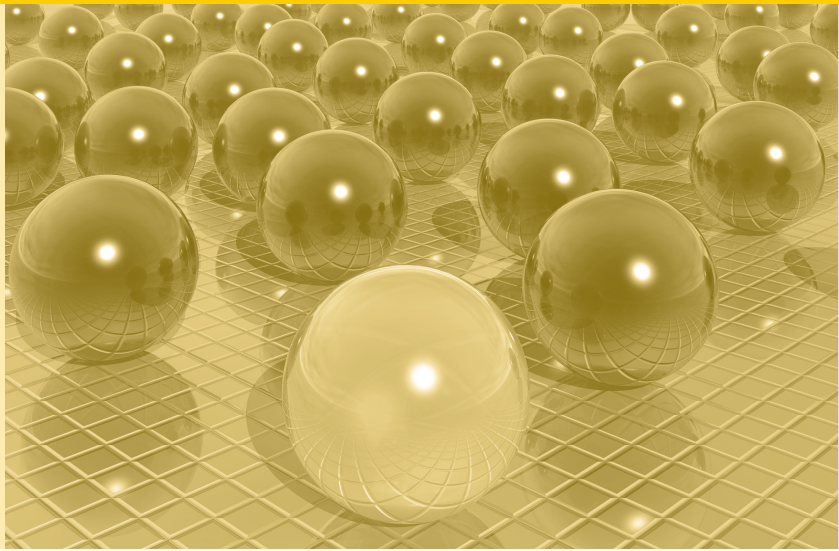


Metadatenreport



Produkt 3 - Gesamtdatensatz Zensus 2011

Impressum

Herausgeber: Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder
Herstellung: Information und Technik Nordrhein-Westfalen
Telefon 0211 9449-01 • Telefax 0211 442006
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@it.nrw.de

Fachliche Informationen

zu dieser Veröffentlichung:
Forschungsdatenzentrum der
Statistischen Ämter der Länder
– Standort Bayern –
Tel.: 0911 98208-185

Informationen zum Datenangebot

Forschungsdatenzentrum der
Statistischen Ämter der Länder
Tel.: 0211 9449-2873
Fax: 0211 9449-8087
E-Mail: forschungsdatenzentrum@it.nrw.de

Erschienen im Juni 2016

Diese Publikation wird kostenlos als **PDF-Datei** zum Download unter www.forschungsdatenzentrum.de angeboten.

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2016
(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, nur auszugsweise, mit Quellenangabe gestattet. Alle übrigen Rechte bleiben vorbehalten.

Inhalt

1.	Der Zensus 2011	5
1.1.	Rechtsgrundlagen.....	5
1.2.	Methode des Zensus 2011	6
2.	Methodik.....	8
2.1.	Anschriften- und Gebäuderegister.....	8
2.2.	Daten der Einwohnermeldeämter	9
2.3.	Erwerbsstatistische Registerdaten	9
2.4.	Haushaltebefragung.....	10
2.5.	Erhebungen an Adressen mit Sonderbereichen.....	11
2.6.	Befragung zur Klärung von Unstimmigkeiten.....	11
2.7.	Prüfung von Mehrfachfällen	12
2.8.	Zusammenführung der Daten aus den verschiedenen Datenquellen im Referenzdatenbestand.....	12
2.9.	Gebäude- und Wohnungszählung.....	13
3.	Die Methodik der Haushaltegenerierung und Korrekturverfahren	14
3.1.	Datenbasis der Haushaltegenerierung.....	14
3.2.	Bildung von Haushaltszusammenhängen	15
3.3.	Verknüpfungen von Haushalten mit Wohnungen	15
3.4.	Bildung weiterer Haushaltszusammenhänge aus Registerinformationen.....	16
3.5.	Integriertes Korrekturverfahren.....	17
3.6.	Bildung von Haushalten und Zuweisung zu Wohnungen nach statistischen Generierungskriterien	18
4.	Datenaufbereitung	20
4.1.	Vollständigkeit des AGR	20
4.2.	Melderegisterdaten.....	20
4.3.	Erwerbsstatistische Registerdaten.....	21
4.4.	Daten aus den Erhebungen an Adressen mit Sonderbereichen.....	21
4.5.	Daten für Deutsche im Ausland.....	21
4.6.	Daten aus der Befragung zur Klärung von Unstimmigkeiten	22

5.	Kurzbeschreibung Produkt 3 – „Gesamtdatensatz Zensus 2011“	24
5.1.	Weitere Aufbereitung	24
5.1.1.	Teiltabellen	24
5.1.2.	Merkmalsauswahl	25
5.1.3.	Stichprobenziehung für die Nutzung am Gastwissenschaftlerarbeitsplatz (GWAP).....	26
5.1.4.	Anpassung der Hochrechnung für die GWAP-Stichprobe	26
5.2.	Erhebungsmerkmale des Gesamtdatensatzes des Zensus 2011.....	27
5.3.	Regionale Einheit.....	31
5.4.	Mengengerüst	31
5.5.	Datennutzung	33
6.	Geheimhaltung des Gesamtdatenbestandes des Zensus 2011	35
6.1.	Tabellengeheimhaltung.....	35
6.2.	Geheimhaltung bei Verhältniszahlen	36
6.3.	Geheimhaltung bei multivariaten Auswertungen.....	37
7.	Datenqualität und Probleme.....	38
7.1	Abweichende Ergebnisse.....	38
7.2	Abweichung zu Ergebnissen der Zensusdatenbank	38
7.3	Fehlende Werte bei Identifikatoren.....	38
7.3.1	Fehlende Werte beim Gebäude- und Wohnungsidentifikator	38
7.3.2	Fehlende Werte bei der Haushaltsnummer	39
7.4	EU-Merkmale.....	39
7.5	Merkmal Glaubensrichtung.....	39
7.6	Angaben zu Studierenden	39
7.7	Ergebnisse zu Haushalten und Familien.....	40
8.	FAQ.....	41
Anhang	44
Anhang 1:	Abkürzungsverzeichnis.....	45
Anhang 2:	Weiterführende Quellen und Literatur.....	46
Anhang 3:	Fragebögen der Erhebungsteile.....	47

1. Der Zensus 2011

Der Zensus 2011 war eine registergestützte Bevölkerungszählung, die durch eine Haushaltebefragung auf Stichprobenbasis und eine Vollerhebung aller an Adressen mit Sonderbereichen lebenden Personen (Gemeinschaftsunterkünfte, Anstalten, Notunterkünfte, Wohnheime und ähnliche Einrichtungen) ergänzt wurde, und kombiniert mit einer Gebäude- und Wohnungszählung (GWZ) zum Stichtag 9. Mai 2011 stattfand. Der Begriff „Zensus“ löst die früher genutzte Bezeichnung „Volkszählung“ ab. Beide Bezeichnungen können synonym verwendet werden. Die letzten Volkszählungen fanden in der Bundesrepublik Deutschland im Jahre 1987 und in der Deutschen Demokratischen Republik im Jahre 1981 statt.

Der Zensus 2011 wurde als Bundesstatistik von den Statistischen Ämtern des Bundes und der Länder durchgeführt. Mit ihm nahm Deutschland an einer EU-weiten Zensusrunde teil. Er ist national wie international ein wesentliches Fundament der amtlichen Statistik. Der Zensus liefert Basisdaten zur Bevölkerung, Erwerbstätigkeit und Wohnsituation, auf denen alle politischen, wirtschaftlichen und gesellschaftlichen Planungsprozesse bei Bund, Ländern und Gemeinden aufbauen. Wichtig ist er auch für das statistische Gesamtsystem, denn er ist Fortschreibungs- und Auswahlgrundlage für viele Statistiken. Zentrale Aufgabe jedes Zensus ist die Ermittlung der Einwohnerzahl, die in vielen Zusammenhängen – z. B. beim horizontalen und vertikalen Finanzausgleich, der Sitzverteilung im Bundesrat sowie bei der Einteilung der Wahlkreise – als maßgebliche Bemessungsgrundlage dient.

Die Zählung im Jahr 2011 bestand aus verschiedenen Erhebungsteilen – der Übernahme der Daten aus bereits vorhandenen Verwaltungsregistern, der Haushaltebefragung auf Stichprobenbasis, der GWZ sowie der Befragung an Adressen mit Sonderbereichen – die im Rahmen der Haushaltsgenerierung (HHGen) miteinander verknüpft wurden. Dieses Verfahren war notwendig, da Informationen über Haushaltszusammenhänge in den Melderegistern nicht unmittelbar vorhanden sind. Deshalb haben die Statistischen Ämter des Bundes und der Länder ein Verfahren entwickelt, bei dem anhand statistisch auswertbarer Merkmale aus dem Melderegister, der Haushaltsstichprobe, der Befragung an Adressen mit Sonderbereichen und der GWZ Haushaltszusammenhänge abgeleitet werden.

Weiterführendes zum Datenangebot der Forschungsdatenzentren zum Zensus 2011

FDZ-Arbeitspapier Nr. 48: Bereitstellung der Daten des Zensus 2011 über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, S. Raab/ C. Meisdrock, September 2015.

http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-48.pdf

1.1. Rechtsgrundlagen

In Deutschland basierte der Zensus 2011 auf einem System verschiedener Rechtsgrundlagen der Europäischen Union (EU), Deutschlands und der einzelnen Bundesländer. Er wurde auf der Basis der *Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates vom 9. Juli 2008 über Volks- und Wohnungszählungen* durchgeführt. Diese wurde ergänzt durch drei Kommissionsverordnungen: Dies waren die (1) *Verordnung (EU) Nr. 519/2010* der Kommission vom 16. Juni 2010 zur Annahme des Programms der statistischen Daten und der Metadaten für Volks- und Wohnungszählungen, (2) die *Verordnung (EG) Nr. 1201/2009* der Kommission vom 30. November 2009 in Bezug auf die technischen Spezifikationen für die Themen sowie deren Untergliederungen und (3) die Verord-

nung (EU) Nr.1151/2010 der Kommission vom 8. Dezember 2010 in Bezug auf die Modalitäten und die Struktur der Qualitätsberichte sowie das technische Format der Datenübermittlung.

