



Flächennutzungsmonitoring IV Genauere Daten – informierte Akteure – praktisches Handeln

IÖR Schriften Band 60 · 2012

ISBN: 978-3-944101-03-3

OpenStreetMap – Datenqualität und Nutzungspotenzial für Gebäudebestandsanalysen

Marcus Götz

Götz, M. (2012): OpenStreetMap – Datenqualität und Nutzungspotenzial für Gebäudebestandsanalysen. In: Meinel, G.; Schumacher, U.; Behnisch, M. (Hrsg.): Flächennutzungsmonitoring IV. Genauere Daten – informierte Akteure – praktisches Handeln. Berlin: Rhombos, IÖR Schriften 60, S. 143-150.

OpenStreetMap – Datenqualität und Nutzungspotenzial für Gebäudebestandsanalysen

Marcus Götz

Zusammenfassung

Das Phänomen des sog. geo-crowdsourcing oder Volunteered Geographic Information (VGI) hat in den letzten Jahren zunehmend an Bedeutung gewonnen. VGI bedeutet, dass eine Gruppe bzw. eine Gemeinschaft freiwillig und gemeinsam unterschiedliche Arten von geo-referenzierten Daten sammelt und in einer Web Community frei zur Verfügung stellt. So kann – ähnlich wie bereits in Wikipedia – von dem Wissen einer großen Gemeinschaft profitiert werden. Eines der bekanntesten Projekte für VGI ist OpenStreetMap (OSM). Es wurde 2004 gegründet und ist seitdem ständig gewachsen – sowohl in Bezug auf die Mitgliederanzahl als auch auf die Datenmenge. Interessant zu sehen ist, dass die Community nicht mehr nur Straßenzüge und Landschaftsflächen digitalisiert, sondern auch ein zunehmendes Interesse an Gebäudegrundrissen und Gebäudeinformationen entwickelt. Besonders bemerkenswert ist hierbei, dass die Anzahl an Gebäudegrundrissen, mit weltweit derzeit knapp 58,96 Millionen, die etwa 52,72 Millionen Straßen bei weitem übersteigt. Und der Abstand wird größer, denn jede Woche mapen die OSM-Mitglieder etwa 450 000 neue Gebäudegrundrisse. Neben diesen zweidimensionalen Geometrien sammeln aber auch immer mehr OSM-Mitglieder zusätzliche Gebäudeinformationen, wie etwa die Höhe des Gebäudes, die Dachform, die Dachhöhe, die Dachfarbe oder die Fassadenfarbe. Ziel dieses Beitrags ist es zu analysieren, wie in Deutschland die derzeitige Datensituation in OpenStreetMap aussieht. Insbesondere soll untersucht werden, wie vollständig und flächendeckend Gebäude derzeit erfasst sind und welche zusätzlichen Informationen am häufigsten an die jeweiligen Geometrien angehängt werden. Verschiedene Untersuchungen zu den Eigenschaften der digitalisierten Grundrisse, wie etwa die erfasste Fläche oder die Anzahl der digitalisierten Punkte, beleuchten den Datenbestand in OSM weiter. Darüber hinaus soll gezeigt werden, welche Nutzungspotenziale die Daten in OpenStreetMap für verschiedene Anwendungsfälle bieten.

1 Einführung

Der Bedarf für unterschiedliche Geodaten – sowohl in der öffentlichen Verwaltung als auch in Unternehmen – wird typischerweise durch amtliche Geodatenätze, wie etwa ATKIS oder ALK, gestillt. Nichtsdestotrotz hat in den letzten Jahren das Phänomen des VGI ein zunehmendes Wachstum und Interesse aus den unterschiedlichsten Bereichen und Geschäftsfeldern erfahren. VGI beschreibt den Trend, dass Benutzer (insbesondere

