

Zugang der Forschung zu den Mikrodaten des ZENSUS 2011

Ausführungen zum Punkt ‚Datenzugang‘ in der Stellungnahme der Zensuskommission zum Merkmalskatalog des Zensusanordnungsgesetzes (ZensusG2011)

Heike Wirth/Gert G. Wagner

2. April 2009

Wie in der Stellungnahme der Zensuskommission zum Merkmalskatalog des Zensusanordnungsgesetzes¹ erläutert, stellen die Mikrodaten des Zensus 2011 eine wichtige Datenquelle für die Forschung dar. Eine wesentliche Voraussetzung für die Nutzung dieser Daten ist dabei allerdings ein an den Forschungsbedürfnissen orientierter Datenzugang. Dies bezieht sich sowohl auf den Zeitpunkt wie auch den Umfang der Datenbereitstellung. Diese Punkte sind bislang nicht geregelt und bieten damit einen erheblichen Gestaltungsspielraum, der sich für die Forschung in der Praxis als negativ erweisen kann. Die Zensuskommission erachtet es daher als wichtig, die grundlegenden Bedürfnisse der Forschung bereits zum jetzigen Zeitpunkt zu artikulieren, mit dem Ziel, innerhalb der datenschutzrechtlichen Rahmenbedingungen einen möglichst optimalen Datenzugang und optimale Datenqualität zu erzielen. Im Folgenden werden diese Punkte näher erläutert.

Der aktuelle Entwurf des ZensusG 2011² sieht keine explizite Regelung des Zugangs der Forschung zu Zensus-Mikrodaten vor. Da es sich beim Zensus um eine Bundesstatistik handelt, die dem Bundesstatistikgesetz unterliegt, wird der Datenzugang demnach unter den Randbedingungen des §16(6) BStatG erfolgen. Danach dürfen *“für die Durchführung wissenschaftlicher Vorhaben (...) vom Statistischen Bundesamt und den statistischen Ämtern der Länder Einzelangaben an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können (...)*”.

¹ Stellungnahme der Zensuskommission zum Merkmalskatalog im Kabinettsentwurf des Zensusanordnungsgesetz (ZensusG2011). Verabschiedet am 22. Januar 2009.

² Entwurf eines Gesetzes zur Anordnung des Zensus 2011 sowie zur Änderung von Statistikgesetzen.

Da §16(6) BStatG nur auf den ‚Grad‘ der Anonymisierung (faktische Anonymität) Bezug nimmt, sieht die Zensuskommission bei einer Datenweitergabe, die sich ausschließlich auf §16(6) BStatG stützt, insofern Probleme³ als

- (1) der Zeitpunkt des Zugangs bzw. der Bereitstellung von Zensus-Mikrodaten völlig offen ist;
- (2) unklar ist, welche Zensus-Erhebungsmerkmale für die Forschung verfügbar sein werden;
- (3) das Risiko besteht, dass – ohne eine spezifische Wissenschaftsklausel im ZensG 2011 – letztendlich nur das Zensus-‚Endprodukt‘ für Forschungsvorhaben verfügbar sein wird.

Mit Zensus-‚Endprodukt‘ sind die Daten gemeint, die nicht nur in Hinblick auf Über- und Untererfassungen korrigiert sind, sondern auch nur noch plausibilisierte Merkmale und im ungünstigsten Szenario nur noch die entsprechend dem EU-Pflichtkatalog harmonisierten Merkmale enthalten, die Ursprungsmerkmale hingegen nicht mehr enthalten sind.

Das *Interesse der Forschung* (das auch der *Qualitätskontrolle des Zensus durch und für die Ämter dient*) an einem zeitnahen Datenzugang und vollständigen Merkmalsumfang (Erhebungsmerkmale wie auch hieraus generierter Merkmale, z. B. ISCO, ISCED, ILO-Erwerbsstatus) ist sowohl inhaltlicher als auch methodischer Art.

