

GeoBTAA Metadata Handbook

None

Table of contents

1. GeoBTAA Metadata Handbook	3
1.1 Who is this for?	3
1.2 Contents:	3
1.3 Version History	3
1.4 Credits	5
2. Reference	6
2.1 GeoBTAA Metadata Application Profile	6
2.2 Custom Elements	7
2.3 Local Input Guidelines	15
2.4 Editing Template	17
3. How-to Guides	18
3.1 How to submit resources to the BTAA Geoportal	18
3.2 Manually add bounding boxes	20
4. Explanation	21
4.1 Resource Lifecycle	21
4.2 Curation Priorities	23
4.3 BTAA Geoportal Collection Development Policy	25
4.4 Content Organization Model	27
4.5 Resource Classes	28
4.6 Supplemental Metadata	29
4.7 GEOMG Metadata Toolkit	30

1. GeoBTAA Metadata Handbook

This handbook describes how to curate metadata records for the [BTAA Geoportal](#).

[Download a PDF of this site here.](#)

1.1 Who is this for?

- Team Members in the [Big Ten Academic Alliance Geospatial Information Network](#) (BTAA-GIN)
- Development & Operations Staff in the BTAA-GIN
- Users & developers of open-source geospatial projects, such as [OpenGeoMetadata](#) and [GeoBlacklight](#)
- Contributors to the [BTAA Geoportal](#)
- Users of the [BTAA Geoportal](#)

1.2 Contents:

1.2.1 Reference

Information about the GeoBTAA Metadata Application Profile

1.2.2 How-to guides

Guidelines for identifying, submitting

1.2.3 Explanation

Descriptions and clarifications of processes, policies, and tools

1.3 Version History

Changes for Version 4.6 (March 15, 2023)

- New page for manually adding bounding boxes
- Restructure using Diataxis framework
- Remove some GEOMG how to guidelines (moved to GEOMG Wiki)
- Clarify Editing Template differences from OGM-Aaardvark documentation
- Added Collection Development Policy and Curation Priorities documents
- Update input guidelines for Spatial Coverage (FAST IDs)

Changes for Version 4.5.1 (February 28, 2023)

- Update version documentation
- Add link to generated PDF

Changes for Version 4.5 (February 28, 2023)

- Add Creator ID
- Update input guidelines for Creator, Creator ID
- Remove Harvesting Guide info (migrating to separate site)
- Edit Submitting Metadata page
- Minor copy editing
- Add PDF export capability

Changes for Version 4.4 (August 23, 2022)

- updated theme
- reorganized and expanded navigation menu
- new sections for Harvesting Guide and using GEOMG

Changes for Version 4.3 (August 15, 2022)

- migrate to MkDocs.org platform
- update bounding box entry guidelines
- add GEOMG page

Changes for Version 4.2 (March 24, 2022)

- New Entry and Usage Guidelines page
- Expands content organization model documentation
- Changes the name of the schema from 'Aardvark' to 'OpenGeoMetadata (version Aardvark)'
- Cleans up outdated links

Changes for Version 4.1 (Jan 2022)

- updates Status as optional; removes controlled vocabulary
- Clarifies relationship model

Changes for Version 4.0 (July 2021)

- Incorporation of GEOMG Metadata Editor
- Upgrade to Aardvark Metadata Schema for GeoBlacklight

Changes for version 3.3 (May 13, 2020)

- Added University of Nebraska
- Reorganized Metadata Elements to match editing template
- Updated the “Update the Collections” section to match new administrative process for tracking records

Changes for version 3.2 (Jan 8, 2020)

- Added Date Range element

Changes for version 3.1 (Dec 19, 2019)

- Added collection level records metadata schema

Changes for version 3 (Oct 2019)

- GeoNetwork and Omeka deprecated
- all GeoBlacklight records are stored in a spreadsheet in Google Sheets
- records are transformed from CSV to GeoBlacklight JSON with a Python script
- additional metadata fields were added for administrative purposes
- IsPartOf field now holds a code pointing to the collection record
- Administrative groupings such as “State agencies geospatial data” are now subjects, not a Collection
- updated editing templates available
- all supplemental metadata can be stored as XML or HTML in project hosted folder
- updated links to collections database

1.4 Credits

Handbook prepared by:

- Karen Majewicz, Geospatial Product Manager
- Ziyang (Gene) Cheng - Graduate Research Assistant


 March 16, 2023

2. Reference

2.1 GeoBTAA Metadata Application Profile

The GeoBTAA Metadata Application Profile consists of the following components:

1. OpenGeoMetadata Elements

- The BTAA Geoportal uses the OpenGeoMetadata Schema for each resource.
- The current version of OpenGeoMetadata is called 'Aardvark'.
- This lightweight schema was designed specifically for the GeoBlacklight application and is geared towards discoverability.
- The GeoBTAA Metadata Profile aligns with all of the guidelines and recommendations in the official OpenGeoMetadata documentation.
- The schema is documented on the [OpenGeoMetadata website](#) .

