Quantitative Studies on the Data Quality of OpenStreetMap in Germany

D. Zielstra[1], A. Zipf[2]

[1]University of Bonn (Germany), Department of Geography, Cartography Research Group
Meckenheimer Allee 172, D-53115 Bonn, Germany

[2]University of Heidelberg (Germany), Department of Geography, Chair of Geoinformatics
Berliner Straße 48, D-69120 Heidelberg, Germany

1. Introduction

As availability and complexity of Volunteered Geographic Information (VGI) (Goodchild, 2007) on the Internet are on the rise, questions about the quality of this information still remain. One of the most promising projects is OpenStreetMap.org (OSM). Initiated in 2004 OpenStreetMap gives all users the opportunity to use spatial data for their own projects without fees. The goal is to create a map of the world that will contain as much detailed information as possible, and this information is being collected by volunteers.

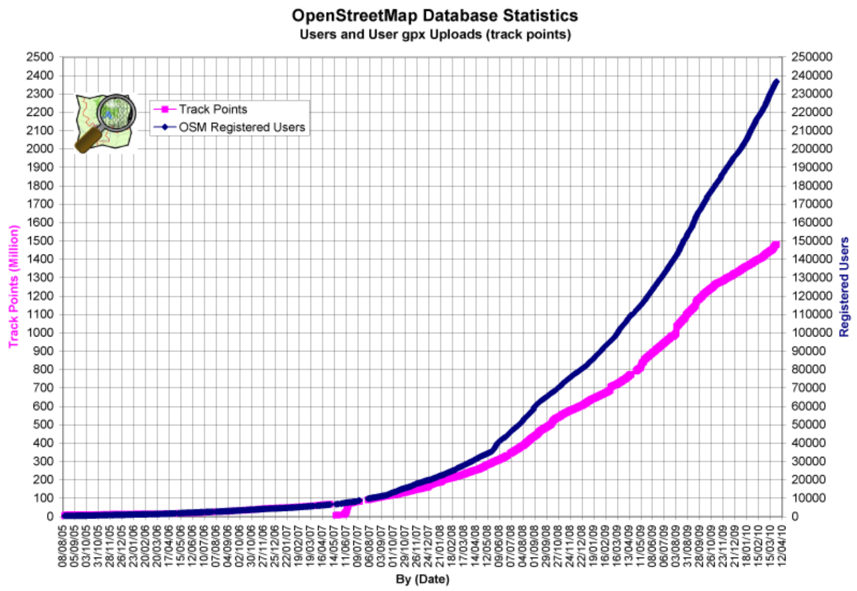

Figure 1. User- and Data development of OpenStreetMap (2005-2010)[1]

As figure 1 illustrates, both the membership numbers and the amount of data rose rapidly in an impressive manner. Since 2004, OpenStreetMap has collected, particularly in Europe, a large amount of geodata, with the greatest gains coming within the last two years.

However, to give objective information about the potential of this free data, it needs to be compared to professional datasets (e.g. Tele Atlas Multinet, Navteq) and applications based on this data need to be developed and tested (e.g. Routing, Geocoding etc.). In recent years the GIScience Research Group of the Geography Department at the University of Heidelberg (former Cartography Research Group at

[1] http://wiki.openstreetmap.org/wiki/Statistics

the University of Bonn), has conducted a variety of research, regarding these issues and created applications based on OpenStreetMap (Schmitz et al. 2008, Auer & Zipf 2009 or Amelunxen 2010).

Here we give an overview of selected quantitative research findings on the quality of OpenStreetMap in comparison to commercial datasets. It covers the timeframe of one year (April 2009 - April 2010). Therefore it includes the latest analysis that covers also the timeframe after the end of the original diploma thesis. Adapted from initial studies in England (Haklay, 2008; Ather, 2009), an updated analysis of the data covering all of Germany is presented (Zielstra, 2009, Zielsta & Zipf 2009). Further work within our group has focused on analysing Points of Interests (Strunck 2010) or Navteq data (Ludwig 2010).

