DIGITAL PUBLIC LIBRARY OF AMERICA

http://bit.ly/dpla-geo-styleguide

DPLA Geographic and Temporal Guidelines for MAP 4.0

DPLA's most complex enrichments involve the enhancement of geographic and temporal information. This guide is intended to help Hubs develop content standards for themselves and their partners that will be easily parsed by DPLA.

Geographic Enrichment

DPLA's geographic enrichment attempts to match values in partner-supplied data with known geographic vocabularies, typically GeoNames. Matches can be made from string values, geospatial coordinates, or URIs. Once a match is found, the enrichment writes the coordinates, URI, and preferred label back to the DPLA MAP record in the spatial class. The original partner-supplied data is stored as the provided label.

During initial mapping, DPLA transforms spatial terms in your record to properties in the Place class of the DPLA MAP. In particular, we populate the *dpla:providedLabel* with the text from the incoming field. If the supplied data has structured spatial elements, such as latitude and longitude we populate other DPLA MAP fields accordingly.

During enrichment we take the value that has been mapped to *dpla:providedLabel* and use geocoding tool (currently TwoFishes) to compare it to a set of "candidate interpretations" based on data from GeoNames and LC authorities. Candidates are weighted slightly toward US results. Before accepting an "interpretation" we compare its coordinates (or bounding box) to any coordinates provided at mapping time to rule out false positives.

If a match is found, we populate prefLabel, latitude, and longitude for the Place. The GeoNames URI is added as a *skos:exactMatch*, and LC authorities returned by the geocoder (if any) are added as *skos:closeMatch* concepts.

Recommendations for Creating Geographic Metadata

The best way to ensure a correct match is to use unambiguous data. A GeoNames URI is great if you have the ability to add them to your data. Geospatial coordinates are also useful. If you are unable to supply these, consistently using place names in a hierarchy (Erie, Pennsylvania, or even Erie, Erie County, Pennsylvania, United States) will result in better matches.

A rich standard like MODS will allow you to arrange geographic locations hierarchically. If you are using a simpler standard like Dublin Core, the best way to include geographical names in a hierarchy is to put the values in a single instance of the property, in ascending or descending order, separated by commas or semi-colons:

✓ ***<dcterms:spatial>Erie, Pennsylvania, United States of America</dcterms:spatial>***

OR

✓ ***<dc:coverage>United States, Pennsylvania, Erie</dc:coverage>***

It is less desirable for these places to be encoded in their own individual properties (in other words for Erie, Pennsylvania, and United States to be placed in separate <dc:coverage> fields). If records are formatted this way we would still be able to use the data to find a place in geonames, but the likelihood of finding an unambiguous match would decrease whenever you added a new place. In that case, it becomes impossible to tell which place names are related to which:

✗ *<dc:coverage>United States</dc:coverage>*
<dc:coverage>Pennsylvania</dc:coverage>
<dc:coverage>Erie</dc:coverage>
<dc:coverage>Illinois</dc:coverage>
<dc:coverage>Washington</dc:coverage>

The example above could be for Erie, Pennsylvania, and Washington, Illinois . . . but there is also an Erie in Illinois and a Washington in Pennsylvania. Without hierarchy, we have trouble disambiguating terms.

If you are able to supply geospatial coordinates together in the same property, we would prefer if they were ordered with latitude first, and the two coordinates separated by a comma.

✓ *6.703, 46.627*

You may also supply coordinates in separate <lat> and <long> properties if supported by your metadata schema.

Coordinates should be limited to whole numbers and decimals (i.e., no degree-minutes-seconds or degree-minutes), and compass directions should appear after the number if present, not before:

- ✓ *45.52N*
- ✓ *45.52 N*
- ✓ *-81.13° S*
- ✗ *45 30 1 N*
- ✗ *N45.52*

If you are supplying information about a bounding box, please contact the Data Services Coordinator to work on a crosswalk.

Coordinates may be included in a hierarchical location value, as in:

- ✓ *<dcterms:spatial>United States, Pennsylvania, Erie, 42.1167, -80.07315</dcterms:spatial>*
- ✗ *<dcterms:spatial>United States, Pennsylvania, Erie</dcterms:spatial> <dcterms:spatial>42.1167, -80.07315</dcterms:spatial>*

The coordinates can be included before or after the place name.

URIs for locations in a linked data vocabulary, such as geonames, may also be included. Again, to prevent duplicates or ambiguity, include these in the same hierarchical string when using a simple schema like Dublin Core:

- ✓ *<dcterms:spatial>United States, Pennsylvania, Erie, 42.1167, -80.07315, http://www.geonames.org/5188867/city-of-erie.html</dcterms:spatial>*
- ✗ *<dcterms:spatial>United States, Pennsylvania, Erie, 42.1167, -80.07315</dcterms:spatial> <dcterms:spatial>http://www.geonames.org/5188867/city-of-erie.html</dcterms:spatial>*

Temporal Enrichment

DPLA's temporal enrichment is extremely flexible and can handle a large variety of date formats. The enrichment itself first parses the value of any date or temporal field and looks for an identifiable date or date range. Examples of different date formats that can be parsed are below:

- 1992
- 1992-12-01
- 1992-12
- 12-1992
- 12-01-1992
- 12.01.1992
- 12/01/1992
- 1992-12-01?
- 1992-12-01~
- 1992-12-uu
- 1992.12.01
- 1992-12-01
- 1992.12
- 1992?
- circa 1992
- ca 1992
- c 1992
- ca. 1992
- c. 1992
- 1992/1993
- 199x
- 19--
- [1992]
- NEARLY c 1992 *[this is to indicate that the date can be parsed even if words other than some form of "circa" precede it]*
- 1990s
- late 1990s
- 199-
- Dec 1992
- Dec 01 1992

- Dec 01, 1992

Temporal information in the DPLA MAP is stored within the TimeSpan class. This class has three sub-properties: date, begin, and end. Once the date is identified, an identical value in the EDTF (YYYY-MM-DD) format is created and stored in the begin property. If it is a single date, it is also stored in the end property. If the supplied date is a range, then the appropriate begin and end dates in EDTF format are recorded. The original date format, as supplied by the provider, remains unchanged in the date property. This is the property that will be displayed in the DPLA portal.

For example, if the date “May 22nd, 1973” was supplied the Timespan property would include:

Date: May 22nd, 1973
Begin: 1973-05-22
End: 1973-05-22

If a date range of “1730-1750” was supplied the Timespan property would include:

Date: 1730-1750
Begin: 1730
End: 1750