

POTENTIALE UND ANWENDUNGEN GEOREFERENZIERTER DATEN

28. Wissenschaftliches Kolloquium

gemeinsam mit der Deutschen Statistischen Gesellschaft am 14. und 15. November 2019 in Bonn

Kurzfassung:

Verbindung von Surveydaten mit Geodaten: Möglichkeiten, Mehrwert und Probleme am Beispiel des SOEP

Dr. Jan Göbel

studierte Soziologie und Volkswirtschaftslehre an der Johannes Gutenberg-Universität Mainz und der Technischen Universität Berlin. Er promovierte 2007 ebenfalls an der TU Berlin in Wirtschaftswissenschaften. Herr Dr. Göbel ist seit 2001 Mitarbeiter am DIW Berlin in der Abteilung Sozio-oekonomisches Panel (SOEP). Nachdem er 2012 die Leitung des Forschungsdatenzentrums des SOEP übernahm, ist er seit 2018 auch Mitglied im neugeschaffenen Direktorium des SOEP. Die Schwerpunkte seiner wissenschaftlichen Arbeit sind im Bereich Einkommensverteilung und Ungleichheit, insbesondere auch in Hinsicht regionaler Ausprägungen.

Der Beitrag zeigt auf, welche zusätzlichen Analysemöglichkeiten in der Verbindung von Daten aus Bevölkerungsumfragen und Geodaten entstehen. Aber auch welche neuen Herausforderungen durch die Verbindung dieser beiden Datenquellen zu beachten sind.

Eine zentrale Herausforderung im Umgang und insbesondere in der Bereitstellung von georeferenzierten Befragungsdaten ist der zu garantierende Datenschutz gegenüber den Befragten. Am Forschungsdatenzentrum des SOEP ist bereits seit mehreren Jahren eine erprobte und erfolgreiche Infrastruktur in Betrieb, die es Wissenschaftlern ermöglicht, solche Daten zu nutzen. Zentraler Teil des entwickelten Datenschutzkonzeptes ist dabei, dass die Geo-Koordinaten der SOEP-Haushalte von den Befragungsdaten der Personen und Haushalte grundsätzlich getrennt gehalten werden. Forschende haben zu keinem Zeitpunkt gleichzeitig Zugriff auf Koordinate und Befragungsinformation. Für die allermeisten Analyseszenarien ist dies auch nicht notwendig. Gegenwärtig werden aber auch Anstrengungen unternommen, um die Hürden zur Nutzung dieser Daten für Wissenschaftler zu verringern, bei gleichzeitiger Aufrechterhaltung datenschutzrechtlicher Standards. Die bisher am SOEP bereitgestellte Infrastruktur erlaubt es Forschenden zwar die Daten zu nutzen, ist aber sehr voraussetzungsintensiv. So müssen die Forschenden das nötige Wissen besitzen, wie sie mit Geodaten in einem GIS System arbeiten können und sie müssen ihre Analysen in zwei aufeinander folgenden Schritten aufteilen. Auf der Seite des Datenanbieters ist der jeweils anfallende Schritt der manuellen Anspielung und der Prüfung von neu erstellten Indikatoren sehr zeitaufwendig. Beides stellt eine deutliche Einschränkung dar und soll im Rahmen eines DFG-Projektes "Sozial-Raumwissenschaftliche Forschungsdateninfrastruktur" (SoRa) in Kooperation mit GESIS – Leibniz-Institut für Sozialwissenschaften, dem Karlsruher Institut für Technologie (KIT) und dem Leibniz-Institut für ökologische Raumentwicklung (IÖR) verbessert werden.

Im Hauptteil des Vortrags wird anhand beispielhafter Analysen aufgezeigt, welche neuen Analysemöglichkeiten mit den vorhandenen Daten bereits umgesetzt werden konnten. Eine wissenschaftliche Nutzung von georeferenzierten Surveydaten kann auf unterschiedlichster Art und Weise erfolgen. Grundlegend ist jedoch, dass jede Art von geokodierter Information

mit der Lage des Wohnortes in Beziehung gebracht werden kann. Im einfachsten Fall ist dies lediglich die Berechnung einer einfachen Distanz des Wohnortes zu einem oder mehreren anderen georeferenzierten Punkten. Mit Hilfe von georeferenzierten Wohnorten ist es jedoch möglich, für jeden Haushalt quasi "individuelle Nachbarschaften" zum Beispiel mit einem variablen Radius zu bestimmen. Dadurch kann nicht nur der Indikator der jeweiligen Region, in der ein Haushalt wohnt, berücksichtigt werden, sondern auch ob der Haushalt an der Grenze dieser Region liegt. Die Werte der benachbarten Regionen können dann über eine Distanzfunktion gewichtet in die Beschreibung der sogenannten egozentrierten Nachbarschaft eingehen. Über eine Variation des Radius kann sich empirisch dem bisher recht unbestimmten Konzept von Nachbarschaft angenähert werden.

Die empirische Sozialforschung kann aber vom Zusammenspiel von Befragungs- und Geodaten nicht nur dadurch profitieren, dass bereits vorhandene Befragungsdaten georeferenziert werden, auch für die Erhebung von Befragungsdaten können verfügbare Geodaten nützlich sein. Das SOEP hat im Jahr 2018 erstmals eine Stichprobe aus den amtlichen Hauskoordinaten des Bundesamtes für Kartographie und Geodäsie (BKG) gezogen. Der Vortrag wird daher auch kurz darstellen, wie aus der Verbindung von kleinräumigen Zensusdaten und Informationen des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) eine Stichprobe von Privathaushalten in Gebieten des Städtebauförderprogramms "Soziale Stadt" gezogen werden konnte.