

Article

The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011

Pascal Neis ^{1,*}, Dennis Zielstra ² and Alexander Zipf ¹

¹ Geoinformatics Research Group, Department of Geography, University of Heidelberg, Berliner Street 48, D-69120 Heidelberg, Germany; E-Mail: zipf@uni-heidelberg.de

² Geomatics Program, University of Florida, 3205 College Avenue, Fort Lauderdale, FL 33314, USA; E-Mail: dzielstra@ufl.edu

* Author to whom correspondence should be addressed; E-Mail: neis@uni-heidelberg.de; Tel.: +49-6221-54-5504; Fax: +49-6221-54-4529.

Received: 1 December 2011; in revised form: 16 December 2011 / Accepted: 19 December 2011 /

Published: 29 December 2011

Abstract: The OpenStreetMap (OSM) project is a prime example in the field of Volunteered Geographic Information (VGI). Worldwide, several hundred thousand people are currently contributing information to the “free” geodatabase. However, the data contributions show a geographically heterogeneous pattern around the globe. Germany counts as one of the most active countries in OSM; thus, the German street network has undergone an extensive development in recent years. The question that remains is this: How does the street network perform in a relative comparison with a commercial dataset? By means of a variety of studies, we show that the difference between the OSM street network for car navigation in Germany and a comparable proprietary dataset was only 9% in June 2011. The results of our analysis regarding the entire street network showed that OSM even exceeds the information provided by the proprietary dataset by 27%. Further analyses show on what scale errors can be reckoned with in the topology of the street network, and the completeness of turn restrictions and street name information. In addition to the analyses conducted over the past few years, projections have additionally been made about the point in time by which the OSM dataset for Germany can be considered “complete” in relative comparison to a commercial dataset.

Keywords: Volunteered Geographic Information (VGI); OpenStreetMap; geodata; quality assessment; Germany; street network

1. Introduction

The OpenStreetMap (OSM) project has a history of nearly seven years now (2011). Similar to Wikipedia, the information gathered can be described as User-Generated Content (UGC) [1,2]. However, unlike Wikipedia, it is not encyclopedia information that is being gathered; instead, users are contributing their geodata to OSM. This data of geographical relevance is compiled by volunteers, saved in a database, and made ‘freely’ available to everyone via the World Wide Web (WWW) [3,4]. OSM is a well-known project in the field of Volunteered Geographic Information (VGI) [5–7], which others also describe as crowd sourced (geodata) [8–11]. Furthermore, especially in connection with the term of Web 2.0 [12], it is also referred to as Neogeography [13–17]. Others again describe it as Collaborative Mapping [18] or the Wikification of GIS [19]. The successful development of user generated content in recent years had an increasing impact on a variety of research fields and particularly OSM has been the focus of many new developments such as routing applications, 3D city models and Location-Based Services (LBS) [20–23].

Within the past few years, the OSM membership numbers have rapidly developed from a few hundred in mid-2004 to more than half a million registered members in November 2011. But what drives such a large number of people to participate in a voluntary project, and where lies their motivation? A few suggestions have been made to describe the motivation of the volunteers, such as a certain need for self-representation by the members or the project’s fun factor, and a degree of interest in technical terminology, equipment, and the WWW [7]. Either way, according to our research, approximately 150 new active members have joined the project each day since the beginning of 2011. Despite these numbers, OSM tends to experience a problem that is similar to other online portals that are based on UGC: participation inequality. This term describes the phenomenon of a 90-9-1 rule that most of the projects inherit [24]. Ninety percent of all members merely consume and are described as lurkers, 9% contribute to the project at irregular intervals, and just 1% of the members is actively involved and counts for the largest number of contributions to the project. A similar scenario exists within the OSM project. At the beginning of 2008 there were 30,000 total members in OSM, and about 10% of them actively contributed to the project [11]. In 2009, there were approximately 200,000 registered members, yet still, only roughly 10% of these were active members [25] and contributed to the project. In the same year 98% of all project data was provided by approximately 5% about 10,000 members. In 2010, only 5% of the 330,000 members were active contributors, and 98% of the data was provided by about 3.5%, which represents 12,000 of the total registered members.

While some countries such as the USA and France imported large datasets from other providers of freely available data (e.g., from governmental agencies such as the US Census Bureau), within the scope of the project, in Germany, OSM relies on its large number of participants. In 2009 nearly 50% of the entire changes in the OSM database were made within Germany [26]. However, in 2010 this value lowered to approximately 30%.

Despite this decrease, the aforementioned numbers give a first impression of how the OSM project has become a potential competitor to public and commercial geodata providers, not just in Germany but also worldwide. We see a revolutionary paradigm shift on how map data is being collected. MapQuest and Bing Maps are some of the first international companies that are reacting to this trend and have started offering maps based on OSM data (e.g., open.mapquest.com) or additional LBS's such as route planning, and address and area search functions on their websites.

Although the OSM project shows a high membership number and contributors are active, it needs to be noted that most of the VGI projects, and thus also OSM, rely on volunteers that do not necessarily have professional qualifications and background in geodata collection or surveying [6]. Furthermore, contributing to the project depends largely on technical aspects such as specific equipment, e.g., a PC/laptop, an internet connection, or potentially also a GPS receiver or Smartphone. The population density of the specific areas naturally plays a role too. Thus, the probability that more densely populated areas are mapped or more complete than sparsely populated areas is beyond question. However, the local knowledge of most participants should in fact make them local experts [27]. This raises numerous questions: How does the data perform in comparison (relative completeness and attribute accuracy) with other proprietary geodata providers? Can a difference be detected between urban and rural areas? How has the OSM project data developed in recent years? Can a projection be made for the data's future development?

The remainder of the paper is structured as follows: The following section gives an overview of prior OSM quality research conducted in recent years. Also, the study area and data preparation steps applied during our research are being discussed. The next section provides the results conducted during our research with regards to the OSM street network evolution for the years 2007 to 2011. This is followed by a discussion of the results and aspects of future work.

2. OpenStreetMap Quality Assessment History

Numerous scientists have investigated the quality of VGI and particularly OSM in recent years, and further research is currently still being conducted. In 2008 a discussion about the need for research with regard to the accuracy and correctness of the compiled information within the world of Web 2.0 was sparked [4,7,28,29]. Preliminary direct comparison analyses with regard to OSM were conducted for Great Britain in 2008 comparing Ordnance Survey (OS) geodata with OSM [30,31]; in Germany in 2009 [32] OSM data has been compared with the commercial Multinet dataset from TomTom (still known as TeleAtlas at the time). A few months later a similar comparison for Germany was conducted, but with the street dataset from Navteq, a different proprietary geodata provider [33]. Both analyses came to similar conclusions despite using slightly different methods: OpenStreetMap data shows a high degree of detail in urban areas; however, this detail richness declines significantly in rural areas. The main difference between the two studies was that one analysis included methodology to show the geographical discrepancies within Germany [32], while the other merely advised how complete OSM is in relative comparison with another dataset.

In France, a similar approach was used to analyze OSM data and further studies were conducted at the same time [34]. The results showed the advantage and flexibility, but also the problem of the

heterogeneity of the data specifically for this country. The latter is the result of the different data sources that have been used in OSM and also the differences in the work by the project participants in France.

In 2011, the first studies that analyzed the quality of OSM outside of Europe were conducted [35]. In this particular case the OSM project data has been compared with proprietary data from TomTom (TeleAtlas) and Navteq for the entire state of Florida (USA) and four specific cities within the USA. In comparison to the results for Germany or England, the discrepancies between the rural and urban areas in the USA showed an opposite tendency. In Florida, the rural data was, in parts, even more complete than that of the proprietary datasets in the relative comparison conducted. This can probably be attributed to the TIGER (U.S. Census Bureau) street data import to the OSM database for the entire United States. Other analyses were conducted comparing the impact of OSM on shortest path generation for pedestrians in Germany and the US [36,37]. Apart from England, no studies have been conducted to date over a period of several years and for an entire country [38].