Das inhaltliche Analysepotenzial der Zensusdaten ist durch die – von Seiten des Gesetzgebers – bewusste Konzentration des Erhebungsprogramms auf die EU-Pflichtmerkmale auf wenige thematische Schwerpunkte begrenzt. Dabei ist zu berücksichtigen, dass der Zensus primär der Ermittlung amtlicher Einwohnerzahlen und sekundär der Bereitstellung von Grundinformationen für die politische Planung dient. Er kann und soll nicht die thematische Breite von Bevölkerungserhebungen der amtlichen Statistik oder der Wissenschaft aufweisen. Die inhaltliche Relevanz des Zensus ergibt sich daher vor allem aufgrund des Datenumfangs (Vollerhebung bei Registerdaten, relativ großer Auswahlatz bei der Haushaltsstichprobe), der sowohl für spezifische und kleine Subpopulationen relativ präzise Schätzungen ermöglicht, wie auch tief regionalisierte Analysen zulässt.

Der Zensus dient auch der Übermittlung von sozialstrukturellen Grunddaten an Eurostat. Zu diesem Zweck müssen Originalmerkmale z. T. entsprechend der Vorgaben von Eurostat in international harmonisierte Merkmale (z. B. ISCO, ISCED,

³ Die konkrete Umsetzung der faktischen Anonymisierung sowie der technischen und organisatorischen Randbedingungen der Datenbereitstellung werden im Folgenden ausgeklammert, da die Zensuskommission bereits empfohlen hat, hierfür Mittel für ein entsprechendes Forschungsprojekt bereitzustellen (Ergebnisniederschrift der Sitzung der Zensuskommission am 22. Januar 2009).

ILO-Erwerbsstatus) transformiert werden. Diese von Eurostat geforderte Harmonisierung ist zentral für den europäischen und internationalen Vergleich, allerdings wird sie den nationalen Randbedingungen nicht immer gerecht, d. h. *einige der nationalen Original-Merkmale werden aussagekräftiger als die EU-harmonisierten Daten sein*. Für die Untersuchung einer Vielzahl von Forschungsfragen ist es aber unabdingbar, dass nicht nur die harmonisierten ‚Exportvariablen‘, sondern weiterhin auch die Originalmerkmale verfügbar sind. Nur so bleibt die Analyseflexibilität der Daten erhalten.

Bedingt durch die vorrangige Zielsetzung des Zensus auf die ‚Festlegung gerichtsfester Einwohnerzahlen‘ wird die Datenqualität von Seiten des Gesetzgebers anhand des Kriteriums einer möglichst präzisen Schätzung der Bevölkerungszahl gemessen. Aus der wissenschaftlichen Perspektive ergibt sich durch das Modell des registergestützten Zensus eine ergänzende Sichtweise von Datenqualität. Denn die Kombination von Registerdaten und Primärerhebungen (Gebäude- und Wohnungszählung, Haushaltsstichprobe; Erhebung in Sonderbereichen) liefert nicht nur Erkenntnisse hinsichtlich Über- und Untererfassung, sondern es bestehen auch partielle Überschneidungen in den Erhebungsmerkmalen von Registerdaten und Primärerhebungen (z. B. Ausbildung, Schulbildung, ausgeübter Beruf, Wirtschaftszweig, Familienstand, Staatsangehörigkeiten).⁴ Über den systematischen Vergleich dieser Überschneidungsmerkmale lassen sich weitergehende Erkenntnisse über das Ausmaß und die Richtung von Abweichungen zwischen Register- und Primärdaten gewinnen. Diese sind sowohl für *zukünftige registergestützte Zensen* wie auch für die *Survey-Methodenforschung* von grundsätzlichem Interesse, da sich hieraus Anhaltspunkte für eine verbesserte Datenerhebung sowohl bei Registern wie auch Primärerhebungen ergeben.