2. Data Quality of OpenStreetMap in comparison to Tele Atlas

2.1 Relative Positional Accuracy

There are a variety of aspects that are important for geodata quality (van Oort 2006). In 2003 the International Organisation for Standardization (ISO) developed a new Standard, which includes the quality principles for geographic information (ISO 19113). One of the most obvious criteria of geodata quality is the positional accuracy of geometries. According to Tele Atlas, the positional accuracy of the Tele Atlas data is determined through a combination of laser scanner and GPS, which allows an accuracy of up to 1 meter (http://www.teleatlas.com/WhyTeleAtlas/FAQs/index.htm). On the other hand OpenStreetMap users typically have access to GPS only (with typically >+-5 meter accuracy, dependent from many factors). While we do not imply, that the positional accuracy of Tele Atlas is better in all situations we need to investigate the relative performance of OSM in relation to some other data set. As Tele Atals was available and is strongly used in the market (in particular the traffic and navigation sector) it is interesting how much (if at all) OSM differs from this “industry standard”. In Germany street data from government sources are typically not used for routing or LBS applications, but mostly for visualisation on maps. Therefore a comparison to the typical data vedors like TeleAtlas or Navteq is of more interest. Based on the positional accuracy measure for linear features (Goodchild and Hunter 1997), the relative (not absolute!) accuracy of OpenStreetMap street-network-data for five major cities and five medium-sized towns in Germany has been compared to the Tele Atlas MultiNet Dataset.

The results show a high overlapping percentage (>=80%) even within a ten meter buffer around the TeleAtlas street data in the bigger cities (Figure 2 & 3). This seems very good for many routing applications given the typical width of streets. On the other side the medium sized towns show fluctuation between fifty and almost eighty percent, which shows a first trend towards a correlation between OSM data quality and number of people (= mappers) within an area.

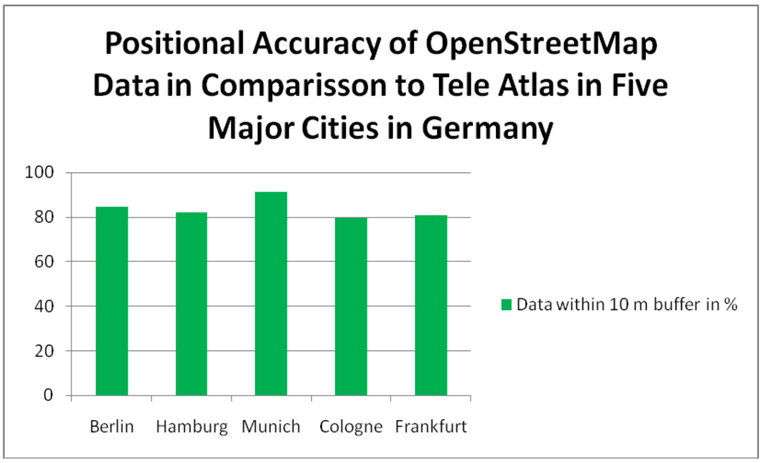

Figure 2. Relative Positional Accuracy of OpenStreetMap in Five Major German Cities

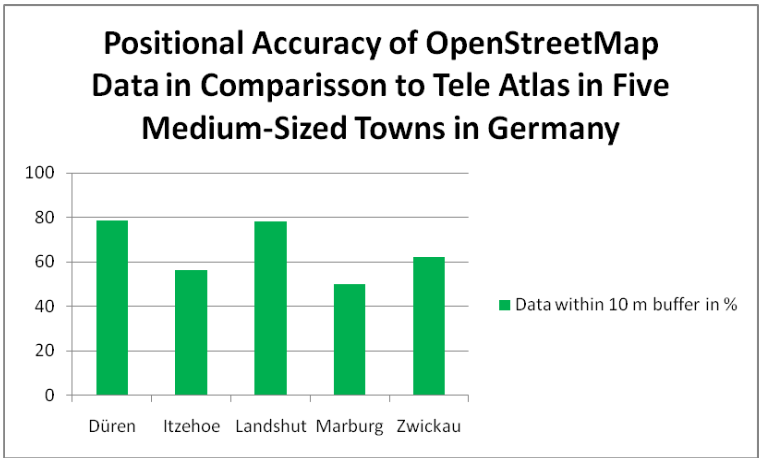

Figure 3. Relative Positional Accuracy of OpenStreetMap in Five Medium-Sized Germany Towns