2.1. Study Area and Data Preparation

A long history in geodata quality research has provided a variety of publications dealing with the definition of characteristics of geodata that can be used as quality parameters [39,40]. In 2002 the International Organization of Standardization (ISO) set a standard that defines the quality attributes of geodata in ISO 19113:2002 (principles for describing the quality of geographic data) and ISO 19114:2003 (framework for procedures for determining and evaluating quality). The defined parameters for the quality of geodata according to ISO 19113:2002 are completeness, logical consistency, positional accuracy, temporal accuracy, and thematic accuracy.

Within the scope of this article, we shall consider all these parameters with the exceptions of positional and thematic accuracy. The completeness of the street network is determined via a relative comparison between OSM and a commercial dataset provider. Furthermore, we will show the development in the urban and rural areas over a certain time period. The logical consistency will subsequently be evaluated with the help of an internal test, whereby topological and thematic consistency will be determined. The temporal accuracy will be verified in a simple form by means of an object's time stamp in the OSM dataset.

The study area for this article relates to all of Germany; however, each of the studies conducted takes place on different scales with the smallest scale being the municipal level. A variety of OSM datasets with three-month intervals were used starting from January 2007 to June 2011. Overall, 19 datasets were prepared for Germany, representing the different dates. In addition to clipping the Germany dataset from the entire OSM database dump-file (<http://planet.openstreetmap.org/>) for each point in time, it proved to be a challenge to work with the different API (application programming interface) versions to extract the data. The first dataset included in the analysis (2007) was taken from OSM API Version 0.4; however, Version 0.6 is the latest version (2011) of the API. The data could be converted, but unfortunately it did not always feature all of the latest attributes. With the older version of the API ways would still consist of segments in the OSM database, while with API 0.5 ways were mapped by referencing to a node. Furthermore, since API 0.6 appeared in April 2009, anonymous edits in OSM are no longer permitted, and a user ID and user name have since been included with every object. This means that at certain points during our analysis it was not possible to retrieve information

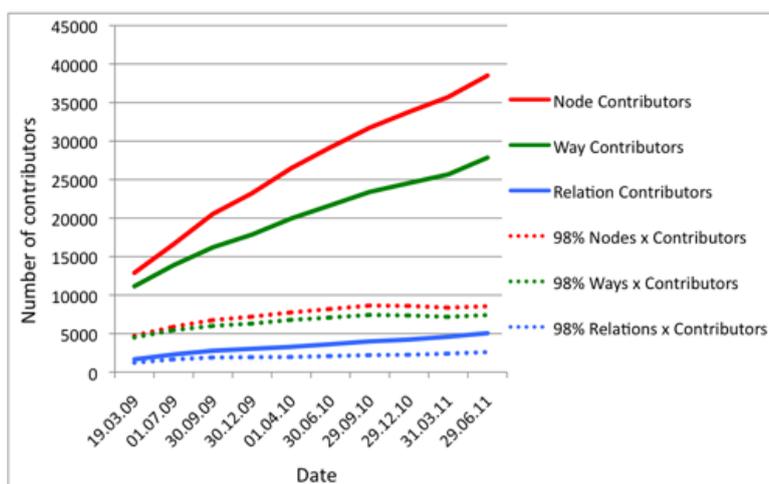
for the entire time frame up to 2007, but instead only to 2009. For the comparison analyses, the TomTom Multinet 2011 commercial dataset has been used. The respective data was imported into a PostgreSQL/PostGIS database with OSM's OSMOSIS program. In contrast to other available OSM programs that import the data into a database, the application used has the advantage of not filtering, preparing, or optimizing the data during the import procedure to the database. All analysis procedures in this paper are either based on PostgreSQL/PostGIS functions or specific tools developed in Java implementing the GeoTools open source toolkit.

3. OSM Street Network Evolution

3.1. User Activity and Data Development

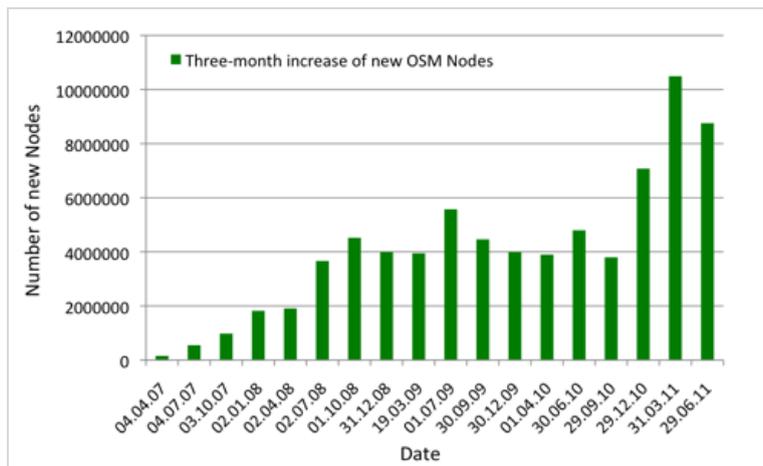
The number of OSM participants in Germany increases from year to year. To date (June 2011), a total of more than 40,000 different members have actively contributed to the project. Slightly different numbers of contributors have generated the three OSM object types: Nodes, Ways and Relations (Figure 1). Another interesting fact, also presented in Figure 1, is shown by the lines in orange, light blue, and purple, representing the users that generated a total of 98% of the data volume of each respective object type. Here 98% of about 74 million Nodes can be attributed to approximately 8,500 members, 98% of about 11 million Ways to approximately 7500, and 98% of about 171,000 Relations to approximately 2,600. These numbers are based on the information in the database on who has been saved as the last owner of each Node, Way, or Relation.

Figure 1. Number of OSM Contributors in Germany from 2009 to 2011.

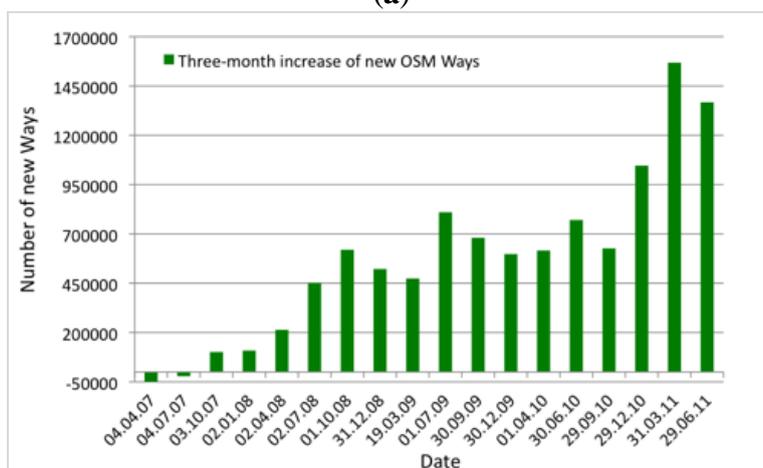


To be able to give more information about the results conducted during our analysis, we need to introduce the three object types that are used in the OSM project/database in greater detail. A Node is the basic object in the database and constitutes a coordinate. Ways represent lines or surface objects and constitute references to Nodes. Objects can be linked via Relations, which relate to each other. The results of our analysis showed a general pattern of an increase in the number of OSM Node and Way objects over the past few years (2007–2011) (Figure 2a,b), which was expected due to the general trend that OSM showed in Germany in recent years.

Figure 2. (a) Development of OSM Nodes in Germany; (b) Development of OSM Ways in Germany.



