

Über den wechselseitigen Einfluss von offenen Geodaten bezogen auf ihre Qualität – ein Fallbeispiel mit Standorten der Landespolizei in Deutschland

Towards the Mutual Influence of Open Geodata on Data Quality – A Case Study with Locations of the State Police in Germany

Pascal Neis¹

¹Hochschule Mainz – University of Applied Sciences · pascal.neis@hs-mainz.de

Zusammenfassung: Für Deutschland sind vermehrt offene Geodaten aus dem Bereich der öffentlichen Verwaltung und anderen gemeinnützigen Projekten erhältlich. Die Datensätze können von unterschiedlicher Qualität sein, wobei eine Kombination vereinzelt nützlich erscheint. Die in diesem Beitrag vorgestellte Methode zur Untersuchung wechselseitiger Einflüsse von Datensätzen hinsichtlich ihrer Qualität zeigt, dass die Methode geeignet ist. Das Ergebnis zeigt am Fallbeispiel der verwendeten Standorte der Landespolizei für Deutschland allerdings, dass die aufgestellte These jedoch nicht bestätigt werden kann. Bei der Untersuchung zeigen Qualitätsmerkmale in Bezug auf die thematische Genauigkeit die größten Unterschiede, wodurch die beiden Datensätze bzw. deren Institutionen nicht voneinander profitieren können.

Schlüsselwörter: Offene Daten, Geodaten, Qualität, Qualitätssicherung

Abstract: *In Germany, open geodata is increasingly available from public administration and other non-profit organizations. These data sets can vary in quality, and sometimes it can be useful to combine them. The method presented in this paper for measuring the mutual influence of data sets on their quality shows that the method is suitable for this purpose, but the results using the available open data sets do not confirm the hypothesis. During the investigation, the greatest differences in quality were found in terms of thematic accuracy, which means that the two data sets or their respective institutions cannot benefit from each other.*

Keywords: *Open data, geodata, quality, quality assurance*

1 Einleitung

Mit dem „Open Data Gesetz“ aus dem Jahr 2017 (vgl. § 12a EGovG, 2021), müssen zumindest Bundesbehörden in Deutschland ihre Daten in einer unbearbeiteten und maschinenlesbaren Form kostenfrei und für jeden zugänglich und nutzbar bereitstellen. In den letzten Jahren gibt es daneben mehrere Novellen in einzelnen Bundesländern, die die „Förderung der elektronischen Verwaltung“ weiter vorantreiben (vgl. z. B. HEGovG, 2018). Dadurch werden inzwischen auch in Deutschland offizielle Geobasisdaten, wie z. B. in Hessen seit dem 1. Februar 2022, öffentlich über das Internet als offene Geodaten zur freien Nutzung und ohne Einschränkungen zur Verfügung gestellt (s. HVBG, 2022). Für die Suche nach verfügbaren Datensätzen des Bundes und der Länder bietet sich das GovData Portal für Deutschland an. Die dort zu findenden Datensätze können sich hierbei durch Kategorie, Herkunft, Datenformat oder auch Lizenzierung unterscheiden. Insbesondere die Lizenzierung hat zumeist je-

doch einen direkten Einfluss auf die Nutzung des Datensatzes (vgl. Khayyat & Bannister, 2015). Als Lizenzform finden sich auf dem GovData Portal Varianten wie Datenlizenz Deutschland (DL-DE, 2023) oder die Open Database License (ODbL, 2023). Neben den vorgestellten verfügbaren Daten von öffentlichen Verwaltungen existieren neben kommerziellen Anbietern auch beispielsweise Projekte im Internet, welche durch meist Freiwillige gesammelte räumliche Informationen (engl. Volunteered Geographic Information (VGI), vgl. Goodchild, 2007) der Allgemeinheit zur Verfügung stellen. Bei dem bekanntesten weltweiten VGI-Projekt handelt es sich um OpenStreetMap (OSM) (vgl. Neis & Zielstra, 2014). Bei der Suche nach verfügbaren raumbezogenen Daten können Anwender von Geoinformationssystemen in Deutschland, je nach Anwendungsfall und Untersuchungsgebiet, auf unterschiedlichen Quellen zurückgreifen. Die Konstellation der Datenlage in Deutschland ist aber im Vergleich mit anderen Ländern, wie z. B. den USA, anders: In den USA stehen viele Geodaten per se frei zur Verfügung und können beispielsweise für Datenimporte in VGI-Projekte genutzt werden (vgl. Zielstra et al., 2013; Witt et al., 2021). Im Unterschied dazu wurden offene Daten, aus den Bereichen der öffentlichen Verwaltungen in Deutschland, noch vor wenigen Jahren nur sehr begrenzt bereitgestellt. Geprägt durch eine sehr aktive Gemeinschaft von Mitwirkenden (vgl. Neis & Zipf, 2012) entstand durch das OSM-Projekt für Deutschland eine alternative und offene Datenquelle mit einer großen Datenvielfalt (Neis & Zielstra, 2014). Mit dieser Studie wird die folgende Hypothese untersucht: *„Offene Daten der kommunalen Verwaltungen haben positive Auswirkungen auf die Datenqualität von anderen offenen Datensätzen und profitieren im Gegenzug von einer Qualitätssicherung“*. Um dies messen zu können, wird eine Methode vorgestellt, die auf Basis einer intrinsischen Bewertung die Entwicklung eines offenen Datensatzes analysiert und diesen mittels einer extrinsischen Qualitätsuntersuchung mit einem offenen Referenzdatensatz vergleicht. Die so gewonnenen Ergebnisse ermöglichen Aussagen über wechselseitige Einflüsse bzgl. der untersuchten Qualitätsmerkmale der herangezogenen Datensätze. Untersucht wird ein Datensatz aus dem OSM-Projekt. Als Referenz wird ein Datensatz der Standorte der Landespolizei vom Bundesamt für Kartographie und Geodäsie (BKG) für Deutschland eingesetzt.

