

Empirische Untersuchungen zur Datenqualität von OpenStreetMap – Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste

Pascal NEIS, Dennis ZIELSTRA, Alexander ZIPF, Alexander STRUNCK

Universität Heidelberg, Geographisches Institut, Lehrstuhl Geoinformatik
<http://www.geog.uni-heidelberg.de/giscience.html>

Einleitung

Mit OpenRouteService (ORS) liegen zur AGIT 2010 schon über 2 Jahre an Erfahrungen mit dem Betrieb von Online-Diensten auf Basis der Daten der freien Wiki-Weltkarte OpenStreetMap (OSM) vor (Neis & Zipf 2008). Dabei bietet ORS mittlerweile natürlich sowohl eine größere Funktionsfülle als auch räumliche Abdeckung. Von knapp 400.000 Strassen-segmenten für Deutschland in der ersten Version wuchs der Service auf nun weit über 20 Millionen Straßensegmente. Dies impliziert natürlich besondere technische Herausforderungen, die im Beitrag dargelegt werden. Auch OSM-3D.org hat Geburtstag und einher geht mit beiden die wachsende Erfahrung mit dem Preprocessing dieser Daten für unterschiedliche Anwendungsgebiete. Der Schwerpunkt der Untersuchungen bezieht sich auf die „fitness for use“ der verwendeten Grundlagendaten und deren räumliche, thematische und zeitliche Differenzierung. Dabei kann mittlerweile auf Erfahrungen mit mehreren unterschiedlichen Routingspezialanwendungen zurückgegriffen werden. Neben PWK (schnellste, kürzeste), Fußgänger und Radfahrer (unterschieden nach MTB, Rennrad, kürzeste und sicherste) sind dies u.a. Rollstuhlrouting, 3D-Routing, Notfall-Routing und Routing für landwirtschaftliche Maschinen (Mähdrescher, Erntezüge) sowie Evakuierungssimulationen. Neben dem eigentlichen Routenplaner und seinen Derivaten wurden am Lehrstuhl eine Vielzahl weiterer OGC- und weiterer GI-Dienste realisiert, die jeweils eigene Ansprüche an die Daten haben. Diese und wesentliche Erfahrungen mit dem Online-Betrieb dieser Dienste werden am Beispiel empirischer Untersuchungen der Daten diskutiert. Zu nennen sind u.a. folgende:

- Web Feature Service (OGC)
- Geocoder & Reverse-Geocoder (gemäß OGC OpenLS Standard)
- POI-Umkreissuche (gemäß OGC OpenLS Standard, Directory Service)
- Erreichbarkeitsdienst (Accessibility Analysis Service) (Neis et al. (2007))
- Web -3D Service (www.osm-3d.org)
- OSM-WMS.de
- HistOSM.org (Historische und archäologische Daten in OSM)
- Höhenprofilberechnung als OGC Web Processing Service (WPS) – in ORS
- Testdaten für diverse WPS-Prozesse auf www.geoprocessing.org

Vergleich Straßennetzwerk OpenStreetMap zu Teletlas

Für eine allgemeine Einführung in OpenStreetMap wird auf Ramm & Topf (2008) verwiesen. Im Folgenden werden ausgewählte Aspekte der Berechnungen zur Datenqualität von OSM dargestellt: Schon 2008 wurden erste Untersuchungen bzgl. der Datenqualität und Vollständigkeit in OSM durchgeführt. Haklay (2008) war einer der ersten, der die Daten in England mit amtlichen Datensätzen, nämlich OS Meridian 2 (Ordnance Survey 2009) verglich. Allerdings schnitten dabei die OSM Daten außerhalb von Städten aus 2008 nicht gut ab. Den Daten wurde aber von Haklay ein großes Potential für die Zukunft voraus gesagt. Nachfolgende Analysen im Oktober 2009 kamen zu sehr erstaunlichen Ergebnissen. In 2008 waren nur 27% von England im OSM Projekt erfasst und im Oktober 2009 waren es bereits 65%, was einen beeindruckenden Zuwachs innerhalb eines Jahres bedeutet (Haklay 2009). In der Arbeit von Ather (2009) wurde der OSMDatensatz mit dem Ordnance Survey's MasterMap-Datensatz verglichen. Dies bietet insofern Vorteile, da dieser Datensatz keiner Generalisierung unterzogen wurde. Dadurch ist die Positionsgenauigkeit höher als bei der vorangegangenen Untersuchung von Haklay mit dem Meridian 2-Datensatz. Die Untersuchung der Positionsgenauigkeit wurde für Autobahnen, Bundesstraßen und Landstraßen in vier 25 Quadratkilometer großen Bereichen in London berechnet. Für die jeweilige Straßenklasse wurden angepasste Puffer verwendet (8m für Autobahnen, 5,6m für Bundesstraßen und 3,75m für Landstraßen). Des Weiteren wurden die Straßen auf Vollständigkeit der Attribute und die an der Aufnahme der Straße beteiligten Nutzer analysiert.

