

SOEP Survey Papers

Series C – Data Documentation

SOEP – The German Socio-Economic Panel study at DIW Berlin

2016

Das DIW-IAB-RWI-Nachbarschaftspanel: Ein Scientific-Use-File mit lokalen Aggregatdaten und dessen Verknüpfung mit dem deutschen Sozio-oekonomischen Panel

Elisabeth Bügelmeyer, Sandra Schaffner, Norbert Schanne, Theresa Scholz

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at <http://www.diw.de/soepsurveypapers>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Elisabeth Bügelmeyer, Sandra Schaffner, Norbert Schanne, Theresa Scholz. 2016. Das DIW-IAB-RWI-Nachbarschaftspanel: Ein Scientific-Use-File mit lokalen Aggregatdaten und dessen Verknüpfung mit dem deutschen Sozio-oekonomischen Panel. SOEP Survey Papers 364: Series C. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

© 2016 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

soeppapers@diw.de

Das DIW-IAB-RWI-Nachbarschaftspanel: Ein Scientific-Use-File mit lokalen Aggregatdaten und dessen Verknüpfung mit dem deutschen Sozio-oekonomischen Panel

Elisabeth Bügelmeier, Sandra Schaffner, Norbert Schanne, Theresa Scholz

Essen, 2015

Diese Dokumentation beschreibt einen Datensatz, der unter der DOI
10.7807/DIWIABRWI:V1 registriert ist und im FDZ IAB, FDZ Ruhr und FDZ SOEP zugänglich ist.

Inhaltsverzeichnis

1.	Einleitung	3
2.	Datenquellen.....	3
2.1	Die Integrierten Erwerbsbiographien des IAB (IEB).....	3
2.1.1	Die Beschäftigtenhistorik (BeH).....	4
2.1.2	Leistungsempfängerhistorik (LeH).....	4
2.1.3	Leistungs-Historiken Grundsicherung (LHG/XLHG)	4
2.2	Adressdaten der BA	5
2.3	Variablenbeschreibung	5
3.	Sozialdatenschutz und Anonymisierung	9
4.	Datenqualität und Probleme.....	11
5.	Zusammenspielen mit SOEP und weitere Verknüpfungspotentiale.....	11
5.1	Regionalanalysen mit dem SOEP	11
5.2	Weiteres Verknüpfungspotenzial	12
6	Datenzugang	18
7.	Literaturverzeichnis	18

Verzeichnis der Tabellen

Tabelle 1:	Variablenliste	6
Tabelle 2	Anonymisierte Beobachtungen 2004 – 2010	10
Tabelle 3	Weitere Daten auf kleinräumiger Ebene	14
Tabelle 4	Anonymisierte PLZ nach dem Zusammenspielen mit dem SOEP	17

1. Einleitung

Das Rheinisch-Westfälische Institut für Wirtschaftsforschung (RWI), das Deutsche Institut für Wirtschaftsforschung (DIW) und das Institut für Arbeitsmarkt- und Berufsforschung (IAB) haben in den Jahren 2011 bis 2014 gemeinsam in dem Projekt „Nachbarschaftseffekte: Die Analyse individuell-rationalen Verhaltens im sozialen Kontext“ gearbeitet. Ziel des Projektes war es die Folgen sozialer Interaktionen in Deutschland zu untersuchen. Aufgrund eines Mangels an Individualdaten, die gleichzeitig auch ausreichend Informationen zur Nachbarschaft zulassen, war ein Bestandteil des Projekts kleinräumige Daten zu erzeugen, die an das Sozio-oekonomische Panel (SOEP) angespielt werden können. Im Gegensatz zu vielen bestehenden Daten soll dabei aber die Aggregationsebene nicht das Bundesland, der Kreis oder die Gemeinde sondern das Postleitzahlengebiet sein.

Die Postleitzahlebene bietet gegenüber der Gemeindeebene den Vorteil, dass sie auch innerhalb der Städte eine Differenzierung nach Nachbarschaften zulässt. Eine Stadt, meist eine einzelne Gemeinde, besteht aber aus mehreren Postleitzahlgebieten (PLZ). Die Gemeinde Berlin besteht beispielsweise aus 191 verschiedenen PLZ Gebieten. Der Nachteil der PLZ Ebene gegenüber der Gemeindeebene ist, dass sie keine Einheit der amtlichen Statistik ist. Nachbarschaftsindikatoren auf PLZ Ebene zu finden hat sich daher als sehr schwierig erwiesen. Darüber hinaus kann die Deutsche Post, je nachdem wie es für die Zustellung der Post effizient ist, PLZ Grenzen verschieben oder ganze PLZ Gebiete schaffen oder auflösen (Knies/Spieß 2007). Änderungen sind dann nur sehr schwer nachzuvollziehen.

Aus diesen Gründen wurden aus den Integrierten Erwerbsbiografien (IEB) aggregierte Daten auf Postleitzahlen (PLZ) generiert. Dabei handelt es sich zum einen um sozialversicherungspflichtig Beschäftigte und zum anderen um Arbeitslose jeweils insgesamt und nach bestimmten Merkmalen.

Im Folgenden wird zunächst die Datenquelle, die IEB, beschrieben. Anschließend wird die Generierung der Daten und die Anonymisierung vorgestellt. Im vierten Abschnitt werden die Datenqualität und Probleme kurz erläutert. Die erzeugten Daten wurden letztendlich an das SOEP angespielt, so dass Analysen zu Nachbarschaftseffekten möglich sind. Das Zusammenspielen wird in Abschnitt 5 dargestellt. Allerdings ist es möglich, diese Daten auch mit weiteren Datensätzen auf PLZ Ebene zu verknüpfen bzw. noch weitere Datenquellen anzuspüren. Eine Auswahl solcher Datenquellen wird abschließend kurz vorgestellt. In Abschnitt 6 wird der Datenzugang zu den Daten kurz erläutert.

2. Datenquellen

Die generierten Daten zu Erwerbstätigkeit und Arbeitslosigkeit auf PLZ Ebene beruhen auf den Integrierten Erwerbsbiographien des IAB (IEB). Diese Daten wurden wiederum mit der Adressdatei der Bundesagentur für Arbeit verknüpft.

2.1 Die Integrierten Erwerbsbiographien des IAB (IEB)

Die IEB führt Individualdaten aus den Meldungen zur Sozialversicherung (Beschäftigten-Historik, BeH) und Daten aus den Geschäftsprozessen der BA sowie der SGB II-Träger (Arbeitsuchenden-Historiken, ASU, Leistungsempfänger-Historik, LeH, Leistungs-Historiken Grundsicherung, LHG/XLHG, Maßnahmen-Teilnahme-Gesamtdatensatz, MTG) zusammen. Die IEB wird einführend in IAB (2012) beschrieben, eine weitergehende Beschreibung einer Stichprobe aus der IEB findet sich in vom Berge/Burghard/Trenkle (2013).

Alle genannten Datenquellen bestehen aus Sätzen, die Zeiträume abbilden. In den hier beschriebenen Daten werden allerdings nur solche Sätze verwendet, die den 30.06. des jeweiligen Jahres umfassen.