2 Stand der Forschung

Seit Entstehung des VGI-Phänomens steht die zentrale Frage nach der Glaubwürdigkeit (siehe Flanagan & Metzger, 2008) und der Qualitätssicherung (Goodchild & Li, 2012) im Mittelpunkt und ist bis heute Teil vieler wissenschaftlicher Studien (vgl. Neis & Zielstra, 2014; Senaratne et al., 2015; Yan et al., 2020; Hongchao et al., 2022). Die zusammengetragenen Daten des OSM-Projektes waren in den vergangenen 15 Jahren häufig Bestandteil von Vergleichsstudien zur Ermittlung der Datenqualität in Deutschland (Zielstra & Zipf, 2010, Neis et al., 2011), Großbritannien (Haklay, 2010), Frankreich (Girres & Touya, 2010) und einigen weiteren Ländern (z. B. Forghani & Delavar, 2014; Tian et al., 2019). Bei den Untersuchungen kann generell zwischen einer intrinsischen und extrinsischen Vorgehensweise unterschieden werden. Die extrinsische Bewertung bzw. Qualitätsuntersuchung erfolgt mittels eines Referenzdatensatz, welcher i. d. R. als „Goldstandard“ von einem kommerziellen Anbieter erhoben wurde. Um die Qualität von Geodaten im Detail zu bestimmen wird sich meist an den ISO-Normen 19113 „Qualität von Geoinformationen“ und 19114 „Verfahren zur Ermittlung der Geodatenqualität (ISO)“ orientiert. Im Jahr 2013 wurden diese beiden Normen durch die ISO 19157:2013 „Geoinformation – Datenqualität“ ersetzt. Die Datenqualität kann nach der ISO-Norm in folgende fünf Eigenschaften untergliedert werden: Vollständigkeit,

logische Konsistenz, Positionsgenauigkeit, zeitliche und thematische Genauigkeit. Neben der Vorgehensweise mittels eines Referenzdatensatzes, zeigen Barron et al. (2014) wie ein Geodatenatz intrinsisch in Bezug auf dessen Datenqualität untersucht werden kann. Bei der vorgestellten Methode wurden verschiedene Indikatoren und Verfahren eingesetzt, um Daten anhand ihrer historischen Entwicklung sowie in Bezug auf ihre inhärente Struktur beurteilen zu können. Für die Untersuchungen in diesem Beitrag sind insbesondere Indikatoren in Bezug auf die Qualitätsmerkmale von Point of Interest (POI) Objekten und Informationen über Beitragende (vgl. Barron et al., 2014; Touya et al., 2017) bei der intrinsischen Bewertung relevant. Dazu zählen im Einzelnen die historische Entwicklung bzgl. der Summe des POI im Datensatz, die durchschnittliche Anzahl an POI-Attributen und deren Vollständigkeit. Aufseiten der Mitwirkenden ist deren Anzahl, Aktivität und Aktivitätsraum von Bedeutung, die dazu beitragen haben, den Datensatz zu erstellen. Die zuletzt genannten Metainformationen der Beitragenden sind entweder direkt im OSM-Datensatz enthalten oder können entsprechend aus den Daten abgeleitet werden. Neis & Zipf (2012) zeigen dafür verschiedene Methoden auf, wie beispielsweise die Aktivität in Form von „Mapping Days“ und der aktuelle oder übergreifende geografische Aktivitätsraum eines Mitwirkenden an einem VGI-Projekt ermittelt werden kann.

