

---

# iOSMANalyzer – ein Werkzeug für intrinsische OSM Qualitätsuntersuchungen

Christopher BARRON, Pascal NEIS und Alexander ZIPF

## Zusammenfassung

Die zunehmende Verfügbarkeit von freiwillig und kollaborativ gesammelten Geodaten führte zu verschiedenen wissenschaftlichen Studien mit dem Ziel, die Qualität dieser Daten genauer bestimmen zu können. Dabei werden unterschiedliche Verfahren eingesetzt und weiterentwickelt, um die verschiedenen Aspekte von Datenqualität abbilden zu können. In vielen Fällen beruht die Analyse dabei auf einem Vergleich mit einem vorhandenen Referenzdatensatz. Daneben gibt es – sogenannte „intrinsische“ – Maße, die aus den Daten selbst gewonnen werden können. Bisher existiert kein Rahmenwerk, mit dem die diversen Indikatoren für die Qualität von Geodaten gemeinsam analysiert werden können. Stattdessen konzentrieren sich die Autoren i. d. R. auf einzelne Untersuchungen. Die hier vorgestellte Arbeit hat zum Ziel diesen Missstand zu beheben und ein integriertes Werkzeug zur Qualitätsanalyse der Geodaten in *OpenStreetMap* (OSM) zu entwickeln. Dabei werden unterschiedliche Verfahren aus der Literatur aufgegriffen und weiterentwickelt. Der Schwerpunkt liegt auf intrinsischen Indikatoren der Datenqualität unter Berücksichtigung der vollständigen Historie der Geodaten. Bisherige Studien verwendeten zumeist einzelne zeitliche Snapshots. Das Ergebnis ist das Softwarewerkzeug *iOSMANalyzer* für Endanwender, das aus OSM-Full-History Daten für beliebige Gebiete umfassende Statistiken, Diagramme und Karten zu mehreren intrinsischen Qualitätsindikatoren erstellen kann. Im Sinne des „Fitness for Purpose“ werden diese in „Übersichtsinformationen“, sowie die Einsatzbereiche „Routing & Navigation“, „Adresssuche“, „Points of Interest-Suche“, „Kartenanwendungen“ und „Aktivität & Verhalten der Mitwirkenden“ unterteilt.

## 1 Einleitung

Bei der Erfassung von geographischen Informationen, durch die Mitwirkenden des OSM Projektes, gibt es keine formelle Instanz, welche die Richtigkeit oder die Qualität der Beiträge beurteilt. Grundsätzlich darf jeder Mitwirkende die Daten nach seinem Ermessen erfassen und bearbeiten. Dies hat zur Folge, dass potentiell qualitativ große Unterschiede in einem Gebiet entstehen können. Ein wichtiger Aspekt bei OSM ist daher die Bewertung der Datenqualität. Hierfür gibt es zahlreiche Online-Tools (u. a. OSMMatrix, OpenStreetBugs, KeepRight oder OSMInspector). Neben diesen Anwendungen kann die Qualität auch mit Hilfe von Referenzdatensätzen bestimmt werden. Qualitativ hochwertige Referenzdatensätze unterliegen jedoch oft hohen Lizenz- und Beschaffungsgebühren oder sind schlichtweg für ein Gebiet nicht verfügbar. Folglich sind Methoden und Analysen notwendig, mit denen in einem intrinsischen Ansatz die Qualität von OSM Daten analysiert werden kann. Dies

bedeutet, dass nur die zu untersuchenden Daten für eine Qualitätsanalyse herangezogen werden und kein Vergleichsdatensatz benötigt wird.

Die Motivation für diese Arbeit lautet daher: „*Wie kann die Qualität von OpenStreetMap unter Einbeziehung ihrer Datenhistorie bewertet werden?*“ Hierfür werden Methoden und Verfahren für eine derartige Qualitätsanalyse vorgestellt. Umgesetzt werden diese in einem prototypisch implementierten Softwarewerkzeug mittels Python und unter der Verwendung einer PostgreSQL/PostGIS Datenbank. Als Untersuchungsgrundlage dient die vollständige Datenhistorie frei vom Nutzer bestimmbarer Gebiete. Da OSM Daten für unterschiedliche Zwecke eingesetzt werden, müssen die Analysemethoden den potentiellen Anwendungsfällen angepasst werden.

Der Beitrag gliedert sich wie folgt: Abschnitt 2 zeigt den aktuellen Stand der Forschung auf. Nach allgemeinen Geodatenqualitätsparametern wird dann auf bereits bestehende Qualitätsuntersuchungen mit Bezug zu OSM eingegangen. In Abschnitt 3 wird die Umsetzung vorgestellt und in Abschnitt 4 an zwei ausgewählten Beispielen diskutiert. Eine Diskussion mitsamt Ausblick schließt den Beitrag ab.

