



Flächennutzungsmonitoring XI Flächenmanagement – Bodenversiegelung – Stadtgrün

IÖR Schriften Band 77 · 2019

ISBN: 978-3-944101-77-4

Die Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur SoRa – Potenziale und Implementierung

Stefan Jünger, Loren Mucha, Gotthard Meinel

Jünger, S.; Mucha, L.; Meinel, G. (2019): Die Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur SoRa – Potenziale und Implementierung. In: Meinel, G.; Schumacher, U.; Behnisch, M.; Krüger, T. (Hrsg.): Flächennutzungsmonitoring XI. Flächenmanagement – Bodenversiegelung – Stadtgrün. Berlin: Rhombos, IÖR Schriften 77, S. 139-148.

Die Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur SoRa – Potenziale und Implementierung

Stefan Jünger, Loren Mucha, Gotthard Meinel

Zusammenfassung

In den Sozialwissenschaften hat die Analyse des räumlichen Kontextes von individuellem Verhalten und Einstellungen Hochkonjunktur. Das zeigt sich besonders an den vielfältigen Arbeiten zu umweltbezogenen Einflüssen der Wohnumgebung auf individuelle Merkmale von Personen. Auf der einen Seite stehen Charakteristika jener Wohnumgebungen, wie etwa die Grünausstattung oder die bauliche Dichte, und auf der anderen Seite die Gesundheit, das Wohlbefinden oder die Einstellung von Menschen. Im gleichen Zuge besitzen diese Zusammenhänge oft auch eine soziale Ungleichheitsdimension, die sich vor allem darin niederschlägt, dass Personen mit niedrigem Einkommen oder Mitglieder ethnischer Minderheiten weniger von positiven Einflüssen der Wohnumgebung profitieren können und mehr von negativen Einflüssen betroffen sind. Während im US-amerikanischen Kontext auf eine traditionsreiche Forschung zurückgeblickt werden kann, fehlt in Deutschland bisher eine systematische Erschließung dieses Themas, u. a. weil dafür notwendige Daten fehlen. In dem von der Deutschen Forschungsgemeinschaft geförderten Projekt „Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur“ (SoRa) wird diese strukturelle Lücke adressiert und die hierfür notwendige Forschungsdateninfrastruktur entwickelt. Zusammen mit den Projektpartnern GESIS – Leibniz-Institut für Sozialwissenschaften, Leibniz-Institut für ökologische Raumentwicklung (IÖR), Karlsruher Institut für Technologie (KIT) und dem Sozio-oekonomischen Panel (SOEP) wird gegenwärtig eine Infrastruktur zur Verknüpfung von raumwissenschaftlichen Daten aus dem Monitor der Siedlungs- und Freiraumentwicklung (IÖR-Monitor) und georeferenzierten Umfragedaten (SOEP, ALLBUS) geschaffen. Der Beitrag stellt die allgemeinen und spezifischen Potenziale dieser Infrastruktur vor, erläutert ihre technische Implementierung und stellt erste Ergebnisse aus Analysen mit den verknüpften Daten vor.

1 Einleitung

Dieser Beitrag stellt die im Aufbau von der Deutschen Forschungsgemeinschaft (DFG) unterstützte Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur (SoRa) vor. Ziel von SoRa ist es, eine interdisziplinäre Forschungsdateninfrastruktur interoperabel zu entwickeln und die gemeinsame Nutzung der Daten für innovative Forschungsfragestellungen zu ermöglichen. Realisiert wurde der Aufbau von SoRa durch GESIS – Leibniz-

Institut für Sozialwissenschaften, dem Leibniz-Institut für ökologische Raumentwicklung (IÖR), dem Sozio-oekonomischen Panel (SOEP) und dem Karlsruhe Institut für Technologie (KIT). In der ersten Projektphase stammen exemplarische Fragestellungen aus dem Bereich der Sozialwissenschaften mit dem Schwerpunkt auf soziale Ungleichheiten im Zusammenhang mit Umweltbelastungen.

Im Folgenden gehen wir kurz auf die Motivation und den Hintergrund zu dem Projekt ein (Abschnitt 2). Danach beschreiben wir die technische Implementierung der Infrastruktur (Abschnitt 3), um dann eine erste inhaltliche Anwendung mit den Daten aus der Infrastruktur vorzustellen (Abschnitt 4). In einem letzten Schritt (Abschnitt 5) fassen wir die Hauptergebnisse zusammen und skizzieren die geplante Weiterentwicklung der Forschungsdateninfrastruktur.

