

Running notes for Geo4LibCamp 2024

<https://z.umn.edu/geo4lib-2024-notes>

<https://z.umn.edu/geo4lib-2024-slides>

Shortcuts

[Summary of Sessions](#) | [Monday](#) | [Tuesday](#) | [Wednesday](#) | [Thursday](#) | [Friday](#)

Schedule	Main Room	Breakout Room	Room 308 (3rd Floor)
Monday Workshops	Medley of Metadata Workflows		GeoBlacklight for Beginners ---- Advanced GeoBlacklight: Voyaging Beyond Core
Tues 1:30pm - 3:00pm	Map Library Tour / Cavern Tour (2:15)		
Tues 3:30pm - 5:00pm	Map Search UX Workgroup / filtering by scale & extent / user interactions: what we've learned with new metadata (aardvark) and tools (Allmaps, COGS, IIIF) Huda / Melinda / Amy		
Wed 9:00am - 10:30am	PM Tiles & COGS / COGS what are they good for (and other cool	GIS metadata remediation	

	formats) Huda / Alex GeoTIFF / Image compression Melinda	Zehong	
Wed 11:00am - 12:00pm	Strategies for Collecting & Archiving Historical Geodata / Polar Geospatial Center Archives Karen / Kelsey	OpenIndexMaps Stephen & Tom	
Wed 1:30pm - 3:00pm	Georeferencing Historical Aerial Imagery Jim Georeferencing Interest Group Adam & Eliot Handling complex content while mosaicking historical atlases Adam	AI or API Enhanced Metadata, what can computers do for me? Eric	
Wed 3:30pm - 4:30pm	OpenGeoMetadata / Metadata discussion (aardvark field for collection string) Huda / Amy	What is LandScan? Ben Using STAC (Spatio Temporal Asset Catalogs)	Ethical description, but not quite: advisory warnings for maps with potentially harmful content Maggie

		Danny	
Thursday			
Friday			

Summary of Sessions

[Lightning Talks - Slides Links](#)

Summary of Geo4LibCamp 2024 Meeting Notes

Monday, May 20, 2024

- Welcome and Introductions: Attendees introduced themselves despite a power outage. The power was restored during Laura's introduction, enabling the display of slides and links.
- Lightning Talks:
 - Open-Source Georeferencing and Curating with Allmaps: Presented by Stephen Appel. Allmaps is an open-source, browser-based tool for georeferencing maps, developed in collaboration with Boston Public Library and Leventhal Map and Education Center. Future plans include the Allmaps Curator project for managing georeferenced maps.
 - Blacklight::Allmaps Plugin: Presented by Eric Larson. A new Ruby on Rails gem for integrating Allmaps with Blacklight/GeoBlacklight, allowing display of georeferenced maps with a tabbed viewer.
 - GeoServer and Security Challenges: Presented by Maura Carbone. Discussed addressing issues of serving restricted and public content in GeoServer and plans for future upgrades.
 - Ephemeral Geospatial Data: Presented by Karen Majewicz. Highlighted the issue of data ephemerality and the importance of libraries transitioning from storing print maps to digital data to avoid data loss.

Tuesday, May 21, 2024

- Keynote Session by Dr. David Haynes:
 - Discussed the role of data and scale in health research and the importance of computational approaches in handling large datasets. Emphasized the benefits of using raster data over vector data for timeliness and repeatability, particularly in health geography.
 - Explored challenges in spatial demography and defining the geographic scale of neighborhoods for assessing risks.

Wednesday, May 22, 2024

- Unconference Sessions:
 - PM Tiles & COGS / GeoTIFF / Image Compression: Discussed the use of PM Tiles and Cloud Optimized GeoTIFFs (COGs) for efficient data handling without needing a separate web server.
 - Strategies for Collecting & Archiving Historical Geodata: Discussed methods for handling complex historical geodata and the potential of AI or API enhanced metadata.
 - OpenGeoMetadata / Metadata Discussion: Focused on improving metadata practices, including the use of the aardvark field for collection strings.
 - Georeferencing Historical Aerial Imagery: Discussed challenges and methods in georeferencing historical aerial imagery and the benefits of using advanced technologies like AI.

Thursday, May 23, 2024

- Best Practices for Authoritative Data Providers: Emphasized the importance of creating a profile for authoritative data collections and explored tools like GeoMonitor for ensuring the availability and quality of geospatial data services.
- OpenGeometadata API Registry: Proposed creating a centralized registry for enhancing and serving geospatial metadata, incorporating tools like Who's on First for place names and hierarchical data.

Friday, May 24, 2024

- Closing Discussions: Reflected on the sessions and discussed plans for future improvements, including more diverse workshop options and pre-conference preparation.

These notes highlight key presentations, discussions, and proposals aimed at enhancing geospatial data management and collaboration among institutions.

Monday, May 20, 2024

Welcome and introductions

- Slides - <https://z.umn.edu/geo4lib-2024-slides>
- Attendees:
 - Ryan Mattke - Map & Geospatial Information Librarian; Big Ten Program Lead - University of Minnesota
 - Dave Mayo - Senior Digital Library Software Engineer - Harvard University -> HUIT -> LTS
 - Melinda Kernik - Spatial Data Analyst and Curator; University of Minnesota (BTAA-GIN)
 - Felipe Valdez - GIS Specialist - Temple University
 - Stephen Appel - Geospatial Information Librarian, University of Wisconsin-Milwaukee
 - Jaime Martindale - Map and Geospatial Data Librarian, University of Wisconsin-Madison
 - Jim Lacy - Associate State Cartographer, University of Wisconsin—Madison
 - Maura Carbone, Harvard University
 - Amy Work, GIS Librarian, UC San Diego
 - Maggie Tarmey, Librarian for Geography, Maps, and Economics at UCLA
 - Kelsey Zimmerman, Geospatial Support Specialist, Polar Geospatial Center @ University of Minnesota
 - Laura Krueger, Cartographic Metadata Librarian, David Rumsey Map Center at Stanford Libraries
 - Adam Cox, UIUC, Healthy Regions and Policies Lab
 - Samuel Kim, GIS and Geospatial Librarian, University at Buffalo
 - Huda Khan, Software Developer, Stanford University
 - Lara Adair, Map & Metadata Librarian, East View Geospatial
 - Alex Marden, GIS and Geospatial Data Coordinator, University of Texas at Austin
 - Girmaye Misgna, Mapping & Geospatial Data Librarian, University of Pennsylvania
 - Ben Liebersohn, Geospatial & Mapping Fellow, University of Pennsylvania
 - Tom Brittnacher, Curator for Geospatial and Digital Collections, UC Santa Barbara

- Zoe Dilles, Map Librarian, Branner Earth Sciences Library at Stanford Libraries
- Allison Anderson, GIS Specialist, University of Washington
- Eric Larson, Lead Developer for the Big Ten Academic Alliance Geoportal
- Heroically, most attendees had to introduce themselves when we had a power outage. Power came back part way through Laura's introduction (thank you Laura). We could then see the slides and links and such :)

Lightning Talks

In addition to in-person attendees, we had 19 online attendees.

Open-Source Georeferencing and Curating with Allmaps - Stephen Appel

- Digital humanities advancement grant, in collaboration with Boston Public Library, Leventhal Map and education center
- Continue working on AllMaps tools and develop new AllMaps tools
- Challenge: Want to enhance collections with georeferencing, but georeferencing can be challenging. Had tried vendor solutions, but these also have cost, access to data challenges. Smaller institutions have barriers to overcome before they can implement georeferencing projects.
- Allmaps: free, easy to use, open source. Looks similar to georeferencing in desktop GIS. Use already geolocated layer and create ground control points. Differs: Not creating derivative files like GeoTIFFs or additional files. All happens within the browser.
- Data structure: Georeference annotation. Tying ground control points in IIIF image with GIS/geographic coordinates.
- Georeferencing map automatically creates a georeference annotation which is currently stored in the vendor's database, but trying to move away from that using Allmaps curator project.
- AGSL: collections as data philosophy. Want to make paper maps more available to those who want to use them
 - Digital asset management - CONTENTdm and IIIF
 - Geographic linked data
 - GeoBlacklight: thinking of how to integrate maps into this portal
- Example project: [Louise A. Boyd's exploration](#)
- Allmaps Curator: Addresses question of how institutions manage georeferenced maps
 - Wanting to learn about curatorial tools people are interested in
 - Exploration of management of georeference annotation and manifest

- latest.allmaps.org: Shows latest maps edited with Allmaps editor
- Consortium for public geography
 - Many map holding institutions using IIIF. Exposing IIIF manifest sufficient to be able to also work with it in tools like Allmaps
 - Pushing forward for Allmaps development as well as other related tools - similarities between this process and what we see as the GeoBlacklight community process of pushing forward development
- Anyone interested in using Allmaps can apply for support from Allmaps (staff?) if they want to use Allmaps in their research
- Acknowledgements: Allmaps dev team, IIIF consortium and the two main institutional groups
- Information about the Allmaps Fellowships:

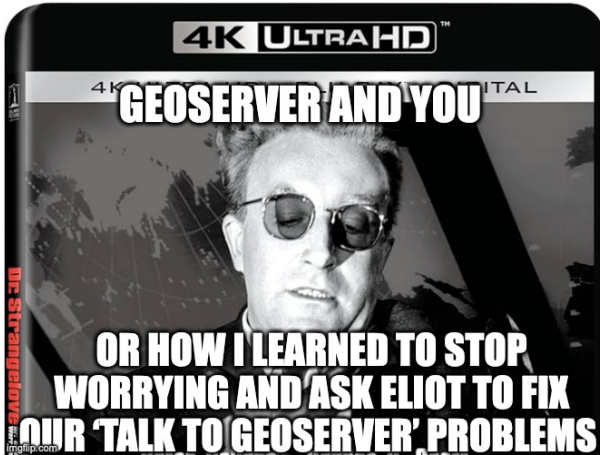
<https://www.leventhalmap.org/projects/digital-projects/allmaps/#allmaps-research-fellowships>

Blacklight::Allmaps Plugin - Eric Larson

- New Ruby on Rails gem for installing Allmaps to Blacklight or GeoBlacklight instance
- allmaps.org
- Working with Allmaps IIIF annotation data and Leaflet plugin for Allmaps
 - Take the maps that are... AllMapped... and make them visible in Blacklight/GeoBlacklight
- GeoBlacklight example
 - Adds tab to “Georeferenced Map” as second tab (next to Item Viewer which shows the usual view)
 - If the map is not georeferenced, the “Item Viewer” tab will show. The right hand side column will include a “Georeferencing” section which will link to Allmaps to enable georeferencing the item.
- How this works
 - Looks for IIIF manifest information in the Solr document (dct_references)
 - If it finds them
 - Displays the tabbed viewer with item view and georeferenced map
 - If not
 - Displays the right hand Georeferencing section
- IIIF annotation sidecars: The Allmaps information is saved in the local Blacklight/GeoBlacklight database
- Configuration options: e.g. if IIIF manifest information is in a field different from the default expectation

- Big 10 Geoportal: geo.btaa.org:
 - Facet “Georeferenced via Allmaps plugin” shows which items have been georeferenced in this fashion and those that are not.

GeoServer and You: Or how I learned to stop worrying and ask Eliot to fix our ‘talk to GeoServer’ problems - Maura Carbone



- Problem: Restricted and public content in GeoServer. How do we serve both?
 - Security mandates, e.g. 2 factor authentication
- Initial solution aka the proxy of proxies
 - Ended up with 3 proxies
 - Two: Geoserver restricted and Geoserver public.
 - Nginx proxy would sit in front of these, and would pass through the request to the appropriate proxy back to GeoServer/forward to GeoBlacklight
 - Lots of dropped requests
 - If something broke, not easy to figure out which proxy was the main problem
 - Team was still learning about the ins and outs of GeoServer performance and fine tuning. Not always clear what/where the problems were
- A very poignant image of Maura’s cat retreating from strangers was used to indicate Maura’s emotional reaction to being told about GeoServer issues. What could she do but restart and hope for the best?
- Plan: Conversations on GeoBlacklight slack helped with information needed for the plan
 - Eliot to the rescue! Had set up two files of code in GeoBlacklight itself (the glory of it all!). Seemed much better than Schrodinger’s three proxies
- harvard-its/HGL4-1: this plan also brought them in line with how many Harvard apps run (notetaker’s Q: with respect to security? Security and structure! Most of our apps did not

have/require 3 proxies to deliver data, and use the download service we have for pre-packaged files)

- Ops team elated that we went from 3 proxies, 1 GBL, 2 Geoservers, and 1 download service to 1 GBL and 1 Geoserver (huzzah!)
- When upgrading to GBL 4, want to use PMTiles and COGS. Align better with other Harvard apps work. Will document this work, and happy to share!