Abweichend von den Ausgangsdaten sind die hier verwendeten Daten als Stichtagsdaten zum 30.06. zu verstehen.

Oft bestehen für eine Person gleichzeitig mehrere Sätze aus verschiedenen Datenquellen (etwa, weil ein Arbeitslosengeldempfänger neben dem LeH- oder LHG-Satz nahezu immer einen ASU-Eintrag hat) oder aus derselben Datenquelle (etwa, wenn sie neben ihrer Haupttätigkeit auch einen Nebenjob ausübt). Hier wird je Person nur ein Datensatz, ihre sogenannte Hauptmeldung, verwendet. Im Normalfall ist dies die Beschäftigungsmeldung mit dem höchsten für die Person vorliegenden Bruttotagesentgelt. Falls keine Beschäftigung vorliegt, wird der letztgültige LeH-Satz oder zuletzt der jeweils gültige (X)LHG-Satz verwendet. Aufgrund dieser Logik sind Datensätze mit Leistungsempfang und vor allem solche aus dem Geltungsbereich des SGB II unterrepräsentiert.

Die in der IEB enthaltenen, in die regionale Aggregation übernommenen Prozessdaten sollen im Folgenden genauer beschrieben werden.

2.1.1 Die Beschäftigtenhistorik (BeH)

Die Quelle für Informationen über Beschäftigung ist die Beschäftigten-Historik des IAB. Die rechtliche Grundlage dieser Daten ist das mit Wirkung vom 1. Januar 1973 eingeführte (und seit dem 1. Januar 1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, das durch das Kürzel DEÜV (früher DEVO/DÜVO) bezeichnet ist (vgl. für weitere Details: Bender et al. 1996, S. 4 ff.; Wermter/Cramer 1988). Es verlangt von den Arbeitgebern Meldungen aller sozialversicherungspflichtig beschäftigten Arbeitnehmer an die jeweils zuständigen Sozialversicherungsträger. Die BeH erfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind.

Dies bedeutet, dass Beamte, Selbstständige und ordentliche Studierende (vgl. Wermter/Cramer 1988) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 1. Januar 1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (erst ab 1. April 1999 enthalten). Jedes Jahr, in dem ein Beschäftigungsverhältnis für eine Person existiert, ist durch mindestens eine Meldung abgebildet. Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit in einer laufenden Datei gesammelt und anschließend vom IAB in seine Historik-Datei integriert.

2.1.2 Leistungsempfängerhistorik (LeH)

Die Leistungsempfänger-Historik des IAB erfasst Zeiträume, in denen Personen Lohnersatzleistungen von der Bundesagentur für Arbeit (BA) beziehen. Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, also ausschließlich Leistungen nach dem Rechtskreis SGB III. Da der Leistungsanspruch von der Erfüllung gesetzlicher Voraussetzungen abhängt, werden Arbeitslosigkeitsperioden, in denen die Voraussetzungen nicht vorliegen (z.B. fehlende Bedürftigkeit im Falle der Arbeitslosenhilfe oder Nichterfüllung der Anwartschaftszeit beim Arbeitslosengeld), nicht in der Leistungsempfänger-Historik berichtet.

2.1.3 Leistungs-Historiken Grundsicherung (LHG/XLHG)

Während die LeH die Leistungen des Rechtskreises SGB III enthält, beinhalten LHG und XLHG Leistungsbezug nach SGB II. Dies umfasst sowohl Leistungen zur Sicherung des Lebensunterhalts (z. B. ALG II) wie auch Zuschläge zum Arbeitslosengeld oder auch zusätzliche Leistungen. Im Gegensatz zu den Leistungen aus dem Rechtskreis des SGB III ist die Bundesagentur für Arbeit nicht mehr zwingend alleiniger Träger der Leistungen. Es wird demnach nach den drei möglichen Trägerarten grundsätzlich unterschieden:

- Gemeinsame Einrichtungen (gE) bzw. Arbeitsgemeinschaften (ARGE) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune.
- Bei getrennter Trägerschaft bleibt die Aufgabenteilung zwischen BA und Kommune bestehen.
- Bei zugelassenen kommunalen Trägern, die auch als optierende Kommunen oder Optionskommunen bezeichnet werden, übernimmt die Kommune Aufgaben aus dem Bereich des SGB II.

Die Leistungs-Historiken Grundsicherung umfassen zwei Quellen, die sich im Übermittlungsverfahren der Daten unterscheiden: Die Daten der „Leistungs-Historik Grundsicherung aus A2LL“ (LHG) stammen aus dem BA-Verfahren A2LL, während die „Leistungs-Historik-Grundsicherung aus XLHG-BA-SGBII“ (XLHG) Daten erfasst, welche nach dem Übermittlungsstandard X-Sozial gemeldet wurden. In der Regel findet A2LL bei ARGE n Verwendung, während der Standard X-Sozial von den optierenden Kommunen genutzt wird. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz. Die Ermittlung der Höhe des Leistungsbezugs erfolgt nicht auf individueller Ebene, sondern auf der Ebene von Bedarfsgemeinschaften.

In LHG und XLHG sind grundsätzlich alle gemeldeten Berechtigten nach § 7 SGB II, also erwerbsfähige Hilfebedürftige und ggf. ihre nichterwerbsfähigen Bedarfsgemeinschaftsmitglieder enthalten. Es wurden alle Personen geocodiert, in die regionalen Aggregatdaten gehen allerdings nur die erwerbsfähigen Hilfebedürftigen ein. Es sind jeweils überschneidungsfreie Zeiträume der Leistungsberechtigung einer Person in einer bestimmten Bedarfsgemeinschaft abgebildet. Die Sätze werden jedoch an transferrelevanten Geburtstagen der Mitglieder (14, 15, 18, 25 und 65) einer Bedarfsgemeinschaft, bei Veränderungen in der Zusammensetzung einer Bedarfsgemeinschaft (z.B. durch Zu-/ Abgänge), bei Merkmalsänderungen des Bedarfsgemeinschafts-Kunden sowie bei Beginn und Ende eines Sanktionssachverhalts für Sätze ab dem 1. April 2006 gesplittet.

2.2 Adressdaten der BA

Die kleinräumigste in der IEB berichtete Ortsangabe bezieht sich auf die Gemeinde, in der eine Person wohnt oder arbeitet, bzw. die zuständige Dienststelle der Bundesagentur für Arbeit. Zur Bestimmung der Postleitzahl (PLZ), unter der eine Person wohnt, werden als zweite Datenquelle die „Adressdaten der Bundesagentur für Arbeit“ herangezogen. Diese berichten für jede Person die der BA gemeldeten Postanschriften, also unter anderem auch die PLZ, mit jeweiligem Gültigkeitszeitraum. Die Adressangaben beruhen wiederum auf Meldungen nach DEÜV, aus dem Verfahren A2LL und aus XSOZIAL (vgl. auch den Abschnitt zu „Kontaktdaten für Befragungen“ in IAB, 2012, S. 37).

