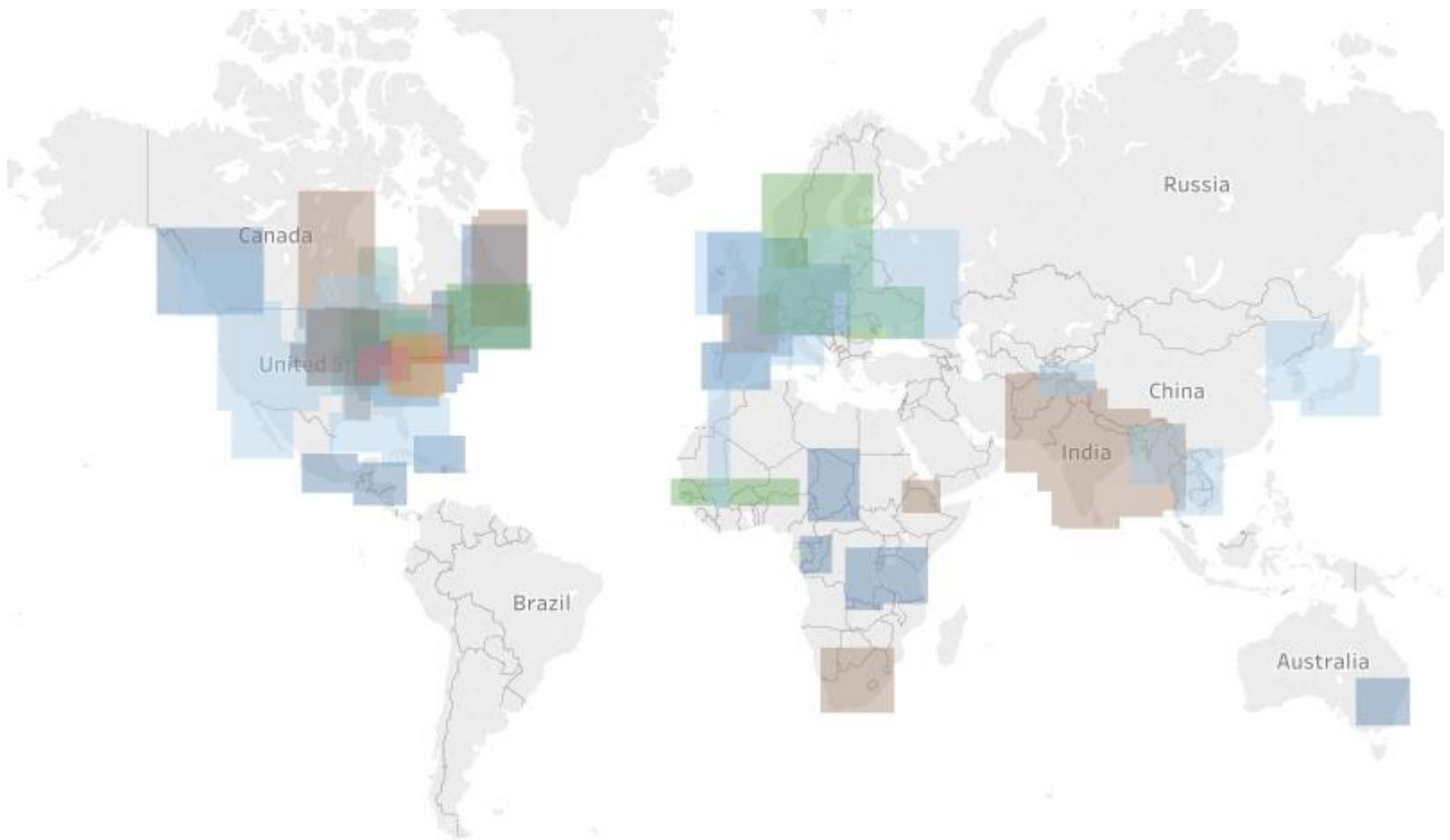
Can highlight individual institutions – this shows Minnesota's concentration in the midwest, but also in India

