Assessing OpenStreetMap Data Quality Using the Data Reviewer Extension

Dennis Zielstra and Hartwig H. Hochmair

Geomatics Program – University of Florida
Fort Lauderdale Research and Education Center

ESRI International User Conference, San Diego (CA), July 8-12, 2013
Background

- Several studies within recent years about OSM data quality in comparison to governmental or proprietary data.

- Mostly with focus on Europe (i.e. UK, Germany).

- Usually good data quality results in countries with active OSM community.

- First research results about OSM data quality in the US in recent years (Zielstra and Hochmair (2011), Zielstra and Hochmair (2012), Zielstra et al. (2013)).

- Tools to conduct analyses are complicated and not user friendly.
Presentation Outline

- Development of OSM in the US
- Main Goal
- Case Study Preparation
- Case Study Results
- Summary
General US Data and Member Development

Development of Nodes, Ways and Relations in the US

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>0%</td>
<td>0%</td>
<td>47%</td>
<td>49%</td>
<td>76%</td>
<td>93%</td>
<td>100%</td>
</tr>
<tr>
<td>Ways</td>
<td>0%</td>
<td>0%</td>
<td>61%</td>
<td>63%</td>
<td>78%</td>
<td>92%</td>
<td>100%</td>
</tr>
<tr>
<td>Relations</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>43%</td>
<td>69%</td>
<td>100%</td>
</tr>
</tbody>
</table>
OSM Contributor Distribution in the US

Members with their first node created in the US (October 2012)
Geometry Errors and Positional Inaccuracy

- Problems with imported data:
  - TIGER/Line 2005 data is at times inaccurate (positional inaccuracy)
  - Some areas show topological errors (blue circle)
  - Data not up-to-date

Example: West Virginia

OSM 2012 → TIGER 2012
OpenStreetMap Data Quality Tools

- Geofabrik OSM Inspector (http://tools.geofabrik.de/osmi/)
- Keepright OSM (http://keepright.ipax.at/)
- Maproulette (http://maproulette.org/)
OpenStreetMap Community Effort in the US

- **Strong decrease of data errors in the US**
  - Unconnected OSM ways reduced in the past 17 months from around 141,000 to 19,000.
  - Duplicate way errors reduced from 17,500 to 11,500.

Source: [http://neis-one.org/2013/06/osm-us-2013/](http://neis-one.org/2013/06/osm-us-2013/)
Main Goal

- Evaluate OSM data quality for selected regions in the US

ArcGIS Editor for OpenStreetMap

ArcGIS Data Reviewer
Case Study Preparation

- **Retrieve OSM data with ArcGIS Editor for OpenStreetMap**
  - Add OSM basemap and zoom to desired extent
  - API limits size of downloadable content
  - Version 2.0 and higher of the OSM Editor (download to PostgreSQL SDE repository available)
  - [http://esriosmeditor.codeplex.com/documentation](http://esriosmeditor.codeplex.com/documentation)
Case Study Preparation

- **Visualize data and create topology**
  - OSM Editor tool creates a GDB with feature classes for points, lines and polygons
  - New Topology is required for Data Reviewer
Case Study Preparation

- **Define ArcGIS Data Reviewer rules**
  - What type of quality measures should be tested?
  - Create batch file with Data Reviewer to run all tests at once
  - Review and edit results
Case Study Results

- Eight data quality measures were tested for five selected areas in the US

1. Lines
   - Duplicate Geometries
   - Dangles
   - Orphans
   - Unnecessary Nodes
   - Cutbacks
   - Duplicate Vertex (Duplicate Nodes)

2. Polygons
   - Polygon Closed
   - Unnecessary Polygon Boundaries
Case Study Results

- 65 total errors found for five selected areas

- Visual detection sometimes impossible
- Large number of false positives
- Inexperienced mapper and lack of community
- Increase with complex features

<table>
<thead>
<tr>
<th>City</th>
<th>Duplicate Geometries</th>
<th>Dangles</th>
<th>Orphans</th>
<th>Unnecessary Nodes</th>
<th>Cutbacks</th>
<th>Duplicate Vertex</th>
<th>Polygon Closed</th>
<th>Unnecessary Polygon Boundaries</th>
<th>Total</th>
<th>Area Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miami</td>
<td>2</td>
<td>3</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>14.0 km²</td>
</tr>
<tr>
<td>Washington D.C.</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>9.0 km²</td>
</tr>
<tr>
<td>San Francisco</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>17.5 km²</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>9.0 km²</td>
</tr>
<tr>
<td>New York</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>12.5 km²</td>
</tr>
</tbody>
</table>

- False positives due to pedestrian data
- Good results due to OSM editors?
Case Study Results

- Example error types found in case study areas
  - Dangles
    - [Image of a dangle]
  - Duplicate Geometries
    - [Image of duplicate geometries]
What Now?

- **Make corrections to erroneous data**
  - Enable “Editor” toolbar and proceed as usual (e.g. snap dangles to nodes)

  and / or

- **Add new data with attributes to the OSM database**
Summary

- Combination of ArcGIS Editor for OpenStreetMap and ArcGIS Data Reviewer proved to be successful

- In total 65 data errors were found for five selected areas in the US

- Automated data error detection needs manual confirmation and correction (false positives)

- Data Reviewer allows not just visualization of errors but allows for systematic correction efforts

- In comparison to OSM error detection tools, ArcGIS Data Reviewer is not tailored to OSM specific requirements i.e. tags or features
References

- ArcGIS Editor for OpenStreetMap: http://www.esri.com/software/arcgis/extensions/openstreetmap