3 Untersuchung der Qualität

Das BKG hat Anfang Dezember 2021 durch eine Kooperation mit dem FOSSGIS e. V. und der Arbeitsgruppe „Open Data“ zwei Datensätze von Standorten der Landespolizei und Gesundheitsämtern in Deutschland veröffentlicht (FOSSGIS OD, 2021). Üblicherweise werden Daten in Deutschland bzw. im GovData Portal häufig unter einer der Varianten der „Datenlizenz Deutschland“ veröffentlicht, was die Übernahme beispielsweise in das OSM-Projekt, durch die Inkompatibilität mit der dortigen ODbL, verhindert. Bei den beiden veröffentlichten Datensätzen mit den erwähnten Standorten wurde eine Ergänzung bei den Nutzungsbedingungen hinzugefügt, die folgendes beinhaltet: „... die Verwendung des Datensatzes für die Pflege und Erweiterung der Daten des OpenStreetMap-Projektes wird ausdrücklich erlaubt“ (DL-DE 2.0 Ergänzung, 2021). Für die Untersuchung im Rahmen dieses Beitrages werden die veröffentlichten Standorte der Landespolizei vom BKG verwendet. Details zu dem Datensatz, wie Beschreibung der Attribute oder Aktualität, können einer Dokumentation im Github Projekt entnommen werden (FOSSGIS OD, 2021). Der zweite Datensatz zur Untersuchung der These stammt wie erwähnt aus dem OSM-Projekt. Für den Export der Daten wurde ein eigener Workflow zur Datenaufbereitung umgesetzt (vgl. Abb. 1).

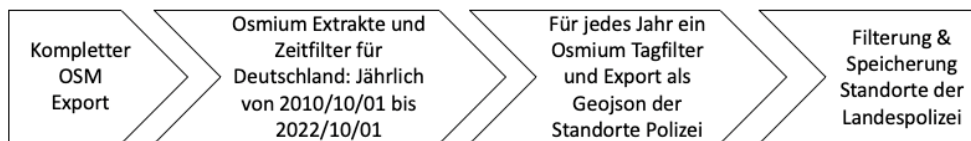


Abb. 1: Workflow der Datenaufbereitung der Standorte der Polizei aus dem OSM-Export

Die Grundlage bildet der komplette OSM-Export (2022), der alle erfassten Features und deren Entstehungshistorie der gesamten Datenbank des OSM-Projektes beinhaltet. Für die Extraktion und Filterung der Polizeistationen für Deutschland kommt das Osmium-Tool (2022) zum Einsatz. So können beginnend für das Jahr 2010 bis 2022 jährliche Datensätze generiert

werden, welche jeweils den Datenstand zum 1. Oktober des Jahres der Polizeistandorte in Deutschland beinhalten. Die Extraktion der Standorte aus dem jeweiligen OSM-Datensatz pro Jahr erfolgt im ersten Schritt über alle Entitäten (Node, Way oder Relation), welche entweder die Eigenschaft *amenity=police* und oder den Schlüssel (Key) *police* beinhalten. Da im OSM-Datenbestand aber auch zusätzlich Bundesbehörden eingetragen werden, müssen diese Entitäten noch aus den Extrakten entfernt werden. Diese Filterung erfolgt entweder über den Schlüssel (Key) *name* oder *operator*, falls einer den Begriff „bund“ im Wert (Value) beinhalten sollte.

3.1 Vollständigkeit und zeitliche Entwicklung/Genauigkeit

Insgesamt beinhaltet der BKG-Datensatz 4.257 Standorte der Landespolizei in Deutschland. Zum Vergleich sind im aktuellen OSM-Datensatz mit der beschriebenen Datenaufbereitung zum Stichtag 1. Oktober 2022 insgesamt 3.680 Standorte zu finden. Auf die Vollständigkeit bezogen bedeutet dies, dass die OSM-Datenbank 14 % weniger Dienststellen vorhält, als der zum Vergleich herangezogene Datensatz des BKG. Über ein Attribut wird weiterhin der Typ der Einrichtung im BKG-Datensatz angegeben. Abbildung 2 (a) zeigt links die Verteilung der insgesamt 3.680 Einrichtungen aus dem OSM-Datensatz, wonach 97 % der Standorte vom Schlüssel und Wert *amenity=police* sind. Laut Beschreibung soll die verwendete Schlüssel-Kombination für eine Polizeidienststelle verwendet werden, „die für den Publikumsverkehr geöffnet ist und bei der man beispielsweise Anzeige erstatten kann“. (vgl. OSM Wiki Tag:amenity 2022). Daneben ist in Abbildung 2 (b) rechts eine aggregierte Verteilung der Typen vom BKG-Datensatz zu sehen (z. B. wurden Außenstellen mit Dienststellen eines Typen addiert). Demnach handelt es sich bei 83 % der Standorte um Einrichtungen der Schutzpolizei. Der Rest verteilt sich auf Kriminal-, Verkehrs-, Wasserschutzpolizei und Sonstiges.