Ephemeral Geospatial Data: An Impending Dark Age - Karen Majewicz

- Data ephemerality: Resources that are available one day disappear the next
- Public geo data: Many revisions which overwrite yesterday's records
- Problem: Researchers rely on data from a specific time.
 - If users do not find the data online, they assume it does not exist
- Map libraries: print maps stored in libraries. Centuries of cartographic history.
 - Many libraries have not been able to expand to use digital formats
 - Eg. FDLR. Designates repository libraries can provide public access to federal government documents. Also physical media such as CDs.
 - Many libraries now don't collect the actual maps, but rather catalog records that point to digital resources (such as in government websites/repositories)
- Commercial map companies use local data to create maps, but many of these companies have gone out of business - so if state/county has info, will just provide their info in data portal
- Collection analysis
 - Road maps and data across BTAA region
 - Looked at
 - Library held print maps
 - Public open geo data
 - Over 100 years
 - What is time period depicted in these resources?
 - BTAA Collection Analysis Tool
 - Can do a faceted search across all the data
 - Physical maps depicting roads
 - 20s high
 - Drop in the 40s
 - Up again
 - Down in the recent past
 - Digital datasets

- Time period: More recent past. (Uptick?)
- Combined print and digital resources
 - Going by year, seeing if print or digital resource available
 - Gap in the 2000s where none available
- Ephemeral datasets study:
 - Data for a particular year in 2017/2018 and then data for that same year in 2023
 - Shows a drop in 2023 i.e. data gets sunsetted/removed
- U Wisconsin Madison Robinson Library
 - Proactively collecting data from all WI counties
 - For roads and WI: Making huge inroads, especially for 2010s. Their value will grow especially if only state in region that has comprehensive data that can be used for research
- BTAA-GIN incremental steps to a geodata collection: Various steps starting from 2015 onwards. Building towards data curation
- Call to action:
 - Start mall: snapshots even if no repository
 - Minimum viable solutions
- Forecast
 - Maps fading out in 2000s, and then data really last couple of years
 - If we don't do anything about it, "dark age of data"
 - But if libraries step in - can fill in this gap
- Paper that folks wrote about this topic (can someone drop in a reference? [MK -> It's still in review (first draft submitted 5/10/24) - hopefully will eventually be available from the Journal of Map and Geography Libraries])

(Some really great questions around openstreetmap/geofabrik and accessing historical data, etc./tracking changes in the data instead of having to download the entire data, etc. Did not capture that in the notes) What does diffing geo data look like?

Issues of grant-based funding models vs continuous funding - for continuity it's important to have an ongoing funding stream.

Workshops

Medley of Metadata Workflows

Metadata workshop repo: <https://github.com/opengeometadata/workshop>

[OGM Aardvark](#) is a relatively flat schema that incorporates some fields from Dublin Core.

GeoBlacklight schema was expanded to account for additional use cases after being first developed by Stanford folks in ~2014

- Adds fields for rights (standardized these when everyone seemed to have their own implementation earlier), relations
- Current way relations are handled is parent-child relationship based on ID, but this isn't perfect for all use cases.
- Geometry field opens up more potential for spatial info beyond bounding box
- References: JSON Hash field with key value pairs. Multiple types of keys
<https://opengeometadata.org/reference-uris/> . OpenIndexMap is one of the keys.
 - Create a separate record for the index map, and then use this field to reference that record. May also want a relationship pointing back from the index map to the other record.
- BoundingBox no longer required. This field is now optional (whereas it was required before)
 - Helpful for collection level record, or if incorrect/missing bounding box
 - Solr_geom field is what search uses, whereas bounding box used to calculate overlap between display window and individual results
 - If the bounding box is not present, the document just doesn't show up (in search? In general?)
 - Info about spatial fields: <https://opengeometadata.org/spatial-fields/>
- How to make an Aardvark record based on a CSV spreadsheet.
 - Templates and scripts in: [workshop-main\geo4libcamp-2024\01_CSV-2-JSON](#)
 - Fill out the template and then run the associated python notebook
 - The script uses a [crosswalk](#) to transform the labels into the aardvark schema and structure
- OpenGeoMetadata (github.com/OpenGeoMetadata): where institutions upload their GeoBlacklight schema records.
 - First pull down the records, turn into CSV, and then re-ingest into metadata editor
- Harvesting from ArcGIS Hub APIs ((DCAT metadata) into CSV ([geo4libcamp-2024/03_harvest-DCAT](#))
 - The scripts attempts to clean up records including:
 - cleaning titles
 - sorting out things that aren't spatial data layers

- The script creates a CSV
- 503 Errors are common on these sites - so sometimes the scripts don't work
- Esri often changes their schema without announcing it - so you have to constantly adjust your scripts to keep up
 - Idea: write a CRON job to check key, value against the expected values
- Scripts to [Clean and Validate](#)
 - The script checks:
 - For some required fields
 - Order of dates
 - Missing formats but download link available
 - Bounding boxes
 - Rounds coordinate fields to save space
 - Checks that they are in the correct order
 - Checks it is a box not a point (and makes a tiny box if it is)
 - Creates a CSV with cleaned values
 - Sometimes the Bounding Boxes still are incorrect
 - Probably an issue from the source
 - Might spend time during the working days this week to add in some additional quality checks
- If multiple institutions are scraping the same data portals, we should check to make sure we don't have duplicate records for these items
 - If harvesting from OpenGeoMetadata, could be a problem because the records aren't identical.
 - Share a stub record?
 - Danger of overwriting items if IDs match?
 - Recommendations for namespace to keep records separate
- [iso metadata](#) (Kim Durante - Stanford)
 - checkData :
 - When importing data, several issues can cause things to break (eg. no projection, no coordinates)
 - Pulls technical information needed from the files into a spreadsheet to make it easier to compare and look for issues
 - createMetadata :
 - Takes an XML template with placeholder blanks and fills in the values from a CSV

- India dataset: Very large and expensive dataset. Attribute field names not self explanatory. Have a script that generates metadata (in ISO) that help to explain what the field names represent.
- There is also a script that converts ISO to Aardvark schema (even though Stanford does not translate ISO directly to Aardvark).
 - More challenging/areas to review e.g. Generating “is part” fields.
- In Earthworks, the iso metadata file is made available as part of the data folder. Moving forward they are going to make the file available separately (but have to re-accession thousands of file)
- In GeoBlacklight, these iso XMLs can be linked to as a dct_references_s
- Discussion:
 - Is Aardvark still a discovery schema or is it moving towards more descriptive?
 - At first it was just for data, but now including a wider range of items (such as maps, portals)
 - Some folks are moving away from creating iso metadata - considering just creating Aardvark or linking out to metadata in other formats
 - Some currently use ISO because it is supported by ArcGIS Desktop, but if there was an alternative editor for guided metadata creation might consider it
 - Can't create metadata for GeoJSON in ArcGIS (a problem for things like OpenIndexMaps)
 - Advantage of ArcGIS - it automatically pulls some technical information

GeoBlacklight for Beginners and Advanced

- Workshop Tutorial Blog Post:
 - <https://geoblacklight.org/blog/2024/05/geo4libcamp-2024-geoblacklight-workshop/>
- Example GitHub repo of the workshop codebase:
 - <https://github.com/ewlarson/gbl-dev>

Tuesday, May 21, 2024

Keynote session

Dr. David Haynes – Bridging the Need Gap: Designing a Digital Social Care Referral Platform to Address Health-Related Social Needs

- Takeaways
 - How do data and scale inform health research problems?
 - Understanding how research teams and problems may be limited by data and scale
- Data
 - What are data structure, types, formats?
 - People working with larger and bigger sets of data
 - Asking myself: What is the computational way to handle data?
 - As health geographer, working with researchers, and can see they are not really thinking about computation - which will become a limiting factor for them later
 - Data models
 - Vector: Lines, polygons, points. Discrete conceptualization of space
 - What researchers may not understand: what info will come into this discrete conceptualization
 - Raster: Satellite or sensor collected data. Continuous concept of space
 - Try to push researchers to using raster data. Vector is a good start, but important for people to understand more benefits (timeliness, repeatability) with satellite data. (Important for areas like disease research).
 - Health geography
 - Driven from vector model.
 - E.g. COVID. Want to know about maps. Not good at mapping that amount of data. Timeliness: Making maps that change continuously over time. Time issue b/c of volume of data
 - Worked at the Institute for Social Research and Data Innovation
 - Exposed to different types of social science data: Survey data, panel data microdata, big data computation

- How do we think about aggregation?
 - What happens when we aggregate microdata?
 - Different kinds of questions that can be asked?
 - Proximity to pollution site VS who is really at risk
- Spatial demography
 - Who is at risk? How do you define the geographic scale of the neighborhood?
 - Neighborhood: Fuzzy question to answer for geographers
- For these kinds of analyses, what tools should you use?
 - People don't consider this until they are deep into using these already and can see these tools do not work for them
 - People can't master all tools.
 - If starting a new research project, start thinking about tools that will be more useful for adaption (e.g. regression - need to rerun the numbers so let's create a script that can do that).
 - Projects with heavy investment and heavy computation - what are the big data tools available? Some of these tools may not be overly complicated to use, and will tell students about these tools.
 - Helping people to understand the continuum of tools where that project might be
- Use cases for big geo data
- No SQL databases for spatial
 - PostGRES is great but also other systems: neo4j, mongoDB, rasdaman SciDB
 - Can keep track of your data without necessarily having to learn the tool completely
- The goal is better spatial science
 - The more you can document in some way, the better off you will be
 - Data preprocessing iceberg metaphor
 - Info that is reported is the visible part of the iceberg
 - What is not reported can make up the majority of the iceberg (which is not visible)
- Scale
 - Modifiable areal unit problem
 - (https://en.wikipedia.org/wiki/Modifiable_areal_unit_problem)
 - E.g. Childhood cancer. Need 15 years of data.
 - Challenge: Two different censuses and population is different
 - Researchers who don't work with geospatial data may not understand the nuances/challenges

- Understanding the unit of analysis may affect their work e.g. do you need the county level?
- Sage client distribution
 - Free breast cancer screening for women without insurance (or under-insured/low insurance)
 - Question that tried to help them understand: If they screened 15,000 women, did they screen all the women?
 - Need to think about where those women live
 - If screening low-income women, need to know where those women are
 - If look at using areal unit - with the population density for these points will make it difficult
 - Need a tool/method that can handle low population densities and high population densities at the same time
 - Synthetic population dataset made by Research Triangular Institute (RTI). Household and person population dataset based on the 2010 Census. (Can download the 2010 data by contacting the keynote speaker. 2020 needs to be purchased)
 - Took microdata survey - assigned household characteristics
 - Large dataset, but would crash most GIS software
 - Spatially adaptive filters
 - Computation method: Need to calculate how many people need to go in a circle
 - Makes cool maps - taken vector data and moved into raster process. Can now understand where high levels of people who are getting screened and where people who are eligible to be screened are also living
 - Using smoothing techniques - can understand processes at very low levels
 - Minnesota: ~40%. Most states between 20 and 40 percent
 - Zooming in: Looking at Native communities. Reciprocity agreement.
 - High rates. In fact 104 in some places (so know it's not a perfect map)
 - Folks are being geocoded to the post office
 - Lots of people who live in reservations don't have street addresses, so artifact where aggregating people's addresses to the same point
 - Another map
 - Lower-income communities appear to have lower screening/utilization rates

- Helpful for the project since this data shows them where they need to work
- Use the same data for the county level map
 - Seems like we are doing pretty good
- Use the same data for zip codes
 - Now the results don't look so good
- The units we use will affect the story/the results you see.
- Strong advocate for telling researchers: use census tract
 - Lots of researchers like to start with zip codes b/c it's easy. Will not be as accurate as you'd like it to be.
 - Slide: Census tract divides the population into multiple groups, whereas only two zipcodes represented in the same area.
 - Created different maps to show different unit analysis
 - BUT in this particular example, turns out the zip code analysis was in fact better than census tracts (using correlation between maps as the criterion)
- A lot of public health is about identifying problems that exist in public health
 - Wanted to change approach where we actually help (and not just identify)
 - Breast cancer champions: Identifying communities of African American women, and then going out and helping screen
- Health (social determinants of health)
 - Brief history of health geography
 - When being trained in this area, lots of vector epidemiological model
 - Then focus shifted to the environment - bringing that into GIS to understand people's health
 - Last 5-10 years: Genetics. No, actually epigenetics (how the environment affects genes). Not just that: Healthcare system with biases/structural and organizational barriers. If not that: Social Determinants of Health (which can cover anything)
 - How do we understand the "anything" of Social Determinants of Health
 - Trying to use the census tracts/census data
 - But: Not quite working. Any individual is more than their zip code, etc.
 - Motivations: Fragmentation/disconnect of service providers and those seeking health-related services
 - Difficult for people to find the information they need

