The Big Ten Academic Alliance Collection Analysis Tool

Getting Started

[V1 - 23 October 2023]

Basic Information

- Each institution deposits their full record set monthly. Collectively, the system contains over 100 million records, representing over 43 million unique records. Deposited records are processed by Gold Rush using both institution-specific mapping (provided by members) and Gold Rush algorithms. See Appendices for information about the local to global processing. The initial dataset is available as we continue to evaluate and adjust the extract, normalization and mapping processes. The dataset will become more accurate and comprehensive in the coming months as this work continues.
- For any given bibliographic entity, Gold Rush chooses the record with the highest level of encoding to designate as the primary record. That record is used to generate a Tool-specific unique identifier called the “matchKey”. Primary records are amended with non-bibliographic information (e.g., holding library, branch, circulation information, etc.).
- While only primary records display in search results, individual member records remain in the system and are available for view and download.
- Many fields, such as format, location, branch, library and others are multi-value. That is, one record can have multiple owning libraries, be held in multiple locations, be an electronic book, etc. Electronic versions are treated as separate from physical versions, i.e., they have distinct primary records.
- Only a few members chose to contribute circulation data.
- While the tool is a primary and important way to look at the collective collection, many use cases will involve exporting information from the system for use in other tools such as Excel, Python, R, OpenRefine, Tableau, etc.

Getting Started - Quick Tips

- For most use cases, start with the “all” search by clicking on the Search button with nothing entered in the search box.
- Always wait for search results and facets to completely load before proceeding to next steps. Due to the size of the dataset, it can take a little while for data to fully load in the tool.
- Refining by facets is a quick and easy way to narrow a search.
- Excluding or limiting to only Electronic in the format facet is a great way to refine a search. For example, try using the “X” next to “Electronic” in the Format facet.
- Clicking on an institution name within a record allows you to see the corresponding MARC record submitted by that institution.
- Note that the present search components are always displayed at the top of the screen above the search box. This is where you can remove a given component and also learn the search syntax for a given field.
Using Facets/Field Searching

Facets are a critical way to explore the tool. Here are some good things to keep in mind when using facets.

- **General Concepts:**
  - “X” beside a facet value
    - Excluding removes all items with that matchkey. *E.g.*, excluding a library will remove those titles from all results.
  - Clicking on a value in a facet limits to that value. This is identical to choosing one checkbox.
  - In most cases, you must click the “Apply” button to manifest your facet choices.
  - The tool loads the top 50 values for each facet based on the present search results. You will see the message “WARNING: MAX LIMIT REACHED” as the last facet value in cases of this limit. To see more granularity, reduce the size of the result set.

- **Call Number:**
  - Not case-sensitive. Do not put spaces in the call numbers.

- **Publication Year:**
  - Year “0000” is the value assigned when Gold Rush did not find a usable year of publication.
  - When limiting to date ranges, make sure and include a value of 1 or greater in the first box.

- **Publisher:**
  - This is a keyword field. *I.e.*, use “stanford” instead of “stanford university press”.

- **Publication Location:**
  - You can use the in-facet search box for “usa” and “canada” as a way to help narrow search results and/or subsequently exclude either/both.

- **Picking multiple entries on one facet list with the check box performs an OR. To use AND in that context, limit to one by clicking on the facet value, hit “apply”, and repeat with next value after results are displayed. This is also a context where typing the search directly into the search box may be prudent.**

Search Operators

- Default search operator is “OR”.
- Other operators (the Tool uses default Blacklight syntax):
  - AND for Boolean “and”
  - “*” is the wild card for 0 or more characters
  - Parentheses for nested searches
  - Double quotes for exact strings
Exports, Visualizations, Pivot Tables