2 Motivation

2.1 Georeferenzierte Daten in den Sozialwissenschaften

In den Sozialwissenschaften ist in den vergangenen Jahren ein deutlicher Trend zur Nutzung georeferenzierter Daten zu verzeichnen. Verschiedene Subdisziplinen, zum Beispiel im Bereich der politischen Einstellungen (Klinger et al. 2017) oder sozialer Ungleichheit (Zwickl et al. 2014), profitieren von einer kleinräumigen Perspektive auf bestehende Fragestellungen. Erst wenn Forschende Geoinformationen (z. B. Wohnumfeldbeschreibungen, Erreichbarkeiten usw.) zu bestehenden Umfragedaten hinzufügen können, wird eine Analyse des räumlichen Kontextes sozialen Handelns ermöglicht. Georeferenzierte Daten helfen dabei, bestehende Ergebnisse zu hinterfragen bzw. aus anderen Perspektiven zu beleuchten oder aber auch ganz neue Fragestellungen zu formulieren (Jünger 2019).

Möglich wird diese Unternehmung insbesondere durch die Methode der räumlichen Verknüpfung. Geokodierte Adressdaten von Befragten werden gemeinsam mit Geodaten in einem Raum projiziert und die Attribute der Geodaten basierend auf gleichen Lokalitäten extrahiert. Abbildung 1 zeigt das Beispiel des Indikators Bodenversiegelungsgrad, dessen 100 m x 100 m-Rasterzellwert über die räumliche Verknüpfung den Adresskoordinaten von Befragten hinzugefügt wird. Damit ergeben sich neue Fragestellungen, etwa in der Untersuchung sozialer Ungleichheiten in der Betroffenheit von Bodenversiegelung in Deutschland (Abschnitt 4).

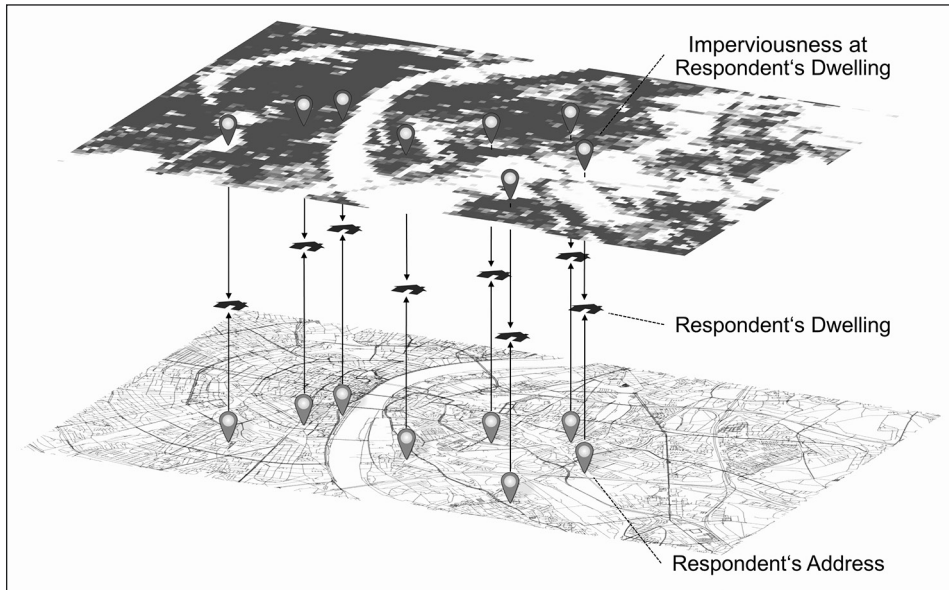


Abb. 1: Räumliche Verknüpfung von georeferenzierten Umfragedaten mit kleinräumigen Geodaten (Quelle: www.sora-projekt.de)

2.2 Herausforderungen

Zunächst sind Sozialforschende, die sich für georeferenzierte Daten interessieren, vor das Problem gestellt, dass die Anwendung dieser Daten oft nicht dem Curriculum ihres Fachbereichs entspricht. Im universitären Studium wird nur an sehr wenigen Hochschulen angeboten, zum Beispiel den Einsatz von Geographischen Informationssystemen (GIS) zu erlernen (Müller 2019). Aus diesem Grund ist die Bereitschaft zur Interdisziplinarität notwendig, um die Methoden aus anderen Fachbereichen zu erlernen und damit auch der erforderlichen Anwendung von Daten aus diesen Fachbereichen gerecht zu werden. Andererseits sollte es Sozialforschenden so leicht wie möglich gemacht werden, räumliche Kontexte in ihre Untersuchungen einzubinden.