- No uniform way to access and learn about services - with updated real-time information. Don't have a Yellow Pages anymore - and even if we did, would not be updated beyond publication.
- "Self-service" model for patients and community organizations
- Smart community health app and dashboard
 - Mobile app: helping people to access this information on their own
 - 2021: focus on particular patient population. 1 in 4 people will experience cancer.
 - >9,000,000 individuals with cancer struggle with social needs
 - Cancer: has a huge financial impact
 - Providers know this. They provide a list of providers. Health support. But the lists are outdated. System is broken
 - As needs change, have to go back to the provider to communicate different needs, then need to assess which programs will fit their needs (e.g. certain programs for
 - HIPAA limits the information that healthcare professionals can share with other organizations
 - ConnectedNest - an app that assists with connecting community members with resources
 - People are more comfortable disclosing needs on their own phones. The more success they see with having their needs met, the more they will come back to use the system
 - Also built support for providers
 - Story: Kiara downloads the app. Agnostic to healthcare system. Complete the assessment for their specific healthcare system.
 - Identifies she has food needs as well. Can connect with services.
 - Sonja (staff member at food4life): uses the community portal to create programs.
 - Want to think about timing and capacity. The app allows Sonja to identify capacities for the program, and when that number of people have clicked on that program, then it is no longer available for selection. Trying to be very clear about what is/is not possible with these programs.
 - Clinical team: can see (manage?) referrals
 - Able to see the information for Sonja and can see which programs will work

- Info is HIPPA protected, so will come back first to Sonja i.e. “Doctors reviewed your info, and suggested these services/referrals”.
- Financial assistance allows Kiara to complete her treatment
- Want to be able to more clearly define what programs are, what needs they serve, etc. to bridge the existing gap between program definitions and people assessing which programs apply to them
- Promote access and equity
 - Minimize bias, integrate community support, and improve survivorship
- Two-sided market
 - Patients looking for info. Going back to healthcare providers and letting them know what will help their patients
- Mixed-method pilot
 - Process
 - 3 months
 - 41 cancer patients, 14 cancer specific community organizations
 - Ask users to complete assessment of needs
 - Usability surveys
 - Community based organizations were then receiving requests and were told to be responsive
 - Social needs collected through ConnectedNest
 - Many different types that still need to be addressed
 - Were able to help 39% of participants to reach out and get their needs met by community organizations
 - Physical activity: most used program. For people who were past treatment and wanted to get back to better health.
 - Interviewed patients
 - Motivation: Connecting to resources, help others in the future, etc.
- Next steps
 - Continuing research in Minnesota - expand beyond to the Midwest (and eventually the US)
 - Continuing to develop tools
- Takeaways review from beginning of presentation

Questions

- People, like some of the elderly, who may not be comfortable with apps?
 - Individual going through treatment already overwhelmed trying to stay healthy
 - Legal way to get someone else to assist
 - For elderly, thinking of other ways to work with technologies
- ConnectedNest: health provider buy in? Another task on overworked population
 - Health care provider buy in changes. Lot of organizations that we are competing with. E.g. 2-1-1s. Also trying to position ourselves as tool that can help with social workers.
 - Not a tool to replace you. More like checking email - see clients and move them along.
 - Thinking working with social care workers will be a helpful way to move forward
- Cancer: condition that we don't know everything about. Needs for recovery. Asking about specific case related to dermatology. Usually based on needs of white skin, so dermatologist may give treatment plan that may not fit your needs. Some dermatologists revisiting fundamentals and rebuild with not just white populations in mind. Advice for specialists in this area - don't really have much in the way of treatment plans or community organizations
 - How to get researchers to engage with community: long hard process. Research is an extractive process. Communities may feel that they are at the end of a long process. Researchers need to connect with communities - may take up to 6 months. Find community organizations that align with your agenda at some point, and try to start working with them as early as possible. With communities: need to build a relationship and learn how to work with them.
- What do you see as role of data providers?
 - Where are they on spatial analysis software continuum (open source/enterprise + desktop/geospatial data science/big data spatial analysis)
 - What's the investment - what grants
 - If educating the researcher a lot, perhaps at the beginning spectrum. Maybe ask do you know geospatial researchers who can help you build those tools.
- How do you approach educating researchers about biases in census data?
 - Learned a lot about census data through working at population center
 - E.g. age group: will move people to make cell size match.
 - Understand some examples - the mean or other variable is not what you want to look for.
 - Census is more than population data. > 18,000 variables. Need to see which ones you need.

Lightning Talks

In addition to in-person attendees, we had 18 online attendees.

Enhancing Access to Digital Aerial Photography at the University of Wisconsin-Madison - Jaime Martindale and Jim Lacy

- Largest print archive of historic aerial imagery in Wisconsin
 - Post pandemic, doing a lot of scanning
- Historic Aerial Photography:
 - Sourced from film or are prints on film
 - Federally produced is main focus, but some state
 - ??-2000 is the the first priority, then NAIP (digital) 2004-
 - ~38,000 images
 - Print indexes
 - Want a user to be able to go to an area and get all the photos
 - USDA+USGS have a massive film scanning effort underway
 - USGS: earthexplorer.usgs.gov
 - USDA-APFO single frame from 1950s+1960s have been acquired by Wisconsin staff
 - What is on the horizon?
 - We need a way to track all this
 - Tack and index photos including
 - Attributes, description, data type
 - Index processing - ArcGIS DataInteroperability Extension
 - Pulling data out of filenames etc. and putting it in the metadata file
 - Helps people who can code but aren't masters in Python the like to increase what they can do
 - Data Storage - Research Data Services - helps make 20TB+ of imagery available to the public
 - \$60 per TB per year for data storage
 - Globus -
 - Need cloud optimized geotiffs (COGs) which can take 1.5+ months on local traditional tools (GDAL/ArcGIS)
 - But by using the CHTC resources, they can rerun 38k images in no time at all!
 - Dark Archive
 - Want this to ensure data for future generations, extra backup to campus's backup (what about human error or amass hardware failure?)
 - AWS Glacier ~\$12 per TB per year (with many caveats)

- WHAIFinder 3.0
 - Using COGs things go right into the map view
 - Using ESRI tools
 - Want a time slider to filter dots relevant to timeframe for user
 - Moving away from IIIF, because other groups are doing that and no one ever used the small/medium/large IIIF downloads
- What about GIS users?
 - Can use download or script to get to them
- Question: glacier protects against ransomware? Yes, also that.
- Question: Orthrectifying? Not something happening right now but other counties did it
- Question: Where do center points come from? Depends, sometimes it comes with the data (and had to be fixed), sometimes people with free time created indexes for them
- Question: Sharing the interface code? PPlan is yes! But caveat is it is esri ecosystem

OpenIndexMaps for 19th-century atlases of the upper Mississippi River - Owen Connolly

- OpenIndex maps build on traditional index maps which stores index map geometry and download links in GeoJSON and can be viewed in an online map viewer
- Click on the square, you can get the original map record and map
- Did a study to experiment with OpenIndex maps to make things 'useful for geographers'
 - Worked mostly on historical charts for the Mississippi River
- Process
 - These maps did come with a traditional index map
 - Then imported index map scan into qgis and georeferenced
 - Then had to manually trace each of the polygons in the index map for each map that would go in
 - Next, added the metadata to each polygon (title of the map, a label, a note, and then links for download and thumbnail)
- Visualizations:
 - BTAA GeoPortal
 - Load things to github and then load to the geoportal, where geoblacklight can visualize
 - ArcGIS Online
 - Can be loaded and turned into hosted feature layers
- Limitations:

- Process can take a long time without a starting index map (but honestly having to trace is also time consuming)
- Overlapping polygons can get confusing

Example from presentation: <https://geo.btaa.org/catalog/p16022coll247:83-index>

What is the National Geospatial Advisory Committee? - Ryan Mattke

- National Spatial Data Infrastructure
 - “Promote geospatial data sharing”
 - “Empowering a geo-enabled Nation and world for place-based decision making”
 - Membership across many spaces
- Academic Library Perspective
 - Digital stewardship Pyramid
 - Get to say the word archiving a lot in these meetings :-)
 - Looking at long term preservation
- Enabling the future of GIS
 - Want the data for students etc. in a digital world!
- NSDI Strategic plan
 - Going to be a 10-year plan
 - Providing final feedback, calling for public comments
 - Has to be high-level enough that federal agencies can connect to it, but detailed enough with action items
 - The Federal Government is HUGE

The UC approach: Single code base, separate deployments - Amy Work and Tom Brittnacher

- Project Surfliner:
 - Collaboration Effort
 - Digital library products
 - SUITE of products, not monolithic install
 - Community-based open source infrastructure
 - Shared code but separate installations (so no forks)
 - So GeoData as an example
 - Single Code Base, SINGLE deployment example
 - Then Single code base, SEPARATE deployments example (teams are all named after trains) (we’ll want to link to the slides since trying to describe the chart is hard)
 - Wanted extra collaboration for geodata

- Data and metadata processing
- Went with M3 metadata, using derived properties to normalize differing base properties
- What We've Learned:
 - Collaborating made them question everything rather than just do what's always been done
 - Challenging and time intensive, but rewarding and sustainable because of shared resources

Using Aardvark to Drive Public Health Data Discovery - Adam Cox

- Grant Goals
 - Build capacity and community
 - Community toolkit and set of guides/fellowship programs in web dev and data creation
 - And data discovery app (which led to GeoBlacklight)
- Using and building on Aardvark
 - Added extra fields
 - Handle multiple publication years as Collection records
 - "Is version of", parent-child
 - Spatial Resolution
 - State, city, county, block group, census tract
 - Data variables
 - All variables in the dataset
 - Methods variables
 - Variables used in creation of data (eg: Social Vulnerability index) to provide extra information and context
 - Metadata Manager Web Application
 - Flask app that allows for editing, basic login etc.
 - Make a nice interface for creating a new record and viewing/editing metadata
 - Running from a single JSON schema, indexed directly to SOLR
 - View/export JSON for each record as a Solr doc
 - Plugin for search enhancement
 - Enhancing search via query expansion
 - So search is "social vulnerability index" but this can use LLM to expand that query into multiples to get more/better results

- Put lambda API gateway that then goes to solr and augment response from solr before going back to GBL

Using Text Recognition to Georeference Aerial Photos? - Kevin Dyke

- Aerial photographs are a huge source of use/requests
- Can we create a google maps like interface to put in an address and get the picture of the farm in 1933?
- Trying to accommodate a variety of needs is difficult
- Manual georeferencing? NO THANKS
- Historical imagery lacks metadata necessary for this work
- Made a laydown index
- DeepLearning in ArcGIS
 - training a model to identify labels and create polygons using ArcGIS Pro Deep Learning
 - Accomplishes the goal of creating center point for individual aerals
 - ESRI has an OCR model for Deep Learning, but it does NOT do well for this application
- Map Kurator:
 - <https://knowledge-computing.github.io/mapkurator-doc/#/>
 - Getting started can be a struggle but the team is super helpful and eventually got it working
 - Performs quite well given this was not explicitly trained for this
 - Not always perfect (CUA-6B-15 detected as CUA-GB-IS), but students can do this in their downtime
 - Flight Directions
 - It is confusing
 - Made a multipage explanation and done demonstrations
 - Still have to do manual georeferencing but this helps
 - (will get the rest of the slides up later!)
 - Sadly no free things, but they have ESRI
 - Tool in ArcGIS to create seamlines

Unconference Sessions

Map Library Tour / Cavern Tour

- [Map Library digital collections](#)
- [Minnesota Library Access Center](#)

Map Search UX Workgroup / filtering by scale & extent / user interactions: what we've learned with new metadata (aardvark) and tools (Allmaps, COGS, IIIF)

Roles:

- Timekeeper: Tom
- Gatekeeper: Dave
- Notetaker: Huda / Maura
- Facilitator(s): Huda / Melinda / Amy

[GeoBlacklight Map Search UX Workgroup](#)

- [Produced Design Document](#)
- Landing page (from above):
<https://xd.adobe.com/view/5c8645c6-782a-49bc-b458-d1a0523ffdcf-09d5/>
- For reference, known map search issues:
https://docs.google.com/document/d/1Ne0MLT73ltBm1B5gW7WISZDFNMS_NUgLLi-kzwjIEEc/edit#heading=h.y7qbpo2he7uh