Ein typisches und aktuelles Beispiel, für welche die Zensusdaten eine hervorragend geeignete Datenbasis bieten würden, wäre etwa der Vergleich des von Seiten der Arbeitgeber an die Bundesagentur für Arbeit gemeldeten ‚Ausbildungsniveaus‘ der Beschäftigten mit den Selbstauskünften zur Schul- und Berufsausbildung aus der Haushaltsstichprobe. Dies insbesondere deshalb, weil den Registerdaten der BA in den empirischen Wirtschaftswissenschaften eine herausgehobene Bedeutung zukommt, auch in Hinblick auf politische Planung, die Qualität des Merkmals

⁴ **Exkurs zu §7 Absatz 1 Satz 3 ZensG 2011:** „Die Feststellung umfasst nicht die Berichtigung der aus den Melderegistern übernommenen Angaben zur Person, wie die Angabe ‚üblicher Aufenthaltsort‘ “. Nach dieser Festlegung haben die in Melderegistern vorliegenden Informationen Vorrang vor den primär erhobenen Angaben der Haushaltsstichprobe. Soweit sich dies auf das Merkmal üblicher Aufenthaltsort bezieht, ist der Vorrang des Melderegisters nachvollziehbar. Da die Erhebungsbeauftragten einen verkürzten Melderegisterauszug (Namen, Geschlecht, Tag der Geburt, Staatsangehörigkeiten der Befragungsperson) erhalten, sollten die Abweichungen zwischen Melderegister und Primärerhebung eher gering ausfallen. Was aber ist, wenn offensichtliche Abweichungen (z. B. Geburtsjahr) auftreten? Welche Angabe hat dann Vorrang, bzw. werden solche Abweichungen in einem Zwischenschritt (Zusammengeführte Daten nach §9) dokumentiert und stehen für Forschungszwecke zur Verfügung?

„Ausbildungsniveau“ jedoch als durchaus verbesserungsfähig eingeschätzt wird.⁵ Eine unabdingbare Voraussetzung für diese Art der Methodenforschung ist jedoch, dass in den Zensusdaten (bzw. in einer Arbeitsversion) die vorhandenen Überschneidungsmerkmale erhalten bleiben und der Wissenschaft auch zugänglich sind (vgl. hierzu auch Fußnote 4). Gleichfalls von erhöhtem Interesse ist auch, dass die methodischen Variablen für die Haushaltegenerierung ebenso wie die so genannten Prozessmerkmale in den Daten verfügbar bleiben.

Bei den oben angeführten Punkten handelt es sich um Detailfragen, die sicherlich nicht Gegenstand gesetzlicher Regelungen sein können und sollten. Sie wurden hier nur beispielhaft skizziert, um zu verdeutlichen, dass das Interesse der Forschung an der Nutzung dieser Daten nicht auf das (anonymisierte) Zensus-„Endprodukt“ beschränkt ist. Vielmehr ergibt sich das eigentliche methodische und inhaltliche Forschungspotenzial dieser Daten (im Vergleich zu anderen Datenquellen) erst aus der Kombination der verschiedenen Erhebungsteile und ihrer partiellen Überschneidungen, weshalb diese zusammengeführten Daten (in anonymisierter Form) grundsätzlich für Forschungsvorhaben zur Verfügung stehen sollten.

Die Kommission ist sich nicht sicher, ob der Zugang zu einem solchen Zensus-„Zwischenprodukt“ im Rahmen von §16(6) BStatG möglich ist. Dies dürfte vermutlich davon abhängen, ob ein „Zensus-Zwischenprodukt“ bereits als Bundesstatistik interpretiert wird oder nicht. Um derartige Unwägbarkeiten, die den Datenzugang um Jahre verzögern können, auszuschließen, empfiehlt die Kommission daher die Aufnahme einer Forschungsklausel im ZensG 2011, die sicherstellt, dass die Forschung Zugang zu den aus den verschiedenen Erhebungsteilen zusammengeführten (anonymisierten) Zensus-Mikrodaten erhält. Die Beantragung und der Aufbau von thematischen und methodischen Zensus-Forschungsprojekten (z. B. bei der DFG) sind nur möglich, wenn der zur Verfügung stehende Merkmalskatalog zumindest in den Grundzügen klar ist. Die Zensuskommission erachtet es in diesem Zusammenhang gleichfalls als wichtig, dass der Datenzugang zu einem möglichst frühen Zeitpunkt gewährt wird, optimalerweise nach Abschluss der Datenaufbereitungsarbeiten.

⁵ Vgl. u. a.: Fitzenberger, Bernd, Aderonke Osikominu und Robert Völter (2005), Imputation Rules to Improve the Education Variable in the IAB Employment Subsample, ZEW Discussion Paper No. 05-10, Mannheim.