Bit.ly/geo4libcamp2020-notes
Geo4LibCamp 2020
Unconference Notes

Schedule	Hartley	Branner (upstairs) - Teaching Corner	Branner (upstairs) - Map Corner
Monday Workshops	Making GeoJSON Index Maps with QGIS https://kgjenkins.github.io/openindexmaps-workshop/	Geospatial Metadata https://z.umn.edu/geo20 bit.ly/2Uk4hC6	Open Index Maps
Tues 2:40-3:40 pm	[Geoserver](#) (Keith) Notetaker: Reina TimeKeeper: Kim Durante Gatekeeper: Andrew B.	[Open Formats for Preservation](#) (Susan Powell) 14 people attended Notetaker: Kasey Wilson TimeKeeper: Eliot Gatekeeper: Chris Thiry	[Dataverse Integration](#) (Mara & Evan) Notetaker: Eric TimeKeeper: Dave M Gatekeeper:Paul
Tues 3:45 pm-4:45 pm	[Geometadata Manager](#) (Eric & Karen) Notetaker: TimeKeeper: Jack Gatekeeper: Eliot	GBL Usability () Notetaker: TimeKeeper: Gatekeeper:	GBL 2.0 Upgrades (??) Notetaker: TimeKeeper: Gatekeeper:

Wed 9:00 - 10:30 am	IIIF Introduction (Jack) Notetaker: TimeKeeper: Gatekeeper:	OpenIndexMaps (Chris) Notetaker: TimeKeeper: Gatekeeper:	
Wed 11:00 am - 12:00 pm	Data extraction from scanned paper maps (Kasey) Notetaker: Huda TimeKeeper: Eliot Gatekeeper:	OpenGeoMetadata (Melinda) Notetaker: Eric TimeKeeper: Gatekeeper:Reina	Geocoding (Stace) Notetaker: Amy TimeKeeper: Gatekeeper:
Governance and Road Map			
Wed 1:30 pm - 3:00 pm	Relationships are hard (between data) (Ryan, Mara & Keith) Notetaker: Kara TimeKeeper: Keith Gatekeeper: Reina	Geo Repos & Where's the data? (Kean) Notetaker: James TimeKeeper: Gatekeeper:	Analytics (Phil) Notetaker: TimeKeeper: Gatekeeper:
Wed 3:30 - 4:30 pm	Building Data Ingest relationships with GOV/Esri Open Data sites (Jim L) Notetaker: Kim TimeKeeper: Mara Gatekeeper: Amy	Cloud-Optimized GeoTIFF (COG) (Eliot) Notetaker: Keith TimeKeeper: Gatekeeper:	GBL DevOps & Support (Reina) Notetaker: James TimeKeeper: Andrew Gatekeeper: Eric
Thu/Fri	Collection Development Writing Group (Susan)		

Summary of Sessions

1. Geoserver
    a. Key takeaway: High cost in terms of knowledge, staff time and expertise. Opportunity to jointly document workflows and share common challenges and resolutions.

      b.   Next steps: begin document. Kim Durante to write up Stanford's workflow for customizing styles for raster datasets to start the process.

2. [Dataverse Integration](#)
   a. Some questions: Dataverse organizational governance? Funding? Consortia negotiated purchases?

3. [Open Formats for Preservation](#)
   a. Next steps: Ryan Mattke to convene an interest group to look into developing a list of formats and develop standards for them. If you are interested in joining, [add name to the list](#).

4. [Metadata Manager](#)
   a. Key takeaway: Eric and Karen working on developing tool over course of the year.
   b. Next steps: Sign up to stay updated on its development.

5. [OpenGeoMetadata](#)
   a. Next steps -- for group:
      i. Flush out https://github.com/OpenGeoMetadata/metadatarepository README as a template for other institutions (Andrew has already started)
      ii. Begin discussing potential GBL schema changes -- see [Geoblacklight Calendar and Roadmap](#)
   b. Recommendation for individuals involved in OpenGeoMetadata:
      i. Add documentation to your metadata using the README file in GitHub to document schema, usage, vocabularies
   c. Other projects: Karen and Andrew working on Friday at creating records to add to GBL for proprietary data like Social Explorer, Policy Map, etc that institutions have access to.

6. [Governance](#)
   a. Next Steps: See [Roadmap draft](#)
   b. Suggestion: Perhaps creating a running list/separate document for potential new features or more generically a summary of discussion

7. [Relationships are hard](#)
   a. Suggestion: Re-investigate [Issue 547](#) in github repo.
   b. Next steps: designate an upcoming community call (March 11) to discuss parent-child relationships. Bring up potentially switching use of source and ispartof in the schema.

8. [Geo Repos and Where's the Data](#)
   a. Next steps: James to provide a video for accessioning spatial content from Princeton U
   b. Suggestion: [Overview of SDI Technology Stacks](#) -- Big 10 Academic Alliance has a Google Drive where documentation can be found outlining their findings from other institutions. It would be excellent if this could be updated and expanded upon for reference in the community.

9. [Usability Study Group](#)

     a. Next Steps: Sign up if interested in developing cross platform usability study for GeoBlacklight

10. [GBL DevOps and Support](#)
     a. Key takeaway: there is a community need for more documentation on this subject. It would be wonderful if this were added to the GeoBlacklight wiki in the future. Are people interested?

11. [Conference Plus / Delta](#)

12. [Collection Development Policy](#) (Thursday working session)
     a. Next Steps: Susan to send email to geo4libcamp group to follow up. Make draft plans, group review of collection plans

Monday, Feb 3, 2020

- Welcome and introductions
  - Intro slides: Code of conduct, land acknowledgment, URL for notes (this page)
  - Thank the planning committee and especially Jib who deserves a standing ovation
- People intros:
  - Dave Mayo (Harvard, Developer working on Geoblacklight implementation)
  - Huda Khan (Cornell, Developer, human, working on CUGIR)
  - Jim Lacy (Associate State Cartographer, University of Wisconsin-Madison)
  - Jaime Martindale (Map & Geospatial Data Librarian, University of Wisconsin-Madison)
  - Keith Jenkins (GIS library, researcher support + CUGIR repository)
  - James Griffin (Princeton University Library, Developer)
  - Diane Fritz (Auraria Library / University of Colorado Denver, Geospatial Specialist)
  - Christina Chortaria (Princeton University Library, Developer)
  - Kara Handren (Scholars Portal / OCUL, Metadata Librarian)
  - Susan Powell (UC Berkeley)
  - Kasey Wilson (Michigan State University Map Library, GIS staff)
  - Karen Majewicz (University of Minnesota/BTAA Geoportal; metadata coordinator)
  - Ryan Mattke (University of Minnesota/BTAA Geoportal; Project Lead)
  - David Triebwasser (UC Berkeley, Head of Projects)
  - Kean McDermott (GWU Libraries, GIS and Data Visualization Specialist)
  - Melinda Kernik (University of Minnesota/BTAA Geoportal; Spatial Data Analyst & Curator)

- ○ Paul Aloisio (Harvard University, Systems Librarian working on GBL implementation)
  - ○ Amy Work (GIS Librarian, UC San Diego)
  - ○ Andrew Battista (GIS Librarian, New York University)
  - ○ Yuchai Zhou (GIS Developer, UC Berkeley)
  - ○ Josh Hadro (IIIF-C) (arrived on Wednesday)
  - ○ Evan Thornberry (GIS Librarian, University of British Columbia)
  - ○ Reina Murray (Geospatial Data Curator and Applications Administrator, Johns Hopkins University)
  - ○ Mara Blake (Manager of Data Services, Johns Hopkins University)
  - ○ Mara Sedlins (Data Management Specialist, Colorado State University Libraries)
  - ○ Tom Brittnacher (Geospatial Data Curator, UC Santa Barbara)
- ● Ideas for unconference sessions (braindump)
  - ○ Last Year's session tabulation sheet (ideas) https://docs.google.com/spreadsheets/d/17X8xteJCiKlZ4NI0mLnMquTf5S-m5OIT4eK7Ov_dyHA/edit#gid=0
  - ○ Geocoding services (Stace is happy to lead this)
  - ○ Data curation
    - ■ annual/etc archiving of regularly updated public datasets
  - ○ Data cooperation with U.S. gov agencies - pipelining into GBL
  - ○ authentication/authorization for licensed datasets - CAS/SAML?
  - ○ Linked Spatial Data / Linked Data Fatigue
  - ○ Dataverse as part of GIS/GeoBlacklight Workflows
  - ○ Workflows for getting records into Solr
    - ■ Tools, techniques, best practices
    - ■ Where do you start? ISO metadata?
  - ○ Dealing with Esri open data portals, or other geoportals
    - ■ E.g., local, state, federal
    - ■ Do you try to sync with these for your region/area of interest?
  - ○ Customizing geoblacklight - where do you draw the line?
    - ■ Need to be careful about over-customizing, because you then "own" those changes forever
    - ■ Mini-version of Monday workshop, but not hands on?
    - ■ proposing ideas to community (how) if you aren't a developer
  - ○ Analytics -how and why
  - ○ Relations between datasets
  - ○ COG (Cloud-Optimized GeoTIFF) and STAC (Spatio-Temporal Asset Catalog)
  - ○ GeoServer
- ● Workshop pitches
  - ○ Introduction to GeoBlacklight and GeoBlacklight Metadata Customization with GitHub (led by Christina Chortaria, James Griffin)
  - ○ Creating GeoJSON for OpenIndexMaps (led by Keith Jenkins)

- - ○ Geospatial metadata (led by Kim Durante, Melinda Kernik, and Karen Majewicz):preview of the first half of the metadata workshop: z.umn.edu/geo20
    - ■ Isometadata: bit.ly/2Uk4hC6 (notes for workshop below)
- Unconference session brainstorming to prep for tomorrow's planning
  - ○ 20ish minutes to write ideas down
  - ○ Session tabulation votes from last year: https://docs.google.com/spreadsheets/d/17X8xteJCiKlZ4NI0mLnMquTf5S-m5OlT4eK7Ov_dyHA/edit?usp=sharing
- Project team intros
  - ○ Big10 Geospatial Data Alliance Project
    - ■ 1 full time, 26 people working (~2 team members from each institution)
    - ■ Focusing on public data (public data, imagery, public scanned maps
  - ○ UC Santa Barbara / UC San Diego
  - ○ Ontario Council of Libraries
  - ○ John Hopkins
    - ■ Grant-funded project to showcase historical map collection -- focusing on this map collection first, then geospatial data in phase 2
    - ■ Starting implementation next week
    - ■ Metadata practices are aligned with Big10
  - ○ Harvard
    - ■ Esri to geoserver to geoblacklight
  - ○ University of Colorado / Colorado State
  - ○ Princeton
    - ■ Use both local geoserver and local ArcGIS server instance
  - ○ University of Wisconsin
    - ■ Part of Big10, but also has their own instance
    - ■ Only contains Wisconsin data.
      - .

Pre-Register for the David Rumsey Map Center.

