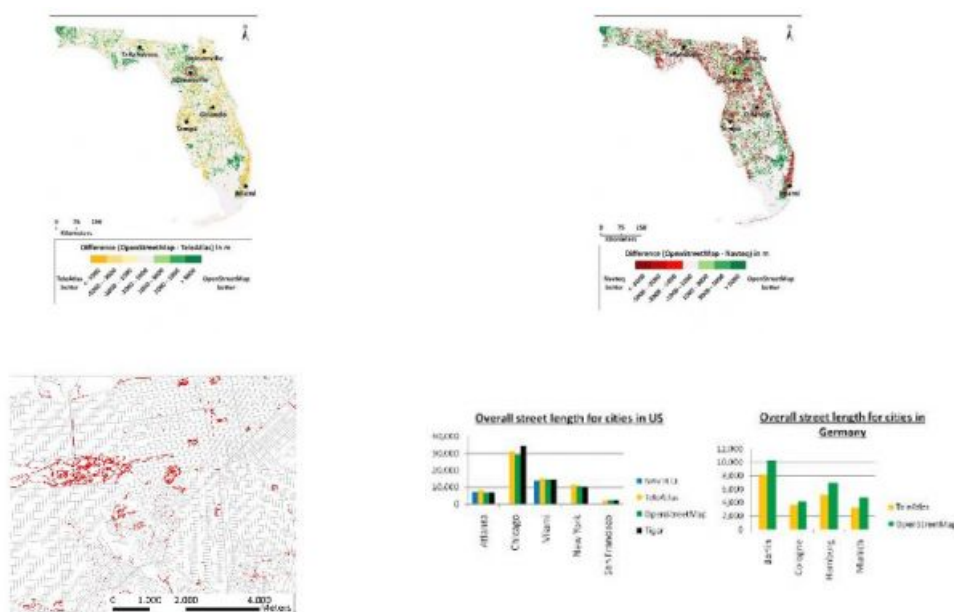


FREE VERSUS PROPRIETARY

Digital Street Data



Strong demand in recent years for freely available spatial data has boosted the availability of Volunteered Geographic Information (VGI) on the internet. The use of VGI in large-scale projects such as trip planners means an increasing need for quality assessment. While commercial data providers selling digital street maps provide a certain level of quality assurance there are no quality regulatory measures incorporated in VGI portals. The authors compare the completeness of freely available and proprietary street-network data for the state of Florida, and selected cities in the US and Germany.

The development of Web 2.0 and the Global Positioning System (GPS) and its integration into mobile phones, photo cameras and other mobile devices, allows web community members to interact, provide information to central sites and thus become a significant source of geographic information. VGI can be found in various web services and other digital data sources. Prominent examples of VGI include geo-tagged entries in Wikipedia, place descriptions in Wikimapia, and photographs in Flickr and Panoramio. Freely available street data have been collected in a collaborative volunteer effort in the OpenStreetMap (OSM) project over the past seven years and are available under certain licensing conditions.

Since VGI is primarily contributed by non-professional individuals generally with little experience or training and no protocols or standards are followed, the quality of VGI data may be lower than that of publicly administered or commercial datasets. Quality checks on VGI are therefore of particular importance in relation to data used in geospatial applications. This is also the case for data collections that combine data contributed by professional organisations with that provided collaboratively by volunteers. An example of such hybrid lineage is the integration of US Census TIGER/Line data with OSM for the United States.

Data Quality

The quality of geodata, and especially VGI, has become a major research topic over the past few years. The quality of one of the most successful VGI projects, OpenStreetMap, has already been tested for a few European countries. For example, data completeness, positional accuracy, attribute accuracy, and the participation inequality of OSM data were examined for selected regions in England and Germany.

The major motivation behind voluntary contribution to VGI data-collection efforts in European countries is that geospatial data layers such as land-use or street data are not generally provided by agencies free of charge. Various geospatial databases need therefore to be started from scratch, and contributors would see significant growth in the data layer through their personal contribution. This motivation may not be as high in the US, where selected base layers are already made publicly available

through federal agencies such as the US census bureau (e.g. TIGER/Line data) or US Geological Survey (e.g. Digital Line Graph data). Given these differing US and European policies, it is of interest to analyse whether the level of community-based contribution to street data also differs between the US and Europe, and how overall street lengths vary between the diverse data sources in various countries.

Analysis of Datasets

Although US agencies provide free street datasets, there are also a variety of commercial data providers offering proprietary datasets to paying customers. We use data from two such major players, NAVTEQ and TeleAtlas, and compare the completeness of these commercial data with freely available TIGER/Line and OSM data.