Institution: **Minnesota**
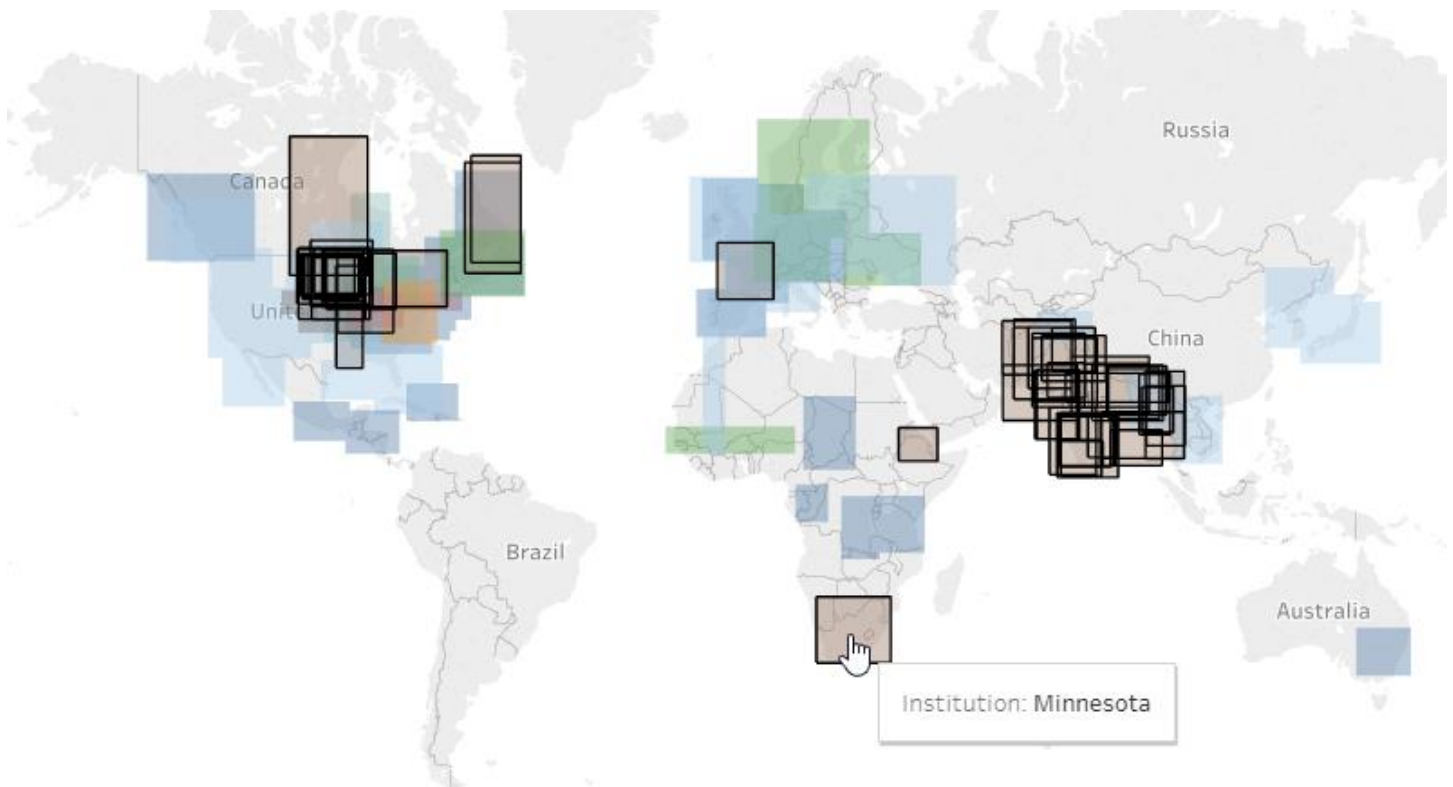
Can get more specific – this shows coverage of geospatial datasets at the county level – can spot states that