Scale/extent:

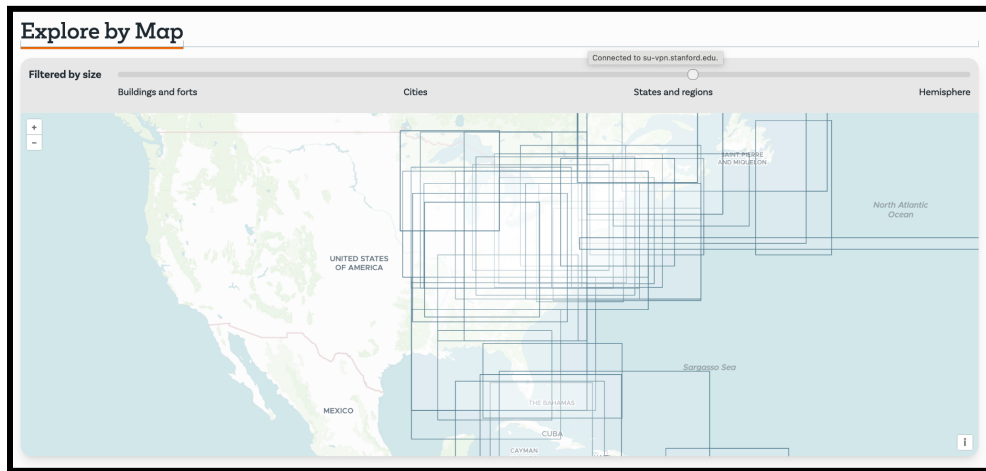
[BTAA Bounding Box quality check spreadsheet](#) (categories for the size of maps by area and flags for potentially suspiciously shaped bounding boxes for manual review)

Example scale finder: <https://www.argomaps.org>

NOTES

- UX Search WorkGroup met 2022-2023 (linked above)
- Wanted to evaluate the user experience with the interface/various topics around map search
 - Produced a design document (also linked above), mockups of possible designs
- Taking a look at breaking things down, keywords, interactive and advanced search with comments from the working group

- Keywords (google-like just searching like “American Map”)
- Location search
- Advanced searchbuilder (adding multiple fields like country, resource class etc.)
 - Includes limiting by geographical area
- Item page has image at top, metadata below
- Map-based search (location search) view options, shows where you’re nested in terms of geographies
- Question: Considered making advanced search the homepage that a person sees first? Yes, that is ONE option or ability to easily navigate (question person would not want to see that as the landing page as it isn’t simple/can be overwhelming)
- Question/Suggestion: Are you considering saving searches? That’s actually a feature already in geoblacklight! As a guest you do get that list too (just ephemeral), but not everyone has that turned on. And not everyone has logins (don’t have data on how much folks use bookmarks)
 - There is some UX data that folks prefer using the browser bookmark instead of needing accounts for each thing
- From user angle: struggling with three search bars.
 - AllMaps search: starts with a general search bar
 - Are the boxes each searches or just explainer boxes?
 - Distinguishing between a keyword and location in a search
- Stephen: When watching students search - if they search for a place name, they expect the map to go there. Would be interesting to see how many people actually use the map search
- Stats on big10? Analytics and user testing both indicate people don’t use the map very much, but what could we do to facilitate that?
 - A paper Mara Blake(name right?) put out said the same and that people were typing in place names but if place name isn’t in metadata...
- Question: Drop a pin model? Versus zooming in? Or circle area? Also could we toggle off anything at the global scale?
 - Turning off at global scale is like if I want stuff in Minnesota, I will get global population maps because Minnesota is PART of that but I don’t need that level
 - Argo (argomaps.org) will let you see the scale



-
- BTAA categorizing by Area Category (like zoom levels, 1-World, 8-Neighborhood etc.)
- Cornell limits itself to NY to start, rather than defaulting to worlds
- There is an overlap boost ratio setting which will prefer overlapping within your view
- Additional challenge: scale is separate from resolution, so it could still be useful if it's higher up than their area (so a state map that's high res might just be good and can be clipped for use about a city)
- Resolution would be nice as an option!
- Idea: connect to gazetteer locations to use preset hierarchy
 - Has come up with machine learning too
 - This will be a breakout talk!
 - Better metadata would lead to better search results
- The bounding boxes for a global scale does make the map less usable
 - Because the world is IN the search, the map must show the world because it is fitting those 2 world results that hit
 - Can also look at this configuration: https://github.com/geoblacklight/geoblacklight/blob/032ad8e024cf02928afce2da3ab58bfaaeb1d879/app/models/concerns/geoblacklight/bbox_filter_query.rb#L26 and increase overlap ratio boost
- Question: Biased average of bounding boxes? Problem is doing analyzation on the fly, moving map places new bounding box queries in your search
 - BTAA did a rough look at analytics and only about 17% are using that map in search based on bbox in search url
- Did look at Zillow/Airbnb, didn't replicate but they mostly have pins and not boxes
- Layers at zoom levels/being able to navigate hierarchy
- There is infra work that's done to prepare us (Yay eliot)

- Some of this is sort of dreams vs realism, but we want to address some of the UX pain points
 - Map and facets aren't always friends
- Could determine a scale
- Wanted to take entire corpus of OpenGeometadata that was enhanced and re-uploaded
- Wanted to also make configurable download option so that you could offer on the fly derivatives if you WANT to (wms derivatives), because some folks don't want to offer it
 - Kind of confusing for users who aren't in the know
 - May get different files based on what the institution may offer for a download (zip vs. some other format)
- Scale/extent question: Do we need to do this in the metadata or have GeoBlacklight figure out in the applications?
 - Metadata remediation is different
 - At Solr level, may create a classifier
 - Don't think we'd need Java :)
- When you get closer to the poles, bounding boxes are not really square any more
- All the data living in the solr instance can make things hard for people to make edits (since if you don't have direct solr access)
- Do a spike (an issue you need to learn about, much narrower than a sprint so it is a research spike)
- We are limited in that we're building on top of blacklight
- Does aardvark allow for changes because it has more in-depth relationship modeling?
- We have a good structure of meeting frequently, having community sprints etc.
 - Things happen when people dev for their own and then contribute back
- Using some more reference information within map to help some of the map interaction
- <https://opengeometadata.org/aardvark-gbl-1-crosswalk/> shows what has been added/expanded
- We liked some of the aardvark+ adam did from the lightning talk
 - Indication of resolution/scale
- Question: User studies are happening at Stanford? UX is going to come up with designs and do the usability testing, so yes something will be coming, summer is timeframe for design (maybe)
- Question: do we have shared map collections across institutions? There is a pilot Zoe is doing for index maps (breakout session)!
 - Could create collection record itself that has the identifiers, so someone would have to make that record
 - How would we identify? How do we bring things in?

- Question: when we bring things in from other places, people will ask about it but they need to go to the other site – how do we facilitate that?
 - Talked about one geoportal but for restricted data it won't work
 - How do we balance discovery vs use of each site?
- Cogs are fast
- Changing metadata has made it so folks have been a bit blocked by moving to aardvark.

Example of Aardvark's geometry + bounding box: <https://geo.btaa.org/catalog/p16022coll230:2937>
(the dot thing is a customized BTAA but it can go to core!)

<https://geo.btaa.org/catalog/p16022coll230:2937/admin> is fancy and the room is impressed

Wednesday, May 22, 2024

Unconference Sessions

PM Tiles & COGS / COGS what are they good for (and other cool formats) / GeoTIFF / Image compression

Roles:

- Timekeeper: Ben
- Gatekeeper: Stephen
- Notetaker: Huda/Maura
- Facilitator(s): Huda / Alex / Melinda

NOTES

- PM Tiles
 - <https://github.com/protomaps/PMTiles>
 - <https://docs.protomaps.com/pmtiles/>
 - <https://gdal.org/drivers/vector/pmtiles.html>
 -

- At UT, relying on a web mapping server. Excited when heard about GeoTIFFs and PM Tiles. Transferring services to COGs. Working on PM Tiles still. Relied on Eliot to get it up and working.
- Would be useful to talk over the background, step by step process for how to produce COGs
- Cloud optimized GeoTIFF: <https://www.cogeo.org/>
 - Pyramidal tiff (zoom levels built into the Tiff)
 - Building tiled images into GeoTIFF
- If you build a TIFF in this way, also have metadata about the tiles. This means you do not need a separate web server.
 - Client knows to ask for front part of the file - can get information about where that tile is
 - If zoomed into the image, can only download that image portion and not the entire GeoTIFF
- Makes it lightweight solution - not having to run a server
- Princeton example
 - <https://maps.princeton.edu/catalog/princeton-n583xz799>
- Can be really big. Came out of satellite imagery - huge datasets that needed a better way of serving up that data
- Q: Comparison with tile cache
 - Before: Upload GeoTIFF into GeoServer. WMS or X/Y tiles - would chop into smaller tiles. When zooming in, request tile at zoom level which GeoServer would serve up and cache.
 - Q: Are COGs cached?
 - They could be, depending on how the images are served. For example, using AWS, then it would be cached
 - In COG - the image is not chopped into tiles. Client libraries use HTTP range request that request a specific piece of the file. Uses existing HTTP technology to find specific parts of the file and show those. File format that makes that easy.
 - https://developer.mozilla.org/en-US/docs/Web/HTTP/Range_requests
 - Pieces are small enough that data is sent fast, even though not caching (as part of COGs)
- Q: How do we utilize COGs?
 - Generate COG from original GeoTIFF
 - Can use GDAL: handles COG natively.
 - Can do a batch script
 - ArcGIS PRO will export as COG if you have mosaic dataset
 - Example for GDAL: <https://gdal.org/drivers/raster/cog.html>

- Need to put them on the web: e.g. Amazon S3 buckets. Needs to be able to request little parts of the file (a lot of cloud storage can support that)
 - In GeoBlacklight - use dct references to point to the COG
 - Note: COGs will be bigger than the GeoTIFF files
- GeoServer handles COG formats better. Even before COG specification, GeoServer (with some configuration options) may already have been storing the GeoTIFF in this fashion
- Princeton
 - Pyramidall TIFF
 - IIIF server also uses this model (or integrates into how Princeton set this up)
- Harvard notes COGs has enabled a big performance boost
- Need to support range requests to enable this
 - Ruby on Rails, Django does not
- GeoBlacklight dct_references_s key/value example:
 -\"<https://github.com/cogeotiff/cog-spec>\" : \"https://geodata.lib.princeton.edu/e2/07/f4/e207f44edbac4f7f94df8811a40186f7/display_raster.tif\" }
- Q: GeoServer and COGs: can they work together?
 - Depends on what you want. If you want WFS/WMS, yes.
 - Would use the same COG file for GeoServer as well - can also point GeoServer to S3 bucket for instance
- PMTiles
 - Spec for file format - tile vector data
 - <https://github.com/protomaps/PMTiles>
 - Multiple zoom levels - can ask for parts of the file and get that part back
 - Somewhat different from COG
 - E.g. <https://maps.princeton.edu/catalog/princeton-1r66j4624>
 - Client reading file. As we zoom in: ask for little parts of the file
 - When zoomed out: don't want to return every shape.
 - Part of processing pipeline:
 - Merging polygons for the zoomed out level, so don't need to render all the shapes
 - Works really well at Princeton - processed 1000s of datasets.
 - Some challenging very large datasets. ~4 out of 10,000 that are challenging.
- Q: Support WMS style queries using PMTiles
 - Yes.
- Q: Put something in references?
 - Key for PM Tile in dct_references.