(a)



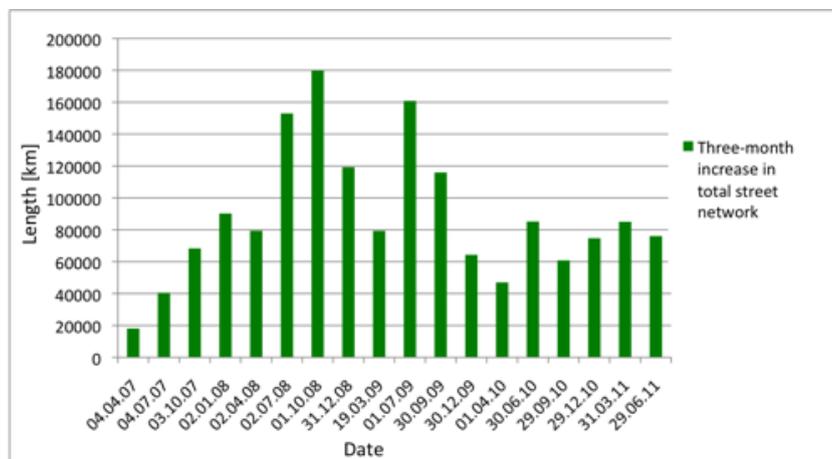
(b)

However, due to the three-month interval used in this analysis, other factors can be interpreted and distinguished. The data clearly shows that during the summer months the members are more active than during the winter months. Also, an above-average increase in data can be noticed at the turn of the year 2010/2011 and in spring 2011. The high proportion of new objects during these points in time can be attributed to the release of the Bing aerial images for digitalization purposes. The negative trend for Ways in OSM at the beginning of 2007 is due to the API changes during that year. With these changes the data schema and representation has been adjusted, and the total number of Ways was affected in this way; however, no data was lost by this change.

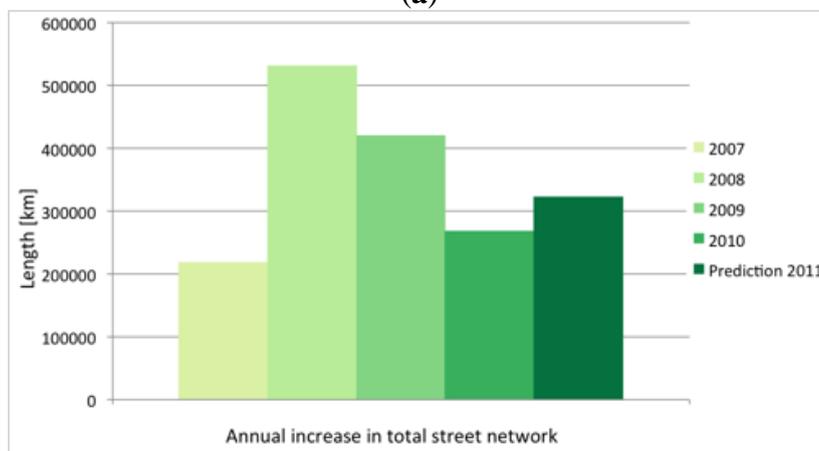
In prior studies the development of the German street network has only been compared to a commercial dataset (TomTom) for a period of eight months [32]. Neither definite statements about when different street types can be considered relatively complete nor a projection for the future were given. In our analysis the strongest increase in transportation-related routes in OSM for Germany to date could be distinguished in the third quarter of 2008 (+180,000 km) (cf. Figure 3a,b). The year 2008 was also when the most transport routes were added to the OSM database in general with a total length of almost 530,000 km. Since 2008 the annual expansion has decreased over time, and a slight change is discernible for the first half of 2011, where the trend slightly increases again. However, if this

tendency continues for the rest of 2011, a small increase in the total street network will be detected for this specific year.

Figure 3. (a) Increase in German OSM Street Network (three-month interval); (b) Annual Increase in German OSM Street Network (2007–2011).



(a)



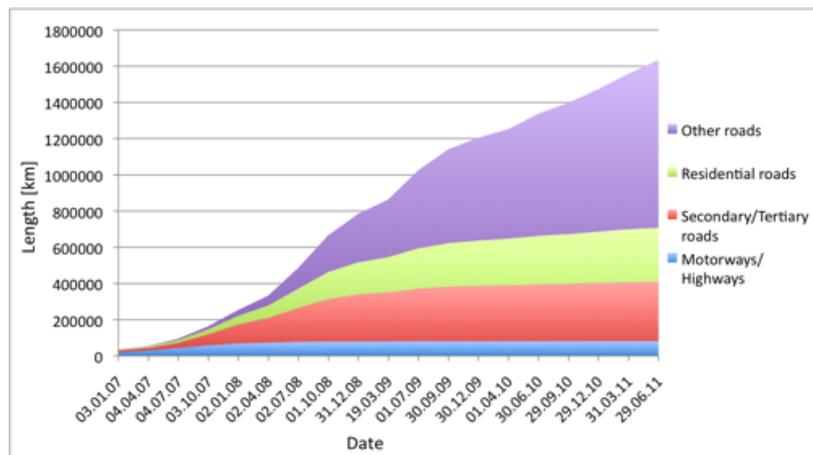
(b)

After gaining this first general impression about the development of the German OSM dataset, we conducted further analyses to give more detailed information about the street network. Due to the fact that every country’s street network consists of several different street categories, it seemed mandatory to consider these in our analysis. Thus, the various different street categories, which can also be found on the OSM Map Features web page (http://wiki.openstreetmap.org/wiki/Map_Features), have been divided into four groups for the sake of clarity and for enhanced research and comparison methods; namely, motorways/dual carriageway, district/municipal roads, roads to/in residential areas, and other roads such as service roads and dirt/forest trails (Figure 4).

Tracing the growth of the different categories, it can be noted that from a specific point in time, most categories do not expand any further. This indicates which category should be close to “completion” or in which category there are still new streets being added. It needs to be noted though that for comparison the TomTom dataset is suitable only for street network data for car-specific navigation, three out of the four categories. The “other routes” category can be compared only to

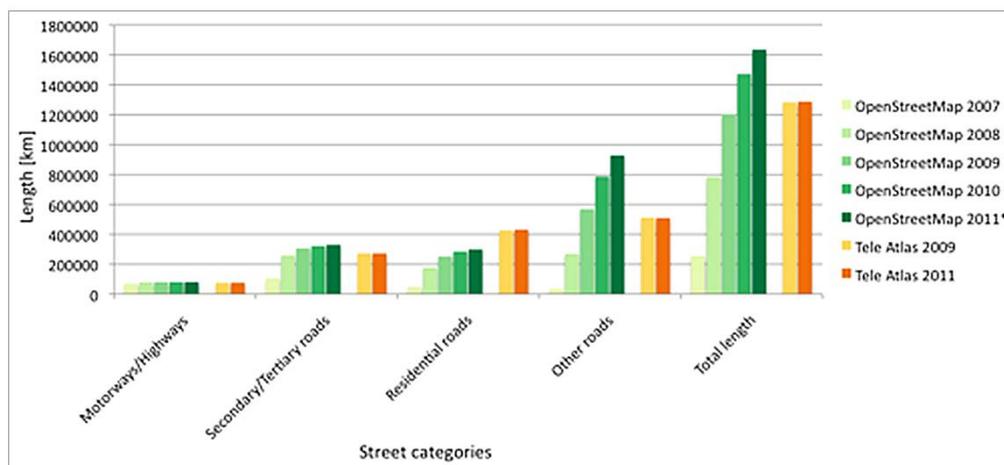
TomTom to a certain degree. In this fourth category, OSM has a far higher street network than the commercial provider. Based on the presumptions stated above and the comparison with the corresponding TomTom category street lengths, we reached the following conclusions. First, motorways and expressways were completely recorded for Germany by the middle of 2008. Second, all municipal roads for all of Germany were recorded by the middle of 2009. Third, streets that are close to or within residential areas are not fully recorded yet. Fourth, at the end of 2009 there were more segments in the “other routes” class in OSM than in the total TomTom commercial dataset. Fifth, in the middle of 2010 OSM surpassed TomTom in the total number of streets recorded. However, a high number of field and forest trails caused this advantage for OSM. Finally, most data contributions in 2011 are isolated street networks close to or within residential areas, but “other route” data, such as forest and field trails, are also increasing.