Laien) nicht mehr nur Konsument von Geodaten sind, sondern vielmehr auch Produzent bzw. Lieferant. Dabei folgt VGI dem populären und bereits erprobten Prinzip des User-Generated Content (UGC) des Web 2.0, welches zum Beispiel bei Wikipedia zum Erfolg geführt hat. VGI bedeutet, dass Mitglieder einer Web Community unterschiedliche Arten (hinsichtlich Granularität, Typ usw.) von Geodaten erfassen und aufbereiten, um diese anschließend in einer Internetplattform mit anderen Mitgliedern zu teilen bzw. die Informationen anderen Interessierten (z. B. Anwendungsentwicklern) kostenlos zur Verfügung zu stellen. Goodchild (2007) spricht VGI ein enormes Potenzial zu, resultierend aus „billions of humans acting as remote sensors“. Die zunehmende Verfügbarkeit und ansteigende Qualität von VGI hat letztendlich dazu geführt, dass VGI auch zunehmend häufiger in professionellen Anwendungsgebieten des Urban Data Managements Anwendung findet.

Eines der bekanntesten und vielfältigsten Quellen von VGI ist das 2004 ins Leben gerufene OSM-Projekt. Das ursprüngliche Ziel war die Erschaffung einer freien Weltkarte. Heute bezeichnen viele OSMer das Projekt aber vielmehr als eine freie weltweite Geodatenbank. Um möglichst viele aktive Mitglieder zu bekommen (und um Frust bei der Datenerfassung vorzubeugen), ist das Datenschema in OSM so einfach wie möglich gehalten. Prinzipiell sammeln die OSMer georeferenzierte Punkte (sog. Nodes) mit Informationen über deren Längen- und Breitengrad. Diese Punkte stammen oftmals von selbstständigen Erhebungen im Gelände mittels eines GPS-Empfängers, immer öfters aber auch durch das Abzeichnen von Kartenelementen aus Luftbildern. Letzteres wurde zunehmend populärer seit Microsoft Bing Maps ihre Luftbilder OSM zur Verfügung gestellt hat. Insbesondere ermöglicht das luftbildbasierte Mappen die Erfassung von Daten in Gebieten, in denen sich der OSMer nicht aufhält. So kann zum Beispiel ein Mitglied aus Deutschland auch Daten in Frankreich erfassen. Mehrere Nodes können weiterhin zu einer geordneten Menge von Nodes zusammengefasst werden. Diese sogenannten Ways repräsentieren entweder eine Linie oder eine Fläche (Polygon). Bei Ersterem ist der Way nicht geschlossen, das heißt der erste Node ist ungleich dem letzten Node, wohingegen bei Letzterem der Way geschlossen ist, also der erste Node gleich dem letzten Node ist. Mehrere Objekte können in Gruppen (sog. Relations) zusammengefasst werden, wobei jedes einzelne Objekt eine bestimmte Rolle (role) innerhalb der Gruppe hat. Dies ermöglicht zum Beispiel die Beschreibung von Flächen mit Löchern oder Abbiegevorschriften für Straßen. Die aktuellen OSM (Stand Juni 2012, eigene OSM-Datenbank) enthalten etwa 1,47 Milliarden Nodes, 141 Millionen Ways und 1,48 Millionen Relations.

Doch OSM beinhaltet nicht nur die pure Geometrie eines Features, sondern vielmehr auch noch zusätzliche (semantische) Informationen, welche das jeweilige Feature typisieren oder näher beschreiben. Dafür bietet OSM die Möglichkeit, sogenannte Key-Value-Paare zu vergeben (Taggen). Der Key beschreibt eine bestimmte Information oder Informationsklasse und der Value verfeinert diese. Wird zum Beispiel ein Way mit dem Key highway getagged, so beschreibt dies, dass die Linie eine Straße repräsentiert. Durch

den Value cycleway, wird diese Information noch weiter verfeinert, denn es handelt sich um eine Straße für Fahrräder. Weitere Beispiele sind der Key building mit dem Value university, welche einen (geschlossenen) Way als Universitätsgebäude identifiziert, oder aber etwa der Key addr:street für die Definition der Adresse (den Straßennamen) eines Gebäudes. Sowohl die Anzahl als auch die Key-Value-Paare selbst sind unbeschränkt, d. h. ein OSMer kann prinzipiell jede beliebige Menge an Informationen zu einem Feature zur Verfügung stellen. Die geläufigsten und meistgenutzten Key-Value-Paare sind in der Map Feature List im OSM Wiki mehrsprachig dokumentiert (OSM 2012).