2. Custom Elements

- The GeoBTAA profile includes custom fields for lifecycle tracking and administration
- These elements are generally added to the record by admin staff. When they appear on editing templates, they are grayed out.
- They all start with the namespace `b1g`
- See the [Custom Elements page](#) for more detail

3. GeoBTAA Input Guidelines

- For the content in some fields, the GeoBTAA profile has specific guidelines that extends or differs from what is documented in the OpenGeoMetadata schema.
- See the [GeoBTAA Input Guidelines page](#) for more detail



The GeoBTAA Metadata Template can be found at <https://z.umn.edu/b1g-template>

2.2 Custom Elements

This page documents the custom metadata elements for the GeoBTAA Metadata Profile. These elements extend the [official OpenGeoMetadata \(Aardvark\)](#) schema.

b1g-id	Label	URI	Obligation
b1g-01	Code	<code>blg_code_s</code>	Required
b1g-02	Status	<code>blg_status_s</code>	Optional
b1g-03	Accrual Method	<code>blg_dct_accrualMethod_s</code>	Required
b1g-04	Accrual Periodicity	<code>blg_dct_accrualPeriodicity_s</code>	Optional
b1g-05	Date Accessioned	<code>blg_dateAccessioned_s</code>	Required
b1g-06	Date Retired	<code>blg_dateRetired_s</code>	Conditional
b1g-07	Child Record	<code>blg_child_record_b</code>	Conditional
b1g-08	Mediator	<code>blg_dct_mediator_sm</code>	Conditional
b1g-09	Access	<code>blg_access_s</code>	Conditional
b1g-10	Image	<code>blg_image_ss</code>	Optional
b1g-11	GeoNames	<code>blg_geonames_sm</code>	Optional
b1g-12	Publication State	<code>blg_publication_state_s</code>	Required
b1g-13	Language String	<code>blg_language_sm</code>	Required
b1g-14	Creator ID	<code>blg_creatorID_sm</code>	Optional

2.2.1 Code

Label	Code
URI	<code>blg_code_s</code>
Profile ID	b1g-01
Obligation	Required
Multiplicity	1-1
Field type	string
Purpose	To group records based upon their source
Entry Guidelines	Codes are developed by the metadata coordinator and indicate the provider, the type of institution hosting the resources, and a numeric sequence number. For more details, see Code Naming Schema
Commentary	This administrative field is used to group and track records based upon where they are harvested. This is frequently an identical value to "Member Of". The value will differ for records that are retired (these are removed from "Member Of") and records that are part of a subcollection.
Controlled Vocabulary	yes-strict
Example value	12d-01
Element Set	B1G

2.2.2 Status

Label	Status
URI	<code>blg_status_s</code>
Profile ID	blg-02
Obligation	Optional
Multiplicity	0-1
Field type	string
Purpose	To indicate if a record is currently active, retired, or unknown. It can also be used to indicate if individual data layers from website has been indexed in the Geoportal.
Entry Guidelines	Plain text string is acceptable
Commentary	This is a legacy admin field that was previously used to track published vs retired items. The current needs are still TBD.
Controlled Vocabulary	no
Example value	Active
Element Set	BIG

2.2.3 Accrual Method

Label	Accrual Method
URI	<code>blg_dct_accrualMethod_s</code>
Profile ID	blg-03
Obligation	Required
Multiplicity	1-1
Field type	string
Purpose	To describe how the record was obtained
Entry Guidelines	Some values, such as "ArcGIS Hub" should be entered consistently. Others may be more descriptive, such as "Manually entered from text file."
Commentary	This allows us to find all of the ArcGIS records in one search. It also can help track records that have been harvested via different methods within the same collection.
Controlled Vocabulary	no
Example value	ArcGIS Hub
Element Set	BIG/ Dublin Core

2.2.4 Accrual Periodicity

Label	Accrual Periodicity
URI	<code>blg_dct_accrualPeriodicity_s</code>
Profile ID	blg-04
Obligation	Optional
Multiplicity	0-1
Field type	string
Purpose	To indicate how often a collection is reaccessioned
Entry Guidelines	Enter one of the following values: Annually, Semiannually, Quarterly, Monthly, As Needed
Commentary	This field is primarily for collection level records.
Controlled Vocabulary	yes-not strict
Example value	As Needed
Element Set	B1G/ Dublin Core

2.2.5 Date Accessioned

Label	Date Accessioned
URI	<code>blg_dateAccessioned_s</code>
Profile ID	blg-05
Obligation	Required
Multiplicity	1-1
Field type	string
Purpose	To store the date a record was harvested
Entry Guidelines	Enter the date a record was harvested OR when it was added to the geoportal using the format yyyy-mm-dd
Commentary	This field allows us to track how many records are ingested into the portal in a given time period and to which collections.
Controlled Vocabulary	no
Example value	2021-01-01
Element Set	B1G

2.2.6 Date Retired

Label	Date Retired
URI	<code>blg_dateRetired_s</code>
Profile ID	blg-06
Obligation	Conditional
Multiplicity	0-1
Field type	string
Purpose	To store the date the record was removed from the geoportal public interface
Entry Guidelines	Enter the date a record was removed from the geoportal
Commentary	This field allows us to track how many records have been removed from the geoportal interface by time period and collection.
Controlled Vocabulary	no
Example value	2021-01-02
Element Set	B1G

2.2.7 Child Record

Label	Child Record
URI	<code>blg_child_record_b</code>
Profile ID	blg-07
Obligation	Optional
Multiplicity	0-1
Field type	string boolean
Purpose	To apply an algorithm to the record that causes it to appear lower in search results
Entry Guidelines	Only one of two values are allowed: true or false
Commentary	This is used to lower a record's placement in search results. This can be useful for a large collection with many similar metadata values that might clutter a user's experience.
Controlled Vocabulary	string boolean
Example value	true
Element Set	B1G