Figure 4. Development of OSM Street Network in Germany by Street Category (2007–2011).



The development of the individual categories in comparison to prior research results [32] and the commercial TomTom Multinet dataset from 2011 is depicted in Figure 5. The assumptions mentioned above with regard to the development and completeness of the total street network can here be confirmed too.

Figure 5. Development of OSM Street Network in Comparison to TomTom.



In June 2011, our studies for Germany showed that OSM had provided a street network for car navigation that is approximately 9% smaller than that of TomTom (Table 1). However, OSM’s total street network is approximately 27% larger in comparison with TomTom’s. In terms of pedestrian-related data and information, the OSM Germany dataset is even approximately 31% larger.

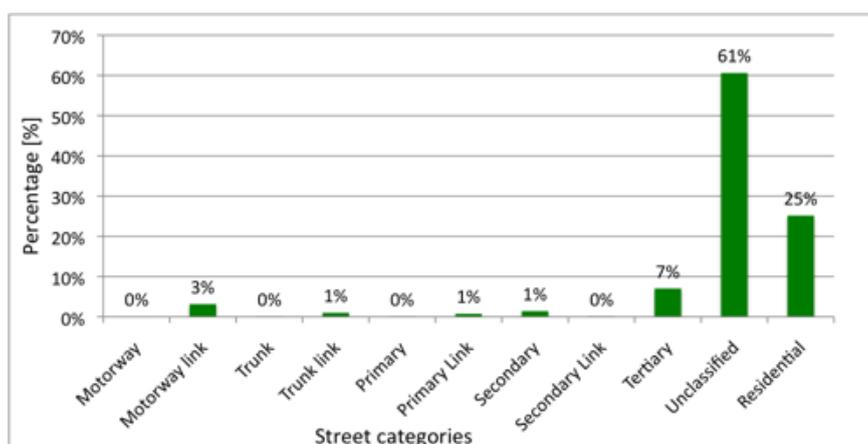
Table 1. Total Street Length of TomTom Multinet 2011 and OSM in June 2011.

Street Network	TomTom Multinet 2011	OSM June 2011	%
Total street network	Approximately 1,283,000 km	Approximately 1,630,000 km	OSM 27% longer street network
Street network for car navigation	Approximately 777,000 km	Approximately 705,000 km	TomTom 9% longer street network
Street network for pedestrian navigation	Approximately 1,185,000 km	Approximately 1,552,000 km	OSM 31% longer street network

In addition to the relative geometric completeness in comparison with another dataset, the internal completeness within the street network with regard to the street names is also important. This factor, sometimes also referred to as attribute accuracy [30], plays a significant role in applications such as routing applications that are being built on the specific dataset. Our results showed that a total of approximately 16% of streets in OSM have neither a name nor a route number (e.g., A 61) that could be used for car navigation. However, these results vary by street type in significant ways (Figure 6).

The results clearly show that the majority of the unnamed streets are streets that are either within or close to residential areas. The “unclassified” street category could lead to confusion in this case, since streets that have a linking function between villages are included within this category. Another reason for this high value could be the fact that many of these particular routes (e.g., country lanes) have been digitized from satellite images, thus the local knowledge to add the specific name of each route is missing.

Figure 6. Distribution of Streets without Name or Route Number Attribute Information by Street Category (June 2011).



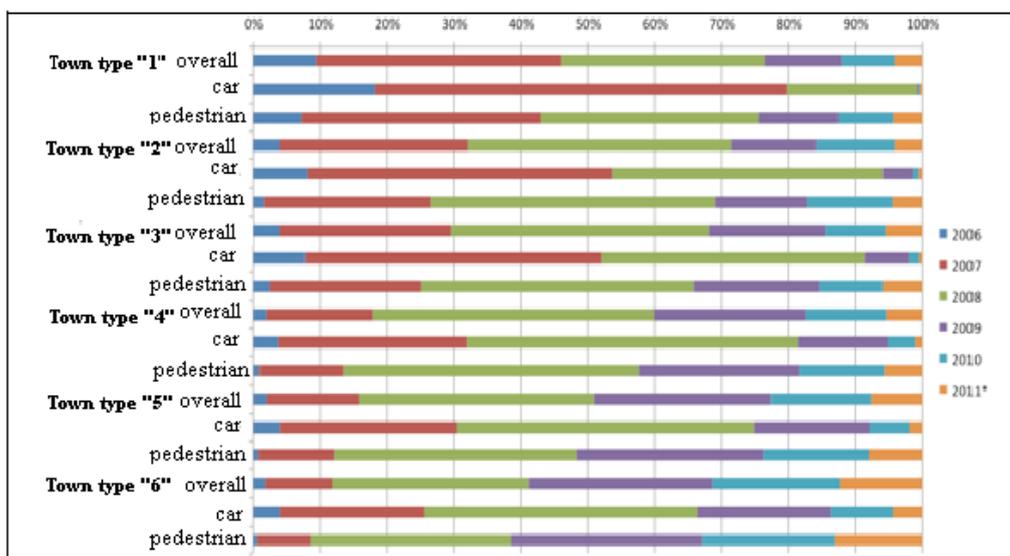
3.2. Data Completeness and Population Density

For further, more detailed studies of the route length of the total street network, the dataset was divided into the smallest possible German administrative units: municipalities and town boundaries. Detailed presumptions about the data development and the relative completeness with regard to population and area can be provided as a consequence of calculating the length of the route network for the different modes of transportation within the specified boundaries.

The administrative areas used in our analysis (12,387 in total) feature the number of inhabitants for the years 2008 and 2009 and were obtained from the TomTom Multinet dataset. The entire administrative area dataset is subdivided into six groups considering different population numbers. The first group ($\geq 1,000,000$) represents metropolitan areas; the second group ($\geq 500,000$ and $< 1,000,000$) large towns; the third ($\geq 100,000$ and $< 500,000$) towns; the fourth ($\geq 50,000$ and $< 100,000$) medium-sized towns; the fifth ($\geq 10,000$ and $< 50,000$) small towns; and the last ($< 10,000$) rural towns. With regard to the entire administrative area of Germany, this means that approximately 73% of the entire population lives in population groups one to five, covering one third of the entire area of Germany. Conversely, around 27% of the population lives in population group six (rural towns) and is distributed over two-thirds of the total area of Germany.

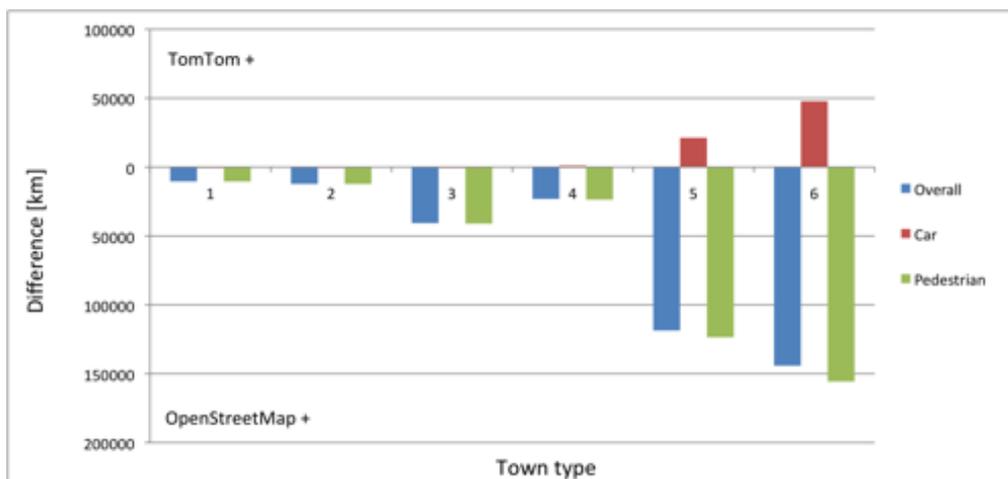
For our analysis we considered the development of three different street networks: total street network, and car and pedestrian networks (Figure 7a). The rows in Figure 7a visualize the expansion for each individual network and percentage of new data per year. It is evident that over the past four years, the route network of the individual groups has developed in correlation to their population density. While the general route network has been less active in the more densely populated areas, an increase in new data can still be seen in the more sparsely populated areas. It is also clearly discernible that the largest overall increase in new streets occurred in 2008.