Sowohl die Quantität als auch die Qualität von Daten in OSM weisen eine sehr heterogene Verteilung auf, wobei Deutschland als eines der aktivsten Länder gilt. Die jüngste Untersuchung von Neis et al. (2012) hat gezeigt, dass in Deutschland mittlerweile 91 % aller Straßen (im Vergleich zu kommerziellen Daten von TomTom) erfasst sind. Darüber hinaus beinhaltet OSM aber noch 27 % mehr Wege (insbesondere Feldwege, Pfade etc.), welche in TomTom gar nicht erfasst sind. Die Attributierung sowie Informationen über Abbiegevorschriften fehlen allerdings noch deutlich. Ähnliche Vergleiche sowohl in quantitativer als auch in qualitativer Form für andere OSM-Features gibt es bisher allerdings noch nicht. In den folgenden Abschnitten werde daher verschiedene Analysen hinsichtlich der Quantität sowie der geometrischen und semantischen Eigenschaften von Gebäuden unternommen.

2 Quantitative Bewertung des Gebäudebestands in OSM

Betrachtet man den weltweiten OSM-Datenbestand und dessen Entwicklung in den letzten Wochen und Monaten wird deutlich, dass immer mehr Gebäude erfasst werden, d. h., dass immer mehr Polygone mit dem Key building markiert werden. Bemerkenswert ist, dass es in OSM mittlerweile mehr Gebäude als Straßen gibt. Mitte Juni gab es etwa 58,96 Millionen Gebäude und 52,72 Millionen Straßen. Jede Woche werden etwa 450 000 neue Gebäudegrundrisse erfasst und der Abstand zwischen Gebäuden und Straßen wird zunehmend größer. Abbildung 1 (a) zeigt die Entwicklung der Anzahl von Gebäudegrundrissen in OSM von 2007 bis Juni 2012.

Da sich diese Datenmenge auf den gesamten Globus verteilt, stellt sich nun die Frage, wie die Situation denn in Deutschland aussieht und ob (oder wann) mit ähnlichen Vollständigkeitszahlen, wie bei Straßen, zu rechnen ist. Mitte Juni gab es in Deutschland etwa 6,5 Millionen Gebäudegrundrisse, das bedeutet, dass etwa 11 % aller OSM-Gebäude in Deutschland gemappt wurden. Zusammengefasst beinhalten diese eine Fläche von 2 022 km². Die Verteilung in Deutschland ist in Abbildung 1 (b) visualisiert. Dabei ist deutlich zu erkennen, dass in Ballungsgebieten (z. B. Berlin oder das Ruhrgebiet) sehr viele Gebäudegrundrisse erfasst sind, wohingegen in ländlichen Gebieten noch jede Menge freie Flächen vorhanden sind.