- Q (paraphrasing): MapBoxTiles relationship
 - MapBoxTiles is a SQL database. Not good for tile access.
- <https://github.com/felt/tippecanoe> : linked from PM Tiles recipes section.
 - “Builds [vector tilesets](#) from large (or small) collections of [GeoJSON](#), [FlatGeobuf](#), or [CSV](#) features, [like these](#).”
 - Can use feature flags for how it generates different zoom levels
 - Upload shape file which we then use this software for
 - Q: Are your transform scripts online?
 - Yes.
- Main point: should be cheap to serve data that doesn’t change. PMTiles examples: open street map world, census tracts, etc.
 - Very attractive for news organizations. Don’t need to run a server. Optimized for storage
 - Don’t need complicated server side software - formats in a way that just runs
 - Not necessarily created to be a distribution software - but rather low cost web service
- Can symbolize - has default symbols.
- Geo data and attribute data - WMS may allow access and integration into other tools (ArcMap?)
 - COGs may not be able to support that kind of functionality
- Q: Can we have COGs to expose that endpoint (to allow that to be used elsewhere)
 - Yes.
- Princeton focus was on data preview. GeoServer was challenging - and not enough time to fix it all. This was a great solution.
- Can put as many attributes for a feature - but that also makes the dataset bigger
- Example:

<https://protomaps.github.io/PMTiles/?url=https%3A%2F%2Fdata.source.coop%2Fprotomaps%2Fopenstreetmap%2Ftiles%2Fv3.pmtiles#map=0.75/0/0>



-
- Q: GeoBlacklight uses Leaflet. Does Leaflet support PMTiles?
 - No.
 - Princeton GBL: Uses OpenLayers instead for PMTile records specifically.
 - JavaScript handles this
 - Part of JavaScript upgrade would be to assess whether we want to stay with Leaflet or move to something like OpenLayers
 - Currently app is very tied in with Leaflet
 - Leaflet is more lightweight and used more widely. Simpler. Uses plugin model.
 - OpenLayers is a more swiss army knife: handles everything. Less of a plugin model
 - Does appear to have some support? <https://docs.protomaps.com/pmtiles/leaflet>
- Q: Which version of PMTiles?
 - Not sure.
 - 3 seems to have additional features/options
- Eliot suggests just trying things out - creating some COGs/PMTiles - linking them to GeoBlacklight.

GeoTIFF compression

- For Melinda's specific case, COGs won't work
 - Migrated from DSpace 5 to 7
 - Repository is a generalist repository. Not specific for geodata.
 - Get all sorts of files.
- Repo: Overall limitations on file sizes. Very large sizes may not be downloadable.
 - Examples
 - Submissions that are very high resolution imagery

- Vegetation index for entire state of Minnesota
 - People trying to get around file size issues using lossless compression
- Impression was we do not want compression for long term storage
- General question: When is compression acceptable? Appropriate? Better or worse compression formats?
 - Also applies to COGs - can also compression for COGs when creating them
- Archive: Using lossless/balanced compression for images.
- Globus: May not help here since requires endpoints at both ends, and would not be possible for patrons.
- If lossless, shouldn't be an issue.
 - When creating COGs, Princeton stores the original image and then creates the derivatives (like COGs)
- Do compression formats change frequently?
 - LZW is ancient - as old as GIFs.
 - Any lossless compression format: can uncompress if need be.
- Lossy compression: changes resolution of file to save space. Lossless: doesn't reduce resolution as part of compression.
- Example of lossy compression use for historical maps
 - Not storing raw elevation values, or aerial imagery. Storing image.
 - JPEG compression most common lossy algorithm.
 - Lossy compression may make sense for some applications/use cases
 - Blog post links:
 - From Koko Alberti:
 - <https://kokoalberti.com/articles/geotiff-compression-optimization-guide/>
 - Has detailed breakdown of different lossless algorithms, write-speed vs. read-speed, LZW, Deflate, etc.
 - From Paul Ramsey
 - <http://blog.cleverelephant.ca/2015/02/geotiff-compression-for-dummies.html>
 - Good comparison of JPEG compression and overview of GDAL support for COGs
-
- With any physical medium, have to make choices around storage as time passes. May have to “pick your poison” whatever approach you take.
- With lossy format - a small change can corrupt file. Fixity may be an issue.

- Lossy for distribution only
- Institutional policy: Do not want file above a certain size - but don't compress. Becomes problematic for people.
 - Best practices around digital preservation around data
- [DPC technology watch report on preserving Geospatial](#)
 - See section 4.9 on compression.
- Q: What if file is too large even after compression? Split data/raster?
 - Have suggested splitting for certain files
 - Trying to understand which options will work in general
 - Broader question for repository with respect to file size limitation. What can we handle? Under what circumstances could we exceed limitation?
- Even with respect to splitting files, not everyone will understand how to do this effectively (or in general)
- HPSS: is it a robot that picks up magnetic tape and loads it? (perhaps).
 - When a project is being closed out, Petabytes of data will need to be stored. Goes onto magnetic storage. Very reliable, slow to degrade, stable. And cheaper as time goes on.
 - https://en.wikipedia.org/wiki/High_Performance_Storage_System
- Q: Where does this repository file size limitation come from? Technological?
 - Mix. Realistically most people can't download very (very) large files
 - Overall: Policy. Everything is backed up three times
- Princeton: Working on how to support faculty/students in accessing Globus
- Make a standard/policy around image upload
 - Image uploaded - that is itself not stored but rather compressed version
 - Expose link in a way where any viewer would be able to display it

Heard during this session multiple times:

“Not a librarian”

“Librarian, not an archivist”

GIS metadata remediation - session was moved to a small discussion next week

Mainly talking about “MLInfomap” datasets issue and remediation workflow.

E.g. [Stanford record](#), [NYU record](#)

Roles:

- Timekeeper:
- Gatekeeper:
- Notetaker:
- Facilitator(s): Zehong

Strategies for Collecting & Archiving Historical Geodata / Polar Geospatial Center Archives

Roles:

- Timekeeper: Kim
- Gatekeeper: Alex
- Notetaker: Melinda
- Facilitator(s): Karen / Kelsey

Overview of the Polar Geospatial Center (<https://www.pgc.umn.edu/>) archives and holdings

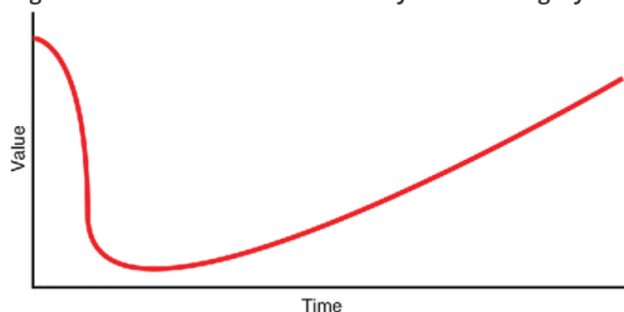
- Most of their data is satellite imagery (many petabytes)
- Use tape storage to keep this
- DEMs and satellite imagery available through a (new-ish) custom tool - [FRIDGE](#)
 - Includes access to a STAC catalog item
 - JSON files are stored in an S3 Bucket
 - Can log-in to access commercial data viewer
 - Must have federal funding (from NASA, NSF) with a backend managing access
- Map Catalog (hasn't been updated since 2019)
 - Maps primarily obtained from other institutions
 - Interest in making these more interoperable, closer connection to Libraries that they originated from
 - U of MN Libraries scanned these originally
- Aerial Photography
 - LIDAR campaign 2014-2015 - finally published May 2024
 - Also Single Frame Photos (1946-200)
 - Students hired to georeference these this summer
 - Hoping to add more functionality around these items - interactive viewers etc

- Discussion:
 - Capacity issues - a lot of ideas and initiatives that start but then aren't sustained. Is this
 - Downloadable file index for LIDAR files current
 - WI does something similar but with OpenIndexMap
 - Are people using Globus?
 - Stanford (it's a lifesaver)
 - PGC for certain workflows - but still using FTP transfer for some types of aerial imagery
 - Currently their data is organized by resource types - is there a long-term goal to unify these types?
 - Yes - planning to do a large web revamp
 - Are there other access points?
 - Sharing data in: Google Earth Engine, Esri Living Atlas, Open Topography
 - Can construct time series on Google Earth Engine - any plans to incorporate that functionality on PGC website?
 - Would be nice
 - Google Earth Engine/Living Atlas - using older versions of the data
 - PGC has data that is considered a public good hosted for free in Amazons S3 Bucket
 - Claire Porter organizing
 - [Open Data Sponsorship Program](#)
 - Saves a lot of money but concerns about whether the program will persist so data is also backed up on tape
 - Primarily storing the DEMs
- Interest in building out data time series
 - What are different ways of gathering historical data and presenting it in a way that people can understand?
 - Data providers sometimes don't want to maintain older data because of concern people will use it when there are more recent layers
 - OpenTopography - new tool for change in time between two images
 - How much do things change every year?
 - With remotely sensed data, Google Earth Engine is great for creating image collection - created graphs of spectral images → how could we make that easier for users?
 - What kinds of things is BTAA interested in collecting?
 - In development

- Probably starting with foundational layers
- Not a focus on research data (already in repo)
- Does WI have a mandate for sharing data?
 - Only parcels are mandated by Dept of Administration, but since 2017 one request for ten data layers
 - Argument is that counties are getting funding, so therefore should be made public
 - State cartographer office compiles and archives it
- NC Emergency Management was collating information, but stopped (making/sharing?) in 2020
- UC / Stanford Common Knowledge Group - try to coordinate
 - [split up the counties between UC institutions + Stanford](#)
 - [Map and air photo collections by county/map of collection responsibility](#)
 - Once a year they download specific foundational data layers (if available)
 - Still some counties that ask you to write in - may not respond or ask to pay
 - Wish they had more of a relationship - not just passively harvesting data online (and then rehosting which feels awkward)
 - WI - it was a long process to build relationships and piecemeal at the start
 - Geodatabase with several layers - lots of processing splitting it up and metadata (~400 layers in 2024)
 - Don't need to resubmit it if doesn't change (Except PLSS Corners, Roads, Address Points, Parcels)
 - Can search for historical data in Geodata@wisc.edu
 - In future, will there need to be collection records to organize the many years?
 - Serving as Geodatabases - concerned about possible data loss if changing to other formats (esp shapefile)
 - Geopackage would be possible - but the metadata doesn't come with
 - Could Aardvark metadata and/or "Metadata" link be bundled within the download? [Yes!]
 - Might be harder to update
 - When collecting data from ArcGIS Hub, it's hard to tell when things have changed (update dates can be unreliable) and a lot of junk files
 - UC - quickly had to reduce scope to a few layers
 - BTAA effort - we talk about doing foundational layers many of which are published by counties, but some states have very few layers

- Should we focus on state produced data
- When do we start filling in the gap
 - (coverage/e00 can no longer be opened in ArcGIS Pro)
- Who is telling us to do this [we are just taking personal initiative]
 - Digital object storage for now, not saying “Archiving” in front of administration
 - Went to the Library Deans 6 years ago, turned us down because each institution has their own repository
 - New set of Library Deans who gave us the go ahead, so we are moving forward quickly!
 - Have tried to set up stuff within Minnesota (have lots of reports about the possibility) but no funding to do it
 - We are trying to get a pilot /proof of concept going that we then could shop around to different states about what it could look like
 - MN took a lot of imagery data from the State recently
 - The data was being stored on harddrives
 - Some of the drives failed to load (don’t know what was lost)
 - Some of the drives can only be opened on Windows7 (and the state no longer has a device that can open it)
 - Now they are trying to see whether the Libraries will keep the data so they can get rid of the harddrives [not without a deeper conversation]
 - “Archive” is scary for administrators, so starting lightweight
- A lot of counties have data but not necessarily available, people don’t know that they can ask for it
- Have to be playing to long game - data will become valuable in the future

Figure 1 Perceived value of remotely sensed imagery over time.



(Dyke et al.