2.2 Relative Completeness

Another major factor for geodata quality is the completeness of the data. Within the past year, four different OpenStreetMap data sets (April 2009, July 2009, December 2009 and April 2010) have been compared to TeleAtlas Multinet 2009. The relative completeness of a road network can be determined by calculating the total length of the roads of one dataset within a predefined area and then comparing it with the total length of roads of the other provider within the same area. If there is a difference in the overall lengths, it indicates that one of the datasets is more complete than the other (see also Haklay 2008). In the following diagrams and maps negative numbers represent a certain amount of missing data in the OpenStreetMap database compared to the Tele Atlas database, while positive numbers represent a higher amount of data in the OpenStreetMap Dataset, the latter meaning the OSM isin that regions "more complete" the Tele Atlas. This again shows that we deal here only with a relative comparison and can't give absolute statements about the completeness of one or the other data set in general.

As figure 4 shows the tremendous amount of data being collected by the volunteers has a big effect on the differences between the two datasets. Within just one year the difference between the entire street network has been reduced from almost thirty percent to almost zero percent, so it can be foreseen easily, that the total length of the entire street network of OSM soon will be larger than that of TeleAtlas data.

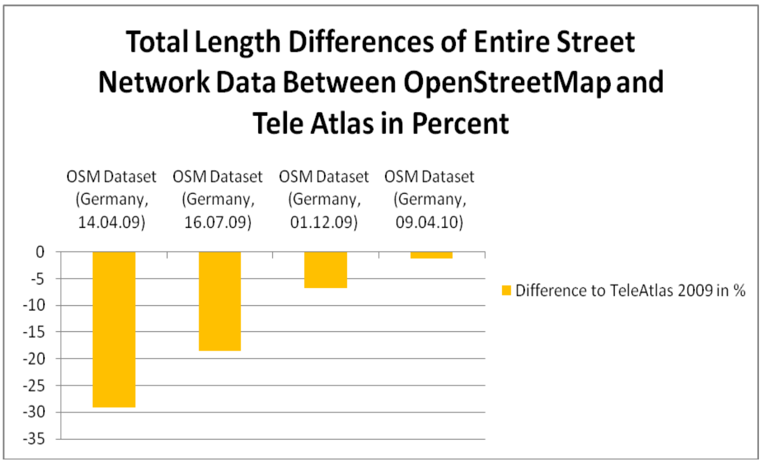

Figure 4. Comparison between OpenStreetMap and TeleAtlas with respect to the entire street network

Also the pedestrian related data in Figure 5 shows a data development during the past year and shows the focus of OpenStreetMap on smaller ways, alleys and streets that are typically not being covered by the professional data provider.

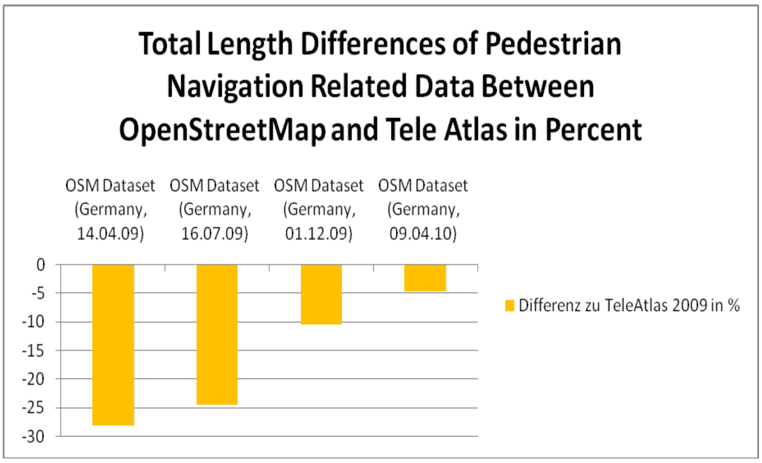

Figure 5. Comparison between OpenStreetMap and TeleAtlas with respect to pedestrian navigation

Only the data related to car navigation in figure 6 still shows that OSM is not "complete" yet and that the development here is much slower. This can be explained later when we look at the spatial distribution of the OSM data. The lack of completeness in rural areas leads to this result when calculating the values of Germany as a whole.