Angaben zum Erwerbsverlauf sind in den Adressdaten der BA nicht enthalten. Eine Verknüpfung mit der IEB ist über einen in beiden Datenquellen berichteten nichtsystematischen Zuordnungsschlüssel, die sogenannte „Einheitliche Statistische Person“ (ESP), möglich.

2.3 Variablenbeschreibung

Aus den oben beschriebenen Daten wurden verschiedene Kennzahlen auf Postleitzahlenebene zum 30.06. generiert. Dabei wurde jeweils die Zahl der Personen innerhalb eines Postleitzahlengebietes aufsummiert, auf die bestimmte Charakteristika zutreffen. Tabelle 1 gibt einen Überblick über die einzelnen Variablen.

Tabelle 1:

Variablenliste

Name	Beschreibung	Definition
Beschäftigte		
BN	Anzahl Beschäftigte: insgesamt	Basiert auf Entgeltmeldungen, die von den Betrieben über jede sozialversicherungspflichtige Beschäftigung und geringfügige Beschäftigungen an die Rentenversicherungsträger übermittelt werden müssen. Gezählt werden also sowohl sozialversicherungspflichtige als auch geringfügig Beschäftigte.
BMANN	Beschäftigte: Männer	s. Anmerkung zu "BN"
BFRAU	Beschäftigte: Frauen	s. Anmerkung zu "BN"
BALT20	Beschäftigte: < 20 Jahre	Auf Basis des Geburtsdatums werden das Alter und die Zahl der Beschäftigten in den jeweiligen Altersklassen berechnet.
BALT25	Beschäftigte: 20 - 24 Jahre	
BALT30	Beschäftigte: 25 - 29 Jahre	
BALT35	Beschäftigte: 30 - 34 Jahre	
BALT40	Beschäftigte: 35 - 39 Jahre	
BALT45	Beschäftigte: 40 - 44 Jahre	
BALT50	Beschäftigte: 45 - 49 Jahre	
BALT55	Beschäftigte: 50 - 54 Jahre	
BALT60	Beschäftigte: 55 - 59 Jahre	Basiert auf der vierten Stelle des Tätigkeitsschlüssels, des so genannten B1-Schlüssels. Dieser informiert über den Status eines Versicherten im Berufsleben. Gezählt werden Personen, die die Ausprägung „Auszubildender (Lehrling, Auszubildender, Praktikant, Volontär)“ aufweisen.
BALT65	Beschäftigte: 60 - 65 Jahre	
BDEUTSCH	Beschäftigte: Deutsche Staatsangehörigkeit	
BAZUBI	Beschäftigte: Azubis	Basiert auf der vierten Stelle des Tätigkeitsschlüssels, des so genannten B1-Schlüssels. Dieser informiert über den Status eines Versicherten im Berufsleben. Gezählt werden Personen, die die Ausprägung „Auszubildender (Lehrling, Auszubildender, Praktikant, Volontär)“ aufweisen.
BUNGA	Anzahl Beschäftigte: Un- und angelernte Arbeiter	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Arbeiter, der nicht als Facharbeiter tätig ist“ aufweisen.
BFACH	Anzahl Beschäftigte: Facharbeiter	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Arbeiter, der als Facharbeiter tätig ist“ aufweisen.
BMEIST	Anzahl Beschäftigte: Meister / Poliere	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Meister/Polier“ aufweisen.
BANG	Anzahl Beschäftigte: Angestellte	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Angestellter“ aufweisen.
BTEIL1	Anzahl Beschäftigte: Wochenarbeitszeit < 18 Stunden	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Wochenarbeitszeit weniger als 18 Stunden“ aufweisen.

BTEIL2	Anzahl Beschäftigte: Wochenarbeitszeit 18 Stunden und mehr, aber nicht vollzeitbeschäftigt	Basiert auf der vierten Stelle des Tätigkeitsschlüssels. Gezählt werden Personen, die die Ausprägung „Wochenarbeitszeit 18 Stunden und mehr, aber nicht vollzeitbeschäftigt“ aufweisen.
BBILDUNK	Anzahl Beschäftigte: Ausbildung unbekannt	Basiert auf dem Merkmal Ausbildung. Das Merkmal enthält die Schul- und Berufsausbildung des Beschäftigten, die vom Arbeitgeber bei der Beschäftigtenmeldung eingetragen werden. Gezählt werden Personen, die die Ausprägung „Ausbildung unbekannt, Angabe nicht möglich“ aufweisen.
BHSOHNE	Anzahl Beschäftigte: Haupt- oder Realschule ohne Berufsabschluss	Basiert auf dem Merkmal Ausbildung. Gezählt werden Personen, die die Ausprägung „Volks-/ Hauptschule, Mittlere Reife oder gleichwertige Schulbildung ohne Berufsausbildung“ aufweisen.
BHSMIT	Anzahl Beschäftigte: Haupt- oder Realschule mit Berufsabschluss	Basiert auf dem Merkmal Ausbildung. Gezählt werden Personen, die die Ausprägung „Volks-/ Hauptschule, Mittlere Reife oder gleichwertige Schulbildung mit Berufsausbildung“ aufweisen.
BABIOHNE	Anzahl Beschäftigte: Abitur ohne Berufsabschluss	Basiert auf dem Merkmal Ausbildung. Gezählt werden Personen, die die Ausprägung „Abitur (Hochschulreife allgemein und fachgebunden) ohne Berufsausbildung“ aufweisen.
BABIMIT	Anzahl Beschäftigte: Abitur mit Berufsabschluss	Basiert auf dem Merkmal Ausbildung. Gezählt werden Personen, die die Ausprägung „Abitur (Hochschulreife allgemein und fachgebunden) mit Berufsausbildung“ aufweisen.
BUNI	Anzahl Beschäftigte: Uni oder FH-Abschluss	Basiert auf dem Merkmal Ausbildung. Gezählt werden Personen, die die Ausprägung „Abschluss einer Fachhochschule“ oder „Hochschul-/ Universitätsabschluss“ aufweisen.
BNARB	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: niedere, manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Agrarberufe“ oder „Einfache manuelle Berufe“ aufweisen.
BMARB	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: mittlere, manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Qualifizierte manuelle Berufe“ aufweisen.
BHARB	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: hohe, manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Techniker“ oder „Ingenieure“ aufweisen.
BNANG	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: niedrige, nicht-manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Einfache Dienste“ oder „Einfache kaufmännische Verwaltungsberufe“ aufweisen.
BMANG	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: mittlere, nicht-manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Qualifizierte Dienste“ oder „Qualifizierte kaufmännische und Verwaltungsberufe“ aufweisen.