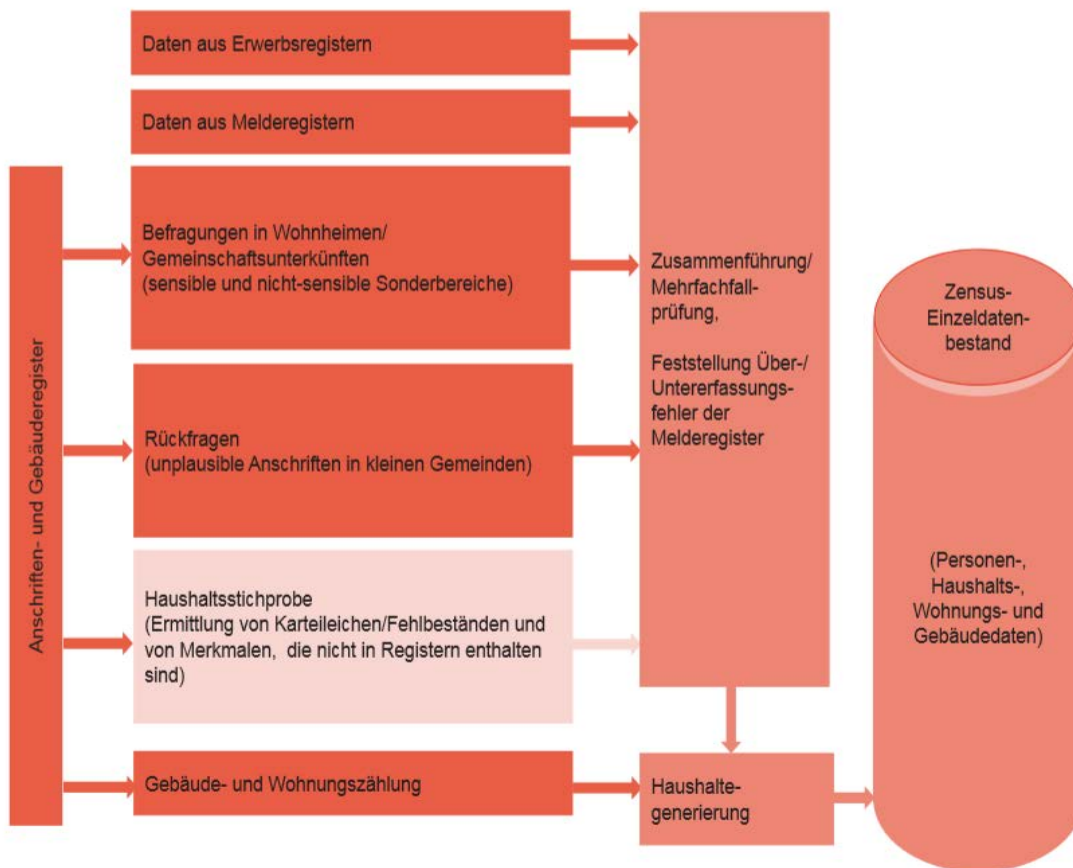
Grundlage für die Vorbereitung des Zensus 2011 in Deutschland war das *Gesetz zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung (Zensusvorbereitungsgesetz 2011 – ZensVorbG 2011)*. Kern dieses Gesetzes war der Aufbau des Anschriften- und Gebäuderegisters (AGR), dem zentralen Steuerungsinstrument der Datenerhebung und -verarbeitung. Weiterhin wurden die Methode und die zu nutzenden Datenquellen bestimmt. Das *Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011)* legte u. a. die zu erhebenden Merkmale fest, regelte die Auskunftspflichten, die Organisation der Erhebung und die Zusammenführung der Daten (HHGen). Für die Umsetzung der Durchführung der Bevölkerungsbefragung wurde am 25. Juni 2010 eine *Verordnung über Verfahren und Umfang der Haushal- tebefragung auf Stichprobenbasis zum Zensusgesetz 2011 (Stichprobenverordnung Zensusgesetz 2011 – StichprobenV)* verabschiedet. Hier wurden das Stichprobenverfahren und der Stichprobenumfang festgelegt. Die konkrete Durchführung der Erhebungen in den 16 Bundesländern wurde in verschiedenen *Zensus-Ausführungsgesetzen der Länder* oder – wie in Bayern – in einer Anpassung des *Bayerischen Statistikgesetzes* geregelt.

Zur Durchführung des Zensus 2011 waren drei weitere EU-Verordnungen zu beachten. Dies waren die (1) Verordnung der Europäischen Union (EU) Nummer 519/2010 der Kommission vom 16. Juni 2010 zur Annahme des Programms der statistischen Daten und der Metadaten für Volks- und Wohnungszählungen, (2) die Verordnung Europäische Gemeinschaft (EG) Nummer 1201/2009 der Kommission vom 30. November 2009 in Bezug auf die technischen Spezifikationen für die Themen sowie deren Untergliederungen und (3) die Verordnung Europäische Union (EU) Nummer 1151/2010 der Kommission vom 8. Dezember 2010 in Bezug auf die Modalitäten und die Struktur der Qualitätsberichte sowie das technische Format der Datenübermittlung.

1.2. Methode des Zensus 2011

Die EU schreibt ab dem Jahr 2011 für alle Mitgliedstaaten die Durchführung von Volks-, Gebäude- und Wohnungszählungen im Abstand von zehn Jahren vor. Damit die Ergebnisse auf europäischer Ebene vergleichbar sind, müssen alle Mitgliedstaaten einen festgelegten Umfang von Merkmalen liefern. Ihnen bleibt dabei jedoch eine große Wahlfreiheit in der Methode, mit der diese Informationen gewonnen werden. Der Zensus 2011 wurde in Deutschland registergestützt durchgeführt. Das bedeutet, dass bereits vorhandene Verwaltungsregister als Datenquellen genutzt wurden, die um Informationen aus Vollerhebungen und Stichprobenerhebungen ergänzt bzw. korrigiert wurden. Da im Bereich der Gebäude und Wohnungen kein zentrales Register existierte, wurden diese Daten durch eine primärstatistische Vollerhebung gewonnen. Der Zensus 2011 war eine Stichtagserhebung. Der Stichtag dieser Zählung war der 9. Mai 2011.

Abbildung 1: Das Zensusmodell im Überblick



Quelle: Statistische Ämter des Bundes und der Länder (2015): Zensus 2011: Methoden und Verfahren, S. 13.

2. Methodik

Beim Zensus 2011 kam in Deutschland erstmals ein neues registergestütztes Verfahren zum Einsatz. Dabei wurden bereits vorhandene Registerdaten verwendet und mit den Ergebnissen unterschiedlicher Befragungen ergänzt und korrigiert. Zur Feststellung sowie zur statistischen Korrektur von Über- und Untererfassungsfehlern¹ in den Melderegistern für alle Gemeinden mit mindestens 10 000 Einwohnern und zur Gewinnung von Angaben, die nicht aus Registern gewonnen werden konnten, wurden bundesweit knapp 10 % aller Personen im Rahmen der Haushalbefragung auf Stichprobenbasis befragt. Dafür wurden nach einem mathematischen Zufallsverfahren Anschriften ausgewählt, an denen alle dort zum Stichtag lebenden Personen befragt wurden.

Durch das statistische Verfahren der HHGen stehen zusätzlich zu den Informationen zu Bevölkerung, Gebäuden und Wohnungen, auch Angaben zu Haushalts- und Familienstrukturen zur Verfügung. Dabei wurden anschriftenweise Personen zu Haushalten zusammengefasst und den in der GWZ erfassten Wohnungen zugeordnet. Die Ergebnisse der HHGen bilden daher die Wohnhaushalte an einer Anschrift ab.

2.1. Anschriften- und Gebäuderegister

Grundlage des Zensus 2011 war das Anschriften- und Gebäuderegister (AGR). Dieses Register entstand aus der Zusammenführung der Daten der Meldebehörden, der Daten der Personalregister der Bundesagentur für Arbeit bzw. der öffentlichen Arbeitgeber und aus den Georeferenzierten Adressdaten Bund (GAB) der Vermessungsbehörden. Jede Anschrift, die in mindestens zwei der Datenquellen gefunden wurde, wurde als potenzielle Anschrift mit Wohnraum klassifiziert. Anschriften, die nur in einer der drei Quellen gefunden werden konnten, wurden gemäß § 14 ZensG 2011 durch die Statistischen Ämter der Länder auf das Vorhandensein von Wohnraum geprüft. Das AGR wurde seit 2008 durch die Statistischen Ämter des Bundes und der Länder im Statistischen Bundesamt aufgebaut und durch weitere Datenlieferungen und Informationen aus anderen Erhebungsteilen kontinuierlich aktuell gehalten. Es enthielt alle am Zensusstichtag existierenden Anschriften mit Wohnraum und bewohnten Unterkünften und somit alle Anschriften, an denen sich zum Zensusstichtag zensusrelevante Gebäude befanden oder zensusrelevante Personen lebten. Das AGR war somit über den gesamten Zensusablauf das Bindeglied zwischen allen Erhebungsteilen.

¹ Die Angaben in den Melderegistern sind nicht fehlerfrei: Zum einen gibt es Personen, die laut Melderegister an einer Anschrift gemeldet sind, aber nicht (mehr) tatsächlich dort wohnen (sogenannte „Karteileichen“). Zum anderen gibt es auch Personen, die an einer Anschrift wohnen, ohne dort gemeldet zu sein (sogenannte „Fehlbestände“). Diese Fehler konnten mit statistischen Methoden bereinigt werden.

Weiterführendes zum Anschriften- und Gebäuderegister:

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011: Methoden und Verfahren, 2015 (Kapitel 3.2).

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethode_nVerfahren5121105119004.pdf?__blob=publicationFile

Kleber, Birgit; Maldonado, Andrea; Scheuregger, Daniel und Ziprik, Katja: Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011. In: Wirtschaft und Statistik 7/2009, S.629–640.

https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Zensus/Zensus.pdf?__blob=publicationFile

2.2. Daten der Einwohnermeldeämter

Der registergestützte Zensus 2011 basierte im Wesentlichen auf den Daten der Einwohnermelderegister. Auszüge dieser Registerdaten wurden – mit Stand zum 9. Mai 2011 – zunächst an das jeweils zuständige Statistische Landesamt und von dort an das Statistische Bundesamt übermittelt. Dort wurden sie im abgeschotteten und gesicherten Bereich der amtlichen Statistik bundesweit zusammengestellt.

Aus den Melderegistern kamen zu jeder Person neben der Anschrift auch Merkmale, wie zum Beispiel Name, Geschlecht, Tag, Monat und Jahr der Geburt, Familienstand und Staatsangehörigkeit(en). Zusätzlich enthielten die Registerauszüge das Zuzugsdatum an die am 9. Mai 2011 bewohnte Anschrift, den Wohnstatus und sogenannte „Verzeigerungen“. Das sind Angaben über Ehegatten, eingetragene Lebenspartner oder Kinder. Sie werden für die HHGen benötigt. Schließlich wurde noch die rechtliche Zugehörigkeit zu einer öffentlich-rechtlichen Religionsgesellschaft übermittelt. Die genaue Auflistung aller aus den Melderegistern übermittelten Merkmale findet sich in § 3 ZensG 2011.

