

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

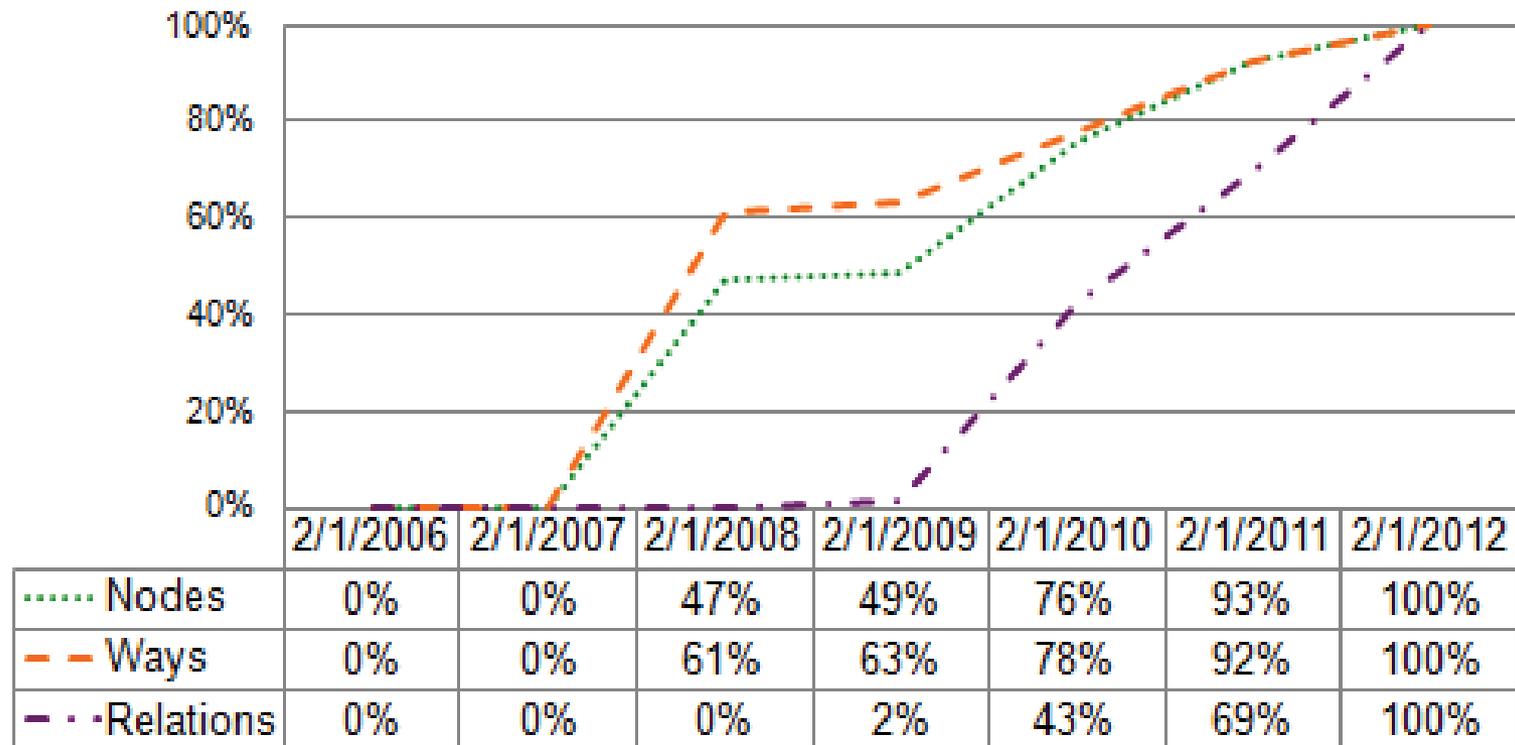
- ❑ Several studies within recent years about OSM data quality in comparison to governmental or proprietary data.
- ❑ Mostly with focus on Europe (i.e. UK, Germany).
- ❑ Usually good data quality results in countries with active OSM community
- ❑ First research results about OSM data quality in the US 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



