Name / Organization: CACI

Contact: Jason Rowe, [jrowe@caci.com](mailto:jrowe@caci.com), 970-224-4996

Dataset solution accesses: Humanitarian OpenStreetMap Team (HOT) data (as explained below, our solution works with the complete raw data from which the HOT data is derived).

“Link” to working product: The complete product, along with sample configuration files, input data and correspond output data are included with this submission.

OpenStreetMap (OSM) – from which HOT data is extracted – is a collaborative project to create a free editable map of the world. It is one of the largest and fastest growing sources for geospatial feature data. It contains over 485 million features and has over 3 million registered contributors. In the last month over 38,000 of those contributors made at least one edit to the OSM data. OpenStreetMap is an especially good data source in the face of a humanitarian disaster as volunteer contributors are able to be mobilized within hours of an event. After the April 2015 Nepal earthquake 800 volunteers had already mapped 10,000 buildings and 24,000 road segments within the first 24 hours after the quake striking. The Humanitarian OpenStreetMap Team (HOT) coordinates these efforts helping to ensure that the data collected by remote volunteers meets the needs of the responders on the ground.

The challenge with OpenStreetMap data, and therefore HOT data, is that it is only semi structured and not suitable for search or analysis within a traditional GIS. The data, having been entered by volunteers, is influenced by their skill level, their commitment to the project, their personal opinion as to how a given feature should be represented, and, of course, human error. For example, an embassy may be tagged in the data as amenity=embassy, diplomatic=embassy, building=embassy or any number of other ways, including misspellings of the preceding ways. Further, it may not be explicitly tagged as an embassy at all, but only be tagged with a name, such as name=Embassy of Sweden. Numeric values can be a special challenge as OSM may record the height of a feature as height=10, height=10 m, height=10’, height=10ft, or any number of forms that were perfectly sensible to the individual entering the data, but are problematic for a machine to interpret. Names are another challenge and opportunity when working with OSM. Over 53 million OSM features have a name. This name may be in the local language, English, or any other language (for example the language of a former colonial power). Further, additional names may be explicitly specified for a given language. These are in addition to the 53 million cited above. For example, there are 1.6 million explicitly tagged as English (name:en). OSM also contains many additional types of names, such as alternate names (alt\_name), building names (building:name) and bridge names (bridge:name). Depending on the intelligence problem being addressed, an analyst may want English names (explicitly specified and otherwise), a specific foreign language name (e.g. Arabic), or all names. The CACI solution solves the types of problems illustrated by the height, embassy and name examples above and our submission for the Disparate Data Challenge contains two examples that illustrate how our solution can be used to extract and convert heights from OSM, and how our solution can extract embassies and their names from OSM.

A number of sources, such as market.trimbledata.com, offer OSM data in standard GIS formats, but these services are a “one size fits all” interpretation of how OSM should be crammed in to GIS and do not allow the analyst to adjust the translation to address their immediate problem. Further they do not take into account the nuances of the data illustrated by the embassy, height and name examples above. The HOT “OSM Export Tool” does offer some customization, but it cannot address the issues outlined above. Further, it only operates over some parts of the world and key areas in which NGA may be interested, such as the Ukraine, are not available.

The CACI solution is a “conversion engine” which takes as input a raw OSM XML (.osm file extension) data file and outputs a file geodatabase. With this solution there is no need to rely on secondary intermediate formats and schemas controlled by entities outside the US Government as it allows one to work with the full, complete and up-to-date actual OSM data in its native format. However, the real power of this solution is that the exact way the conversion takes place can be completely controlled by the user by editing, or specifying an existing, Python based file. The “heavy lifting” of the conversion is done by the conversion engine, while the user can control the feature class where an OSM feature belongs, and how attribution is set. The majority of GIS analysts today are versed in Python, with many using it in their day-to-day work, thus making the power of the conversion engine accessible to almost anyone who would have a need to use OSM data. The use of Python in this solution makes the full power of that language, as well as 100’s of third party modules, available to solve the types of problems illustrated earlier in this document.