2.2.8 Mediator

Label	Mediator
URI	<code>blg_dct_mediator_sm</code>
Profile ID	blg-08
Obligation	Conditional
Multiplicity	0-0 or 1-*
Field type	string
Purpose	To indicate the universities that have licensed access to a record
Entry Guidelines	The value for this field should be one of the names for each institution that have been coded in the GeoBlacklight application.
Commentary	This populates a facet on the search page so that users can filter to only databases that they are able log into based upon their institutional affiliation.
Controlled Vocabulary	yes
Example value	University of Wisconsin-Madison
Element Set	B1G/ Dublin Core

2.2.9 Access

Label	Access
URI	<code>blg_access_s</code>
Profile ID	blg-09
Obligation	Conditional
Multiplicity	0-0 or 1-1
Field type	string JSON
Purpose	To supply the links for restricted records
Entry Guidelines	The field value is an array of key/value pairs, with keys representing an institution code and values the URL for the library catalog record. See the Access Template for entry.
Commentary	This field is challenging to construct manually, it is a JSON string of institutional codes and URLs. The codes are used instead of the actual names in order to make the length of the field more manageable and to avoid spaces.
Controlled Vocabulary	no
Example value	<code>{\"03\": \"https://purl.lib.uiowa.edu/PolicyMap\", \"04\": \"https://www.lib.umd.edu/dbfinder/id/UMD09180\", \"05\": \"https://primo.lib.umn.edu/permalink/f/1q7ssba/UMN_ALMA51581932400001701\", \"06\": \"http://catalog.lib.msu.edu/record=b10238077~S39a\", \"07\": \"https://search.lib.umich.edu/databases/record/39117\", \"09\": \"https://libraries.psu.edu/databases/psu01762\", \"10\": \"https://digital.library.wisc.edu/1711.web/policymap\", \"11\": \"https://library.ohio-state.edu/record=b7869979~S7\"}</code>
Element Set	B1G

2.2.10 Image

Label	Image
URI	<code>blg_image_ss</code>
Profile ID	blg-10
Obligation	Optional
Multiplicity	0-0 or 0-1
Field type	stored string (URL)
Purpose	To show a thumbnail on the search results page
Entry Guidelines	Enter an image file using a secure link (https). Acceptable file types are JPEG or PNG
Commentary	This link is used to harvest an image into the Geoportal server for storage and display. Once it has been harvested, it will remain in storage, even if the original link to the image stops working.
Controlled Vocabulary	no
Example value	https://gis.allencountyohio.com/GIS/Image/countyseal.jpg
Element Set	B1G

2.2.11 GeoNames

Label	GeoNames
URI	<code>blg_geonames_sm</code>
Profile ID	blg-11
Obligation	Optional
Multiplicity	0-*
Field type	stored string (URI)
Purpose	To indicate a URI for a place name from the GeoNames database
Entry Guidelines	Enter a value in the format "http://sws.geonames.org/ <code>URI</code> "
Commentary	This URI provides a linked data value for one or more place names. It is optional as there is currently no functionality tied to it in the GeoBlacklight application
Controlled Vocabulary	yes
Example value	https://sws.geonames.org/2988507
Element Set	B1G

2.2.12 Publication State

Label	Publication State
URI	<code>blg_publication_state_s</code>
Profile ID	blg-12
Obligation	Required
Multiplicity	1-1
Field type	string
Purpose	To communicate to Solr if the item is public or hidden
Entry Guidelines	Use the dropdown or batch editing functions to change the state
Commentary	When items are first added to GEOMG, they are set as Draft by default. When they are ready to be published, they can be manually changed to Published. If the record is retired or needs to be hidden, it can be changed to Unpublished
Controlled Vocabulary	yes
Example value	Draft
Element Set	B1G

2.2.13 Language string

Label	Language string
URI	<code>blg_language_sm</code>
Profile ID	blg-13
Obligation	Required
Multiplicity	1-*
Field type	string
Purpose	To display the spelled out string (in English) of a language code to users
Entry Guidelines	This field is automatically generated from the Language field in the main form
Commentary	The OGM schema specified using a 3-digit code to indicate lanuage. In order to display this to users, it needs to be translated into a human-readable string.
Controlled Vocabulary	yes
Example value	French
Element Set	B1G

2.2.14 Creator ID

Label	Creator ID
URI	<code>blg_creatorID_sm</code>
Profile ID	blg-14
Obligation	Optional
Multiplicity	0-*
Field type	string
Purpose	To track the URI of a creator value
Entry Guidelines	This field is entered as a URI representing an authority record
Commentary	These best practices recommend consulting one or two name registries when deciding how to standardize names of creators: the Faceted Application of Subject Terminology (FAST) or the Library of Congress Name Authority File (LCNAF). FAST is a controlled vocabulary based on the Library of Congress Subject Headings (LCSH) that is well-suited to the faceted navigation of the Geoportal. The LCNAF is an authoritative list of names, events, geographic locations and organizations used by libraries and other organizations to collocate authorized creator names to make searching and browsing easier.
Controlled Vocabulary	yes
Example value	fst02013467
Element Set	B1G

 March 16, 2023

2.3 Local Input Guidelines

For the following elements, the GeoBTAA Metadata Profile has input guidelines beyond what is documented in the OpenGeoMetadata schema:

Title

Maps: The title for scanned maps is generally left as it was originally cataloged by a participating library. MARC subfields are omitted and can be inserted in the Description field.

Datasets: Harvested datasets often are lacking descriptive titles and may need to be augmented with place names. Dates may also be added to the end, but if the dataset is subject to updates, the data should be left off. Acronyms should be spelled out. The preferred format for dataset titles is: `Theme [place] {date}`. This punctuation allows for batch processing and splitting title elements.