## 2 Qualität von Geodaten - Stand der Forschung

Im Bereich des Geoinformationswesens können die Prinzipien der ISO („International Organization for Standardization“) 19113 und 19114 herangezogen werden. Diese beschreiben allgemeine Grundsätze zur Qualität von Geoinformationen (ISO 19113) und Verfahren zur Ermittlung der Geodatenqualität (ISO 19114). Sie beinhalten u. a. folgende Parameter, anhand derer die Qualität von Geodaten bestimmt werden können (vgl. VAN OORT 2006):

1. Entwicklung („Lineage“): die historische Entwicklung eines Datensatzes.
2. Vollständigkeit („Completeness“): die Vollständigkeit eines Datensatzes.
3. Logische Konsistenz („Logical Consistency“): beinhaltet sowohl topologische Konsistenz als auch die Attributierung und die Beziehungen innerhalb der DB.
4. Positionsgenauigkeit („Positional Accuracy“): auf Koordinaten bezogen.
5. Zeitliche Genauigkeit („Temporal Accuracy“): beinhaltet:
  - „Accuracy of Measurement“: Genauigkeit von Zeitangaben.
  - „Temporal Consistency“: Korrekte zeitliche Reihenfolge der Daten.
  - „Temporal Validity“: Gültigkeit der Daten in Bezug auf einen bestimmten Zeitpunkt.
6. Thematische Genauigkeit („Thematic Accuracy“): die richtige syntaktische Attributierung („Syntactic Accuracy“) der Daten, als auch die korrekte Zuordnung von Objekten zu ihren Objektklassen („Semantic Accuracy“).

Die Qualität von Geodaten muss jedoch immer auch in Relation zum Einsatzzweck betrachtet werden. So ist beispielsweise die Vollständigkeit der Attribute „maximal zulässige Geschwindigkeit“ oder „Einbahnstraße: ja/nein“ in einem Geodatensatz, der für Navigation bestimmt ist, von großer Bedeutung, aber weniger bei reinen Kartenanwendungen.

## 2.1 Qualitätsuntersuchungen mit Bezug zu OpenStreetMap

Erste Untersuchungen zur Qualität von OSM wurden von HAKLAY (2008) durch einen Vergleich von OSM Daten mit einem Referenzdatensatz durchgeführt. Dabei verglich er OSM mit dem Referenz-Geodatensatz „Ordnance Survey Meridian 2“ (OS Meridian 2) für England (insbesondere den Großraum London). Seine Analyse zur Positionsgenauigkeit der OSM Daten basiert auf der „Puffermethode“ von GOODCHILD & HUNTER (2007). In einer ähnlichen Untersuchung verglich ATHER (2009) ebenfalls für Großbritannien die OSM-Daten mit dem „OS MasterMap“ Referenzdatensatz. Neben der Positionsgenauigkeit untersuchte KOUNADY (2009) auch die Qualitätsparameter „Vollständigkeit“ und „thematische Genauigkeit“. Hierfür verglich sie die OSM-Straßendaten Athens mit denen des *Hellenic Military Geographical Service* (HMGS). Für Deutschland vergleichen erstmalig ZIELSTRA & ZIPF (2010) das OSM-Straßennetz mit einem kommerziellen Datensatz (TeleAtlas-MultiNet), um Aussagen zur Positionsgenauigkeit, Vollständigkeit und Anwendbarkeit der OSM Daten treffen zu können. Ebenfalls für Deutschland untersucht LUDWIG (2010) das OSM Straßennetz mit Hilfe der Navteq-Straßendaten als Referenz. Die Ergebnisse der Studien von ZIELSTRA & ZIPF (2010) und LUDWIG (2010) bestätigen, dass im Bereich urbaner Großräumen sich OSM Daten durch großen Detailreichtum, Vollständigkeit und eine hohe Positionsgenauigkeit auszeichnen, wohingegen in ländlicheren Gebieten die Positionsgenauigkeit der Daten deutlich abnimmt. NEIS ET AL. (2012) analysieren die zeitliche Entwicklung des OSM Straßennetzwerkes für Deutschland zwischen 2007 und 2011. GIRRES & TOUYA (2010) untersuchen den französischen OSM Datensatz. Als Referenzdatensatz wurde der *BD TOPO Large Scale Referential* (RGE) Datensatz des *Institut Géographique National* (IGN) herangezogen. In ihrem Ansatz überprüfen sie erstmalig einen OSM Datensatz nach den in der ISO 19113:2002 definierten Kriterien. Neben Straßen untersuchen sie unter anderem auch Küstenlinien, Seen und administrative Grenzen. Auch MOONEY ET AL. (2010) und MOONEY ET AL. (2010a) analysieren Polygone für natürliche Objekte wie Wasserflächen, Wälder, etc. in OSM. Letztere richten das Augenmerk ihrer Untersuchung vor allem auf den Detailgrad der Repräsentation dieser Polygone. Hierfür nutzen sie die durchschnittliche Distanz zwischen benachbarten Nodes und die Äquidistanz aller Nodes in einem Polygon als Indikator. Ein anderer Weg zur Qualitätsbewertung von Polygonrepräsentationen sind sog. „Shape Similarity Tests“ (MOONEY & CORCORAN 2011). REHRL ET AL. (2012) entwickeln ein Verfahren, mit dem sie den Verkehrsgraphen für die Berechnung von länderübergreifenden Erreichbarkeitspotentialen auf der Basis von OSM Daten evaluieren. Sie zeigen auf, dass ausgewählte Siedlungsgebiete in einem hohen Maß an den Verkehrsgraphen angebunden sind. HELBICH ET AL. (2012) führen eine Analyse zur Positionsgenauigkeit von OSM Daten mit mehreren proprietären Referenzdatensätzen (TomTom, TeleAtlas und amtliche Daten) mit einem eigenen methodischen Ansatz durch. NEIS & ZIPF (2012) untersuchen das Verhalten der Community als wichtigen Indikator für die Qualität der Daten. In einer weiteren Studie auch speziell Verfahren zur Detektierung von Vandalismus (NEIS ET AL. 2012).