- User initiated exports are limited to 5M records. If for some reason you need a larger export, contact goldrush@coalliance.org for assistance.
- “Delimited” file exports are ASCII. Use “Excel” for non-MARC exports where non-Roman vernacular matters.
- Exports include all submitted records, not just the primary record.
- Excel and Delimited exports contain the following fields, plus the match key if you select that when creating your export: Title, Author, Control Number(001), Library, Branch, ISBN, ISSN, Publisher Name, Publisher Number, LC Call Number, Dewey Call Number, SuDocs Call Number, Report Number, General Report Number, Edition, Action Note(583$f), Action Note(583$h), Publication Date, Language, 035 field.
- Try the icon within an expanded facet for basic visualization of that facet for the displayed results. That same icon appears at the top of the facet list, and will display counts for all facets for the present result set.
Appendix A – Creating the Tool: From Local Records to BTAA Tool

BTAA Collection Analysis Tool – from local to global

Source
- Institution 1
- Institution 2
- ...
- Institution 15

Gold Rush

Process
- Ingest/Map/Normalize
- Identify Primary Record
- Create matchKey
- Index

BTAA Tool

Explore
- Search
- Refine
- Export
- Visualize

BTAA Collection Analysis Tool – Source Preparation

Institutional Steps
- Specify MARC position for local bib ID, item record MARC field and subfields for branch, shelving location, barcode, serials/journals enumeration, and (optionally) circulation count.
- Specify local name mappings for values in 913 subfields for branch and shelving location.
- Prepare/automate monthly extracts from local system(s)

Gold Rush
BTAA Collection Analysis Tool – Processing Records

1. Ingest
   - SolrMARC + custom Gold Rush fields + institution-specific mappings

2. Identify primary record
   - Record with highest encoding level

3. Create matchKey
   - Create unique identifier for primary record

4. Finish Index and Publish to Tool
## Appendix B – Fields and facets from local to Tool

<table>
<thead>
<tr>
<th>GR Facet/Field</th>
<th>Sourcing Notes</th>
</tr>
</thead>
</table>
| LC Call Number              | 1. Look in the 050(a) field and append the 050(b).  
2. If not found (i.e., d in 050), look in the 090(a) and append the 090(b).                                                                                                                                                                                                         |
| *MVF*                       | Populated based on data found in all of the following:  
1. Handle pre-RDA by adding data contained in 245(h), 590(a), 533(a), 300(a) and 007 fields and leader bit 6  
2. Handle RDA by adding data in 337(a)  
3. Handle government document by adding the "Gov. Doc." facet if the 086 field exists. If the 086 field is found, look for an existing 856 field to add the "electronic" facet value.  
4. Look for the strings 'electronic or 'ebook' in the name of the file being indexed. If found, add the "Electronic" facet value.                                                                                                                                 |
| Location Code and Branch Code| Both *MVF*  
Branch is for building location. Location is for shelf location. Both are built from institution-specific mapping tables.  
For Branch Code, “No item data” means there is no value in the record OR there was not a match between the institution-specified field and the institution-specified mapping table. If the value in the institution-specified field is not in the mapping table.                                                                                                                                 |
| Language                    | language_facet = 008[35-37]:041a:041d, language_map.properties. See https://github.com/solrmarc/solrmarc/wiki                                                                                                                                                                                                                      |
| Region                      | Gold Rush custom mapping, removeTrailingPunct(651a:650z). See https://github.com/solrmarc/solrmarc/wiki                                                                                                                                                                                                                               |
| Overlap                     | Number of libraries with the same generated matchKey                                                                                                                                                                                                                                                                                        |
| Publisher                   | 1. Look for publisher in RDA 264b then In 260b.  
2. Normalize results with Java language routine java.text.Normalizer.normalize(text, Form.NFD).replaceAll("\\p{InCombiningDiacriticalMarks}+", ").  
3. Capitalize the first letter and make the rest lower case.                                                                                                                                                                                                                   |
| Dewey Call Number           | Use 082(a) field to get the class prior to "." ie 010, 020, 100. This class is used as the key to lookup values in dewey_facet_map. If not found, return 082(a) contents.                                                                                                                                                                           |
| SuDocs Call Number          | Use field 086(a) Decode the portion prior to the ":" ie 'GS 2.22:' If input does not have a : one is added to the end of the input String. Class letters are changed to uppercase. If a space, does not exist between the class and the #s one is inserted. Use the final results as a key to lookup the agency name in agencyNameMap. |