OSM Contributor Distribution in the US

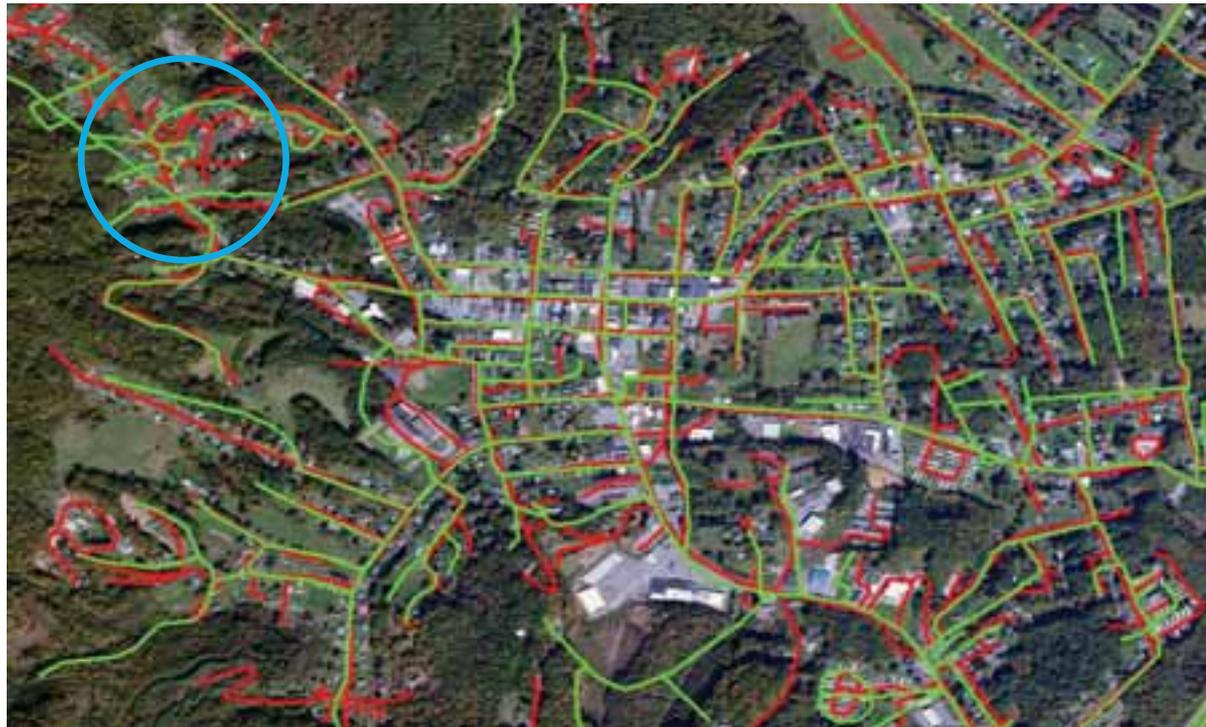
Members with their first node created in the US (October 2012)



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/>)