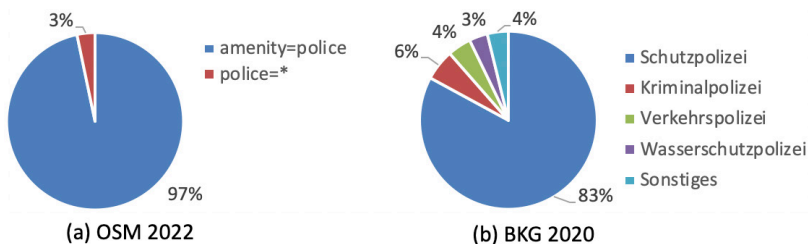


Abb. 2: Verteilung nach Typ der Einrichtung im (a) OSM- und (b) BKG-Datensatz

Zur Untersuchung der aufgestellten These sind Veränderungen zwischen den einzelnen Datenbeständen von OSM über die vergangenen Jahre von Interesse. Abbildung 3 enthält daher die Entwicklung bei der Anzahl der Standorte der Landespolizei des OSM-Projektes in Deutschland für die jeweiligen Jahre ab 2011 im Vergleich zu der Anzahl des BKG-Datensatzes. Für jedes Jahr ist daneben der prozentuale Zuwachs angegeben. Die Jahre zwischen 2011 und 2020 sollen dabei die Entwicklung des Datenbestandes zeigen, um eine Annahme bzgl. der Vollständigkeit durch den Zuwachs bzw. das Stagnieren treffen zu können (vgl. Barron et al., 2014). Hierbei ist für die Jahre 2011 bis 2017 sehr gut ersichtlich, dass die Anzahl stetig zunimmt, bis sie dann ab dem Jahr 2018 stagniert. Dies könnte ein Indikator für eine gute Vollständigkeit sein (vgl. Neis et al., 2011). Für das Jahr 2022 ist wiederum ein Wachstum von 3 % ermittelt worden. Wie bereits erwähnt, ist im Vergleich zwischen den Datensätzen für das Jahr 2022 (OSM) und 2020 (BKG) jedoch weiterhin die Differenz von

ca. 570 Standorten bei der Vollständigkeit ersichtlich. Abbildung 2 enthält zusätzlich eine Farbskala, wann zuletzt eine Entität in der OSM-Datenbank geändert wurde. Für das Jahr 2022 bedeutet dies, dass ca. 55 % der Standorte des OSM-Extraktes in den letzten drei Jahren im Idealfall aktualisiert und/oder geprüft wurden. Allerdings wurden auch ca. 20 % der Entitäten seit dem Jahr 2016 nicht mehr geändert.

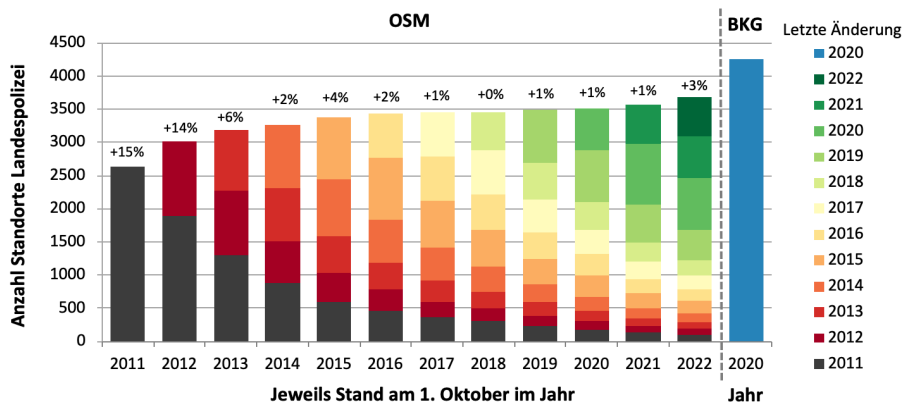


Abb. 3: Entwicklung des OSM-Datenbestandes bzgl. der Anzahl der Landespolizeistandorte für die Jahre 2011 bis 2022 im Vergleich zur Anzahl des BKG-Datensatzes

3.2 Logische Konsistenz und Vollständigkeit der Attribute

Die Untersuchung der logischen Konsistenz, inwieweit die Daten Regeln einer konzeptionellen, logischen und physikalischen Datenstruktur einhalten, erfolgte im Rahmen dieses Beitrags über die Prüfung der jeweiligen Attribute der beiden Datenätze. Der Datensatz der Landespolizei vom BKG besitzt insgesamt 11 Sachattribute. Laut Datenbeschreibung sollte als 12. Attribut noch die Angabe einer „Homepage“ enthalten sein, dieser ist im Datensatz aber nicht auffindbar. Beim OSM-Datensatz ist de facto keine definierte Anzahl von Key-Value-Paaren (Tags) vorgeschrieben. Theoretisch kann ein Standort einer Dienststelle nur über die Angabe *amenity=police* angezeigt werden. Für den Vergleich der logischen Konsistenz wurde die Anzahl bzw. der Schlüssel des Attributes durch den BKG-Datensatz vorgegeben. Die Ergebnisse bzgl. der Angabe der verglichenen Attribute sind in Abbildung 4 zu sehen.

