

Ableitung demographischer Indikatoren aus offenen Datenquellen

Dominik Visca, Max Hoppe, Pascal Neis

Zusammenfassung

Die zunehmende Verfügbarkeit von offenen (Geo-)Daten aus Beständen des öffentlichen Sektors ermöglicht die Generierung neuer Datensätze. Dieser Beitrag geht auf den Bedarf an hoch aufgelösten Daten ein, auf deren Basis kleinräumige Dynamiken abgebildet und ausgewertet werden können. Daraus lassen sich z. B. demographische Indikatoren aus offenen Datenquellen ableiten. Unter Einhaltung datenschutzrechtlicher Bestimmungen kann ein auf Gebäudeebene georeferenzierter Datensatz erstellt werden, der, trotz Abstrichen bezüglich Qualität, Aktualität und Belastbarkeit der Daten, als zweckmäßige Lösung bzw. Ergänzung für verschiedene Forschungsfragen dienen kann.

Schlagwörter: Demographie, Open Data, Disaggregation, Zensus, OpenStreetMap

1 Einführung

Die Öffnung und Verknüpfung von Datenbeständen des öffentlichen Sektors hin zu standardisierten, offenen (Geo-)Daten trägt zu einer transparenteren Aufgabenerfüllung staatlichen Handelns bei. In der Folge können Akteure aus unterschiedlichen Bereichen, beispielsweise aus Wissenschaft oder Wirtschaft, Wissen generieren, um Potenziale zu aktivieren und Fragen von großer gesellschaftlicher Tragweite zu beantworten (u. a. von Lucke, Gollasch 2022; Siebenlist, Mainka 2019).

Demographische Transformationsprozesse sind Beschäftigungsfelder verschiedenster Disziplinen – z. B. der Raumplanung, der Geografie oder der Soziologie. Um diesbezüglich aussagekräftige Indikatoren ableiten zu können, werden feingranulare Geodaten-sätze benötigt. Bestimmungen des Datenschutzes und divergierende Rahmenbedingungen führen jedoch dazu, dass Informationen, insbesondere mit Blick auf Kommunen im ländlichen Raum, zumeist nur auf Ebene der Gesamtgemeinde zugänglich sind.

Mit Methoden der Geoinformatik und im Rückgriff auf frei verfügbare Daten, lassen sich Datensätze generieren, die, unter Beachtung des Datenschutzes, Auswertungen auf kleinräumiger Ebene zulassen. Dafür werden die im 100-m-Gittersystem bereitgestellten Ergebnisse der Zensus-2011-Volkszählung disaggregiert, um auf Basis von OpenStreetMap-Daten einen hochauflösenden Gebäudegeometriedatensatz zu erzeugen. Aufgrund der den Projekten im Kontext von Volunteered geographic information (VGI) inhärenten Problemen bzgl. Aktualität, Richtigkeit, Vollständigkeit sowie dem

SAFE-Anonymisierungsverfahren und dem Erhebungszeitraum der Zensusdaten nimmt diese Vorgehensweise Einschränkungen hinsichtlich Anwendbarkeit und Verwertbarkeit in Kauf (Statistische Ämter des Bundes und der Länder 2014). Analysen und Ergebnisse sollten daher vor allem als zweckmäßiges Mittel bzw. potentielle Ergänzung für Forschungsvorhaben verstanden werden.

2 Hintergrund und Herausforderungen

Für zahlreiche Fragestellungen der vergleichenden Raum- und Stadtbeobachtung besteht ein großes Interesse am Zugang zu möglichst kleinräumigen Daten und Indikatoren. Kleinräumig bedeutet in diesem Zusammenhang zunächst die Abgrenzung von Teilräumen unterhalb der Gesamtadministration (u. a. Großmann et al. 2021). Viele Autoren verweisen auf Daten aus öffentlichen Registern, z. B. diskutieren Dittrich-Wesbuer & Sturm (2020) im Rahmen von Multilokalität über Melderegister; Schaffert & Höcht (2018) und Oku et al. (2016) machen darauf aufmerksam, dass in Registern hinterlegte Adressen als indirekte Geokoordinaten Gebäude referenzierbar sind. Ausgehend von der Gebäudeebene können in der Folge Geodatenätze erzeugt werden, die sich auf unterschiedlichste Flächenbezüge aggregieren lassen.