Language

Although [Language is optional in the OGM schema](#), a three-digit code is required for the BTAA Geoportal.

Creator

When possible, Creators should be drawn from a value in the [Faceted Application of Subject Terminology \(FAST\)](#).

Creator ID

If the Creator value is from a name authority, insert the ID in this field.

Publisher

Maps: Publisher values for maps are pulled from the original catalog record. Remove subfields for place names and dates.

Datasets: The BTAA Geoportal does not use the Publisher field for Datasets.

Provider

This is the name of the organization hosting the resources. If the organization is part of the BTAA library network, a university icon will display next to the resource's title. However, most Providers will not have an icon.

Spatial Coverage

This should be in the format used by the [Faceted Application of Subject Terminology \(FAST\)](#).

For US counties and cities, the format should be `state--county` or `state--city`. The state itself should also be included. Examples:

Example

- Wisconsin--Dane County
- Wisconsin--Madison
- Wisconsin

Bounding Box

On the Metadata Editing Template, provide Bounding Boxes in this format: **W,S,E,N** This order matches the DCAT API and is how the Klokkan Bounding Box provides coordinates with their "CSV" setting.

This format will be programmatically converted to other formats when it is published to the Geoportal:

- The OpenGeoMetadata Bounding Box field (`dcat_bbox_s`) uses this order: `ENVELOPE (W,E,N,S)`
- The OpenGeoMetadata Geometry field (`locn_geometry`) uses a WKT format and the coordinate order will be converted to this layout: `POLYGON ((W N, E N, E S, W S, W N))`
- The OpenGeoMetadata Centroid field (`dcat_centroid`) will be calculated to display longitude,latitude.

Example

Metadata CSV: **-120,10,-80,35**

converts to

```
dcat_bbox_s: ENVELOPE(-120,-80,35,10)
```

```
locn_geometry: POLYGON((-120 35, -80 35, -80 10, -120 10, -120 35))
```

```
dcat_centroid: "22.5,-100.0"
```

 March 16, 2023

2.4 Editing Template

The [GeoBTAA Metadata Template](https://z.umn.edu/b1g-template) (<https://z.umn.edu/b1g-template>) is a set of spreadsheets that are formatted for our metadata editor, GEOMG.

Users will need to make a copy of the spreadsheet to use for editing. In some cases, the Metadata Coordinator can provide a customized version of the sheets for specific collections.

The Template contains the following tabs:

- Map Template
- Dataset Template
- Website Record Template
- Values: All of the controlled vocabulary values for the associated fields.
- Access Links and Multiple Downloads: Fields for adding [secondary tables](#)



Note

The input format for some fields in this template may differ from how the field is documented in [OpenGeoMetadata](#). These differences are intended to make it easier to enter values, which will be transformed when we upload the record to GEOMG.

- **Bounding Box** coordinates should be entered as `W,S,E,N`. The coordinates are automatically transformed to a different order `ENVELOPE (W,E,N,S)`. Read more under the [Local Input Guidelines](#).
- **Date Range** should be entered as `YYYY-YYYY`. This is automatically transformed to `[yyyy TO yyyy]`.
- External links are added separately under column headers for the type of link. These are combined into the `dct_references_s` field as a string of key:value pairs.

3. How-to Guides

3.1 How to submit resources to the BTAA Geoportal

3.1.1 1. Identify Resources

It is the role of the Team members to seek out new content for the geoportal.



Places to find public domain collections

- State GIS clearinghouses
- State agencies (especially DNRs and DOTs)
- County or city GIS departments
- Library digital collections
- Research institutes
- Nonprofit organizations

Review the [Curation Priorities](#) and the [Collection Development Policy](#) for guidelines on selecting resources.

Optional: Contact the organization

Use [this template](#) to inform the organization that we plan to include their resources in our geoportal.



Tip

If metadata for the resources are not readily available, the organization may be able to send you an API, metadata documents, or a spreadsheet export.

3.1.2 2. Investigate metadata harvesting options

Metadata records can be submitted directly or we can harvest it using parsing and transformation scripts.

Here are the most common methods of obtaining metadata for the BTAA Geoportal:

Spreadsheets

This method is preferred, because the submitters can control which metadata values are exported and because format transformations by UMN Staff are not necessary. The [GeoBTAA Metadata Template](#) shows all of the fields needed for the Geoportal.

API Harvesting or HTML Parsing

Most data portals have APIs or HTML structures that can be programmatically parsed to obtain metadata for each record.

DCAT enabled portals

ArcGIS Open Data Portals (HUB), Socrata portals, and some others share metadata in the DCAT standard.

CKAN / DKAN portals

This application uses a custom metadata schema for their API.

HTML Parsing

If a data portal or website does not have an API, we may be able to parse the HTML pages to obtain the metadata needed to create GeoBlacklight schema records.

OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a framework that can be used to harvest metadata records from enabled repositories. The records are usually available as a simple Dublin Core XML format. If the protocol is not set up to include extra fields, such as the map image's download link or bounding box, this method may not be sufficient on its own.

Individual Metadata files

Geospatial metadata standards are expressed in the XML or JSON format, which can be parsed to extract metadata needed to create GeoBlacklight schema records. Common standards for geospatial resources include:

- ISO 19139
- FGDC
- ArcGIS 1.0
- MARC
- MODS



The best way to transfer MARC records is to send a single file containing multiple records in the .MRC or MARC XML format. The Metadata Coordinator will use MarcEdit or XML parsing to transform the records.