To provide more information about the regional development of OpenStreetMap both the total lengths of the street network of each dataset per square km as well as the differences by square km (OpenStreetMap minus TeleAtlas = difference in each grid area) have been calculated similar to the work by Haklay (2008).

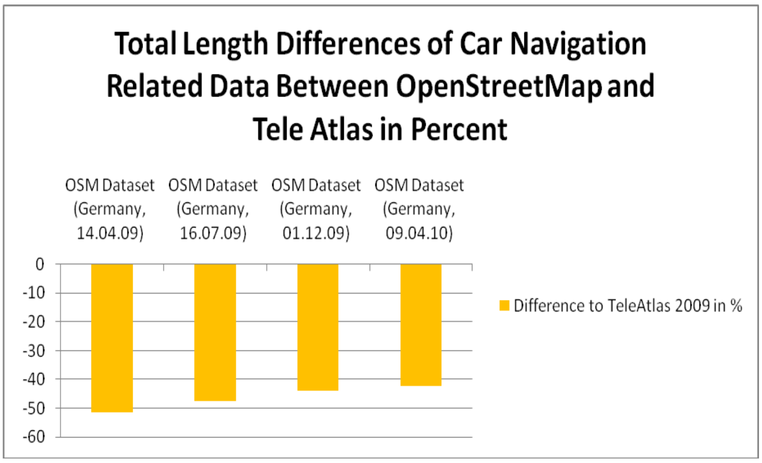

Figure 6. Comparison OSM to Tele Atlas with respect to car navigation

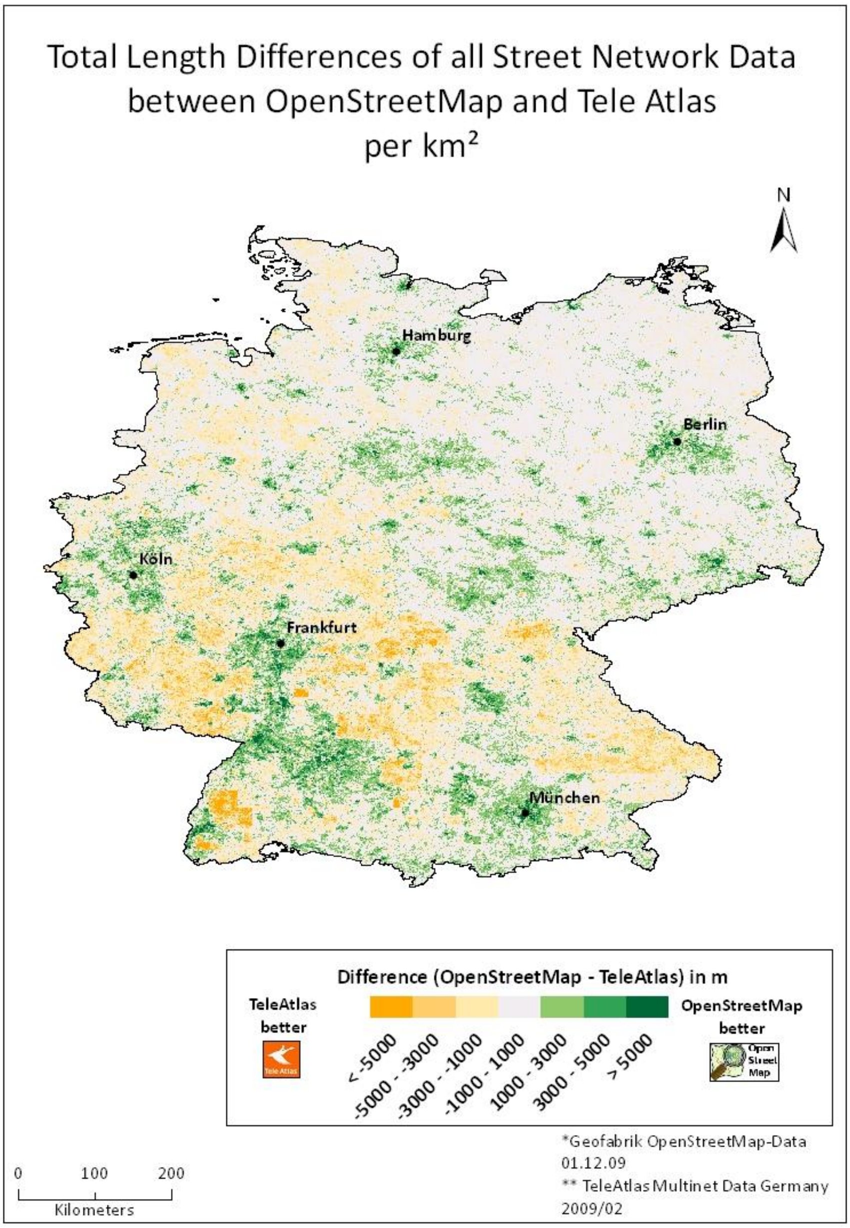

Figure 7: Length difference (OSM-TA) in absolute values (Dec. 2009) in Germany

Figure 7 shows the map with the results of the difference of the OpenStreetMap dataset of Dezember 2009 and Tele Atlas 2009 with respect to the total street network of Germany. The green areas in the map show a stronger concentration of OpenStreetMap data in the urban areas, compared to the rural areas, which are coloredin orange, indicating more TeleAtlas data per square kilometer.

Figure 8 gives this information as relative percentage with respect to the reference dataset Tele Atlas in order to unmask the heterogeneous length of the underlying street network.

