

Zur Generierung eines Melderegisters als Pseudo-Derivat aus frei verfügbaren Informationen

Generating a Pseudo Resident Register by Using Open Data

Dominik Visca¹, Max Hoppe¹, Pascal Neis¹

¹Institut für Raumbezogene Informations- und Messtechnik, Hochschule Mainz – University of Applied Sciences · dominik.visca@hs-mainz.de

Zusammenfassung: Kombination und Verknüpfung offener (Geo-)Daten bieten eine Möglichkeit zur (Teil-)Rekonstruktion von Registerdatensätzen des öffentlichen Sektors. In diesem Beitrag wird eine Methode vorgestellt, die unter Einhaltung datenschutzrechtlicher Bestimmungen ein auf Gebäudeebene georeferenziertes Einwohnermelderegister als Pseudo-Derivat nachbildet. Im Rahmen einer zweckmäßigen Lösung als potenzielle Ergänzung für Vorhaben der Forschung wird somit die zunehmende Relevanz und Diskussion über Datensätze zur Abbildung individueller Dynamiken auf kleinräumiger Ebene aufgegriffen. Potenzial hinsichtlich Aktualität und Belastbarkeit der Daten sind mit Blick auf den Zensus 2022 und hinzukommende offene Daten gegeben.

Schlüsselwörter: Melderegister, Disaggregation, Zensus, Open Data, Registerforschung

Abstract: *With respect to data protection and privacy regulations, combining and linking open (geo-) data offers a possibility for a (partial) reconstruction of a public sector register dataset. Here, a method is proposed that simulates a building-level georeferenced resident register as a pseudo-derivative. Understood as a potential supplement for research projects, the increasing relevance and discussion of datasets for analysing individual dynamics on a small scale is taken up. Further improvements are conceivable with latest census and additional open data.*

Keywords: *Resident register, disaggregation, census data, open data, register-based research*

1 Einleitung

Die Datenbestände des öffentlichen Sektors (Bund, Länder, Gemeinden) sind wichtige Quellen für kleinräumige Auswertungen. Zumeist sind die Daten der statistischen Ämter der Länder und des Bundes jedoch auf Gemeinde- oder einer höheren administrativen Ebene aggregiert. Viele gesellschaftliche Fragestellungen lassen sich nur schwer auf Basis dieser Zahlen beantworten, da eine Folgerung aus Summen und Durchschnittswerten auf individuelle Dynamiken nicht möglich ist (Redaktion, 2020). Nutzung und Verwertbarkeit der Daten sind daher begrenzt. Insbesondere in den Sozialwissenschaften ist der Bedarf an kombinierten Datensätzen für die Modellierung von Forschungsfragen auf möglichst kleinräumiger Ebene, beispielsweise zur Untersuchung von lokalen Ungleichheiten, von Lebensstilen oder der Wohnstandortwahl bestimmter Gruppen, sehr hoch (vgl. Heldt & Heinrich, 2017). Auch unter dem Eindruck der Corona-Pandemie ist in der Folge die Diskussion um eine Öffnung von Registerdaten für die Forschung präsent.

Viele private Unternehmen bieten mikrogeographische Datenvariablen an, die im kommerziellen Bereich unter anderem zur Analyse von Kundenpotenzialen, für zielgruppenspezifische Werbekampagnen oder zur Standort- und Filialnetzplanung genutzt werden. Diese Variablen liegen zum Teil adressengenau, wenigstens aber auf Ebene von Siedlungseinheiten oder

städtischen Quartieren vor. Sie enthalten Angaben zu Haushaltsgrößen und Haushaltstyp, aber auch abgeleitete Kennziffern zur sozialen Schicht, zu Einkommensklassen oder Kaufkraft (Milbert & Fina, 2021). Die Methodik, die von privaten Unternehmen zur Erstellung kostenpflichtiger mikrogeographischer Daten eingesetzt wird, ist allgemein jedoch nicht offengelegt. Für wissenschaftliches Arbeiten bedeutet dies eine Einschränkung hinsichtlich einer nicht durchzuführenden tiefergreifenden Qualitätskontrolle. Benötigt werden daher transparente Methoden, die mit den Anforderungen des Datenschutzes kompatibel und trotzdem flexibel genug für die Modellierung unterschiedlicher Forschungsfragen aber auch sonstiger Anwendungen sind (vgl. Heldt & Heinrich, 2017).