Eine weitere Herausforderung im Umgang mit georeferenzierten Daten in den Sozialwissenschaften ist die (Daten-)Verfügbarkeit von Geodaten. Das betrifft insbesondere Daten öffentlicher Anbieter, da sich die Geodatenhaltung bedingt durch die föderale Struktur Deutschlands auf die Geodatenzentren der Bundesländer verteilen. Diese fragmentierte Datenlandschaft kann zu nicht-harmonisierten, fehlerhaften oder gar fehlenden Daten führen (Schweers et al. 2016).

Auch aufseiten georeferenzierter Umfragedaten ist dieser Effekt zu beobachten. Georeferenzierte Daten sind zunächst einfache Umfragedaten, die über Kataloge bei

Datenanbietern recherchiert werden können. Den Katalogen wiederum fehlen oft angemessene Beschreibungen über die verfügbaren Georeferenzen, sodass Forschende ein unvollständiges Bild über alle verfügbaren Daten erlangen. Während es einige prominente Beispiele gibt, wie zum Beispiel das Forschungsdatenzentrum (FDZ) des Sozio-ökonomischen Panels (SOEP) (Goebel 2017), ist der Zugriff auf kleinere und spezialisiertere Befragungen limitiert. Forschende müssen aus diesem Grund einige Zeit für die Datenrecherche investieren.

Die größte Herausforderung, die im Zusammenhang mit der Datenanalyse steht, ist die des Datenschutzes. Georeferenzierte Umfragedaten sind besonders schützenswert, da sie kleinräumige Informationen über die Wohnorte von Befragten enthalten. Nach deutschem Recht ist es nicht erlaubt, Umfragedaten mit Wohnortsinformationen der Befragten gemeinsam zu speichern. Datenanbieter reglementieren somit den Zugang zu georeferenzierten Umfragedaten, etwa durch den Einsatz spezieller Zugangswege in geschützten und von außen abgeschirmten Arbeitsumgebungen.

Die im SoRa Projekt entwickelte Forschungsdateninfrastruktur widmet sich diesen Herausforderungen, um einen großen Teil dieser sich ergebenden Schwierigkeiten für Sozialforschende abzumildern. Erreicht wird dies mit der Entwicklung einer technischen Plattform zur räumlichen Verknüpfung von georeferenzierten Umfragedaten mit Geodaten, die keine GIS-Kenntnisse erfordert und die alle datenschutzrechtlichen Anforderungen automatisiert berücksichtigt. Diese Infrastruktur reduziert somit Fehler, die sich aus der versehentlichen Speicherung von sensitiven Daten wie eben georeferenzierte Umfragedaten ergeben könnten.

3 Technische Implementierung

3.1 Ziele und Lösungsansatz

Ziel von SoRa ist, eine interoperable Architektur zwischen sozial- und raumwissenschaftlichen Forschungsdateninfrastrukturen zu entwickeln und bereitzustellen. Dieses Ziel impliziert die Interoperabilität sowie die Konformität mit internationalen Standards aus den Sozialwissenschaften und den Raumwissenschaften. Dadurch wird ein einfacher Zugang für Forschende zu den Daten geschaffen, der auf einer transparenten und verständlichen Kombination der Daten beruht.

Datengrundlage der raumwissenschaftlichen Seite in SoRa sind die Daten des Monitors der Siedlungs- und Freiraumentwicklung (IÖR-Monitor). Diese raumwissenschaftliche Forschungsdateninfrastruktur informiert seit 2010 deutschlandweit, indikatorbasiert, hochauflösend und in Form von Zeitreihen über die Flächennutzung und deren Entwicklung in Deutschland. Auf Grundlage dieser Daten sind u. a. auch kleinräumige Wohn-

umfeldbeschreibungen durch Indikatoren zur baulichen Dichte, zur Grünausstattung und -erreichbarkeit und zur Verkehrswegenetzdichte möglich.