Geo4Lib:
http://bit.ly/Geo4Lib2020

IIIF+Maps:
http://bit.ly/IIIF_Maps

Geoblacklight Slack
https://geoblacklight.slack.com/

Workshops

Introduction to GeoBlacklight and GeoBlacklight Metadata Customization with GitHub

Geospatial Metadata
Link: z.umn.edu/geo20

- Overview
  - We will review GeoBlacklight metadata schema, convert from JSON to CSV and back.
  - Will then start with DCAT metadata json then transform to CSV and then to GeoBlacklight schema
  - Kim will review ISO standards: Isometadata: bit.ly/2Uk4hC6
- Quick poll to check people's experience with topics.
  - Scale: 0 - 5 for Python experience
  - GeoBlacklight metadata experience
  - Metadata in general
- Clone GitHub onto desktop, with Jupyter Notebooks installed and running
- Use post-its to indicate whether stuck/need help or whether progressing as you wish
  - Yellow= happy and great
  - Purple=still happy but need help
- Steps for downloading/setting up Jupyter Notebook
  - Download https://github.com/BTAA-Geospatial-Data-Project/geo4libcamp-metadata-workshop and unzip
  - Start Jupyter Notebook (instructions differ for Windows/Macs) and go to URL where you should see the tree (list of files)
  - Within this window, navigate to the directory where you downloaded and unzipped the files for the workshop
- GeoBlacklight metadata schema
  - For metadata discovery, search for text and facets and spatially within map
  - Not designed for complete metadata documentation (can't store attribute table, etc.)
    - That metadata is kept separately and can be referenced from the GeoBlacklight schema
    - Some similarity to DublinCore but some elements that are completely separate for spatial values, etc.
    - Indexing in Solr relies on each record is its own JSON file
  - Creation: could write from scratch, crosswalk from similar schema (such as DCAT) or OpenGeoPortal or extract values from ISO, FGDC, or MARC

- ○ When importing metadata from other formats, it can be good to have an intermediary step for review with a spreadsheet.
  - ○ Big 10 geoportal: resources are hosted elsewhere with the metadata converted to GeoBlacklight schema. Will share/explain scripts and processes others may be able to use
  - ○ OpenGeoMetadata (github.com/OpenGeoMetadata): GitHub organization where every school/project has their own repository.
    - ■ Metadata is organized differently within each repo
    - ■ Big10: organized by collection code. Inside each code, can see individual GeoBlacklight JSON
    - ■ Cornell: pair tree structure. Long identifiers are separated out into sets of two digits, so would have to navigate multiple folders before seeing the FGDC and GeoBlacklight record
- ● In folder for this workshop, have collection of records from Big 10
- ● Activity 1
  - ○ GBL folder - 02a-01
  - ○ Flat schema
  - ○ DC stands for Dublin Core
  - ○ b1g is an example of a custom entry for Big10
  - ○ DCT references: external links, download, web services, etc.
  - ○ In GBL folder, scripts for converting JSON to CSV
  - ○ Cell -> Run All will result in this script being run and a new file being generated in the GBL folder
  - ○ Can download the CSV to review in a spreadsheet program like Google/Excel
- ● Activity 2
  - ○ Download CSV and open
  - ○ Some metadata elements only accept some values while others may be more flexible (separated by a |)
  - ○ Do some manual data-cleaning tasks such as replace the name "USGS" with expansion in the document and save with "_edited" added to the name
  - ○ Upload this to the GBL folder
  - ○ Update the JSON to CSV script to correct the mapping of the field for "b1g_code_s" to "b1g_collection_sm" and run the script
- ● Schema
  - ○ Which links show up where is informed by what is included in the dct_references field
  - ○ Out of the total fields in the schema, there are a few required fields with a few recommendations
  - ○ People sometimes do add their own custom metadata fields. They will then need to consider maintenance for those fields when they upgrade to a different version of GeoBlacklight
  - ○ Can refer to the BTAA Metadata handbook which describes the custom metadata elements they added

- The fewer custom elements we have, the easier it would be to share out metadata
- Adding "/raw" to the URL for a GeoBlacklight record will show the json metadata if you wish to understand the metadata schema behind the record

Part II - parsing DCAT schema

Overview
- Differences in DCAT schema
  - Add data.json at the end of a URL to see dcat schema
  - There is a lot of overlap with dublin core, but it's less flat
    - Dictionaries within lists adds more scripting
  - Plain text format (GBL can't recognized the tags within the text)
  - Compatibility with ArcGIS hub portals - There can be different structures between other portals like CKAN, etc.
  - Open data portals - items are frequently added and removed, so there is a LOT of turnover. This workshop is looking at some jsons that are two years old. The scripts don't solve the problems these updates cause. So this is for comparing snapshots so you don't have to re-accession the whole portal
- Metadata Quality is another challenge
  - Remediation for records can be required.

- DCAT harvesting Simplified
  - Activity 3 - largely doing a comparison of new json to old json
  - https://github.com/BTAA-Geospatial-Data-Project/geo4libcamp-metadata-workshop/wiki/Harvest-metadata-from-DCAT-enabled-portals
  - From the description of the activity in Jupyter Notebook:
    - For this demonstration, we'll look at a script that: 1) steps through a list of geospatial data portals 2) takes a snapshot of DCAT metadata for the items available in each portal, and 3) compares it to a previous snapshot of DCAT metadata from the same portals to see what has changed
- DCAT harvesting Full
  - First script was just to pull the un-nested comparison - and shows what is new
  - Second script goes deeper. For example, it digs into nested elements and will also show what used to exist and no longer does.
    - As a note says within script "The script currently prints two combined reports - one of new items and one with deleted items. The script also prints a status report giving the total number of resources in the portal, as well as the numbers of added and deleted items."
  - Melinda has done work to filter out datasets their group knows they do not want (Big10 folks)
  - Pull keyword from data providers

- - - Not always helpful or consistent
      - Sometimes include acronyms that may not be found useful by others
      - Preserve the keywords but create ISO topic category which is what is displayed
- Many times, content that is not geospatial metadata is often added to these repositories
  - PDFs
  - Links to other websites
  - CSVs that you could not open in a geospatial platform
  - Have a filter that will assign type = shapefile, will pull out image server content and identify as imagery
    - ArcGIS, CSV, error, Shapefile, web application
    - Allows them to pull out the types of content they need to be looking at
- Going over fields that merit examination, like dates, etc. this code addresses
  - Bounding coordinates? - common error that's hard to fix, but it's not in this code - (suggested solution is to kick out a broad bounding area, eg a county, for each item in a set like NAIP imagery)
- Use workflow to keep track of what has been added (removed) and then making only those changes as appropriate.
- Matching based on identifiers. If identifier has changed, that is treated as an entirely new item
  - Script gives list of all that has been deleted
  - Add a "suppressed" field so that the item is still available if a user already has a direct link but it doesn't show up in the search. Add a sentence like "this record may have been deleted from the source"
- Script also prints status report:
  - Portal name, total number of records present, new items, and deleted items
- Converting harvested DCAT records to GeoBlacklight JSONs
  - https://github.com/BTAA-Geospatial-Data-Project/geo4libcamp-metadata-workshop/wiki/Convert-harvested-records-into-GeoBlacklight-JSONs
  - Accomplishes:
    - Converts DCAT elements to GBL (single and multivalue fields)
    - Reordering of elements to be GBL compatible
- Discussion
  - How often to harvest data
  - Metadata is messy - what do you want to keep and how often
  - The above are hard questions and this exercise only showed HOW to harvest metadata.
- Isometadata: bit.ly/2Uk4hC6
  - Download this similarly to how you downloaded the previous content to view in Jupyter Notebook. Unzip in same directory as used for previous exercise
  - All the files exist already, running through this exercise will just overwrite them.
  - Templates - this folder is mostly xml files, but there is an easy-to-read html template.

- ○ Need GDAL to run this script (on your machine and not in Jupyter Notebook) which has binding dependencies
- ○ GDAL: reads native properties of the data
  - ■ Filename, spatial reference, geometry type, coordinates
  - ■ These are the elements ArcGIS pulls in for you
  - ■ Use this as a way to visualize the collection and see where the gaps may be
- ○ Scripts
  - ■ Create dictionary. Use variable names to store field names (easier to track/visualize when reviewing code/update)
  - ■ Date/times:
    - ● Spreadsheets will often normalize to a format (that you may not want)
    - ● Simple and common examples. Translate dates
  - ■ Running script will match metadata values to dictionary and add them
    - ● ISO advantage - carries metadata for codebooks
  - ■ Feature records generated separately from the main metadata (please correct if I got this wrong)
- ○ ArcGISXML - there is a GUI editing interface
  - ■ Files have endings like shp.xml or metadata.xml so can be distinguished from ISO
  - ■ Can't accession items w/o a thumbnail at Stanford, so there is a quick thumbnail generator that creates it and inserts into the xml record
- ○ Convert ArcGIS Metadata to ISO19139 through bash script
- ○ ISO to GBL
- ● Discussion
  - ○ Applying templates - within or outside of Arc?
    - ■ Overwriting still happens even if you apply the templates outside of Arc
    - ■ Things don't update if you don't directly click on them
    - ■ How to preserve what already exists? One approach: copy what is already there and paste in a text editor to put back in instead of redoing it all.
    - ■ Idea - toggle syncing T/F?
    - ■ Run script to export original metadata in csv, then do re-import of the important items
  - ○ Geodatabases?
    - ■ Presently Stanford breaks them into individual shapefiles
    - ■ Issue with field length and maintaining full information
    - ■ WI creates a GDB for each feature set
    - ■ Sometimes a GDB is a collection with odd files like pdfs, etc.
    - ■ Work with a parent - child relationship for these?
  - ○ Geopackages - they've just started showing up!
    - ■ Need to join the tables together to make them usable
    - ■ Princeton has a whole page on the GBL wiki on geopackages
    - ■ Is it a good preservation format?

Creating GeoJSON for OpenIndexMaps

Workshop URL: https://kgjenkins.github.io/openindexmaps-workshop/

Working in QGIS and installing plugin - QuickMap Services

Tuesday, Feb 4, 2020

- Keynote Sessions
  - Building & Sharing a Historical GIS: Insights from working with twentieth-century Mongolian maps, Susan Powell (Slides: https://docs.google.com/presentation/d/1DHjOm4FqCSWZTvrbfuOCzAoP2kaDdEP3-x4RLyIRJLs/edit?usp=sharing)
  - Mapping Prejudice: Cartographic Activism and Primary Sources, Ryan Mattke (slides: https://z.umn.edu/geo4lib2020)
- Lightning and other talks
  - Jim Lacy: Working with ESRI open data
  - James Griffin: Linked data, https://tinyurl.com/geo4libcamp-linkeddata
  - Amy Work: Shoreline/Surfliner: https://lib.ucsd.edu/shoreline-geo4lib2020
  - Kevin Dyke, Sideby, (demo) (slides) (github)
  - Andrew Battista, Groundhog Day, http://tiny.cc/groundhog2020
- Afternoon:
  - Unconference session pitches

Unconference Sessions

GeoServer, 2:40 - 3:40 pm

Notetaker: Reina
TimeKeeper: Kim Durante
Gatekeeper: Andrew B.

Suggested topics to cover
- Overview of Geoserver
- Tips and Tricks

- - - Default symbologies
    - Organizing layers with workspaces
  - Optimizing performance
  - [Extensions](#) (official) and [community modules](#) -- so far doesn't seem like anyone has tested out these extensions:
    - S3 Blobstore/Geotiff
    - DXF Output (for CAD)
    - Vector tiles
    - Geopackage
    - JP2 ← (Harvard is using extension)
  - Using the API
  - Effective integrations -- how to use geoserver
  - Geoweb cache
  - Infrastructure -- Is anyone using a hosted geoserver solution?
    - Cornell, NYU -- uses AWS EC2
    - Geosolutions -- Italian company that contributes to geoserver, may provide hosting services
  - Upgrading versions -- any experiences?

Overview of [Geoserver](#) (Andrew Battista, Keith Jenkins)
- Open source software developed originally by TOPP (The Open Planning Project, later becoming Development Seed). Now maintained by international open source community.
- Used as part of the stack for geoblacklight -- holds copies of the data; enables conversions to other data types on the fly; provides preview of the dataset in question and call up attribute data
- Think of it as a lens that refracts vector data
- Everyone using it together with PostGIS, but not necessary
- Highly configurable
  - Stanford manages layer settings in XML (custom XML setup) that syncs
- Can use GUI interface (styling, small tweaks) or API (loading files)

- Basic resource info: name, title, abstract (kept the same as in geoblacklight metadata)
  - General thought: populating fields in geoserver -- doesn't hurt to do
- Store = Dataset

Styles
- Each layer has an associated style; some schools have delved into developing custom styles (NYU -- lidar data, Cornell -- chesapeake land cover datasets)
- **Unanswered question:** People have encountered trouble customizing styles for raster datasets. Need to declare max and min; Stanford has a workflow where they define a custom sld (styled layer descriptor) -- way to express styling of a layer, vector and raster, using XML -- by calculating the max and min values and then applying a custom sld.

- **Challenges -- ingesting files with 3 bands:** https://earthworks.stanford.edu/catalog/stanford-zg150pg1648
    - Convoluted workflow to add 4th band/channel including gdal (transformation library for geospatial data) -- convert into another file format, convert back into a tiff and load it into geoserver
        - To do: Kim at Stanford willing to write up workflow to start process
- QGIS -- can style/tweak symbology and then export into sld
    - SLD4Raster plugin

GeoServer Workspaces
- **Question to consider** -- how many workspaces?
    - url must be in geoblacklight schema
    - Analogous to a namespace, a workspace is a container which organizes other items. In GeoServer, a workspace is often used to group similar layers together. Individual layers are often referred to by their workspace name, colon, then store.
    - May want to have multiple workspace for testing, or for thematic categorizations (eg: transportation data)
    - Can you use workspaces to split access? Eg public vs restricted data
        - Stanford -- 2 geoservers to handle access
        - Geoserver does have authentication settings
        - Princeton -- Uses two workspaces, both for the repository resources
            - They create and access the workspace itself over the REST API, which uses HTTP authentication with a password
            - One workspace is access-restricted, where a Single Sign On session is used to provide access to the WMS endpoint
            - One workspace is publicly accessible, and anyone can request data from the WMS endpoint

- **Question to consider** -- Use of docker images to create geoserver?