2.3. Erwerbsstatistische Registerdaten

Mit dem Zensus 2011 sollten nicht nur die amtlichen Einwohnerzahlen festgestellt, sondern auch Informationen über die Beteiligung der Bevölkerung in Deutschland am Erwerbsleben und zu den ausgeübten Berufen gewonnen werden. Ein Teil dieses sogenannten erwerbsstatistischen Gesamtbildes konnte aus vorhandenen Registern entnommen werden. Bei der Bundesagentur für Arbeit in Nürnberg sind sowohl alle etwa 29 Millionen sozialversicherungspflichtig Beschäftigten, als auch die Empfängerinnen und Empfänger von Arbeitslosengeld sowie Arbeitssuchende registriert.

Neben Angaben wie Name, Anschrift und Tag der Geburt wurden zu jeder Person auch der Wirtschaftszweig, in der die Person tätig ist, der Arbeitsort und ihr Status – also zum Beispiel beschäftigt oder arbeitssuchend gemeldet – übermittelt. Die genaue Auflistung aller von der Bundesagentur für Arbeit übermittelten Merkmale findet sich in § 4 ZensG 2011.

Die Register der Bundesagentur für Arbeit lieferten zwar einen großen Teil der Informationen zur Erwerbstätigkeit abhängig Beschäftigter, aber längst nicht alle. Denn die rund 1,8 Millionen Beamtinnen und Beamte, Richterinnen und Richter, Soldatinnen und Soldaten gehören nicht zu den sozialversicherungspflichtig Beschäftigten. Die Informationen zur Erwerbstätigkeit dieser Berufsgruppen mussten demnach auf andere Art und Weise – nämlich über die öffentlichen Arbeitgeber – erhoben

werden. Die genaue Auflistung aller von den öffentlichen Arbeitgebern übermittelten Merkmale findet sich in § 5 ZensG 2011.

2.4. Haushaltebefragung

Die Ergebnisse der Haushaltebefragung spielen eine wichtige Rolle bei der Ermittlung der Einwohnerzahlen. Mit der Haushaltebefragung wurden außerdem Informationen gewonnen, die in keinem Register enthalten sind.

Die Haushaltebefragung diente zur Ermittlung der Einwohnerzahlen in großen Gemeinden mit 10 000 und mehr Einwohnern². Die Ergebnisse des Zensus 2001 zeigten, dass es in großen Gemeinden mehr Ungenauigkeiten in den Melderegistern gab als in kleinen Gemeinden. Der Prozentsatz an Über- und Untererfassungen war tendenziell umso größer, je größer eine Gemeinde war. Mit Hilfe der Haushaltebefragung konnten sogenannte Karteileichen (Übererfassungen) und Fehlbestände (Untererfassungen) in den Melderegistern ermittelt und statistisch korrigiert werden.

Mit der Haushaltebefragung, die als Stichprobenerhebung nur bei einem Teil der Bevölkerung durchgeführt wurde, wurden Merkmale erhoben, die in Deutschland nicht oder nicht ausreichend in Registern vorliegen. Hierzu gehören zum Beispiel Angaben zu Schul- und Berufsabschlüssen, zur Erwerbstätigkeit insbesondere von Selbstständigen, zur Religionszugehörigkeit und zum Migrationshintergrund. Darüber hinaus fordert die EU Angaben zur Erwerbstätigkeit nach der Definition der Internationalen Arbeitsorganisation (ILO). Da diese Anforderung nicht allein mit den Angaben der Bundesagentur für Arbeit bedient werden kann, wurden diese Angaben im Rahmen der Haushaltebefragung erhoben. Diese Merkmale stehen für Gemeinden mit mehr als 10 000 Einwohnern und Kreise zur Verfügung.

Bei der Haushaltebefragung des Zensus 2011 wurden knapp 10 % der Bevölkerung, das sind insgesamt rund 7,9 Millionen Menschen, zur Auskunft verpflichtet. Die Befragten wurden per mathematisch-statistischem Zufallsverfahren ausgewählt. Dabei wurden zunächst zufällig Adressen gezogen. Anschließend wurden alle Personen, die am 9. Mai 2011 an diesen Adressen gewohnt haben, von Erhebungsbeauftragten befragt.

Da die Angaben nur bei einem Teil der Bevölkerung erfragt wurden, wurden diese im Anschluss an die Befragung auf die gesamte Bevölkerung³ der jeweiligen Gemeinde oder des Kreises hochgerechnet und bilden so eine verlässliche Datengrundlage sowohl für die Gemeinden als auch für die gesamte Bundesrepublik.

² Die Zuordnung der Gemeinden zu einer Gemeindegrößenklasse nach § 2 Abs. 6 ZensG 2011 erfolgte anhand der amtlich fortgeschriebenen Bevölkerungszahl zum 31. Dezember 2009 bzw. 1. Januar 2010.

³ Für die Bevölkerung in bestimmten Gemeinschafts- und Anstaltsunterkünften (sogenannte „sensible Sonderbereiche“) liegen keine Stichprobeninformationen vor. Aus diesem Grund können die Summen der Merkmalsausprägungen von der Gesamteinwohnerzahl der ausgewählten Einheit abweichen.

Weiterführendes zur Haushaltebefragung:

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 1 – Haushaltsstichprobe Zensus 2011.

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_t_1.pdf

Berg, Andreas und Bihler, Wolf (2011): Das Stichprobendesign der Haushaltsstichprobe des Zensus 2011. In: Wirtschaft und Statistik 4/2011, S. 317-328.

https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2011_04_Destatis_Das_Stichprobendesign_der_Haushaltsstichprobe_des_Zensus_2011.pdf?_blob=publicationFile&v=22

2.5. Erhebungen an Anschriften mit Sonderbereichen

Der Zensustest 2001 und die Volkszählung 1987 haben gezeigt, dass die Melderegister bei Wohnheimen und Gemeinschaftsunterkünften aus verschiedensten Gründen über mehr Karteileichen und Fehlbestände verfügen als die sogenannten „Normalanschriften“. Um die Einwohnerzahlen in Gemeinden und Städten verlässlich ermitteln zu können, wurden aus diesem Grund beim Zensus 2011 Daten bei allen Bewohnerinnen und Bewohnern von Wohnheimen und Gemeinschaftsunterkünften vollständig erhoben. Mit diesen Angaben wurden die zum Stichtag erhobenen Melderegisterdaten statistisch korrigiert.

Die Wohnheime und Gemeinschaftsunterkünfte wurden dabei in Anschriften mit nicht-sensiblen und sensiblen Sonderbereichen unterteilt. In den nicht-sensiblen Gemeinschaftsunterkünften, wie zum Beispiel Internaten, Studierendenwohnheimen, Klöstern und Seniorenwohnheimen, wurden die Bewohner persönlich befragt.

Zu den sensiblen Einrichtungen zählen in erster Linie Unterkünfte, bei denen die Information über die Zugehörigkeit der Personen zu diesen Bereichen die Gefahr einer sozialen Benachteiligung hervorrufen könnte. Deshalb wurden bei dieser Befragung besonders hohe Anforderungen an den Datenschutz gestellt. In sensiblen Einrichtungen, wie zum Beispiel Justizvollzugsanstalten, Notunterkünften für Obdachlose oder psychiatrischen Einrichtungen, gab stellvertretend die Einrichtungsleitung gegenüber Interviewerinnen und Interviewern oder per Online-Meldung Auskunft.

Gemäß § 8 ZensG 2011 wurde nach Name und Vorname, Geschlecht, Familienstand, Geburtsdatum, Geburtsort, Geburtsstaat, Staatsangehörigkeiten, Datum des Einzugs in das Wohnheim oder die Gemeinschaftsunterkunft und Wohnungsstatus (alleinige Wohnung beziehungsweise Haupt- oder Nebenwohnsitz) gefragt, was zur Ermittlung der Einwohnerzahl ausreichend war.

Die Anschrift eines nicht-sensiblen Wohnheims oder einer Gemeinschaftsunterkunft konnte auch für die Haushaltebefragung zufällig ausgewählt werden.

2.6. Befragung zur Klärung von Unstimmigkeiten

Trotz der erfahrungsgemäß – auch durch den Zensustest 2001 bestätigten – besseren Qualität der Einwohnermelderegister in Gemeinden mit weniger als 10 000 Einwohnern, traten bei der Zusam-

menführung der Melderegisterdaten mit den Ergebnissen der GWZ Unstimmigkeiten bzgl. der gemeldeten Personen an einer Anschrift auf.

Diese konnten – neben der Befragung zur Klärung des Wohnsitzes und der Erhebung an Adressen mit Sonderbereichen – mit der Befragung zur Klärung von Unstimmigkeiten (BKU) geklärt werden, was zur Ermittlung der Einwohnerzahl aus in Gemeinden mit weniger als 10 000 Einwohnern ausreichte. Hierbei wurden die an den betroffenen Adressen wohnenden Personen mit einem verkürzten Fragebogen befragt. Diese Befragung fand in der Regel an Adressen mit nur einer bewohnten Wohnung statt.

2.7. Prüfung von Mehrfachfällen

Da die Einwohnermelderegister dezentral in jeder Gemeinde geführt werden, kann es fälschlicherweise vorkommen, dass eine Person gleichzeitig in mehreren Gemeinden ihren Hauptwohnsitz oder ihren alleinigen Wohnsitz angemeldet hat. Zudem ist es möglich, dass eine Person ausschließlich mit einer oder mehreren Nebenwohnungen gemeldet ist. Die Ermittlung und ggf. Korrektur dieser Unstimmigkeiten war das zentrale Ziel der Mehrfachfallprüfung. Denn jede Person war mit genau einer Hauptwohnung beziehungsweise alleinigen Wohnung zu zählen.

In Gemeinden mit 10 000 oder mehr Einwohnern wurden die Registerdaten von Personen mit mehreren Hauptwohnungen oder mehreren alleinigen Wohnungen ausschließlich maschinell korrigiert. Personen mit ausschließlich einer oder mehreren gemeldeten Nebenwohnungen konnten mit Hilfe des Fragebogens zur Klärung des Wohnsitzes nach ihrem Wohnungsstatus befragt werden.

In Gemeinden mit weniger als 10 000 Einwohnern wurden bei der Befragung zur Klärung des Wohnsitzes alle diejenigen um Klärung gebeten, die am 9. Mai 2011 mit mehreren Hauptwohnsitzen oder ausschließlich mit Nebenwohnsitz in den Melderegistern gemeldet waren.

Somit konnte sichergestellt werden, dass alle Personen, bei denen unplausible Angaben in den Melderegistern festgestellt wurden, korrekt an ihrem Haupt- oder alleinigen Wohnsitz gezählt wurden. Die Ergebnisse der Befragung flossen in die Berechnung der amtlichen Einwohnerzahl ein.

Eine Rückmeldung an die Meldebehörden oder andere Behörden erfolgte nicht: Beim Zensus 2011 gibt es - basierend auf dem Volkszählungsurteil von 1983 des Bundesverfassungsgerichts - das sogenannte **Rückspielverbot**.