In der Praxis erschweren Datenschutz sowie organisatorische Hürden und personelle Aufwendungen jedoch die Bereitstellung von (Geo-)Datensätzen unterhalb der Gemeindeebene. Porsche et al. (2021) benennen hierfür das Fehlen einer leistungsfähigen Kommunalstatistik als Kernproblem: Kleinstädte bzw. kleine Kommunen haben in der Regel keine eigene Statistikstelle bzw. genügend Personal, um Daten gezielt aufzubereiten. Dies spiegelt sich in Versorgung und Tiefe wider: Da viele Daten, wenn überhaupt, nur auf Gemeindeebene verfügbar sind, ist eine Differenzierung beispielsweise von demographischen Entwicklungspfaden nach Ortsteilen nur in Ausnahmefällen möglich bzw. nur dann, wenn Informationen tatsächlich untergemeindlich zur Verfügung gestellt werden (Porsche et al. 2021). Für amtliche statistische Daten sind Bundesstatistikgesetz (BstatG¹) und Datenschutzgrundverordnung (DSGVO²) maßgebliche rechtliche Grundlagen. Bei Geodaten stellt sich zudem immer die Frage, ob und inwieweit diese vollends als personenbezogene bzw. personenbeziehbare Daten betrachtet werden können, oder ob sie lediglich als datenschutzrechtlich unerhebliche Sachdaten zu verstehen sind (Kriesten 2021). Werden beispielsweise geokodierte Adressdaten mit zusätzlichen Informationen kombiniert, kann über die räumliche Verortung ein Re-Identifikationsrisiko entstehen bzw. eine Deanonymisierung von Personen möglich werden.

Viele Privatunternehmen nutzen diese Informationslücke und bieten kostenpflichtige mikrogeographische Kennziffern an. Die Methodik, mit der Unternehmen solche Daten

¹ https://www.gesetze-im-internet.de/bstatg_1987/

² <https://dsgvo-gesetz.de/>

erstellen, wird zumeist jedoch nicht offengelegt. Für Analysen auf Basis dieser Daten ist eine tiefgreifende Qualitätskontrolle folglich nicht möglich (Milbert, Fina 2021). Es besteht somit ein Bedarf an Methoden, die unter Einhaltung des Datenschutzes ausreichend flexibel und transparent sind, um eine kleinräumige Modellierung hinsichtlich verschiedener Fragestellungen der Raum- und Stadtbeobachtung zu erlauben.

3 Methodik

Zur Ableitung von demographischen Indikatoren aus offenen Daten werden die Ergebnisse der Zensus 2011-Volkszählung (Statistische Ämter des Bundes und der Länder 2022) sowie die frei nutzbaren Geodaten vom OpenStreetMap-Projekt genutzt (OpenStreetMap 2022). Tabelle 1 zeigt eine Übersicht möglicher demographischer Merkmale, die aus den Zensusdaten abgeleitet und später zu aussagekräftigeren Indikatoren verarbeitet werden können.