Other
Be wary of bad file names -- certain characters (& for example) - It's been noted that filenames cannot begin with a number, have spaces or special characters in them, or they won't be enabled in GeoServer

Geoweb cache -- Cornell experience issues with the amount of space being used by tile layers not adding up correctly. Make sure to clear the cache when re-ingesting layers.
There is a plugin to move geoweb caches to aws bucket -- hasn't been tested by anyone in group

Geoserver upgrades -- generally pretty mild development over the last few years; Cornell reported relatively smooth upgrades

To consider: high cost in terms of knowledge, staff time and expertise. Potential opportunity for documentation of workflows and common challenges/resolutions?

Mapserver as an alternative? Doesn't seem like anyone is using/has investigated using it with Geoblacklight over geoserver.

Dataverse Integration

Roles:
- Timekeeper: Dave M
- Gatekeeper: Paul A
- Notetaker: Eric Larson
- Facilitators: Mara B and Evan

Dataverse, self-depositing data repo for researchers.

Ottawa / Dataverse presentation / Slidedeck
Slide Deck from most recent Geodisy presentation:
https://docs.google.com/presentation/d/1_QB5R5cfZQRanHbfF-DThBYKcYrL-7Jqwk_6DJK_FZE/edit?usp=sharing
Geodisy video from Texas Digital Library webinar:
https://www.youtube.com/watch?v=B3IZnMRhdqA&feature=youtu.be
Geodisy Github repo: https://github.com/ubc-library/geodisy

UBC uses Dataverse (named Abacus) as the university's only data repository. In Canada, federal support for extending data resource/infrastructure. Scholar's Portal (centralized dataverse instance). UBC is creating a GBL discovery instance for Scholar's Portal. Grant funded through March. Portage is another funding organization behind Geodisy ("**Geo**spatial" + "**Dis**cover**y**"). Additional plans include French lang translation, functionality. Connectors to CKAN, turnkey data repository for municipalities.

Auth mechanisms for Dataverse CAS SAML

UBC project layout: Have Dataverse moving data into GeoBlacklight via DDI source data crosswalk to GBL JSON. Dataverse, integrated with Datacite for minting DOIs. UBC creates bounding boxes for "filled-out" geospatial metadata fields for non-geospatial data. Geospatial files bounding boxes are created via GDAL. Moving Dataverse the data to GeoServer, providing

enhancement. Translating DDI to ISO for downloading in GBL. Discovery happens via Dataverse and via GBL. Not all Dataverse items are present in GBL.

What does metadata going into Dataverse look like? Researcher provided, sub-par, non-optimal descriptions. Configurable submission form.

Harvard has multiple Dataverses within Dataverse and are looking to build a Harvard "curated" Dataverse instance. Workflow: digital repo for some things, GeoServer for other things. Middleware for pulling items together (ex. Med school repo). Challenge of centralizing data across data repos and dataverse into single place. Planning for next version of digital repository.

Questions: Dataverse organizational governance? Funding? Consortia negotiated purchases?

Johns Hopkins: Dataverse in place for 10 years. Only doing mediated deposit. Metadata is fairly standard, with many fewer data quality issues. JH "Box" branded data to be migrated to Dataverse. Setting up GBL. Geospatial data in Dataverse may not be described as well as needed.

Discussion of data ingest flow and canonical representation.
Dataverse is Archivmatica "friendly".

UBC Data flow: Dataverse > GeoServer > Geocombine > GBL
GeoServer exposes endpoints for Dataverse persistent identifiers

Open Formats for Preservation

Roles:
- Timekeeper:
- Gatekeeper:
- Notetaker:
- Facilitator: Susan Powell

- When thinking about the future, how do we handle proprietary data types (ie shapefile) for long term preservation as we build repositories and portals.
- What makes data open vs proprietary?
  - It requires a spec to be open data format
- The use of zips for shapefiles is common for serving data out but not good for preservation
- Shapefile has a PRONOM entry: https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=328

- GeoJSON might be useful but not good for larger files and doesn't preserve topological relationships. TopoJSON does preserve those relationships.
- Coverage files might not be useable in ArcGIS Pro and
- GeoPackage is an open standard that might serve these needs but not many attendees have used them. http://www.geopackage.org/spec/
- One of the best ways to maintain useability of data types is for people to use it!
- How can we document the changes or losses in data when converting to a chosen format for serving data vs preservation?
- LOC recommended formats: https://www.loc.gov/preservation/resources/rfs/data.html. See section on geospatial data.
- Data Curation Network building lists to help make these decisions (Curation Primers)
- Where can we go to find information on best practices on these preservation methods?
  - Geoblacklight and Geo4Lib community
  - Data Curation Network
  - Library of Congress map and geospatial division (GeoMapp Project, NDIP were previous efforts)
  - GIS&T Body of Knowledge: https://gistbok.ucgis.org/
  -
- There needs to be a social aspect to preservation projects to make efforts sustainable and not fail after retirements, etc.
- Raster formats
  - TIFF, GeoTIFF, MrSID,
- Library of Congress - GeoTIFF
- What about world files?
  - What is the advantage or disadvantage of having a single file
    - Use of TIFF files in non-GIS software might be easier with a world file and can be used with other formats (Jpeg)
- What about other file types?
  - Network datasets
  - Topology
- What workflows do people use to handle their archive files?
  - Typically take in whatever file type is given to them and convert to what is best for sharing
  - Issues of controlling and updating data once it's archived
    - Data can be gotten out but not converted in the dark archive (revise and resubmit)
    - The spec should be included with the archive file to assist with future use
    - Zipped files can be an issue since you can't know the type until opening it
- Repositories conversation should include discussion of preservation formats
- Semvera community has a working group on preservation formats if people are interested\
  - https://wiki.lyrasis.org/display/samvera/Samvera+Geospatial+Interest+Group

- Should Geo4Lib/GeoBlacklight community develop a list of formats and develop standards?
- Joining an interest group:
    - Ryan Mattke will convene a meeting
    - Eliot Jordan - eliotj@princeton.edu
    - Susan Powell - smpowell@berkeley.edu
    - Melinda Kernik - kerni016@umn.edu
    - Andy Irving - Andrew.Irving@bl.uk
    - Kara Handren - kara.handren@utoronto.ca
    - Abby Pennington - apennington@ucsd.edu
    - Jaime Martindale - jmartindale@wisc.edu
    - Tom Brittnacher tombritt@ucsb.edu
    - James Griffin: jrg5@princeton.edu
    - Kasey Wilson wils1301@msu.edu
    - Phil White philip.white@colorado.edu
    - Yuchai Zhou zhouyu@berkeley.edu
    - Reina Murray - reinamurray@jhu.edu
    - Jack Reed - pjreed@stanford.edu
    - Evan Thornberry - evan.thornberry@ubc.ca
    - Christina Chortaria - cc62@princeton.edu
    - Amy Work - awork@ucsd.edu
    - Paul Aloisio - paul_aloisio@harvard.edu

Metadata Manager 3:45-4:45

Roles:
- Timekeeper: Jack Reed
- Gatekeeper: Eliot Jordan
- Notetaker: Kim Durante
- Facilitators: Karen & Eric

Designing/Creating a metadata editor for GeoBlacklight

Administrative interface for creating, editing and managing metadata.
Object creation, Normalization, and CSV import.

What are some issues in current metadata workflows that could be addressed by a tool?
- Validation
- Sync between solr and GB instance
- Comparison between institutions - inventory
- Filtering out (ignore or delete) records
- Filtering record/evaluating records from other institutions
- Filter by specific values (place,publisher, etc...) to ingest from OpenGeoMetadata (OGM)
- Support for custom fields
- Support for standardized vocabularies (LCSH/LCNAF, ISO Topics, File Formats, Data Types)
- Active exchange of metadata vs. searching OGM for specific records
- Autogeneration of metadata fields - bounding box, data type...
- Webform for creating and editing
    - Example: https://www.mdeditor.org/
    - Scholars Portal has a custom tool, IP restricted. kara.handren@utoronto.ca (willing to open for viewing !)

How to manage changes in GB metadata? If a field/value is changed, does it need to be changed in the source metadata (ISO, FGDC, etc)? It depends where the canonical metadata are stored. Could canonical metadata be propagated with changes to GB metadata?

GeoCombine - tool used to ingest records from OGM into local instances. Used for metadata transformations. SDRFriend (NYU) is a similar application.

Traject - newer tool to manage metadata pipelines (Stanford, Princeton uses for Blacklight)

Read from a dataset technical metadata (point, line, polygon, geometry)

Should this be a standalone tool? Web-based? Intention is to couple the tool with the UI. Will include support for user authentication.

Indicate metadata that are unpublished and ready for review. (needs review, flags, take down)

UCB wants a magic button to harvest all records from all institutions

Faster way of generating fixtures for testing/sampling new features.

Custom fields **currently in use:**

Big10 - collection codes (internal), genre (scanned maps vs. geo-data), thumbnail path

NYU - nyu_addl_downloads_s, nyu_addl_formats_sm, nyu_addl_dspace_id

Wisconsin - uw_supplemental_s (Additional info about archived data vs. producer copy in geoportal), uw_notice_s (used to display a custom warning message on a per-dataset basis)

Stanford - use and reproduction

Cornell -

```
  "cugir_filesize_s": "2.68 MB"

  "cugir_addl_downloads "cugir_addl_downloads_s":
"{\"PDF\":\"https://cugir-data.s3.amazonaws.com/00/79/97/agWAYN.pdf\",\"KML\":\
"https://cugir-data.s3.amazonaws.com/00/79/97/agWAYN2017.kmz\"}"
```
  (modeled after NYU)

```
  "cugir_category_sm": [ "landcover", "climate" ]
```
  (human-readable ISO Topics)

Please sign up to stay updated on Metadata Management Tool development!

- eliotj@princeton.edu
- reinamurray@jhu.edu
- tombritt@ucsb.edu
- cc62@princeton.edu
- smpowell@berkeley.edu
- davidtriebwasser@berkeley.edu
- yzhou@berkeley.edu
- jmartindale@wisc.edu
- awork@ucsd.edu
- lacy@wisc.edu
- paul_aloisio@harvard.edu
- kgj2@cornell.edu

-

Wednesday, Feb 5, 2020

Introduction to IIIF

Notetaker: Reina Murray
Timekeeper: Amy Work
Gatekeeper: Eliot Jordan

Starting with adapted workshop http://iiif-ml-workshop.netlify.com
Question on setting up a IIIF server

Josh -- IIIF managing director is here (so ask him questions)

Overview: Starting with https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/problem
- Images = fundamental information carriers for cultural heritage
- Without standards content becomes siloed -- data gets tied to the platform, and what happens to that data when platforms inevitably change? Interoperability is key.
- IIIF -- solution for image content.
  - Provides specifications for **serving** and **consuming** images on the web
  - Like WMS for images
  - Community that develops shared APIs together
    - Also implement them in software
    - Try to expose content in interoperable ways so that content can be used and reused without having to rehost it on the web
  - **Viewing and usage experience is not tied to data model at local institution**
- See diagram and example on https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/solution
  - Watch youtube video at end of page (under "more resources")
- **Technical Specifications**: 4 stable specifications with distinct use cases that they serve (https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/iiif-apis):
  - IIIF Image API - Used to deliver "pixels" to end users
    - Deliver high res images viewing on the web without having to have user download the image
  - IIIF Presentation API - "Just enough metadata to drive a viewing experience"

- - ■ Not a metadata standard, but there's metadata in it, enough to drive a viewing experience (see image on the top link showing what aspects come from which API)
    - ○ [IIIF Authentication API](#) - An authentication specification and workflow for providing authentication for IIIF images
      - ■ Eg: have content that can only be viewed in a certain restricted data room, or can only be viewed by authenticated users. This API provides workflows for those scenarios.
    - ○ [IIIF Search API](#) - Used to "search within" IIIF content
      - ■ Not searching across content, but within content (eg: look through newspaper image for all references to a topic)
- **IIIF Community:** see https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/community-software and https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/more-presentation-api#request--response-cycle
  - ○ Image Servers
    - ■ ContentDM -- any content in this server now gets IIIF automatically
- See **[IIIF Awesome List](#)** for examples
- To make a geo analogy:
  - ○ Geoserver = image server (where you host the data)
  - ○ WMS = IIIF API (specifications)
  - ○ Leaflet = image client
  - ○ Geoblacklight = image app (how you view it all)
    - ■ Note -- geoblacklight has IIIF support
- **Image API** (see https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/image-api)
  - ○ Standardized way to request and deliver images
  - ○ See link above for breakdown of a request
  - ○ Specified pixel region (see second image) - input starting x value, starting y value, width and height **{x, y, w, h}**
  - ○ Play around with parameters at https://www.learniiif.org/image-api/playground
  - ○ **Don't need to create multiple copies of an image for a header banner, thumbnail, etc -- can use IIIF to serve all of this.**
- Potential project: If we have bounding box coordinates to look at a specific portion of a scanned map -- would need to find the pixel version / community looking into better supporting georeferenced maps with IIIF
  - ○ Pixel cropper apps -- Stanford uses it (embedded in Spotlight) to be able to quickly define and pull the pixel parameters for an image if you want to crop it
- **Examples:** see (https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/image-api-examples)
  - ○ Key takeaway: *The IIIF Image API allows you to serve out derivative versions of these images in a standardized way, without having to generate derivates in advance.*
    - ■ Think of it as tiling images