Figure 7. (a) Development of OSM Street Network by Town Type; (b) Relative Difference by Town Type and Street Network (June 2011).



(a)

Figure 7. Cont.



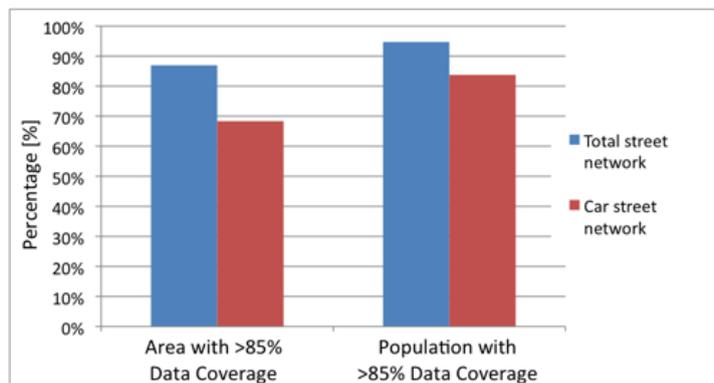
(b)

Another aspect that has been included in our analysis was the difference in total length of the route networks by town or municipality (Figure 7b). The results showed that the aforementioned approximately 9% of missing data is mainly distributed over the sparsely populated areas. It is also clearly discernible that OSM provides more overall data in comparison to the proprietary dataset with regard to the total and pedestrian route network length.

When expressed in route network lengths, this means that in mid-2011, OSM was still lacking approximately 3% (21,000 km) in the small-town population group and approximately 6% (48,000 km) in the rural-town population group. Using these highly detailed studies for the increase in street data for the different town types and the analyses of the differences in route network lengths in comparison with a commercial dataset, projections could be made of the time frame within which the dataset could be completed (at least in a relative comparison to another dataset, since neither of the datasets represent ground truth). As Figure 7b indicates, there is currently still a lack of data for less densely populated areas in OSM. Figure 7a shows the development of data by population group. In line with the expansion rate of this graph, 6% (14,000 km) of new streets were added to group 5 for car navigation in 2010 and nearly 10% (33,000 km) to group 6. By mid-2011, 2% (5,000 km) of new streets were added to group 5 and slightly less than 4% (16,000 km) to group 6. This means that if there is an increase in new street data that remains at least at the same level and does not decline as shown in Figure 3, the German street network for population groups 5 and 6 will be almost completely covered by the middle to end of 2012.

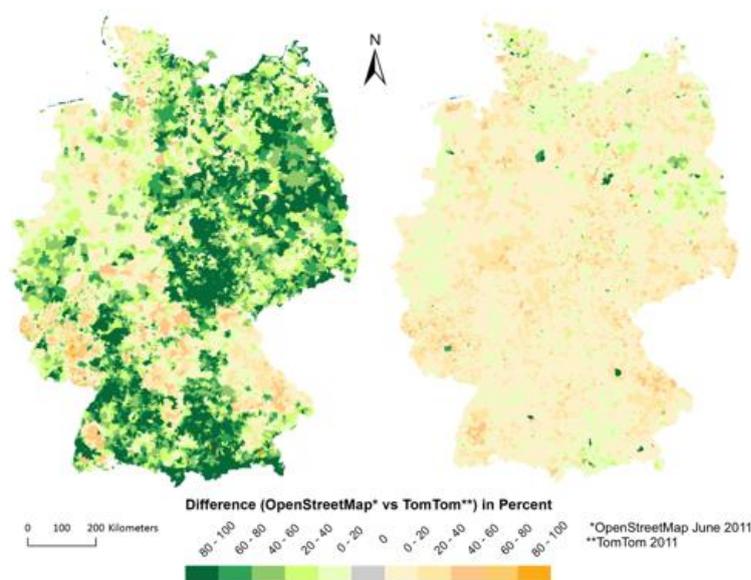
With regard to the correlation between TomTom’s commercial dataset and OSM, and the relative route network comparisons by town or municipality area, the following statement can be made (cf. Figure 8). Overall there exists an 85% correlation in total length between the OSM and the TomTom dataset for a total of 87% of the area of Germany. For data related to car navigation, this value decreases to approximately 69%. Considering the population density, this means that nearly 95% of the inhabitants of Germany are covered by 85% data coverage. In the case of car navigation data, this value decreases again to nearly 84% of the population.

Figure 8. Correlation between OSM Data Coverage and Area, and OSM Data Coverage and Population (June 2011).



Although OSM’s total route length already exceeds that of TomTom, there are still areas in Germany within which TomTom has more data present than does OSM. According to the previous results gathered with regard to population density, these are typically areas in which the population tends to be low. The following two maps show where the differences in the total route network (Figure 9, left) and the route network for car navigation (Figure 9, right) can be found, based on the administrative areas for municipalities and towns.

Figure 9. Relative Difference between TomTom and OSM for Total Route Network (left) and for Car Navigation Network (right) (June 2011).

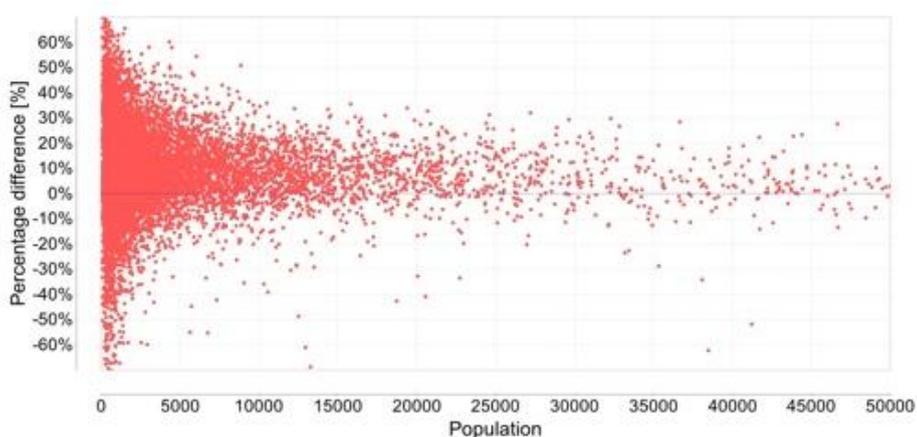


The results gathered from several analyses over time showed that data collections in municipalities in the southeast of Germany show a good total route network; however, the same areas still lack data specific to car navigation. Upon closer examination, routes within these areas showed that although they were geometrically present in the dataset, attributes associated with these routes would not give a definitive street category. This error occurs often when streets are digitized by a contributor from aerial images, but due to the lack of local knowledge about the area, no statement can be made on the category of the street. The second information that could be derived from the maps was that TomTom

has less data available in the total route network for the eastern part of Germany, while OSM generally shows a higher total route network length in this area. Overall, with the exception of a few areas, this statement can be made for large parts of all of Germany. However, with regard to the route network for car navigation, this situation is, as mentioned before, somewhat worse.

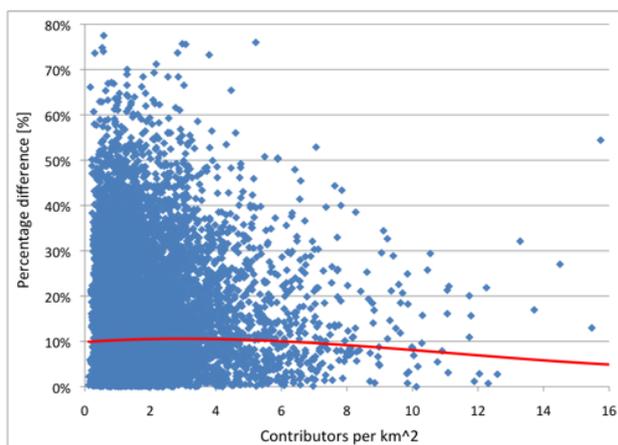
A cloud diagram allows us to visualize the towns and municipalities according to their population and relative differences between the TomTom and OSM datasets, in particular, the network for car navigation (Figure 10). The graph clearly shows a decrease in discrepancies between the datasets with growing population density. These discrepancies can be positive and negative for each dataset. Additionally we can see that data differences in the class of rural towns (10,000–50,000 inhabitants) can vary between 10% and 20%.