## 3 Umsetzung

Die Vorgehensweise in dieser Arbeit unterscheidet sich in mehrfacher Hinsicht von bisherigen Untersuchungen: Da die komplette Historie eines OSM Datensatzes in die Analysen einbezogen wird, kann die gesamte zeitliche Dimension berücksichtigt werden. Andere

Untersuchungen haben in der Vergangenheit zwar auch die zeitliche Entwicklung bestimmter Merkmale in OSM betrachtet, dies jedoch lediglich unter Zuhilfenahme von OSM Datenbank Snapshots verschiedener Zeitstempel. Damit können ebenfalls Entwicklungen aufgezeigt und Qualitätsaussagen getätigt werden, jedoch nicht in dem Detailgrad, wie es durch die Verwendung eines kompletten OSM-Full-History-Dumps möglich ist. Des Weiteren wird in dieser Arbeit kein Referenzdatensatz zum Vergleich herangezogen, sondern ein intrinsischer Ansatz zur Untersuchung der Datenqualität verwendet. Dies besagt, dass nur die OSM Daten an sich analysiert und nicht mit einem anderen Datensatz verglichen werden (extrinsische Bewertung). Dies hat zur Folge, dass bei zahlreichen Verfahren keine absoluten Aussagen zu den einzelnen Qualitätsparametern möglich sind, sondern einige nur annähernd bestimmt werden können.

Nach Recherche der in der Literatur zu findenden Ansätze zu Datenqualitätsanalysen in OSM wurden derzeit 27 Indikatoren mit jeweils zahlreichen Optionen und Darstellungsvarianten ausgewählt. Anschließend wurden diese in einem gemeinsamen Software-Werkzeug implementiert. Die Gliederung der verschiedenen Analysen des Werkzeuges ergibt sich aus den unterschiedlichen Einsatzbereichen. Nach „Übersichtsinformationen“ sind dies die Einsatzbereiche „Routing & Navigation“, „Adresssuche“, „Points of Interest-Suche“, „Kartenanwendungen“ und „Aktivität & Verhalten der Mitwirkenden“. Folgende Qualitätsindikatoren werden im Werkzeug berücksichtigt:

### **1. Übersichtsinformationen**

- Entwicklung der Punkte, Linien & Polygone
- Aktualität des Datensatzes
- Gegenüberstellung von erstellten Daten & Datenbearbeitungen
- Quantitative Analysen zur Attributierung von OSM Objekten
- Syntaktische Attributgenauigkeit
- Analyse der Positionsgenauigkeit von Straßenkreuzungen vers. Zeitstempel

### **2. Routing & Navigation**

- Länge des Gesamtwege- und Straßennetzwerks
- Entwicklung der syntaktischen Attributgenauigkeit bei Straßen & Wegen
- Aktualität des Gesamtwegenetzwerks
- Logische Konsistenz des Gesamtwegenetzwerks
- Straßen ohne Namen oder Routenbeschriftung
- Grid-basierte Analyse der Positionsgenauigkeit des Gesamtwegenetzwerks