Unter Einhaltung der zu garantierenden Regelungen des Datenschutzes und mit Zugeständnissen an Qualität, Aktualität und Belastbarkeit der Daten können für Beobachtungen unterhalb der Gemeindeebene zweckmäßige Lösungen eine potenzielle Ergänzung für Vorhaben der Forschung bieten. Eine Methodik zur (Teil-)Rekonstruktion eines auf Gebäudeebene georeferenzierten Einwohnermelderegisters aus frei verfügbaren Informationen für die Verwendung im Rahmen eines kleinräumigen Geomonitorings auf Basis demografischer Analysen erörtert dieser Beitrag. Dabei werden die gitterzellenbasierten Ergebnisse der Zensus 2011-Volkszählung durch die Anreicherung aus Daten des OpenStreetMap-Projekts disaggregiert. Über die Zwischenschritte von Gebäuden und dazugehörigen Haushalten werden Einwohner-Datensätze gebildet, welche durch weitere Zensus 2011-Attribute ergänzt werden.

2 Herausforderungen

In Deutschland unterliegen die im Melderegister erfassten Einwohnermeldedaten dem Bundesmeldegesetz (BMG). Gegenwärtig speichern die Meldebehörden nach § 3 des BMGs (2015) 19 Merkmale und 11 erforderliche Hinweise für meldepflichtige Personen. Neben der Adresse enthalten Einwohnermelderegister u. a. Informationen über Geburtsdatum, Geschlecht oder Zuzugs- und Wegzugsdatum. In seiner ursprünglichen Fassung stammt das BMG vom 3. Mai 1933. In Kraft getreten ist es am 1. November 2015 (BGBl. I, 2015). Mit dem Meldewesen machte der Bund von seiner ausschließlichen Gesetzgebungskompetenz Gebrauch, die der Bundesebene durch die Föderalismusreform (I) übertragen wurde. Die Länder haben nur noch Regelungskompetenz, sofern sie dazu im Bundesmeldegesetz oder in einem anderen Bundesgesetz explizit berechtigt sind (Ehmann, 2017). In seinem gesetzlich definierten Beschreibungsumfang hat das Melderegister weder die Wissenschaft noch die Nutzung der Meldedaten für ein kleinräumiges Geomonitoring im Fokus (vgl. Schaffert & Höcht, 2018). Dass jedoch die positiven Effekte der Registerforschung in der Gesetzgebung erkannt sind, zeigt sich etwa im Erwägungsgrund 157 der EU-Datenschutz-Grundverordnung (DSGVO) (2016): „Durch die Verknüpfung von Informationen aus Registern können Forscher neue Erkenntnisse von großem Wert in Bezug auf weitverbreiteten Krankheiten wie Herz-Kreislaufkrankungen, Krebs und Depression erhalten. Durch die Verwendung von Registern können bessere Forschungsergebnisse erzielt werden, da sie auf einen größeren Bevölkerungsanteil gestützt sind. Im Bereich der Sozialwissenschaften ermöglicht die Forschung anhand von Registern es den Forschern, entscheidende Erkenntnisse über den langfristigen Zusammenhang einer Reihe sozialer Umstände zu erlangen, wie Arbeitslosigkeit und Bildung mit anderen Lebensumständen.“

Anhand von drei Beispielen aus unterschiedlichen Bereichen verweisen Schmoigl & König (2021) dahin gehend ganz allgemein auf das Potenzial einer reglementierten Öffnung von Registerdaten für die Wissenschaft. Pajares et al. (2021) untersuchen mit zum Teil ähnlichen Datenquellen eine in Teilen vergleichbare Vorgehensweise wie in der hier vorgestellten Methodik. Im Ergebnis erhalten sie ebenfalls disaggregierte Bevölkerungsdaten auf Gebäudeebene, die sie allerdings im Kontext der Erreichbarkeit von Haushalten und der Bestimmung von Gebäudeeingängen verwenden. Den Wert von Einwohnermeldedaten diskutieren Othen- grafen et al. (2021) im Rahmen der Erforschung von Multilokalität in ländlichen Räumen. Einen Ansatz zur kleinräumigen Bevölkerungsmodellierung für Europa, abgeleitet aus Registerzählungen, auf Rasterbasis zeigen Steinnocher et al. (2011) auf. Im Sinne eines kleinräumigen Monitorings machen schließlich sowohl Schaffert & Höcht (2018) als auch Oku et al. (2016) auf die Möglichkeiten von Einwohnermeldedaten für eine genauere räumliche Betrachtung aufmerksam.

Durch die Zuordnung von Geokoordinaten zu Daten, als Georeferenzierung oder Geokodierung bezeichnet, können Beobachtungen im Raum in Beziehung zueinander gesetzt und konkrete Analysen basierend auf diesem Raumbezug möglich werden, wie beispielsweise die Berechnung von geografischen Distanzen zwischen Punkten oder auch eine präzisere Ermittlung von Versorgungsbedarfen und Flächenverbräuchen (vgl. Meyer et al., 2013, zitiert nach Müller, 2019; Schaffert & Höcht, 2018).