<https://doi.org/10.1080/15420353.2015.1073655>)

- Complexity of licensing/permissions:
 - BTAA would like to talk directly to the data providers
 - Stanford - if it's online it is open/public so they just take it
 - Generally people have been okay with it
 - Under-resourced staff may prefer not having to manage relationships
 - How do we avoid duplicating data that is still available for download
 - WI geoportal - it is a snapshot in time, so no way to be sure if it aligns with the most recent data on the individual portals
- Open data sites are so difficult to use... also a lot of custom county sites
- Hardest piece is building relationships and acquisition piece
 - More trust if they see it in action rather than abstract concepts

NOTES NOTES NOTES

OpenIndexMaps

Roles:

- Timekeeper: Maura
- Gatekeeper: Kevin
- Notetaker: Maggie
- Facilitator(s): Stephen & Tom

[OpenIndexMaps.org](https://openindexmaps.org)

[Workshop from WAML 2023](#)—presentation by Tom and Stephen

Example from Owen's presentation yesterday: <https://geo.btaa.org/catalog/p16022coll247:83-index>

Changes to GeoBlacklight from 2024 Winter Sprint:

<https://github.com/geoblacklight/geoblacklight/pull/1333>

- 11am: overview
- Index map: if there's a map series, the creator will put a grid over the larger area with each rectangle/polygon being an individual map in the larger series
 - Can also be point-based (popular for air photos) where there's an annotated point on a map that marks the center point of the image
 - Can also be lines showing air photo flights paths
- Index maps are hard to read and we tend to have lots of them
- Goal is to change these index maps into a GIS format to make them more readable and usable
- See linked article above for how to create a GIS format index map
- OpenIndexMaps is born and raised from Geo4Lib 2017
 - What if we all made these GIS index maps the same way?
 - A community standard for encoding an index map with details and access information that was completed in 2021 after 4 years of work
- All GeoJSON format
- Few requirements, many recommendations
- OpenIndexMap is a minimum specification, but you can add as much more information as you'd like
- Element names are the attributes in the header that are clearly defined in OIM
- All of this metadata is for a single frame (map or aerial image)
- We haven't yet standardized description at the set level
- There's a tool to convert from shapefiles to GeoJSON
- One component is to get these index maps to work in GeoBlacklight
- Standards were created based on an ideal situation rather than current technology
 - If you have a map with 3 editions, you have three overlapping polygons which describes each of the three editions as opposed to one polygon full of overlapping information
 - GeoBlacklight still can't handle overlapping polygons, but we think it will get there
- How do we share these out?
 - Github site [add link]
 - You don't have to recreate the wheel—many of us have the same sets of maps
- Chose GeoJSON because it's lightweight and interoperable as well as human-readable
 - You can visually verify and modify these index maps
 - As long as you have a valid GeoJSON file, you meet that part of the requirement and can spend your time actually populating the metadata
- Think about precision when entering coordinates—don't need 7 digits of decimals as that's about 1 cm of precision

- Shapefiles are more limiting than GeoJSONs
- GeoJSON.io will help you validate, edit, and convert formats
- GeoJSONLint validates and previews GeoJSON
- GeoBlacklight 2024 winter sprint–modified some of the code base
 - Some field names changed
 - The examples in the repository used to point out to external github repositories, so they are now baked in
- Some questions that remain to be answered
 - What are the default fields we want to show in a Geoblacklight instance?
 - How do folks customize what fields show?
- 11:20am–moving to questions and discussion
- Question: why can't GeoBlacklight handle overlapping polygons?
 - It's a leaflet problem but not sure about more info beyond that
 - Maybe move to an open layers viewer in the future to resolve this?
- Q: Have the element/field names changed a bit or have they been solid since day 1?
 - Solid since day 1
 - Will ID as a separate version if the standards change
 - When GBL codebase updates happened, legacy field names that predate the standard are sorta backwards compatible but you should update to element names to the standard as that's not a guarantee
 - Stanford, Cornell, UWM have been using the old one
- Comment: The question of describing sets is important. Particularly air photo sets with project numbers, county codes. Kevin Dyke is interested in talking about this.
 - Tom realized when they started doing this for air photos that they never included the flight number. Might be in the future. Set levels/flight numbers are tricky because do you want that same thing in every single row, or should it be higher level somewhere?
 - GeoJSON is a nested structure
 - A feature is an individual map or image, a featured collection is the set
 - You can also assign properties to the featured collection–maybe in the future we could start defining collection level properties
 - File naming?? Standardizing that in the future? Tricky
 - There are other use cases outside of GBL so can't rely on Solr information
- Other use cases–the UC system might have the same map sets but not all of the same sheets
 - We could ID what campus has what sheets in the OIM framework

- If a campus has multiple copies of certain sheets and someone is missing something, could use that information to share the extra copies to complete the incomplete sets
- Embedding collection level records in the GeoJSON keeps information in one place rather than splitting that information across multiple locations
 - Most people aren't reading the straight GeoJSON—how to make sure people know to look there before it's in whatever viewer the information is showing in
- Q: How specifically tied would a specific OIM file be to a specific Aardvark record?
 - When you put these into GBL, these are linked with the Aardvark record. However that is still split and if this is anywhere outside of GBL it gets tricky
 - Perhaps workflows could be developed to write a crosswalk that generates a Solr document out of OIM data or vice versa
- Apparently in GeoJSON specs, a featured collection cannot have additional properties?
 - More research needed
 - Can it be above the featured collection layer, as in in the header as opposed to the body?
 - It won't break anything but also won't be recognized outside of the GeoJSON record itself
 - Putting series level information into every individual feature feels redundant but??
 - Is there a GeoJSON community who can help develop this or otherwise help?
 - GeoJSON is stripping things down to be as simple as possible, but the featured collection is designed as a lightweight structure while the meat is in the features
 - To summarize (Huda)—if we could put more information into the index map specification itself, it would just live in the GeoJSON. In the GBL instance, it would point to this record and also read it to populate the GBL instance.
 - Turns out that properties are restricted to features in GeoJSON but if you call them something else it could be okay (no double dipping)
 - As long as we don't call it properties or any existing member names, we could make this work possible?
- Does an old school index map display in GBL where you can see it but then click into it to access the individual maps? As opposed to looking at it in only a digital form?
 - Do a COG of an individual map scan?
 - Good idea! Would be cool
 - Make it transparent w/ a slider?
 - People are interested in old index maps since they can be really cool looking
- Stephen worked on creating a GeoJSON schema file with field specifications and recommendations/requirements during the last GBL spring

- Important distinction: fields for north, south, east, and west as minimum bounding for things like complex polygons and lines
 - These serve more as an envelope and if you have a simple bounding box they'll be the same, but you might have more complex geometries inside of the envelope
- Q: We looked at a LIDAR index map earlier. How are people using geometries with flight line maps?
 - UWM builds out geometries
 - Real edge case but interesting
 - Air photos on a line get split into points
 - Individual points are probably best if you're a researcher, but GeoJSON could theoretically use any of the styles it supports
- Q: What's the roadmap for later?
 - Haven't had regular meetings for a few years
 - Busy learning about how it's working out for folks as they implement it
 - Set level metadata is probably the next step
 - GBL Slack has an OIM channel
 - Maybe we pre-plan some working time during the working days at a future Geo4Lib

Issue describing a possible set-level metadata implementation for OpenIndexMaps:

<https://github.com/OpenIndexMaps/openindexmaps.github.io/issues/43>

Georeferencing Historical Aerial Imagery / Georeferencing Interest Group / Handling complex content while mosaicking historical atlases

Roles:

- Timekeeper: Dave
- Gatekeeper: Zoe
- Notetaker: Ben Liebersohn
- Facilitator(s): Jim / Adam / Elliot

Jim Lacy

Context from Jim Lacy: Yesterday we talked about the Wisconsin Historic Aerial Imagery Finder.

When you are using the website, there are points representing the center of each image. When you open the image, you just open the .tiff, there is no viewer, and UW collections does not work with IIIF.

Today we are looking at a prototype of the new WHAI Finder, with the slider to adjust the years. The prototype has a bar on the left with results that you can use to preview a georeferenced image. Some of these are georeferenced such that they overlay really well. We use one point to georeference! No time for control points!

Some don't match as nicely - ballpark but not perfect. That's acceptable (as long as it's not at 0,0), so at some level it needs to be georeferenced, but it doesn't need to be perfect. When we look at a 1.8GB file, it loads pretty fast!

The flight lines are never perfectly straight - there's also yaw in the photos (no photos are perfect!) so we must accept some limitations. We look for the center of the photo, the scale, and place a centerpoint. It might not look great, but it's in the ballpark! The accuracy of the georeference output corresponds to the georeference index points. The process is quite simple - each record is grabbed, find the scale, scan resolution, find the NE, SW corners, and dump the plane coordinate system (UTM zone kinda) into a csv (a WTM).

So if we know the corner coordinates, GDAL can read it in and output a COG using `gdal_translate $infile tempfile.tif -stats -of COG -CD compress=lzw -a_ullr $ulx $uly ...` (ask Jim 😊)

Jim's dream plan: the Allmaps model: In the future it would be cool if we could get a slightly better georeferencing workflow, a transparency slider to compare images
When you hover over a point, a box appears to show the extent (area covered). If you click on a dot, all the results in that dot area are visible.

Projection problems were expected, but it seems ESRI basemap is being reprojected on the fly. Index points are WTM, the photos are WTM, the basemap is being warped. At this point, jim is in the "get it working, we will perfect it later" model.

Jim has not tried using the Orthomapping Wizard, as it doesn't require control points (it requires center point, metadata for flights, focal length). You can get the calibration report from USDA. The APFO or now USDA Geo should have this information. Jim's objective at this point is to get it sourced and get it out to people who might be able to do the full ortho - the audience at UW for this is geared towards public users (who might want a photo of their farm). If you used this Ortho Wizard it will automatically georeference and generate a bounding box!

Adam Cox: oldinsurancemaps.net

Handling complex content from Sanborn Maps built using svelte (JS) and django application. They look uniform but they have complexities, such as multiple maps on one page and other challenges. Context is that this started as a masters thesis and a passion project. People now use it for many new areas it wasn't intended for (it was about Louisiana) and now people are looking at other areas. Funded by NIH and UVA, who are the biggest users. 31000 is the number of sessions, not the number of layers. People use it at different institutions for different reasons. The purpose of this includes distribution.

Some want to use it for OCR, some want to use it for digital heritage. The application that facilitates this is the OHMG (online historical map georeferencer) (ohmg.dev).

It was messy to work with each individual sheet. The entry point is a simple map, a single scanned document. This is the simple case, but if you have a multi page atlas, you have multi page documents with possibly multiple layers each. The more complicated case is a multi-volume atlas, where the discovery takes place at the volume level.

Example: if we look at Neworleans, we see there are tens of maps that are georeferenced from 5 volumes, which are each part of a hierarchy, and the georeferencing is tracked at the volume level, so you don't search across all the sheets - you search at the volume level! This is important for the multimask, where you georeference and then trim. This is tedious at the volume level - instead, we create a single unified multimask using snapping for a seamless connection between layers. This allows masks to manage the visibility at the volume level (if content such as a street name is not visible, you can adjust the multimask to reveal the content in another sheet)

In the last 6 months, Adam has been working on making an abstraction level between the individual layers and the collection, which he calls annotation sets. Between the volume and the keymaps, you can have the ability to use keymaps, however the skeleton maps are not included in this grouping. Now there is a multimask not for a volume, but for a category or level of content. This is a more robust model because you can use one page or one category (simple) or more complex collections without leaving anything behind. People in the past have ignored the skeleton maps (uncolored maps) which may still be desirable to have.

The sheets can be turned into COGs with one layer volume (rather than 70 pages per volume) that is a downloadable GeoTIFF or xyz tiles. Xyz tiles are created using [TiTiler](https://titiler.org/), an open source project which is a dynamic tile server. This was fairly easy and way better than GeoServer. TiTiler can now generate WMS endpoints.

This can also be used to handle inset maps! There is more work to figure out how to a georeferencing spec from AllMaps. Adam is trying to keep the annotations clean until there's

somewhere to store them. Ryan asks about who pays for the server, which U of Michigan pays for. Mapwarper will just start deleting stuff eventually - is there a promise of longevity here? Adam says no promise, things are going well at the moment. Adam was inspired by the idea that interns were georeferencing small areas but they would be repeating their efforts. It was suggested to give the geotiffs or control points to Lib of Congress but it's unclear that they care. The infrastructure or process is the exciting part, not just the output product. The next step might be to use AllMaps plugins to display the geotiff from Lib of Congress IIIF images. Georeference annotation has a beautiful simplicity (pixel coordinates to geographic coordinates) and this is portable, but resolution dependent.

Allmaps as a discovery platform - how will it be used?

How do you choose to display? What do you prefer, and what do you do with insets? What are the logistics of deciding what sheet goes on top?

First, we get rid of inset maps so there's no confusion over what is the inset. Next, we display the mosaic - "here's the mosaic for the skeleton maps." There is no formal way for decide what to display. There are a lot of ways to decide how to handle that. There isn't a plan to add georeferencing information import. This wasn't designed with IIIF in mind, so importing directly isn't possible. There is a digital scholarship lab at UVA that's trying to extract text information specifically related to health information. Ghost Neighborhoods at OSU is looking into extracting footprints.

AI or API Enhanced Metadata, what can computers do for me?

Roles:

- Timekeeper: Amy
- Gatekeeper: Huda
- Notetaker: Kim
- Facilitator(s): Eric

Ways to enhance metadata for the Big 10 geoportal. Ai or API enhanced metadata.

LLM - Large Language Model ([IBM YouTube Video](#)). LLMs are composed of data, architecture, and training.

Instances of Foundation models applied to text. Trained on large datasets. LLMs are reading info across the internet. Bots can affect performance time on websites, applications.

Highlights of ChatGPT: writes python code, helps with troubleshooting, like talking to a colleague. Use for programming - working with older javascript.