- To keep in mind: Different strategies for how to optimize performance and balance budget, goals, etc
- **Internals**: (see https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/image-api-internals)
- **Presentation API** (see https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/presentation-api)
  - Structure that contains important information like (image title, order to present images, etc)
  - Note the descriptions for manifest, collections (can have infinitely nested collections in IIIF -- don't do it), etc
  - Example:
    - Atlas = collection
      - Map 1 = manifest
      - Map 2 = manifest
  - Manifest - most used concept, but collections are also important to know about
    - Manifest -- has a sequence (eg: book -- sequence of pages)
    - See structure of a manifest: https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/manifest-structure
      - Think of canvas as a powerpoint slide (canvas has a width and a height, may or may not be the same as the image that is on the canvas) -- eg: have an old scanned image with annotations, planning to re-scan at higher res in a few years. Set canvas to the resolution of the new; when you swap out the image with the new scanned image, the annotations will scale appropriately.
  - See https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/more-presentation-api
    - Note -- image api and presentation api in json
    - Linked data
  - See https://iiif-ml-workshop.netlify.com/docs/intro-to-iiif/more-presentation-api#request--response-cycle
- **Coming Soon/In the Works:**
  - Presentation API v3
  - Discovery API

Example of IIIF from CONTENTdm:

- Multipage (using Presentation manifest)
  - Page for Nobles Atlas in a CONTENTdm instance: http://cdm16022.contentdm.oclc.org/cdm/compoundobject/collection/p16022coll231/id/2304/rec/2
  - The IIIF Presentation Manifest JSON API: (This is all the atlas pages with some metadata - https://cdm16022.contentdm.oclc.org/iiif/info/p16022coll231/2304/manifest.json

- ○ How it looks in the Big Ten Geoportal with the Mirador Viewer: https://geo.btaa.org/catalog/3f8fef84-74d3-4e5a-aadd-44350c6419ce
- Single page (using Image API)
    - ○ Page for geological map in a CONTENTdm instance: http://cdm16022.contentdm.oclc.org/cdm/singleitem/collection/mgs/id/926/rec/1
    - ○ Example of a IIIF Image API: (single image) https://cdm16022.contentdm.oclc.org/digital/iiif/mgs/926/info.json
    - ○ How it looks in the Big Ten Geoportal with the Leaflet IIIF Viewer: https://geo.btaa.org/catalog/04d04cb1-2523-4c8a-bcd1-1f6588dde05f
    - ○

Questions:
- Balance bandwidth -- Stanford has load balance IIIF image servers with another layer in front to do authentication control/mitigate any bad actors. Stanford also does pre-warming caches -- pre-emptively run caches.
- Can you throttle operations on a session basis -- Stanford uses rackattack- yes
- What goes into setting up a IIIF image server? Are there resources for setting up a basic one for proof of concept? What's a good workflow?
    - ○ See training linked on IIIF Awesome page
    - ○ Cantalope - Java IIIF image server
    - ○ Klokan technologies -- also provides IIIF hosting (in addition to lots of other services)
- How long did it take to transition over to IIIF?
    - ○ Relatively simple -- transfer files over to IIIF-enabled image server
    - ○ Hard part -- how futzy do you have to get on your front end?

Data extraction from scanned maps

Notetaker:Huda
Gatekeeper:
Time keeper: Eliot
Facilitator:Kasey

- Princeton Paper: http://www.e-perimetron.org/Vol_12_1/Shawa.pdf
- https://gisportal.princeton.edu/portal/apps/MapJournal/index.html?appid=9e3a743a0eef441fbc7e45fd244dcead
- ●
- University of Calgary paper is great reference https://www.researchgate.net/publication/329637837_From_Paper_Map_to_Geospatial_Vector_Layer
- ●

- Discussion: What people are working on? AI tools? What works/doesn't?
  - Kasey: read two papers. Several historical maps at Michigan with ethnic info. Tried semi-automated methods. Not working with as many maps but later on digitization will be what's doe.
  - Bahar: Boston Center Institute Project. Lots of maps that have not been georectified yet. Paper attached to maps are still in index cards and are at MA historical society.
    - Students are creating excel sheets to join additional data with maps
    - Georectification by hand. Some info changed (streets etc) - remember Boston is old city
    - Problem: Time consuming, georectification not easy, digitizing index cards is difficult. Need to supervise students and the info needs to be validated.
    - Susan Powell's talk was very interesting - if we could use automation to validate the info (even if creating manually)
    - Creating this info for future researchers and community usage and want to make sure people can add data later. Creating rational (hope I got this right) dataset behind this info
  - Collection of 2300 maps covering state of Oklahoma. Each map represents township (6 * 6 mile square). Tried some algorithms
    - Seems like should be able to find largest square on sheet and clip it out
    - After that, could say use these four points (township intersection points), and then automating georeferencing
    - Maps made in 1930s. Just non-standard enough that algorithms didn't work out.
    - Students georeferencing
    - Would like to talk about computer vision and how that would help
  - Bahar: any active group working on such a project? Automated georectification?
    - Seems like it's just a dream/goal
  - Susan: Stanford working on project to do automated vectorization of old highway maps. Could follow up
  - [QUAD-G+](#) ([User manual](#))
    - Software used to automatically georeference USGS topo surveys
    - Someone at McMaster working with this to look at Canadian topo survey
  - Facebook initiative to extract road outlines from satellite imagery - could be related (http://mapwith.ai, http://deepglobe.org, https://github.com/zlkanata/DeepGlobe-Road-Extraction-Challenge… satellite imagery based but could it be useful with other training data?)
  - If a researcher goes to Mongolia and does this for a map, etc. - how do we find others who have done this kind of work?
  - Bahar: have digital repo saving info in zipped formats but cannot exhibit shape files b/c don't have tools for that repo
    - Don't have GeoBlacklight or PostGres
    - Saved info in EsriCloud

- ■ People can go to ArcGIS online and extract shapefile as needed
  - ■ Wish to eventually have a web service to provide to researchers so they can preview/view the info before downloading
- ○ Eliot: Considered partnering with CS department?
  - ■ Kasey: Have high performance computing department so could get their help.
  - ■ Bahar: integrated into CS. During semester, students work with repo but not focused on mapping
- ○ Kasey: What kind of scale are you considering? As far as collection size
- ○ Eliot: Sanborn maps for insurance for NJ - not georeferenced, 30,000 sheets. Would take a long time to manually georeference these. Would be a great idea to partner with Princeton undergrads.
  - ■ Good collection since they are from one state so narrowed down.
  - ■ Do OCR on them
- ○ Matt Miller: need training models. If there was a corpus of hand made vectors - could become basis for training data
- ○ Susan: American Association of Geographers: historical clearinghouse with GIS info. Could browse to see what approaches were taken to create historical GIS http://www.aag.org/cs/projects_and_programs/historical_gis_clearinghouse/hgis_databases
- ○ Kasey: Waste of time?
  - ■ At a certain scale, is it no longer of interest?
  - ■ How do you target collections? Where do you start?
  - ■ African study librarians: would go to them first to see what their most needed stuff was
  - ■ Not just technical angle - have limited resources
  - ■ Mara: Have rich water archives collection at CSU. How good to have the quality have to be before georectification? How do you target collections?
  - ■ Kasey: Set of standards or recommendations that could come out of this group?
  - ■ Kasey: If have map of something we think is unique and could be valuable for research, would target that for georectification
    - ● Would need to go hand in hand with strategic plan and policy
    - ● Better to do something like the BTA geoportal where resources be pooled?
  - ■ Ryan: Got resources to do some GIS projects. Found some faculty - for your history class, can overlay these maps and you can provide testimonials of how cool this is. After some years of this, got funding for digitizing content
    - ● Had religious studies professors come in - Map of mosques in the US in 1980s - useful to see how that changed over time

- ■ Kasey: Ethics of the info we make available. E.g. locations of natural resources where people who own maps don't necessarily want to share this info.
  - ● Is this something we decide on our own with stakeholders?
- ■ Ryan: issue with extracting data from copyrighted material?
  - ● Susan: No. The map is copyrightable but not the information
- ■ Kasey: had that same question about Sanborn. EBSCO owns a lot of this info. Is the data transformational enough that that is ok?
  - ● Eliot: a lot in our collection are out of copyright. Ever January handle many that are coming out of copyright
- ■ Josh: Worked at NYPL. Any data extracted would be CC0.
- ■ Susan: Could make argument that not showing the whole map and the data results in transformative piece
  - ● Josh: doesn't replace utility of the original
  - ● Susan: e.g. people can copy chapters
  - ● Josh: for educational and non-commercial use and transformative. Caveat: not a lawyer.
- ○ Kasey: how many people have already done some work in this area?
  - ■ UBC: not a priority yet
    - ● Request by researcher. Subset of forest inventory scanned maps. 149 covering province.
      - ○ Thought of partnering on project to automate georeferencing
      - ○ Researcher did vectorization on own with their own resources and students
      - ○ Was looking at Quad-G
      - ○ Used crowdsourcing
      - ○ Working to get scanned georeferenced surrogates into repository. Not done yet
- ○ Susan: Forestry grad student who identified soil and survey maps in California. Had team of undergrads to georeference. Vectorized half of them.
  - ■ End goal: make available through geoportal and digital library
  - ■ Need to work on setting up system to handle this content/process
- ○ Jamie: Researcher applied for grant to digitize maps but did not get funded.
  - ■ If way to save time and money, could help save on budget in case that was an issue for funding
- ○ Glen: Before working on IIIF, worked at national library of Wales. Maps covering all of Wales digitized. Worked with Klokan (links below)
  - ■ Info: who owned the field, etc.
  - ■ Output: Click to see landowner, etc. and then can see original map (boundary numbers, etc.)
  - ■ Difficult to see which fields owned by a particular landowner

- - Tried to use computer vision to identify boundaries of fields to then say this field is owned by this person
  - Used [others' work](#) with computer vision to identify buildings.
    - First step: black and white
    - Controlled finding - disregard any boxes that are too small and then draws boxes around fields
    - Some of the results are quite good but others not so much (e.g. instead of finding one field, combined two together - this may be because of pencil lines)
  - Josh: Did this at NYPL. Reduces cost of problem space. Getting people to correct this as opposed to getting them to draw bounding boxes: easier and faster.
  - Christina: land use project. Were these ortho-maps or were they geo-referenced? (Ortho-rectified i.e. they consider the elevation - need to request triangular points which is a solid concrete point on the ground - usually the army or survey uses these points b/c reliable).
    - Glen: Not sure about this but Klokan got points from old maps and in new and then used that info. About 1000 maps.
      - Can georeference and transcribe in same system
  - Links:
    - Crowdsourcing system: https://www.georeferencer.com/
    - Public website: https://places.library.wales/
    - Details on the project: https://www.library.wales/discover/projects/end-of-project/cynefin/cynefin-local-projects/
  - Kasey: Need to consider audience for data?
    - Sending to survey team versus people who want to know where their grandparents live
    - Christina: who owns what land - needs to be accuracy b/c could request money from government on this basis.
  - Susan: Creating layers: what is the standard reference layer?
    - Christina: have same projection and georeference system. Pseudo-mercator (which thinks earth is a globe, which it isn't)
      - For web, use WGS-84. Intent to display map in 2D surface.
      - Have to have same projection across all the layers
    - Susan: current modern reference layers don't line up with each other. Cities may appear in slightly different locations International boundaries may shift. What is the definitive layer for the world?
      - Produced by governments? May differ
    - Christina: Scanned map has no projection, just scale. Possible to project it. If projected in specific country's region projection system, won't be the same as WGS-84 which would be considered global.