Die im Kontext von Melderegistern hinterlegten Adressen machen Gebäude zunächst referenzierbar. Sie stellen als solche nicht zwingend personenbezogene Daten dar (Müller, 2019). Problematisch kann jedoch die Zuordnung einer Adresse zu einem alleinstehenden Gebäude werden, wenn dieses nur von einer Person bewohnt wird. Die Identität dieser Person kann dann aufgedeckt werden (vgl. Müller, 2019). Sollten geokodierte Adressdaten mit weiteren Informationen (z. B. einfachen Umfragedaten ohne Namensbezug) kombiniert werden, gilt es den Datenschutz besonders zu beachten, da die räumliche Verortung ein Re-Identifikationsrisiko darstellt bzw. eine Deanonymisierung von Personen möglich machen kann (vgl. van der Meer et al., 2018; Müller, 2019). Diese Herausforderung wird auch von den Produzenten der deutschen amtlichen Statistik im Kontext der Integration statistischer und räumliche Daten genannt: Je kleinräumiger statistische Daten angegeben und dargestellt werden, desto relevanter wird das Problem des Aufdeckungsrisikos (vgl. Schnorr-Bäcker & Etienne, 2018). § 16 des Bundesstatistikgesetzes (2016) besagt in diesem Zusammenhang das Einzelangaben über persönliche und sachliche Verhältnisse geheim zu halten sind bzw. eine Re-Identifikation von Individuen verboten ist.

3 Methodik

Das Bundesmeldegesetz (BMG) listet in § 3 die im Melderegister zu führenden Attribute auf. Eine vollständige Rekonstruktion eines Melderegisters durch offene Daten ist nicht möglich, da Attribute wie Namen, Personalausweis-Nummern oder Steuer-IDs aus Gründen des Datenschutzes nicht in frei verfügbaren Datensätzen geführt werden dürfen. Andere Attribute können mit Informationsverlust rekonstruiert werden. Statt des im Melderegister geführten Geburtsdatums enthalten offene Daten z. B. Altersklassen. Im Hinblick auf die Verwendung im Rahmen eines kleinräumigen Geomonitorings auf Basis demografischer Analysen ist eine vollständige Rekonstruktion aber auch nicht notwendig. Als Grundlage für die Teilrekon-

struktion werden in der hier vorgestellten Methodik die offenen Daten der Zensus 2011-Volkszählung (Statistische Ämter des Bundes und der Länder, 2022) sowie die frei nutzbaren Geodaten vom OpenStreetMap-Projekt verwendet (OpenStreetMap, 2022). Tabelle 1 zeigt eine Übersicht der für die demografischen Analysen relevanten und im Melderegister geführten Attribute sowie die Datenquellen als Grundlage der Ableitung.

Tabelle 1: Relevante Melderegisterattribute und ihre Ableitung aus offenen Daten

Attribut	Open-Data-Quelle
Geburtsdatum	Zensus 2011 (Altersklassen)
Derzeitige Anschrift	OpenStreetMap
Angabe, ob Ehegatte oder Lebenspartner existiert	Zensus 2011 (Familientypen)
Angabe, ob minderjährige Kinder existieren	Zensus 2011 (Familientypen)

Kleinräumige, verortete Ergebnisse des Zensus 2011 liegen in tabellarischer Form für die Verknüpfung mit einem geografischen Gittersystem vor. Diese Form der Datenlage wurde erst durch eine 2013 erfolgte Änderung des Bundesstatistikgesetzes (BStatG) möglich und liegt daher nicht für alle im Zensus 2011 erhobenen Merkmale vor (Neutze, 2015). Um eine geografische Zuordnung vornehmen zu können, ist die Verwendung eines passenden Gitter-Datensatzes notwendig. Das Bundesamt für Kartographie und Geodäsie stellt geografische Gitter in verschiedenen räumlichen Auflösungen zum Gebrauch bereit. Maßgeblich für die Verwendung mit den Zensus 2011-Datensätzen ist das 100-m-Gittersystem in der Georeferenzierung ETRS89-LAEA entsprechend der INSPIRE-Durchführungsbestimmung zur Interoperabilität von Geodatenätzen und -diensten (Anhang II, 2.2.1 1089/2010 EG). Nicht alle Zensus-2011-Merkmale liegen in der räumlichen Auflösung des 100-m-Gittersystems vor. Sogenannte „Gitterzellenbasierte Ergebnisse“ stehen auf dem Webportal des Zensus 2011 für die Themen „Bevölkerung“, „Demographie“, „Familien“, „Haushalte“, „Gebäude“ und „Wohnungen“ zur Verfügung. Dabei handelt es sich um CSV-Tabellen, die mittels ID mit den geografischen Gittern verbunden werden können (Statistische Ämter des Bundes und der Länder, 2020).