Für Deutschland wiederholte erstmals Zielstra (2009) die verschiedenen Vergleichs- und Analysemethoden in angepasster und erweiterter Form und kommt zu raumzeitlich differenzierten quantitativen Ergebnissen. Strunck (2010) adaptiert die Methode zusätzlich für den Vergleich von Points of Interests (POIs). Dabei wurden nicht staatliche, sondern Daten des kommerziellen Anbieters Teletlas verwendet. Hier können aus Platzgründen nur kleine Ausschnitte der Ergebnisse angedeutet werden. Ähnlich untersuchte Ludwig (2010) in einer weiteren Diplomarbeit wie OSM-Daten auf Navteq-Daten abgebildet werden können. Aktuelle Untersuchungen von Stroh (2011 in Arbeit) beschäftigen sich verstärkt mit der Rolle des „Mappers“. Bei den Ergebnissen für Deutschland zeigte sich, wie bereits bei Haklay, dass die OSM Datenqualität in weniger dicht besiedelten Gebieten abnimmt. Starke Differenzen erschienen sowohl bei der relativen Positionsgenauigkeit, wie auch bei der relativen Vollständigkeit zwischen den Daten bei Groß- und Mittelstädten (Zielstra 2009). Bei den Untersuchungen wurde deutlich, dass zumindest in Deutschland die Datenmenge der frei verfügbaren Daten in Großstädten bereits so umfangreich ist, dass sie eine Alternative zu kommerziellen Daten darstellen können, da sie z.B. die Gesamtlänge des Straßennetzwerks der kommerziellen Daten in Ballungsgebieten schon übertreffen. Allerdings gilt das für die unterschiedlichen Straßennetze für die unterschiedlichen Verkehrstypen in unterschiedlichem Maße. Insgesamt hat Teletlas speziell für das PKW-Routing noch fast flächendeckend ein umfangreicheres Datennetz, unter Einbeziehung von Fußgänger- und Radrouting-Daten wird das Ergebnis aber vielschichtiger. Die Geschwindigkeit der Annäherung zwischen OSM und beispielsweise Teletlas zeigt Abbildung 1. Bei Beginn der ersten Untersuchungen im April 2009 bot OSM noch ein um fast 1/3 kürzeres Gesamtstraßennetz auf.

Dieser Unterschied hatte sich bis Jahresende auf ca. 6 Prozent verringert und dürfte zum heutigen Zeitpunkt zumindest ausgeglichen sein, wie unsere Extrapolationsberechnungen zeigen.

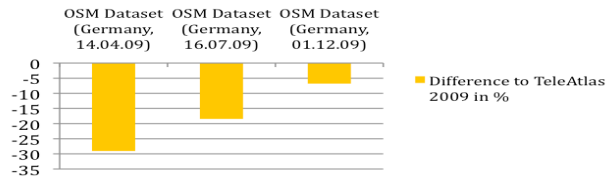


Abb.1: Zeitliche Entwicklung der Differenz der Gesamtlänge des gesamten Straßennetzes von OSM und Teleatlas in Prozent.