### **3. Adresssuche**

- Art & Entwicklung der vollständigen Adressierung
- Vollständigkeit der Hausnummern aller Gebäudepolygone

### **4. „Points of Interest“-Suche**

- Entwicklung der POI-Gesamtzahl & der durchschnittlichen Anzahl an Attributen
- Attributive Vollständigkeit verschiedener POI-Kategorien

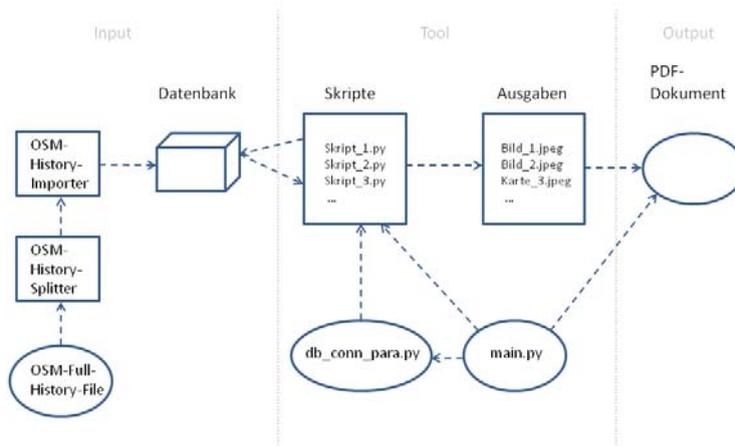
### **5. Kartenanwendungen**

- Entwicklung der Geometrie-Repräsentation von Polygonen
- Unbearbeitete Punkte, Linien & Polygone
- Entwicklung invalider Polygon-Geometrien
- Logische Konsistenz von Landnutzungs-Polygonen
- Quantitative Entwicklung relevanter Polygone

## 6. Aktivität & Verhalten der Mitwirkenden

- Anzahl & Aktivität der Mitwirkenden
- (Un)gleichverteilung der Beiträge
- Beitragsverhalten der Mitwirkenden

Mit der Hilfe des neu entwickelten Werkzeugs *iOSMAnalyzer* wird eine Zusammenfassung der genannten Indikatoren generiert. Diese beinhaltet zahlreiche Diagramme, Statistiken, Karten und Tabellen und wird als einzelnes PDF-Dokument ausgegeben. Die Gliederung des Dokuments erfolgt analog zu dieser Arbeit nach verschiedenen Anwendungsfällen der OSM Daten. Einen Überblick über die Architektur des *iOSMAnalyzers* bietet Abbildung 1. Grundlage für das Werkzeug ist eine PostgreSQL/PostGIS Datenbank. Das Extrahieren der Daten aus einem OSM-Full-History-Dump erfolgt durch den OSM-History-Splitter<sup>1</sup>.



**Abb. 1:** Architektur des prototypisch implementierten *iOSMAnalyzer*

## 4 Ausgewählte Beispiele für Analyseoptionen der Datenqualität

Aufgrund des eingeschränkten Platzes sollen in diesem Beitrag lediglich zwei Beispiele für Analyse-Optionen des neuen *iOSMAnalyzers* veranschaulicht werden.

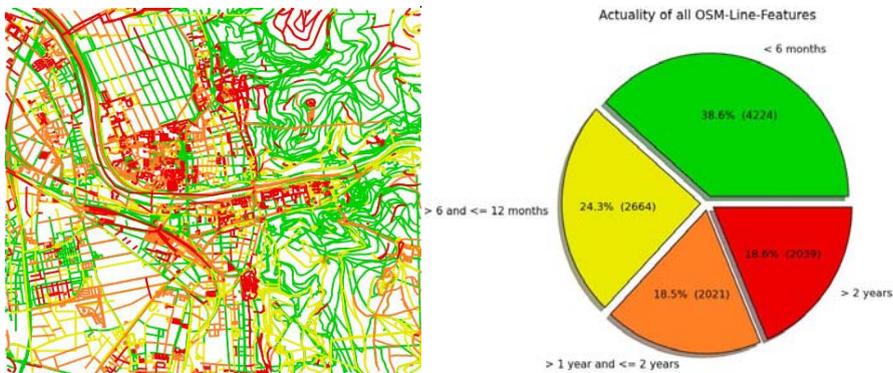
### 4.1 Beispiel 1: Aktualität des Wegenetzwerks