Tab. 1: Mögliche Demographie-Merkmale und die jeweiligen Zensus-Datenquellen

Demographisches Merkmal	Zensus-Datenquelle
Alter	Bevölkerung je Hektar
Anzahl Kinder	Familie je Hektar
Eigentumsverhältnis der Wohnung/des Gebäudes	Wohnungen und Gebäude je Hektar
Familienstand	Bevölkerung je Hektar
Geschlecht	Bevölkerung je Hektar
Geburtsland	Bevölkerung je Hektar
Staatsangehörigkeit	Bevölkerung je Hektar
Religion	Bevölkerung je Hektar
Wohnfläche	Wohnungen und Gebäude je Hektar

Während die Zensusergebnisse als CSV-Tabellen zur Verknüpfung mittels ID mit einem geografischen 100-m-Gittersystem verfügbar und visualisierbar sind, werden die demographischen Merkmale in Anlehnung an Visca et al. (2022) zunächst auf Wohngäudeebene disaggregiert (Statistische Ämter des Bundes und der Länder 2020). Eine Abbildung auf dieser Ebene erlaubt eine Aggregation auf unterschiedliche Flächeneinheiten, um auf verschiedene Ansprüche hinsichtlich kleinräumiger Auswertungen möglichst dynamisch reagieren zu können.

Abbildung 1 zeigt eine Übersicht der jeweiligen Verarbeitungsschritte zur Erstellung eines Datensatzes auf Grundlage der frei verfügbaren OSM-Daten. Für die Ableitung von demografischen Merkmalen für ein bestimmtes Gebiet kann ein Polygon als Grenzbereich zugewiesen werden.

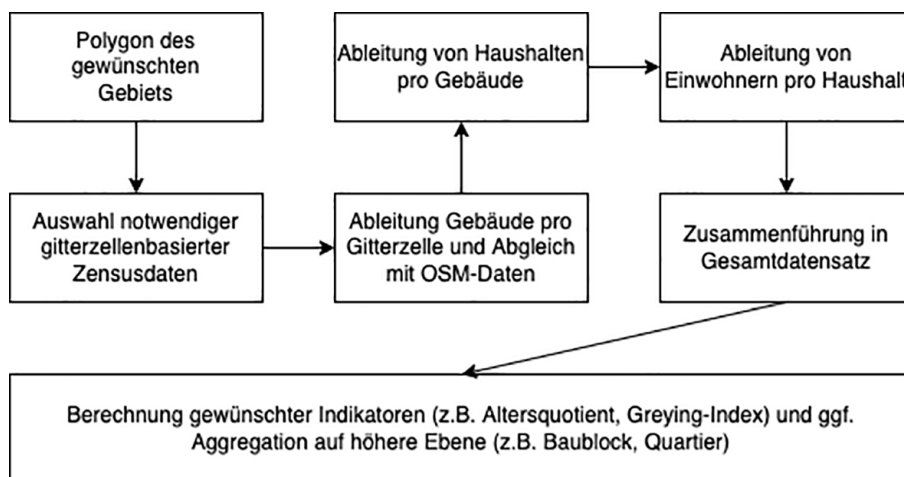


Abb. 1: Übersicht der Verarbeitungsschritte zur Berechnung aggregierter Indikatoren (Quelle: eigene Darstellung nach Visca et al. 2022)

Zentrales Attribut zur Verknüpfung mit dem OpenStreetMap-Datensatz ist das Attribut „Gebäudetyp (Größe)“ der Zensus-Ergebnistabellen. Mit diesem Attribut lässt sich aufschlüsseln, ob beispielsweise ein Ein- oder Mehrfamilienhaus vorliegt. Die Grundgesamtheit der wohnlich genutzten Gebäude einer Gitterzelle resultiert aus der aufsummierten Menge an Gebäuden pro Attributsausprägung. Mit dem Werkzeug Overpass API³ können aus den OpenStreetMap-Daten für ein räumlich definiertes Gebiet Adressinformationen entnommen werden (Olbricht 2015). Darüber hinaus kann mit der Overpass Query Language (Overpass QL) auch eine inhaltliche Auswahl auf Grundlage des Datenmodelles und der Tagging-Richtlinien von OpenStreetMap vorgenommen werden. Weil nach der Erstellung eines Benutzerkontos Mitwirkende direkt Änderungen vornehmen können, finden sich in den Datenbeständen des OpenStreetMap Projektes auch falsche oder irrtümliche Informationen (Neis, Zielstra 2014). Adressinformationen werden nicht nur am Polygon des Gebäudes vorgehalten, sondern können auch an einem Punktobjekt inmitten des Gebäudepolygons vorliegen. Eine entsprechende Abfrage muss dies berücksichtigen. Werden Datensätze entdeckt, die einen *amenity*-Tag oder einen nicht verwendbaren *landuse*-Wert beinhalten, werden diese selektiert. So werden ausschließlich Wohn- oder Siedlungsfläche einbezogen, da die im Zensus-Datensatz enthaltenen Gebäude nur Gebäude mit (Teil-)Wohnnutzung umfassen (Statistische Ämter des Bundes und der Länder 2015).