The raw OSM data is available directly from OpenStreetMap, through the Overpass Server, and from many third parties. However, CACI maintains its own mirror of the OSM data which is automatically updated with minute by minute difference (diff) files. This allows us to extract OSM data without suffering the timeouts and limits imposed by the equivalent public resources, and without potentially telegraphing our, or our customer’s, interests to third parties.

The conversion engine is currently a command line tool, but this technology could form the basis for an interactive web or desktop app that allows the user to select an AOI, select a conversion file (or submit a custom file), and receive either a file geodatabase or a web feature service (WFS) containing the results.

Instructions for accessing our Submission (not included in 1,000 word limit)

What is included with the submission:

* CACI.zip
  + osm\_converter.exe, python27.dll & res directory - Conversion engine
  + Embassy.py – illustrates how to extract embassies from OSM data
  + Height.py – illustrates how to extract heights in consistent units from OSM data
  + Syracuse.osm – Raw OSM data over Syracuse NY US. This was selected because it illustrates our ability to convert both meters; and feet and inches to a common unit (feet in this case).
  + Turkmenistan.osm – Raw OSM data over Turkmenistan
  + SyracuseHeights.gdb – blank file geodatabase which will be populated with features with heights in Syracuse when the converter is run.
  + SyracuseHeights\_Populated.gdb – Version of the above database upon which the converter has been run. If NGA does not wish to run the converter, or for some reason is unable to, it may examine the results of doing so by inspecting this file.
  + SyracuseHeights.bat – Windows batch file to run the conversion of features in OSM over Syracuse that have heights to file geodatabase (populates SyracuseHeights.gdb). Provided for NGA’s convenience when testing.
  + TurkmenistanEmbassies.gdb – blank file geodatabase which will be populated with embassy and consulate features in Turkmenistan
  + TurkmenistanEmbassies\_Populated.gdb – Version of the above database upon which the converter has been run. If NGA does not wish to run the converter, or for some reason is unable to, it may examine the results of doing so by inspecting this file.
  + TurkmenistanEmbassies.bat – Windows batch file to run the conversion of features in OSM over Turkmenistan that are embassies to file geodatabase (populates TurkmenistanEmbassies.gdb)

Instructions

* Unzip caci.zip to a location on your computer’s hard drive. There is no need to install any software.
* Using a GIS, such as ArcMap or QGIS, examine SyracuseHeights.gdb and TurkmenistanEmbassies.gdb. Note that they are blank. Close the GIS (to release the file lock).
* From windows explorer, double click on each of the provided batch files (SyracuseHeights.bat and TurkmenistanEmbassies.bat). Follow the prompts from the command line window. Some warnings are normal and are usually due to geometry errors in OSM.
* Again using a GIS, open the file geodatabase associated with the bat file that you double clicked above. Open the attribute table for each feature and notice the table is now populated. Since the default view is set to the world extent, highlight a feature in the attribute table and select “zoom to feature”. Further note that in the height database the heights are populated with real world units (feet). A comment attribute is provided to explain how the converter arrived at the real world values (e.g. “Converting meters to feet” or “Converting feet and inches to feet”. In the Embassies file geodatabase the comment attribute explains both how the determination was made that the feature was an embassy (or related facility), as well as how the name was determined (e.g. “name taken from name tag”, “name taken from name:en tag”). Note that different methods of selecting the embassies were used (e.g. “amenity key”, “embassy key”, “diplomatic key”). In Both databases the RawTags attribute lists all of the OSM tags associated with the given feature in pipe delimited text format. Provided are also prepopulated databases that were ran at our Fort Collins CACI facility (SyracuseHeights\_Populated.gdb and TurkmenistanEmbassies\_Populated.gdb). You may use these to compare the results as they should be identical.