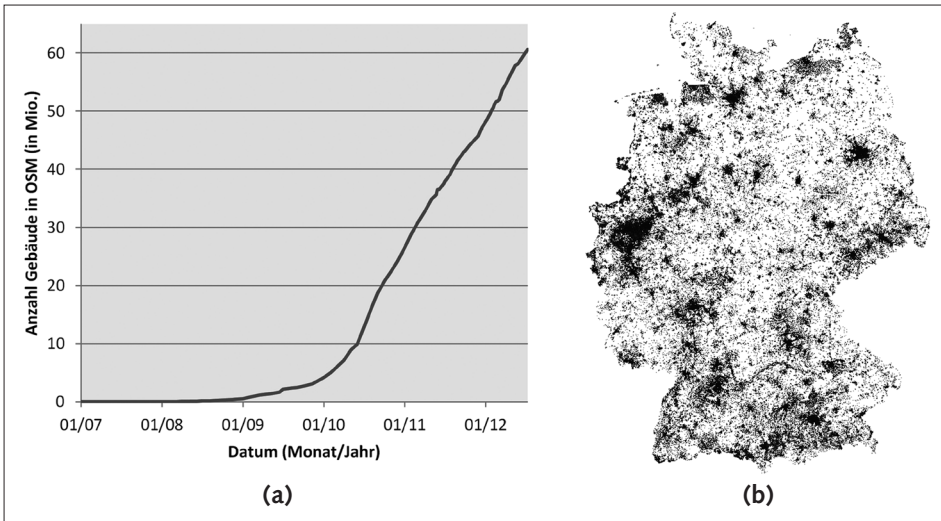


Abb. 1: Entwicklung des globalen Gebäudebestands in OSM seit 2007 (a) und Visualisierung der OSM-Gebäudepolygone in Deutschland (b) (beides Quelle: Eigene OSM-Datenbank)

Vergleicht man die absoluten Zahlen von OSM mit der Realität (sprich mit amtlichen Daten, siehe Beitrag von Behnisch et al. in diesem Band), wird allerdings deutlich, dass – rein quantitativ gesehen – nur etwa 13 % aller Gebäude von Deutschland in OSM erfasst sind. Diese Zahl wird allerdings negativ von einem in OSM weit verbreitetem Phänomen beeinflusst: dem Mapping for the Renderers Prinzip, welches beschreibt, dass viele OSMer nur beitragen, was auf der Karte auch sichtbar ist und das insbesondere nur die Darstellung auf der Karte entscheidend ist. Dies hat zur Folge, dass – gerade in städtischen Gebieten – mehrere nebeneinanderliegende Gebäude zu einem großen Gebäudeblock zusammengefasst werden. Für die Karte – in welcher Gebäude oftmals nur als Verschönerung und nicht als wichtige Information wahrgenommen werden – scheint diese Art der Erfassung zu genügen. Benötigt man aber Informationen über die einzelnen Gebäude oder möchte Statistiken erheben, so führt dies zu Problemen. Oftmals kann es daher sein, dass z. B. 50 einzelne Gebäude eines Blocks in OSM lediglich als ein Gebäude gemappt sind, weswegen die o. g. Vollständigkeitszahlen negativ beeinflusst werden. Diese Problematik wurde von OSM auch schon wahrgenommen und einige Mitglieder haben begonnen, die Gebäudekonglomerate mühsam aufzuteilen und somit diesem negativen Effekt entgegen zu wirken.

3 Detaillierte Betrachtung der OSM-Gebäudedaten

Nach der rein quantitativen Betrachtung des Gebäudebestands für Deutschland in OSM unternimmt dieses Kapitel eine detailliertere Betrachtung der Gebäudedaten. Dabei wird zum einen auf die Geometrie (vgl. 3.1) sowie auf die Attributierung (vgl. 3.2) eingegangen.

3.1 Geometrische Eigenschaften

Insgesamt gehören in Deutschland knapp 35,32 Millionen Punkte zu Gebäuden, wobei Gebäude typischerweise als geschlossener Weg (Way) repräsentiert werden. Bei den knapp 6,5 Millionen Gebäudegrundrissen handelt es sich zum größten Teil um einfache Flächen ohne Löcher. Lediglich 10 087 Gebäude bestehen aus komplexeren Polygonen mit Löchern (wie etwa einem Hof o. ä.). 75,44 % dieser komplexen Gebäude (knapp 7 600) haben genau ein solches Loch, wohingegen es ein Gebäude mit insgesamt 26 Löchern gibt.