It must be noted that the results of the following analyses show a relative comparison of overall street lengths between the four available datasets, but what is not revealed is the absolute completeness of network data with respect to the real world (ground truth). However, the two commercial datasets we analyse, i.e., NAVTEQ NAVSTREETS and Tele Atlas Multinet street data, are widely used in commercial applications such as GPS car navigation systems. They can therefore be used as a relative reference for purposes of comparison, in particular with respect to navigation tasks. All four datasets, besides street geometries, offer a wider range of feature classes, such as traffic signals, which will not be further discussed here.

Data Completeness

We determine relative completeness of a road network by comparing total road length in grid-cells from different data providers. A difference in overall street length per grid-cell indicates that one dataset is more complete than another. We applied this method for the State of Florida using a 1km² grid. Visualisation of length differences in such a grid allows identification of local variation in completeness between the datasets. Figure 1 shows differences between OSM and TeleAtlas data for Florida, computed as street length for OSM minus street length for TeleAtlas in 1km² grid cells. Negative values (shown in orange) indicate stronger coverage in the commercial dataset, as is found in most urban areas. There are also areas with a higher street density in OSM data (green), such as the Gainesville area (circled). A similar result can be seen in Figure 2, where NAVTEQ shows stronger data coverage in urban areas (red) than OSM, which has a higher density of streets in rural areas (green).

Results in Context

These results provide new insight into the patterns of street coverage given by the OSM dataset in the US. In contrast to the OSM coverage pattern observed for England and Germany, OSM data coverage is generally higher in rural than in urban areas when compared to commercial datasets. The good coverage results for OSM in rural areas results not from user contribution, but is primarily due to TIGER/Line import. This is because TIGER/Line contains more data, especially for agricultural areas, than provided by those commercial data providers who do not import any TIGER/Line street data.

Despite the lower coverage of OSM in urban areas compared to commercial datasets, comparison between OSM and TIGER data reveals that in some urban areas OSM data are actively collected by the web community, in particular alleys and pedestrian segments. Figure 3 shows the TIGER/Line street network for San Francisco in black, with additional OSM pedestrian segments overlaid in red.

We can also compare street lengths for complete cities. Figure 4 (top) shows the overall length of all used street types for entire urban street networks in five US cities, comprising segments accessible to cars/pedestrians and pedestrians only. Differences in street lengths between the four data providers are small for US cities, with no apparent dominance in coverage for a specific data provider. The somewhat higher overall length value for TIGER/Line in Chicago is the result of a particular classification scheme in TIGER/Line whereby private streets in industrial areas, such as quarries, are also classified as local neighbourhood and rural streets. A more distinct pattern can be found for cities in Germany, where total lengths are clearly higher for OSM than for TeleAtlas (NAVTEQ was not available for comparison). Differences in total length range between 13% for Cologne and 44% for Munich (Figure 4, bottom). The reason for this difference is the apparent abundance of alleys and pedestrian paths in German cities compared to the US cities, and a more active OSM community.

Concluding Remarks

As the two Florida maps suggest visually, there is strong heterogeneity of OpenStreetMap data for the US in terms of its completeness. Significant differences were observed between rural and urban areas, but in a contrasting pattern to that shown by European tests results. However, overall segment lengths in US cities are similar for all data providers. This may be explained by the fact that, although urban areas are better covered by commercial data, calculations of overall lengths for the cities were based on the entire county in which each was situated and so included length counts from both rural and urban areas.

Some users explain the generally lower level of OSM activity in the US by the large volume of already freely available datasets. However, the analysis also indicates significant development of OSM over recent months in some cities, such as San Francisco.

Acknowledgements

The authors thank NAVTEQ and Tele Atlas for their generosity in providing US sample datasets.

Further Reading

- Flanagan, A.J. and Metzger, M., 2008. The Credibility of Volunteered Geographic Information. *GeoJournal*, 72(3), pp137-148. - Goodchild, M. F., 2007. Citizens as Sensors: the World of Volunteered Geography. *GeoJournal*, (69), pp211-221.
- Haklay, M., 2010. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, Vol. 37, 4, pp682 - 703.
- Zielstra, D. and Hochmair, H., H 2011. A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. Transportation Research Record: *Journal of the Transportation Research Board*.
- Zielstra, D. and Zipf, A., 2010. OpenStreetMap Data Quality Research in Germany. GIScience 2010. Sixth international conference on Geographic Information Science. Zurich, Switzerland.

<https://www.gim-international.com/content/article/digital-street-data>