3.1.3 3. Contact the BTAA-GIN Product Manager

Send an email, Slack message to the Product Manager / Metadata Coordinator.

Minimum information to include:

- Title and Description of the collection
- a link to the website
- (If known) information about how to harvest the metadata or construct access links.

The submission will be added to our collections processing queue.



Metadata processing tasks are tracked on our public [GitHub project dashboard](#).

3.2 Manually add bounding boxes

3.2.1 Summary

This page describes a process for obtaining bounding box coordinates for our scanned maps. The coordinates will be used for indexing the records in the Big Ten Academic Alliance Geoportal.

3.2.2 About bounding box coordinates for the BTAA Geoportal

- Bounding boxes enable users to search for items with a map interface.
- The format is 4 coordinates in decimal degrees
- Provide the coordinates in this order: West, South, East, North.
- The bounding boxes do not need to be exact, particularly with old maps that may not be very precise anyways.

3.2.3 Steps for obtaining bounding box coordinates

Part A: Setup

1. Open and inspect the image file.
2. Try to identify a single / combined region that the map or atlas represents
3. You can also check to see if the map has the bounding coordinates printed in the text anywhere or you are able to find the bounds by inspecting the edges.
4. Open another window with the [Klokan Bounding Box](#) tool.
5. Set the Copy & Paste section to CSV.

Part B: Find the coordinates

OPTION 1: SEARCH FOR A PLACE NAME

1. Use the search boxes on the Klokan Bounding Box tool to zoom to the region. (For example, search for “Illinois”.
2. Manually adjust the grey overlay box in the Klokan site to line up the edges to the edges of the map.
3. Try to line it up reasonably closely

OPTION 2: DRAW A SHAPE

1. Switch to the Polygon tool by clicking on the pentagon icon
2. Click as many points on the screen as needed to approximate the map extent.
3. Click on the first point to close the polygon
4. The interface will display a dotted line showing the bounding box rectangle.

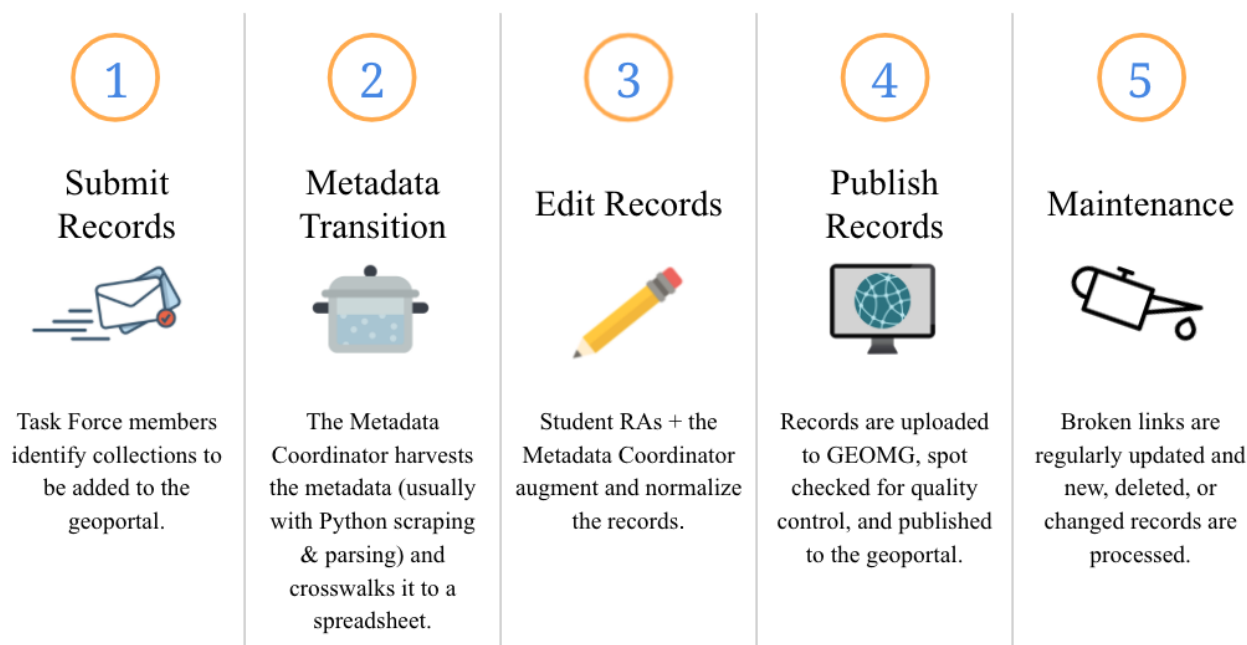
Part C: Copy back to GeoBTAA metadata

1. Click the “Copy to Clipboard” icon on the Klokan site.
2. Paste the coordinates into the Bounding Box field in the GeoBTAA metadata template or in the GEOMG metadata editor.

 March 7, 2023

4. Explanation

4.1 Resource Lifecycle



4.1.1 1. Submit Records

It is the role of the Team members to seek out new content for the geoportal. See the page [How to Submit Resources to the BTAA Geoportal](#) for more information.

4.1.2 2. Metadata Transition

This stage involves batch processing of the records, including harvesting, transformations, crosswalking information. This stage is carried out by the Metadata Coordinator, who may contact Team members for assistance.

Regardless of the method used for acquiring the metadata, it is always transformed into a spreadsheet for editing. These spreadsheets are uploaded to GEOMG Metadata Editor.