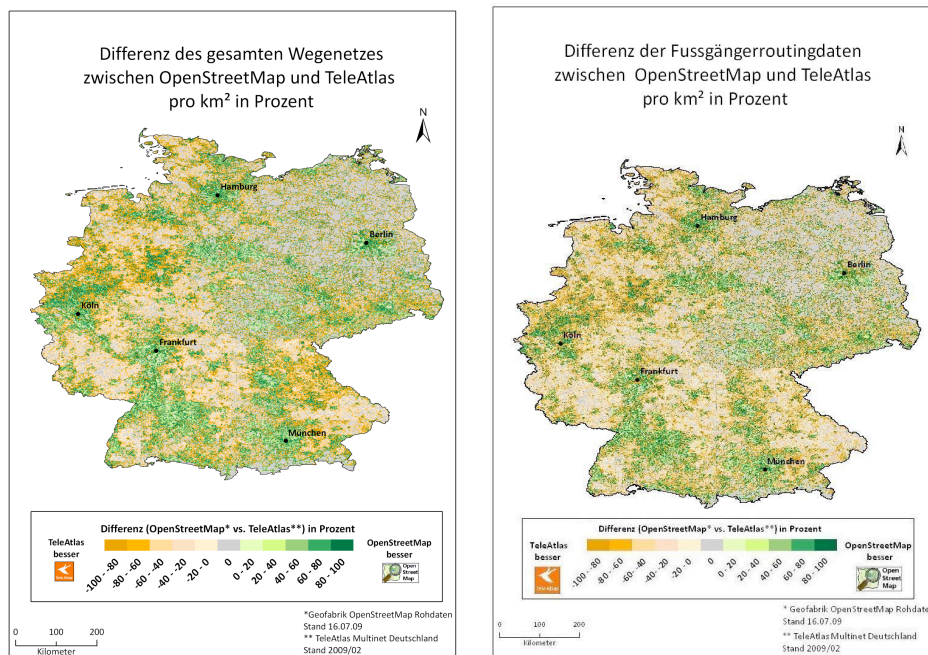


Abb. 2: Differenzkarten OSM versus Teleatlas (Gesamtnetzwerk links, Fußgängerrouting rechts), OSM Stand 16.07.2009. Teleatlas Stand 2009/Q2.

Um regionale flächenhafte Vergleiche zu ermöglichen, wurde in Anlehnung an Hacklay die Differenz der Länge des Netzwerkes von OS und Teleatlas für unterschiedliche Verkehrsarten zu unterschiedlichen Zeitpunkten berechnet (Auswahl siehe Abb. 2). Wie die Karten zeigen, ist eine klare Abnahme der OSM Datenmenge von den Ballungsgebieten und Stadtzentren hin zum Umland zu erkennen. Um diese Problematik genauer zu untersuchen, wurden vom Mittelpunkt ausgewählter Städte unterschiedlicher Größenklassen aus kreisrunde

Puffer mit unterschiedlichen Distanzen erstellt (Großstädte: 0-10 km, 10-25 km, 25-50 km; Mittelstädte: 0-5 km, 5-15 km, 15-25 km). Innerhalb der Puffer wurde die Differenz zwischen OSM und TeleAtlas berechnet, um die Differenz zum Umland genauer zu beziffern.

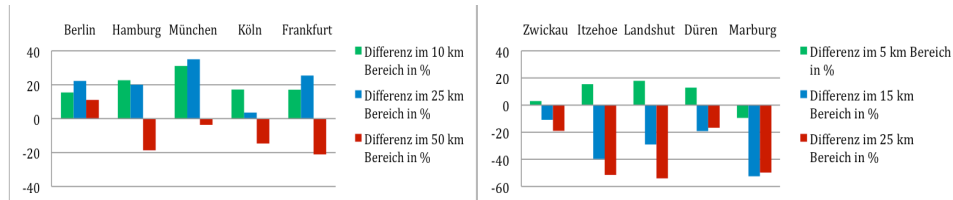


Abb. 3: Differenz des gesamten Wegenetzes von OSM und Teleatlas im Stadtgebiet und im Umland fünf ausgewählter Großstädte (links) und im Stadtgebiet und Umland fünf ausgewählter Mittelstädte (rechts). (OpenStreetMap Stand: 16.07.09, TeleAtlas Stand: 02/2009).