2.8. Zusammenführung der Daten aus den verschiedenen Datenquellen im Referenzdatenbestand

Aufgrund der Heterogenität der beim Zensus 2011 verwendeten Datenquellen – Register, Vollerhebungen und Stichprobenerhebungen – war es notwendig, die Erhebungsbefunde der verschiedenen Datenquellen zusammenzuführen und abzustimmen. Hierzu wurde der Referenzdatenbestand (RDB) erstellt.

In den §§ 9 und 12 ZensG 2011 wird dem RDB als Zusammenschluss des AGR, der Melderegisterdaten und der erwerbsstatistischen Register eine koordinierende Rolle im Zensus 2011 zugewiesen. Der RDB bildete unterschiedliche Bezugsgrößen ab, wie Informationen auf Adressen- und Personenebene. Informationen zur Anschrift enthielt das AGR. Informationen auf Personenebene stammten aus den Angaben der Melderegister, der Erwerbsregister und der Befunde aus den Bevölkerungser-

hebungen. Über ein Ordnungsnummernsystem war jeder Personendatensatz eindeutig einer Anschrift aus dem AGR zugeordnet.

Die zentrale Quelle für die Personendatensätze waren die Melderegisterdaten. Sie stellten den Grundstock des RDB dar und wurden sowohl zur Vorbereitung und Organisation der primärstatistischen Erhebungen als auch zur Ermittlung der Einwohnerzahl verwendet. In diesen Ausgangsdatenbestand des RDB wurden in einem nächsten Schritt die Daten aus den Erwerbsregistern integriert.

Weiterführendes zum Referenzdatenbestand:

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011: Methoden und Verfahren, 2015 (Kapitel 3.3.7).

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethode_nVerfahren5121105119004.pdf?__blob=publicationFile

2.9. Gebäude- und Wohnungszählung

Da in Deutschland kein flächendeckendes Verwaltungsregister existiert, aus dem der Gebäude- und Wohnungsbestand und seine Nutzung festgestellt werden kann, mussten die entsprechenden Informationen mit Hilfe einer Vollerhebung gewonnen werden. Von Ende April 2011 bis Juli 2012 wurden deshalb etwa 19 Millionen Eigentümerinnen und Eigentümer beziehungsweise Verwalterinnen und Verwalter von Gebäuden und Wohnungen im Rahmen der GWZ postalisch befragt. Damit war diese Erhebung der umfangreichste Erhebungsteil im Zensus 2011.

Die gebäude- und wohnungsstatistischen Ergebnisse einer solchen Zählung liefern die Basis für die Fortschreibung des Gebäude- und Wohnungsbestandes bis zur nächsten Zählung. Außerdem bilden deren Ergebnisse unter anderem eine notwendige Entscheidungsgrundlage für wohnungspolitische und raumplanerische Fragestellungen in Bund, Ländern und Kommunen.

Der Fragebogen der GWZ enthielt beispielsweise Fragen nach der Art des Gebäudes, dem Baujahr, der Zahl der Wohnungen im Gebäude und der Heizungsart. Bei Wohnungen wurde unter anderem nach der Wohnfläche, der Zahl der Räume und der Ausstattung der Wohnung gefragt.

Weiterführendes zur Gebäude- und Wohnungszählung:

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 2 – Gebäude- und Wohnungszählung Zensus 2011.

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_2.pdf

3. Die Methodik der Haushaltegenerierung und Korrekturverfahren

Ein wichtiges Ziel des Zensus 2011 war es – neben der Berechnung der amtlichen Einwohnerzahl – Daten zu Zahl und Struktur von Haushalten und deren Wohnsituation zu gewinnen. Da diese Informationen jedoch nicht in den Melderegistern vorhanden sind, wurde mit der HHGen ein neuartiges regelbasiertes Verfahren entwickelt, um diese Haushaltszusammenhänge zu ermitteln, weitere kombinierte Auswertungen zu ermöglichen und einen zur ermittelten amtlichen Einwohnerzahl entsprechenden Datensatz zu erhalten. Dabei wurden – in einer Vielzahl von Prozessschritten – die für eine Anschrift vorliegenden Personeninformationen zu Haushalten zusammengefasst und den in der GWZ erfassten Wohnungen an der Anschrift zugeordnet. Die HHGen bildet also die Wohnhaushalte an einer Anschrift ab, für die einerseits reine Haushaltsergebnisse (z. B. Anzahl der Ein- und Mehrpersonenhaushalte), andererseits kombinierte Ergebnisse (z. B. durchschnittliche Wohnungsgröße von Haushalten mit Kindern) ausgewiesen werden können.

3.1. Datenbasis der Haushaltegenerierung

Auf Grund des für einen Zensus erstmalig in Deutschland durchgeführten registergestützten Verfahrens bediente sich auch die HHGen einer Vielzahl unterschiedlicher Datenquellen. Basis aller Erhebungsteile bildete das eigens für den Zensus aufgebaute AGR, das alle am Zensusstichtag existierenden Anschriften mit Wohnraum und bewohnten Unterkünften enthielt und somit alle Anschriften, an denen sich zum Zensusstichtag zensusrelevante Gebäude befanden oder zensusrelevante Personen lebten. Dieses Register ermöglichte es, in allen anderen Erhebungs- und Aufbereitungsschritten – speziell auch im Verfahren der HHGen – die Anschrift selbst als kleinste regionale Einheit zu erfassen.

Weiterhin bildeten die Registerdaten der Einwohnermeldeämter die Grundlage für einen Einzeldatensatz auf Personenebene, der bereits vor der HHGen verschiedenen Aufbereitungs- und Bereinigungsschritten zugeführt wurde, aber auch noch im Haushaltegenerierungsverfahren selbst weiteren Bereinigungsschritten unterlag.

So lagen die korrigierten Melderegisterdaten bereits zu Beginn der HHGen vor, bereinigt um die Ergebnisse der Befragung in Wohnheimen und Gemeinschaftsunterkünften (gemäß § 8 ZensG 2011) und der Befragung zur Klärung des Wohnsitzes (gemäß § 15 ZensG 2011). Als weitere ergänzende Befragungen zur Ermittlung der amtlichen Einwohnerzahlen dienten die Haushaltebefragung (gemäß § 7 ZensG 2011) in Gemeinden ab 10 000 Einwohnern und die Befragung zur Klärung von Unstimmigkeiten (gemäß § 16 ZensG 2011) in Gemeinden mit weniger als 10 000 Einwohnern. Die Ergebnisse dieser beiden primärstatistischen Erhebungen wurden im Haushaltegenerierungsverfahren korrigierend in den Einzelpersonenbestand eingearbeitet.

Für die Ermittlung von Zahl und Struktur von Haushalten und deren Wohnsituation wurden Daten zu Gebäuden und Wohnungen benötigt. Da in Deutschland kein flächendeckendes Register mit diesen Informationen existiert, wurde in einer Vollerhebung, der GWZ, Daten zum Gesamtbestand des Wohnraums erhoben. Diese Daten wurden von der HHGen genutzt, um die Personen zu Haushalten zusammenzufassen und den Wohnungen zuzuordnen.

Man kann also zusammenfassen: Im Rahmen der HHGen wurden sämtliche Datenquellen des Zensus 2011 zusammengeführt und zu einem Datensatz verknüpft, der fachlich und regional tief gegliederte Auswertungen zulässt. Dieser wurde im Rahmen des in der HHGen eingebetteten Korrekturverfahrens mit den berechneten amtlichen Einwohnerzahlen synchronisiert. Hierzu flossen hochgerechnete Ergebnisse aus der Haushaltebefragung in das Verfahren ein.

Weiterführendes zur Haushaltgenerierung:

Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltgenerierung. In: Bayern in Zahlen 9\2012, S. 615–624.

<https://www.statistik.bayern.de/veroeffentlichungen/epaper.php?pid=40830&t=1>

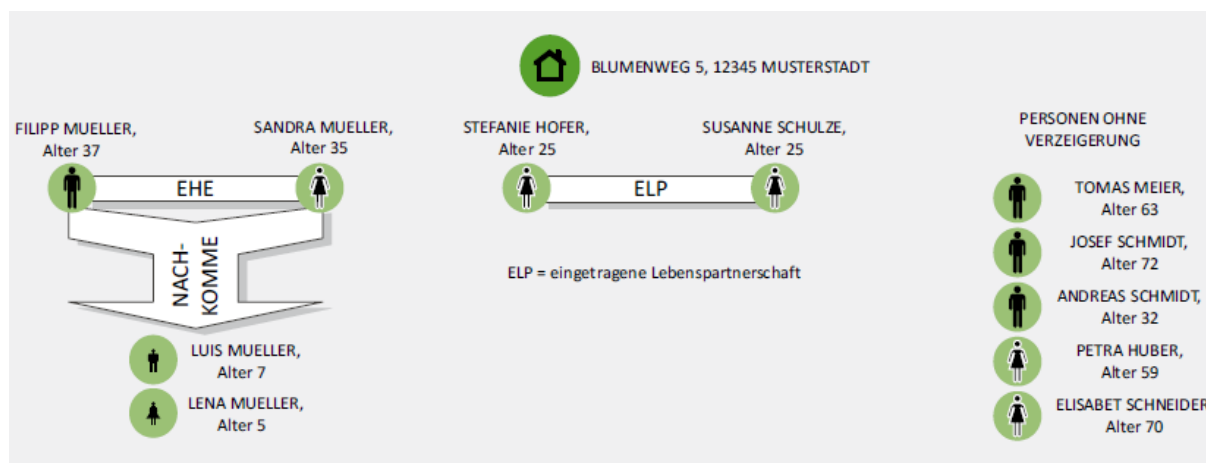
Die HHGen erfolgte in mehreren Stufen, die nachstehend vereinfacht beschrieben werden.⁴

3.2. Bildung von Haushaltszusammenhängen

Das Melderegister enthält Verweise zwischen Personen, die Verzeigerungen genannt werden. Diese Verzeigerungen geben eindeutige Hinweise auf Beziehungen zwischen an der gleichen Anschrift gemeldeten Personen. Zum einen werden mit Hilfe von Verzeigerungen eingetragene Partnerschaften zwischen zwei Personen gekennzeichnet. Dies sind im Einzelnen die Ehe und die eingetragene gleichgeschlechtliche Lebenspartnerschaft. Zum anderen werden mit Hilfe von Verzeigerungen gesetzliche Vertretungen gekennzeichnet. Dabei kann es sich um Eltern-Kind-Beziehungen, aber auch um sonstige gesetzliche Vertretungen handeln, wobei Kinder üblicherweise lediglich bis zur Vollendung des achtzehnten Lebensjahres verzeigert sind.