Die IÖR-Monitoraten werden mit georeferenzierten Umfragedaten des GESIS Panels und des SOEP räumlich verknüpft. Da es sich bei letzteren um datenschutzrechtlich sehr sensible Daten handelt, wurde bei der Lösungsarchitektur ein besonderes Augenmerk auf den Datenschutz gelegt. Diese Herausforderung wird mit einer verteilten Lösung in Form verschiedener, lose gekoppelter Webservices gelöst, die zentral orchestriert werden und durch die Anwendung standardisierter Schnittstellen miteinander kommunizieren. Die einzelnen Komponenten werden im Folgenden kurz vorgestellt.

3.2 Architektur

Die SoRa-Architektur setzt sich aus den folgenden vier Komponenten zusammen, die miteinander kommunizieren:

1. SoRa-App
2. Survey Data Discovery
3. Participant Data Linking
4. Spatial Linking

Der zentrale Zugriffspunkt für Forschende bildet dabei die SoRa-App (1.), über welche Anfragen zur räumlichen Verknüpfung von IÖR-Monitor-Daten und Umfragedaten gestellt werden. Die Survey Data Discovery Komponente (2.) erlaubt den Zugriff auf allgemeine Metadaten aus den Umfragedaten, die für die Auswahl der Umfragedaten von Interesse sind. Im Hintergrund wählt die Participant Data Linking Komponente (3.) entsprechende Geokoordinaten der georeferenzierten Umfragedaten aus, um sie in einem vorerst letzten Schritt an die Spatial Linking Komponente (4.) beim IÖR zu senden. Eine Übersicht dazu zeigt Abbildung 2.

Generell beinhaltet der Endpunkt dieses Prozesses das Zurückspielen der Daten an die Forschenden. Da die zurückgespielten Daten jedoch weiterhin sensitive Daten beinhalten können, sieht die gegenwärtige Infrastruktur vor, dies lokal bei den Datenhaltenden zu lösen. Im Fall des Projektkonsortiums können somit Forschende auf die Daten des SOEP im FDZ SOEP in Berlin und auf die Daten des GESIS Panels im GESIS Secure Data Center in Köln zugreifen. Eine potentielle Anwendung der verknüpften Daten wird im folgenden Abschnitt vorgestellt.

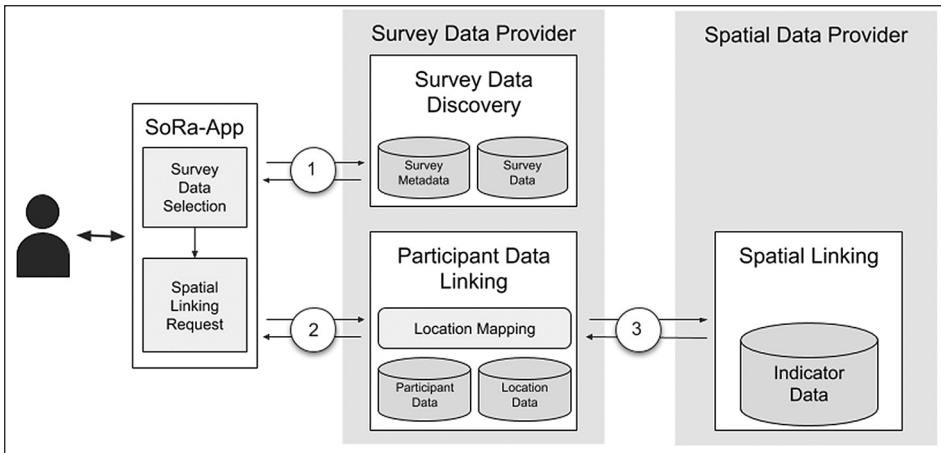


Abb. 2: Übersicht über die verteilte Architektur lose gekoppelter Webservices in der Sozial-Raumwissenschaftlichen Forschungsdateninfrastruktur SoRa (Quelle: Bensmann et al. 2019)

4 Sozialwissenschaftliche Anwendungsbeispiele

4.1 Sozialräumliche Ungleichheit

Die sozial ungleiche Verteilung von Umweltbelastungen ist schon länger Gegenstand zahlreicher Untersuchungen, vor allem in den Vereinigten Staaten von Amerika. Es zeigte sich, dass Personen mit niedrigem Einkommen oder Mitglieder ethnischer Minderheiten überproportional oft von Umweltbelastungen betroffen sind gegenüber Personen mit hohem Einkommen bzw. Mitgliedern der ethnischen Mehrheit. Auch in Deutschland ist dieses Thema in den letzten Jahren in den Fokus gerückt, wenngleich noch verhältnismäßig wenige empirische Belege auf kleinräumigem Niveau existieren.