Eine Übersicht der einzelnen Verarbeitungsschritte zur Generierung eines Melderegisters aus den verwendeten freien Informationen ist in der folgenden Abbildung 1 zu sehen. Das Gebiet, für welches ein abgeleitetes Melderegister erstellt werden soll, kann über ein Polygon als Grenzbereich festgelegt werden. Dieses wird im ersten Schritt mit den vollständigen Gitterzellen der Zensusdaten verschnitten. Die Sachinformationen aus den Zensusdaten werden anschließend mit den Gitterzellen verknüpft. Strukturell unterscheidet sich die Ergebnistabelle „Bevölkerung“ von den anderen. Die Ergebnistabellen der Themen „Demographie“, „Familien“, „Haushalte“, „Gebäude“ und „Wohnungen“ beinhalten für jede vorhandene Merkmalsausprägung jedes Merkmals pro Gitterzelle jeweils eine eigene Zeile. Um einen 1:1-Join zwischen Gitterzellen und Ergebnistabellen zu ermöglichen, sind die Tabellen zu pivotieren. Die so entstandenen Tabellen enthalten für jede Gitterzelle eine Zeile und für jede vorhandene Kombination aus Merkmal und Merkmalsausprägung eine entsprechende Spalte. Kontrastierend dazu beinhaltet die Ergebnistabelle „Bevölkerung“ für jede Gitterzelle eine Spalte mit der Bevölkerungsanzahl der jeweiligen Zelle, sodass keine Vorverarbeitung notwendig ist.

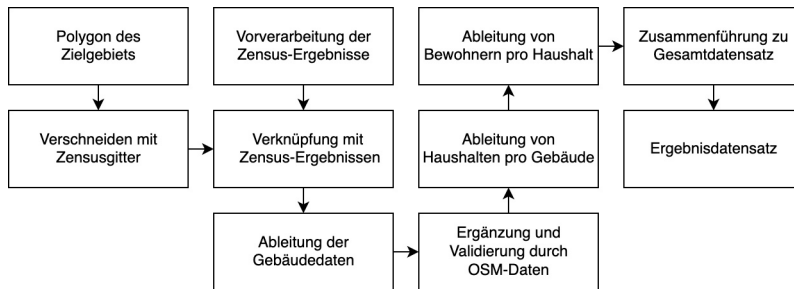


Abb. 1: Schematische Darstellung der einzelnen Verarbeitungsschritte

Anschließend werden Datensätze der gitterzellenbasierten Ergebnisse des Zensus 2011 anhand ihrer Gitter-ID mit den geografischen Gitterzellen verbunden. Zu beachten ist die Verwendung des geografischen Gitters in der korrekten Projektion, da sich die Gitter-ID aus den Koordinaten der jeweiligen Zelle ableitet (BKG, 2020). Eine Zuordnung zu den Zensus 2011-Ergebnissen ist nur mit den geografischen Gittern in der Projektion ETRS89-LAEA möglich. Für die weitere Verarbeitung sind die einzelnen Zellen des geografischen Gitters die maßgebliche räumliche Bezugseinheit. D. h., die folgenden Arbeitsschritte finden individuell für jede Gitterzelle statt. Das Ergebnis jeder Iteration mündet in einem Gesamtdatensatz (s. Abb. 2). Für jede Iteration der nachfolgenden Verarbeitungsschritte wird eine Datenstruktur befüllt. Die Gitter-ID wird als Metadatensatz und räumliche Bezugsinformation der Datenstruktur übergeben. Anschließend wird die Gesamtanzahl der Gebäude und Haushalte aus den verknüpften Tabellen übernommen. Diese dienen als Kontrollfeld.