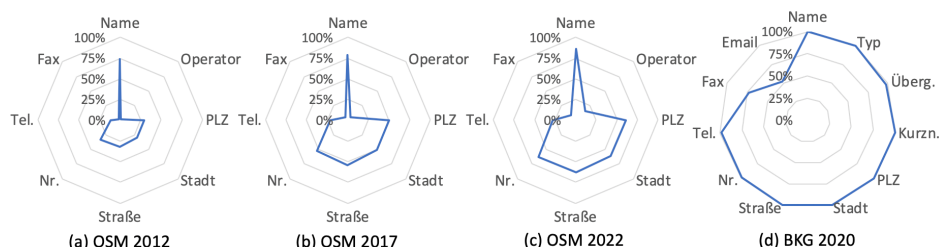


Abb. 4: Entwicklung der logischen Konsistenz der Attribute im OSM-Datenbestand für die Jahre 2012 (a), 2017 (b), 2022 (c) und für den BKG-Datensatz von 2020 (d)

Angefangen mit dem OSM-Datensatz für das Jahr 2012 (a), ist auch in den Folgejahren für 2017 (b) und ebenfalls für das Jahr 2022 (c) ersichtlich, dass die Angabe der Attribute ledig-

lich bei rund 50 % der Standorte gegeben ist. Nur der Name-Key mit rund 80 % bildet eine Ausnahme bei den untersuchten Standorten der Landespolizei von OSM. Im Vergleich sind beim BKG-Datensatz für das Jahr 2020 (d) fast alle Attribute zu 100 % angegeben. Ausnahmen sind nur die beiden Sachattribute Fax (74 %) und Email (53 %).

3.3 Positions- und thematische Genauigkeit

Um die Genauigkeit der Lage (Positionsgenauigkeit) zu vergleichen, wurden für die jeweiligen Standorte der Landespolizei im BKG-Datensatz 100-m-Puffer erzeugt und in der Folge nach vergleichbaren Entitäten im OSM-Datensatz gesucht. Diese Analyse ergab, dass für 86 % der BKG-Standorte im Umfeld ein Standort der Landespolizei im OSM-Datensatz gefunden werden kann. Die Größe des Puffers ergab sich aus der Gegebenheit, dass im BKG-Datensatz die Standorte als Punktgeometrie vorliegen, während diese im OSM-Datensatz als Punkt- oder Flächengeometrie enthalten sein können. Die Prüfung der thematischen Genauigkeit erfolgte über den Namen und den Typ der Einrichtung. Hierfür wurden die beiden Attribute auf Basis des Ergebnisses bzw. durch die hergestellte Verknüpfung der Untersuchung der Positionsgenauigkeit miteinander verglichen. Im Ergebnis konnte der angegebene Name vom BKG-Datensatz zu 34 % identisch in der verknüpften OSM-Entität gefunden werden. Ein weiterer Versuch mit einer Levenshtein-Distanz von bis zu fünf Zeichen ergab 37 % an Übereinstimmung. Beim Vergleich des angegebenen Typs des Standortes der Landespolizei konnte in Summe eine Übereinstimmung von 72 % erreicht werden. Ähnlich der Analyse zur Überprüfung der Vollständigkeit der Attribute zeigt sich auch hier eine mögliche Herausforderung bei der Verwendung des OSM-Datensatzes. Bei der historischen Entwicklung des OSM-Datensatzes, in Bezug auf die thematische Genauigkeit, sind keine Ausfälligkeiten im Vergleich zum Vorjahr oder den letzten Jahren feststellbar: Die Zunahme von Attributwerten erfolgt im Bereich von bis zu 2 % jährlich.

3.4 Beitragende

In der Dokumentation des verwendeten BKG-Datensatzes sind verschiedene Informationen über den Herstellungsprozess, z. B. die teilweise automatisierte und manuelle Datenerhebung, zu finden. Zusätzliche Angaben wer oder welche Abteilung die Daten erhoben haben, sind dort, wie i. d. R. auch in anderen öffentlichen Datensätzen, nicht zu finden. Anders sieht dies für die Standorte des OSM-Datensatzes aus. Hier enthält jede Entität, neben den untersuchten Sachattributen und der Geometrie, weitere Metadaten, wie z. B. die Kennung der Mitwirkenden des Projektes, welche die Entität erstellt oder beispielsweise zuletzt geändert haben. Daraus ergibt sich, dass insgesamt 1.413 registrierte Benutzerkonten bei der Erfassung oder Aktualisierung von den 3.680 Standorten beteiligt waren. Rund 25 % der Beteiligten stehen dabei für einen Anteil von knapp 70 % der Daten. Werden die Benutzerkonten, welche mindestens 10 oder mehr Standorte der Landespolizei im OSM-Datensatz geändert haben, detaillierter und manuell analysiert, fällt auf, dass sich unter diesen 44 Konten vier Accounts für mechanische Änderungen und vier Accounts von bezahlten Beitragenden (vgl. Neis 2019) befinden. Dies kann die Frage bzgl. der Datenqualität der Beitragenden durch fehlendes „Lokales Fachwissen“ aufgreifen (vgl. Neis & Zielstra, 2014). Für die Prüfung, ob ein Benutzerkonto innerhalb einer seiner „lokalen Gebiete“ aktiv ist, eignet sich die Erstellung von Aktivitätsräumen aus den jeweiligen Beiträgen (vgl. Neis & Zipf, 2012). Abbildung 5 enthält zwei Visualisierungen für Deutschland: Auf der linken Seite sind die (a) Aktivitätsräume der Benutzerkonten visualisiert, welche bei den Standorten beteiligt waren. Auf der