Im Folgenden zeigen wir eine kurze Analyse, wie mit der neugeschaffenen Forschungsdateninfrastruktur SoRa diesen Fragen nachgegangen werden kann. Denn Flächennutzungsindikatoren können Umweltbelastungen anzeigen, wenn sie Freiraum eingrenzen und den Zugang zu Ressourcen wie zum Beispiel Grünflächen erschweren. Unsere Hypothese ist, dass auch in Deutschland Einkommen negativ mit solchen Indikatoren korreliert. Das heißt, ärmere Bevölkerungsgruppen sind überproportional oft Umweltbelastungen, die sich aus der Flächennutzung ergeben, ausgesetzt.

4.2 Daten und Methode

Für die Analyse nutzen wir georeferenzierte Umfragedaten, die mit Indikatoren des IÖR-Monitors räumlich verknüpft wurden. Es handelt sich dabei um zwei verschiedene Umfrage-Datenquellen: das GESIS Panel aus dem Jahr 2014 mit insgesamt 3 852 Befragten

und das Sozio-ökonomische Panel aus dem Jahr 2014 mit insgesamt 27 813 Befragten. Bei beiden Datenquellen handelt es sich um sogenannte Bevölkerungsstichproben, welche durch ihre Form der Stichprobenziehung versuchen, ein möglichst repräsentatives Bild der deutschen Bevölkerung abzubilden.

Aus dem IÖR Monitor benutzen wir die Indikatoren Bodenversiegelungsgrad, Gesamtkraftverkehrsnetzdichte und Industrie- und Gewerbeflächenanteil an der Gebietsfläche. Es ist davon auszugehen, dass die drei gewählten Indikatoren negativ mit der Grün-ausstattung korrelieren, einen hohen Wert an Lärm und Abgasen bedingen und damit unmittelbar als Umweltbelastungen auf die dort wohnenden Menschen einwirken. Von diesen Indikatoren wurden jeweils der Indikatorwert einer 100 m x 100 m großen Rasterzelle und der Mittelwert einer Pufferzone von 500, 1 000 und 2 000 Metern um die befragte Adresse zugespielt. Damit wird gesichert, dass das gesamte Wohnumfeld der befragten Personen in die Untersuchungen einbezogen wird.

4.3 Ergebnisse

Sind Personen mit einem niedrigen Einkommen häufiger Umweltbelastungen aus der Flächennutzung ausgesetzt als Personen mit einem hohen Einkommen? Um dieser Frage nachzugehen, wurden basierend auf einem linearen Regressionsmodell Vorhersagen für die jeweiligen Umweltbelastungen in Abhängigkeit des Einkommens errechnet.

Abbildung 3 zeigt die Ergebnisse der Analyse zunächst für die Daten des GESIS Panels für die IÖR-Indikatoren Bodenversiegelung (Imperviousness), Gesamtkraftverkehrsnetzdichte (Density of total roadway network) und die Anteil Industrie- und Gewerbeflächenanteil an der Gebietsfläche (Percentage of industrial and commercial site to geographic area). Auf den X-Achsen aller drei Unterabbildungen befindet sich das Einkommen der Befragten (Income) in Kategorien von 1 bis 17 und auf den Y-Achsen der jeweilige vorhergesagte Indikatorwert in Prozent (Estimate). In den Abbildungen selbst befinden sich die Vorhersagen für die verschiedenen geographischen Größen der Indikatorwerte mit deren Unsicherheitsintervallen (95 %ige Konfidenzintervalle).

Bezüglich des Bodenversiegelungsgrades zeigt sich, dass mit höherem Einkommen der Bodenversiegelungsgrad sinkt. Im Vergleich zur Gruppe mit den niedrigsten Einkommen sinkt der Indikatorwert der 100 m x 100 m-Rasterzelle um 14 % gegenüber der Gruppe mit dem höchsten Einkommen. Dieser Zusammenhang ist natürlich weniger ausgeprägt, wenn der Untersuchungsraum durch Pufferbildung vergrößert wird. Die gleiche Tendenz zeigt sich bei den Indikatoren Gesamtverkehrsnetzdichte und Industrie- und Gewerbeflächenanteil an der Gebietsfläche. Jedoch handelt es sich dabei um wenig prägnante Ergebnisse, sodass sich kaum ein Zusammenhang mit dem Einkommen zeigt.