have great coverage and some that don't, due to lack of data availability or open data policies
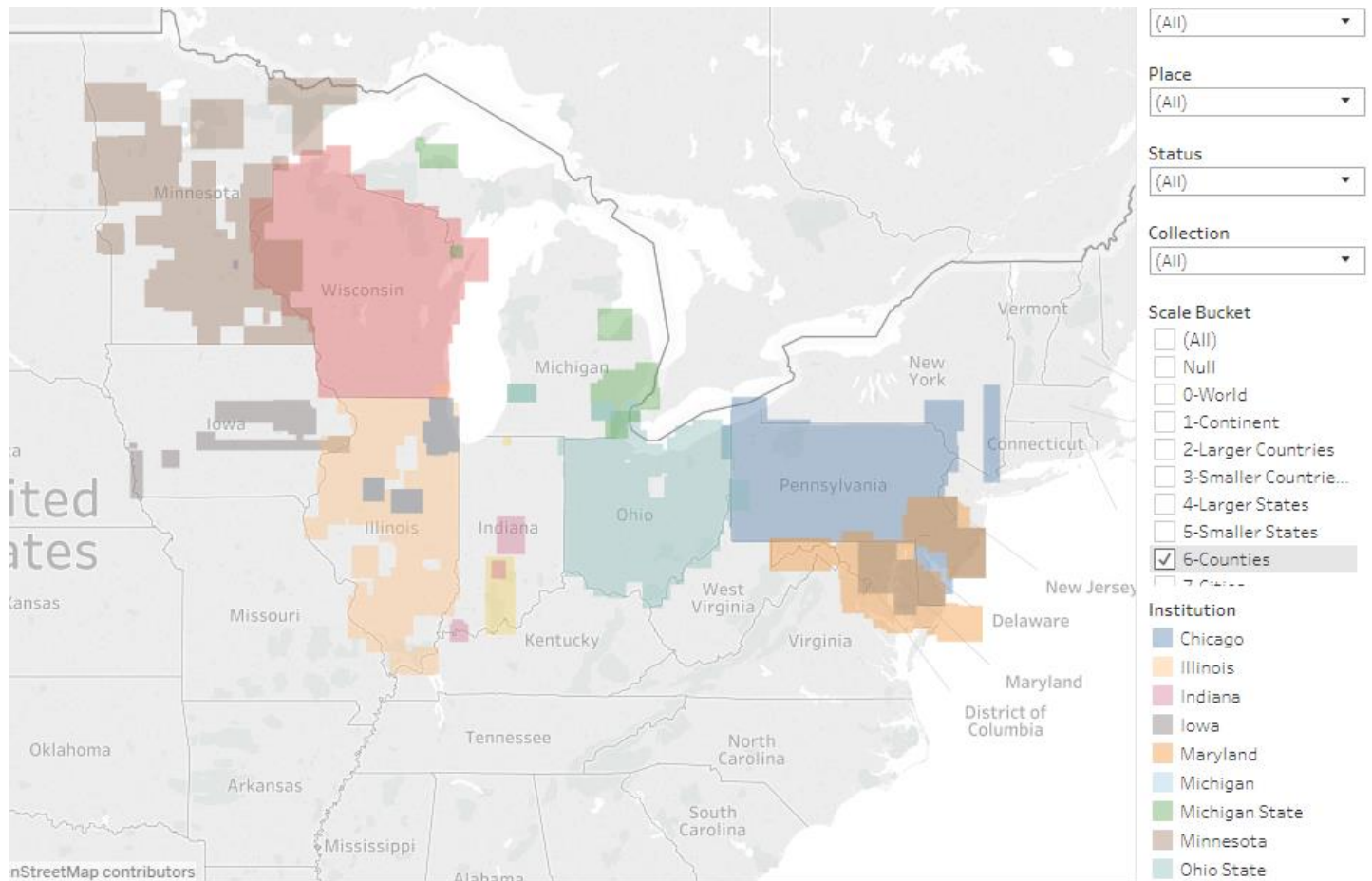
Future Proposal & Possible Expansion

