Licensed Data Implementation Working Group

Final Report

December 2020

Caroline Kayko
Karen Majewicz
Ryan Mattke (Lead)
Kathleen Weessies

The Licensed Data Working Group examined five licensed dataset types from six vendors that are held by most or all of the institutions participating in the BTAA Geoportal Project.[1] Our purpose was to assess the advisability and best practices for incorporating records into the Geoportal. Establishing the 'best' approach for these complex resources is not easy due to their dissimilarity, the varied nature of their interfaces, and the different providers that supply access to these datasets.

Structural Questions

We started from easiest to add (Landscan) and progressed to most difficult (SimplyAnalytics). With one exception, none of the products we studied allows the geoportal to link out to individual data points within their interfaces but rather force us to link only to the front page of each resource. We debated whether to describe a single geoportal record that linked to the whole resource, or if we ought to populate the geoportal with multiple records per product to capture the variety of data points available within each. We further weighed pros and cons of linking to the resource itself or to a library catalog record for each institution. We decided to link to the library catalog record itself since that is the most stable link and the least likely to change and require an update. We chose to create parent collections for each entire resource and multiple child records per parent to describe individual data points.

Structuring Records

The overarching theme for entering the datasets into the portal is to try to mimic the navigation of the web interface of the resource as much as possible in terms of layer names or layout. This is to prepare the user for how the source of the licensed data will be structured when they are directed to an external site.

---

[1]Landscan from Eastview, SimplyAnalytics from Geographic Research Inc., fire insurance maps from Proquest and FIMO, PolicyMap from The Reinvestment Fund, and Social Explorer from Social Explorer Company.

Depending on the structure of the licensed data source and what is most logical for the user to access, the records are either organized:

- by year (Landscan)
- by state (fire insurance maps)
- by topic category as defined by the vendor (Social Explorer, SimplyAnalytics and Policymap)

Considerations

Any dataset that already exists in the geoportal is disqualified from being added as a record, especially if it exists in a "better" form. We give preference to original data over derived data products. For example, Social Explorer provides raw census data, while SimplyAnalytics offers census data that has been reconfigured by a third party vendor. We also set aside some datasets of uncertain origin, such as SimplyAnalytics' five-year population projections.

We also determined that some records are too complicated to add based on the data structure (for instance PolicyMap, see below), and we may just add a parent record and direct the user to the resource itself.

Landscan

The Landscan collection is made available to institutions in annual datasets beginning in 2000. We tested creating records by institution, but ultimately found it was less complicated to go by year. We settled on adding 19 records to the geoportal, one for each year 2000 to 2018.

Database description: Landscan is a world population dataset at approximately 1 km resolution (30" X 30"). It represents an ambient population (average over 24 hours). The LandScan algorithm uses spatial data and imagery analysis technologies and a multi-variable dasymetric modeling approach to disaggregate census counts within an administrative boundary. Since no single population distribution model can account for the differences in spatial data availability, quality, scale, and accuracy as well as the differences in cultural settlement practices, LandScan population distribution models are tailored to match the data conditions and geographical nature of each individual country and region. Note: Due to methodology changes from year to year, direct comparison of individual pixels between years is not recommended. Click on a pixel in the map to see the population for that pixel. For larger geographic areas, click "search" to select either a political unit or to draw a polygon on the map. A pop up window will provide population with gender and age breakdown for that area.

Digital Sanborn Maps/FIMo

The Digital Sanborn Maps from Proquest and FIMo are each separated into records by individual cities which are further collected into states. We decided to create state-level records for the geoportal, especially because many institutions only own maps for a certain state and not for the whole country. To improve searching, we added the many hundreds or thousands of place names mapped within each state into the keyword metadata field. We will create separate

records for each vendor and state, we combine the comments here into one entry due to the similarity of the data structure.

FIMO Database description: This resource contains full color detailed maps of numerous towns and cities, however generally of only the more densely settled parts of each place. They contain detailed information about city properties including building characteristics, their addresses, and in commercial areas the type of business establishment in each building. Each town center was mapped one or even several times between 1870-1960. Most content is from the Sanborn Fire Insurance Company, but this growing database will continue to add fire insurance maps from other publishers and other large scale maps. The database is searchable by address, by coordinates, or by a clickable map interface.

Digital Sanborn Maps Database description: The Sanborn Maps provide detailed information about city properties including all the buildings, their addresses, and in commercial areas the type of business establishment in each building. Each town center was mapped one or even several times between 1870-1960.   Each volume begins with a graphic index to guide the user to the correct page; each page only covers a few blocks. This database is black-and-white only, while the original paper maps used color to indicate building materials**.**

PolicyMap

The PolicyMap records are organized by the topmost grouping in the hierarchy by major topic areas (Demographics, Health, Quality of Life, etc), which was given to us by a representative from the company. This decision was made because there are only 32 major topic areas as opposed to hundreds of categories and thousands of individual data layers. The individual layers would be more difficult to track changes or added layers from the provider. Each of those 32 parent records has the layer names as child records. We elected to not include PolicyMap's census data since it is available more readily (for GIS users) from Social Explorer.

Database description: Policymap is an online data and mapping application that gives access to over 15,000 indicators related to housing, crime, mortgages, health, jobs, demographics and education. Data are available at all common geographies (address, block group, census tract, zip code, county, city, state, MSA) as well as unique geographies like school districts and political boundaries. Data come from both public and proprietary sources. Many of the public files are available for download. Through PolicyMap, you can access all data in interactive maps, tables, charts, reports and through a unique analytic tool. You may also upload your own address level data and share your data and interactive maps with others.