BHANG	Anzahl Beschäftigte: Berufsklassifikation nach Blossfeld: hohe, nicht-manuelle Tätigkeiten	Basiert auf der Berufsklassifikation nach Blossfeld (1985). Gezählt werden Personen, die die Ausprägung „Semiprofessionen“ oder „Professionen“ oder „Manager“ aufweisen.
Arbeitslose		
AN	Anzahl Arbeitslose: insgesamt	Als Arbeitslos gezählt werden Personen aus der a) Leistungsempfängerhistorik, d.h. Bezieher von Arbeitslosengeld, Arbeitslosenhilfe, Unterhaltsgeld, Beiträgen zur Pflegeversicherung bei Personen, die privat versichert sind b) Arbeitssuchendenhistorik mit Erwerbsstatus „arbeit-suchend und arbeitslos“ oder „krank während arbeits-los“ c) Leistungshistorik Grundsicherung mit Erwerbsstatus „Erwerbsfähiger Hilfebedürftiger“ oder „Minderjähriger Erwerbsfähiger Hilfebedürftiger“ s. Bemerkung zu "AN"
AMANN	Anzahl Arbeitslose: Männer	s. Bemerkung zu "AN"
AFRAU	Anzahl Arbeitslose: Frauen	s. Bemerkung zu "AN"
AALT20	Arbeitslose: < 20 Jahre	Auf Basis des Geburtsdatums wird das Alter berechnet und die Zahl der Arbeitslosen in den jeweiligen Alters-klassen berechnet.
AALT25	Arbeitslose: 20 - 24 Jahre	
AALT30	Arbeitslose: 25 - 29 Jahre	
AALT35	Arbeitslose: 30 - 34 Jahre	
AALT40	Arbeitslose: 35 - 39 Jahre	
AALT45	Arbeitslose: 40 - 44 Jahre	
AALT50	Arbeitslose: 45 - 49 Jahre	
AALT55	Arbeitslose: 50 - 54 Jahre	
AALT60	Arbeitslose: 55 - 59 Jahre	
ADEUTSCH	Arbeitslose: Deutsche Staatsan-gehörigkeit	
AL00	Anzahl Arbeitslose: Dauer der Arbeitslosigkeit Null oder unbe-kannt	Die Dauer der Arbeitslosigkeit wird aus der tagesge-nauen Differenz von Ende- und Beginndatum des Leis-tungsbezugs / der Arbeitslosigkeitsmeldung berechnet.
ALOK1J	Anzahl Arbeitslose: Dauer der Arbeitslosigkeit bis ein Jahr	Die Dauer der Arbeitslosigkeit wird aus der tagesge-nauen Differenz von Ende- und Beginndatum des Leis-tungsbezugs / der Arbeitslosigkeitsmeldung berechnet.
ALOG1J	Anzahl Arbeitslose: Dauer der Arbeitslosigkeit über ein Jahr	Die Dauer der Arbeitslosigkeit wird aus der tagesge-nauen Differenz von Ende- und Beginndatum des Leis-tungsbezugs / der Arbeitslosigkeitsmeldung berechnet.
ASOZH	Anzahl Arbeitslose: Empfänger von Leistungen gem. SGB II	Basiert auf dem Merkmal Quelle. Gezählt werden Per-sonen, die als Quelle Leistungshistorik Grundsicherung aufweisen. Aufgrund der Einführung der zugelassenen kommuna-len Träger (Options- kommunen) im Jahr 2005 ist von einer deutlichen Untererfassung der Empfänger von Sozialhilfe im Jahr 2005 und einer Untererfassung im Jahr 2006 auszugehen.

Quelle: Eigene Darstellung.

3. Sozialdatenschutz und Anonymisierung

Ziel der im Folgenden beschriebenen Anonymisierung der Arbeitsmarktdaten ist es, einen Zugang zu aggregierten Mikrodaten nach §16(6) BstatG zur Nutzung außerhalb der Räumlichkeiten der Bundesagentur für Arbeit zu Zwecken der Arbeitsmarkt- und Berufsforschung im Rahmen des §282 SGB III gewähren zu können.

Aggregationsebene sind Postleitzahlbezirke. Diese sind flächenmäßig erheblich kleiner als die Kreis- oder Mikrozensusregionsebene, die bei den Mikrodaten des IAB sonst als Regionalinformation in den schwach anonymisierten bzw. faktisch anonymisierten Daten verfügbar sind und ebenfalls kleinräumiger als die von der Statistik der BA angebotenen Regionaldaten. Die Re-identifikation einzelner Personen wird allerdings dadurch erschwert, dass jede einzelne Variable aus den Mikrodaten aggregiert wurde und nur in Ausnahmefällen Aggregate aus Kombinationen von Merkmalen gebildet wurden. Ein Risiko bezüglich der Re-identifikation ergibt sich (im Gegensatz zu den Mikrodaten des IAB) weniger aus der Kombination von Merkmalen als aus dem Verschneiden mit anderen Datenquellen, die regionale Arbeitsmarktinformationen auf anderen kleinräumigen Gebietsständen unterhalb der Kreisebene enthalten. Deswegen ist eine (über die unten beschriebene Verknüpfung mit dem SOEP hinausgehende) Verknüpfung mit Personen- oder Haushaltsbezogenen Informationen auf einem anderen Gebietsstand als den Postleitzahlbezirken unter §282 SGB III nutzungsrechtlich ausgeschlossen (und höchstens mit einer Genehmigung für schwach anonymisierte Daten nach §75 SGB X zulässig). Deswegen ist eine Verknüpfung mit personen- oder haushaltsbezogenen Informationen auf einem anderen Gebietsstand als den Postleitzahlbezirken nur mit Daten des FDZ IAB, FDZ SOEP und des FDZ Ruhr unter bestimmten Voraussetzungen zulässig.

Um das Re-identifikationsrisiko zu minimieren, werden zunächst alle Informationen je Postleitzahlbezirk gesperrt (bzw. auf unbekannt gesetzt), hinter denen weniger als 20 Personen stehen (Primärspernung). Bei einigen Variablen muss dabei auch darauf geachtet werden, dass sich die entfernte Angabe nicht aus einer komplementären Menge und einem übergeordneten Aggregat herleiten lässt: bspw. aus der Anzahl aller Beschäftigten und der Anzahl beschäftigter Männer, wenn die Anzahl der beschäftigten Frauen weniger als 20 beträgt. In diesem Fall wird die komplementäre Variable ebenfalls als „unbekannt“ kodiert (Sekundärspernung); die Information für die übergeordnete Variable bleibt nach Möglichkeit erhalten. In ähnlicher Weise wird verfahren, wenn in einer Gruppe von Aggregaten (bspw. den Beschäftigten nach Altersklassen) eine einzige Kategorie gesperrt wurde: hier wird das nächst kleinste Aggregat ebenfalls gesperrt.

Unbekannte, nicht-vorhandene bzw. entfernte Informationen werden mit einem Wert „-9“ kodiert. Eine Unterscheidung zwischen Primär- und Sekundärspernung ist nicht möglich.

Die Zahl der Postleitzahlen mit nicht ausreichend Beobachtungen, so dass es zu einer Anonymisierung kommt, ist in Tabelle 2 dargestellt. Insgesamt gibt es zwischen 4 502 (2010) und 4 524 (2004) Postleitzahlen in den einzelnen Beobachtungsjahren.