 Islands

up to 1m distance to
highway=footway

up to 2m distance to
highway=footway

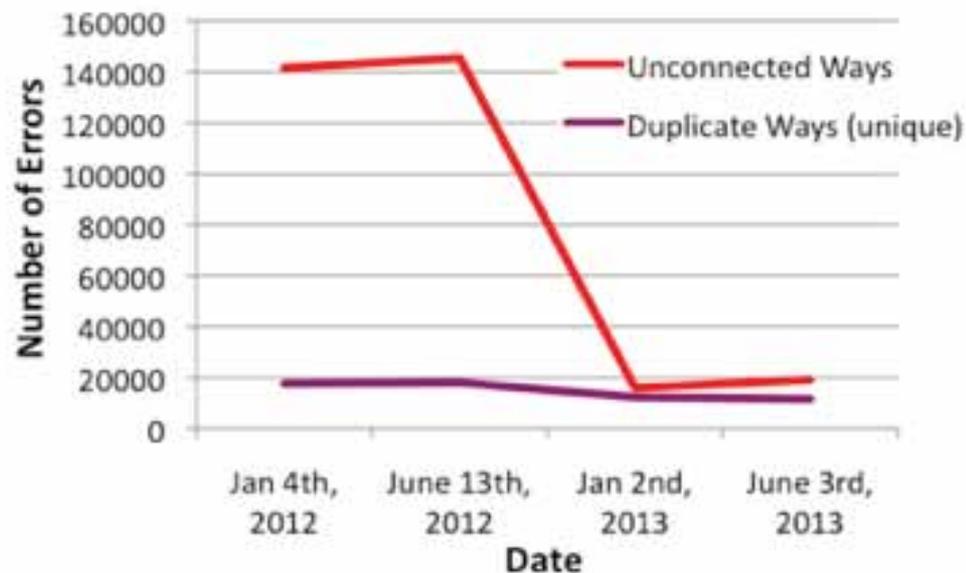
up to 5m distance to
highway=footway



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.