What are the specific considerations for a future proposal?

The current, status-quo proposal is in effect until June 2021. However, we are considering possible expansions in scope and membership and would put forth a new proposal to take effect in July 2020.

Specific considerations:

❖ Cost – current cost is $15,000 per institution per year
❖ Governance – we would like to formalize a governance structure
❖ Expansion – possible expansion of services and/or membership

What are we proposing for expansion of services?

★ Licensed Data
★ Geospatial Web Services
★ IIIF Services
★ Public Geospatial Data Archiving

What do we mean by expansion of membership?

There are non-BTAA institutions that have expressed interest in joining the project. While there are obvious challenges associated with expanding the membership beyond the BTAA, there are also benefits, including a larger funding base and a wider array of geospatial data available through a single portal.

These institutions have all expressed interest in geospatial web services and would like to see licensed data in the geoportal; most have expressed interest in the discovery interface; some have expressed interest in the IIIF services.

What are some things to consider?

❖ Cost

➢ What level of increase (if any) would be tolerable to BTAA Library Directors?

- ➢ Consider grant funding for startup costs for expanded services?

  - ■ Licensed Data

  - ■ Geospatial Web Services

  - ■ IIIF Services

  - ■ Public Geospatial Data Archiving

- ❖ Governance

  - ➢ Current governance includes local sponsors (at UMN), Strategic Leadership Group (directions/planning), and Steering Committee (comprised of Project Lead, Product Owner, Committee chairs) – does this model still make sense?

  - ➢ Strategic Leadership Group to be sunsetted after new proposal – should we create an advisory board (or similar) to take its place?

  - ➢ Get more of a firm commitment from participating institutions for how much time each Task Force member is expected to spend on the project?

  - ➢ If we expand the membership beyond the BTAA, how would we give voice to new member institutions?

- ❖ Expansion of services

  - ➢ Which of the possible expansions in services are we recommending?

    - ■ Licensed Data

    - ■ Geospatial Web Services

    - ■ IIIF Services

    - ■ Public Geospatial Data Archiving

- ❖ Expansion of membership

  - ➢ What is the BTAA community reaction to that?

  - ➢ Should we float this by the BTAA Library Directors at the November meeting to get their reaction?

  - ➢ What would this mean for the project?

  - ➢ Do we foresee specific issues?

- ■ Branding, e.g.

Geospatial Web Services

What are we proposing?

Collective **geospatial web services** for the BTAA Geoportal.

What is a geospatial web service?

A geospatial web service delivers GIS data to users over the web. It allows users to preview, visualize, and repurpose data in online maps and mapping applications. The BTAA Geoportal currently has a viewer through which users can display and interact with geospatial web services. There is also a button to "Open in ArcGIS" that allows users to quickly start making maps with data they find in the Geoportal.

Example: ArcGIS Feature Service Layer

Why do we want this?

◎ The ability to **preview resources** before downloading them is one of the best features of GeoBlacklight.

- ○ Web browsers cannot interpret and display GIS data in their common native file formats (e.g. shapefiles)

- ○ Transforming the data into a web service allows users to preview and query a dataset online, without needing to download and load the data into a desktop application

◎ More **consistent display and user experience** for geospatial data.

- ○ A large number of the geospatial datasets available through the geoportal already have associated web services.

- ○ Hosting web services for data that do not currently have them would support a more consistent user experience for records from across institutions and data sources.

◎ **Incentive for researchers** to use the geoportal

- ○ Geospatial web services can be easily added to a web map in ArcGIS Online, even by novice users

- ○ Researchers who deposit their data in an institutional repository would gain access to web services – Stanford reports that this is a major draw for convincing researchers to deposit

◎ Creating web services will **improve the reliability of the geoportal**, because we would be able to control the resources

   ○ Data from government data portals and their associated web services are volatile (see tables below). In order to keep our records functioning well, we currently have to reaccession data portals every month (about 22 hrs/week of work).

   ○ If we host our own copy of these web services, we would not need to reaccession as frequently and we would be able to provide more persistent access to public data.

Geospatial Data Records in the BTAA Geoportal by Institution (Oct 2019)

Institution	With Web Services	Without Web Services	Percent w/o web services
Wisconsin	41	2,902	99%
Penn State	436	1,515	78%
Maryland	2,111	824	28%
Minnesota	347	710	67%
Chicago	168	285	63%
Illinois	69	120	63%
Michigan	11	106	91%
Iowa	854	99	10%
Ohio State	1,107	27	2%
Purdue	273	18	6%
Indiana	159	0	0%
Michigan State	450	0	0%
Total	**6,026**	**6,606**	**52%**

Deleted records per month over a 4 month period for *selected* ArcGIS Hub collections in the BTAA Geoportal*

ArcGIS HUB	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Sum
Lake County Illinois Open Data Portal	362	156	361	398	1277
Indianapolis Open Data	0	2	2	242	246
IowaDOT Open Data	203	16	19	3	241
District of Columbia Open Data	137	33	42	28	240
GeOhio	174	20	21	23	238
Harford County GIS Data	0	33	33	7	73
State of Michigan Open Data Portal	5	8	10	14	37
Minneapolis Open Data	0	16	0	9	25
Maryland iMap	5	3	2	13	23
Delaware Valley	0	1	4	4	9

	Sum	886	288	494	741	**2409**

*These numbers typically represent anywhere from 5-30% of a Hub's total records.