Tabelle 2
Anonymisierte Beobachtungen 2004 – 2010

	2004	2005	2006	2007	2008	2009	2010
N	4 524	4 518	4 510	4 509	4 506	4 509	4 502
Beschäftigte							
BN	47 (1.0%)	41 (0.9%)	32 (0.7%)	31 (0.7%)	27 (0.6%)	30 (0.7%)	25 (0.6%)
BMANN	47 (1.0%)	42 (0.9%)	33 (0.7%)	34 (0.8%)	31 (0.7%)	34 (0.8%)	28 (0.6%)
BFRAU	47 (1.0%)	42 (0.9%)	33 (0.7%)	34 (0.8%)	31 (0.7%)	34 (0.8%)	28 (0.6%)
BALT20	127 (2.8%)	138 (3.1%)	126 (2.8%)	124 (2.8%)	112 (2.5%)	126 (2.8%)	150 (3.3%)
BALT25	70 (1.5%)	63 (1.4%)	50 (1.1%)	49 (1.1%)	48 (1.1%)	55 (1.2%)	59 (1.3%)
BALT30	62 (1.4%)	57 (1.3%)	54 (1.2%)	49 (1.1%)	46 (1.0%)	50 (1.1%)	53 (1.2%)
BALT35	55 (1.2%)	54 (1.2%)	50 (1.1%)	49 (1.1%)	49 (1.1%)	55 (1.2%)	56 (1.2%)
BALT40	51 (1.1%)	46 (1.0%)	39 (0.9%)	38 (0.8%)	37 (0.8%)	42 (0.9%)	38 (0.8%)
BALT45	51 (1.1%)	45 (1.0%)	38 (0.8%)	37 (0.8%)	34 (0.8%)	39 (0.9%)	34 (0.8%)
BALT50	51 (1.1%)	46 (1.0%)	36 (0.8%)	37 (0.8%)	34 (0.8%)	38 (0.8%)	34 (0.8%)
BALT55	53 (1.2%)	47 (1.0%)	39 (0.9%)	40 (0.9%)	34 (0.8%)	40 (0.9%)	35 (0.8%)
BALT60	75 (1.7%)	62 (1.4%)	50 (1.1%)	41 (0.9%)	38 (0.8%)	42 (0.9%)	39 (0.9%)
BALT65	122 (2.7%)	145 (3.2%)	129 (2.9%)	122 (2.7%)	106 (2.4%)	114 (2.5%)	109 (2.4%)
BDEUTSCH	785 (17.4%)	778 (17.2%)	735 (16.3%)	705 (15.6%)	694 (15.4%)	686 (15.2%)	656 (14.6%)
BAZUBI	750 (16.6%)	786 (17.4%)	837 (18.6%)	864 (19.2%)	899 (20.0%)	923 (20.5%)	961 (21.3%)
BUNGA	63 (1.4%)	59 (1.3%)	46 (1.0%)	43 (1.0%)	40 (0.9%)	45 (1.0%)	37 (0.8%)
BFACH	57 (1.3%)	54 (1.2%)	43 (1.0%)	43 (1.0%)	42 (0.9%)	45 (1.0%)	45 (1.0%)
BMEIST	780 (17.2%)	818 (18.1%)	859 (19.0%)	878 (19.5%)	917 (20.4%)	937 (20.8%)	976 (21.7%)
BANG	47 (1.0%)	44 (1.0%)	35 (0.8%)	33 (0.7%)	32 (0.7%)	35 (0.8%)	29 (0.6%)
BTEIL1	56 (1.2%)	52 (1.2%)	43 (1.0%)	41 (0.9%)	40 (0.9%)	41 (0.9%)	35 (0.8%)
BTEIL2	62 (1.4%)	53 (1.2%)	44 (1.0%)	42 (0.9%)	40 (0.9%)	42 (0.9%)	35 (0.8%)
BBILDUNK	51 (1.1%)	46 (1.0%)	38 (0.8%)	37 (0.8%)	35 (0.8%)	40 (0.9%)	34 (0.8%)
BHSHONE	53 (1.2%)	48 (1.1%)	40 (0.9%)	42 (0.9%)	40 (0.9%)	45 (1.0%)	41 (0.9%)
BHSMIT	47 (1.0%)	43 (1.0%)	34 (0.8%)	33 (0.7%)	30 (0.7%)	35 (0.8%)	28 (0.6%)
BABIOHNE	871 (19.3%)	843 (18.7%)	788 (17.5%)	748 (16.6%)	669 (14.8%)	682 (15.1%)	693 (15.4%)
BABIMIT	836 (18.5%)	811 (18.0%)	756 (16.8%)	723 (16.0%)	638 (14.2%)	657 (14.6%)	660 (14.7%)
BUNI	170 (3.8%)	161 (3.6%)	138 (3.1%)	127 (2.8%)	119 (2.6%)	112 (2.5%)	111 (2.5%)
BNARB	54 (1.2%)	49 (1.1%)	37 (0.8%)	39 (0.9%)	37 (0.8%)	41 (0.9%)	33 (0.7%)
BMARB	54 (1.2%)	46 (1.0%)	37 (0.8%)	40 (0.9%)	38 (0.8%)	42 (0.9%)	38 (0.8%)
BHARB	81 (1.8%)	77 (1.7%)	63 (1.4%)	66 (1.5%)	61 (1.4%)	67 (1.5%)	54 (1.2%)
BNANG	49 (1.1%)	44 (1.0%)	35 (0.8%)	37 (0.8%)	33 (0.7%)	38 (0.8%)	33 (0.7%)
BMANG	49 (1.1%)	45 (1.0%)	35 (0.8%)	35 (0.8%)	33 (0.7%)	38 (0.8%)	33 (0.7%)
BHANG	76 (1.7%)	72 (1.6%)	60 (1.3%)	62 (1.4%)	58 (1.3%)	62 (1.4%)	49 (1.1%)
Arbeitslose							
AN	66 (1.5%)	59 (1.3%)	57 (1.3%)	84 (1.9%)	118 (2.6%)	88 (2.0%)	102 (2.3%)
AMANN	194 (4.3%)	167 (3.7%)	195 (4.3%)	279 (6.2%)	360 (8.0%)	318 (7.1%)	350 (7.8%)
AFRAU	194 (4.3%)	167 (3.7%)	195 (4.3%)	279 (6.2%)	360 (8.0%)	318 (7.1%)	350 (7.8%)
AALT20	4 200 (92.8%)	1 929 (42.7%)	1 703 (37.8%)	2 042 (45.3%)	2 495 (55.4%)	2 910 (64.5%)	3 137 (69.7%)
AALT25	2 617 (57.8%)	1 199 (26.5%)	1 289 (28.6%)	1 684 (37.3%)	2 010 (44.6%)	1 806 (40.1%)	2 157 (47.9%)
AALT30	1 636 (36.2%)	1 087 (24.1%)	1 113 (24.7%)	1 291 (28.6%)	1 435 (31.8%)	1 187 (26.3%)	1 398 (31.1%)
AALT35	1 456 (32.2%)	1 199 (26.5%)	1 248 (27.7%)	1 450 (32.2%)	1 616 (35.9%)	1 521 (33.7%)	1 548 (34.4%)
AALT40	823 (18.2%)	845 (18.7%)	975 (21.6%)	1 203 (26.7%)	1 469 (32.6%)	1 619 (35.9%)	1 801 (40.0%)
AALT45	800 (17.7%)	777 (17.2%)	888 (19.7%)	1 042 (23.1%)	1 216 (27.0%)	1 149 (25.5%)	1 293 (28.7%)
AALT50	931 (20.6%)	884 (19.6%)	905 (20.1%)	1 078 (23.9%)	1 195 (26.5%)	1 067 (23.7%)	1 146 (25.5%)
AALT55	934 (20.6%)	856 (18.9%)	876 (19.4%)	1 057 (23.4%)	1 247 (27.7%)	1 224 (27.1%)	1 236 (27.5%)
AALT60	990 (21.9%)	779 (17.2%)	722 (16.0%)	882 (19.6%)	1 108 (24.6%)	1 190 (26.4%)	1 221 (27.1%)
ADEUTSCH	1 574 (34.8%)	1 438 (31.8%)	1 331 (29.5%)	1 502 (33.3%)	1 647 (36.6%)	1 631 (36.2%)	1 724 (38.3%)
ALOK1J	80 (1.8%)	87 (1.9%)	119 (2.6%)	198 (4.4%)	247 (5.5%)	144 (3.2%)	204 (4.5%)
ALOG1J	534 (11.8%)	239 (5.3%)	215 (4.8%)	290 (6.4%)	435 (9.7%)	484 (10.7%)	470 (10.4%)
ASOZH	4 524 (100.0%)	1 280 (28.3%)	1 142 (25.3%)	1 568 (34.8%)	1 956 (43.4%)	2 373 (52.6%)	2 413 (53.6%)