Als wesentliches Vergleichsmerkmal für den OpenStreetMap-Datensatz wurde das Attribut Gebäudetyp (Größe) der Zensus-Ergebnistabellen herangezogen. Dieses Attribut klassifiziert u. a. ob es sich um Ein- oder Mehrfamilienhäuser handelt. Die aufsummierte Anzahl der Gebäude pro Attribut bildet die Grundgesamtheit der wohnlich genutzten Gebäude in der Gitterzelle. Die einzelnen Gebäude werden mit Adressinformationen aus OpenStreetMap angereichert. Zu diesem Zweck wird das Werkzeug Overpass API verwendet, um die Adressinformationen aus einem räumlich umgrenzten Gebiet, d. h. der Ausdehnung der aktuellen Gitterzelle, zu extrahieren (Olbricht, 2015). Zusätzlich zur räumlichen Eingrenzung erlaubt die Overpass Query Language (Overpass QL) auch eine thematische Auswahl, basierend auf dem Datenmodell und den Tagging-Richtlinien von OpenStreetMap (ebd.). Gebäude werden in OpenStreetMap als *way* vorgehalten, an denen Sachinformationen wie die Adresse oder die Gebäudeart in Form von *tags* geführt wird. Infolge der freien Editiermöglichkeit von OpenStreetMap finden sich auch fehlerhafte oder fehlende Informationen im Datenbestand (Neis & Zielstra, 2014). U. a. werden Adressinformationen nicht am Gebäudepolygon, sondern an einem Punktobjekt innerhalb des Gebäudepolygons geführt. Daher werden zusätzlich auch die *Adresstags* aller *nodes* und *landuse*-Tag der *ways* in der jeweiligen Gitterzelle abgefragt. Datensätze, welche ein *amenity*-Tag oder einen nicht kompatiblen *landuse*-Value (also anders als Wohn- oder Siedlungsfläche) aufweisen, werden herausgefiltert, sodass gewerblich genutzte Gebäude nicht berücksichtigt werden. Das ermöglicht eine genauere Zuordnung zu den im Zensus-Datensatz geführten Gebäuden, die nur Gebäude mit (Teil-)Wohnnutzung enthalten (Statistische Ämter des Bundes und der Länder, 2015).

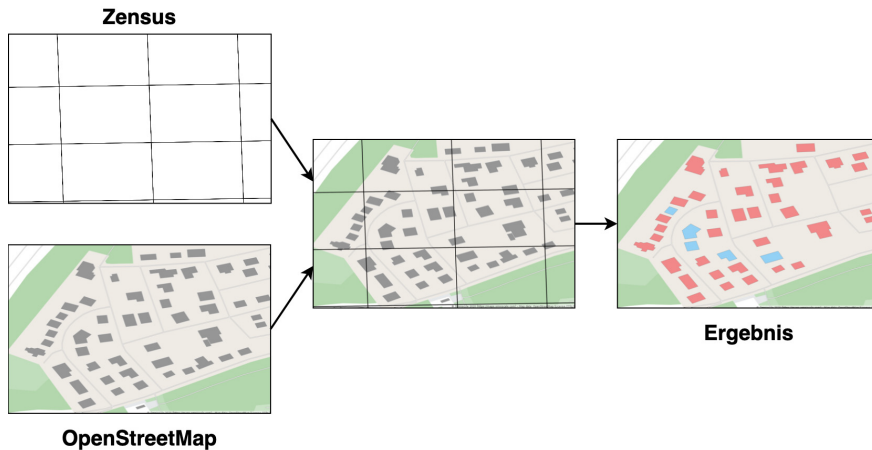


Abb. 2: Ableitung des Pseudo-Derivats (© OpenStreetMap-Mitwirkende)

Die Zuordnung der Adressdaten aus OpenStreetMap und den Gebäudeinformationen der Zensus-Datensätze erfolgt, wenn möglich, über die Gebäudetypen, andernfalls randomisiert. Basierend auf dem Gebäudetyp und dem Zensusdaten-Attribut „Zahl der Wohnungen im Gebäude“ werden Wohnungen bzw. Haushalte auf die Gebäude verteilt. Die Grundgesamtheit der Haushalte basiert auf den aufsummierten Haushalten des Merkmals „Typ des privaten Haushalts“, welchem jeweils passende Sachdaten des Merkmals „Größe des privaten Haushalts“ zugeordnet wurden. Damit werden einem Gebäude Haushalte mit einer plausiblen Anzahl Personen zugewiesen. Eine Differenzierung ob es sich bei den Haushaltsmitgliedern um Erwachsene oder Kinder handelt, wird durch den Haushaltstyp, d. h. „Einpersonenhaushalt“ oder „Paare mit Kind(ern)“, durchgeführt. Für jeden Haushalt werden anschließend entsprechend des Attributs der Haushaltsgröße eine bestimmte Anzahl Personen-Datensätze generiert. Diesen werden passende Altersklassen aus den Zensusdaten zugewiesen. Die abschließende Zellen-Datenstruktur wird nach erfolgreicher Iteration einem Gesamtdatensatz zugeführt.

4 Evaluierung und Diskussion

Die vorgestellte Methodik wurde für das in Abbildung 3 aufgeführte Testgebiet in der Stadt Herborn (Deutschland, Bundesland Hessen) angewendet. Das Testgebiet umfasst eine Fläche von ca. 4,52 km², wobei insgesamt 257 wohnlich genutzte Gitterzellen und 1.992 Häuser abgeleitet und mit für demografische Analyse relevanten Attributen angereichert werden konnten.