Because of the variety of platforms and standards, this process can take many forms. The Metadata Coordinator will contact Team members if they need to supply metadata directly.

4.1.3 3. Edit Records

Once the metadata is in spreadsheet form, it is ready to be normalized and augmented. UMN Staff will add template information and use spreadsheet functions or scripts to programmatically complete the metadata records.

- [The GeoBTAA Metadata Template](#) is for creating GeoBlacklight metadata.
- Refer to the documentation for the [OpenGeoMetadata, version Aardvark fields](#) and the [GeoBTAA Custom Elements](#) for guidance on values and formats.

4.1.4 4. Publish Records

Once the editing spreadsheets are completed, UMN Staff uploads the records to [GEOMG](#), a metadata management tool. GEOMG validates records and performs any needed field transformations. Once the records are satisfactory, they are published and available in the BTAA Geoportal.

[Read more on the GEOMG documentation page.](#)

4.1.5 5. Maintenance

General Maintenance

All project team members are encouraged to review the geoportal records assigned to their institutions periodically to check for issues. Use the feedback form at the top of each page in the geoportal to report errors or suggestions. This submission will include the URL of the last page you were on, and it will be sent to the Metadata Coordinator.

Broken Links

The geoportal will be programmatically checked for broken links on a monthly basis. Systematic errors will be fixed by UMN Staff. Some records from this report may be referred back to Team Members for investigating broken links.

Subsequent Accessions

- Portals that utilize the DCAT metadata standard will be re-accessioned monthly.
- Other GIS data portals will be periodically re-accessioned by the Metadata Coordinator at least once per year.
- Team members may be asked to review this work and provide input on decisions for problematic records.

Retired Records

When an external resource has been moved, deleted, or versioned to a new access link, the original record is retired from the BTAA Geoportal. This is done by converting the Publication State of the record from 'Published' to 'Unpublished'. The record is not deleted from the database and can still be accessed via a direct link. However, it will not show up in any search queries.

 March 16, 2023

4.2 Curation Priorities

 Authors: BTAA Collection Development & Education Outreach Committee; Product Manager

There are three distinct but related aspects of prioritizing the addition of new collections: content/theme, administration, and technology.

These priorities will affect how quickly the items are processed or where they fall in line within our processing queue.

4.2.1 Content/Theme

When it comes to scanned maps, prioritization based on content or theme is primarily a local effort. However, there are opportunities for internal collaborations, including with Special Collections librarians or other local digital collections initiatives. These collaborations can allow for unique and distinctive maps to be harvested into the geoportal across our universities.

For geospatial data, datasets created in association with research projects at our institutions may be a high priority based on content or theme. Additionally, resources that provide access to foundational datasets, such as administrative boundaries, parcels, road networks, address points, and land use, should also be considered.

Regardless of the content type, special consideration should be given to highly relevant content, especially to current events. For example, in April 2020, a call went out to all task force members to identify and submit content related to COVID-19 for harvesting into the geoportal. Content that aligns with other ongoing BTAA-GIN program efforts, such as the Diverse Collections Working Group, will also be a higher priority as these efforts are further developed.

4.2.2 Administration

Collections may be prioritized based on the organization responsible for creating and maintaining content, which impacts the types of maps or datasets available to be harvested, spatial and temporal coverage, and stability. Based on these considerations, current priorities in terms of administration are:

1. University libraries and archives

- Links to these resources are likely to be stable
- Resources will likely be documented with a metadata standard
- Represent our core audience

2. States and counties

- Produce most foundational geospatial datasets (e.g., roads and parcels) and are currently our largest source of geospatial data
- Technology and open data policies vary widely resulting in patchwork coverage

3. Regional organizations and research institutes

- Often special organizations with funding to create geospatial data across political boundaries
- Higher risk of harvesting duplicate datasets, as these organizations sometimes aggregate records from cities, counties, or state agencies

4. Cities

- less likely to produce and share data in geospatial formats and more likely to share tabular data
- prioritized cities: major metropolitan areas and the locations of our university campuses

4.2.3 Technology

The source's hosting platform influences the ease of harvesting, the quality of the metadata, and the stability of the access links. Based on these considerations, current priorities in terms of technology are:

1. Published via known portal or digital library platforms, including:

- Blacklight/GeoBlacklight
- Islandora
- Preservica
- ArcGIS Hubs
- Socrata
- CKAN
- Sites with OAI-PMH enabled APIs

2. Custom portals

- each portal requires a customized script for HTML web parsing
- writing and maintaining custom scripts takes extra time

3. Static webpages with download links

- at a minimum, a title is required for each item
- static sites with nested links take a long time to process and may require an extensive amount of manual work

4. Database websites

- require the user to perform interactive queries to extract data
- not realistic to make Geoportal records for individual datasets
- usually results in a single "website" record in the Geoportal to represent the database

 March 16, 2023

4.3 BTAA Geoportal Collection Development Policy

 Authors: BTAA Collection Development & Education Outreach Committee

4.3.1 Purpose

The BTAA Geospatial Information Network is a collaborative project to enhance discoverability, facilitate access, and connect scholars across the Big Ten Academic Alliance (BTAA) to scanned maps, geospatial data, and aerial imagery resources. The project's main output is the [BTAA Geoportal](#), which serves as a platform through which participating libraries can share materials from their collections to make them more easily discoverable and accessible to varied user communities. Collections within the Geoportal primarily support the research, teaching, learning, and information needs of faculty, staff, and students at participating institutions and beyond.