Figure 10. Correlation between Dataset Differences and Population Density (June 2011).



Different numbers of members have been gathering data for OSM in each administrative area that we analyzed. A simplified number of participants per square kilometer can be calculated by dividing the total number of participants per administrative area by the size of the area. Our results showed that with an increasing number of participants, the relative difference between the datasets decreased (Figure 11). However, what is more important, a statement can be made on how many participants are required to gather all data to receive a sophisticated dataset.

Figure 11. Correlation between Data Completeness and Number of Contributors (June 2011).

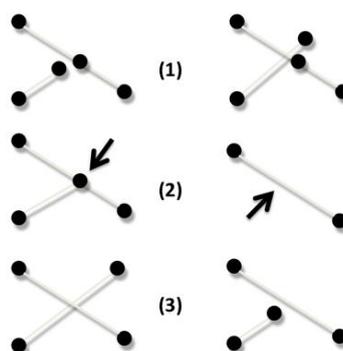


Bearing in mind with the current data collection trend in Germany, completeness for car navigation data of more than 90% could already be achieved in relative comparison to the commercial dataset with an average of two project participants per square kilometer. According to the trend line, more than six participants are required to achieve a dataset that is close to “complete”.

3.3. Topology Errors and Turn Restrictions

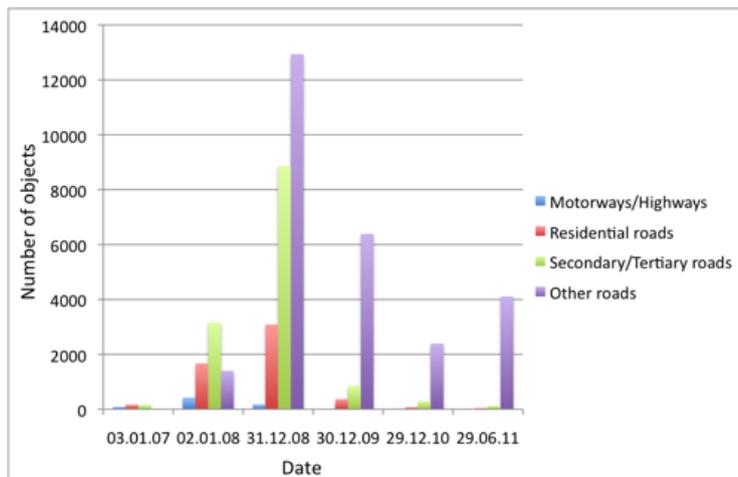
A graph is generally required for a routing application that represents a street network and also comprises nodes and edges. Due to this fact, it is essential that the graph is topologically correct and that it does not contain any errors. OSM data is not routable in its standard form [41,42]; however, within the OSM project, attempts are being made to record the street data correctly topologically, but this topology cannot be used directly for routing without additional data preparation. During this preparation, procedure junctions must be localized by searching for nodes that are used by several streets, and streets must be attributed to these nodes accordingly. However, errors do occur in the OSM dataset. We have examined the entire route network for Germany to find possible topology errors. In doing so, we identified errors in the topology similar to those visualized in Figure 12. The first possible error that can occur is that the junction cannot be determined as such, as the ways do not share a common node (1). Second, duplicate nodes or ways can cause an error (2), and third, the streets do not cross or lack information and they simply overlap (3).

Figure 12. OSM Topology Error Types.

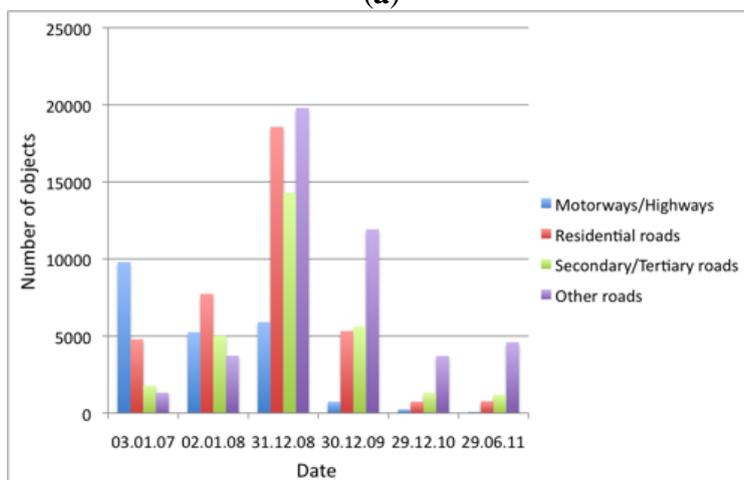


We converted the annual datasets (2007–2011) of OSM into routable street networks and searched for possible topology errors. The topology errors for non-linked streets were determined by measuring the distance between the two applicable streets, which should not be greater than 1 m. It can be clearly seen that the number of such errors has decreased over the years and remains high only for routes of cyclists or pedestrians (Figure 13a). The results of the second analysis for possible double streets also showed that the quality has continually improved here, at least in the street network for car navigation (Figure 13b). The number of errors for the third analysis, which shows the results of the error for intersecting streets without any shared nodes (Figure 13c), remains relatively constant, with the exception of the “other routes” data group. During random sampling, it happened that some of the errors that were identified were based on attribute errors in the dataset. For example, the information that the street is in fact a bridge was missing.

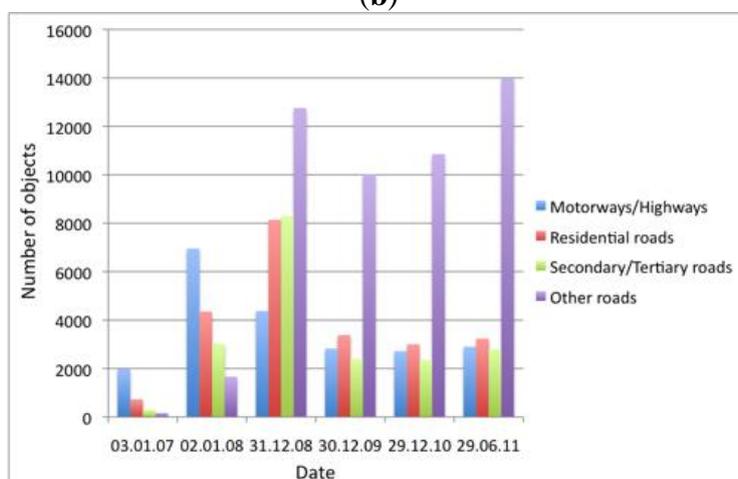
Figure 13. (a) OSM Topology Errors; (b) OSM Duplicate Nodes or Ways Errors; (c) Lack of Information Errors.



(a)



(b)



(c)

Turn restrictions constitute an essential component of routing applications. In a worst-case scenario, serious street accidents can occur should they be absent or incorrect. There are several different types of turn restrictions. In general, two types can be differentiated: requirements and prohibitions. Requirements prescribe the only possible way(s) to turn or travel at a junction. Prohibitions, on the

other hand, indicate where it is not permitted to travel. In the following preliminary comparison (Table 2), the total number of turn restrictions of TomTom and OSM for Germany are compared.