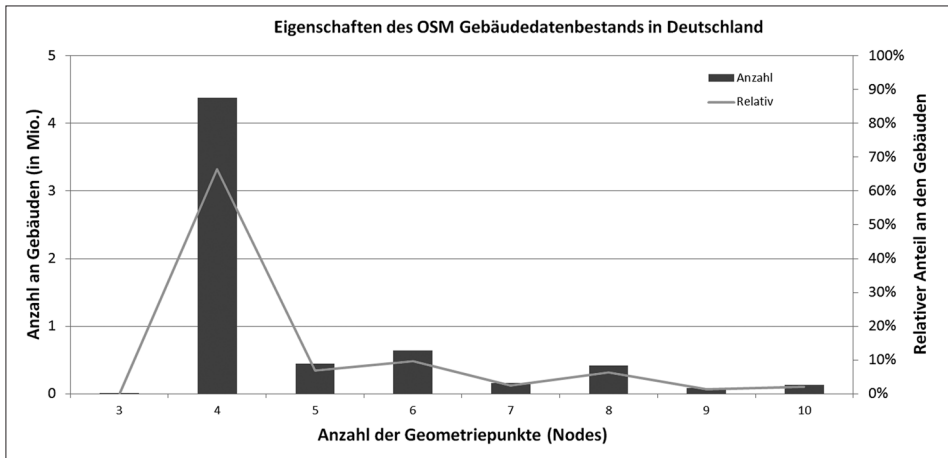


Abb. 2: Anzahl der Geometriepunkte von Gebäuden (Quelle: Eigene OSM-Datenbank)

Ein Gebäude besteht im Durchschnitt aus 5,38 Punkten pro Grundriss, wobei das Minimum bei drei Punkten liegt und das Maximum bei 1 332 Punkten (Landgericht Bremen). Betrachtet man den gesamten Gebäudebestand von Deutschland ergibt sich, dass 95 % aller Grundrisse maximal aus 10 Punkten bestehen. Sehr gut zu erkennen ist, dass der Großteil der Gebäudegrundrisse als Viereck repräsentiert wird (knapp 66 %). Die genaue Verteilung – sowohl absolut als auch relativ gesehen – ist in Abbildung 2 visualisiert. Das gehäufte Vorkommen von viereckigen Gebäudegrundrissen lässt darauf schließen, dass die vorhandenen Daten oftmals eine vereinfachte Darstellung des tatsächlichen Gebäudes sind. Weiter ist diese Eigenschaft auch dadurch bekräftigt, dass immer häufiger Luftbilder als Grundlage für Mapping-Aktivitäten dienen, sodass OSMer vielmehr die Dachgeometrie als den konkreten Gebäudegrundriss mappen. Da sich kleine Erker und Kanten des Grundrisses oftmals nicht in der Dachform widerspiegeln, kommt es zur Erfassung einer vereinfachten Gebäudegeometrie.

Ein Gebäude bedeckt im Durchschnitt eine Grundfläche von 307,64 m². Es gibt einige wenige Artefakte (453) mit einer Grundfläche die kleiner als 1 m² ist – diese können aber wohl nicht als Gebäude bezeichnet werden. Das flächenmäßig größte Gebäude hat eine

Gebäudegrundfläche von knapp 445 000 m². Dies stellt aber eher eine Ausnahme dar, denn 95 % aller Gebäude haben eine Fläche von weniger als 1 000 m². Betrachtet man Gebäude mit einer Fläche zwischen 0 m² und 100 m² (vgl. Abb. 3) so wird deutlich, dass 5 % aller Gebäudegrundrisse weniger als 30 m² umfassen, was darauf hindeutet, dass OSM auch kleine „Gebäude“, wie etwa Hütten oder Garagen, beinhaltet, welche in der Realität wohl eher als Nebengebäude oder Teil eines Gebäudekomplexes zu bewerten sind.