The project supports the creation and aggregation of discovery-focused metadata describing geospatial resources from participating institutions and public sources across the Big Ten region and makes those resources discoverable via an open source portal. For more information and a list of participating BTAA institutions, [please visit our project site](#).

4.3.2 Summary of Collection Scope

Access to the BTAA Geoportal is open to all. This collection consists of geospatial resources relevant to all disciplines. Access to resources is curated based on their authoritativeness, currency, comprehensiveness, ease of use, and relevancy. Materials included are generally publicly available geospatial datasets (vector/raster), scanned maps (georeferenced or with bounding box coordinates), and aerial imagery. Scanned maps protected by copyright are not included in the Geoportal. Access to licensed resources may be restricted to users affiliated with a participating institution.


1. Geographic areas: Items in the collection vary in scale based on subject and range from global to local. Geographic areas vary based on subject and may refer to biomes/ecosystems, political boundaries, cultural boundaries, economic boundaries, or land use types. In addition to a geographic focus on the Big Ten region (i.e., the states where participating institutions are located), the collection will emphasize resources and topics relevant to faculty and student research interests and reflect the strengths of participating library collections.
2. Time periods: All time periods are collected, with an ability to accommodate both current and historical versions of datasets.
3. Format: The collection consists of geospatial datasets, georeferenced maps, scanned maps with bounding box coordinates, and aerial imagery. Records for web mapping applications may also be included, with priority given to applications with datasets that are also accessible for download through the Geoportal. Preference is given to open and open source formats, but other formats are accepted as required to facilitate ease of use. When possible, resources are presented in formats that allow for download capabilities. Additional software may be needed to view datasets after download.
4. Language(s): The collection primarily consists of English language content. Some non-English language content may be available for certain regions, reflecting the collection strengths and research/curricular interests of participating institutions.
5. Diversity: The Geoportal and its participating institutions aspire to collect and provide access to geospatial resources that represent diverse perspectives, abilities, and experience levels. We will strive to apply [best practices for diverse collection development](#) as they relate to geospatial resources, including but not limited to:
 - a. considering resources from small, independent, and local producers
 - b. seeking content created by and representative of marginalized and underrepresented groups.
6. Preservation and life cycle: Digital file preservation for discovery metadata is managed by BTAA Geoportal staff. Digital file preservation for resources is the responsibility of the content provider. Resources may cease to be accessible through the Geoportal if access from the original provider is no longer available.

4.3.3 Statement of Communication

The members of the Geoportal project team will continue to communicate with the creators of other geoportals (GeoBlacklight Community, etc.), with data providers in our respective regions, and across Big Ten institutions to build a comprehensive and robust collection.

Implementation and Revision Schedule: This policy will be reviewed annually by the Collection Development & Education Outreach Committee and is considered effective on the date indicated below. It will be revised as needed to reflect new collection needs and identify new areas of study as well as those areas that may be excluded.