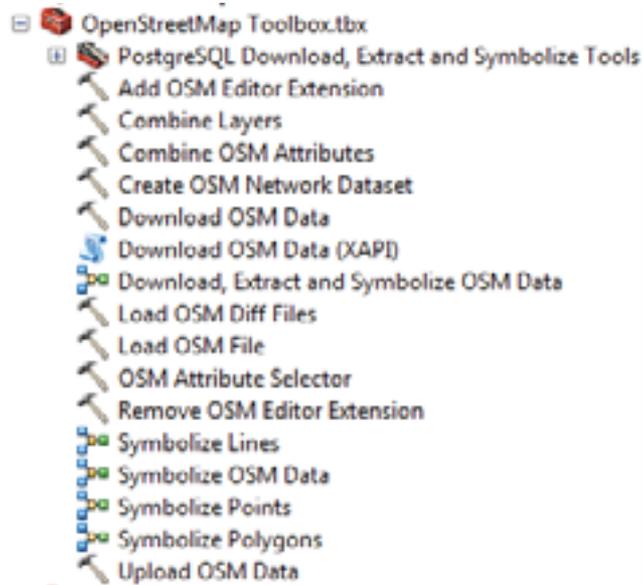


Source: <http://neis-one.org/2013/06/osm-us-2013/>

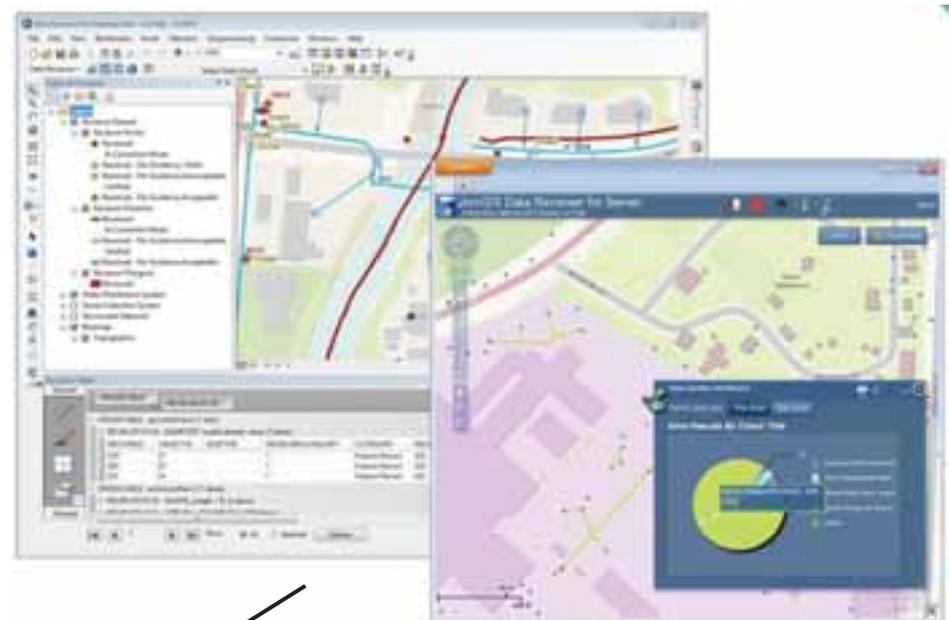
Main Goal

- ❑ Evaluate OSM data quality for selected regions in the US

ArcGIS Editor for OpenStreetMap



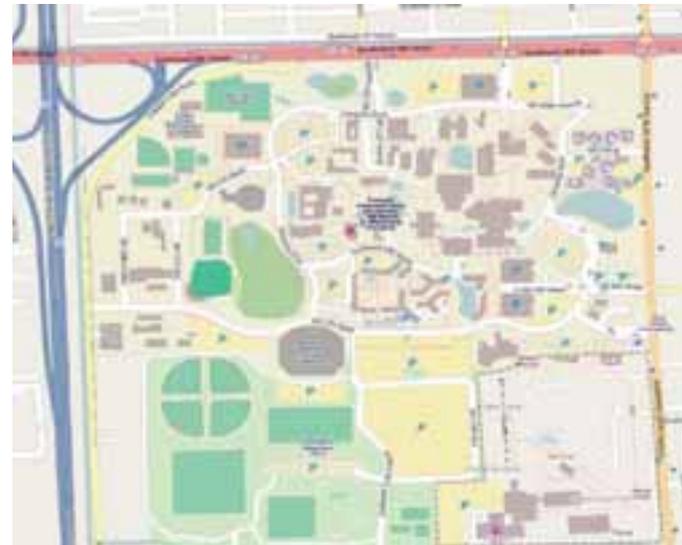
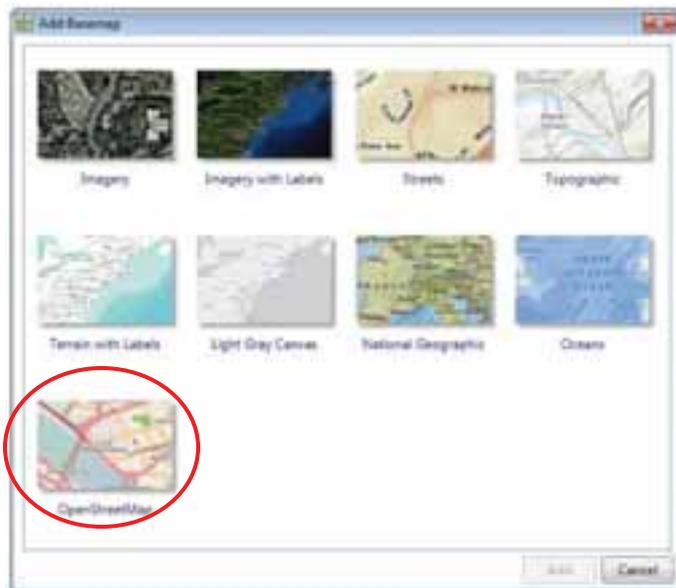
ArcGIS Data Reviewer



Data Analysis

Case Study Preparation

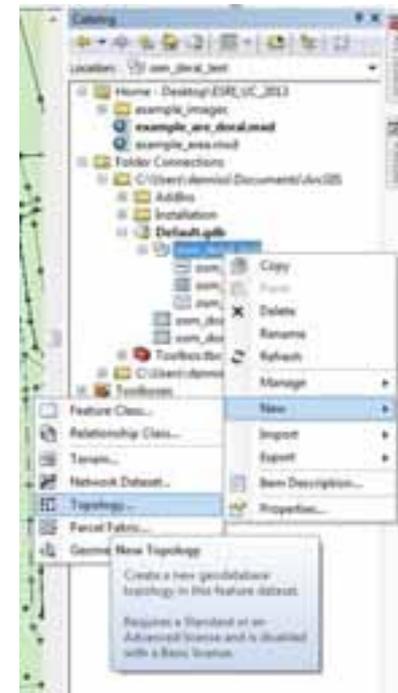
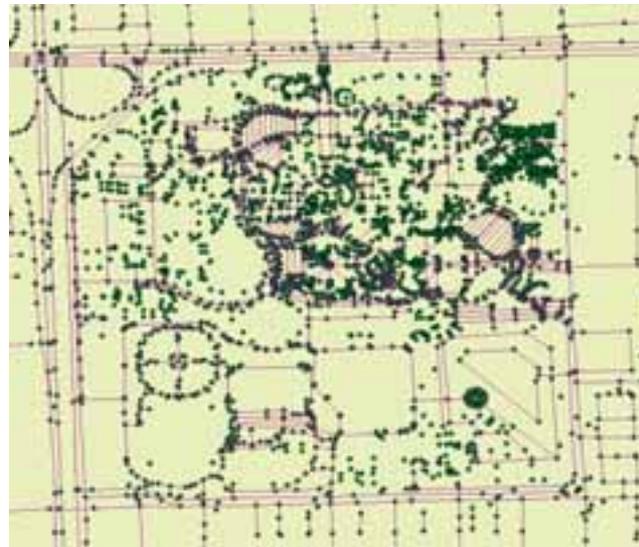
- ❑ **Retrieve OSM data with ArcGIS Editor for OpenStreetMap**
 - Add OSM basemap and zoom to desired extent
 - API limits size of downloadable content
 - Version 2.0 and higher of the OSM Editor (download to PostgreSQL SDE repository available)
 - <http://esriosmeditor.codeplex.com/documentation>