How would we implement it?

Model 1: Open source locally hosted server with GeoServer
- ◉ Technology: Geoserver application; PostgreSQL to store data
- ◉ Staff activities
  - ◦ Curation
    - Develop a collection plan
    - Collect digital files and metadata
    - Deposit resources into database
    - Publish data with GeoServer
  - ◦ Technology development
    - Set up servers
    - Install PostGreSQL and GeoServer
- ◉ BTAA staff time
  - ◦ Metadata Coordinator: 20 hours/week
  - ◦ Development time: TBD

Model 2: Commercially managed server with ArcGIS

- ◉ Technology: ArcGIS Online with Esri's cloud infrastructure
- ◉ Staff activities:
  - ◦ Curation
    - Develop a collection plan
    - Collect digital files and metadata
    - Create discovery metadata for ArcGIS
    - Upload data to ArcGIS Online
  - ◦ Technology:
    - Communicate with Esri representatives about performance and storage issues
- ◉ BTAA Staff time
  - ◦ Metadata Coordinator time: 20 hours/week
  - ◦ Development time: TBD

More information available in the full Web Services Report.

What are some things to consider?/ Discussion topics

❖ Task force members have varying levels of access and permission to ArcGIS organizational accounts at their institutions. Task force members who leave their institution may not be able to continue to host the web services uploaded to their local ArcGIS organization and transferring ownership of content between organizations is very challenging. Setting up a local GeoServer server is out of scope for many of our contributing institutions.

❖ Depending on how we set up our web services, users would not need to visit the original source. We do not currently communicate actively with government data portals, because we are enhancing discovery to their resources not making versions of their content. We may need better outreach if we are hosting web services.

❖ Workflows and challenges creating web services will differ depending on the source of the data. We could focus first on stable research data held in institutional repositories and then public data that does not have an associated web service. Once we have a process in place, we could consider making more persistent web services for data in public portals with frequent turnover.

❖ We currently only maintain discovery-level metadata for most of the geospatial data in the Geoportal. More comprehensive metadata work has not been feasible due to the frequent turnover of records on open data portals. If we keep copies of data and host web services, more intensive metadata processing may be appropriate.

IIIF (International Image Interoperability Framework)

What are we proposing?

A collective **IIIF image server** for the BTAA Geoportal.

What is IIIF?

The International Image Interoperability Framework (IIIF) defines a standardized method of describing and delivering images over the web. A IIIF image server offers an API endpoint that can be loaded into several different viewing applications. The BTAA Geoportal currently uses open source viewer plug-ins that can display IIIF images from anywhere. Examples

- o [Single map using a IIIF service with the Leaflet viewer](#)
- o [Multipage atlas using a IIIF service with the Mirador viewer](#)

Why do we want this?

◎ The ability to **preview resources** before downloading them is one of the best features of GeoBlacklight.

- o Image previews encompass panning or zooming in on a map or dataset.

- o This is accomplished by using web services, and IIIF is a type of web service for scanned images.

◎ This functionality currently relies upon each contributing institution **setting up individual local IIIF servers**. Currently, only 3 institutions offer IIIF services and the remaining do not.

Map Records in the BTAA Geoportal with and without IIIF (Oct 2019) *updated Oct 31 with PSU maps*

With IIIF	Map records
Minnesota	2428
Michigan	803
Maryland	36
Penn State	2416
Total	5683

Without IIIF	Map records
Iowa	746
Indiana	665
Michigan State	127
Chicago	70
Purdue	67
Total	1675

◎ Unfortunately, the Geoportal **user experience** of discovering items from various schools that exhibit different behavior on the detail record view is very confusing.

  ○ Items with IIIF have a viewer that pans and zooms

  ○ Items without IIIF, but with an accessible image link will display a static image

  ○ Items without an accessible image link display nothing beyond the metadata.

  ○ **Adding a IIIF server for B1G Geoportal assets would allow us to standardize the user experience for assets from all contributors.**

How would we implement it?

◎ Technology
  ○ IIIF Image Server: local UMN installation (already tested with the application called Cantaloupe)
  ○ File storage on an external server: likely Amazon Web Services (AWS) contracted through UMN IT

◎ Staff time
  ○ Development time: ~30 hours
  ○ DevOps time: ~20 hours
  ○ Metadata Coordinator time: ~40 hours

What are some things to consider?/ Discussion topics

❖ Contributing institutions to the Geoportal have different local digital collection technologies and services, and setting up a local IIIF server is out of scope for many of them.