- - - Susan: if with the same projection, places where they don't match. Georectifying scanned map to some kind of consistent visual layer. If have 5 scanned maps and georectifying all of them, layers will align
      - Kasey: zones of uncertainty (as discussed in keynote yesterday). Where Mongolia considers its boundaries vs how China may identify the boundaries
      - Susan: Even when accounting for projections etc., still discrepancy
      - UBC: As long as trying to follow best practices for georectification, may need to live with some level of discrepancy
      - Jamie: georeferencing the natural earth while someone else georeferences another source, can't say which one is authoritative - as long as you document which one you're using
      - UBC: normalize back in time - not using something so old that don't have roads, etc.
      - Kasey: start with something recent and move back in time
      - Mara: automatically move from something georeferenced with one standard to another?
        - Reprojection? Yes
      - UBC: NYPL Map Warper? Josh?
        - I left a year and a half ago. But last I checked, version 2 was out
        - UBC: Vectorizer - used to create preliminary features which could be put into building inspector (which is a project? That is the notetaker asking a question)
        - Matt: repo which doesn't appear to be maintained but if you put in a sanborn map you get data (or something) back
    - Susan: Photoscan (proprietary software): ortho mosaics (https://www.agisoft.com/features/professional-edition/).
      - Heather Ross' WAML 2019 Lightning Talk: https://drive.google.com/a/berkeley.edu/file/d/1LGXy7M-GKmbDtLksHHxj362xsbbX5C5S/view?usp=sharing
    - Ryan: open drone map. Gave it directory of 1940 county maps, used tool and it generated map of county (from pictures of separate overlapping areas). Created GeoTIFF. Overall off by 10-20 meters. https://www.opendronemap.org/ Contact is David Bittner
      - Related blog post at https://www.opendronemap.org/2019/05/stitching-historical-aerial-images-from-1940-using-opendronemap/

Related article:
- Automatic Feature Extraction and Text Recognition From Scanned Topographic Maps Pezeshk, A ; Tutwiler, R. L
  - IEEE Transactions on Geoscience and Remote Sensing, December 2011, Vol.49(12), pp.5047-5063

DOI: 10.1109/TGRS.2011.2157697
- Boston Research Center prototype website: https://bostonresearchcenter.org/

Tool description:Semi-automatic delineation of visible cadastral boundaries from high-resolution remote sensing data
https://its4land.com/automate-it-wp5/

Geocoding

Geocoding definition: taking something implicit and translating into something explicit
- Types of geocoding
- Place name - text into geographic location (e.g. lat/long)
- Photos - along a line, start and end points. Place first photo, place last photo and then interpolate along the line

Topics of interest
- Data Sources: Esri, open
- Infrastructure to support those
- Stand alone, home-brewed geocoders for sensitive data
- Fundamentals and concepts

Fundamentals
- Types of geocoders
  - Matching
  - Percent along - the method for street address level since 1967
  - Text search - street address, roof point, tax lot, etc
  - Composite - most specific to least specific (street address, goes to zip code)
    - Example Esri Streetmap Premium (HERE Locators)
      - You can partner with Esri + HERE to get "custom roads" dataset to allow integration of other datasets like GTFS.
- Sources of geocoders
  - Geolocate: out of Tulane, special for biological collection (e.g., bird tag found 5 miles north of Stanford). Bakes in circle of uncertainty = point + uncertainty, in addition to other isoteric styles. Free and unrestricted.
  - OpenRefine: great for doing geocoding. Easiest to do for small or medium sized datasets, just add a table. Up to 100,000 addresses. Anything more than that should use python & api/programmatic approach.
  - ArcGIS for APIs
  - GeoNames: Sign up as an individual, get an API key, 40,000 geocodes per day. Good for place names, but not street addresses. Stace sends individuals here whenever possible.
  - Interesting approach to deliver geocorders for PHI applications as single virtual machines within an isolated network (Stace has some examples of packing a Windows install with tools needed for researchers, researcher burns it when they are done).

- ○ StreetMap Premium from Esri: better documentation from the World Geocoders rather than the documentation for the API references.
  - ■ At regional level: North America, Europe, Africa, Asia ($5,000 per region under an Higher Education Enterprise License)
  - ■ In conversation with Esri for a discount of $20,000 for world. Still not confirmed.
    - ● This will effectively give you a world geocoding service on premise without using credits in ArcGIS Online.
  - ■ Add it to your license. Work with Esri Rep.
  - ■ Stanford manages this on ArcGIS Server as part of Enterprise
    - ● ArcGIS Server Manager is managing all the services and geocoding services.
    - ● Built system because at the time only had one seat.
    - ● Typical usage of 1million addresses at a time, a few times a week.
  - ■ Not logging anything on the server. using SSL. Can use for human subjects, but not PII data.
  - ■ No throddles
  - ■ Setting up ArcGIS Enterprise - good documentation from Esri.
    - ● ArcGIS Enterprise, Web Adaptor (managing through web page), data store (similar to the old SDE).
    - ● Server: 16gb RAM, 64bit operating system
    - ● Results
      - ○ 600,000 an hour using Desktop or Pro → submit entire record
      - ○ 1 million an hour using server geocoding
    - ● Have it locked at the service level as well at the site so that unless someone is authorized they won't be able to mess with it.
    - ● Locator properties can be turned off to get 4X boost.
      - ○ Reference data ID
      - ○ User specified output field
      - ○ Write elapsed time
      - ○ Write score per component
- ○ If you are on your own, have limited IT, Esri StreetMap premium is the easiest way to get started.
- ○ Don't do geocoding in ArcGIS Online, You don't get all the results back. Scores, quality, matching and unmatching. Use the Desktop or Pro and link to the World Geocoding Services.
- ● QGIS: MMQGIS plugin for geocoding
  - ○ Can now use the Esri API and run geocoding with QGIS
  - ○ Also supports US Census Geocoder and OSM Nominatim

Open Source Geocoding
- ● Pelias - picked up by the open source community, MapBox.
  - ○ Based on open source data. Installation pretty straight forward, can be containerized. Setting everything up is based on the installation processes.
  - ○ Get OpenAddresses data - collecting address level data at the highest resolution.
    - ■ Doesn't have StreetMap USA in there, but pulls from where it can
- ● TAMU - Texas A&M Geocoding service

- [Getlon.lat](#)
  - Geocoding service APIs from dozens of different providers analysed to help you choose. Lists free geocoding services as well as those that $$.
  - Nominatim - great for place names, not optimal for addresses
- [Who's on First](#)
  - Gazetteer rather than a geocoder
    - Authoritative
    - Hierarchy baked into it
  - Interested in using it in the cataloging infrastructure. Place names for catalog records. LOC headings included.
  - [Quatro shapes](#)
- Nextzen (maintains infrastructure that mapzen built)
- [US Census Geocoder](#)
  - No api key needed, a little slow, mix of percent along and point, bulk upload is supported 10,000 per job.
  - accessible via QGIS MMQGIS plugin
- [Geomarker](#) - uses TAMU infrastructure, takes list of place names or (lat/lon) but then pulls out census data like avg income for census tract.

Privacy & logging concerns
- Be mindful of what gets logged as data is submitted and processed against geocoders.

Scalability of teaching geocoding
- Added as a part to the intro to GIS course, good general exposure
- Guidance on larger jobs, Python notebook in github as shared code template
- Big question to help answer - what level of patch is good enough 85%
  - Statistically sufficient or how will you iterate the geocode to get needed results
    - Parcel centroid, post office points, PO box, etc -> implications for bias introduced into research

Invented at Yale School of Forestry. Integration of topology along a line, integrated into street network.
- FromAddressLeft, FromAddressRight.
  - These encode adjacency into the data
- Finds the street first, then identifies the "percent along" that segment based on the underlying attribute numbers.
- Works really well in urban environments. Rural area where you may have ½ mile between intersections, that can cause problems.
- Toxic Release Inventory has also Lat/Long address in addition to address due to many of those facilities being located in rural areas

Why select Esri vs. setting up your own geocoding services.
- Support staff is a big reason to by the data and leverage StreetMap premium
- Good results, but without needing all the backend support.

GIS Index Map Creation Recommendations and Union Catalog

(Wed. 9am, Teaching Corner)

Document:
[GIS Index Map Creation Requirements and Recommendations document](#)

Notetaker: Phil White
Timekeeper: Kasey Wilson
Gatekeeper: Susan Powell

Background: discussion of OpenIndexMaps github repository and effort to standardize index map metadata began at Geo4Lib 2019(?).

[OpenIndexMaps.org](#)
https://github.com/OpenIndexMaps
Check ownership, get another admin to OpenIndexMaps repo.
- Index maps in OpenIndexMaps now don't follow a standard yet.

Review of Requirements and Recommendations [Doc](#)
- Wanted to avoid unnecessary fields.
- Requirements: required by GeoBlackLight
    - We don't want to prevent folks from participating because they don't use geoblacklight, we just want their records to be usable by others
    - ***Needs to be geojson format!
- Recommendations: this would be nice for us all to do the same way

Should each of these files be like a union catalog of what we all hold? Or should each index represent just what we hold at our individual institutions (not a union catalog)
- For openindexmaps--separate for institutions.
- Does not preclude the idea of a larger union catalog system that pulls from these records.

Discussion of naming conventions.
- Tom: we don't see a reason to specify file naming conventions. Does anybody know of a reason why we should?
    - Would it be easier to merge things later?
    - There is likely a way to uniquely define each institutions' files anyway.
    - Could be multiple things with the same name ("Uganda 250k" for example)
- Tom: you can put whatever you want in the header area of geojson.
    - Good place for collection-level metadata, call number, etc.
- Could we build a lookup table from a script for everything in the Repo?

- - If I don't want to recreate something that already exists, for example?
    - A scheduled task (cron job) could maybe do something like that
  - Some simple searching of the github repo is fairly effective. (After searching, click on the "code" results to find individual .geojson index maps that contain your search terms)

Discussion of projections/coordinate systems field:
- Ultimately the geojson file should be WGS84 (4326)

Discussion of element name character length:
- 10 or less for recommended and required elements (this is a shapefile holdover)
- Some discussion about the 10 character rule… we should have 10 characters because "shapefile"

Discussion of Element name format & character length
- camelCase (underscores wastes a space)
- Should this be alphanumeric only?
- Added "latin alphanumeric characters only" for recommended and required elements. Other elements don't need this restriction

Bounding boxes:
- Long discussions have been had.
- How do we deal with multiple editions of the same map? One polygon for each map extent or air photo extent. If it's a multiple editions of the same map/same extent, there should be polygons/feature for each even if they overlap partially or totally
- Should we rename this something other than "Bounding Box" to avoid some confusion?
  - Yes… later? "Geometry" suggested.
- ***Geoblacklight can't currently recognize multiple polygons when you click on them, (like,... for example, if features overlap Esri will let you cycle through). We would like GeoBlacklight to have the ability to cycle through more than the top feature clicked on.
- Discussion of multipart polygons for inset maps… how do we know some of these things are all in the same sheet?
  - Main recommendation is multi-part polygon for these complex/multiple polygons pointing to same URL
- Data Entry: use decimal degrees stored as separate fields/columns. (this is a recommendation)
  - What if the maps you're representing are something else like state plane?
  - If you have a polygon/geometry field, why do we need this "ultimate bounding box"? Bounding box can extend beyond the geometry.
  - How are people using bounding box?
    - No one knows!

Discussion of Holdings Information:
- We have 3 holdings field as of now.

- ○ Available, physHold, digHold
- ○ physHold/digHold basically is our inventory
- ○ Available is used to color code the index in GBL (relying on boolean)
  - ■ TODO: check the red/green used by GBL for color blindness contrast

Discussion of Sheet/Frame number: none!

Discussion of Date:
- ● ISO 8601: (YYYY or YYYY-MM-DD)

Discussion of Websites: none

Discussion of Element Names:
- ● Label: special meaning in GBL used for mouse over effects
  - ○ Should these special fields be noted?
  - ○ Add to the description of this field that it is specifically used for the tool tip in GeoBlackLight
  - ○ Added labelAlt field over email discussion earlier.

Tom did a brief overview of suggested fields.
- ● **Added that bathLines should also be a boolean
- ● Should there be another note field that is for non-institutions-pecific notes?

Discussion of Union Catalog (Chris Thiry)

Chris Thiry is proposing a union catalog of indexes.
- ● We don't have time to catalog at sheet level, we catalog a set.
- ● What library has sheet J23 of this set?
- ● When user clicks on an index, pop up displays all who own it. Styled by color ramp that displays how many copies of a map is owned by other libraries.
- ● Click a country, view sets in popup, then go to the set index
- ● Three requirements:
  1. Do you own it?
  2. (I missed the rest)

Chris: is this useful? How difficult is this from a tech perspective? Discuss:
- ● Kim: noble goal but hard to implement.
- ● Chris: if you think it's worthwhile, I would try to get grant funding
- ● Keith: seems like all the data is there. We just need some way to identify who has all the same items
- ● Susan: could be useful for reference
  - ○ Editions are important

- ○ Is there some expectation that these need to be kept up to date if say a library gets rid of a sheet?
  - ○ Chris: Yes.
  - ○ Susan: not likely
- Chris's current sites are available here: http://libguides.mines.edu/maps/topos
  - ○ Some general discussion of how these index interfaces work
- Chris's map tools for librarians: http://libguides.mines.edu/maps/tools
- Chris's Clearinghouse site: http://csmgis.maps.arcgis.com/apps/PublicGallery/index.html?appid=044815b03f4c4e86bfb9f0d7ca0de44a&group=427f021a56f9449dbba24fbb4b915f55

USGS Geologic Map Database (flash): https://ngmdb.usgs.gov/ngmdb/ngmdb_home.html

OpenGeoMetadata

Notetaker: Eric Larson
Timekeeper:
Gatekeeper: Reina Murray

https://github.com/OpenGeoMetadata
How can we filter through OpenGeoMetadata?
What are Best Practices?
Should we share our metadata with custom fields or strip them out?