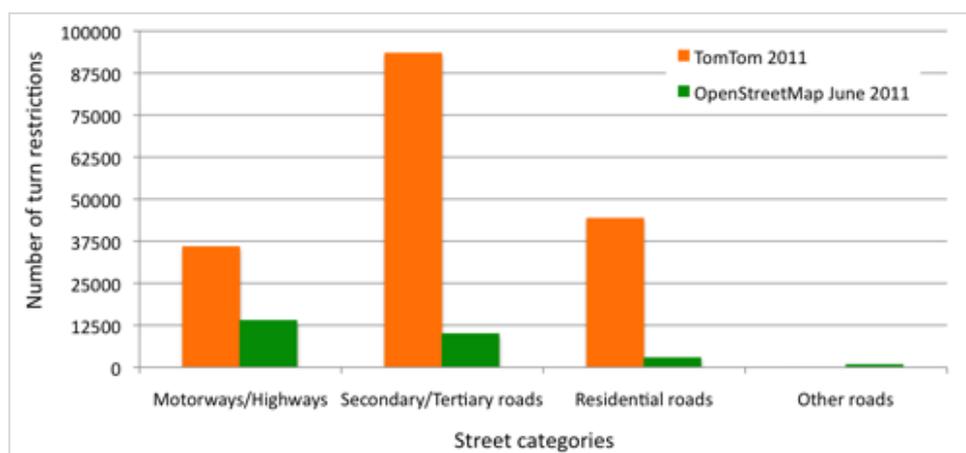
Table 2. Total Number of TomTom and OSM Turn Restrictions in Germany (June 2011).

Data Provider	Date	Total	Standardized
TomTom	2011	Approximately 176,000	Approximately 174,000
OpenStreetMap	June 2011	Approximately 21,000	Approximately 28,000

The difference between TomTom and OSM totals almost 146,000. As such, TomTom currently has five times more turn restrictions available for Germany than does OSM. Although the number of turn restrictions available in the OSM dataset is continually increasing, it will probably take several more years before OSM achieves the same level as TomTom, based on the current status and development.

The biggest issue during this analysis was to adjust TomTom’s dataset, read the turn restrictions, and convert them in such a way that the OSM data would be applicable for a comparison. In addition to the distribution of information for turn restrictions over several attribute tables and datasets in TomTom, the existing restrictions also had to be filtered. For example “automatically calculated” turn restrictions or those prohibiting turning into a “residents only” street were among the restrictions that have been filtered out of the TomTom dataset. In addition to the total number of differences described above, a comparison by street category was also conducted (Figure 14).

Figure 14. Number of Turn Restrictions by Street Category in Germany for TomTom and OSM (June 2011).



For a further analysis, we organized the standardized turn restrictions according to their appearance in the different population groups (Figure 15). The results showed that a large number of missing objects fall into the rural groups. However, the graph also shows that objects are missing in urban areas as well.

A further important quality parameter, and the final aspect of our analysis, is the temporal accuracy of the geodata. The OSM dataset allowed us to analyze this accuracy factor by identifying the street time stamp of each object in the dataset. According to the information retrieved from the dataset, which included the time stamp of each route network object, approximately one third of the data originated during 2011 and 2010, and another third during 2009 and 2008 (cf. Figure 16).

Figure 15. Number of Turn Restrictions by Town Type in Germany for TomTom and OSM (June 2011).

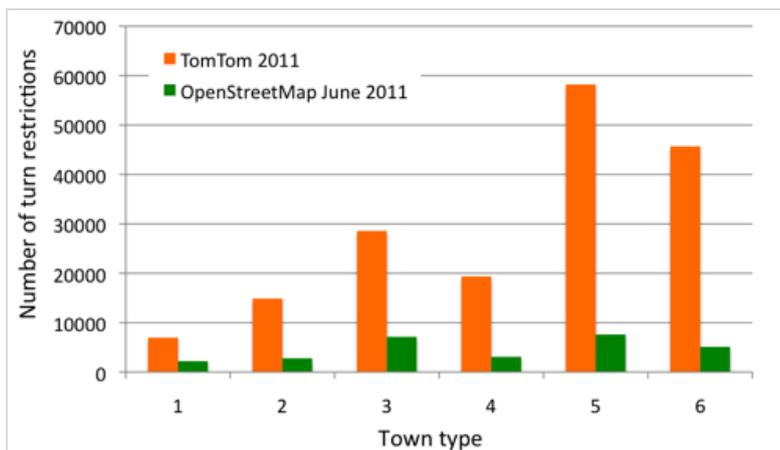
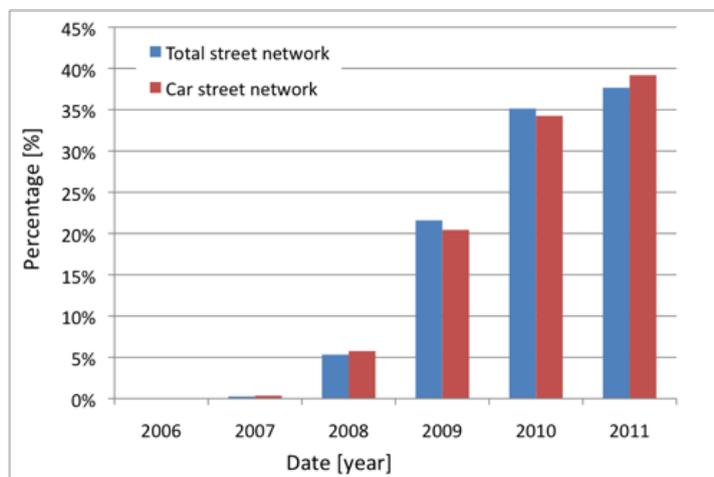


Figure 16. Actuality of the OSM Route Network.



4. Conclusions and Future Work

In this article, we outlined the development of Volunteered Geographic Information in Germany from 2007 to 2011, using the OpenStreetMap project as an example. Specifically, we considered the expansion of the total street network and the route network for car navigation. With a relative completeness comparison between the OSM database and TomTom’s commercial dataset, we proved that OSM provides 27% more data within Germany with regard to the total street network and route information for pedestrians. On the contrary, OSM is still missing about 9% of data related to car navigation. According to our projection for the future, this discrepancy should disappear by the middle or end of 2012, and the OSM dataset for Germany should then feature a comparative route network for cars as provided by TomTom.

In addition to the route network comparisons, we conducted further analyses regarding topology errors and the completeness of street name information. The results showed that the OSM dataset is not flawless; however, the trend shows that the relative and absolute number of errors is decreasing. Thus, it can also be discerned that not only is new data being added to the project database but also quality assurance is becoming a major factor within the OSM community. Our findings with regard to turn

restrictions within the OSM database, which are of critical importance to navigation, showed that based on the current development rate and activity, it will take more than five years for OSM to catch up with the information found in the proprietary dataset used in our analysis. This slower development in comparison to the regular street data collection can have several reasons. It can be based on the fact that turn restrictions cannot be seen in the regular OSM map and thus are less appealing for contributors to be added. Some members might also not be familiar with the importance of turn restrictions for the dataset or do not understand how to implement them correctly.

Overall, a certain trend can be distinguished from our studies, as well as in all other studies conducted to date for the countries that were examined. Preliminary statements and conclusions in the past were that OSM data is sufficient for use with map applications. Today we can say that, at least in countries in which the OSM project is well developed, the data is becoming comparable in quality to other geodata from commercial providers regarding the different factors analyzed in this paper such as temporal accuracy and geometric accuracy.

However, several questions remain and further research is still needed. One important factor that has not been addressed yet is the importance of whether users who contribute data to OSM should also maintain it. Also, it is unclear whether missing attribute information, such as street types or names, if added at a later date, could be analyzed and provided useful insights. So far it seems as if processing within the OSM project is closely related to visual factors, meaning that most data is collected in areas where there are white spots on the map, and thus no information is available. We will investigate specific questions regarding this user behavior in detail in the near future. It will be important to obtain further information on the project's participants and data contributors. These are some of the questions that need to be addressed: Are OSM mainly long-term contributors or are most of them so-called "submarine users"; that is, do they appear for a short period, add information, and then disappear again? Do members only add new data, or do they also edit existing information? Can an activity radius or area be determined for the participants of the project? Is the administrative area of an entire country completely covered by volunteers of the project or are data contributions by agencies playing a major role in certain areas?