Ein wichtiger Faktor bei der Bestimmung der Qualität von Geodaten, der bei VGI-Projekten (Volunteered Geographic Information) von großer Bedeutung ist, ist die Berücksichtigung des zeitlichen Aspekts. Nachdem die Daten anfänglich erhoben wurden, ist es für die Qualität der Daten von großer Bedeutung, dass diese in der Folge von der Community gepflegt und aktualisiert werden. Je nach Anwendung spielt die Aktualität eine mehr

<sup>1</sup> <https://github.com/MaZderMind/osm-history-splitter>

oder minder wichtige Rolle. Bei Routing- und Navigationsanwendungen kommt ihr beispielsweise große Bedeutung zu, da sich Änderungen in Geschwindigkeitsbegrenzungen, Fahrtrichtungen oder Abbiegevorschriften unmittelbar auf die Qualität der Anwendung auswirken und daher so aktuell wie möglich sein sollten.

Im Idealfall werden Geometrie und Attribute der Daten von der Community fortlaufend aktualisiert und somit auf dem neuesten Stand gehalten. Bei diesem Prozess ist es zudem wichtig, dass die Aktualisierung im Datensatz homogen geschieht und sich nicht nur auf bestimmte Daten beschränkt (NEIS & ZIPF 2012, MOONEY ET AL. 2010). Eine Möglichkeit die Aktualität von OSM Daten näher zu untersuchen besteht darin, das Datum der letztmaligen Bearbeitung des jeweiligen Objekts zu betrachten. VAN EXEL (2012) greift dies auf und spricht dabei von der „*data staleness*“, d. h. „Abgestandenheit der Daten“.



**Abb. 2:** Aktualität aller Linien in Heidelberg (Karte und Diagramm)

Der Zeitpunkt der letztmaligen Bearbeitung jedes Punkts, jeder Linie bzw. jedes Polygons wird als jeweils aktuellster interpretiert. Die Darstellung der Aktualität erfolgt mit Hilfe der prozentualen Verteilung der gegenwärtig gültigen Daten (vgl. Abb. 2). Hierfür werden folgende Klassen gebildet:

- Die letzte Bearbeitung liegt weniger als 6 Monate zurück (grün).
- Die letzte Bearbeitung liegt zwischen 6 und 12 Monaten zurück (gelb).
- Die letzte Bearbeitung liegt zwischen 1 und 2 Jahren zurück (orange).
- Die letzte Bearbeitung liegt mehr als 2 Jahre zurück (rot).

Grundsätzlich spricht es für eine bessere Datenqualität, wenn der Großteil der Daten möglichst aktuell ist (Objekterstellung oder letztmalige Bearbeitung liegt nicht länger als 6 Monate zurück). Es kann argumentiert werden, dass der Mitwirkende mit der letztmaligen Änderung zudem für die Richtigkeit des Objekts verantwortlich ist und dies mit seinem Datenupload bestätigt. Je größer der prozentuale Anteil der übrigen Kategorien ist, desto weniger aktuell ist der Bestand. Des Weiteren können bei der Analyse der Objektaktualität auf der Mikroebene möglicherweise Muster der Fragmentierung erkennbar werden, die sich u. U. sogar kleinräumig in funktional ähnlichen städtischen Bereichen eines Gebiets zeigen. Der Grund hierfür könnte sein, dass Mitwirkende im OSM Projekt Gebiete nicht gleichermaßen flächendeckend aktualisieren, sondern sich bei ihrer Arbeit auf bestimmte Objekte

fokussieren, bzw. diese bei ihrer Arbeit bevorzugt bearbeiten. Dies wird dann deutlich, wenn die Aktualität verschiedener Objekte in einem Gebiet betrachtet wird. Hierfür eignen sich v. a. kartographische Darstellungen der Aktualität (vgl. Abb. 2). Für die Interpretation dieser ist jedoch auch lokales Wissen über die Community bzw. das Gebiet vorteilhaft.

Abschließend kann festgehalten werden, dass es schwierig ist, die Aktualität eines Datensatzes als Ganzes zu betrachten. Stattdessen muss dies differenziert anhand von unterschiedlichen Objekttypen geschehen, da die Mitwirkenden Gebiete nicht „en bloc“ bearbeiten, sondern ihre Beiträge eher auf unterschiedliche Objekte beziehen. Das Werkzeug unterscheidet zunächst Punkte, Linien und Polygone. Dabei wird ersichtlich, dass sich die Aktualität derselben z. T. deutlich voneinander unterscheidet. Für detaillierte Informationen zur Aktualität wäre es sinnvoll, die Daten in weitere Objektgruppen aufzuteilen, die für den jeweiligen Anwendungsbereich von Interesse sind.