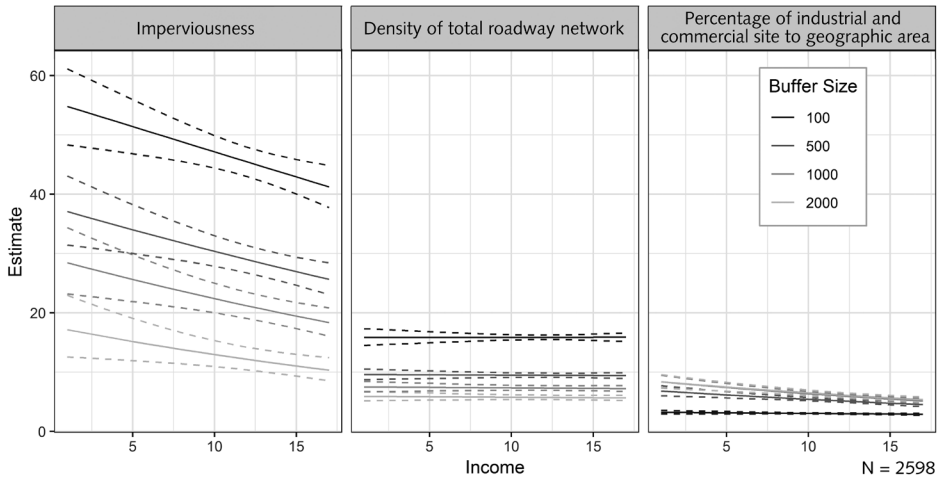


Abb. 3: Ergebnisse der linearen Vorhersage als Funktion von Einkommen und den IÖR-Indikatoren Bodenversiegelungsgrad, Gesamtverkehrsnetzdicke und Anteil an Industrie- und Gewerbeflächenanteil an der Gebietsfläche für das GESIS Panel (N = 2598) (Quelle: Bensmann et al. 2019)

Diese Ergebnisse bewähren sich in der Analyse des SOEP, dargestellt in Abbildung 4. Da hier die Stichprobe größer ist, sind auch die Vorhersagen präziser. Im Vergleich zur Gruppe mit dem niedrigsten Einkommen ist der Bodenversiegelungsgrad auf 100 m x 100 m gar um 19 % höher als in der Gruppe mit dem höchsten Einkommen. Die auf den linearen Trend bezogenen Vorhersagen für die anderen beiden Indikatoren sind ähnlich zu denen des GESIS Panels.

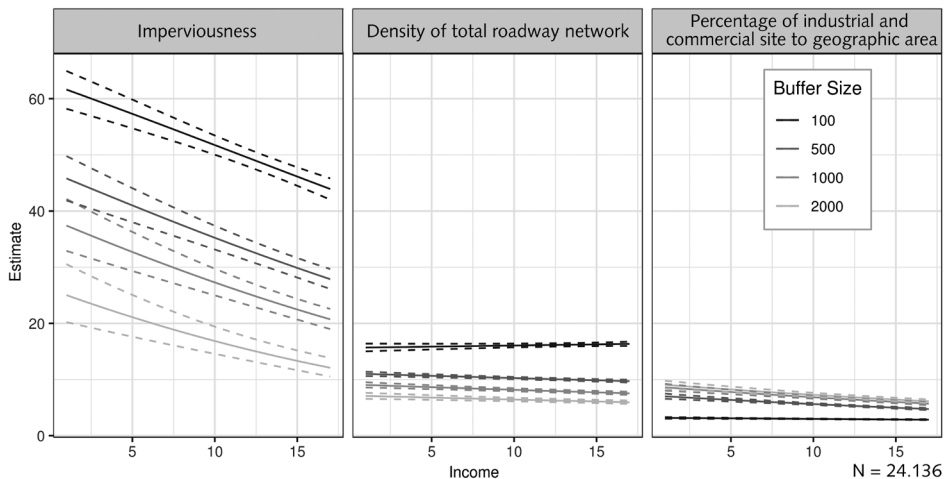


Abb. 4: Ergebnisse der linearen Vorhersage als Funktion von Einkommen und den IÖR-Indikatoren Bodenversiegelungsgrad, Gesamtverkehrsnetzdicke und Anteil an Industrie- und Gewerbeflächenanteil an der Gebietsfläche für das SOEP (N = 24.136) (Quelle: Bensmann et al. 2019)