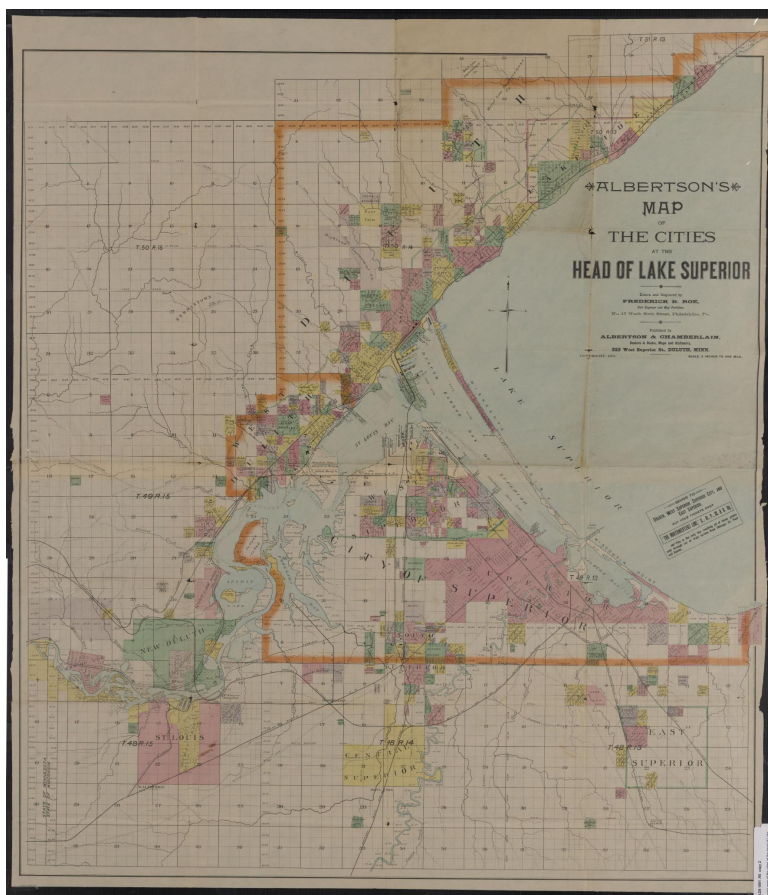
Limitations: sometimes it makes up stories, sensitive to how questions are phrased. Can exhibit biased behavior, or respond to harmful instructions.

Recent discussion at Code4Lib - using ChatGPT in cataloging (enhancing metadata, looking for inconsistencies, adding consistency). Adam Cox's Social determinants of Health Project at UIUC is using AI as a query enhancer.

[GitHub Co-pilot](#) plugin for [VS Code](#). Writing inline documentation for code.

How can we leverage ChatGPT to make geoportals better? ChatGPT4 can respond to image prompts. ChatGPT can look at an image and describe it, which can aid in accessibility issues. Useful for providing descriptions of resources lacking metadata. Can provide estimated digitized map bounding box coordinates. Can extract place names from maps. Can provide coordinates of place names or points of interest on maps. Will require evaluation - sometimes gives incorrect answers. Can provide some correct GeoNames identifiers from place names.

Example map image:



Prompt: "Describe this map in a short paragraph."

ChatGPT response: "The map titled "Albertson's Map of the Cities at the Head of Lake Superior" depicts the region surrounding the western end of Lake Superior, including the cities of Duluth, Minnesota, and Superior, Wisconsin. Created by Frederick B. Roe and published by Albertson & Chamberlain in 1887, the map features a detailed grid system, highlighting property divisions and land ownership. It includes railway lines, roads, and other significant geographical and man-made features. Various regions and townships are color-coded, indicating different administrative boundaries and land uses. The map serves as a historical document, providing insights into the urban and rural development of the Lake Superior region during the late 19th century."

This response could be an accessible text-based alternative to the ItemViewer media object in a GeoBlacklight application.

Prompt: "For the map I uploaded, please generate a list of placenames and geographic coordinates."

ChatGPT response: “Here's a list of placenames from the "Albertson's Map of the Cities at the Head of Lake Superior" and their approximate geographic coordinates:

Duluth, MN: (46.7833, -92.1066)

Superior, WI: (46.7208, -92.1041)

West Duluth, MN: (46.7487, -92.1807)

New Duluth, MN: (46.7083, -92.2007)

Fond du Lac, MN: (46.6533, -92.3110)

Rice's Point: (46.7640, -92.1067)

Smithville, MN: (46.6845, -92.2760)

Allouez, WI: (46.6810, -92.0404)

South Superior, WI: (46.6833, -92.0934)

East End, Duluth, MN: (46.7984, -92.0675)

Lake Superior: (47.7000, -87.5000) (general reference point)

St. Louis River: (46.7200, -92.2000) (general reference point along the river)

Spirit Lake: (46.7200, -92.2300) (general reference point for the lake along the river)

These coordinates are approximate and based on the modern-day locations overlaid onto the historical map.

Prompt: “Do you know GeoNames placename identifiers?”

ChatGPT response: “Yes, GeoNames provides a database of geographic information that includes placename identifiers, also known as GeoNames IDs or geonamelds. These unique identifiers are used to reference specific locations within the GeoNames database, which includes a wide range of information such as names, coordinates, and administrative divisions.

For the placenames extracted from the "Albertson's Map of the Cities at the Head of Lake Superior," I can provide their approximate GeoNames IDs where available:

Duluth, MN:

GeoNames ID: 5024719

Superior, WI:

GeoNames ID: 5275191

West Duluth, MN:

GeoNames ID: (part of Duluth)

New Duluth, MN:

GeoNames ID: (part of Duluth)

Fond du Lac, MN:

GeoNames ID: (part of Duluth)

Rice's Point:

GeoNames ID: (part of Duluth)

Smithville, MN:

GeoNames ID: (part of Duluth)

Allouez, WI:

GeoNames ID: 5244638

South Superior, WI:

GeoNames ID: (part of Superior)

East End, Duluth, MN:

GeoNames ID: (part of Duluth)

Lake Superior:

GeoNames ID: 5128638

St. Louis River:

GeoNames ID: 5043359

Spirit Lake:

GeoNames ID: 5048033

Please note that some locations, particularly smaller neighborhoods or regions within larger cities, may not have distinct GeoNames IDs and are often referenced as parts of the larger city.”

OpenAI provides API access to models. You pay by transaction. Very cheap.

Google Cloud Vision does a great job of spotting text. Can identify colors. Can output text, which can be used to build descriptions, add place names.

Machines Reading Maps - text on maps are now searchable, beyond place names

Outputs always need to be sanity-checked.

Would AL generated metadata be added to Aardvark records? Or integrated somewhere else?

How does this affect the search experience? Could potentially overwhelm the search experience.

This technology is still very new but will likely be standard library practice.

OpenGeoMetadata / Metadata discussion (aardvark field for collection string)

Roles:

- Timekeeper: Eliot
- Gatekeeper: Dave
- Notetaker: Maura
- Facilitator(s): Huda / Amy

NOTES

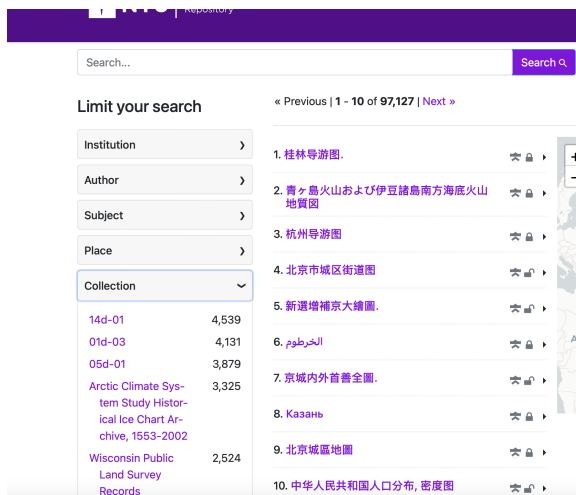
OpenGeoMetadata:

- Fever dream of Eliot and Jack Reed ‘what if we had all the geodata in one place?’
 - Talking about metadata, what if we put it all in a git repo to share with each other, and that’s what we did!
 - Then we had geocombine tool which is for harvesting that data
- Many institutions use this for everyone’s data to be indexed so we can all show things at other institutions
 - Stanford stopped putting iso stuff up
- Maybe AI/ChatGPT can do stuff for pulling out metadata
- 2 questions:
 - With the way things are now, can we think of other use cases for combined data?
 - Could data be used to cross-check? If 2 institutions hold the same thing how do we keep from double-harvesting things that aren’t unique.
 - Each university makes their own metadata and identifiers but sometimes it’s the same dataset we all bought. Example: [PolicyMap \(BTAA\)](#)
 - Would be nice to attach holdings, one bucket to rule them all (geoworldcat)
 - We don’t have a TON of duplication
 - Distributed but form one thematic collection (Uni 1 has 5 maps from a series, and uni 2 has 6 maps)
 - Parcel data? If we aren’t using a traditional boundary (New Mexico, Arizona), so parcels for a region. Spread out, one university is doing Arizona, one might be doing NM, but a researcher would like the parcel data aggregated by region Southwest
 - Are you ingesting data from other institutions? Hope is yes, UC is starting with the California schools and then hoping to expand
 - Need to be consistent in vocabularies
 - What happens when we delete records? How do we let others know?
 - Some people are converting gbl1.0 to aardvark because they can’t index other metadata
 - Otherwise have to periodically purge and then reindex a university’s metadata
 - Could GeoCombine do anything to help? Make a list of things?
 - BTAA does a flag for publishing but that isn’t in aardvark for everyone to act on
 - Could be added to aardvark, this is easy to filter on
 - Deleting was messing up the analytics
 - Sometimes they come back

- Could be setting up a problem with deletion? Mark as withdrawn?
- Include a flag for reharvesting?
- Just use a cron job? Purge x school every 1st Monday
- Use git diff? Then the git can make lists
- Are folks using permalinks? Marking withdrawn? Do both?
 - If you do big updates, git can really be pushed to the limit with diff
 - Princeton does some discarding of history to keep things smaller
 - Use git for a sharing source, not version control
- Could we harvest from GCat, other things? And then those go here?
 - The problem is 10% turn over every month so you will end up with lots of broken links
- What else might we do that we haven't done yet that would be useful? (kind of answered above)

Aardvark Changes:

- Member_of in gbl1 was a string, but aardvark this is an id
 - Example: NYU search, look at collections and you see ids not strings



- If we stay with IDs over strings, do we keep the Collection filter? No, probably not. At some point GBL1 added source, then when we went to aardvark we used IDs for everything. Now BTAA does Resource Class – Collection and then there's collection records being used by Member Of, and Is Part of ties to individual smaller source websites.
- Eric came up with a way with a widget to other items in collection (not as intuitive currently)
- Could we just pull the title connected to the id in the filter? Technically yes but it would be a lot of solr overhead (or making maps, see language)

- Do we want to just add a new string field to tie this together? This has come up a lot so it seems we should do it, people don't want to make a solr document for just collection it seems?
 - Easier to crosswalk gbl -> aardvark with this
 - Maybe less overhead for users to find collections
 - The downside, if the collection is its own thing, you don't have to change it in every record, you just change it in the collection level its all connected to
 - Also will it get confusing if we support both?
- Cornell/UW create general collections ('coastal layers')
- Member of -> designed by the institution?, is part of -> an intellectual thing (like coastal)?
 - Definition is different
 - Aardvark uses is part of like part of this same book, or website
 - Member of is something we create a collection and add it
- Looking at Princeton's interface, Sanborn maps, defined a search with Solr (canned search) NOT a collection label
- Some times folks want the collection record AND results, some folks just want the results.
 - So vector data you'd want the collection record to show with the individual results but maybe geotiffs you dont care about that collection record
 - Using related so you can always get back to the collection

What is LandScan? / Using STAC (Spatio Temporal Asset Catalogs)

Roles:

- Timekeeper: Eric
- Gatekeeper: Kelsey
- Notetaker: Adam
- Facilitator(s): Ben / Danny

NOTES NOTES NOTES

STAC

- STAC is a way to describe and catalog data
 - Esp. for raster data, developed by satellite imagery providers
- <https://stacindex.org> has a list of available stac endpoints
- JSON document that describes spatial and temporal aspects of a dataset

- Extensions are used to add extra information, like cloud cover, etc.
- Not very suitable for vector
- There are Collections, Catalogs, and Items, different levels of content (nestable)
- Easy to create, low barrier to entry
- Often stored as JSON files in S3
- There is a STAC API is a server you can run to create a queryable API on top of Postgres or other database endpoints
- Does this serve the same purpose as CWS??
- Question: Who manages the spec and website?
 - Radiant Earth: radiant.earth
 - They manage the catalog website
- Question: Do people have users or customers that want API access to your content?
 - It is more common now that APIs are expected offerings for content in collections
 - If you have an API that can expose and deliver for people, you are multiplying the access abilities
- Gatekeeping check-in:
 - What is JSON?
 - What is API?
 - Programmatic way of interacting with software or
 - Are there free APIs?
 - Example is Library of Congress JSON API
 - Or, National Weather Service has a free API

Land Scan

<https://landscan.ornl.gov/> Explore here!

