

Zensusdatenangebot der Forschungsdatenzentren der Statistischen Ämter

Reicher Datenfundus für die Wissenschaft

Joachim Werner

Die Forschungsdatenzentren des Bundes und der Länder bieten der Wissenschaft rund 100 Statistiken zur Nutzung von Mikrodaten für die vielfältigsten Auswertungszwecke. Seit Ende des Jahres 2015 wurde dieses Datenangebot um die ersten beiden Produkte des umfangreichen und sehr differenzierten Datenmaterials des Zensus 2011 ergänzt. Mit dem Produkt 3, dem Gesamtdatensatz des Zensus, wurde nunmehr das Einzeldatenangebot komplettiert.

Mit den verfügbaren Mikrodaten des Zensus 2011 steht für die Forschung ein außerordentlich großes Datenangebot zu vielfältigen Themen bereit. Das breite Spektrum an Mikrodaten des registergestützt durchgeführten Zensus speist sich aus verschiedenen Verwaltungsregistern wie den Daten aus Erwerbsregistern und Melde-registern. Um die nicht in Verwaltungsregistern vorliegenden Informationen bereitzustellen, wurden im Rahmen des Zensus 2011 ergänzend Erhebungen zu Gebäude- und Wohnungsdaten und eine Stichprobe bei bundesweit rund 10 % aller Haushalte durchgeführt. Im Rahmen der Haushaltsstichprobe wurden unter anderem Angaben zur Bildungssituation und dem Erwerbs-leben erfasst.¹

Mikrodaten der Haushaltsstichprobe und der Gebäude- und Wohnungszählung

Das Mikrodatenangebot des Zensus 2011 zu Produkt 1, der Haushaltsstichprobe, umfasst Daten auf Personenebene für rund 10 % der Bevölkerung. Die hochgerechneten Daten sind repräsentativ für die Gemeinden, die mindestens 10 000 Einwohner haben, und auf Kreisebene können durchgehend repräsentative Ergebnisse erstellt werden. Die Haushaltsstichprobe enthält zahlreiche Merkmale zu den Bereichen Demografie, Religion, Zuwanderung und Migration sowie zu den Themen Bildung, Ausbildung, Erwerbstätigkeit und Arbeitssuche. Insgesamt umfasst das Zensus-Produkt 1 rund 7,6 Mill. Fälle mit 104 Merkmalen pro Datensatz.

Die Mikrodaten der Gebäude- und Wohnungszählung (GWZ) des Zensus 2011 mit über 40 Mill. Datensätzen zu dem Produkt 2 beruhen auf einer flächendeckenden und vollzähligen Erfassung aller Gebäude mit Wohnraum und bewohnten Unterküften sowie der in den Gebäuden/Unterküften enthaltenen Wohnungen. Die erhobenen Merkmale beziehen sich auf folgende Inhalte: Baujahr, Eigentumsform, Gebäudetyp, Zahl der Wohnungen im Gebäude, Nutzung der Wohnung (zum Beispiel vermietet), Wohnfläche sowie Aus-



Lic. rer. reg. Joachim Werner ist Leiter der Abteilung „Wirtschaftswissenschaftliche Analysen, Umweltbeobachtung, Agrarwirtschaft“ im Statistischen Landesamt Baden-Württemberg.

¹ Ausführliche Erläuterungen zum Datenangebot des Zensus, auch zur Methode des Zensus, definitorischen Abgrenzungen und Metadaten, sind zu finden unter: www.forschungsdatenzentrum.de/be-stand/zensus_2011/index.asp (Abruf: 24.06.2016).

Ü Zugangswege zu den FDZ-Produkten des Zensus 2011

Produkt	Charakteristik	Zugangswege	
		Gastwissenschaftler-arbeitsplatz	Kontrollierte Datenfern-verarbeitung
Produkt 1 – Haushaltsstichprobe	standardisiertes Produkt	Ergebniserstellung nur über KDFV verpflichtende Einarbeitung am GWAP	
Produkt 2 – Gebäude- und Wohnungszählung	standardisiertes Produkt		
Produkt 3 – Gesamtdatensatz	Produkt bestehend aus 11 Teiltabellen, die untereinander verknüpft werden können.		

stattung der Wohnung mit Bad und WC. Die tiefste regionale Gliederungsebene für diese Merkmale ist die Gemeinde.

Datennutzung am Gastwissenschaftlerarbeitsplatz und über die Kontrollierte Datenfernverarbeitung (KDFV)