Im ersten Schritt der HHGen wurden die Verzeigerungen aus dem Melderegister genutzt, um Partner sowie Eltern und deren Kinder zusammenführen. Nach diesem Schritt war jede Registerperson einem vorläufigen Haushalt zugeordnet, der sich im weiteren Verlauf der HHGen immer wieder verändern konnte. Diese Haushalte bestanden entweder aus Einzelpersonen oder aus einfachen Familienstrukturen resultierend aus den Verzeigerungen.

Abbildung 2: Auswertung der Verzeigerungen



Quelle: Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltgenerierung, in: Bayern in Zahlen 9\2012, S. 618.

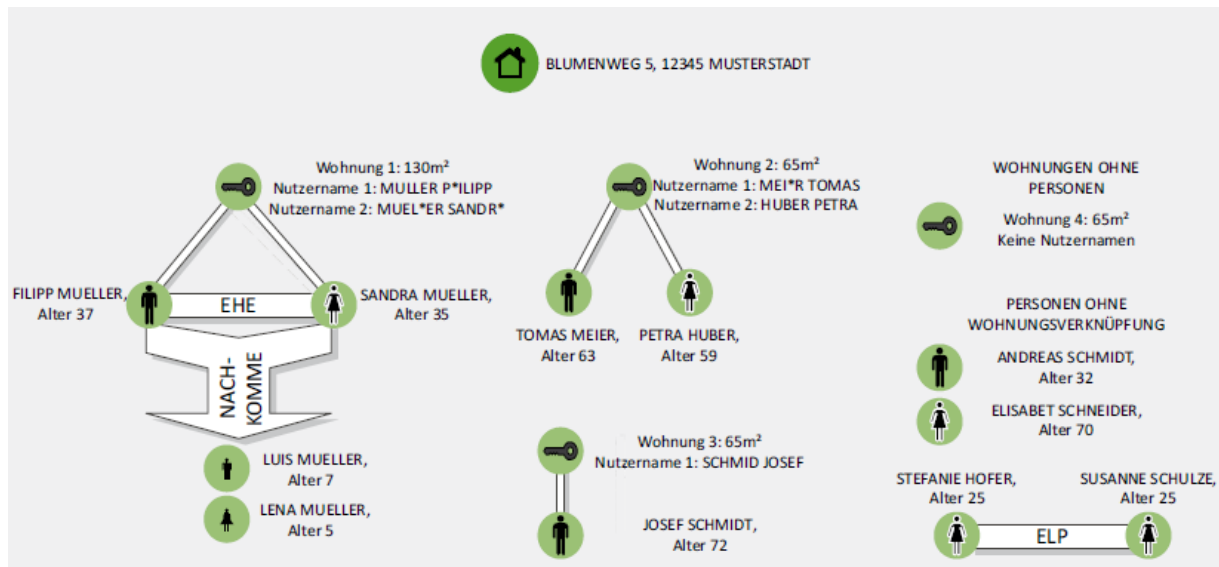
3.3. Verknüpfungen von Haushalten mit Wohnungen

Um die Verknüpfungen zwischen den aus der GWZ gelieferten Wohnungen und den Personen aus den Melderegisterdaten herzustellen, fand ein maschineller Abgleich der Namen aus dem Meldere-

⁴ Die Reihenfolge und Beschreibung der Schritte dient dem methodischen Verständnis und spiegelt nicht im Detail die technische Realisierung wider.

gister mit den in der GWZ als Hilfsmerkmal erhobenen Namen der Bewohner statt. Hierfür wurden im Rahmen der GWZ pro Wohneinheit die Namen von bis zu zwei Bewohnerinnen bzw. Bewohnern erfragt.

Abbildung 3: Verknüpfung von Haushalten und Wohnungen über Wohnungsnutzernamen



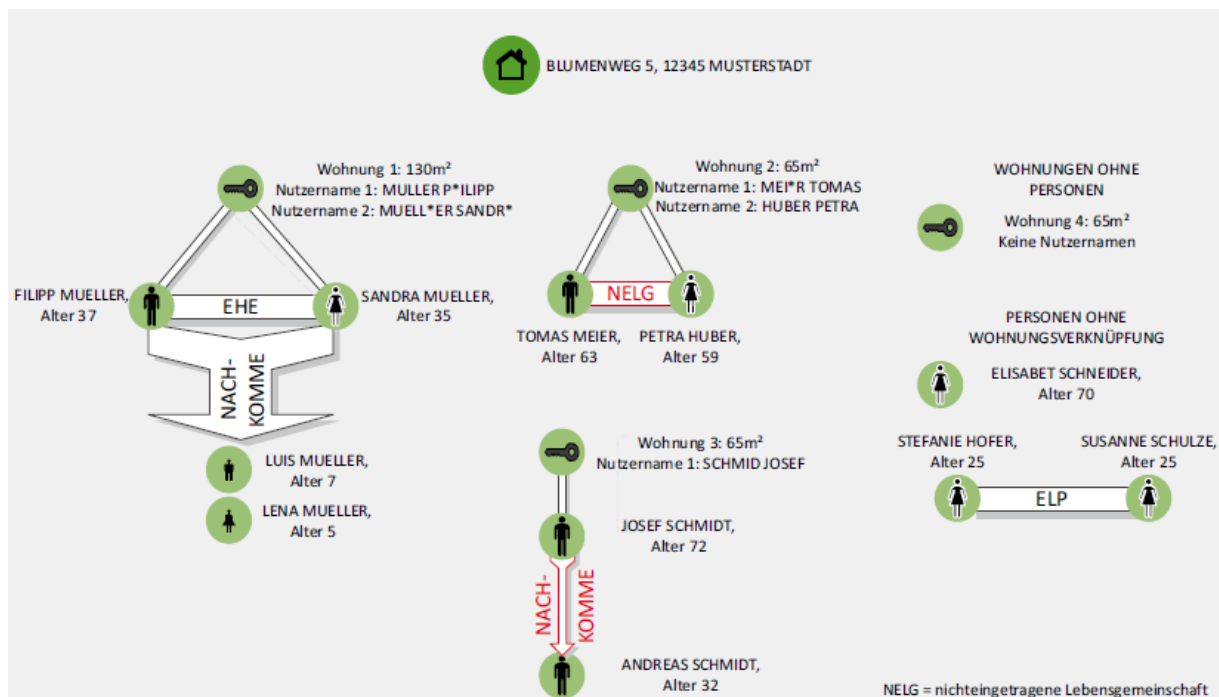
Quelle: Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltgenerierung, in: Bayern in Zahlen 9\2012, S. 619.

3.4. Bildung weiterer Haushaltszusammenhänge aus Registerinformationen

Da die im ersten Generierungsschritt ausgewerteten Verzeigerungen lediglich klassische Haushaltsstrukturen abbildeten – und diese auch nur, wenn sie explizit im Melderegister angegeben waren – war es notwendig, personenbezogene Registerinformationen zu nutzen, um Indizien für weitere Haushaltszugehörigkeiten zu erhalten. Dabei wurden neben den klassischen Haushaltsstrukturen auch nicht eingetragene Partnerschaften oder Großeltern-Enkel-Beziehungen erstellt. Hierzu kam ein mehrstufiges regelbasiertes Verfahren zur Anwendung. Je nachdem, welcher Beziehungstyp gesucht wurde, wurden Kombinationen unterschiedlicher Registerinformationen verwendet. Zum Beispiel wurden nicht eingetragene Partnerschaften anhand der Informationen zu Geschlecht, Alter, Familienstand, Zuzugsdatum und Zuzugsanschrift abgeleitet.

Nach diesem Schritt waren bereits 92 % aller Haushalte fertig generiert. Davon konnten bereits 76 % mit einer Wohnung verknüpft werden, die 16 %, für die zwar Haushalte generiert waren, aber keine Übereinstimmung von Wohnungsnutzernamen gefunden werden konnte (vor allem auch, da keine angegeben waren), wurden im Schritt unter Kapitel 3.6. ebenfalls mit einer Wohnung verknüpft.

Abbildung 4: Referenzbildung anhand weiterer Registerinformationen



Quelle: Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltgenerierung, in: Bayern in Zahlen 9\2012, S. 621.

3.5. Integriertes Korrekturverfahren

Um einen qualitativ hochwertigen, fachlich und regional flexibel auswertbaren Zensuseinzeldatensatz zu erhalten, musste der Registerdatenbestand auf Basis der Einzeldaten um Über- und Untererfassungen bereinigt werden. Für diese statistische Korrektur der Registerfehler wurde eine Differenzierung der mit der Haushaltsstichprobe festgestellten Karteileichen und Fehlbestände nach demografischen Merkmalen sowie nach Haushaltszusammenhängen (entweder wurde für alle Haushaltsmitglieder die Eigenschaft „Karteileiche“ bzw. „Fehlbestand“ festgestellt oder nur für einzelne Haushaltsmitglieder) benötigt, um eine systematische Verzerrung der demografischen bzw. haushaltsstatistischen Zensusergebnisse zu vermeiden.

Zu diesem Zwecke wurde ein Verfahren entwickelt, das die gemeindeweise aggregierten Vorgaben aus der Haushaltsstichprobe möglichst genau umsetzte. Die gemeindeweise Bereinigung der Karteileichen und Fehlbestände fand dabei in einem zwei- bzw. dreistufigen Verfahren statt:

1. Zunächst wurde anhand der aus der Haushaltsstichprobe geschätzten Verteilungen der Karteileichen und Fehlbestände nach einzelnen demografischen Merkmalen (Alter, Geschlecht, Staatsangehörigkeit, Familienstand) sowie nach Erwerbstätigkeit näherungsweise bestimmt, wie viele Personen einer bestimmten Merkmalskombination als Karteileichen zu löschen bzw. als Fehlbestände zu imputieren sind. Die Gesamtzahl der zu löschenden und zu imputierenden Datensätze pro Gemeinde stellte dabei eine strikt einzuhaltende Vorgabe dar.
2. Im nächsten Schritt wurden die konkreten Datensätze bestimmt, die zu löschen bzw. zu doppeln waren. Dabei wurden sowohl die in Schritt 1 vorgenommene Klasseneinteilung in Merkmalskombinationen als auch haushaltsstatistische Randbedingungen berücksichtigt. Das bedeutet zum Beispiel, dass eine Person mit einer „passenden“ Merkmalskombination sinn-

vollerweise an einer Anschrift gelöscht wird, an der die bis zu diesem Prozessschritt generierten Haushalte bereits „voll“ sind.