4.4 Zusammenfassung des Anwendungsbeispiels

Generell zeigt dieses Anwendungsbeispiel, wie sich Daten aus verschiedenen Forschungsdateninfrastrukturen miteinander kombinieren lassen. Ähnliche und doch leicht verschiedene Datenquellen replizieren zudem ähnliche Ergebnisse: Einkommen ist negativ mit Umweltbelastungen resultierend aus der Flächennutzung assoziiert. Personen mit niedrigen Einkommen sind vor allem Belastungen, die sich aus der Bodenversiegelung ergeben, überproportional oft ausgesetzt. Derartige Untersuchungen können in Zukunft verfeinert werden, zum Beispiel indem weitere soziodemographische Angaben, etwa der Migrationshintergrund von Personen einbezogen werden. Es wird spannend werden, wie und mit welchen Fragen potentielle Anwenderinnen und Anwender die entwickelte Forschungsdateninfrastruktur SoRa in dieser Hinsicht nutzen.

5 Fazit und Ausblick

SoRa demonstriert das Potential interoperabler und interdisziplinärer Forschungsdateninfrastrukturen. Die Implementierung aus lose gekoppelten Webservices, die durch standardisierte Schnittstellen miteinander kommunizieren, ermöglicht eine leichte Ergänzung weiterer Datenquellen und auch die Übertragung bzw. Kopplung der Forschungsdateninfrastruktur SoRa auf bzw. mit Forschungsdateninfrastrukturen andere Fachdisziplinen. Derartige Kopplungen von Forschungsdateninfrastrukturen sind das Ziel beim beginnenden Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) für Deutschland.

Die Infrastruktur folgt einem Baukastenprinzip, das beliebig um weitere Komponenten erweiterbar ist und diese nahtlos integriert. Nächste Ziele des Projekts sind die Erweiterung um neue Geodatenquellen (offenen behördliche Geobasisdaten, OSM usw.) und weiterer sozialwissenschaftlicher Datensätze (z. B. GLES, ALLBUS) sowie die Ergänzung weiterer Funktionen und Analysemöglichkeiten.

6 Literatur

- Bensmann, F.; Heling, L.; Jünger, S.; Mucha, L.; Acosta, M.; Goebel, J.; Meinel, G.; Sure-Vetter, Y.; Zapilko, B. (2019): An Infrastructure for Spatial Linking of Survey Data. Manuskript (erhältlich auf Anfrage).
- Goebel, J. (2017): SOEP 2015 – Informationen zu den SOEP-Geocodes in SOEP v32. DIW Berlin, German Institute for Economic Research.
- Jünger, S. (2019): Using Georeferenced Data in Social Science Survey Research. The Method of Spatial Linking and Its Application with the German General Social Survey and the GESIS Panel. GESIS-Schriftenreihe, 24. Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.
<https://nbn-resolving.org/urn:nbn:de:0168-ssoar-63688-7> (Zugriff: 27.06.2019).

- Klinger, J.; Müller, S.; Schaeffer, M. (2017): Der Halo-Effekt in einheimisch-homogenen Nachbarschaften: Steigert die ethnische Diversität angrenzender Nachbarschaften die Xenophobie? *Zeitschrift für Soziologie* 46: 402-419.
- Müller, S. (2019): Räumliche Verknüpfung georeferenzierter Umfragedaten mit Geodaten: Chancen, Herausforderungen und praktische Empfehlungen. In: Jensen, U.; Netscher, S.; Weller, K. (Hrsg.): *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten*. Opladen, Berlin, Toronto: Verlag Barbara Budrich. 211-229.
- Schweers, S.; Kinder-Kurlanda, K.; Müller, S.; Siegers, P. (2016): Conceptualizing a Spatial Data Infrastructure for the Social Sciences: An Example from Germany. *Journal of Map & Geography Libraries* 12: 100-126.
- Zwickl, K., Ash, M.; Boyce, J. K. (2014): Regional Variation in Environmental Inequality: Industrial Air Toxics Exposure in U.S. Cities. *Ecological Economics* 107: 494-509.

Danksagung

Wir danken der Deutschen Forschungsgemeinschaft (DFG) für die Finanzierung des Aufbaues der Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur (SoRa, Projektnummer 316669855).