❖ We would need to make copies of images and store them on a server. Paying for this server would require dedicated funds, but at a level that is in scope for our current technology budget.

❖ Users could preview and download images directly through the geoportal, and would not need to visit the original source. This might diminish traffic to each institution's digital library. However, this aligns with the emerging BTAA Collections ethos of combining BTAA resources.

❖ There would be additional time at the beginning of the workflow to harvest and publish the images. However, the ongoing metadata maintenance of fixing broken links would be minimal. (Several institutions have migrated all of their image URLs since the project began, and we can predict this will continue to occur.)

Licensed Data

What are we proposing?

To incorporate metadata records for **licensed data resources** into the BTAA Geoportal.

What do we mean by Licensed Data?

Geospatial data purchased or obtained by participating institutions that requires authentication or authorization for patron access.

How would we implement it?

- When a user lands on an item page for a licensed data resource, they would be presented with a drop-down listing the institutions that subscribe to that resource
  - Selecting their institution would take them to the library catalog record to login and authenticate
  - Item page would also include explanatory information as to why access is restricted to subscribing institutions
- Adopt a GeoBlacklight feature that presents public versus private availability, create a Solr document schema and metadata pattern to collect institutional holdings and proxy links
  - Development changes to the application logic and views will be moderate to minimal.

- Staff time
  - Development time: ~40 hours
  - Metadata Coordinator time: ~40 hours
  - Task Force member time: ~20 hours per institution

Why do we want this?

- Licensed Data was identified as a particular issue in our Fall 2018 Continuation Proposal:

  - "Many participating institutions hold licensed geospatial data that they wish to submit to the geoportal. Incorporating these types of resources would present challenges in terms of enabling multiple institutional authentication protocols that will require specialized technical research."

- Participating institutions expend collections funds on Licensed Data regularly and, other than the library catalog and libguides, do not have a good way to highlight and promote discovery of these (often expensive) resources.

BTAA Geoportal Member Institutions - Licensed Data Inventory (September 2019)

- [Inventory of Licensed Data resources](#)

    - 14 subscription databases and 15 data set resources were identified by Task Force members

    - Subscription databases

        - Five databases are subscribed to by at least nine of the 12 institutions:

            - Social Explorer, PolicyMap, and SimplyAnalytics are databases of geospatial data; data available to the subscribing institution is the same in all cases

            - East View LandScan is a database of geospatial data; data available to the subscribing institution varies based on the subscription

            - Digital Sanborn Maps is a database of scanned maps; maps available to the subscribing institution varies – some have access to maps from all states, some to just their own state

    - Recommendation would be to pilot the addition of Licensed Data using one of the three databases in the first bullet above; next steps would be to move to a database where the data available varies by institution

- **Adding a Licensed Data would allow for a more streamlined discovery experience for (often expensive) geospatial data.**

What are some things to consider?/ Discussion topics

- ❖ Some users will not have access to the licensed data – we will have explanatory text, but this may cause some confusion.

- ❖ Different databases / licensed data sources will require different approaches – how to prioritize?

BTAA GDP Outreach 2019- 2020

Current Efforts: Increasing awareness and use of geoportal

- Target Audience: [Geoportal users](#)

- Lead: Communications Committee

- Key Messages
  - The geoportal saves researcher time by centralizing regional geospatial data discovery into a single interface
  - The geoportal contains a wide variety of geospatial resources
  - The geoportal is unique in that it provides discovery to the most up-to-date resources.
  - The geoportal is unique in that it allows users to search by What, Where, and When, without needing to know Who or Why

- Outreach Methods
  - Creating and sharing regular [blog posts](#)
  - Contacting instructors and including in course presentations
  - Tracking and expanding listings in library websites

Future Plans: Strengthen connections with data providers

- Target Audience: Data providers at public agencies

- Lead: Metadata Committee and Collections Development Committee

- Collaborate with local, county, regional, and state government agencies who produce the geospatial data. Creating formal archiving agreements with government data providers would enhance reliability, but add complexity.