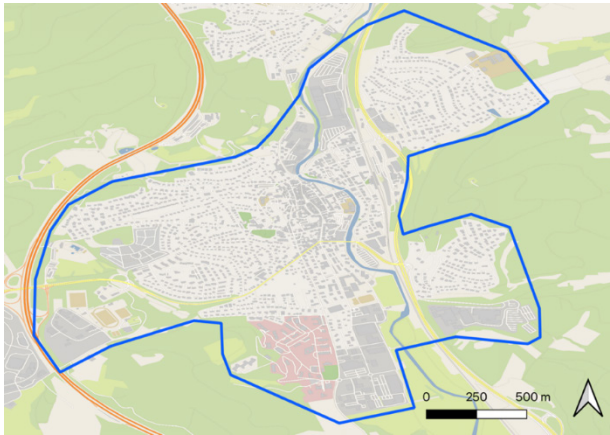


Abb. 3: Überblick über das Testgebiet der Stadt Herborn (© OpenStreetMap-Mitwirkende)

Die Abbildung 4 zeigt links eine Klassifizierung nach Gebäudetyp (Größe) auf Basis der Zensusdaten. Innerhalb der Gitterzelle wurden alle neun, in OpenStreetMap vorhandenen Wohngebäude erkannt, acht von diesen als Einfamilienhaus (in rot markiert) und eines als Mehrfamilienhaus (blau markiert). Ein manueller Abgleich mit den in OpenStreetMap geführten Keys liefert für diesen Bereich sieben Einfamilien- und zwei Mehrfamilienhäuser. Ein Feldvergleich vor Ort kommt ebenfalls zu diesem Ergebnis.



Abb. 4: Klassifizierungsergebnis nach Gebäudeart (Größe) (© OpenStreetMap-Mitwirkende)

Insgesamt verteilen sich in der Gitterzelle 25 Bewohner auf zwölf Haushalte. Konkret drei Haushalte bestehend aus Paaren mit einem Kind, zwei Haushalte mit einem alleinerziehenden Elternteil und einem Kind, zwei Einpersonenhaushalte, sowie fünf Paaren ohne Kinder. Für jeden Bewohner sind die in Tabelle 1 aufgeführten Merkmale vorhanden. Die abgebildete Gitterzelle besticht durch seine relativ homogene Siedlungsstruktur. Es ist anzunehmen, dass dies ein Faktor für die hohe Übereinstimmung zwischen Zensus- und OpenStreetMap-Daten ist. Dies gilt es aber in einer Folgeuntersuchung näher zu untersuchen.

Erwähnenswert ist, dass Teile des Zensusdatensatzes zwei weitere Wohngebäude für diese Gitterzelle vorsehen. Das Anonymisierungsverfahren bei der Aufbereitung des Zensus hat zur Konsequenz, dass Summenbildung über verschiedene Merkmale hinweg nicht zwangsläufig zu identischen Ergebnissen führt (Statistische Ämter des Bundes und Länder, 2015). Die Aufsummierung aller Ausprägungen des Merkmals „Gebäudetyp (Größe)“ ergibt lediglich 1.805 Gebäude. Das ist eine Abweichung von ca. 9 %. Insgesamt weisen 161 der 257 Gitterzellen, also fast 63 % eine Abweichung mit einem Mittelwert von 1,75 Gebäuden und einer Standardabweichung von 0,89, dieser beiden Merkmale auf, die sowohl positiv als auch negativ ausschlagen kann. Diese Abweichungen finden sich auch für andere Merkmale des Zensusdatensatzes und relativieren die theoretisch erreichbare Genauigkeit der Disaggregation. Damit erfüllt sie auch den Zweck des SAFE-Verfahrens zur Anonymisierung (ebd.). Abbildung 4 zeigt rechts ein Resultat des Anonymisierungsverfahrens. Die Gitterzelle enthält laut OpenStreetMap vier Gebäude, der Zensusdatensatz gibt 76 Einwohner an, allerdings fehlen jegliche Angaben zu Gebäuden oder Haushalten. Ein Feldvergleich vor Ort zeigt neben einem Schulgebäude das Vorhandensein von drei mehrstöckigen Häusern mit Senioren- und Seniorinnenresidenzen. Die hier beschriebene Methode kann die 76 Einwohner dieser Gitterzelle ohne Informationen zu Gebäuden und Haushalten bisher nicht berücksichtigen.