Diese Berechnungen wurden zunächst für April 2009 durchgeführt und für Juni 2009 wiederholt. Aktuellere Berechnungen werden derzeit ausgewertet. Bei den Differenzberechnungen der unterschiedlichen Datengrundlagen hat sich gezeigt, dass die rasante Entwicklung des OpenStreetMap-Projektes sich durchaus in diesen Zahlen widerspiegelt. Ein Anstieg der Daten um ca. 10% in nur drei Monaten (April-Juni 2009) stellt eine enorme Datenmenge dar, wobei dies bei den gewählten Kartendarstellungen für Gesamtdeutschland kaum auffällt und daher nur die aktuellere Version abgebildet ist. Die Berechnungen zu den Großstädten ergaben in allen untersuchten Städten, dass OpenStreetMap mehr Daten vorweisen konnte als TeleAtlas. Das heißt, dass besonders hier viele kleine Wege für die Fußgänger, aber auch Seitenstraßen für Autos erfasst werden. Die Ergebnisse der Berechnungen in den ausgewählten Mittelstädten ergaben beim Vergleich des gesamten Wegenetzes durchaus positive Werte (mit Ausnahme von Marburg). Da es sich bei den Untersuchungen von Groß- und Mittelstädten jedoch lediglich um unterschiedliche Stadtklassen handelt, wurde außerdem die Abnahme der Daten von den Innenstädten hin zum Umland untersucht. Die Ergebnisse der Kreispuffermethode zeigen deutlich, dass in den Großstädten im 50-Kilometer-Bereich eine Abnahme der OpenStreetMap Datenmenge zu vermerken ist. Fast alle getesteten Städte zeigen diese Entwicklung und weisen Differenzen von bis zu 23% auf. Auch bei den Mittelstädten wurden trotz der kleineren Puffergrößen sehr große Differenzen von bis zu 57% berechnet. Interessant ist die Prognose der weiteren Entwicklung der Menge der OSM-Daten durch Interpolation in die Zukunft. Diese wurde sowohl absolut für Gesamtdeutschland, als auch für jede einzelnen 1 Quadratkilometer Zelle durchgeführt. Die Ergebnisse werden derzeit noch verarbeitet und analysiert. Im Ergebnis kann gefolgert werden, dass zur AGIT 2010 OSM die Gesamtkilometer von Teleatlas überholt haben wird – ob dies jedoch auch differenziert für die einzelnen Verkehrstypen – also insbesondere das PKW-Straßennetz, bei dem Teleatlas den größten Vorsprung hatte - bleibt genauer zu analysieren. Insgesamt muss festgehalten werden, dass das Muster Stadt-Land (bzw. die Bevölkerungsdichte) noch länger bestehen bleibt, da auch die Rate der Datenzunahme in weniger dicht besiedelten Gebieten geringer ist und der die damit einhergehende geringere relative Datendichte noch längere Zeit zu verzeichnen ist.

Vergleich der Points of Interest zwischen OSM und Teleatlas Multinet

Nachdem in den ersten Arbeiten vorrangig das Straßennetzwerk untersucht wurde, beschäftigt sich die Arbeit von Strunck u.a. mit der Situation bei den Points of Interests (POIs). Hier bestand zunächst auch das Problem, dass die unterschiedlichen Typen von POIs vom OSM und Teleatlas daraufhin untersucht werden mussten, ob sie überhaupt die gleiche oder zumindest eine ähnliche Kategorie entsprachen. Dann wurden ähnliche Untersuchungen zur relativen Positionsgenauigkeit und v.a. Vollständigkeit durchgeführt wie bei den Straßennetzen, d.h. für 1 Quadratkilometer große Rasterzellen die Anzahl sich entsprechender POIs von OSM und TA verglichen. In Abbildung 4 ist zunächst die Gesamtzahl der POIs pro Quadratkilometer graphisch dargestellt.