3. Für die Fehlbestände musste in einem dritten Schritt entschieden werden, in welchen „Empfängerhaushalt“ oder in welche – gegebenenfalls auch noch nicht belegte – „Empfängerwohnung“ die duplizierte Person imputiert wird. Hierzu wurden Schätzungen aus der Haushaltsstichprobe zur Anzahl verschiedener Haushaltsgrößen und -typen herangezogen, d. h. das Imputieren einer Person in einen Haushalt sollte möglichst bis zu diesem Prozessschritt noch überrepräsentierte Haushaltstypen verringern und die Anzahl unterrepräsentierter Haushaltstypen erhöhen.

Eine solche Korrektur der Einzeldaten konnte nur statistisch erfolgen. Es war nicht die buchhalterisch betrachtete „Richtigkeit“ des Einzelfalls relevant, sondern die Produktion möglichst unverzerrter Strukturdaten über die Bevölkerung in Deutschland.

In 5 % aller Haushalte erfolgte durch die Löschung von Karteileichen oder die Imputation von Fehlbeständen noch eine Änderung.

Weiterführendes zum Korrekturverfahren der Haushaltegenerierung:

Hofmeister, Katrin und Dr. Fürnrohr, Michael: Das Korrekturverfahren beim Zensus 2011. In: Bayern in Zahlen 6\2014, S. 310–318.

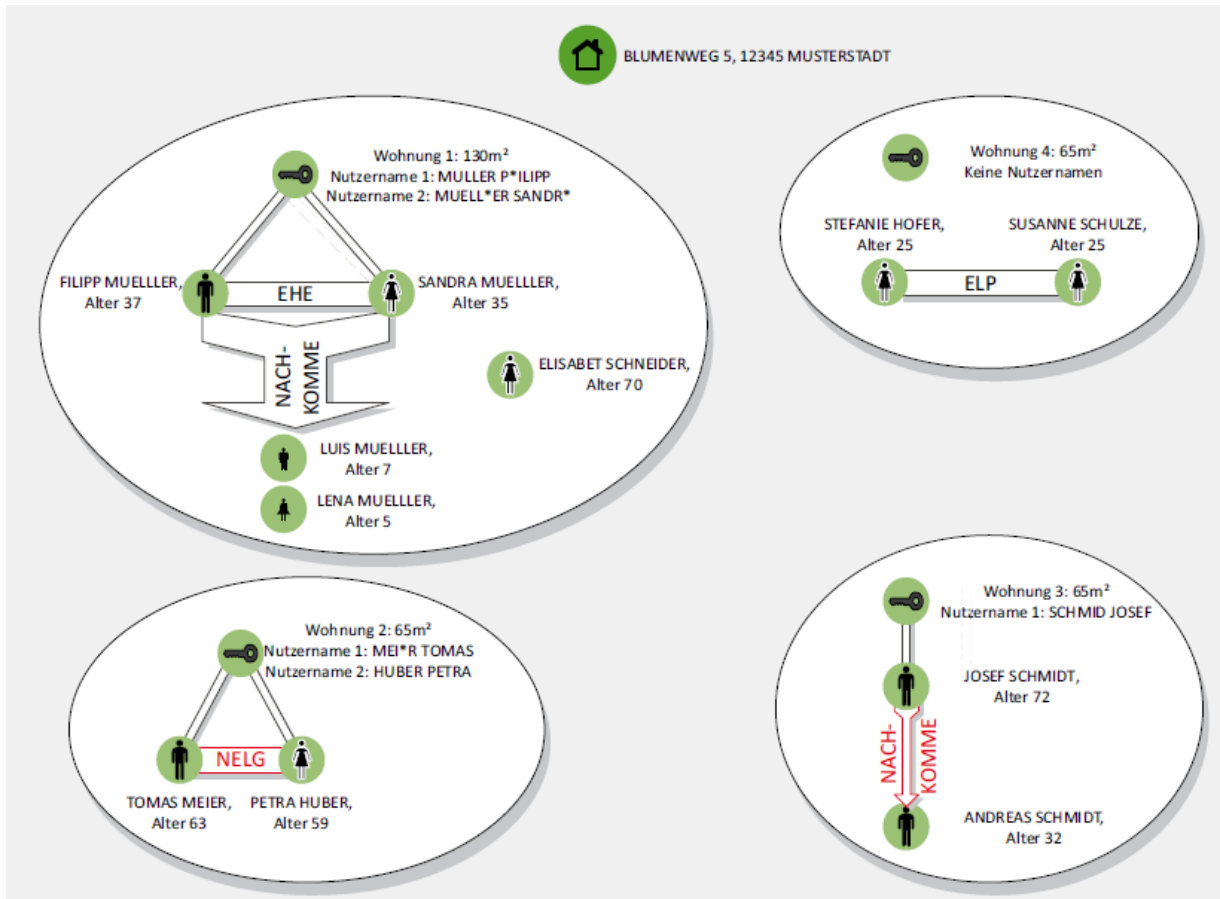
<https://www.statistik.bayern.de/veroeffentlichungen/epaper.php?pid=41876&t=1>

3.6. Bildung von Haushalten und Zuweisung zu Wohnungen nach statistischen Generierungskriterien

In den bisher dargestellten Schritten der HHGen wurden an einer Anschrift – abgesehen von den im Korrekturverfahren (Kapitel 3.5.) mit reinen Fehlbestandshaushalten belegten Wohnungen – nur diejenigen Wohnungen mit den generierten Haushalten zusammengeführt, für die mindestens ein Wohnungsnutzernamen aus der GWZ vorlag, der mit den an dieser Anschrift im Melderegister vorhandenen Personen über den Namensabgleich zusammenzuführen war. Die ggf. noch vorhandene Restmenge aller anderen an dieser Anschrift noch vorhandenen, unverknüpften bewohnten Wohnungen, für die entweder keine Wohnungsnutzernamen angegeben wurden oder keine zu der Angabe der Wohnungsnutzer passende Person an der Anschrift gefunden werden konnte, stand noch für eine Zusammenführung mit den ggf. noch an dieser Anschrift vorhandenen und noch nicht mit einer Wohnung verknüpften Haushalten zur Verfügung. Es konnte also noch bewohnte Wohnungen, die noch mit keinem Haushalt verknüpft waren, und/oder Haushalte, die noch mit keiner Wohnung verknüpft waren, geben. In diesem Schritt der HHGen erfolgte zum einen eine Zusammenführung der noch zur Verfügung stehenden Restmengen an unverknüpften Wohnungen und Haushalten. Waren danach immer noch unverknüpfte Haushalte vorhanden, so erfolgte über statistische Kriterien eine Zusammenführung der noch zur Verfügung stehenden Restmengen an unverknüpften mit bereits verknüpften Haushalten; dabei wurden über Annahmen zu sinnvollen Belegungsdichten an einer Anschrift noch zu verteilende Personen so auf die an der Anschrift bereits generierten Haushalte verteilt, dass keine unplausiblen Personenzahlen, gemessen an der verfügbaren Wohnungsgröße, resultierten. Ferner wurde versucht, für die Gemeinde bzw. den Kreis insgesamt möglichst gut aus der Haushaltsstichprobe geschätzte Eckzahlen zur Anzahl bestimmter Haushaltstypen zu treffen.

Neben den bei Schritt 3.4 bereits erwähnten 16 % aller Haushalte, die bereits fertig generiert waren, für die lediglich noch eine Wohnung benötigt wurde, erfolgte in diesem Schritt die Verarbeitung der letzten 3 % aller Haushalte.

Abbildung 5: Bildung von Haushalten und Zuweisung zu Wohnungen nach statistischen Generierungskriterien



Quelle: Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltegenerierung, in: Bayern in Zahlen 9\2012, S. 622.

Weiterführendes zur Generierung von Haushalten:

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011: Methoden und Verfahren, 2015 (Kapitel 4).

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethode_nVerfahren5121105119004.pdf?__blob=publicationFile

4. Datenaufbereitung

Im Zuge der Datenaufbereitung wurde geprüft, ob zu jeder Untersuchungseinheit vollständige und plausible Daten übermittelt wurden. War dies nicht der Fall, mussten die vorhandenen Fehler korrigiert und die fehlenden Angaben ergänzt werden. Der Prozess der Erkennung und Beseitigung von Fehlern war bundeseinheitlich und wurde überwiegend maschinell durchgeführt.

Im Folgenden werden die Maßnahmen beschrieben, die zur Aufbereitung der erhobenen Daten unternommen wurden, die nicht bereits in den FDZ-Produkten 1 und 2 näher beschrieben wurden.

Weiterführendes zu den FDZ-Produkten 1 und 2:

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 1 – Haushaltsstichprobe Zensus 2011.

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_1.pdf

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 2 – Gebäude- und Wohnungszählung Zensus 2011.

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_2.pdf

4.1. Vollständigkeit des AGR

Größte Aufmerksamkeit erhielt hierbei die vollständige Erfassung aller Anschriften von Personen, die am Stichtag zur zensusrelevanten Bevölkerung gehörten. Dies war besonders für die Qualität der Zensusergebnisse wichtig, da das AGR als Auswahlgrundlage für die Haushaltebefragung des Zensus 2011 genutzt wurde.

Eine Maßnahme der Sicherstellung der Vollständigkeit des AGR war eine zusätzliche Prüfung der Anschriften, die nur in einem der drei Datenbestände (GAB der Vermessungsbehörden, Daten der Meldebehörden und Daten der Personenregister der Bundesagentur für Arbeit) enthalten waren. Zudem wurde durch Update-Datenlieferungen der Vermessungsbehörden sowie der Meldebehörden die Aktualität des AGR-Datenbestandes sichergestellt. Schließlich wurde die Vollständigkeit und Aktualität auch durch die Integration von Erkenntnissen aus den primärstatistischen Erhebungen gewährleistet.

4.2. Melderegisterdaten

Aufgrund der dezentralen Führung der Melderegister war nicht sichergestellt, dass eine zeitgleiche Verarbeitung von Anmeldungen in der Zuzugsgemeinde und Abmeldungen in der Fortzugsgemeinde stattfand. Darüber hinaus gelten für meldepflichtige Personen Fristen, in welchen sie ihren Zuzug oder auch persönliche Datenänderungen melden müssen. Um diesen Differenzen entgegenwirken zu können, waren die Meldebehörden verpflichtet, neben der Datenlieferung zum Zensusstichtag 9. Mai 2011 auch eine weitere Lieferung zum 9. August 2011 zur Verfügung zu stellen. So konnten mit Hilfe des Merkmals „Einzugsdatum“ auch nach dem Stichtag erfolgte An- und Ummeldungen berücksichtigt werden. Beide Lieferungen wurden in den Statistischen Ämtern der Länder zudem auf folgende formale Vorgaben überprüft:

- Prüfung, ob für jede Gemeinde eine Datenlieferung erfolgt ist

- Vergleich der Angaben auf dem Begleitschein mit der Anzahl der Datensätze in der Datenlieferung
- Prüfung, ob der Datensatzaufbau der vorgegebenen Datensatzbeschreibung entspricht
- Prüfung, ob der verwendete Zeichensatz der Datenlieferung zulässig ist

Anschließend erfolgte eine Prüfung der von den Meldebehörden übermittelten Daten auf Vollständigkeit, Vollzähligkeit sowie Plausibilität. Gegebenenfalls waren weitere Bearbeitungsschritte bis hin zu einer Neulieferung der Daten notwendig.