Weiterhin korreliert das Ergebnis der vorgestellten Methodik mit der Adressvollständigkeit in OpenStreetMap. Das ausgewählte Testgebiet weist mit ca. 71 % eine mittlere bis gute Vollständigkeit im Verhältnis zum Adressdatensatz des amtlichen Liegenschaftskatasters (ALKIS) auf. Für Bereiche mit niedriger Adressvollständigkeit ist eine gebäudebezogene Genauigkeit nur in geringem Ausmaß zu erwarten. Ferner wirkt sich die Aktualität der Datensätze ebenfalls auf das Ergebnis aus. Während der Zensusdatensatz den Stand des Jahres 2011 wiedergibt, beinhaltet OpenStreetMap tendenziell einen laufend aktualisierten Datenbestand. Je nach städtebaulicher Dynamik und Bevölkerungsfuktuation ist mit kleineren oder größeren Abweichungen zu rechnen. Daneben muss auch das angedachte Anwendungsgebiet berücksichtigt werden. Beides gilt es wie erwähnt in einer Folgeuntersuchung näher zu untersuchen.

5 Fazit und Ausblick

In diesem Beitrag wird eine Methode vorgestellt, die auf Basis frei verfügbarer räumlicher und öffentlich-statistischer Daten ein auf Gebäudeebene georeferenziertes Einwohnermelderegister als Pseudo-Derivat generiert. Eine vollständige Rekonstruktion und Disaggregation der Zensusdaten ist aufgrund der Anonymisierungsverfahren nicht möglich, für die Forschungspraxis aber i. d. R. auch nicht notwendig. Vorliegend wurden Adresszuordnungen zu demografischen Merkmalen mit plausiblen Altersklassen getroffen. Die Genauigkeit der Zuordnung kann unter Berücksichtigung weiterer Zensusdaten-Merkmale und OpenStreetMap-Tags sowie ggf. zusätzlicher offener Datensätze weiter erhöht werden. Ferner bietet der Zensusdatensatz mit Merkmalen wie dem Geschlecht, Religionszugehörigkeit oder Staatsangehörigkeit weitere Melderegisterattribute. Die Verwendungsmöglichkeiten des Pseudo-Derivats kann damit zusätzlich gesteigert werden.

Die dem SAFE-Verfahren zugrunde liegende, fehlende interne Konsistenz der Zensusdaten sowie die Datenqualität von OpenStreetMap macht eine folgende umfänglichere Evaluierung der Qualität der durch die hier aufgeführte Methode erzeugten Datensatzes notwendig.

Mit Blick auf den Zensus 2022 kann die hier vorgestellte Methode auch in Zukunft einfach zu erzeugende Datensätze für die Forschung bereitstellen ohne dabei datenschutzrechtliche Fragestellung aufzuwerfen. Hinsichtlich der Anreicherung bzw. der Kombination von Daten lässt weiterhin die im letzten Jahr vorgelegte Open-Data-Strategie der Bundesregierung (2021) darauf hoffen, dass in den kommenden Jahren nicht nur weitere Datensätze öffentlicher Verwaltungen bereitgestellt werden, sondern auch die Qualität der Daten stetig steigt. Zugleich ist eine verstärkte Verfügbarkeit von bisher nicht zugänglichen Daten aus Wirtschaft, Wissenschaft und Zivilgesellschaft zu erwarten.

Danksagung

Die Arbeiten an diesem Beitrag entstanden im Rahmen des Projekts „Raumintelligenz für die integrierte Versorgung von Seniorinnen und Senioren in ländlichen Quartieren (RAFVI-NIERT)“, das von der Carl-Zeiss-Stiftung im Programm „Transfer – Intelligente Lösungen für eine älter werdende Gesellschaft“ gefördert wird.