Quelle: Eigene Berechnungen; absolute Zahl und Anteil (in Klammern) der anonymisierten Postleitzahlen.

4. Datenqualität und Probleme

Eine allgemeine Übersicht zu Qualitätsproblemen in der IEB findet sich in vom Berge/Burghard/Trenkle (2013), S. 27ff. Die hier bedeutsamsten Probleme sollen aber gesondert benannt werden:

- In der zugrundeliegenden Version der IEB liegt der Füllgrad der BeH-Meldungen im Jahr 2007 letztmalig bei 100%. Für die Jahre 2008 und 2009 wurden 18-Monats-Dateien verwendet, für 2010 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass die BeH-Meldungen für die Jahre 2008 und 2009 leicht und für 2010 etwas stärker untererfasst sind. Die Auswertbarkeit der Daten auf Personenebene sollte dadurch aber nicht eingeschränkt sein.
- In den Jahren 2005 und 2006 sind LHG und XLHG nicht vollständig gefüllt. Auch ab 2007 kommt es hauptsächlich in der XLHG immer wieder zu meist einmonatiger trägerspezifischer Untererfassung. Auch die LHG weist in Zusammenhang mit Trägerartwechseln Unter-/Übererfassungen auf.
- Ab 2007 ist von einer Vollständigkeit der XASU-Meldungen von den 69 zugelassenen kommunalen Trägern der IEB auszugehen. Davor zeigen sich teils systematische Untererfassungen.
- Aufgrund der Hauptspell-Logik kommt es bei den Aggregaten zu einer leichten Untererfassung von Arbeitslosen (insbesondere aus dem Bereich des SGB II), etwa wenn diese einer (für Arbeitslose erlaubten) geringfügigen Beschäftigung nachgehen. Eine Überprüfung für einzelne Postleitzahlbezirke bzw. eine Aggregation der Postleitzahlbezirke zu Städten zeigte eine Unterschätzung der amtlichen Arbeitslosenzahlen in einer Größenordnung von 5 bis 15 Prozent.
- Es bestehen fehlerhafte oder inkonsistente Adressangaben. Die beim Arbeitgeber vermerkte bzw. von ihm im Rahmen des DEÜV-Verfahrens weitergegebene Adresse sollte zwar dem jeweils gültigen Stand entsprechen. Es kann jedoch geschehen, dass Informationen über Umzüge nicht vermerkt werden und die gemeldete Adresse dem Stand bei Einstellung entspricht, oder dass bei einer Jahresmeldung zur Beschäftigung die Adressinformation dem Stand am Jahresende entspricht, nicht aber dem zur Jahresmitte. Eine weitere mögliche Fehlerquelle in der Postleitzahlinformation kann auch in unterschiedlicher Aktualität der verschiedenen Datenbanken der Bundesagentur für Arbeit oder in einer nicht durchgeführten Gebietsstandsbereinigung bei den Postleitzahlbezirken liegen. In ersten Analysen wurde für ca. 2 % der Personen eine Diskrepanz zwischen Wohnortskreis und der in den Adressdaten zum jeweiligen Stichtag geführten Postleitzahl beobachtet (d.h. einer Person ist eine PLZ zugeordnet, die nicht für den jeweiligen in der IEB berichteten Kreis gültig ist).

5. Zusammenspielen mit SOEP und weitere Verknüpfungspotentiale

Das ursprüngliche Ziel der Bereitstellung der aggregierten Daten aus der IEB war es, diese auf der Ebene der Postleitzahlen an das SOEP heranzuspielen, um diese als zusätzliche Variablen für die jeweilige Nachbarschaft zu nutzen. Das Sozio-oekonomische Panel ist eine repräsentative Befragung privater Haushalte in Deutschland (siehe Wagner et al. 2007). Seit 1984 werden jedes Jahr dieselben Personen in denselben Haushalten zu ihrer Lebenssituation befragt. Die Befragten werden unter anderem zu ihrer Gesundheit, ihrem Einkommen und ihrer Bildung interviewt. Die Stichprobe umfasst ca. 10 000 Haushalte und 20 000 Personen.

5.1 Regionalanalysen mit dem SOEP

Das SOEP bietet regionale Analysepotentiale auf verschiedenen Ebenen. Es werden zwei Arten regionaler Informationen zur Verfügung gestellt (Goebel 2013). Zum einen Charakteristika des Wohnumfelds

der Befragten, welche entweder direkt aus der Befragung der Personen gewonnen werden oder zugespielt werden aus anderen Datenquellen¹. Zum anderen die aktuellen, offiziellen Schlüssel der regionalen Einheiten auf Basis derer externe Daten an das SOEP angespielt werden können. Folgende regionale Schlüssel sind im SOEP verfügbar: europäische NUTS Klassifikation, Bundesland und Regierungsbezirk seit 1984, Kreiskennziffer und Raumordnungsregion seit 1985, Gemeindeschlüssel und Postleitzahl seit 1993. Seit 2000 verfügt das SOEP außerdem über die Straßenabschnittskordinaten der befragten Haushalte. Diese unterliegen einem strengen Datenschutzkonzept das unter anderem festlegt, dass die SOEP Nutzer niemals gleichzeitig auf die SOEP Erhebungsdaten und die Koordinaten der Haushalte zugreifen können und die Daten immer nur anonymisiert bereitgestellt werden. Der Vorteil der Straßenabschnittskordinaten gegenüber anderer administrativer Einheiten ist, dass die Nutzer Nachbarschaften in beliebiger Größe definieren können. Außerdem sind sie nicht betroffen von Änderungen in den administrativen Verwaltungseinheiten (Goebel et al. 2010). Die Datenschutzbestimmungen denen die Nutzung der verschiedenen regionalen Einheiten im SOEP unterliegt sind hierarchisch, d.h. sie werden umso strenger je genauer der Wohnort der Befragten eingegrenzt wird.