Some tools for pulling records from OpenGeometadata:
https://github.com/geoblacklight/geoblacklight/wiki/Metadata-tools-and-resources

https://github.com/geoblacklight/geoblacklight/wiki/GeoBlacklight-1.0-Metadata-Elements

OpenGeoMetadata (OGM): https://github.com/OpenGeoMetadata
GeoBlacklight (GBL): https://geoblacklight.org/

GitHub repo for stashing GBL metadata for other institutions to discover and import into their local installations.

Challenges: updating process, metadata consistency (institutional specific attributes).

Contributors present: UMN/BTAA, Stanford, NYU

Strategies for pulling data:
1. GeoCombine: harvest all of OGM or specific institution
2. SDRFriend: traverses a directory to generate CSV for those records -- helpful for metadata cleanup.
3. UNIX: cat to pull together all the data

Publishing frequency:
Stanford has not been updating data contributed -- workflows for improving frequency contributing are forthcoming. SUL has web services for metadata availability.

Harvesting goals:
Filtering for desired institutions, topics, public/restricted, etc.

Workflow for pushing and pull requests. Originally hoped to use github API features for filtering and searching for course filtering (ex. "Maryland" in a JSON file). OGM could use documentation, or a primer/tutorial on using GH API for these features.

BTAA Metadata Toolkit project could be a wrapper for GH API to help facilitate discovery, filtering, etc.

Jupyter Notebook is also a nice approach for pulling and visualizing documents.

Guidelines for contributing:
- Metadata about our metadata
  - Ex. UMN has keys for collections and origination
  - Helpful if keys are just numbers
  - B1G metadata schema documentation is helpful too
- Wishlist
  - Rights statements
    - Want percentage of available (public) works to be high
    - Restricted material creates a significant barrier
    - Maybe create directory structure for public and restricted -- should also be capable to filter via the GH API
  - Data Age
    - If data hasn't been updated frequently is less desired
  - Custom metadata
    - Shouldn't break the system (GBL)
    - Needs to be documented
    - Helpful to "see" practices at other institutions
    - Are they painful? Need to be cleaned?
      - Maybe GH API / filter might remove custom metadata
    - Should we formalize GBL schema to support custom fields?

- ● Boolean flag for custom attributes
- ● Current practice is namespaced attributes like b1g_ or nyu_
- ● Naming patterns for suffix-based dynamic fields could maybe be changed
- ● Single versus multiple value fields presents a challenge too -- ex. From BTAA is dc_publisher_sm (multi-value), where GBL is dc_publisher_s (singular), language is another example.
  - ■ Ingest and normalization scripts are often localized for institutional need too
  - ■ Who uses what controlled vocabulary, too
    - ● LCSH versus ISO
- ○ GBL Schema
  - ■ Metadata schema mapping to controlled vocabulary
  - ■ Certain fields where everyone is on the same pattern
  - ■ Helps to ensure ingest and discovery (faceting are sane)
  - ■ GBL schema tension between ease of data-creation versus semantic sharing and richness
  - ■ GBL was originally designed for discoverability
- ○ What steps are in consuming?
  - ■ What do I have?
  - ■ What needs to be changed? (normalization, enrichment)
- ○ What steps happen in OCLC/MARC copy-cataloging?
  - ■ Local changes do happen
  - ■ Local call numbers
  - ■ OCLC model would require everyone using ONE basic description record to be shared amongst everyone
- ○ Suggestion: use the README file in GitHub to document schema, usage, vocabularies
  - ■ GBL schema element documentation in GH is maybe a model, too
  - ■ BTAA repo / https://github.com/OpenGeoMetadata/big-ten
    - ● Potential example to follow
  - ■ NYU repo / https://github.com/OpenGeoMetadata/edu.nyu
- ○ Social Explorer records
  - ■ Document for community use
  - ■ With appended holdings
  - ■ Policy Map records
- ○ Long-term goals for OGM?
  - ■ What is good/worthy effort to support next 10 or 15 years
  - ■ ISO or FDGC record should be that
  - ■ More and more reliance on GBL record for sharing
    - ● Should GBL schema become a standard?
- ○ Schema design and Simplicity
  - ■ FDGC, ISO, and MARC

- - - Wanted to share records between multiple institutions
        - Took DC and some Geo fields to make original schema
        - Was designed for discovery and ease of contribution
        - Is there a movement towards DCAT?
            - DCAT as data model
            - DCAT as an endpoint instead of the data model
            - Nobody wants to generate and write DCAT
        - How does GBL schema change?
            - Governance
    - Any blockers for acceptance?
        - Extra steps to find data
        - Clarified update frequency

Todos:
- Flesh out https://github.com/OpenGeoMetadata/metadatarepository README as a template for other institutions
    - Details on pair-tree directory pattern https://datatracker.ietf.org/doc/html/draft-kunze-pairtree-00#section-1
    - Is layers.json still used or useful?
    - Open pull request for contributions template https://github.com/OpenGeoMetadata/metadatarepository/pulls
- Discuss GBL schema changes

Topics:

- Filtering records from OpenGeoMetadata
- Protocol for updating records when the schema changes

GeoBlacklight Governance and Roadmap

Roadmap draft

Btaa roadmap/contract for next year
- Want to be semi sure things will get pulled back upstream
Deciding GBL features/plugins/etc
- Let's document/formalize how these interactions (should) work
- Core values of the software to guide the decisions that get made wrt above
Prioritization for code sprints
Handling additions/changes to the metadata schema

- For example, if there are multiple institutions with similar custom fields

Two years ago code of conduct around contributions developed via short term interest group. This model could be applied to adopt core set of values, formalize decision process for

Eliot: is there an example from other OS comms?

Jack: maybe Ember? BL does not have a great expression of this

Ryan: we don't want to limit contribs by those with certain level of involvement, but should not be a free for all

Build a highlight legend, color coded to see roadmap features?

Similar to Apache's incubator process? Each proj has its own repo. Going from a proposed proj to an official project.

Ryan: Custom code for just one project is not desirable

Perhaps creating a running list/separate document for potential new features or more generically a summary of discussion

Comparative roadmaps for the different instances (Karen has a draft)

Jack: With Surfliner coming online, a chance to test out a new approach to "onboarding" contributors

Mara: How can we account for or otherwise address implementers of GBL who do not (currently) participate in the community.

Karen: It would def be helpful to hear about ongoing development (current list of instances at https://geoblacklight.org/)

Jack: We should add an agenda item to meetings about engaging folks not in the room (or on the calls). Potential outreach efforts, etc.

Amy: Setting aside time to discuss where we are? Separate occasional discussion open to anyone where others could be invited (rather than the down n' dirty month calls)

Karen: What about the monthly calls don't work for this?

Amy: the existing agenda, trying to fit, having basic questions. Onboarding type concerns. Could be envisioned as a companion to Geo4Lib

Karen: Monthly meetings are pretty unstructured at the moment

Eric: Folks frequently just talk about what each institution is working on. Not so much discussion about smaller scale talk/visioning.

Susan: I like the monthly reminders about the community call!

Ryan: Is there interest in a short term group to study values
- Jack, Eliot, Karen, Diane, Kara, Huda volunteered
- And Paul volunteered at the wrap up!

Jack: Eric's talk from a few years ago

Eric: (Github Link) Dev'ing the BTAA project, and how to weigh customization decisions

Jaime: It was helpful to spend a couple days working with Eric to hear about the maintenance and ongoing work with GBL.
Tom: Lots of things we want were already done, but are not part of the code base.
Eric: Good example was taking Eliot's work with thumbnails, adding some and throwing it back to the community.
Diane: tracking plugin adoption might help maintainers
Jack: maintenance, bug fixes, migration is the biggest obstacle for plugin development and longtime support. Modeling that complexity may be useful.

Karen: Usability meet up in the kitchen around 3

Relationships are hard (between data)

1. How to deal with duplicates / multiple copies from institutions
   ○ Is it possible to have the system recognize and group duplicates? Where would this sit in the metadata?
   ○ Karen: Don't most scanned maps have MARC tied to them? Would it be possible to add the OCLC members ?
   ○ Mara: Value in difference between scanned maps, maybe ok to group but not remove? Wouldn't you want to know another copy existed?
   ○ Licensed data would be another case to handle compared to maps
   ○ Paul: Would be great to also know when others have done extra work like cataloguing individual atlas sheets

2. How to represent relationships between data. Ex. Atlas with two volumes, but the entire Atlas is also part of a larger collection etc. Lack of metadata fields in GeoBlacklight to describe these levels
   ○ Source/derivative is what exists now. But maybe there are core relationships that should exist? Cautionary note about taking this too far
   ○ Eliot: Generally similar metadata between atlases and individual sheets. But having item level description allows patrons to find these, download individual sheets and direct link. They have suppressed those records in the search though. Only the parents record shows up
     i. Eg: see item #3 in search result
        https://maps.princeton.edu/?page=1&q=new+jersey+sanborn&search_field=all_fields&utf8=%E2%9C%93
     ii. And the item page, see child records on side:
        https://maps.princeton.edu/catalog/princeton-dv13zx13b

- ○ Kara: Scholars Portal supports this by aggregating records so that the children know who their parent is and metadata in the child record only will pull up the parent record
    - i. Note: possible GeoBlacklight feature request?
- ○ What are good practices for representing this in the current schema though? IS it adequate (seems no) or should this be part of the expansion of the schema?
- ○ NYU is using the DCT relations field. The app recognizes if you put the slug (unique value the GB uses to display a record in its catalogue view) of the parent in the child it makes the connection automatically. In dc:source
- ○ Andrew: best practices with the slug of child records is to use a hyphen with additional identifier after the identifier for the parent
- ○ Others are using `dct_isPartOf_sm` (which displays in GBL as "Collection") to make this connection. But some have differents needs for collection that are meaningful to the library.
- ○ Maps and Atlas causing a lot of problems.
- ○ Maybe if you have a really good record for each sheet, it doesn't matter as much? You likely wouldn't want to suppress in this case as well
- ○ Keith: what if when an indexmap record gets indexed an extra process is run to create another SOLR record that would copy stuff from the main records? May be helpful in some cases
- ○ Spatial search based on bounding box for now. How do you spatially rank the results in that case ? Another thing to test / consider
- ○ Mara: more useability testing! PRevious tests have found users have trouble distinguishing records and don't use the spatial search often
- ○ Paul: Harvard's system. The parent catalogue record has all the layers linked to it using a custom side database. Also thinking about how to represent this in GeoBlacklight
- ○ Eric: Maybe there are some solr features (field collapsing / result grouping) that aren't currently shipped that could be helpful
- ○ Should flag this as an enhancement in the github repo or explain this on the wiki at least so the community is aware of what others are doing
- ○ An issue like this might be good for a code sprint as well
- ○ Also part of the metadata conversation esp. as records are shared.

- ○ Mara: is anyone modelling relationships between scanned maps / vectorization? Most people plan to use source to model this, so that's a possible use to represent derivatives instead of parent/child relationships
- ○ WHat exactly is a parent child relationship is another question. Part/whole vs. derivative. How do users think of it / what is the use case is very important.
- ○ Karen: needed to track collections, and so started making parents records. Use dct_isPartOf_sm to indicate collection code. An entity, not a link.
- ○ Keith: collection search also allows you to see siblings which may be helpful.