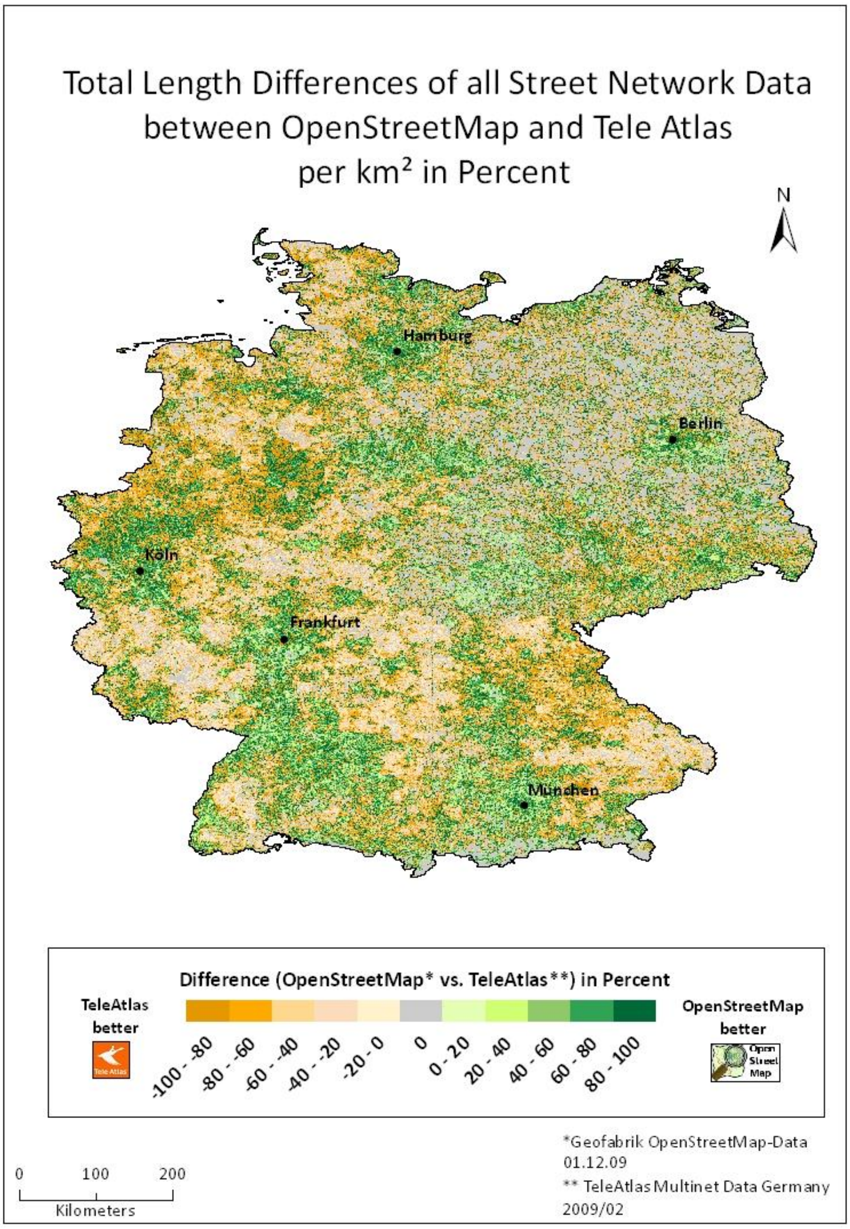

Figure 8: Length difference (OSM-TA) in relative values (Dec. 2009) in Germany

3. Conclusion and Future Work

The results presented show that OpenStreetMap is a very fast growing and expanding project. The amount of data collected by volunteers in Germany has been tremendous and will cause OpenStreetMap to pass Tele Atlas in the near future in the total length of all street network data. As the results do show also, there is still a very strong heterogeneity of the OpenStreetMap data in terms of their regional completeness. It remains to be seen if and when the "wikification of GIS" (Sui 2008) will also cause the OpenStreetMap community to spread to the rural areas and to also concentrate on car navigation related data to create a "complete" and accurate dataset of entire Germany. Although most of the data in OpenStreetMap is being collected by volunteers with just few or no knowledge of professional measurement and probably mediocre equipment the accuracy of the street data was good in major cities and still acceptable in mid-sized towns.

The potential of OpenStreetMap for the future can be big if its current membership continues to stay active and new members can be gained. However, it became clear that the coverage of OpenStreetMap in rural areas is still much lower and also does not grow that fast as the densely populated areas. If coverage is needed mostly in the densely populated urban areas of Germany (e.g., by regional traffic providers or logistics companies), OpenStreetMap may already be an interesting - and very cost-efficient - alternative to commercial datasets.

References

Ather, A., 2009 A Quality Analysis of OpenStreetMap Data, M.Eng. Thesis, London.

Amelunxen, C. 2010 An Approach to gecoding based on volunteered Spatial Data. Zipf, A. et al. (eds): Geoinformatik 2010. Die Welt im Netz. Kiel.