Da die Zahl der SOEP Haushalte in tiefer gegliederten Regionen zu gering ist, kann das SOEP nicht dazu genutzt werden einzelne Regionen zu analysieren und zu vergleichen. Statistisch aussagekräftige Analysen dieser Art sind nur auf Bundesland Ebene möglich. Jedoch ermöglichen die tief gegliederten Regionalschlüssel die im SOEP zur Verfügung stehen eine Verknüpfung mit externen Regionalindikatoren sowie die Klassifizierung von ähnlichen Regionen (Goebel 2013).

Für das vorliegende Projekt wurden die Individualdaten des SOEP mit Informationen aus der IEB auf Postleitzahlebene verknüpft. Seit 1993 werden die PLZ der SOEP Befragten in jedem Erhebungsjahr erfasst (Knies/Spieß 2007). In diesem Jahr wurde in Deutschland das fünfstellige PLZ System eingeführt. Von den annähernd 8 300 PLZ in Deutschland² sind 2012 ca. 64% im SOEP vertreten. Im Jahr 2012 wohnen durchschnittlich 17 SOEP Haushalte in einer PLZ, maximal befinden sich 72 Haushalte in einem PLZ Gebiet. Aus datenschutzrechtlichen Gründen kann mit den PLZ nur im Rahmen eines Gastaufenthaltes im FDZ des SOEP am DIW Berlin gearbeitet werden.

5.2 Weiteres Verknüpfungspotenzial

Die erzeugten Daten lassen sich selbst mit anderen Daten auf PLZ Ebene ergänzen. Außerdem ist es möglich, das SOEP nicht nur um die oben beschriebenen Daten, sondern auch um weitere Daten auf PLZ-Ebene zu ergänzen. Für die Analyse von Nachbarschaftseffekten sind Regionalinformationen auf möglichst kleinräumiger Ebene entscheidend. Daher wurde eine Erweiterung des SOEP um kleinräumige Regionalindikatoren aus externen Datenquellen geprüft. Ziel ist es, die sozio-ökonomische und demographische Struktur sowie die Infrastruktur der Nachbarschaft der SOEP Befragten möglichst umfassend abzubilden. Dies ist für die Mehrheit der SOEP Nutzer bisher nur auf höherer regionaler Ebene, also Bundesland oder Kreisebene möglich. Um jedoch den Einfluss der Nachbarschaft auf das Verhalten von Individuen zu untersuchen ist es wichtig die Nachbarschaft so genau wie möglich, d.h. so kleinräumig wie möglich, abzubilden. Hierfür wurde die Postleitzahlebene gewählt.

¹ Das Mikromarketing Unternehmen „Microm“ stellt kleinräumige Indikatoren zur Verfügung die mit dem SOEP im sogenannten „Microm-SOEP-Datensatz“ verknüpft werden. Für eine Dokumentation dieser Daten siehe Goebel et al. (2007).

² Quelle: http://fa-technik.adfc.de/code/opengeodb/PLZ_2012-10.tab, 25.03.14

Auch das Angebot an Charakteristika der Nachbarschaft auf Koordinatenbasis ist sehr begrenzt. Im Folgenden eine Zusammenfassung der Ergebnisse der Datenrecherche auf PLZ Ebene oder kleiner mit Angabe der Quelle und wenn bekannt der Kosten (Stand 2012).

Das SOEP deckt jedoch nicht alle Postleitzahlengebiete vollständig ab. Des Weiteren sind die befragten Haushalte unterschiedlich über die Postleitzahlen verteilt. Entsprechend variiert der Anteil der Missings aufgrund von Datenschutzgründen beim Zusammenspielen mit dem SOEP auch zu den kompletten IEB-Daten. Tabelle 3 zeigt die Zahl der Haushalte auf, für die aufgrund der Anonymisierung der IEB Aggregate keine Information zu den Arbeitsmarktgrößen vorhanden sind. Es lässt sich erkennen, dass die Zahlen und insbesondere die Anteile für die meisten Variablen sehr klein sind.

Tabelle 3
Weitere Daten auf kleinräumiger Ebene

	Beschreibung	Ebene	Quelle	Kosten	Zugang
Benzinpreise	Benzinpreise mit Adressen der Tankstellen seit September 2013	Koordinaten, PLZ, Quadratkilometer und größere Ebenen	Markttransparenzstelle für Kraftstoff	kostenfrei	FDZ Ruhr am RWI
Immissionsdaten	Ca. 650 Messstationen der Bundesländer und des Bundesumweltamtes erheben bis zu halbstündlich Schadstoffe in der Luft - Feinstaub: ab 2001 - Kohlendioxid: ab 2001 - Stickstoffdioxid: ab 2001 - Ozon: ab 2001 - Schwefeldioxid: ab 2001	Koordinaten der Messstationen	Bundesumweltamt	kostenfrei	Können beim Umweltbundesamt bestellt werden.
Immobilienpreisindizes	Immobilienpreisindizes auf Grundlage hedonsicher Preisfunktionen der ImmobilienScout 24-Daten (Angebotspreise) seit 2007	PLZ, Quadratkilometer, Arbeitsmarktregion etc. individuell	ImmobilienScout24	Bearbeitungsgebühr	FDZ Ruhr am RWI
Microm – Indikatoren FDZ SOEP	- Typus der Wohngegend - Kaufkraft - Sozialstrukturelle Informationen - PKWs - Konsumverhalten - Umzüge - Sinus Milieus - (siehe Anhang für eine Liste an Variablen)	Haus seit 2004; Straßenabschnitt, Marktzelle, PLZ8 seit 2010	Microm	Kostenfrei bei der Zuspierung zum SOEP am FDZ SOEP (nur SOEP Haushalte)	SOEP

Microm – Indikatoren FDZ Ruhr	2003, 2005, 2009, 2010, weitere Wellen geplant - Wohngebäude - Haushalte - Kaufkraft - bebauung - PKWs - Ausländeranteil - Haushaltsstruktur - Arbeitslosenquote - Bevölkerung nach Alter und Geschlecht - Ethnische Zusammensetzung - Zahlungsindex	Quadratkilometer	Microm	Kosten zwischen 50 und 500 € je nach Zahl der Variablen (komplett Deutschland)	FDZ Ruhr
Migration	WebGIS des BA für Migration und Flüchtlinge weist PLZ die nächstgelegenen - Integrationsprojekte - Migrationsberatung - Sportverein - Integrationsangebot im Sport zu	Adressen → Koordinaten	BA f. Migration und Flüchtlinge		Muss angefragt werden
Risikoindex	Statistische Wahrscheinlichkeit von Zahlungsausfällen errechnet auf Basis der Negativmerkmale der Creditreform - Zuordnung eines Hauses zu einer von 100 Risikoklassen	Häuserebene	Microm	Kostenfrei	Können auf Anfrage bei Microm angespielt werden
Schuldnerquoten	2004-2010 Anteil der Personen mit Negativmerkmalen (Gerichtsmerkmale, Mahnverfahren etc.) im Verhältnis zu allen Personen ab 18 - Harte Schuldnerquoten (Fälle die vor Gericht kommen) - Weiche Schuldnerquoten (≥ 3 Inkassomerkmale) - Einteilung in Schuldnerklassen - Personen über 18	PLZ Kreis	Creditreform	Kostenfrei für wissenschaftliche Zwecke	SOEP

