Ingesting Open Geodata: Observations from OpenStreetMap

Alan McConchie
Design Technologist, Stamen Design
PhD Candidate, University of British Columbia
alan@stamen.com
@mappingmashups

Wikimedia Research & Data Showcase
February 18, 2015
Outline

What is OpenStreetMap?
The TIGER import
Problems and complaints
Emerging best practices
Origins of OpenStreetMap

“The Wikipedia of Maps”

Started in 2004 in the UK

Map originally based on nothing but volunteers’ GPS traces and notes
OpenStreetMap today

On-the-ground mapping still important

Tracing from permitted satellite imagery

Mass imports of data

TIGER/Line (U.S. Census Bureau)

Automobile Navigation Data (Netherlands)

Other Imports

OpenStreetMap
License

**CC-BY-SA** until 12 September 2012

Now: **ODbL**

CC-derived Open Database License

[wiki.osm.org/wiki/Open_Database_License](http://wiki.osm.org/wiki/Open_Database_License)
<table>
<thead>
<tr>
<th>Tag</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>highway</td>
<td>secondary</td>
</tr>
<tr>
<td>name</td>
<td>52nd Street</td>
</tr>
<tr>
<td>oneway</td>
<td>yes</td>
</tr>
<tr>
<td>tiger:cfcc</td>
<td>A41</td>
</tr>
<tr>
<td>tiger:county</td>
<td>Alameda, CA</td>
</tr>
<tr>
<td>tiger:name_base</td>
<td>52nd</td>
</tr>
<tr>
<td>tiger:name_type</td>
<td>St</td>
</tr>
<tr>
<td>tiger:reviewed</td>
<td>no</td>
</tr>
<tr>
<td>tiger:source</td>
<td>tiger_import_dch_v0.6_20070809</td>
</tr>
<tr>
<td>tiger:tlid</td>
<td>124994944:124995055:124995070:124995071:124995049</td>
</tr>
<tr>
<td>tiger:upload_uuid</td>
<td>bulk_upload_pl-9d5e179e-e488-4300-b71d-b0a82eacb273</td>
</tr>
<tr>
<td>tiger:zip_left</td>
<td>94608</td>
</tr>
<tr>
<td>tiger:zip_right</td>
<td>94608</td>
</tr>
</tbody>
</table>

Nodes:
- 506458708 (part of ways 52nd Street (41413743) and 52nd Street (6335404))
- 266433535
- 506458789
- 266435373 (part of way 51st Street (41413733))
- 266435374
- 26643540
OSM TIGER Import

Stands for “Topologically Integrated Geographic Encoding and Referencing”

TIGER is a product of the United States Census Bureau, therefore in the Public Domain.

OpenStreetMap imported TIGER’s 2005 data starting in late 2007:

animation: wiki.osm.org/wiki/Tiger
Fixing TIGER
Major 'TIGER' roads
- Reviewed
- Edited recently, but still has 'tiger:reviewed=no'
- Not edited recently and still has 'tiger:reviewed=no'

Minor 'TIGER' roads
- Reviewed
- Edited recently, but still has 'tiger:reviewed=no'
- Not edited recently and still has 'tiger:reviewed=no'

Non-TIGER roads
- Paths and service roads without 'tiger:source'
- Other roads without 'tiger:source'
- State/national borders

A 'reviewed' road is one with 'tiger:source' but without a 'tiger:reviewed=no' tag. A 'major' road is anything from tertiary to motorway. 'Recently' means within the past 3 years.
OpenStreetMap versus TIGER 2014 Battle Grid

Orange squaroids are urban, Green squaroid are non-urban. Black squaroids are marked done.

fix the bright cells first / read more about this / built with ♥ by Martijn van Exel / Telegav
This way has one or more unusually sharp angles. It can probably benefit from some smoothing and aligning with Bing imagery.

This way has one or more unusually sharp angles. It can probably benefit from some smoothing and aligning with Bing imagery.

We are in Gridley, Clarke County, Mississippi, the United States of America.
Imports beyond TIGER
Buildings without addresses
Addresses without buildings
Community backlash
Just last Saturday, I discovered that the edits I made in Aylmer and Hull — suburbs of Gatineau, Quebec that I know very well, because I pass through them every time I go to Ottawa — had been replaced by a batch import from CanVec.

It removed a lot of human intelligence that I had added to the existing streets, such as pedestrian crossings, traffic lights, and turning circles. It also obliterated service roads, and turned all streets into highway=unclassified instead of residential, tertiary, or secondary, and divided highways into two-way streets. And any areas that had shared a node with a street way were now bollixed up.