Literatur

- Bundesamt für Kartographie und Geodäsie (2020). *Dokumentation. Geographische Gitter für Deutschland. GeoGitter*. Retrieved Jan 24, 2022, from https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/geogitter.pdf.
- Bundesministerium des Innern, für Bau und Heimat (2021). Open-Data-Strategie der Bundesregierung. Retrieved Jan 25, 2022, from https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/moderne-verwaltung/open-data-strategie-der-bundesregierung.pdf?__blob=publicationFile&v=4.
- Ehmann, E. (2017). *Mit Meldedaten richtig umgehen* (3rd Ed.). Stuttgart: Boorberg.
- Europäische Kommission (2010). *Verordnung (EG) Nr. 1089/2010 der Kommission vom 23. November 2010 zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich der Interoperabilität von Geodatenätzen und -diensten*.
- Heldt, B., & Heinrichs, D. (2017). Nutzung des Zensus 2011 zur Modellierung der Wohnstandortwahl von Haushalten unter Geheimhaltungsbedingungen. *Zeitschrift für amtliche Statistik Berlin Brandenburg*, 29–33.
- König, T., & Schmoigl, L. (2020). Erfolgreiche Registerforschung in Österreich. Welchen Mehrwert generiert die reglementierte Öffnung von Registerdaten für die wissenschaftliche Forschung? Eine Darstellung anhand von drei Beispielen. *Österreichisches Institut für Wirtschaftsforschung (WIFO) und Institut für höhere Studien (IHS)* (pp. 1–16).
- Milbert, A., & Fina, S. (2021). Methoden der Kleinstadtforschung: Definitionen, Daten und Raumanalysen. In: A. Steinführer, L. Porsche & M. Sondermann (Eds.), *Kompendium Kleinstadtforschung*. Hannover: Forschungsberichte der ARL 16.
- Müller, S. (2019). Räumliche Verknüpfung georeferenzierter Umfragedaten mit Geodaten: Chancen, Herausforderungen und praktische Empfehlungen. In: U. Jensen, S. Netscher & K. Weller (Eds.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten* (pp. 211–229). Barbara Budrich.

- Neis, P., & Zielstra, D. (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, 2014, 6, 76–106.
- Neutze, M. (2015). Gitterbasierte Auswertungen des Zensus 2011. *Stadtforschung und Statistik*, 2, 64–67.
- Oku, N., Nonaka, Y., & Ito, Y. (2016). A Discussion about Effective Ways of Basic Resident Register on GIS. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B4, 175–179, <https://doi.org/10.5194/isprs-archives-XLI-B4-175-2016>, 2016.
- Olbricht, R. M. (2015). Data Retrieval for Small Spatial Regions in OpenStreetMap. In: J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.). *OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography* (pp. 101–122). Heidelberg: Springer).
- OpenStreetMap (2022). *OpenStreetMap*. Retrieved Jan 30, 2022, from <https://www.openstreetmap.org>
- Othengrafen, F., Linda L., & Greinke, L. (Eds.) (2021). *Temporäre An- und Abwesenheiten in ländlichen Räumen: Auswirkungen multilokaler Lebensweisen auf Land und Gesellschaft*. Wiesbaden: Springer Fachmedien.
- Pajares, E., Muñoz Nieto, R., Meng, L., & Wulforst, G. (2021). Population Disaggregation on the Building Level Based on Outdated Census Data. *ISPRS Int. J. Geo-Inf.* 2021, 10, 662. Retrieved Apr 04, 2022, from https://mdpi-res.com/d_attachment/ijgi/ijgi-10-00662/article_deploy/ijgi-10-00662-v2.pdf.
- Redaktion, W. (2020). Registerforschung: Chancen, Risiken und Herausforderungen. *Wirtschaft und Gesellschaft*, 46(3), 315–328.
- Schaffert, M., & Höcht, V. (2018). Geokodierte Meldedaten als Basis bedarfsgerechter Planungen in ländlichen Gemeinden und Regionen. *Raumforschung und Raumordnung* 76(5), 421–35.
- Schnorr-Bäcker, S., & Etienne, M. (2018). Herausforderungen und mögliche Lösungsansätze bei der Kombination statistischer und räumlicher Daten aus Sicht der Bundesstatistik. *Stadtforschung und Statistik: Zeitschrift des Verbandes Deutscher Städtestatistiker*, 3(1), 63–69.
- Statistische Ämter des Bundes und der Länder (2015). *Zensus 2011. Methoden und Verfahren*. Retrieved Jan 22, 2022, from https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf?jsessionid=48260D25A514027445F421D903862F47.live422?__blob=publicationFile&v=2.
- Statistische Ämter des Bundes und der Länder (2020). *Ergebnisse des Zensus 2011 zum Download – erweitert*. Retrieved Jan 24, 2022, from <https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html>.
- Statistische Ämter des Bundes und der Länder (2022). *Zensus 2011*. Retrieved Jan 30, 2022, from <https://www.zensus2011.de>.
- Steinnocher, K., Köstl, M., & Weichselbaum, J. (2011). Kleinräumige Bevölkerungsmodellierung für Europa: Räumliche Disaggregation auf Basis des Versiegelungsgrades. In: J. Strobl, T. Blaschke & G. Griesebner (Eds), *Angewandte Geoinformatik* (pp. 513–520). Berlin/Offenbach: Wichmann.
- Van Der Meer, M., Meissner, F., Merten, M., & Münsterlein, D. (2018). Entwicklung und Potentiale digitaler Raumforschung. Ethische Fragestellungen und Impulse für die Hochschullehre. *RaumPlanung*, 196(2/3), 20–27.