Social Explorer

The records for Social Explorer are organized by their layer names in the interface for attribute data. We also added data from the new SocialExplorer Geodata section to the portal

using the overarching categories, which allows users to download contemporary and historical shapefiles.

Database description: Create maps and reports from selected census statistics. 1790 to 1930: County, state and national data. 1940-2000: Census tract, county, state and national data. 2006- : ACS. Also includes religion data, U.S. Election data, UK Census 2011 data, and Ireland Religion data. Generate thematic maps for census tracts. It offers a growing selection of shapefile downloads.

SimplyAnalytics

SimplyAnalytics was not able to provide us with a list of their layers according to how they are structured on the site. They were only able to provide a list of datasets by source ("Standard Package" and "Additional Data Sets"). We determined that the Standard Package was already covered by other sources like Social Explorer and Policy Map, therefore we only included the Additional Data Sets as records in the geoportal. Because SimplyAnalytics allows export of shapefiles, we were keen to include SimplyAnalytics especially for layers that weren't available elsewhere.

Database description: Formerly named SimplyMap, SimplyAnalytics is a web-based mapping tool that lets users quickly create professional-quality thematic maps and reports using powerful demographic, business, and marketing data. Maps are exportable as shapefiles.

Changes to the Metadata and Interface

We needed to make enhancements to the metadata and the geoportal interface to incorporate licensed data.

Metadata Fields:
1. Mediator (dct_mediator_sm):
    a. Purpose: Allow users to select their institution when searching for restricted records
    b. Format: a multi-valued text field that contains the names of each institution that has access to the record
2. Access (b1g_access_s):
    a. Purpose: Provides links to each institution
    b. Format: This is a single-valued text field that contains a stringified hash of key-value pairs.
3. Access Rights (dct_accessRights_sm)
    a. Purpose: To clarify that these are restricted records and provide external links for more information

     b.   Format: This is a multi-valued text field that can contain plain text and hyperlinks

Metadata Values:
1. Provenance (dct_provenance_s): The licensed data pilot demonstrated that our prior convention of using short labels for institutions (i.e. Minnesota) was confusing for users. We changed this practice to write out full institution names (i.e. University of Minnesota) in both the Provenance and Mediator fields.
2. Date Issued (dct_issued_s): We removed the Date Issued field from all licensed records. Including this date was confusing as it conflicted with temporal coverages.
3. Type (dct_type_sm): The licensed databases require users to query, filter, and combine data within the application. To differentiate these databases from the rest of the BTAA collections, they are given the value "Interactive Resource" in the Type field. This aligns with the [Dublin Core controlled vocabulary](#).
4. Access (b1g_access_s): This new field is made up of key-value pairs, where an institution is the key and the value is the link. As the institution names are long, we used a set of codes that represent each institution. This allows the metadata to be more concise.

> **Code | Institution**
> 01 | Indiana University
> 02 | University of Illinois
> 03 | University of Iowa
> 04 | University of Maryland
> 05 | University of Minnesota
> 06 | Michigan State University
> 07 | University of Michigan
> 08 | Purdue University
> 09 | Pennsylvania State University
> 10 | University of Wisconsin-Madison
> 11 | The Ohio State University
> 12 | University of Chicago
> 13 | University of Nebraska-Lincoln

New Facets/Filters:
1. Public/Restricted: This allows users to filter records between two values (Public or Restricted). These values come from the Rights fields in the metadata, which was always present. Since we only loaded Public records in the past, this filter has not been needed.
2. Institutional Access: This pulls from the new Mediator field and allows users to filter to their institution

Item Show Page features:
1. Access Rights: This displays the values from the Access Rights field and includes clickable hyperlinks

2. Licensed Resource widget: This widget appears in the Links box. At the top is a padlock with the term "Licensed Resource." If a user hovers over the term, the phrase, "*This resource requires authentication through one of the following institutions. Select your institution for access.*" Beneath this heading is a list of institutions and icons. Each link will take a user to a library catalog page for the resource.

Implementation

We will be implementing the changes outlined in this document through an internal reallocation of effort, rather than increased direct costs.

Outcomes & Next Steps

Outcomes

- The recommendations and guidelines in this report will allow for the incorporation of other licensed geospatial content such as scanned & georeferenced topographic map sets.
- Procedure for adding records for licensed resources to the BTAA Geoportal
    - Contact the Project Manager & Metadata Coordinator
    - Coordinate with other project institutions that also subscribe to the resource
    - Based on the decisions described in this report, determine how the item records for the resource should be structured in the BTAA Geoportal
    - Draft a description of the resource to be used in the item record

Next Steps

- Form a small group or organize a focused sprint time to create help pages to help users understand what licensed data is, how it impacts their browsing experience of the portal, and how to access it via their own institutions. Pages to be created include:
    - Licensed data
        - Navigating resources (for interactive applications) - find guides that exist and point to those; also, quick tips on what to use for searching the resource
    - Sanborn - find guides that exist and point to those
    - Public Land Survey System maps
    - Census data (locating and accessing)
    - [GMU guide](#) has example comparisons and text
- Data cleanup and review
    - Review and update the Licensed Data Inventory sheet
    - Ensure the correct links are present for each institution's version of the above resources
- Ask Collection Development Committee to amend the Collection Development Policy to include licensed data resources

- Form a small group to ensure that the descriptions for existing licensed data resources are accurate