In other words, a big mess to revert, or to fix: the CanVec data had some useful information to add to the map — but not at the cost of erasing hours upon hours of existing work. My work. I was pissed[…]

My basic point was, and is, the following: that you’re not going to get local people to contribute to OSM if they believe that their edits are going to be wiped out by the next person to import a pile of data.

Jonathan Crowe, 2011 (emphasis mine)

OSM is part of a greater movement of collaborative productivity, where people all over the world can and do join forces to create something great, something of value.

I believe that in 40 years, probably even in 15, hardly anything of the data we have collected will retain much value - but we will have been part of a great development, and mankind will be the better for it.

Will OSM, instead of being the social endeavor of “a great map that people made themselves”, then be the technical challenge of “the geo database where a few clever guys managed to combine lots of existing data”?

Frederik Ramm, 2012 (emphasis mine)
A popular (although not universally accepted) theory is that massive imports are to blame for slowing down the development of the US community. …instead of the fun task of going out, finding new streets, and filling in a blank map, new contributors in the US were now faced with the relative tedium of correcting repetitive errors in an existing dataset.

Meanwhile, in countries like Germany, almost all data has been collected by volunteer mappers, with only a few small-scale imports conducted in places where there already was an active community. This was not by design - there simply wasn't a comparable dataset that could have been imported - but it was ultimately beneficial. At least if you accept the theory, of course.

Tobias Knerr, literally yesterday (emphasis mine)
Current best practice

wiki.osm.org / wiki / Import / Guidelines
OSM Import Guidelines

Consult with the community
  • Import-focused mailing lists and working groups
  • Mappers in the area and local groups

Documentation
  • License compatibility
  • Import plan
  • Translation code and sample data

Review
  • By imports@openstreetmap.org mailing list
  • By local mappers for ground truthing

wiki.osm.org/wiki/Import/Guidelines
Community Imports

- Divide data into chunks
- Community members manually upload each chunk

Community Imports

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>OSM Seattle Owner</th>
<th>Progress</th>
<th>QA/d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>Glassman</td>
<td>Completed 2/3/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Alki</td>
<td>tbd</td>
<td>pending</td>
<td>pending</td>
</tr>
<tr>
<td>Arbor Heights</td>
<td>Lagerratrobe</td>
<td>Started 3/30/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Atlantic</td>
<td>tbd</td>
<td>pending</td>
<td>pending</td>
</tr>
<tr>
<td>Ballard</td>
<td>Glassman</td>
<td>Completed 2/8/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Bellevue</td>
<td>Richard</td>
<td>pending</td>
<td>-</td>
</tr>
<tr>
<td>Belltown</td>
<td>SeattleImport</td>
<td>Completed 5/31/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Bitter Lake</td>
<td>SeattleImport</td>
<td>Completed 4/29/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Bothell</td>
<td>Chris</td>
<td>pending</td>
<td>pending</td>
</tr>
<tr>
<td>Briarcliff</td>
<td>SeattleImport</td>
<td>Completed 5/18/2013</td>
<td>pending</td>
</tr>
<tr>
<td>Brighton</td>
<td>tbd</td>
<td>pending</td>
<td>pending</td>
</tr>
</tbody>
</table>

NOTE: The term "Import" is highly loaded in the OSM community. "A distributed and curated merge," is a more accurate description of what Seattle OSM planning to do.

— wiki.osm.org/wiki/Seattle_IMPORT
Building community through imports

Four training events and editathons
Five events out walking around
20+ active participants
NYC community building import

- Estimated 1500 hours of work
- Volunteers + Mapbox employees
- Coordination with NYC GIS Dept

animation: mapbox.com/blog/nyc-buildings-openstreetmap
NYC Buildings and Addresses

Ready?

Take a task Or choose one by clicking on the map.

If you’re an experienced mapper, you can also be given a task to validate.
LA County building import
(still in planning stage)

3,141,244 building footprints
6,423 block groups

github.com/osmlab/labuildings
Scholarly research on OSM imports
The impact of the TIGER import on OSM completeness

Comparison of OSM in various cities:

Central London with 1000m grid
Total edits vs “blank spot” edits
London

Greater London: OSM contributors plotted by number of “blank spot” edits and number of total edits. Circles are individuals editors sized according to number of days active.
San Francisco Bay Area: OSM contributors plotted by number of “blank spot” edits and number of total edits. Circles are individuals editors sized according to number of days active.
Seattle-Tacoma: OSM contributors plotted by number of “blank spot” edits and number of total edits. Circles are individuals editors sized according to number of days active.
Metro Vancouver: OSM contributors plotted by number of “blank spot” edits and number of total edits. Circles are individuals editors sized according to number of days active.
Cairo, Egypt: OSM contributors plotted by number of “blank spot” edits and number of total edits. Circles are individuals editors sized according to number of days active.
Thanks!

@mappingmashups
alan@stamen.com