- ○ Will the development on multiple download solve some of the issues when it comes to Atlases in particular?
- ○ How does the application treat these properties (dc is part of and dc source) differently?
    - i. Dc_source is used to identify the source dataset for a derived dataset, and when viewing the source page, a separate query is made against Solr (through the relations controller) that looks for all derived datasets and then lists these datasets
    - ii. Is part of shows up as a field/facet on the item record page, so if a record is identified as part of a collection, the field/facet link would show that value
- ○ What about for updated datasets / letting the user know a more updated version is available? Should there be a space for that in the metadata? Would this require updating all past metadata?
    - i. This could be presented by taking things with the same title and combining them eg. Streets. Click on and expands to all years
        1. Functionally you can do this using source
    - ii. Or you could suggest that other years exist when one is selected?
    - iii. Like amazon! Algorithm, not hardcoded eg. "you may also like.." (privacy issues..)
        1. There is a solr feature looking at metadata similarities
- ○ No one has a hard and fast rule for geodatabases, some have experimented

Summary
- ● [Issue 547](#) in git might be a good place to start.
    - ○ Andrew noted that the work-in-progress on multiple downloads may help us assign many different relationships from a record
- ● How easy it will be to manipulate the records post-production will be a big factor in any solution
- ● Might be instructive to look at DCAT ontology for describing relationships to create a list of relationships that work for our community
- ● Dublin Core: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
- ● DCAT: https://www.w3.org/TR/vocab-dcat-2/ (scroll down to qualifiedRelation)
- ● Might be worth a discussion about how the community is using the metadata fields currently esp. source/derived vs parent/child. Guidance on use and labelling.
    - ○ Separating these is very important esp. Is there is going to be different functionality for parent/child vs. source/derivative
- ● Might also be worth exploring if we care about volumes etc. - yes (just nest!)

Next steps: designate an upcoming community call to discuss this (March)
- ○ Look at examples, schema elements
- ○ Making some fixtures (sample metadata) to address various ways of handling this

- Explore the idea of having the child metadata in search (i.e. enable a user query which matches child metadata but which shows the parent record in the result), but not having the child records show up (as an option)
- Note to bring up switching use of source and ispartof in the schema (see DC)

Geo Repos & Where's the Data?

Notetaker: James Griffin
Timekeeper: Kean
Gatekeeper: (All)

- Setting up geospatial data repositories
  - What are people doing?
  - Earlier in the sessions, there were shared structural documents outlining where assets were stored
- What are some of the human components?
  - Skill-based assessment of those who might be hired?
- Physical requirements for hardware
  - Should this lie within an existing digital repository?
- Most attending this have long-term plans for implementing a spatial repository
  - There is a near-term spatial repository
    - Nearline
    - GeoServer
    - Online Exhibits
  - ...and there are long-term solutions as well
    - DSpace
    - Samvera
  - Repositories can be agnostic buckets
    - Derivative generation and the services involved in serving spatial visualizations are components of the Spatial Data Infrastructure
  - Preservation
    - Preservation copies at the Stanford Digital Repository are submitted to a remote storage system
  - Organizations do *not* need to implement a local GeoBlacklight instance
    - They can submit to OpenGeoMetadata and have GeoBlacklight adopters index the metadata on your behalf
    - Links for download are required, but they need not been previewed immediately
- ArcGIS for Server
  - ArcGIS Online can have a ZIP file registered
  - Provide WMS and WFS endpoints

- ○ Costs of storage and bandwidth with Esri are negligible
- Common Patterns for Spatial Data Infrastructures
  - ○ Samvera did attempt to support a common solution
    - ■ TLDR; Wait for Hyrax 3.0.0 and see how everyone can return to this
    - ■
  - ○ Princeton does support synchronization between a custom Samvera repository and GeoBlacklight
    - ■ This is highly customized and cannot really be shared (Princeton can improve documentation)
  - ○ Kubernetes
    - ■ No support for this yet at the community level
    - ■ Many Samvera adopters are using Docker, but these are local to the institutions
    - ■ James will refer to the Infrastructure Working Group in Samvera (this was newly formed)
- Troubles in working with GeoBlacklight in Spatial Data Repositories
  - ○ Identifying an ingestion workflow
  - ○ Where are ARKs minted more generally?
  - ○ Some want to ensure that systems are built to be as flexible as possible
  - ○ Customized Workflows
    - ■ These vary from different environments
    - ■ GeoBlacklight is limited to discovery only
    - ■ Workflows are more manual in Stanford and a bit more manual
    - ■ Workflows at Princeton are highly automated, and ensuring that this is generalizable is quite hard
- Is Anyone Doing Dockerization?
  - ○ Last year there was an ingestion tool which was built using Docker-based solutions
  - ○ Hard-coded URLs in metadata and Dockerfiles create network connectivity issues
  - ○ This might be best resolved by considering Kubernetes
  - ○ Standardization for Surfliner between UCSD and UCSB may encounter these issues
- GeoBlacklight and Spatial Data Repository Developer Skillsets
  - ○ GIS experience is not required (one major contributor does not have this as a formal background - they are a Rails developer)
  - ○ How much labor (in terms of developer time and number of developers) is required for a spatial data repository?
    - ■ This highly depends upon the scope of the project
      - ● Stanford uses automatic pipelines and maintenance tasks
      - ● Time invested up-front in this is significant, but it ensures that the maintenance is limited moving ahead
  - ○ What is the ideal team for GeoBlacklight?

- - - Community has seen an individual GIS librarian get GeoBlacklight running
      - Andrew Battista (NYU) is someone who was not a software developer by training who managed to successfully run and deploy this
  - UC Berkeley
      - One .NET developer found it very quick to learn and work with GeoBlacklight
  - Remote developers also ensure that flexibility is given and that recruitment can more effective
- Communicating with Developers
  - NYU has wonderful [documentation](#) for discussing the process of implementing a custom GeoBlacklight installation
  - Value of working with developers with a GIS background is present in the exchanges which are found between users, subject domain experts, and developers
- Agile Development
  - Ticket items which are created and placed on a Kanban board
  - Often times ticket items warrant some discussions for the context and detail for each item
      - Geospatial capacity personnel and developers are needed to identify which tickets are absolutely essential for releasing a GeoBlacklight customization into production
  - GeoBlacklight offers design process documentation detailing precisely which steps were taken
      - https://geoblacklight.org/documents/GeoBlacklight%20Concept%20Design%20v0.3.3.pdf
- Long Term Preservation
  - How is long term preservation being addressed?
  - With data served out of GeoServer, a range of 5 years might be enough for some
      - When is preservation needed in the longer term?
  - Stanford
      - Researchers are producing data sets for publication, which require a DOI, and these are then readily accessible on the WWW
      - This requires a long-term preservation strategy
  - Princeton
      - Cloud-based storage strategy uses Google Cloud Storage
      - Most resources are uploaded into a storage bucket, the BagIt format is used for these backups
          - https://en.wikipedia.org/wiki/BagIt
      - Offers fixity checking (Princeton ensures that the content and metadata for each ingested resource are not corrupted)

- - - ○ This is scheduled and performed periodically
    - Cloud-based strategies are becoming more affordable
      - ○ Nearline and Coldline offer better pricing options for content which is infrequently accessed
  - ■ Carbonite founders have developed Wasabi
    - https://wasabi.com/
    - This is relatively affordable, and offers an API compliant with that of the Amazon S3 API
  - ○ Derivative Generation
    - ■ CSU needs to have spatial content from DSpace served through GeoServer
      - Would the content need to be duplicated?
    - ■ Yes, but there are some options for processing the data before it is ingested into GeoServer
    - ■ Stanford reprojects GeoTIFFs, raster and vector datasets into the WGS84 datum
      - This is expensive when done en masse, and these are addressed by using asynchronous jobs
    - ■ Inspecting the data might be necessary
      - Styling visualizations of data can be helpful
      - Part of the accessioning process includes the analysis of the data sets
  - ○ Handling Anomalous Formats
    - ■ GeoServer does not handle geodatabases at all
    - ■ One can extract individual layers for individual ingestion…
      - ...but is one modifying the dataset when this is done?
      - Functionality is definitely lost for some of these cases
    - ■ Shapefiles and GeoTIFFs are the trivial cases
  - ○ Data Versioning Policy
    - ■ This also introduces complexity
      - Automating the workflow saves effort and time with this
      - New versions for the same dataset can then be handled
  - ○ **Request for Princeton to provide a video for accessioning spatial content**
  - ○ Stanford does feed back into the Stanford Digital Repository from the spatial data infrastructure
    - ■ OEmbed requires this
- Code Base Names
  - ○ Difficulty of onboarding
  - ○ Samvera could do better with making code bases more transparent
- Subsets of Geospatial Datasets
  - ○ User wants a collection of tiles which lie within a region
    - ■ Not the entire set of tiles
  - ○ Multidimensional raster data sets

- - ■ How are users provided with the data which are needed?
      - ■ Similar to clip-and-ship
      - ■ Cloud Optimized GeoTIFFs may well address this
        - ● Support for these within GeoServer is limited to S3 in release 2.1.7
        - ● Cornell ~~was supporting this~~ is exploring this
      - ■ How is a portion of that data set provided?
      - ■ This becomes a modeling challenge
        - ● With a 500GB GeoTIFF, will using this to then request a small area of the dataset best?
      - ■ Providing clip-and-ship for datasets could potentially lead to recreating features of an actual GIS application
        - ● GeoBlacklight avoids this by linking out to CartoDB
      - ■ This remains a feature which is in high demand
        - ● This is largely how new vector data sets are being created
  - ● [Overview of SDI Technology Stacks](#)
    - ○ Big 10 Academic Alliance has a Google Drive where documentation can be found outlining their findings from other institutions
      - ■ It would be excellent if this could be updated and expanded upon for reference in the community
      - ■

Usability Study Group

Who is interested in developing a cross platform GeoBlacklight usability study?
- ● Karen
- ● Mara Blake
- ● Amanda
- ● Phil
- ● Susan
- ● Tom
- ● Reina

BTAA Interface Committee is doing a usability study in May that will be fairly simple interviews

We want (later in the year, probably in early fall) to do a cross platform study; compare GeoBlacklight portals

March 31 - timeline for shoreline to get their MVP up; some basic functionality; T-Test

Revisit the personas; match personas to features

Have an online meeting in May for initial planning and coordination

IRBs - may need to get permissions early, but we can still follow through by testing all the sites at each institution

Building Data Ingest relationships with other geoportals

Two components to frame the discussion
1. Technical aspects
   a. The "hows", technology matters
2. Human relationships?
   a. Defining and acknowledging authoritative copies -- what are your experiences?
   b. Ways of sharing without duplicating
   c. Who should share/host what - e.g., NAIP imagery
   d. Everybody has their own "thing" they want to promote = decentralization
      i. Should I do X, just because I can?
   e. Ultimate goal is to simplify life for the data seeker

Wisconsin Model
- Archiving at the library. Snapshots over time when the data will disappear, the Library will still have it.
- Referencing back to the source data. Just pulling the metadata and making the data discoverable. Not actually harvesting the data.
- Does not have formal agreements with data providers. Initiative started in 2004 to archive certain data for research purposes.
- When communicating w/data providers and researchers, emphasize that the rationale for collecting their geodata is similar to collecting their paper maps.
- Not legalling binding. The University is not legally responsible for hosting, collecting data from the agencies. They can turn it off, and nothing will happen.
- UW communicates with open data portals they harvest from. Emphasize that they are not trying to replace their portals, rather they can drive traffic to them

UC Boulder has the GBL instance in CO, there is a drive for coordinating efforts towards statewide data collection. Divide and conquer for the types of data being grabbed by snapshots.

Berkeley (old, from 2015) form letter to send to county governments:
https://drive.google.com/file/d/1xzUFkv8RMJWgQSiUsKdEToj--dT6a41a/view?usp=sharing

Ontario - many municipalities aren't currently sharing their data. Scholars Portal is currently harvesting metadata from open data portals on a bi-weekly basis, as well as taking annual snapshots of municipal data. As more data are being served through portals, there is more acceptance and willingness to share and allow for archiving.

Newer data are made available w/web services, metadata, etc. Older data (5 years) are made available w/o web services as download only (pre-packaged zip), but contain full metadata which is used for indexing.

Relationships with municipalities are formed at the local level by the nearest institution (ex. Western University has a relationship with City of London), so data from many smaller areas in Ontario are not captured or represented as those relationships don't exist.

UMinn - reasons for not having a collection development policy around geoData are due to not have any place to store the data, rights management, etc.…
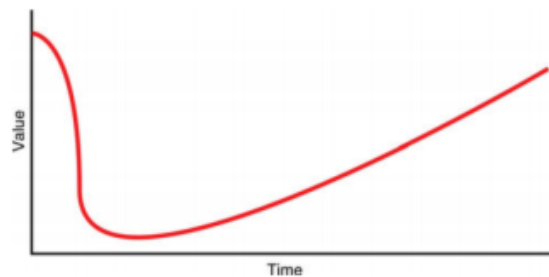
Perceived value of remotely sensed imagery over time

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

Cooperation amongst universities within a region?
Colorado - needs/wants to create statewide collection development policy
UC/Stanford campuses just beginning to develop collection, cataloging, and preservation strategies for digital content

Collection Development Policies
Wisconsin - ("hub of hubs") https://geodata.wisc.edu/pages/contribute

Important to consider keeping data separate from branding, interface. Multiple portals are still difficult for uses to search and find data. Having an 'academic' portal that curates useful data can provide a better user experience and supports preservation efforts.