4.3. Erwerbsstatistische Registerdaten

Bedingt durch das Übermittlungsverfahren waren die Registerdaten der Bundesagentur für Arbeit zunächst als Rohdaten ohne inhaltliche Prüfungen anzusehen. Im Gegensatz dazu wurden die Daten der personalführenden Stellen der öffentlichen Arbeitgeber teilweise, je nach verwendetem Übermittlungsverfahren, bereits während der Übertragung einer ersten inhaltlichen Prüfung unterzogen. Dieses Vorgehen gewährleistete einen qualitativen Mindeststandard für eine erfolgreiche Datenübermittlung.

Um den Qualitätsanforderungen im Zensus 2011 gerecht zu werden, war die Entwicklung eines Softwaresystems notwendig, das die vollständige Erfassung aller Erwerbsregisterdaten sowie die Datenprüfung und -aufbereitung beinhaltete und diese für anschließende Folgeprozesse bereitstellte. Das oberste Ziel war dabei, valide Daten über den Bestand und die Strukturen der Erwerbstätigen in Deutschland zu gewinnen, die personenscharf mit den Daten der Melderegister verbunden werden können. Hierfür wurden zwei elektronische Register, das Erwerbsregister – Teil Bundesagentur für Arbeit (ER-BA) und das Erwerbsregister – Teil Personalstand (ER-PS) konzipiert.

4.4. Daten aus den Erhebungen an Anschriften mit Sonderbereichen

Die Aufbereitung der erhobenen Daten an Anschriften mit Sonderbereichen fand in mehreren Schritten in den Statistischen Ämtern der Länder, in Teilen aber auch im Statistischen Bundesamt statt. In der Aufbereitung wurden alle Prozesse durchgeführt, die notwendig waren, um insbesondere die erhobenen Personendaten mit den Melderegisterbeständen abzugleichen, die Daten zu plausibilisieren und – falls notwendig – zu vervollständigen. Zur Datenaufbereitung gehörten die Schritte Zusammenführung von Fragebogen mit der elektronischen Erhebungsliste⁵, Melderegisterabgleich bzw. -anbindung, Mehrfachfallprüfung, Wohnstatusfeststellung, Plausibilisierung und Imputation sowie die abschließende Datenübermittlung an den RDB. Je nach Erhebungsart (z. B. sensibler bzw. nicht-sensibler Sonderbereich, Kaserne) waren die Prozessabläufe etwas unterschiedlich.

4.5. Daten für Deutsche im Ausland

Die der EU-Zensusverordnung zugrunde liegende Definition des Begriffs „Bevölkerung“ sah vor, dass die ins Ausland entsandten Angehörigen der Streitkräfte und die für die Streitkräfte im Ausland tätigen Personen, die Angehörigen der deutschen Polizeibehörden und des Auswärtigen Dienstes sowie ihre dort ansässigen Familien für die Feststellung der Einwohnerzahl im Heimatland zu zählen waren. Zusätzlich waren zeitlich befristet in den Auswärtigen Dienst übernommene und in das Ausland entsandte Angehörige anderer Bundesbehörden (§ 31 Abs. 1 Gesetz über den Auswärtigen Dienst) zu erfassen. Zu erfassen waren auch Angehörige des Auswärtigen Dienstes, die das Auswärtige Amt

⁵ In der elektronischen Erhebungsliste wurden im Wesentlichen die demografischen Kernangaben und das Befragungsergebnis je Person durch die Erhebungsstellen gespeichert. Die Angaben stammten aus der papierernen Erhebungsliste der Erhebungsbeauftragten, die damit die Existenzfeststellung dokumentierten.

nach den Entsenderichtlinien des Bundes zu internationalen Organisationen beurlaubte und entsendete sowie Angehörige des Auswärtigen Dienstes, die nach § 123a des Beamtenrechtsrahmengesetzes (BRRG) einer öffentlichen Einrichtung außerhalb des Anwendungsbereiches des BRRG einer vorübergehenden Tätigkeit zugewiesen wurden. So wurde sichergestellt, dass auch die Personen zur amtlichen Bevölkerung von Deutschland gezählt wurden, die aufgrund ihrer zeitweisen Verwendung außerhalb des deutschen Staatsgebiets in Deutschland nicht meldepflichtig waren.

Um diese Personengruppe zu erfassen, war in § 3 Abs. 3 und 4 ZensG 2011 eine einmalige Datenlieferung durch das Auswärtige Amt, das Bundesministerium der Verteidigung und das Bundesministerium des Innern angeordnet. Die Daten der ins Ausland entsandten Personen entstammten Verwaltungsunterlagen, die durch die in § 3 Abs. 4 ZensG 2011 benannten Auskunftspflichtigen geführt werden. Diese Unterlagen wurden angelegt, um den internen Datenbedarf der Auskunftspflichtigen zu bedienen. Dementsprechend konnte nicht von einer Detaillierung und Standardisierung der Merkmale wie im Datensatz für das Meldewesen ausgegangen werden. Zusätzlich unterlagen Teile der Daten einem besonderen Schutzbedürfnis, das sich aus den Aufgaben der Personen ergab. Aus diesem Grund wurde bei der Datenerhebung nur auf einen reduzierten Merkmalskranz zurückgegriffen, der zur Ermittlung der amtlichen Einwohnerzahl unabdingbar war.

Die durch die drei Stellen übersandten Daten wurden vor der Weiterverarbeitung einer Eingangskontrolle unterzogen, die folgende Prüfungen umfasste:

1. Vergleich der Angaben auf dem Begleitschein mit Satzanzahl in der Datenlieferung
2. Überprüfung des Datensatzaufbaus gemäß Datensatzbeschreibung
3. Überprüfung des Zeichensatzes der Datenlieferung

Wenn diese Daten die Eingangsprüfung erfolgreich durchlaufen hatten, wurden sie plausibilisiert. Hier war zu prüfen und ggf. zu korrigieren, ob

- die gelieferten Daten stichtagstreu geliefert wurden.
- die Daten vollständig geliefert wurden.
- der Aufbau der Datensätze eine weitere maschinelle Verarbeitung und Auswertung zuließ.
- der aktuelle Staatenschlüssel für das Verwendungsland genutzt wurde, ggf. Korrektur bzw. Vercodung der Klartexteintragungen.
- Personen mehrfach geliefert wurden.

Anschließend erfolgte eine Mehrfachfallprüfung im bundesweiten Melderegisterdatenbestand.

4.6. Daten aus der Befragung zur Klärung von Unstimmigkeiten

Die Verfahren der Aufbereitung der Daten der BKU folgten denen der Haushaltebefragung auf Stichprobenbasis (Produkt 1). Aufgrund des reduzierten Merkmalskranzes der BKU entfiel die Verarbeitung von Klartextangaben. Somit unterteilte sich die Datenaufbereitung in die Prozesse Vollständigkeitskontrolle, Digitalisierung der Papierfragebogen, das Lesen und Interpretieren der digitalisierten Fragebogen, Zusammenführung der Daten der elektronischen Erhebungsliste (Liste aller Personen, für die die Existenz an einer BKU-Anschrift festgestellt wurde) mit denen der digitalisierten Fragebogen und die Plausibilitätsprüfung einschließlich Imputation.

Weiterführendes zur Datenaufbereitung:

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011: Methoden und Verfahren, 2015 (Kapitel 3).

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethodenVerfahren5121105119004.pdf?__blob=publicationFile

5. Kurzbeschreibung Produkt 3 – „Gesamtdatensatz Zensus 2011“

Die Forschungsdatenzentren (FDZ) der Statistischen Ämter des Bundes und der Länder ermöglichen der Wissenschaft den Zugang zu den Zensus-Mikrodaten gemäß § 16 Abs. 6 Bundesstatistikgesetz. Im Gegensatz zum üblichen Vorgehen der FDZ wird für die Daten des Zensus 2011 nicht nur ein Produkt, sondern es werden insgesamt drei Produkte erstellt. Produkt 1 enthält die Daten der Haushaltsstichprobe. Produkt 2 stellt die Daten der GWZ zur Verfügung. Produkt 3 erlaubt nun auch kleinräumige Auswertungen zur Bevölkerung, nachdem aus Produkt 1 Ergebnisse lediglich auf Ebene der Gemeinden mit mindestens 10 000 Einwohner bzw. der Landkreise nachweisbar sind. Zusätzlich zu den Informationen zur Bevölkerung (Produkt 1) und der GWZ (Produkt 2) umfasst Produkt 3 auch Angaben zu den Bevölkerungs-, Haushalts- und Familienstrukturen in Deutschland. Des Weiteren können nun auch Informationen zu themenübergreifenden Fragestellungen wie „Wie wohnt die Bevölkerung Deutschlands“ geliefert werden, oder auch gezielt die Wohnsituation bestimmter Bevölkerungsgruppen untersucht werden.

Im Gegensatz zu den FDZ-Produkten 1 und 2 kann das FDZ-Produkt 3 nicht standardisiert zur Verfügung gestellt werden. Der Gesamtdatensatz des Zensus ist sehr umfangreich und komplex zu bearbeiten, sodass das FDZ-Produkt 3 in mehreren Teiltabellen (entsprechend dem Zensus-Modell) angeboten wird. Der benötigte Datensatz für das jeweilige Forschungsvorhaben wird daher stets individuell zusammengestellt. Im relationalen Datenmodell des FDZ-Produkts 3 können Merkmale aus den folgenden elf Teiltabellen ausgewählt werden: Anschrift, Familie-EU, Familie-National, Gebäude, Generation Edge, Haushalt, Haushaltsstichprobe, Partner Edge, Person, Regionalinformation, Wohnung. Die in den Teiltabellen enthaltenen Erhebungsmerkmale sind unter Abschnitt 5.2. dargestellt. Die Teiltabelle Haushaltsstichprobe entspricht dabei dem FDZ-Produkt 1, die Teiltabellen Gebäude und Wohnung entsprechen dem FDZ-Produkt 2. Diese Teiltabellen können im FDZ-Produkt 3 genutzt und mit anderen Teiltabellen verknüpft werden, können aber natürlich auch weiterhin einzeln als FDZ-Produkt 1 bzw. 2 genutzt werden. Da die Verknüpfung der Teiltabellen sehr komplex ist, werden benötigte Verknüpfungen immer vom fachlich zuständigen FDZ-Standort (Bayern) vorgenommen.