- Key Messages
  - The Geoportal adds value to geospatial data with enhanced discovery metadata
  - The Geoportal brings local geospatial data to a wider audience
  - Engaging in recommended best practices for metadata and data management will improve the discoverability and usability of resources

Assessment

- Increase usage as demonstrated by Google Analytics
- Increase the number of listings to refer users to the geoportal
- Generate user stories

What are some things to consider? - discussion topics

- ❖ What other forms of outreach should the project be pursuing?

- ❖ What types of user stories are the most compelling?

- ❖ What should the ideal number of users of the geoportal look like?

Public Geospatial Data Archive

What are we proposing?

A systematic process for archiving public geospatial data created across the BTAA geographic region.

What do we mean by **public geospatial data**?

Data that is:

◉ Created or published by state, regional, and local government entities and academic institutions
◉ Free and open
◉ In a recognized geospatial file format (e.g. shapefile) or aerial image file (e.g. jpg)

## What do we mean by **archiving**?

A system that includes any or all of the following:

◉ File storage
◉ Archiving strategies: identifiers, replication, verification
◉ Preservation: validation and migration
◉ Curation: evaluation and description

Why do we want this?

Community Benefits

◉ To ensure long term access to a wealth of valuable geospatial data so that it can be preserved and available for the future
◉ Data becomes more valuable over time for use as the basis of time series analyses or longitudinal studies
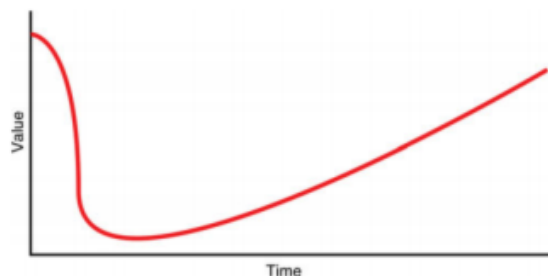
Project Specific Benefits[2]

- ◉ To make staff time more efficient and more valuable
  - ◦ The Metadata Coordinator currently spends about 10 hours/ week and a graduate RA spends about 12 hours/ week just maintaining the collections and fixing broken links
  - ◦ These hours could be redirected towards creating stable, permanent resources

- ◉ To improve the reliability of the Geoportal. Maintaining the status quo risks:
  - ◦ User Attrition
    - • Users may not revisit the site if they click broken access links
    - • Users may attempt to save records that they cannot find later, making replication of studies impossible
  - ◦ Community Reputation
    - • BTAA records may not be seen as adequate by users (few web services or stable download links)
    - • The project may be contributing to the vexing phenomenon of "link rot"
    - • We cannot share majority of records with other geoportals in good faith, because of the maintenance required

How would we implement it?

Model 1: Collective Archive
- ◉ Technology: One repository based upon a digital collections applications (e.g. Samvera/Fedora or DSpace)
- ◉ Staff activities
  - ◦ Curation
    - • Develop a collection plan
    - • Collect digital files and metadata
    - • Assess, repair, and augment data and metadata
    - • Deposit resources
  - ◦ Technology development
    - • Select and develop repository
    - • Provide preservation and migration support
  - ◦ BTAA Staff time
    - • Data Curator/ Metadata Coordinator time: 20 hours/week
    - • Development time: [400 hours + 5 hours/week] - TBD

---

[1] Kevin R. Dyke, Ryan Mattke, Len Kne & Shawn Rounds (2016) Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota, Journal of Map & Geography Libraries, 12:1, 52-72, DOI: 10.1080/15420353.2015.1073655, Image from page 68.

[2] See the document Comparing Staff Time for Maintaining Access to Ephemeral Resources vs Archived Data for more details

Model 2: Distributed Archives

- ◉ Technology: Each contributing institution would archive data in their own local repositories.
- ◉ Staff activities: Same as Model 1, but replicated at each institution
- ◉ BTAA Staff time
  - ◦ Metadata Coordinator time: 10 hours/week aggregating metadata records
  - ◦ Development time: none

What are some things to consider?

- ❖ The interviews during Summer 2019 with AULs/ADs of BTAA libraries revealed that archiving public geospatial data was not seen as a priority for their institutions

- ❖ Not all participating libraries have local data repositories – if participating institutions decide to archive individually, how would these libraries participate?

- ❖ Creating a local data repository or expanding an existing one to include public data may not be in scope for the contributing institutions

- ❖ Achieving this goal may need to be preceded by
  - ○ Educating stakeholders about the need to preserve public geospatial data
  - ○ Implementing web services to build a corpus of data suitable for an archive