Schulen	- Adressen aller Schulen nach Bundesländern - Unterschiedliche Jahre - Unterschiedliche Merkmale: Schülerzahlen, Geschlecht, Nationalität, Träger...	Adressen	Statistische Landesämter	~ 1100 € Kostenfrei für: Berlin Bremen Meck.-Pom. Rheinland-Pfalz Thüringen	Auf Anfrage bei den statistischen Landesämtern
Unternehmensstatistik	Zusammenfassung Unternehmen aus Handelsregister und Nichthandelsregister - Wirtschaftszweig - Mitarbeiter -	Quadratkilometer, PLZ oder höher möglich	Databyte	Preis auf Anfrage	FDZ Ruhr
Nach eigenen Recherchen, sind die folgenden Faktoren auf PLZ Ebene nicht verfügbar:	Frauenerwerbsquote				
	Fertilität				
	Kinderbetreuung				
	DJI Regionaldatenbank				
	FOCUS Schulkompass				
	Ausländerzentralregister				
	Regionaldatenbank Deutschland				

Quelle: Eigene Recherche; die Autoren erheben keinen Anspruch auf Vollständigkeit.

Tabelle 4

Anonymisierte PLZ nach dem Zusammenspielen mit dem SOEP

	2004	2005	2006	2007	2008	2009
Beschäftigte						
BN	27	23	22	22	23	28
BMANN	27	23	22	22	23	29
BFRAU	27	23	22	22	23	29
BALT20	199	198	211	167	144	250
BALT25	165	173	174	114	106	222
BALT30	166	172	182	118	115	226
BALT35	158	166	175	118	109	225
BALT40	155	159	165	105	100	218
BALT45	154	158	164	104	96	214
BALT50	154	158	159	104	95	212
BALT55	156	158	164	105	95	214
BALT60	172	162	168	106	100	215
BALT65	197	201	210	166	143	249
BDEUTSCH	1791	1726	2001	1609	1515	1895
BAZUBI	1571	1620	2066	1765	1779	2283
BUNGA	1528	1577	2027	1729	1746	2247
BFACH	1528	1575	2023	1726	1743	2246
BMEIST	1571	1620	2066	1765	1779	2285
BANG	1525	1574	2022	1722	1739	2240
BTEIL1	1528	1575	2027	1729	1746	2246
BTEIL2	1527	1574	2027	1726	1743	2246
BBILDUNK	1250	1204	1351	1129	941	1168
BHSHNE	1254	1206	1356	1133	946	1173
BHSMIT	1248	1203	1350	1126	938	1162
BABIOHNE	1706	1640	1814	1505	1262	1512
BABIMIT	1691	1635	1808	1503	1255	1503
BUNI	1404	1332	1464	1220	1019	1273
BNARB	75	75	65	74	66	109
BMARB	75	75	65	74	66	109
BHARB	76	76	71	77	66	112
BNANG	74	75	64	74	61	108
BMANG	74	75	64	72	60	105
BHANG	76	75	70	75	65	111

Arbeitslose						
AN	218	129	147	173	215	245
AMANN	274	193	209	274	365	346
AFRAU	274	193	209	274	365	346
AALT20	12614	3849	2994	2700	2771	3537
AALT25	11992	3397	2562	2523	2667	2867
AALT30	12007	3474	2485	2384	2543	2715
AALT35	11828	3517	2684	2572	2689	3049
AALT40	11357	3194	2394	2325	2594	3133
AALT45	11290	3152	2296	2122	2415	2667
AALT50	11502	3239	2316	2182	2391	2570
AALT55	11463	3200	2310	2200	2450	2833
AALT60	11271	3044	1951	1884	2158	2695
ADEUTSCH	3474	2891	3213	2966	2953	3485
ALOO	14154	13857	17423	14379	13625	17853
ALOK1J	88	97	143	212	309	230
ALOG1J	911	249	274	278	293	360
ASOZH	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Quelle: Eigene Berechnung; Zahl der Haushalte mit fehlender Information aufgrund von Anonymisierung

6 Datenzugang

Der Datenzugang erfolgt nur durch Gastaufenthalt bzw. Fernverarbeitung in einem der beteiligten FDZ (FDZ IAB, FDZ Ruhr oder FDZ SOEP). Dort können die Daten mit den Daten des jeweiligen FDZ verknüpft werden.

Für die Nutzung des Datensatzes ist ein entsprechender Nutzungsantrag bei dem jeweiligen FDZ zu stellen. Der Nutzungsantrag und entsprechende Nutzungsbedingungen können bei dem jeweiligen FDZ erfragt werden.

7. Literaturverzeichnis

Bender, Stefan; Hilzendegen, Jürgen; Rohwer, Götz; Rudolph, Helmut (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, IAB, Nürnberg.

vom Berge, Philipp; Burghardt, Anja; Trenkle, Simon; (2013): Stichprobe der Integrierten Arbeitsmarktbio grafien - Regionalfile 1975-2010 (SIAB-R 7510). FDZ-Datenreport 09/2013, IAB, Nürnberg.

Blossfeld, Hans-Peter (1985): Berufseintritt und Berufsverlauf - Eine Kohortenanalyse über die Bedeutung des ersten Berufs in der Erwerbsbiographie. In: MittAB 2, 177-197.

Goebel, Jan (2013): Regionalisierungsmöglichkeiten des Sozio-oekonomischen Panels (SOEP), in: Regionale Standards. GESIS Schriftenreihe Band 12. Hg.: Arbeitsgruppe Regionale Standards.

Goebel Jan; Spieß C. Katharina; Witte Nils; Gerstenberg Susanne (2007): Die Verknüpfung des SOEP mit MICROM-Indikatoren: Der MICROM-SOEP Datensatz, Data Documentation 26, DIW Berlin.

Goebel Jan; Wagner Gert G.; Wurm Michael (2010). Exemplarische Integration raumrelevanter Indikatoren auf Basis von "Fernerkundungsdaten" in das Sozio-oekonomische Panel (SOEP), SOEPpaper 267, DIW Berlin.

IAB (2012), Hrsg.: Datenprodukte des IAB - Katalog des Geschäftsbereichs IT- und Informationsmanagement. Version 4.00, Nürnberg

Knies, Gundi; Spieß C. Katharina (2007): Regional Data in the German Socio-Economic Panel Study (SOEP), Data Documentation 17, DIW Berlin.

Wagner Gert G.; Frick Joachim; Schupp Jürgen (2007). The German socio-economic panel study (SOEP): scope, evolution and enhancements. Schmollers Jahrbuch 127(1):139–169.

Wermter, Winfried; Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt – und Berufsforschung 4/88, S. 468-482.