Sofern machbar, wird die Zuordnung der Adressdaten aus OpenStreetMap und den Gebäudeinformationen der Zensus-Datensätze über die eindeutige Zuordnung von Gebäudetypen oder einem iterativ durchgeführten Ausschlussverfahren vollzogen, respektive durch eine zufallsbedingte Verteilung. Die Zuteilung der Wohnungen bzw. Haushalte auf die Gebäude erfolgt auf Grundlage des Gebäudetyps und dem Zensusdaten-Attribut

³ <https://dev.overpass-api.de/overpass-doc/de/>

„Zahl der Wohnungen im Gebäude“. Die Grundgesamtheit der Haushalte kommt durch die aufsummierten Haushalte des Attributs „Größe des privaten Haushalts“ zustande. Diesem wurden im Einzelnen passende Sachinformationen des Merkmals „Größe des privaten Haushalts“ zugeteilt. Dadurch sind Gebäuden jeweils Haushalte mit einer adäquaten Anzahl an Personen zugeordnet. Im Anschluss wird für jeden Haushalt gemäß dem Attribut der Haushaltsgröße eine verbindliche Anzahl an Personen-Datensätzen erstellt. Auch diese Ableitungen werden entsprechend eindeutiger Zuordnungen und Ausschlussverfahren durchgeführt. Schlussendlich wird die generierte Zellen-Datenstruktur nach gelungener Iteration einem Gesamtdatensatz übergeben, wo sie anschließend auf Basis der georeferenzierten Gebäudeadresse zu beliebigen räumlichen Bezugsebenen wie Baublöcke oder Quartiere aggregiert werden können.

4 Ergebnisse und Diskussion

Die im vorherigen Kapitel aufgezeigte Methodik wurde beispielhaft für die Stadt Herborn (Bundesland Hessen) angewendet. Hierfür wurden für eine Fläche von ca. 4,52 km² insgesamt 257 wohnlich genutzte Gitterzellen und 1 992 Häuser abgeleitet und mit relevanten demografischen Attributen angereichert. Durch das Anonymisierungsverfahren kommt es bei der Summenbildung der unterschiedlichen Merkmale der den Zensusdaten zu Grunde liegenden Informationen zu differenten Ergebnissen (Statistische Ämter des Bundes und Länder 2015). Bemerkbar macht sich dies bei der Aufsummierung der Gesamtausprägungen des Merkmals „Gebäudetyp (Größe)“ mit nur 1 805 Gebäuden. Dies ist eine Differenz von ca. 9 %. Insgesamt lassen sich für 161 der 257 Gitterzellen Abweichungen (63 %), mit einem Mittelwert von 1,75 Gebäuden und einer Standardabweichung von 0,89 der jeweiligen Merkmale ausmachen, die sich sowohl positiv als auch negativ äußern können. Auch in anderen Merkmalen des Zensusdatensatzes sind Abweichungen feststellbar. Die potentiell realisierbare Verlässlichkeit der Disaggregation wird somit vermindert, erfüllt zugleich jedoch das Ziel des SAFE-Verfahrens zur Anonymisierung (ebd.).