rechten Seite wurde ein weiterer Datensatz vom BKG mit den (b) Siedlungsstrukturellen Kreistypen verwendet (vgl. BKG GE250, 2019). Die Intensität der Farbe Grün stellt dabei die Fläche bei (b) von dünn besiedelt, über ländliche hin zu städtischen und kreisfreien Großstädten dar. In den beiden Karten ist daneben ersichtlich, dass keine direkte Korrelation zwischen fehlenden Standorten und Aktivitätsräumen oder siedlungsstrukturellen Gegebenheiten vorliegt. Von den in OSM fehlenden Standorten befinden sich ca. 55 % in den Aktivitätsräumen von Beitragenden, welche zuvor bereits Standorte der Landespolizei bearbeitet haben. Daneben fehlen in OSM ca. 60 % der Standorte in städtischen Kreisen und kreisfreien Großstädten. 37 % der fehlenden Standorte in OSM sind entweder in ländlichen Regionen oder dünn besiedelten Kreisen.

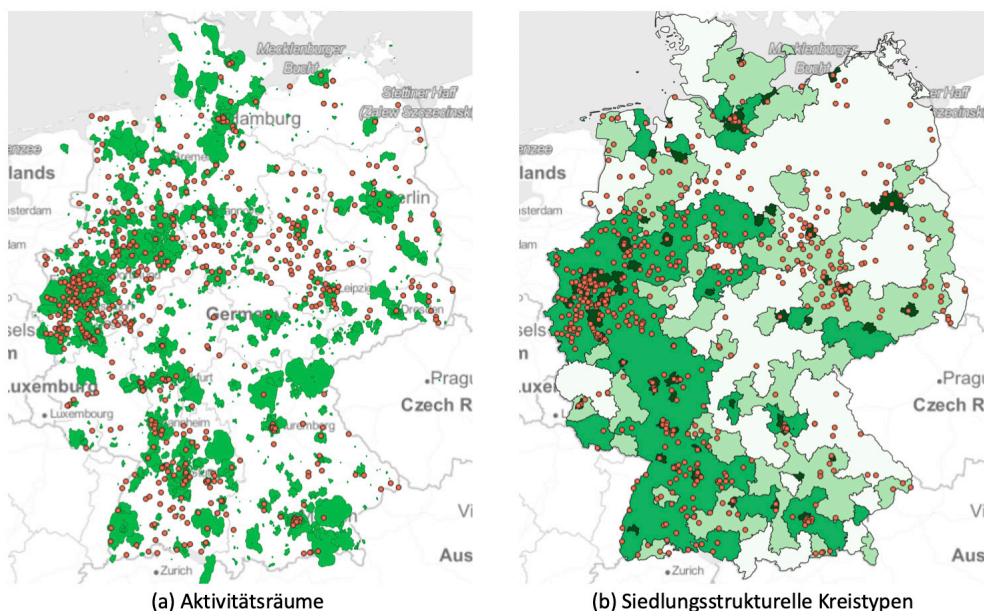


Abb. 5: Fehlende Standorte der Landespolizei (rote Punkte) in OSM zusammen mit den (a) Aktivitätsräumen der Beitragenden und (b) Siedlungsstrukturellen Kreistypen

4 Diskussion

Bei den verschiedenen Analysen dieses Beitrags sind mehrere Besonderheiten und Herausforderungen bei den beiden Datensätzen ersichtlich geworden. Die intrinsische wie auch die extrinsische Untersuchung haben dabei gezeigt, dass beide Datensätze in unterschiedlicher Qualität vorliegen. Für die aufgestellte These bedeutet dies, dass in diesem Fallbeispiel die OSM-Daten nicht quantitativ und qualitativ von dem BKG-Datensatz profitieren konnten. Gerade die Qualitätsmerkmale der Attribute der OSM-Daten kann mögliche Datennutzer:innen vor Herausforderungen stellen, wohingegen im BKG-Datensatz sehr detailliertere Attribute und Standortangaben je Einrichtung enthalten sind. Die beiden Datensätze unterscheiden sich jedoch hinsichtlich der geometrischen Modellierung. Im BKG-Datensatz kann beispielsweise eine Mehrfachnennung einer Punktgeometrie enthalten sein, wenn sich an der geocodierten Adresse mehrere verschiedene Einrichtungen der Landespolizei befinden. In