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

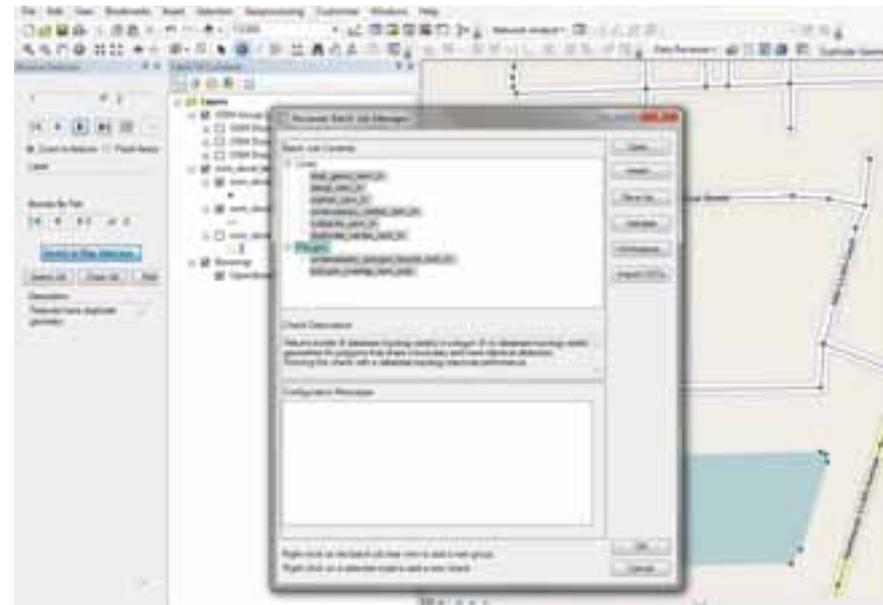
❑ Define ArcGIS Data Reviewer rules

- What type of quality measures should be tested?
- Create batch file with Data Reviewer to run all tests at once
- Review and edit results

Single Lines
Within a database topology, find polyline features that have nodes that are within a tolerance but not connected to other features in the database topology.
Comparison to Topology:
Use must not have single lines

Orphan
Find single polyline features that are not connected in the database topology.

Unnecessary Nodes
Find features that share a node and have identical attributes in editable fields.
Comparison to Topology:
Use must not have pseudo nodes.