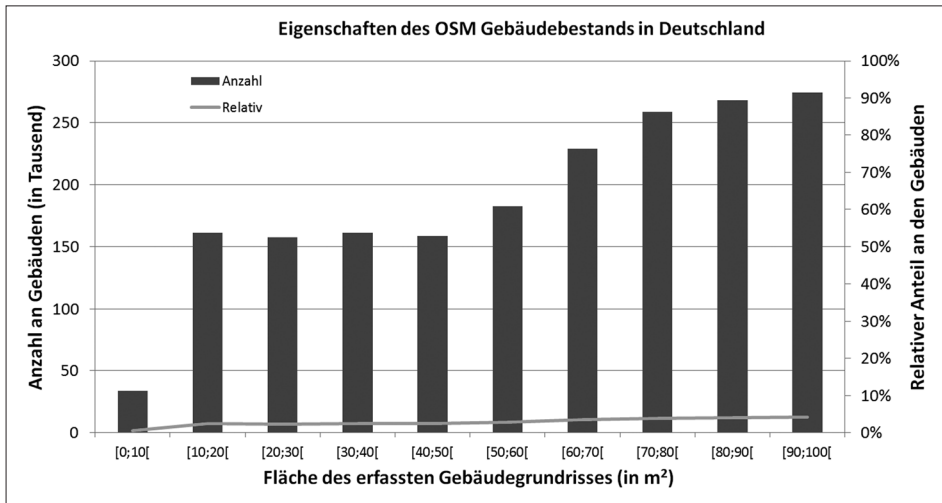


Abb. 3: Gebäude mit einer Fläche zwischen 0 m² und 100 m² (Quelle: Eigene OSM-Datenbank)

3.2 Attributierung

Wie bereits erwähnt, besteht OSM nicht nur aus purer Geometrie, sondern auch aus Tags, welche das betroffene OSM-Feature näher beschreiben bzw. mit zusätzlichen Informationen anreichern. Eine der wichtigsten Informationen über Gebäude (insbesondere für Navigationsanwendungen) ist die Adresse eines Gebäudes. Es hat sich gezeigt, dass etwa 17,67 % der Gebäude mit einer kompletten Adresse, sprich Straße, Hausnummer, Postleitzahl und Ort, angereichert sind. Auffällig ist, dass allerdings 21,82 % über eine Straße und eine Hausnummer verfügen. Dies lässt den Schluss zu, dass OSMer diese Informationen als wichtiger erachten, da sich die Stadt und Postleitzahl oftmals aus der geo-referenzierten Geometrie ableiten ließen. Doch OSMer sammeln nicht nur Adressinformationen, sondern auch weiterreichende Informationen, welche sowohl das Gebäude an sich, z. B. den Namen, den Typ o. ä., aber auch die Geometrie näher beschreiben. Im letzteren Fall ist anzumerken, dass Tags tendenziell dazu benutzt werden, 3D-Informationen (wie etwa die Höhe des Gebäudes oder die Dachform) als Attribut zu speichern. Tabelle 1 gibt einen Überblick über die Häufigkeit solcher Informationen.

Tab. 1: Verfügbarkeit von zusätzlichen (semantischen) Informationen
(Quelle: Eigene OSM-Datenbank mit Stand Mitte Juni 2012)

Attribut	Absolut	Relativ * in Prozent
Name	1 640 682	24,9651
Gebäudetyp	75 591	1,1502
Anzahl Stockwerke	49 147	0,7478
Höhe	33 054	0,5030
Dach	15 387	0,2341
Dachform	12 103	0,1842
Dachfarbe	4 703	0,0716
Dachorientierung	2 962	0,0451
Fassadenfarbe	2 271	0,0346
Dachhöhe	1 067	0,0162
Dachwinkel	666	0,0101

* basierend auf 6 571 891 Gebäudegrundrissen

4 Nutzungspotenzial

OSM birgt, nicht zuletzt wegen den vielen verfügbaren Prozessierungstools und der kostenlosen Nutzung für (kommerzielle) Anwendungen, ein enormes Potenzial für verschiedene Anwendungsfelder. OSM-Daten eignen für die Erstellung von 2D-WebKarten (Goetz et al. 2012), welche auch um Routingfunktionalität erweitert werden können (Neis, Zipf 2008). Das Straßennetz eignet sich für die Durchführung von Verkehrssimulationen (Zilske 2011). Die 3D-Informationen können dazu genutzt werden, um digitale 3D-Stadtmodelle automatisiert zu erzeugen (Over et al. 2010). Außerdem hat sich gezeigt, dass CityGML LoD1- und LoD2-Gebäudemodelle automatisiert erzeugt werden können (Goetz, Zipf 2012), sodass diese in einer Geodateninfrastruktur (GDI) ausgetauscht werden können. Darüber hinaus können Informationen wie etwa Dachform oder Dachneigung für Solarpotenzialanalysen herangezogen werden, wobei diese Informationen momentan noch kaum vorhanden sind (vgl. Tab. 1).