OSM dagegen kann ein komplettes Areal einer Polizeieinrichtung erstellt worden sein, wobei es keine feinere Untergliederung gab. Bei einigen manuellen Prüfungen von fehlenden Standorten in OSM ist häufiger aufgefallen, dass Gebäude und Adressen von Standorten vorhanden waren, die Verwendung des Schlüssels *amenity=police* oder *police=** aber gefehlt hat, wodurch sie für diese Untersuchung nicht beachtet werden konnten. Im Namen war stattdessen beispielsweise „Polizei“ oder „Präsidium“ enthalten. Bei mehreren Stichproben, von fehlenden Standorten in OSM in sehr ländlichen Gegenden, ist auch aufgefallen, dass es sich dabei um sogenannte „Polizei Regionalbereichsbeamte“, „Polizeiposten“ oder „Bezirksdienste“ handelt. Diese Standorttypen der Landespolizei sind dann i. d. R. nicht permanent besetzt, und sollen beispielsweise für eine „Bürgersprechstunde“ oder für die „polizeiliche Präsenz“ dienen (vgl. Regionalbereichsbeamte, 2022). Ob es sich dann lediglich um einen Büroraum, eine Wohnung oder Haus mit oder ohne von außen ersichtlicher Kennzeichnung handelt, wurde im Rahmen dieser Studie nicht weiterverfolgt. Bezüglich der geocodierten Adresse kam es andererseits vereinzelt zu Ungenauigkeiten im BKG-Datensatz. Bei visuellen Vergleichen ist an dieser Stelle aufgefallen, dass die Punktgeometrie mit ihrer Hausnummer laut anderer Datenquellen einem benachbarten Haus hätte zugewiesen werden müssen. Der BKG-Datensatz wurde wie erwähnt auf Github veröffentlicht um Nutzer:innen eine Möglichkeit anzubieten, dort Fehler in den Daten melden zu können. Bzgl. der aufgestellten These des wechselseitigen Einflusses hätte somit das BKG durch Meldungen im Sinne einer Qualitätssicherung profitieren können. Bis zum 15.03.2023 wurden allerdings keine Korrekturvorschläge auf der vorgesehenen Plattform abgegeben.

5 Fazit und Ausblick

Die Untersuchungen in diesem Beitrag haben somit anhand der genutzten offenen Datensätze vom BKG und dem OSM-Projekt am Beispiel der Standorte der Landespolizei gezeigt, dass die aufgestellte These zumindest mit diesem Fallbeispiel nicht bestätigt werden kann. Bei der Vollständigkeit ist in OSM-Datensatz zwar ein Zuwachs an Informationen erkennbar, logische Konsistenz und thematische Genauigkeit sind aber weiterhin nicht gut. Der OSM-Datensatz konnte damit nicht in dem Maße vom BKG-Datensatz profitieren, obwohl die Informationen hätten in das OSM-Projekt eingepflegt werden können. Umgekehrt hat in diesem Fallbeispiel das BKG keine Fehler oder Korrekturvorschläge auf der dafür vorgesehenen Plattform erhalten. Um die hier aufgestellte These breiter betrachten zu können, sind weitere Fallstudien mit zusätzlichen Untersuchungen notwendig. In einer Folgestudie soll untersucht werden, inwieweit Mitwirkende des OSM-Projektes unterstützt werden können, die fehlenden Standorte und unvollständigen Attributangaben zu vervollständigen. Dies könnte über die Bereitstellung der Ergebnisse dieser Arbeit im Rahmen eines Web-GIS erfolgen. Vielversprechender werden aber Ansätze aus dem Bereich „Gamification“ bzw. „Spatial Gamification“ sein (vgl. Martella et al., 2015; Antoniou & Schlieder, 2014). Hier konnten bereits Keller et al. (2016) am Beispiel von Zebrastreifen in der Schweiz zeigen, wie unter Verwendung von sogenannten „Map Challenges“ Daten zur Verbesserung von OSM angeboten werden könnten. Noch ein Schritt weiter, wäre beispielsweise die Integration der Daten in eine Standortbezogene mobile Anwendung wie „Streetcomplete“ (vgl. Lorei et al., 2019) um Beitragende des OSM-Projektes über ein Scoring-System für das Sammeln von Punkten zusätzlich zum Mitwirken zu motivieren. Neben der Bereitstellung der zu korrigieren Daten wäre dann zu untersuchen, inwieweit dann wechselseitige Einflüsse zwischen den offenen Daten beobachtbar sind.