Problematischer zu interpretieren ist der Fall, in dem Objekte in der Vergangenheit vollständig und korrekt erfasst wurden und sich seitdem keine Änderungen in der Realität ergeben haben. In dem Ansatz dieser Arbeit würde ein solches Objekt das Ergebnis der Statistik negativ beeinflussen, trotz der Tatsache, dass es vollständig erfasst in OSM vorliegt. Ein Ansatz, solche Objekte zu identifizieren, besteht in Anlehnung an VAN EXEL (2010) darin, im Umkreis dieser Objekte die Zeitstempel der anderen Objekte zu betrachten. Wenn diese größtenteils auf dem neuesten Stand sind, so könnte daraus abgeleitet werden, dass besagte Objekte trotz einer langen Zeit ohne Veränderung vollständig mit Attributen versehen und v. a. in Lage und Ausdehnung korrekt erfasst wurden. Diese Schlussfolgerung kann deshalb gezogen werden, weil anzunehmen ist, dass die Bearbeiter der umliegenden Objekte diese Tatsache erkannt haben und durch ihr „nicht-tätig werden“ stillschweigend dieses als vollständig bzw. korrekt erfasst beurteilen.

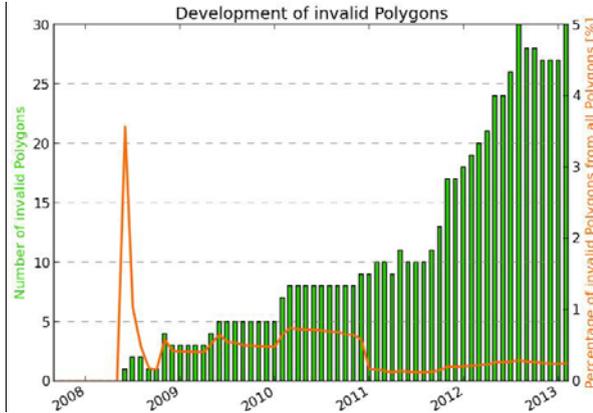
## 4.2 Beispiel 2: Entwicklung invalider Polygon-Geometrien

Die Validität von Geodaten spielt in zahlreichen GIS-Anwendungen und -Operationen eine wichtige Rolle. Invalide Geometrien können zu unerwünschten topologischen Fehlern führen. In Geodatenbanken beschreibt das Simple-Feature-Modell des *Open Geospatial Consortium* (OGC) die Darstellung von zweidimensionalen Vektorgeometrien (BRINKHOFF 2008). Eine Geometrie ist demnach wohlgeformt (und damit valide), wenn sie keine anormalen geometrischen Punkte (z.B. sich selbst-schneidende oder berührende Geometrien) aufweist und 0- bzw. 1-dimensional ist.

Mit der „ST\_IsValid“-Funktion in PostGIS können Geometrien auf Wohlgeformtheit überprüft werden. Die Funktion „ST\_IsValidReason“ gibt die Begründung für die ggf. nicht vorliegende Wohlgeformtheit aus. Beide Funktionen sind konform mit den OGC OpenGIS-Spezifikationen. Sie werden hier eingesetzt, um monatlich die absolute Anzahl an invaliden Polygon-Geometrien über die Historie des Datensatzes zu berechnen. Zudem wird ihr prozentualer Anteil dargestellt, um sie mit der Gesamtzahl aller Polygone vergleichen zu können. Je geringer die Zahl der invaliden Polygone ist, desto besser ist der Datensatz. Trotz einer Zunahme an invaliden Polygonen deutet ein abnehmender prozentualer Anteil, relativ betrachtet, auf weniger invalide Geometrien hin.

Abbildung 3 zeigt für Heidelberg: Obwohl die absolute Anzahl an invaliden Geometrien (grün) stetig steigt, sinkt dennoch ihr prozentualer Anteil (orange) an der Gesamtheit aller

Polygone, nach einem anfänglich sehr hohen Wert (der auf die geringe Gesamtzahl von Polygonen zurückzuführen ist) und ist danach nie größer als 1 %.



**Abb. 2:** Zeitlicher Verlauf der Anzahl an invaliden Polygon-Geometrien & ihr prozentualer Anteil an der Gesamtheit aller Polygone in Heidelberg

Letztlich lässt sich festhalten, dass bei nur 0,2 % (Jan 2012) invaliden Geometrien der größte Teil der Polygone im exemplarischen Heidelberger Untersuchungsgebiet valide ist, was einen positiven Indikator für die Qualität der Polygone darstellt.