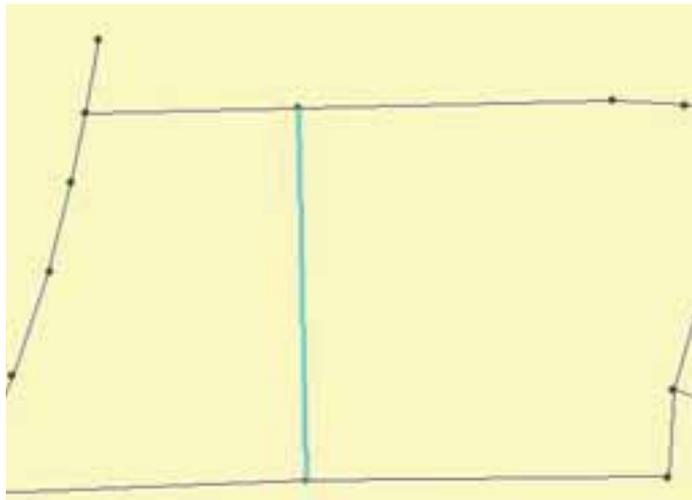


Case Study Results

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

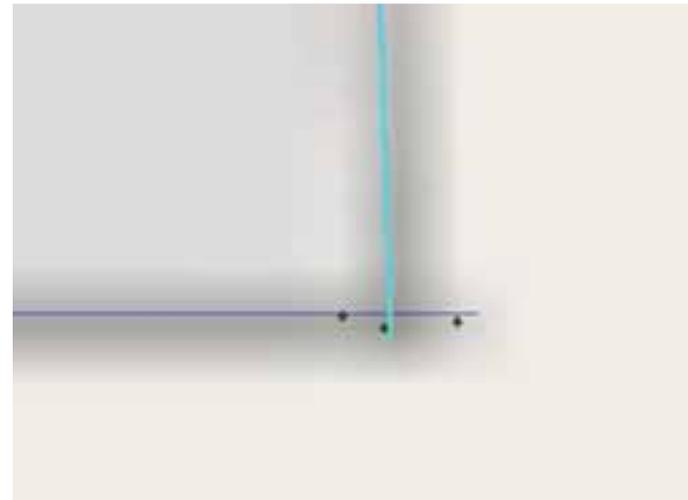
1. Lines

- 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

City	Duplicate Geometries	Dangles	Orphans	Unnecessary Nodes	Cutbacks	Duplicate Vertex	Polygon Closed	Unnecessary Polygon Boundaries	Total	Area Size
Miami	2	3	20	0	0	2	0	0	27	14.0 km ²
Washington D.C.	2	2	1	0	1	2	0	0	8	9.0 km ²
San Francisco	2	2	0	0	0	9	0	0	13	17.5 km ²
Los Angeles	0	1	1	0	0	7	0	0	9	9.0 km ²
New York	0	6	2	0	0	0	0	0	8	12.5 km ²

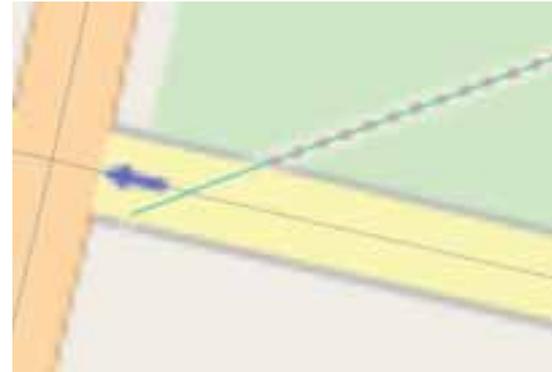
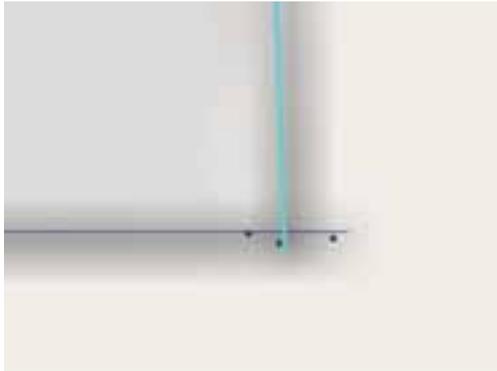
False positives due to pedestrian data

Good results due to OSM editors?