Literatur

- Antoniou, V., & Schlieder, C. (2014). Participation patterns, VGI and gamification. *AGILE 2014* (pp. 3–6).
- EGovG (2021). *Gesetz zur Förderung der elektronischen Verwaltung* (E-Government-Gesetz – EGovG). https://www.gesetze-im-internet.de/egovg/inhalts_bersicht.html.
- Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877–895.
- BKG GE250 (2019). *Gebietseinheiten 1:250 000 (GE250)*. Retrieved from <https://gdz.bkg.bund.de/index.php/default/gebietseinheiten-1-250-000-ge250.html>.
- DL-DE (2023). *Datenlizenz Deutschland – Namensnennung – Version 2.0*. Retrieved from <https://www.govdata.de/dl-de/by-2-0>.
- DL-DE 2.0 Ergänzung (2021). *Ergänzung der „Datenlizenz Deutschland – Namensnennung – Version 2.0“ für die Nutzung von Daten der Behörden durch das OpenStreetMap Projekt*. Retrieved from https://sg.geodatenzentrum.de/web_public/gdz/lizenz/deu/Datenlizenz_Deutschland_Erg%C3%A4nzungstext_Namensnennung.pdf.
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3), 137–148.
- FOSSGIS OD (2021). *Open Data BKG im Github Repository des FOSSGIS e. V.* Retrieved from <https://github.com/fossgis/open-data/tree/main/bkg>.
- Forghani, M., & Delavar, M. R. (2014). A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information*, 3(2), 750–763.
- Girres, J. F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435–459.
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221.
- Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, (1), 110–120.
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4), 682–703.
- HEGovG (2018). *Hessisches Gesetz zur Förderung der elektronischen Verwaltung (Hessisches E-Government-Gesetz – HEGovG)*. Retrieved from <https://www.rv.hessenrecht.hessen.de/bshe/document/jlr-EGovGHEpP1>.
- Hongchao, F. A. N., Gefei, K. O. N. G., & Anran, Y. A. N. G. (2022). Current status and prospects of research for volunteered geographic information. *Acta Geodaetica et Cartographica Sinica*, 51(7), 1653.
- HVBG (2022). *Einführung von Open Data zum 01.02.2022 – Hessische Verwaltung für Bodenmanagement und Geoinformation (HVBG)*. Retrieved from <https://hvbh.hessen.de/open-data>.
- ISO 19157 (2013). *Geoinformation – Datenqualität / Geographic information – Data quality*. Retrieved from <https://www.iso.org/standard/32575.html>.
- Keller, S., Bühler, S., & Kurath, S. (2016). Erkennung von Fußgängerstreifen aus Orthophotos. *AGIT – Journal für Angewandte Geoinformatik*, (2), 162–166.
- Khayyat, M., & Bannister, F. (2015). Open data licensing: more than meets the eye. *Information Polity*, 20(4), 231–252.

- Lorei, H., Westerholt, R., & Zipf, A. (2019). Characterizing player types in gamified geodata acquisition – An exploratory analysis of StreetComplete. *Proceedings of the Academic Track at State of the Map 2019*.
- Martella, R., Kray, C., & Clementini, E. (2015). A gamification framework for volunteered geographic information. *AGILE 2015* (pp. 73–89). Springer, Cham.
- Neis, P., Zielstra, D., & Zipf, A. (2011). The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(1), 1–21.
- Neis, P., & Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project – The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1, 146–165.
- Neis, P., Zielstra, D., & Zipf, A. (2013). Comparison of volunteered geographic information data contributions and community development for selected world regions. *Future internet*, 5(2), 282–300.
- Neis, P., & Zielstra, D. (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, 6(1), 76–106. Retrieved from doi:10.3390/fi6010076.
- Neis, P. (2019). Untersuchung zum bezahlten und organisierten Mapping im OpenStreetMap Projekt – Zahlen und Fakten? In: *FOSSGIS 2019*.
- ODbl (2023). Open Data Commons Open Database License (ODbL). Retrieved from <https://opendatacommons.org/licenses/odbl/>.
- OSM Datenbank Export (2022). Retrieved from <https://planet.osm.org/pbf/full-history/>.
- OSM Wiki Tag:amenity (2022). Retrieved from <https://wiki.openstreetmap.org/wiki/DE:Tag:amenity>.
- Osmium-Tool (2022). Retrieved from <https://osmcode.org/osmium-tool/>.
- Regionalbereichsbeamte (2022). Polizei. Retrieved from <https://www.arneburg-goldbeck.de/region-verwaltung/polizei-regionalbereichsbeamte/>.
- Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139–167.
- Tian, Y., Zhou, Q., & Fu, X. (2019). An analysis of the evolution, completeness and spatial patterns of OpenStreetMap building data in China. *ISPRS International Journal of Geo-Information*, 8(1), 35
- Touya, G., Antoniou, V., Olteanu-Raimond, A.-M., & Van Damme, M.-D. (2017). Assessing Crowdsourced POI Quality: Combining Methods Based on Reference Data, History, and Spatial Relations. *ISPRS International Journal of Geo-Information*, 6(3), 80. Retrieved from <http://dx.doi.org/10.3390/ijgi6030080>.
- Witt, R., Loos, L., & Zipf, A. (2021). Analysing the Impact of Large Data Imports in OpenStreetMap. *ISPRS International Journal of Geo-Information*, 10(8), 528.
- Zielstra, D., & Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. In: *13th AGILE International Conference on Geographical Information Science* (Vol. 2010, pp. 1–15). Portugal: Guimarães.
- Zielstra, D., Hochmair, H. H., & Neis, P. (2013). Assessing the Effect of Data Imports on the Completeness of OpenStreetMap – A United States Case Study. *Transactions in GIS*, 17(3), 315–334.
- Yan, Y., Feng, C. C., Huang, W., Fan, H., Wang, Y. C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34(9), 1765–1791.