Auch korreliert das Resultat der Methodik mit der Adressvollständigkeit in OpenStreet-Map. Für das Testgebiet ist mit 71 % eine mittlere bis gute Vollständigkeit im Vergleich zum Verhältnis des Adressdatensatzes des amtlichen Liegenschaftskatasters (ALKIS) ermittelt worden. Dort wo eine niedrigere Adressvollständigkeit vorliegt, kann eine gebäudebezogene Genauigkeit nur begrenzt vorausgesetzt werden. Ein weiterer wichtiger Faktor ist die Aktualität der Datensätze. Der Zensusdatensatz gibt den Stand von 2011 wieder. OpenStreetMap bietet laufend aktualisierte Daten an. Abweichungen stehen im Zusammenhang mit städtebaulicher Dynamik und Bevölkerungsbewegungen – je nach Ausmaß fällt die Abweichung stärker oder weniger stark aus. Auch ist das vorgesehene Anwendungsgebiet zu beachten. In einer Folgeuntersuchung sollte auf beides eingegangen werden.

Die generierten Merkmale können abhängig von der Fragestellung zu aussagekräftigen Indikatoren wie z. B. Alters- und Jugendquotient oder durchschnittliche Haushaltsgröße weiterverarbeitet werden. Abbildung 2 zeigt beispielhaft den Anteil Kinder pro Gebäude (in Prozent) innerhalb eines ausgewählten Areals. Die Angaben bewegen sich zwischen keinem ermittelbaren (schwarz) und bis zu ca. 83 % (hellgrün) Anteil an Kindern aller Einwohner eines Gebäudes. Für das Gebiet lässt sich festhalten, dass die Siedlungsstruktur vergleichsweise homogen ist. Möglicherweise ist dies ein Faktor für die ausgeprägte Ähnlichkeit zwischen Zensus- und OpenStreetMap-Daten.

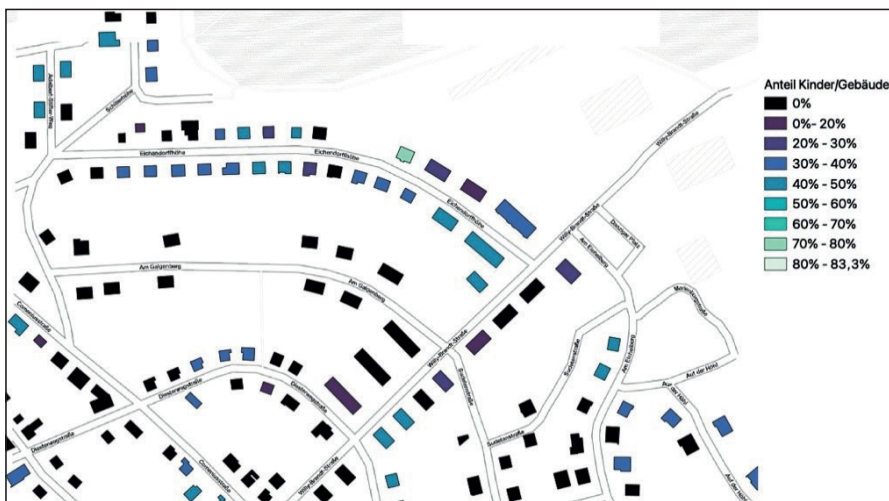


Abb. 2: Anteil Kinder pro Gebäude (in Prozent) (Quelle: Map Tiles von Stamen Design, CC BY 3.0. Daten von OpenStreetMap, ODbL)