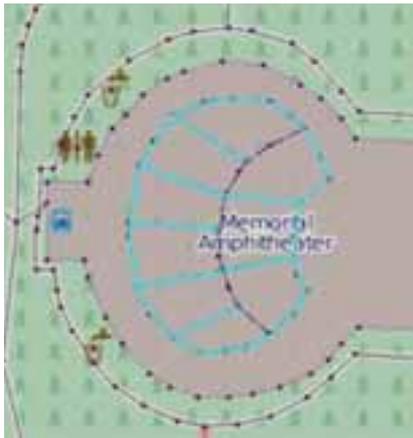
Case Study Results

❑ Example error types found in case study areas

➤ Dangles



➤ Duplicate Geometries



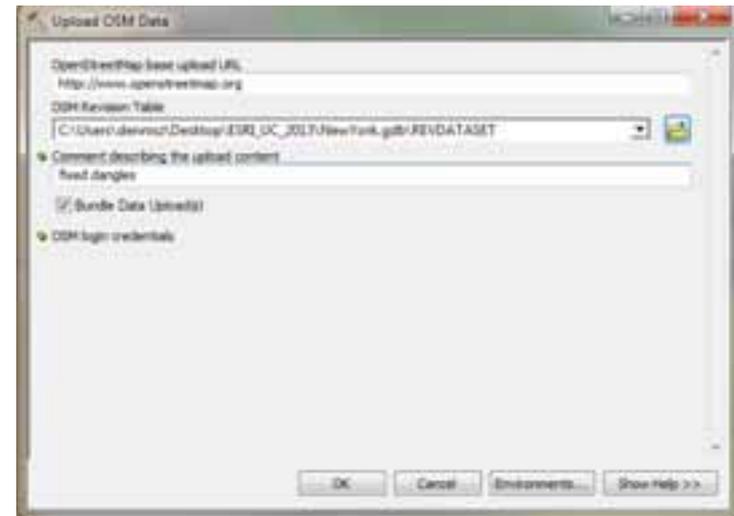
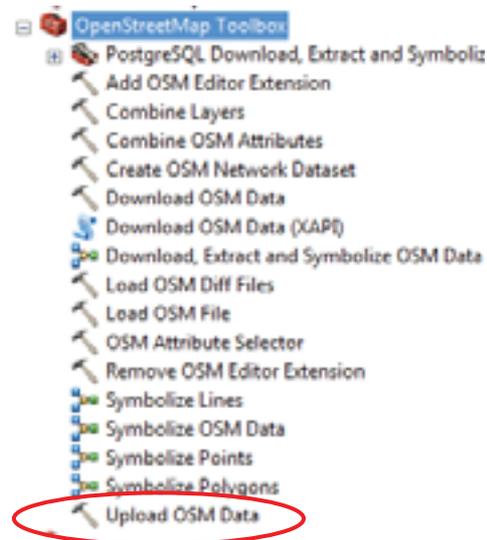
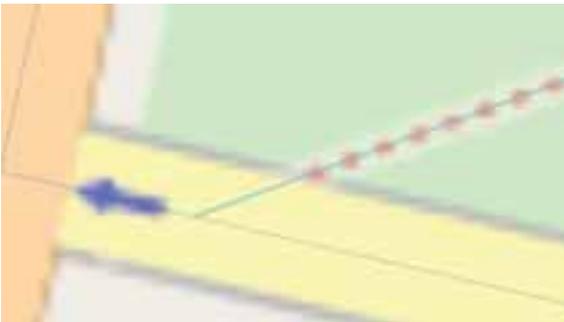
What Now?

❑ Make corrections to erroneous data

- Enable “Editor” toolbar and proceed as usual (e.g. snap danglers 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>
- ❑ ArcGIS Data Reviewer Extension:
<http://www.esri.com/software/arcgis/extensions/arcgis-data-reviewer>

- ❑ Zielstra, D. and Hochmair, H. H. (2011). A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Board: Journal of the Transportation Research Board*, Volume 2117, pp. 145-152.

- ❑ Zielstra, D., and Hochmair, H. H. (2012). Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2299, 41-47.

- ❑ Zielstra, D., Hochmair, H. H., and Neis, P. (2013). Assessing the Effect of Data Imports on the Completeness of OpenStreetMap – A United States Case Study. *Transactions in GIS*, Volume 17, Issue 3, 315-334.