Updated: April 27, 2022

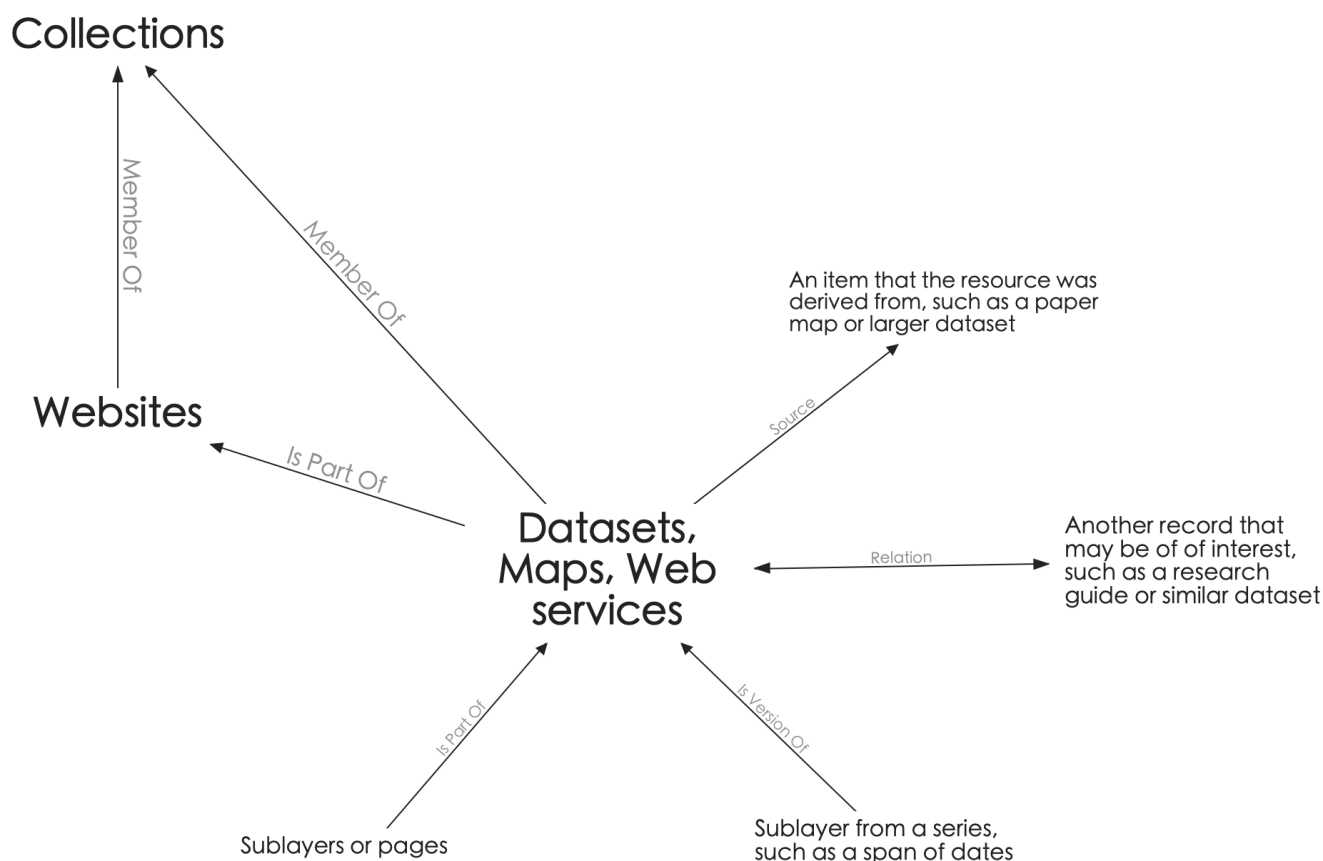
 March 16, 2023

4.4 Content Organization Model

GeoBlacklight organizes records with a network model rather than with a hierarchical model. It is a flat system whereby every database entry is a "Layer" and uses the same metadata fields. Unlike many digital library applications, it does not have different types of records for entities such as "communities," "collections," or "groups." As a result, it does not present a breadcrumb navigation structure, and all records appear in the same catalog directory with the URL of <https://geo.btaa.org/catalog/> ID.

Instead of a hierarchy, GeoBlacklight relates records via metadata fields. These fields include `Member Of`, `Is Part Of`, `Is Version Of`, `Source`, and a general `Relation`. This flexibility allows records to be presented in several different ways. For example, records can have multiple parent/child/grandchild/sibling relationships. In addition, they can be nested (i.e., a collection can belong to another collection). They can also connect data layers about similar topics or represent different years in a series.

The following diagram illustrates how the BTAA Geoportal organizes records. The connecting arrow lines indicate the name of the relationship. The labels reflect each record's **Resource Class** (**Collections**, **Websites**, **Datasets**, **Maps**, **Web services**).



4.5 Resource Classes

4.5.1 Collections

The BTAA Geoportal interprets the Resource Class, **Collections**, as top-level, custom groupings. These reflect our curation activities and priorities.

Other records are linked to Collections using the `Member Of` field. The ID of the parent record is added to the child record only. View all of the current **Collections** in the geoportal at this link: https://geo.btaa.org/?f%5Bgbl_resourceClass_sm%5D%5B%5D=Collections

4.5.2 Websites

The BTAA Geoportal uses the Resource Class, **Websites**, to create parent records for data portals, digital libraries, dashboards, and interactive maps. These often start off as standalone records. Once the items in a website have been indexed, they will have child records.

Individual **Datasets**, **Maps**, or **Web services** are linked to the **Website** they came from using the `Is Part Of` field. The ID of the parent record is added to the child record only.

View all of the current **Websites** in the geoportal at this link: https://geo.btaa.org/?f%5Bgbl_resourceClass_sm%5D%5B%5D=Websites

4.5.3 Datasets, Maps, and Web services

The items in this Resource Class represent individual data layers, scanned map files, and/or geospatial web services. (Some items may have multiple Resource Classes attached to the same record.)

This item class is likely to have the most relationships specified in the metadata. A typical **Datasets** record might have the following:

1. `Member Of` a **Collections** record
 2. `Is Part Of` a **Websites** record
 3. If the data was digitized from a paper map in the geoportal, it can be linked to the **Maps** record via the `Source` relation
 4. a general `Relation` to a research guide or similar dataset
-

4.5.4 Multipart Items

Many items in the geoportal are multipart. There may be individual pages from an atlas, sublayers from a larger project, or datasets broken up into more than one download. In these cases, the `Is Part Of` field is used.

As a result, these items may have multiple `Is Part Of` relationships- (1) the parent for the multipart items and (2) the original website.

4.6 Supplemental Metadata

All other forms of metadata, such as ISO 19139, FGDC Content Standard for Digital Geospatial Metadata, attribute table definitions, or custom schemas are treated as **Supplemental Metadata**.

- Supplemental Metadata is not usually edited directly for inclusion in the project.
- If this metadata is available as XML or HTML, it can be added as a hosted link for the Metadata preview tool in GeoBlacklight.
- XML or HTML files can be parsed to extract metadata that forms the basis for the item's GeoBlacklight schema record.
- The file formats that can be viewed within the geoportal application include:
 - ISO 19139 XML
 - FGDC XML
 - MODS XML
 - HTML (any standard)

 August 23, 2022

4.7 GEOMG Metadata Toolkit

4.7.1 About

What is it?

[GEOMG](#) is a custom tool that functions as a backend metadata editor and manager for the GeoBlacklight application.

Who uses it?

BTAA-GIN [Operations technical staff](#) at the University of Minnesota

Who developed it?

The BTAA Geoportal Lead Developer, Eric Larson, created GEOMG, with direction from the BTAA-GIN. It is based upon the [Kithe framework](#).

Can other GeoBlacklight projects adopt it?

We are currently working on offering this tool as a plugin for GeoBlacklight.

In the meantime, this presentation describes the motivation for building the tool and a few screencasts showing how it works:



 March 16, 2023