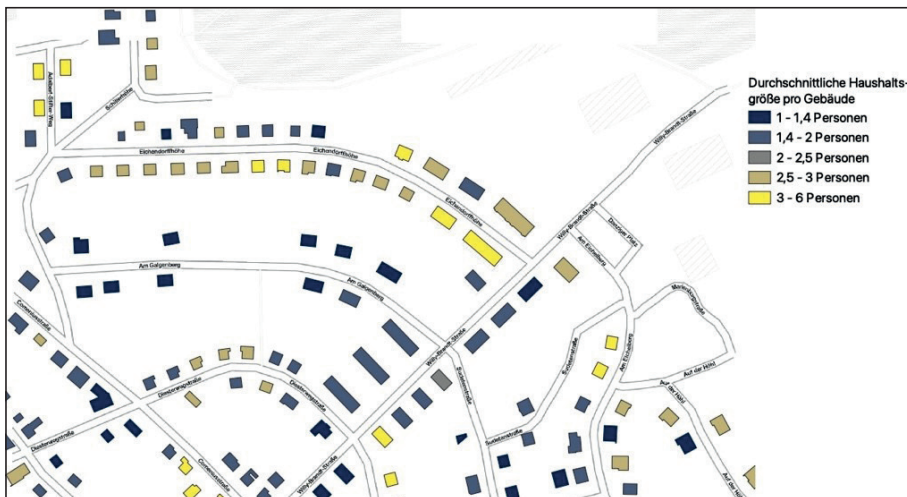


Abb. 3: Durchschnittliche Haushaltsgröße pro Gebäude (Quelle: Map Tiles von Stamen Design, CC BY 3.0. Daten von OpenStreetMap, ODbL)

In Verbindung mit anderen Attributen lassen sich potenziell Auswertungen über Familienstruktur bzw. wenn aggregiert über lebenszyklusbedingte Generationenwechsel innerhalb eines Quartiers vornehmen (vgl. hierzu auch Schaffert, Höcht 2018). Abbildung 3 zeigt die durchschnittliche Haushaltsgröße pro Gebäude. Maximal leben

sechs Personen in einem Haushalt (gelb hervorgehoben). Einpersonenhaushalte bilden die kleinste Einheit (dunkelblau hervorgehoben).

5 Fazit

Auf der Grundlage von frei verfügbaren Daten aus öffentlichen Beständen wurde in diesem Beitrag eine Methode vorgestellt, die ein auf Gebäudeebene georeferenzierten Datensatz erzeugt. Im Ergebnis sind Adresszuordnungen zu demographischen Merkmalen aus den Zensusergebnissen von 2011 als Basis weitergehender Indikatorberechnungen und Aggregationen abgeleitet worden. Verlässlichkeit und Tiefe können unter Hinzunahme weiterer Zensusdaten-Attribute und OpenStreetMap-Tags sowie ggf. zusätzlicher offener Datensätze noch verbessert werden. Eine ausführlichere Evaluierung der Qualitätsmerkmale der hier aufgezeigten Methode ist aufgrund der fehlenden internen Konsistenz der Zensusdaten, bedingt durch die notwendige Anonymisierung durch das SAFE-Verfahren sowie der eingeschränkten Datenqualität von OpenStreetMap erforderlich. Zu diesem Zweck wird ein Abgleich mit realen Melderegisterdaten angestrebt.

Zukünftig bietet die Methode, insbesondere im Hinblick auf die Ergebnisse des Zensus 2022 viel Potenzial, Datensätze ohne die Restriktionen des Datenschutzes für die Forschung bereitzustellen. Darüber hinaus lässt die Zugänglichkeit zu bisher nicht erhältlichen Daten aus Wirtschaft, Wissenschaft und Zivilgesellschaft mit einem breiteren inhaltlichen Spektrum hoffen.