It will continue to be important to carry out studies about the quality assurance of VGI. Preliminary suggestions have been made on how consistency of compiled VGI data could be achieved by improving quality during production and providing quality metadata for the users [43].

References

1. Anderson, P. *What Is Web 2.0? Ideas, Technologies and Implications for Education*; JISC: Bristol, UK, 2007.
2. Diaz, L.; Granell, C.; Gould, M.; Huerta, J. Managing user-generated information in geospatial cyberinfrastructures. *Future Gen. Comput. Syst.* **2011**, *27*, 304–314.
3. Coast, S. OpenStreetMap. Workshop on Volunteered Geographic Information. 2007. Available online: <http://www.ncgia.ucsb.edu/projects/vgi/> (accessed on 11 November 2011).
4. Nelson, A.; de Sherbinin, A.; Pozzi, F. Towards development of a high quality public domain global roads database. *Data Sci. J.* **2006**, *5*, 223–265.

5. Elwood, S. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal* **2008**, *72*, 173–183.
6. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
7. Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.
8. Chilton, S. Crowdsourcing Is Radically Changing the Geodata Landscape: Case Study of OpenStreetMap. In *Proceedings of the UK 24th International Cartography Conference*, Santiago, Chile, 15–21 November 2009.
9. Heipke, C. Crowdsourcing geospatial data. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 550–557.
10. Hudson-Smith, A.; Batty, M.; Crooks, A.; Milton, R. Mapping for the masses: Accessing web 2.0 through crowdsourcing. *Soc. Sci. Comput. Rev.* **2008**, *27*, 524–538.
11. Ramm, F.; Stark, H.-J. Crowdsourcing geodata. *Geomatik Schweiz, Géomatique Suisse* **2008**, *6*, 315–318.
12. O'Reilly, T. *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*; O'Reilly Media: Cambridge, MA, USA, 2005.
13. Goodchild, M.F. Spatial Accuracy 2.0. In *Proceeding of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, China, 25–27 June 2008.
14. Hudson-Smith, A.; Batty, M.; Milton, R.; Batty, M. NeoGeography and web 2.0: Concepts, tools and applications. *J. Locat. Based Serv.* **2009**, *3*, 118–145.
15. Rana, S.; Joliveau, T. NeoGeography: An extension of mainstream geography for everyone made by everyone? *J. Locat. Based Serv.* **2009**, *3*, 75–81.
16. Turner, J.A. *Introduction to Neogeography*; O'Reilly Media: Cambridge, MA, USA, 2006.
17. Walsh, J. The beginning and end of neogeography. *GEOconnexion Int. Mag.* **2008**, *7*, 28–30.
18. Fischer, F. Collaborative mapping—How wikinomics is manifest in the geo-information economy. *GEO Inf.* **2008**, *2*, 28–31.
19. Sui, D.Z. The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Comput. Environ. Urban Syst.* **2008**, *32*, 1–5.
20. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; Grillmayer, R.; Achard, F.; Kraxner, F.; Obersteiner, M. Geo-Wiki.Org: The use of crowdsourcing to improve global land cover. *Remote Sens.* **2009**, *1*, 345–354.
21. Mooney, P.; Corcoran, P. Using OSM for LBS—An Analysis of Changes to Attributes of Spatial Objects. In *Advances in Location-Based Services*; Gartner, G., Ortag, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 34, Issue 6, pp. 496–507.
22. Neis, P.; Zipf, A. *OpenRouteService.org is Three Times “Open”: Combining OpenSource, OpenLS and OpenStreetMap*; GIS Research UK: Manchester, UK, 2008.
23. Over, M.; Schilling, A.; Neubauer, S.; Zipf, A. Generating web-based 3D city models from OpenStreetMap: The current situation in Germany. *Comput. Environ. Urban Syst.* **2010**, *34*, 496–507.

24. Nielsen, J. Participation Inequality: Encouraging More Users to Contribute, Alertbox. 2006. Available online: http://www.useit.com/alertboxparticipation_inequality.html (accessed on 11 November 2011).
25. Ramm, F. Crowdsourcing Geodata. Society of Cartographers Summer School, Southampton. 2008. Available online: <http://www.geofabrik.de/media/2009-09-08-crowdsourcing-geodata.pdf> (accessed on 28 December 2011).
26. Ramm, F. Kraut sourcing 2.0 Beta - The State of Germany. State of the Map, Amsterdam. 2009. Available online: <http://www.geofabrik.de/media/2009-07-11-sotm-state-of-germany.pdf> (accessed on 28 December 2011).
27. Goodchild, M. NeoGeography and the nature of geographic expertise. *J. Locat. Based Serv.* **2009**, *3*, 82–96.
28. Flanagan, A.J.; Metzger, M.J. The credibility of volunteered geographic information. *GeoJournal* **2008**, *72*, 137–148.
29. Mau é P.; Schade, S. Quality of Geographic Information Patchworks. In *Proceedings of AGILE*, Girona, Spain, 5–8 May 2008.
30. Ather, A. A Quality Analysis of OpenStreetMap Data. M.E. Thesis, University College London, London, UK, May 2009.
31. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B* **2010**, *37*, 682–703.
32. Zielstra, D.; Zipf, A. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In *Proceedings of 13th AGILE International Conference on Geographic Information Science*, Guimarães, Portugal, 10–14 May 2010.
33. Ludwig, I.; Voss, A.; Krause-Traudes, M. Wie gut ist Open Street Map? Zur Methodik eines automatisierten objektbasierten Vergleiches der Strassennetze von OSM und NAVTEQ in Deutschland. *GIS Sci.* **2010**, *4*, 148–158.
34. Girres, J.F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459.
35. Zielstra, D.; Hochmair, H.H. Digital street data: Free versus proprietary. *GIM Int.* **2011**, *25*, 29–33.
36. Zielstra, D.; Hochmair, H.H. A comparative study of pedestrian accessibility to transit stations using free and proprietary network data. *Transp. Res. Rec.: J. Transp. Res. Board* **2011**, *2217*, 145–152.
37. Zielstra, D.; Hochmair, H.H. Comparison of Shortest Path Lengths for Pedestrian Routing in Street Networks Using Free and Proprietary Data. In *Proceedings of Transportation Research Board - 91st Annual Meeting*, Washington, DC, USA, 22–26 January 2012.
38. Haklay, M.; Ellul, C. Completeness in volunteered geographical information—the evolution of OpenStreetMap coverage in England (2008–2009). *J. Spat. Inf. Sci.* submitted for publication, 2011.
39. Brassel, K.; Bucher, F.; Stephan, E.; Vckovski, A. Completeness. In *Elements of Spatial Data Quality*; Guptill, S.C., Morrison, J.L., Eds.; Elsevier: Oxford, UK, 1995; pp. 81–108.
40. Oort, P.A.J. Spatial data quality: From Description to Application. Ph.D. Thesis, Wageningen University, Wageningen, Germany, January 2006.

41. Chen, H.; Walter, V. Quality Inspection and Quality Improvement of Large Spatial Datasets. In *Proceedings of the GSDI 11 World Conference: Spatial Data Infrastructure Convergence: Building SDI Bridges to Address Global Challenges*, Rotterdam, The Netherlands, 15–19 June 2009.
42. Schmitz S.; Neis P.; Zipf, A. New Applications Based on Collaborative Geodata—the Case of Routing. In *Proceedings of XXVIII INCA International Congress on Collaborative Mapping and Space Technology*, Gandhinagar, Gujarat, India, 4–6 November 2008.
43. Brando, C.; Bucher, B. Quality in User-Generated Spatial Content: A Matter of Specifications. In *Proceedings of the 13th AGILE International Conference on Geographic Information Science*, Guimarães, Portugal, 10–14 May 2010.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).