Issues with Data Quality. Many data contain errors or anomalies. Sometimes results in take down requests. Open data portals contain many data that are 'open' but designated as 'not for public use.'

Minnesota Geospatial Advisory Group Archiving Workgroup:
http://www.mngeo.state.mn.us/workgroup/archiving/

Cloud-Optimized GeoTIFFs

"COG"

A type of GeoTIFF that is internally organized with tiles and overviews that allows quick access to subsets across the web using HTTP GET range requests. Which means clients can request just parts of the image, at a given scale.

Basic info and recipes for creating COGs at https://www.cogeo.org/

Chris Holmes' on COG:

- https://medium.com/planet-stories/cloud-native-geospatial-part-2-the-cloud-optimized-geotiff-6b3f15c696ed

A Handy Introduction to Cloud Optimized GeoTIFFs

- https://medium.com/planet-stories/a-handy-introduction-to-cloud-optimized-geotiffs-1f2c9e716ec3

Implementations:

- GDAL
    - `gdal_translate in.tif out.tif -co TILED=YES -co COPY_SRC_OVERVIEWS=YES -co COMPRESS=DEFLATE`
    - upcoming GDAL 3.1 will include a "COG" output format that will automatically apply the tiling and overviews
    - Related GDAL github work: https://github.com/OSGeo/gdal/pull/1600 https://lists.osgeo.org/pipermail/gdal-dev/2019-May/050169.html https://lists.osgeo.org/pipermail/gdal-dev/2019-May/050318.html
- QGIS
    - Add data > Raster > HTTP > paste URL to the .tif
    - Acts pretty much like a WMS service, but without needed to have a WMS server
- Esri can read/write (?)
    - Esri has a tool called OptimizeRaster that converts to https://github.com/Esri/OptimizeRasters/blob/master/Documentation/OptimizeRasters_UserDoc.pdf
- COG Explorer website
- Leaflet - https://github.com/GeoTIFF/georaster-layer-for-leaflet
- Azavea GeoTrellis - https://www.azavea.com/announcements/cloud-optimized-geotiff-support-in-geotrellis-2-0-release/

Costs?

- How would it compare to WMS serving costs? or full downloads?
- Serving from local servers vs cloud servers

Ability to clip-and-ship? Using QGIS, could extract by extent. Possible without GIS software?

Anyone have standard practices for whether or not to include overviews (pyramids) in TIFFs (even regular non-COGs) that you distribute?
- distribute "original file" however it had overviews (or not)
- selective overview levels
- overviews can add ~50% to the file size

Support for alpha bands

http://geotiff.io/ Javascript tools for processing raster in a web browser (not necessarily COGs, though)

Do we need Geoserver for COG previews in GeoBlacklight?

Needed to support COG in GeoBlacklight:
- dct_references to indicate COG
- COG should work like regular TIF previews, but perhaps that creates extra overhead
- probably better to use COG-specific viewer (Leaflet plugin "georaster layer")

Do HTTP range requests get cached?
For local serving, is it just a matter of enabling range requests?
Do COGs provide benefits even when just being used locally (like in QGIS)?

What is the vector equivalent of COG?
- vector tiles
- flatgeobuf

Tapalcatl:
https://medium.com/@mojodna/tapalcatl-cloud-optimized-tile-archives-1db8d4577d92

GBL DevOps & Support

- What do we want for maintenance staff to consider priorities?
- How is Change Management handled with respect to GeoBlacklight?
- Is there documentation around guidelines and best practices?

- After deploying GeoBlacklight, what are the best maintenance procedures
- For a few years now, the community hosts a Winter code sprint and a Summer code sprint
  - Commonly there is a release following the code sprint
  - Hence, there might be a need to schedule releases following code sprints
  - Many adopters are not on the latest release following these sprints
- Those on AWS have had to address operating system upgrades
  - AWS infrastructure support for EC2 instances can change
  - New releases for Solr can become burdensome also
  - Load balancing and other infrastructure challenges are not being thoroughly addressed given other more urgent priorities
  - Coordination with other parts of University IT can become necessary
    - NYU is looking to assemble a triage team in order to address this
- Leadership is often needed in order to effectively communicate the needs of teams looking to support GeoBlacklight to any affected parties
  - Establishing a team for addressing infrastructure issues is essential
  - Drafting proposals for administration has been helpful for some adopters
  - Well balanced team membership should be an ideal, where, if possible, support from systems administration/DevOps paired with development and user experience design can greatly improve the ease of success
- Approaches for undertaking upgrades for GeoBlacklight
  - Much of this depends upon institutional resources
  - Requesting virtual machines which are provisioned using Ansible Playbooks is one strategy
    - The Big 10 (U. Minnesota) and Princeton are using Ansible pretty heavily for server provisioning
    - Big 10 provisions a new machine for a new upgrade (using the code base for the new GeoBlacklight release), and then retires that the machine for the deprecated GeoBlacklight installation
  - Maintenance
    - It would be ideal if certain routine upgrades were well documented
    - For example:
      - Clearing search logs in the RDBMS regularly

- - - Clearing the /[RAIL_ROOT]/tmp regularly in order to ensure that requested downloads cached there is essential
  - Docker and Vagrant
    - Problem of maintaining parity requires that upgrades made in production need to be reflected into the virtualized environment
  - Community Need for Documentation
    - It would be wonderful if this were added to the GeoBlacklight Wiki
    - Details on production-related tips, issues, gardening, troubleshooting
  - Princeton:
    - Solr 7 dedicated on its own cluster config server
      - On and off production problems and intermittent failures
      - Testing Solr 8
    - Plug-in precarity: UTF8
  - Security certificates through campus IT
    - These processes are not automated, and require a human to update the certificate on each server at NYU and Princeton
    - Big 10 has had success with using Ansible to work with LetsEncrypt
    - Can AWS provide this service?
      - This can become costly, depending on the organization, but it's definitely possible
  - Monitoring
    - AppSignal
    - DataDog
    - Monit
    - Munin
  - Running Solr as SolrCloud vs. Solr alone
    - SolrCloud as had problems for Princeton
    - Using Solr as a lone, dedicated service
      - Requires quite a bit of memory
      - Isolation ensures that the system resources aren't exhausted with other applications
- Culture of Team Building
  - Big 10 and Princeton use Slack to communicate extremely effectively
  - Princeton uses Zoom paired with Slack in order to try and ensure that knowledge is shared for cases where pairing is essential
  - Cultural should shaped to avoid explicitly blaming anyone for errors or failures
    -
- Princeton uses incident reports when service outages occur
  - James will share this
- Bus Factor
  - JHU uses the "what if one of us wins $1,000,000 prize money" to ensure that knowledge is shared effectively
  -

Conference Plus / Delta

Plus:

+ Unplanned preconf preplanning
+ Pacing of sessions and breaks (wasn't too overwhelming)
+ Workshops on the first day
+ Workshops helped set the tone, was useful for a first timer
+ Indian catering for lunch over sandwiches
+ Covered a lot of topics
+ First time attendees

Delta:

- Another half hour of lightning talks in favor of a reduced session (+1)
- A few lightning talks scattered around
- Ambiguity with the first day proposals
- Lingo thing, brief overview of .. glossary collaborative
- Hearing from the portals that are up, what are major changes that have happened over the years
- Team introductions
- Take group picture earlier in the week or day to capture everyone

Thursday, February 6, 2020

Collection Development

Notetaker: Reina Murray
Timekeeper: Tom
Gatekeeper: Amy Work

Shared drive for collection development policies/plans:
https://drive.google.com/drive/folders/1cVvTyXmsnPKv4jFvVa_dUBnMNjZQBEJ8?usp=sharing

UC/Stanford Maps Common Knowledge Group (CKG)
https://ucsd.libguides.com/ld.php?content_id=3174735

Colorado GeoLIbrary Collection Development Plan
https://docs.google.com/document/d/1jfLNsN_Aziap8fXsOsUQYCeUXipHblq2CVrZJ_fv5OA/edit?usp=sharing

University of Maryland
https://www.lib.umd.edu/binaries/content/assets/public/collections/collection-policies/collection-policies-2018/subject_policy_geog.pdf

Yale Maps and Geospatial Collection Development Statement:
https://web.library.yale.edu/collection-development/maps-and-geospatial-information

Took 15 minutes to read through different policies and guidelines (see links above)

Discussion points:
- What is a collection development policy? Different emphases in the policies above. What are the core elements of such a policy?
  - Theme: being able to tie what you collect with what researchers on your campus need
  - NYU - tied to role and identity of a subject specialist; had a lot of discussion about whether this is a policy or statement. A way to set parameters to protect you from unsolicited gifts, bad actors/rogues, protect from collecting items that don't meet the need of research community. In policy, attempts to get at what does it take to collect and manage such a collection
    - See linkage to https://www.loc.gov/acq/devpol/cpc.html on https://guides.nyu.edu/c.php?g=777582&p=5577916 -- useful codification to follow
- Goal of a collection development policy -- to help define the parameters, provide guidance/explanation for decisions
- UC-Stanford Common Knowledge Group (https://ucsd.libguides.com/ld.php?content_id=3174735) -- to discuss/avoid duplication of efforts.
- Is a policy different from a strategy?
  - Policy -- what we're collecting and why
  - Strategy -- how we're collecting; the details
- If you have a new researcher who comes in with a new area of study -- do you incorporate that into your strategy/policy?

- How often do you revisit these policies?
  - BTA policy (https://drive.google.com/open?id=0Bz0NrVxoaPgaNGxmTEZpai1PNzdHMjA0Mjk4N0l3R0l2cVg0) - says every six months; new one says annually
- Do you have separate policies for your own vs others that you might be bringing in (e.g. open geometadata)?
  - Ty to distinguish between core vs ancillary (see Colorado's)
    - Establish priority
  - Some schools suppress proprietary data from other schools -- others include it
  - Having too many restrictions on policy would make it challenging for pulling in data from other institutions
- Like the Wisconsin one (https://geodata.wisc.edu/pages/contribute)
  - Also outlines all the services (collect, archive, make it discoverable, curate, etc)
  - Includes link to list of foundational layers
- Consider the audience/stakeholders
- Stanford has a lot extensively written on map collection, but not on geospatial data -- getting a lot from researchers covering range of areas and topics. Should that be formalized in policy?
  - What to do when you have a repository for research data?
    - Inclined to include it in policy to have it be a direct call out
- Separate/distinguish between map, geospatial data, etc in policies?
  - Can be beneficial to distinguish between the types of data in your policy
- Consider temporal scope as well as geographic -- see UC-Stanford
- To consider: explicitly stating/looking at/for data that is not heavily digitized/curated, even if it's outside of geographic scope -- "at-risk data"
  - Defining what's at risk/what's priority can be tricky but important
- What are people doing for imagery data? What is the scope?
  - Image consortium involving UCs -- have access going back historically for country, 3 images taken per year; given 1 aggregated snapshot for the year from NearMap which is allowed to be distributed out
  - BTA - did survey of what's the most important to collect -- imagery was the top vote; but hardest to collect and store
  - UC Santa Barbara -- in 80's given huge collection of imagery data; started with a large donated collected → now 2.5 million images. Running out of room now; not actively collecting. Strong geographic focus now.
- Consider need to discuss/involve other stakeholders -- eg: IT.
- Next steps:
  - Susan to send email to geo4libcamp group to follow up (see items below)
  - Follow up -- make draft plans, group review of collection plans
  - Also draft statement/put in opengeometadata? More guidance on what kind of metadata to submit and where/how.

Links of Interest

(Anything of interest but not directly related to daily sessions)

Google's Geo4Good 2020 Summit (Heavy on Earth Engine, but with Maps API, Earth Web, Earth Desktop, and other Google Geo platforms)

What: **Geo for Good Summit**

When: September 14-17, 2020

Where: Google Headquarters, Sunnyvale, California

This year's Geo for Good Summit is a special four-day conference and hands-on technical workshop at the Google headquarters in Sunnyvale, California from September 14-17, 2020. The Summit is intended for nonprofits, scientists and other change-makers who want to leverage technology and use mapping tools (such as Google Earth, Earth Engine, Maps, and Street View) for positive impact in the world.

Application Process: To ensure an effective hands-on experience, there will be a limited number of seats. Interested candidates will need to apply to attend the summit, with the competitive application process starting in April 2020.

Cost: Attendance is free. Accepted participants will be responsible for any travel expenses.

For a glimpse at the Geo for Good Summit experience, please see last year's [Summit Highlights video](#) and [website](#). Stay tuned for an official workshop announcement and invitation to apply in April. In the meantime, please feel free to pass along this save-the-date to potential participants, and encourage them to [sign up for our mailing list](#) to receive the workshop announcement.

[Google Earth Outreach](#)