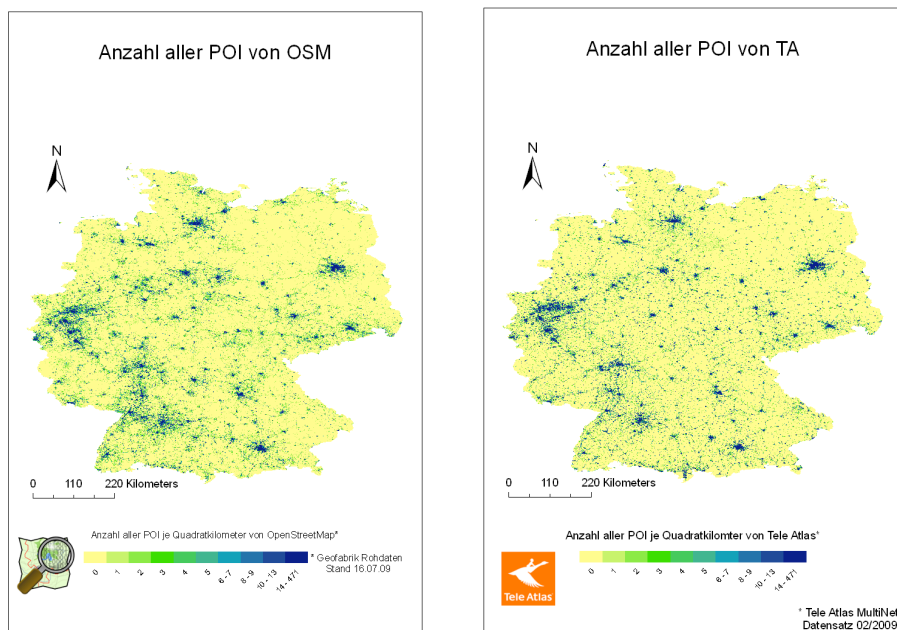


Abb. 4: Anzahl der POIs von OSM und Teleatlas im räumlichen Vergleich

Insgesamt konnten 248.320 POI bei TA (Stand 02/2009) mit 84.545 POI bei OSM (Stand 16.07.2009) mit jeweils übereinstimmenden Tags (Kategorien), die zudem über eine Namensangabe in den Tags verfügen, analysiert werden. TA besitzt einige POI-Typen, die um Größenordnungen höhere Fallzahlen aufweisen als OSM, wo also offensichtlich dedizierte externe Datenquellen in TA aufgenommen wurden. Z.B. Ärzte (OSM: 962, TA: 26.728) und Zahnärzten (OSM: 387, TA: 28.067). Auch ist eine Erhöhung der Anzahl der POI mit relevante Tags für die Autonavigation (z. B. Tankstellen OSM: 7.284, TA: 14.524) möglich. Aber auch hier sind die Steigerungsraten der Menge der POIs in OSM beeindruckend. Wird die Anzahl aller POI der beiden Datensätze in der Quadratkilometerrasterung verglichen ergibt sich, dass die Anzahl der POI auf 39.960 km² von TA höher ist und auf 55.239

km² bei OSM höher ist. Werden lediglich die zwischen den beiden Datenquellen miteinander vergleichbaren POI betrachtet, ergibt sich dass auf 40.606 QKM mehr POI von TA sind und auf 19.311 Rasterzellen mehr von OSM.

Zusammenfassung und Ausblick

Die umfangreichen Untersuchungen des Vergleichs von OSM mit Teleatlas bzw. Navteq-Daten konnten hier nur in wenigen Ausschnitten vorgestellt werden. Neben den genannten Diplomarbeiten wird in einer vierten Arbeit die Untersuchung weitergeführt mit einem neuen Fokus auf die Dichte der Mapper. Neben den Diplomarbeiten wurden die Untersuchungen auch innerhalb der Arbeitsgruppe ausgeweitet, so dass Prognosekarten erstellt werden, die untersuchen wie sich bei gegebenen oder angenommenen Wachstumsraten die Datenmenge von OSM zukünftig im Verhältnis zu den kommerziellen Anbietern weiterentwickeln wird. Spezielle Untersuchungen für dedizierte Anwendungen liefert Amelunxen (2010), der die Eignung von OSM für Geocoding evaluiert und durch die Implementierung von Interpolationsmechanismen verbessert. In Nachfolgearbeiten untersucht er weitere Aspekte der automatischen Verbesserung von OSM-Daten. Weitere Ergebnisse liefern Zielstra und Zipf (2010).

Literatur:

- Amelunxen, C. (2010): An Approach to geocoding based on volunteered Spatial Data.. Zipf et al. (Hrsg.) Geoinformatik 2010. Die Welt im Netz. Kiel.
- 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.
- Haklay, M. (2009): OpenStreetMap and Ordnance Survey Meridian 2 – Progress maps <http://povesham.wordpress.com/2009/11/14/openstreetmap-and-ordnance-survey-meridian-2-progress-maps/>
- Ludwig, I. (2009): Abbildung von Straßendaten für Qualitätsuntersuchungen - Ein Vergleich von OpenStreetMap mit Navteq. Diplomarbeit. Lehrstuhl Prof. Zipf.
- Ramm F., Topf J. (2008): OpenStreetMap, Die freie Weltkarte nutzen und mitgestalten. Lehmanns Media.
- Schmitz, S.; Neis, P.; Zipf, A. (2008): New Applications based on Collaborative Geodata - the Case of Routing. XXVIII INCA Int. Cong. on Collaborative Mapping, India.
- Strunck, A. (2010): Raumzeitliche Qualitätsuntersuchung von OpenStreetMap. Diplomarbeit. Lehrstuhl Prof. Zipf. Nicht publiziert.
- Zielstra, D. (2009): Datenqualität und Anwendbarkeit von Volunteered Geographic Information. Vergleich von proprietären und frei verfügbaren Geodaten. Diplomarbeit.
- Zielstra, D. & Zipf, A. (2009): Datenqualität von OpenStreetMap - Erste Ergebnisse empirischer Untersuchungen. POSTER: AGIT 2009. Salzburg. Austria.
- Zielstra, D. and Zipf, A. (2010): A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. AGILE 2010. The 13th AGILE International Conference on Geographic Information Science. Guimarães, Portugal.