Auer, M. and Zipf, A. 2009 How do Free and Open Geodata and Open Standards fit together? From Sceptisism versus high Potential to real Applications. The First Open Source GIS UK Conference. University of Nottingham. UK.

Goodchild, M. F. und Hunter, G. J., 1997 A simple positional accuracy measure for linear features. International Journal of Geographical Information Science, 11(3), pp. 299-306.

Goodchild, M. F., 2007 Citizens as sensors: the world of volunteered geography. GeoJournal, (69), 211-221.

Haklay, M., 2008 How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England, Under review in Environment & Planning B: Planning and Design.

International Organisation for Standards (2002): ISO 19113 Geographic information – Quality principles.

Ludwig, I. (2009): Abbildung von Straßendaten für Qualitätsuntersuchungen - Ein Vergleich von OpenStreetMap mit Navteq. Diplomarbeit. Lehrstuhl Prof. Zipf.

Oort, P. A. J. van, 2006 Spatial data quality: from description to application, PhD Thesis, Wageningen: Wageningen Universiteit.

Schmitz, S., Neis, P. and A. Zipf, 2008: New Applications based on Collaborative Geodata - the Case of Routing. XXVIII INCA International Congress on Collaborative Mapping & SpaceTechnology, Gandhinagar, Gujarat, India.

Strunck, A. (2010): Raumzeitliche Qualitätsuntersuchung von OpenStreetMap. Diploma Thesis. Prof. Zipf.

Sui, D. Z., 2008 The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. Computers, Environment and Urban Systems, 32(1), p 1-5.

Zielstra, D., 2009 Datenqualität und Anwendbarkeit von Volunteered Geographic Information. Vergleich von proprietären und frei verfügbaren Geodaten. Diploma Thesis, Prof. Zipf.

Zielstra, D. & Zipf, A., 2009: Datenqualität von OpenStreetMap - Erste Ergebnisse empirischer Untersuchungen. AGIT 2009. Symposium für Angewandte Geoinformatik. Salzburg. Austria.