6 Literatur

- Bundesamt für Kartographie und Geodäsie (2020): Dokumentation. Geographische Gitter für Deutschland. GeoGitter. https://sg.geodatenzentrum.de/web_public/gdz/dokumentation/deu/geogitter.pdf (Zugriff: 24.06.2022).
- Dittrich-Wesbuer, A.; Sturm, G. (2020): Quantitative Daten I: Amtliche Daten zur Verbreitung multilokaler Lebensformen in Deutschland, Österreich und der Schweiz. In: Danielzyk, R.; Dittrich-Wesbuer, A.; Hilti, N.; Toppel, C. (Hrsg.): Multilokale Lebensführungen und räumliche Entwicklungen: ein Kompendium. Hannover: Forschungsberichte der ARL 13: 104-111.
- Großmann, K.; Franke, R.; Henkel, L. (2021): Erfahrungsbericht Segregationsanalysen, Teil 1: Die Hürden der Datenbeschaffung. In: Steinführer, A.; Porsche, L.; Sondermann, M. (Hrsg.): Kompendium Kleinstadtforschung. Hannover: Forschungsberichte der ARL 16: 50-61.
- Kriesten, M. (2021): Geodaten im Kontext von Datenschutz, Urheberrecht und Informationsfreiheitsansprüchen. In: zfv – Zeitschrift für Geodäsie, Geoinformation und Landmanagement, Heft 5/2021, 146. Jg.: 340-347.
- Lucke, J. von; Gollasch, K. (2022): Open Government: Offenes Regierungs- und Verwaltungshandeln – Leitbilder, Ziele und Methoden. Springer Fachmedien: Wiesbaden.

- Milbert, A.; Fina, S. (2021): Methoden der Kleinstadtforschung: Definitionen, Daten und Raumanalysen. In: Steinführer, A.; Porsche, L.; Sondermann, M. (Hrsg.): Kompendium Kleinstadtforschung. Hannover: Forschungsberichte der ARL 16: 24-49.
- 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.
- 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.
- Olbricht, R. M. (2015): Data Retrieval for Small Spatial Regions in OpenStreetMap. In: Jokar Arsanjani, J.; Zipf, A.; Mooney, P.; Helbich, M. (Hrsg.): *OpenStreetMap in GIScience. Lecture Notes in Geoinformation and Cartography*: 101-122.
- OpenStreetMap (2022): OpenStreetMap.
<https://www.openstreetmap.org> (Zugriff: 28.06.2022).
- Porsche, L.; Sondermann, M.; Steinführer, A. (2021): Jenseits der "Aufmerksamkeitslücke" – Was wir bisher über Kleinstädte (nicht) wissen. In: Steinführer, A.; Porsche, L.; Sondermann, M. (Hrsg.): *Kompendium Kleinstadtforschung*. Hannover: Forschungsberichte der ARL, 16: 314-338.
- Schaffert, M.; Höcht, V. (2018): Geokodierte Meldedaten als Basis bedarfsgerechter Planungen in ländlichen Gemeinden und Regionen. *Raumforschung und Raumordnung* 76(5): 421-435.
- Siebenlist, T.; Mainka, A. (2019): Digitale Transformation in der Verwaltung: An Open Data geht kein Weg vorbei. In: Büttner, S. (Hrsg.): *Die digitale Transformation in Institutionen des kulturellen Gedächtnisses: Antworten aus der Informationswissenschaft*: 111-133.
- Statistische Ämter des Bundes und der Länder (2014): Nutzer-Kurzinformationen zu SAFE.
https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Nutzerhinweise_safe.pdf;jsessionid=DD285B7C567AF8E32D62F525D094BD.live292?__blob=publicationFile&v=2 (Zugriff: 24.06.2022).
- Statistische Ämter des Bundes und der Länder (2015): Zensus 2011. Methoden und Verfahren. https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=2 (Zugriff: 24.06.2022).
- Statistische Ämter des Bundes und der Länder (2020): Ergebnisse des Zensus 2011 zum Download – erweitert.
<https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html> (Zugriff: 24.06.2022).
- Statistische Ämter des Bundes und der Länder (2022): Zensus 2011.
https://www.zensus2011.de/DE/Zensus2011/zensus2011_node.html (Zugriff: 28.06.2022).
- Visca, D.; Hoppe, M.; Neis, P. (2022): Zur Generierung eines Melderegisters als Pseudo-Derivat aus frei verfügbaren Informationen. In: *AGIT – Journal für Angewandte Geoinformatik* 8: 73-82.