## 5 Synthese und Ausblick

Mit einer zunehmenden Zahl von angemeldeten Mitgliedern und den von ihnen erfassten Daten entwickelt sich das OSM Projekt kontinuierlich weiter. Dabei ist insbesondere die Qualität der Daten von großer Bedeutung. Es wurden verschiedene Indikatoren und Verfahren vorgestellt, mit denen die Qualität von OSM Daten anhand ihrer Historie beurteilt werden können. Der durchgehend intrinsische Ansatz ermöglicht dies ohne einen Referenzdatensatz. Dies ist in mehrfacher Hinsicht vorteilhaft: Es fallen keine Lizenz- oder Beschaffungsgebühren für den proprietären Vergleichsdatensatz an. Da zudem alle verwendeten Technologien frei zugänglich sind, können die vorgestellten Verfahren für jeden beliebigen Ausschnitt in OSM wiederholt werden. Mit dem prototypisch implementierten *iOSMAnalyzer* kann eine große Zahl von Qualitätsuntersuchungen automatisiert durchgeführt werden. Die Ergebnisse werden in Form von Diagrammen, Tabellen und Karten in einem PDF-Dokument zusammengefasst dargestellt und geben einen Überblick über die Qualität der Daten eines ausgewählten Gebiets. Der Tatsache, dass die Anforderungen an die Datenqualität je nach Anwendungsfall variieren, wird durch eine entsprechende Gliederung Rechnung getragen. Es ist zu beachten, dass absolute Aussagen zu bestimmten Qualitätsparametern nicht immer möglich sind. Die Vollständigkeit oder die Positionsgenauigkeit eines OSM Straßennetzwerks kann beispielsweise ohne einen Referenzdatensatz nicht exakt bestimmt werden. Eine Möglichkeit erstere jedoch mit Hilfe der OSM Datenhistorie annähernd zu bestimmen besteht darin, den abnehmenden bzw. stagnierenden Längenzuwachs als Maß für eine potentiell erreichte Vollständigkeit heranzuziehen („*close to completion*“).

Zudem kann die Entwicklung von Attributen an bestimmten Objekten Aufschluss über die attributive Vollständigkeit geben. In den verschiedenen Verfahren kann weiterhin analysiert werden, in welcher Form sich bestimmte Ereignisse in OSM, wie beispielsweise Datenimporte, automatisierte Bots oder Luftbildfreigaben, auf bestimmte Qualitätsparameter auswirken. Insbesondere die Betrachtung der kompletten Datenhistorie ermöglicht detaillierte Aussagen zur Art und Intensität der Weiterbearbeitung von bereits vorhandenen Daten durch die Community. Relative Qualitätsentwicklungen in einem OSM Gebiet können auch dadurch bestimmt werden, dass aktuell gültige Daten mit ihren älteren (oder ersten) Versionen verglichen werden. Aus der Differenz bestimmter Parameter zueinander, beispielsweise der Äquidistanz von Polygonstützpunkten, kann die heutige Qualität der Objekte aus ihrer historischen Entwicklung heraus beurteilt werden. Für künftige Arbeiten mit der OSM Datenhistorie ist die Berücksichtigung von Relationen wichtig. Der verwendete OSM-History-Importer ist gegenwärtig nicht in der Lage Relationen in die Datenbank zu importieren. Neben den verwendeten Analyseverfahren bietet auch der *iOSMANalyzer* weitere Optimierungsmöglichkeiten. Durch den modularen Aufbau des Werkzeuges können weitere Analyseverfahren einfach integriert werden. Wichtig erscheint auch eine engere Verzahnung mit Ansätzen wie der OSMatrix (ROICK ET AL. 2011). Diese Web-basierte Plattform, die ebenfalls zur Analyse intrinsischer OSM Indikatoren konzipiert wurde, könnte durch die zahlreichen Verfahren dieser Arbeit weitere flächendeckende und zeitlich lückenlose Informationen bereitstellen. Hier bestehen Chancen für nachfolgende Aufgaben, um die diversen Ansprüche der Nutzer und die Vor- und Nachteile beider Ansätze zu kombinieren.

## Danksagung

Die Autoren bedanken sich bei allen, die direkt oder indirekt zu diesem Artikel beigetragen haben, insbesondere der GIScience Research Group der Universität Heidelberg und bei allen Mitwirkenden des OSM-Projektes. Diese Arbeit wurde teilweise durch die Klaus-Tschira Stiftung (KTS) Heidelberg finanziert.

