Geospatial Open Data Standards

Draft Geospatial Open Data Standards

Last November, Mayor de Blasio signed Local Law 108 of 2015, mandating the formation of a working group tasked with creating standards for address and geospatial information on the Open Data Portal. Over the last several months, the working group drafted recommendations for geospatial attributes, column headers, and geocoding processes for datasets on the Portal. In the spirit of Open Data for All, we invite the public to join the conversation as we finalize these recommendations into formal requirements.

Open Data promotes transparency by publishing data used internally in the City with as few transformations as possible. While we seek to maximize the usability of data on the Portal, the Open Data team changes the form of raw data only when it presents minimal risk of misrepresenting source information. As a result, the recommended geospatial standards reflect the data fields most frequently captured by City agencies, information that is in highest demand from public users, and attributes that will have the biggest impact on citywide operations once they are standardized.

Please read the standards and leave your comments, suggestions, and questions in the form below. **Public comment will close on Thursday, September 15, 2016.** The working group will consider your feedback when finalizing the standards for inclusion in the Technical Standards Manual, the document of record for the City's Open Data policies and protocol.

Local Law 108 of 2015 (Excerpt)

To amend the administrative code of the city of New York, in relation to the standardization of address and geospatial information on the Open Data Portal:

The [Technical Standards] manual… shall include a technical standard **requiring every public data set containing address information to utilize a standard field layout and presentation of address information and include corresponding community district and geospatial reference data**. If there is a public data set for which an agency cannot utilize such standard field layout and presentation of address information, such agency shall report to the department and to the council the reasons why it cannot, and the date by which the agency expects that it will be able to utilize such standard field layout and presentation of address information, and such information shall be disclosed in the compliance plan prepared pursuant to section 23-506.

Draft Standards

For any dataset on the NYC Open Data Portal that includes row-level address fields, agencies must separate locational information into "core address" and "core geospatial reference" attributes. These attributes will appear on the Portal according to a standard column naming convention.

Agencies will be responsible for separating *core address* information into five standard column fields:

- "HOUSE NUMBER"
- "STREET NAME"
- "APARTMENT NUMBER"
- "ZIPCODE""
- "BOROUGH"

Agencies will also be required, with technical guidance from the Open Data team, to include six standard column fields of *core geospatial reference* information:

- "LATITUDE"
- "LONGITUDE"
- "COMMUNITY BOARD"
- "COUNCIL DISTRICT"
- "BIN" (Building Identification Number)
- "BBL" (Borough-Block-Lot)

Geosupport: We recommend agencies whose datasets do not already contain the six *core geospatial reference* fields to use Geosupport, a publicly available tool that also serves as the City of New York's geocoder of record maintained by the Department of City Planning. *Core address* data entered into Geosupport can return all required *core geospatial reference* data. Agencies may geocode their locational data at the database level or the extraction level. Alternatively, agencies may elect to have the Open Data team establish an automated feed, in which datasets are passed through an ETL where they are geocoded and uploaded directly to the Portal.

When a dataset is geocoded by Geosupport, its data dictionary must designate which attribute fields were reported directly by the agency, and which attribute fields were created by geocoding in order to meet these standards. Data dictionaries must also include the version number of the geocoder and error rates that result from geocoding. Finally, agencies with datasets that do not have address fields but include other locational data are encouraged, but not required, to populate as many *core geospatial reference* fields as possible using Geosupport.