- Question: Why does Ben care so much about an obscure data source?
 - He worked on it for two years and not enough people know about it!
- Showing examples of the data first
 - Syria
 - Raster data, single band, 90m resolution
 - Population density in Syria
 - Population in the data in areas that are outside of Google-marked place names
 - There are clusters out in the desert on the Syrian border
 - Drone strike on Tower 22 in Jordan by the border, and then a cluster of people just across the border in Syria

- Many people have fled from Syria and Aleppo, and they are people and structures in this cluster
 - Rokban Camp
 - Example of use: Select area of interest around camp, and use Zonal Statistics in QGIS to calculate the total population
 - Important to have this information when allocating aid (number of tents, etc) to remote areas of need
 - Many countries are available for download, generally the US, and then areas for war, natural disasters
- Minneapolis
 - Land Scan will have day-time and night-time population
 - Calculate Zonal Statistics by county,
 - Ben calculates 1.26 million at night
 - Does this equal the Census number???
 - Yes!
 - Ben calculates 1.5 million during the day
 - Important! A large increase
 - Example of why this is useful, you could use central points for highest density to help with network analysis
 - Clear difference in downtown when comparing day-time (large population) vs. night-time
- HD datasets are built from place-specific models.
 - Human geographers are used to interpret how many people live in what kinds of structures in a particular region?
- Development of LandScan HD
 - Created on very
- LandScan Global
 - 30 arc seconds
 - Distributed by Eastview Geospatial for a long time
 - Classic distribution
- LandScan USA
 - Has day/night
- LandScan HD (ongoing development)
- An example of a tricky model for a model to recognize
 - When density changes along urban areas

- Port Said
 - Combining densely population areas with less dense
- Settlement Mapping Tool
 - Amy Rose
 - Urban Dynamics Institute at Oak Ridge
 - Detecting settlements around the world
 - Uses **edge-detection** to find and classify settlements
 - The process is super efficient and accurate (93%)
 - Example:
 - Different types of settlement patterns in Albuquerque correspond to socio-economic status
 - Question: Can this model detect abandoned settlements?
 - Yes
 - Example:
 - Damascus, clearly different settlement patterns in different parts of town.
 -
- Population Density Tables
- LandCast
- There is a QGIS plugin to access
- There is an API as well, though it's new and may be hard to register for!

Ethical description, but not quite: advisory warnings for maps with potentially harmful content

Roles:

- Timekeeper: Zoe
- Gatekeeper:
- Notetaker: Maggie
- Facilitator(s): Maggie

NOTES NOTES NOTES

- Discussion about recent MAPS-L listserv content warnings email and response
 - [insert text]
 - Things are much different when you have open stacks
 - Both/and thinking

- Plus digital collections that are publicly available is another issue
 - Maps with Nazi stamps at Stanford
- What is sustainable? Can't update individual records
- B1G 10 approach: recently looked for outdated language that could be updated
 - They have a warning in the BTAA geoportal: <https://gin.btaa.org/library/harmful-language/>
- Stanford has library-wide statement and an anti-racist description ground
 - Lot of discussion about what level the description needs to be present
 - Rare and antiquarian map materials it's basically everything
 - Tentative plan: if you're doing MARC language they have a content advice note at the individual or collection level [insert text]
 - Calls it "content advice" as many were reticent about "content warning" or "trigger warning"
- Los Angeles map with the N-word triggered Maggie to begin this work
- UMN had atlases that referenced the Sea of Japan and asked Ryan to cross all of that out to replace with the East Sea
 - We obviously can't literally censor materials
- Rumsey site lets you text search OCR maps for basically any slur but the N-word
 - Eventually someone is going to notice this and yikes
- It's still important to have this as researching the history of place names, etc. and to continue making it available, but how to give a heads up?
- Ask forgiveness not permission—Stanford metadata
- Worries about getting things approved and jumping through so many hoops and having to get approval from people who don't know about maps
- Big admin DEI talk but actions and behavior is different
- Example MARC using 520 field for content advice: <https://searchworks.stanford.edu/view/13872224>
- Example of content advice in an archival finding aid: https://oac.cdlib.org/findaid/ark:/13030/kt2b69r8fw/entire_text/
 - Material Specific Details: Content warning: This collection includes racist depictions of the "Stanford Indian" mascot. For more information about the history of the "Stanford Indian" mascot, including the campaign of the Stanford American Indian Organization which led to its removal and disavowal in 1972, please see <https://nacc.stanford.edu/about-nacc/history-timelines/stanford-mascot-timeline/removal-indian-mascot-stanford>. Stanford Libraries collects and makes these materials available to facilitate scholarly research and education, and does not endorse the

viewpoints within. Our collections may contain language, images, or content that are offensive or harmful.

- Statement on Stanford Special Collections' policy page:
https://drive.google.com/file/d/1-2U14_QKT3N8FnAgmPK50YhEloCIzVDG/view
- Colorado State statement:
<https://lib.colostate.edu/find/archives-special-collections/describing-historical-materials/>
- Not trying to reinvent the wheel but how to contextualize in a geospatial context
- How to incorporate learning resources instead of only a statement
- So few resources that are targeted at geospatial resources in particular, don't want only general primary source information
- The shift in land acknowledgements
- Google maps almost triggering military conflicts -
<https://www.theatlantic.com/technology/archive/2010/11/google-maps-nearly-starts-a-war/66451/>
- Laura doing some reparative cataloging and highlighting other cultures involved in that region that may not be named on the item or in the original record
- BTAA works with metadata remediation
- Long? Short? Lots of resources? How to fit it on signage? Knowing your audience? How to modify or adapt for a digital platform (Geoportal, Digital library)
- Who is emphasizing metadata versus content?
- Stanford and machine-readable maps—what is searchable?
- UMN talks about harmful content in digital repositories but doesn't acknowledge physical collections - <https://www.lib.umn.edu/collections/harmful-content-in-digital>
 - Except for archives (<https://www.lib.umn.edu/collections/special/historical-language>) but not the “regular” collections
- Blanket popup in the viewer? Like instagram?
- Including resources for library workers who have been harmed by seeing some hurtful material—counseling resources, etc. (such as spending all day cataloging Nazi maps)
- Internal versus external documentation
- Stanford internal documentation for descriptive/reparative content:
 - “As you work with difficult material, you may find that not only is the descriptive language about the material potentially harmful, but that the material itself is potentially harmful—to users and to staff. When encountering difficult content, your first responsibility is to take care of your own emotional and mental well being. If you find yourself struggling with the material you're working on, take breaks, go for a walk, engage with less challenging material for a bit, consult with colleagues—whatever works

best for you in the situation. Members of the Anti-Racist Description Discussion Group are available via the Slack channel #anti-racist-description or individual contacts to consult and commiserate.”

- How are things labeled? Sorting and shifting—even top of one stack and bottom of the previous to split things up
- Leverage the inconveniences of your physical space—Zoe
- Library of Congress G series with bible lands, Israel, Palestine
 - Can we modify these internally?

Conference Plus / Delta

Plus

- + Great venue
- + Cavern tour
- + Catered lunch together
- + Good food choices
- + Great keynote
- + Evening social time
- + Website thorough and registration process was easy
- + Enjoyed it being in-person
- + Time of year (i.e. spring quarter)
- + Great lightning talks (seemed to work okay with Zoom participants as well)
- + Polite, respectful discussions - open to questions and explaining jargon

Delta

- Less rain :)
- Having set times for when happy hour will be (with possible breaks in between)
- If possible, do the breakouts in smaller rooms
 - Awkward for the facilitators to go to the podium
 - Was harder to hear
- An additional step for brainstorming unconference pitches
 - Regroup halfway through
 - Options to continue conversations after the first day of unconference sessions?
- Putting sessions together before the voting
- Could be some spontaneous lightning talks (while people can also still plan workshops in advance)

- Geoblacklight workshop
 - Stressful to have installed all of the software pieces but then not use the tools
 - Having geoblacklight slack ahead of the workshop to get help
 - Pre-conference option for setting things up (or later in conference)
 - Mon. setting up GeoBlacklight, conference Tues-Thurs
 - Trying to do both conceptual (what can it offer you) and technical combined was hard...could it be done separately?
- All workshops were very GeoBlacklight focused - would be nice to have other options
- More questions in registration about what people are interested in learning about
 - Or a virtual preconference brainstorming on a jamboard equivalent
-

Next year ideas

- Workshops Monday afternoon (with optional help installing software Monday morning or maybe a zoom help install session Friday?)
- Main conference Tuesday-Thursday
- Optional working session Friday

Have two tracks for evening socializing

- Happy hour track
- Cultural outing track

Set sign up deadline a month out for T-shirt and catering orders

Thursday, May 23, 2024

Notetaker: Karen

Timekeeper: Eric

Gatekeeper: Stephen

Impromptu session: OpenGeometadata API registry

- desc.

- Creating a registry of OGM,

- automatically enhancing OGM metadata with gazetteers, AI descriptions, other normalization activities; this would add value
- serving it back out as a web service API

Possible tools:

- [Who's on First?](#) -

- an enhanced gazetteer of place names
- hierarchical
- geonames, TGN crosswalks
- includes bounding boxes, boundary (geometry) outlines

Proposing essentially a GeoOCLC: a central repository of metadata

Benefits

- more participation
- better consistency between metadata
- single unified service point
- share developer time
- front-end agnostic - can use any front-end discovery interface
- centralized cost?
- ease metadata remediation in the future? a centralized service creates the “Aardvark” for everyone

What would the API deliver?

- an aardvark **metadata** record
- potentially additional info, such as AI, OCR text, machines reading maps, etc

Subscribing to “events” that alerts people when a change has been made and triggers a notification that change needs to be made in local repositories

Where does GeoCombine fit in?

Where does a single geoportal to rule them all fit in?

This could be like a version of OCLC or a metadata registry just for geospatial metadata.

Incorporating a way of remediating each others metadata;

“Hub record” - layer from an open data portal, most commonly ArcGIS Hub; there is so much data on this hub, but it is continuously changing; This is a recurring theme at Geo4Lib- what are we supposed to be doing about these records? Continue to point to the live services ala BTAA Geoportal?

Best Practices for Authoritative Data Providers

We would create a profile or way for people to opt in to certain collections.

Could GeoMonitor be helpful?: (tool that monitors GeoServer to make sure that it is running and available)

Other tools for querying and monitoring external server, like ArcGIS Hubs, etc.

Include a transactional history, including finding records that were withdrawn

In the short term, what is the goal and focus?

- are we proposing that we stop focusing on GeoBlacklight?
- there is limited GeoBlacklight developer time already
- we may not have the capacity to set this up yet.
- however, software changes; how will we persist our metadata records?
- archive data, functions in proprietary services, such as StoryMaps
- in the “early days” of GBL, the focus was on getting the software up and running; the past few years, it has been more on the metadata side
- Would we need a server to host it? Who pays?
- Do we need to do a SWOT analysis? How would we even start?
- How to share costs? ex. IIIF, Samvera, umbrella nonprofits?
- Consortium for Public Geography
 - part of the Digital humanities grant by LMEC and AGSL (Contact is [Emily Bowe](#)) have proposed to take community ownership of the allmaps, maps kurator; could it be an organized consortium that includes geoblacklight, opengeometadata, a nonprofit a board; dedicated time; a small budget; 2 more years on the grant;
 - who “hosts” this?
 - payments need to go to a structured official entity
 - allmaps
 - could it include metadata as well?
 - how would the user community support it?
- What about about the for-profit community? i.e. East View that charges for cataloging services

- Can we move beyond just GeoBlacklight?
- Grant proposal? How about just starting with a supported slack account?
- allmaps needed big institutional support; utilize local support for grant applications;
- costs: paying developer time, server costs, etc.

This is such a big direction. Can we try these projects without having funding and larger organization first?

What can we experiment with in the meantime? What are the logistics? Is this a community goal? To discuss at the next GeoBlacklight community call? Ask for volunteers to create an MVP?

Ideas for GeoElevator. OpenGeoMetadata API endpoint.

Working Groups to form?

- governance
- technology

Next steps:

1-week research spike for technical & use case aspects

- gather use cases, ideas, sketch out a white paper
- technical details
- fall semester goal?

Piggy back on consortium for public geography

- governance

Send out GeoBlacklight Sprint signup

Timeline 2 years of planning?

NSF - Pathways to Enable Open-Source Ecosystems (POSE)

<https://new.nsf.gov/funding/opportunities/pathways-enable-open-source-ecosystems-pose>

Friday, May 24, 2024