## Literatur

- ATHER, A. (2009), A Quality Analysis of OpenStreetMap Data. Dissertation, University College London, Department of Civil, Environmental & Geomatic Engineering. 81 S.
- BRINKHOFF, T. (2008), Geodatenbanksysteme in Theorie und Praxis. Einführung in objektrelationale Geodatenbanken unter besonderer Berücksichtigung von Oracle Spatial. 2. Auflage. Heidelberg: Wichmann.
- GIRRES, J.-F. & TOUYA, G. (2010), Quality Assessment of the French OpenStreetMap Dataset. In: Transactions in GIS, 14 (4), S. 435 - 459.
- GOODCHILD, M. F. & HUNTER, G. J. (1997), A simple positional accuracy measure for linear features. Int. Journal of Geographical Information Science, 11 (3), S. 299 - 306.
- HAKLAY, M. (2008), How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. In: Environment and Planning B: Planning and Design, 37 (4), S. 682 - 703.
- LUDWIG, I. (2010), Abbildung von Straßendaten für Qualitätsuntersuchungen. Ein Vergleich von OpenStreetMap mit Navteq. Diplomarbeit, Rheinische Friedrich-Wilhelms Universität Bonn, Geographisches Institut. 145 S.

- HELBICH, M., AMELUNXEN, C., NEIS, P. & ZIPF, A. (2012), Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. In: STROBEL, J. ET AL.. (Hg.) (2012): *Angewandte Geoinformatik 2012*. Berlin/Offenbach: Herbert Wichmann Verlag, VDE Verlag GmbH.
- ISO19113:2002. [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=26018](http://www.iso.org/iso/catalogue_detail.htm?csnumber=26018) (26.01.2012).
- ISO19114:2003. [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=26019](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=26019) (26.01.2012).
- KOUNADY, O. (2009), Assessing the quality of OpenStreetMap data. MSc dissertation, University College of London, Dep. of Civil, Environ. & Geomatic Engineering. 80 S.
- MOONEY, P., CORCORAN, P. & WINSTANLEY, A. C. (2010), A study of data representation of natural features in openstreetmap. In: *Proceedings of the 6th GIScience International Conference on Geographic Information Science, GIScience 2010*, University of Zürich.
- MOONEY, P., CORCORAN, P. & WINSTANLEY, A. C. (2010a), Towards quality metrics for openstreetmap. In *Proc. of the 18th SIGSPATIAL Int. Conference on Advances in Geographic Information Systems (New York, NY, USA), GIS '10*, ACM, S. 514 - 517.
- MOONEY, P. & CORCORAN, P. (2011), Using OSM for LBS: An Analysis of Changes to Attributes of Spatial Objects. In: GARTNER, G. & ORTAG, F. (2012): *Advances in Location-Based Services Lecture. 8th Int. Symposium on Location-Based Services*, Vienna 2011 Springer-Verlag. Heidelberg, Dordrecht, London, New York. S. 165 - 175.
- NEIS, P., ZIELSTRA, D. & ZIPF, A. (2012), The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007-2011. In: *Future Internet*, 2012 (4), 1-21.
- NEIS, P. & ZIPF, A. (2012), Analyzing the Contributor Activity of a Volunteered Geographic Information Project - The Case of OpenStreetMap. In: *ISPRS International Journal of Geo-Information*, 1, S. 146 - 165.
- NEIS, P., GÖTZ M. & ZIPF, A. (2012), Towards Automatic Vandalism Detection in OpenStreetMap. *ISPRS International Journal of Geo-Information*. 2012; 1(3):315-332.
- ROICK, O., HAGENAUER, J. & ZIPF, A. (2011), OSMatrix - Grid based analysis and visualization of OpenStreetMap. *State of the Map EU*. Wien, Juli 2011.
- VAN EXEL, M., DIAS, E. & FRUITIER, S. (2010), The impact of crowdsourcing on spatial data quality indicators. In *GIScience 2010: Proceedings of the 6th International Conference in Geographical Information Science*. University of Zurich.
- VAN EXEL (2012), A Look At Stale OpenStreetMap Data.. <http://oegeo.wordpress.com/2012/07/03/a-look-at-stale-openstreetmap-data/> (26.01.2012).
- VAN OORT, P. (2006), *Spatial data quality: from description to application*, PhD Thesis, Wageningen: Wageningen Universiteit. 125 S.
- REHRL, K., EDLINGER, K.-M. & FRIEDWAGNER, A., HAHN, B., LANGTHALER, T., WAGNER, A. & WIMMER, M. (2012), Evaluierung von Verkehrsgraphen für die Berechnung von länderübergreifenden Erreichbarkeitspotenzialen am Beispiel von OpenStreetMap. In: STROBEL, J. ET AL. (Hg.) (2012): *Angewandte Geoinformatik 2012*. Berlin/Offenbach: Herbert Wichmann Verlag, VDE Verlag GmbH.
- ZIELSTRA, D. & 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.