Sowohl bei Produkt 1 (Haushaltsstichprobe) und bei dem Produkt 2 (Gebäude- und Wohnungszählung) handelt es sich um standardisierte Produkte. Für diese Produkte braucht deshalb im Rahmen des Nutzungsantrags keine Auswahl von Erhebungsmerkmalen beantragt werden. Die Auswertung der Daten des Zensus 2011 wird in einer kombinierten Nutzung von Gastwissenschaftlerarbeitsplatz (GWAP) und der Kontrollierten Datenfernverarbeitung (KDFV) angeboten (vgl. Übersicht). Hierbei ist mit der Datennutzung des Zensus 2011 die Verpflichtung verbunden, sich zunächst am Gastwissenschaftlerarbeitsplatz mit den Einzeldaten vertraut zu machen und gegebenenfalls Syntaxen zu erstellen. Die Mikrodaten am Gastwissenschaftlerarbeitsplatz sind faktisch anonymisiert. Am GWAP erstellte Ergebnisse dienen auch dazu, die Verwertbarkeit der Daten zu prüfen, wobei Ergebnisse dem Nutzer nicht zur Verfügung gestellt werden. Die letztendliche Ergebniserstellung erfolgt durch die KDFV. Um eine Deanonymisierung der Daten auszuschließen, erhält der Nutzer dabei keinen Zugang zu diesen Mikrodaten. Im Rahmen der KDFV wird dem Nutzer stattdessen ein Strukturdatensatz zur Verfügung gestellt, der dem Aufbau des formal anonymisierten Datensatzes entspricht, ohne allerdings inhaltliche Informationen aufzuweisen. Somit stehen dem Nutzer zwei Analysewege zur Verfügung. Zum einen die Syntaxerstellung am GWAP mit der Statistiksoftware, zum anderen die Anwendung des erwähnten Strukturdatensatzes. Mit den auf beiden Wegen erzeugten Syntaxen werden im Forschungsdatenzentrum auf der Basis des formal anonymisierten Datenmaterials die Ergebnisse erstellt und diese anschließend auf Geheimhaltung geprüft.²

Gesamtdatensatz für mannigfaltige Auswertungsziele

Der Gesamtdatensatz (Produkt 3) basiert auf dem Datenmaterial nach der Zusammenführung der einzelnen primärstatistischen und registergestützten Zensusbestandteile im Rahmen der Haushaltsgenerierung. Er enthält neben den Informationen der Produkte 1 und 2 weitere Angaben zu den Haushalts- und Familienstrukturen sowie zur Wohnsituation bestimmter Bevölkerungsgruppen. Der Gesamtdatenbestand bietet

für die unterschiedlichen Auswertungszwecke erweiterte Analysemöglichkeiten des Einzeldatenmaterials. Zur Verbesserung der Nutzerfreundlichkeit wurden im Hinblick auf die Nutzung des Merkmalsumfangs und der Vercodung von Merkmalen verschiedene Maßnahmen durchgeführt. Aufgrund der Größe und Komplexität kann das Datenmaterial des Produktes 3 nicht standardisiert bereitgestellt werden. Es können vielmehr Merkmale aus diesen elf Teiltabellen ausgewählt werden: Haushaltsstichprobe, Person, Generation Edge, Partner Edge, Haushalt, Familie-National, Familie-EU, Gebäude, Wohnung, Regionalinformation, Anschrift. Eine Verknüpfung der Teiltabellen ist dabei möglich. Die Ergebnisdarstellung kann bis auf die Ebene der Gemeinden erfolgen, und darüber hinaus können aufgrund der Zuordnung der Einzeldaten zu Gitterzellen (auch unterhalb der Gemeindeebene) eigene räumliche Abgrenzungen für spezielle Untersuchungsziele gebildet werden.³

Wie bei den Produkten 1 und 2 ist für die Nutzung des Gesamtdatenbestandes die Einarbeitung der Wissenschaftler und Wissenschaftlerinnen am Gastwissenschaftlerarbeitsplatz obligatorisch. Die Erstellung von Syntaxen und der Ablauf der Ergebniserstellung erfolgt analog zu den Produkten 1 und 2 wie oben bereits beschrieben. Das heißt die von den Forschern erstellten Syntaxen werden im FDZ auf das Einzeldatenmaterial angewandt und die anschließend vorliegenden Ergebnisse im FDZ, bevor sie dem Nutzer zur Verfügung gestellt werden, auf Geheimhaltung geprüft.

Der große Umfang des Datenmaterials, verbunden mit der hohen Anzahl von Merkmalen quer über ein breites Themenspektrum, das vielfach auch auf tiefer regionaler Ebene auswertbar ist, machen die Mikrodaten des Zensus zu einem reichen Fundus für die Wissenschaft. Bisher vorliegende Forschungsprojekte auf der Grundlage des Zensusmaterials zeigen bereits eine breite Forschungspalette. Sie reichen von sozialen Einflussfaktoren auf die Bildungsabschlüsse über den Einfluss des Humankapitals auf den individuellen und gesellschaftlichen Wohlstand, verschiedene Studien zur Integration von Einwanderern in Deutschland bis hin zu Analysen sozialer Rahmenbedingungen im Kontext zur Kinderarbeit. ■

Weitere Auskünfte erteilt
Joachim Werner, Telefon 0711/641-26 30,
Joachim.Werner@stala.bwl.de

2 Zum Ablauf der Datennutzung und zu den verschiedenen Anonymisierungsstufen in den Forschungsdatenzentren, der Geheimhaltungsprüfung sowie der Syntaxerstellung; siehe hierzu auch: Stefanie Raab/Claudia Meisdrock: FDZ – Arbeitspapier, Nr. 48, S. 15 ff.

3 Konzept zur Erstellung des Zensus-Produktes 3, S. 1 ff.