5 Fazit

Die Entwicklungen des Datenbestandes von OSM innerhalb der letzten Wochen und Monate lassen ein zunehmendes Interesse an Gebäudedaten und insbesondere Gebäudegrundrissen erkennen. Betrachtet man die derzeitigen Wachstumsraten scheinen ähnliche Vollständigkeitszahlen wie für Straße (~91 %), allerdings erst in etwa vier bis fünf Jahren realistisch. Bezüglich der erfassten Geometrien lässt sich annehmen, dass die meisten Gebäudepolygone auf Basis von frei verfügbaren Luftbildern gemappt wur-

den und somit größtenteils nicht direkt Gebäudegrundrisse, sondern eher Dachflächen gemappt werden. Insbesondere lässt dies den Schluss zu, dass viele Gebäudegrundrisse in OSM eher vereinfacht dargestellt werden. Weiterhin wurde deutlich, dass zusätzliche (semantische) Informationen nur selten zur Verfügung gestellt werden. Lediglich Adressinformationen stehen für etwa ein Fünftel aller Gebäudegrundrisse in Deutschland zur Verfügung. Dies ist zum einen daraus begründet, dass zusätzliche (semantische) Informationen oftmals das Wissen von Ortskundigen voraussetzen. Weiterhin ist in OSM (nicht ausschließlich bei Gebäuden, sondern auch bei anderen Objektarten) oftmals ein „Mapping for the renderers“ Paradigma zu erkennen: OSM-Mitglieder tragen nur Informationen bei, welche entweder in der Karte angezeigt werden, oder welche einen weiteren Nutzen (wie etwa Adressen für Navigation) für den Beitragenden haben. Gebäudeinformationen haben, so scheint es, im Moment noch keine große Bedeutung für die OSM-Mitglieder. Nichtsdestotrotz hat sich bereits gezeigt, welches Nutzungspotenzial in Gebäudedaten von OSM für verschiedenste Anwendungsfälle steckt. Durch die Entwicklung und Verbesserung von Anwendungen, die solche Daten (z. B. die Gebäudehöhe) nutzen, kann ein Anreiz für die Erfassung solcher Daten geschaffen werden. Das Problem mit Gebäuden, welche zu einem Polygon zusammengefasst wurden, wurde bereits erkannt und diverse OSMer haben damit begonnen, diese Daten zu korrigieren.

6 Literatur

- Goetz, M.; Lauer, J.; Auer, M. (2012): An Algorithm Based Methodology for the Creation of a Regularly Updated Global Online Map Derived From Volunteered Geographic Information. *GEOProcessing 2012*, Valencia.
- Goetz, M.; Zipf, A. (2012): Towards Defining a Framework for the Automatic Derivation of 3D CityGML Models from Volunteered Geographic Information. In: *International Journal of 3-D Information Modeling* 1(2), 1-16.
- Goodchild, M. (2007): Citizens as sensors: the world of volunteered geography. In: *GeoJournal* 69(4)/2007, 211-221.
- Neis, P.; Zielstra, D.; Zipf, A. (2012): The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007-2011. In: *Future Internet* 4(1)/2012, 1-21.
- Neis, P.; Zipf, A. (2008): OpenRouteService.org is three times „Open“: Combining Open-Source, OpenLS and OpenStreetMaps. *GIS Research UK (GISRUK)*, Manchester.
- OSM – OpenStreetMap (2012): Map Features. http://wiki.openstreetmap.org/wiki/Map_Features (Zugriff: 10.07.2012).
- Over, M.; Schilling, A.; Neubauer, S.; Zipf, A. (2010): Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany. *Computers, Environment and Urban Systems* 34(6)/2010, 496-507.
- Zilske, M.; Neumann, A.; Nagel, K. (2011): OpenStreetMap For Traffic Simulation. *SOTM-EU 2012*, Wien.