5.1. Weitere Aufbereitung

Für die Nutzung des Gesamtdatenbestands des Zensus 2011 wurden weitere Schritte zur Aufbereitung der Daten unternommen. Diese werden im Folgenden vorgestellt.

5.1.1. Teiltabellen

Das FDZ-Produkt 3 liegt aus Gründen der Komplexität und Größe des vorhandenen Datenmaterials in elf Teiltabellen vor und wird für jede Nutzung individuell auf das gewünschte Forschungsvorhaben angepasst. Für jede Nutzung werden somit ausschließlich die benötigten Teiltabellen zur Verfügung gestellt. Diese gliedern sich wie folgt auf:

1. **Anschrift:** Diese Tabelle enthält im Wesentlichen alle anschriftenbezogenen Informationen unterhalb der Gemeindeebene (inkl. Gitterzellen).
2. **Familie-EU:** Die Tabelle enthält die Familienmerkmale nach Definition für die EU.
3. **Familie-National:** Die Tabelle enthält die Familienmerkmale nach nationaler Definition für Standardauswertungen.
4. **Gebäude:** Die Tabelle enthält die Gebäudedatensätze nach Abschluss der HHGen (FDZ-Produkt 2).

5. **Generation Edge:** Diese Tabelle enthält die im Rahmen der HHGen gewonnenen Informationen zu Nachkommensbeziehungen bzw. gesetzlichen Vertretungen zwischen jeweils zwei Personen.
6. **Haushalt:** Die Tabelle enthält die Haushaltsinformationen aus der HHGen und bildet damit die Haushaltszusammenhänge ab.
7. **Haushaltsstichprobe:** Diese Tabelle beinhaltet die aufbereiteten Daten der Haushaltsstichprobe, für die bundesweit knapp 10 % der Bevölkerung im Rahmen der Haushaltebefragung auf Stichprobenbasis befragt wurden (FDZ-Produkt 1).
8. **Partner Edge:** Diese Tabelle enthält die im Rahmen der HHGen gewonnenen Informationen zu Partnerschaftsbeziehungen zwischen jeweils zwei Personen, wobei jede Person lediglich an einer einzigen Partnerschaftsbeziehung teilnehmen kann.
9. **Person:** In dieser Tabelle sind die endgültigen Personeninformationen nach HHGen abgelegt.
10. **Regionalinformation:** In dieser Tabelle sind wesentliche Zusatzinformationen (z.B. Regionalschlüssel der Gemeinde, zugehöriger Kreisschlüssel zu einer Gemeinde etc.) je administrativer Auswertungseinheit abgelegt.
11. **Wohnung:** Die Tabelle enthält die Wohnungsdatensätze nach Abschluss der HHGen (FDZ-Produkt 2).

Die beiden Teiltabellen Anschrift und Regionalinformation beinhalten somit sämtliche Ortsinformationen, die bei der Erstellung von Ergebnissen zusätzlich benötigt werden können. Sollen bspw. untergemeindliche Ergebnisse erstellt werden, so wird die Teiltabelle Anschrift mit den relevanten Merkmalen an die gewünschte Auswertungstabelle angespielt. Eine alleinige Bereitstellung dieser Tabellen ist nicht empfehlenswert.

Die beiden Teiltabellen Generation Edge und Partner Edge dienen ausschließlich zur Verknüpfung mit Personendaten, da sich die darin enthaltenen Informationen auf Nachkommensbeziehungen bzw. gesetzliche Vertretungen zwischen jeweils zwei Personen oder auf Partnerschaftsbeziehungen zwischen jeweils zwei Personen beziehen. Eine alleinige Bereitstellung dieser Tabellen ist ebenfalls nicht empfehlenswert.

Alle weiteren Teiltabellen beinhalten die oben beschriebenen Datensätze, die zur Ergebniserstellung benötigt werden und können somit auch einzeln sowie unabhängig voneinander verwendet werden. Möchten Sie zum Beispiel eine Auswertung erstellen, die Haushalts- und Familienmerkmale beinhaltet, so werden die beiden Teiltabellen miteinander verknüpft und Ihnen zur Verfügung gestellt. Ihnen werden somit ausschließlich die für Ihr Forschungsvorhaben relevanten Teiltabellen angeboten.

5.1.2. Merkmalsauswahl

In das FDZ-Produkt 3 wurden aus Gründen der Geheimhaltung – wie bei FDZ-Produkten üblich – keine direkten Identifikatoren aufgenommen. Lag ein Merkmal mehrfach in identischer Form vor (z. B. sowohl als Erhebungs- als auch als Auswertungsmerkmal), wurde es nur einmal im Datensatz belassen. Da über die FDZ auf Basis der Zensusdaten nur inhaltliche Forschung möglich ist, wurden Merkmale, die hierfür nicht sinnvoll genutzt werden können, nicht in das Datenmaterial aufgenommen. EU-Merkmale werden beibehalten, um die internationale Vergleichbarkeit der Daten sicherzustellen.

5.1.3. Stichprobenziehung für die Nutzung am Gastwissenschaftlerarbeitsplatz (GWAP)

Um die faktische Anonymität der Daten nach § 16 Abs. 6 BStatG und kürzere Programmlaufzeiten zu gewährleisten, wird für die Nutzung am GWAP eine 2%-Substichprobe der Teiltabellen auf Anschri-ftebene gezogen. Davon abweichend wird für die Nutzung der Teiltabelle Haushaltsstichprobe (Produkt 1) am GWAP eine 10%-Substichprobe gezogen.

Zur Wahrung der faktischen Anonymität bei Auswertungen auf Gitterzellen-Ebene (Teiltabelle An-schrift) werden am GWAP ausschließlich Gitterzellen mit einer Seitenlänge von 10 km zur Verfügung gestellt. Bei der Ziehung der 2%-Stichprobe wurde gewährleistet, dass in jedem am GWAP bereitge-stellten „10 km × 10 km“-Raster mindestens drei Anschriften liegen.

5.1.4. Anpassung der Hochrechnung für die GWAP-Stichprobe

Für die am GWAP zur Verfügung gestellten Substichproben wurden die Hochrechnungsfaktoren in den FDZ mittels einer einfachen gebundenen Hochrechnung an das Vollmaterial angepasst. Damit kann auch mit der Sub-Stichprobe auf die Gesamtbevölkerung geschlossen werden. Die durch die gebundene Hochrechnung entstehenden Ergebnis-Ungenauigkeiten werden in Kauf genommen, da am GWAP erzeugte Ergebnisse nicht veröffentlicht werden.⁶

Folgende Merkmale und Merkmalsausprägungen wurden bei der gebundenen Hochrechnung je Teil-tabelle berücksichtigt:

Tabelle 1: Für die gebundene Hochrechnung der GWAP-Stichprobe genutzte Merkmale

Tabelle	Merkmal	Ausprägungen
Familie-EU	Amtlicher Gemeindeschlüssel ⁷	12-stelliger Schlüssel
	Größe der Kernfamilie (EU-Merkmal)	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
	Typ der Kernfamilie (EU-Merkmal)	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Familie-National	Amtlicher Gemeindeschlüssel	12-stelliger Schlüssel
	Größe der Kernfamilie	2 Personen 3 Personen 4 Personen 5 Personen 6 und mehr Personen
	Typ der Kernfamilie (nach Fami-lien)	Paare ohne Kind(er) Paare mit Kind(ern) Alleinerziehende Elternteile
	Typ der Kernfamilie (nach Lebens-form)	Ehepaare Eingetr. Lebenspartnerschaften Nichteheliche Lebensgemeinschaften

⁶ Der GWAP dient aufgrund des großen Datenumfangs des Zensus nur zur Erstellung der Syntaxen und zur ers-ten Sichtung von Ergebnissen. Finale Ergebnisse werden über die Kontrollierte Datenfernverarbeitung (KDFV) erzeugt.

⁷ Im Amtlichen Gemeindeschlüssel (AGS) werden die Gemeinden und gemeindefreien Gebiete bundeseinheitlich verschlüsselt. Der Schlüssel ist 12-stellig. In der 1. und 2. Stelle wird das Bundesland verschlüsselt. Die 3. Stelle bezeichnet den Regierungsbezirk. In der 4. und 5. Stelle werden die kreisfreien Städte und Landkreise verschlüsselt. Die vier folgenden Stellen (6. bis 9. Stelle) kennzeichnen den Gemeindeverband innerhalb eines Landkreises. In den Stellen 10 bis 12 wird die Zuordnung zur Gemeinde verschlüsselt.

		Alleinerziehende Väter Alleinerziehende Mütter
Haushalt	Amtlicher Gemeindeschlüssel	12-stelliger Schlüssel
	Größe des privaten Haushalts	1 Person 2 Personen 3 Personen 4 Personen 5 Personen 6 und mehr Personen
	Typ des privaten Haushalts (nach Familien)	Einpersonenhaushalte (Singlehaushalte) Paare ohne Kind(er) Paare mit Kind(ern) Alleinerziehende Elternteile Mehrpersonenhaushalte ohne Kernfamilie
	Typ des privaten Haushalts (nach Lebensform)	Einpersonenhaushalte (Singlehaushalte) Ehepaare Eingetr. Lebenspartnerschaften Nichteheliche Lebensgemeinschaften Alleinerziehende Mütter Alleinerziehende Väter Mehrpersonenhaushalte ohne Kernfamilie
Person	Amtlicher Gemeindeschlüssel	12-stelliger Schlüssel
	Alter (5 Jahresklassen)	unter 18 18 bis 29 30 bis 49 50 bis 64 65 und älter
	Geschlecht	männlich weiblich
	Familienstand (kurz)	Ledig Verheiratet/Eingetr. Lebenspartnerschaft Verwitwet/Eingetr. Lebenspartner/-in verstorben Geschieden/Eingetr. Lebenspartnerschaft aufgehoben Ohne Angabe

Durch die gebundene Hochrechnung der GWAP-Stichprobe kommt jede Kombination der oben genannten Merkmalsausprägungen in der hochgerechneten Unterstichprobe etwa genauso häufig vor wie im hochgerechneten Vollmaterial der KDFV.

Für die Haushaltsstichprobe und die Gebäude- und Wohnungszählung wurden jeweils entsprechend andere Merkmale und Merkmalsausprägungen genutzt. Eine detaillierte Beschreibung findet sich im Metadatenreport für Produkt 1 bzw. 2.

5.2. Erhebungsmerkmale des Gesamtdatensatzes des Zensus 2011

Im Rahmen der Volkszählung wurden gemäß ZensG 2011 die in Tabelle 2 genannten Merkmale erhoben. Zusätzlich wurden sogenannte Hilfsmerkmale abgefragt, die für die Zusammenführung der verschiedenen Erhebungsteile im Rahmen der Aufbereitung des Zensus 2011 benötigt und nach Abschluss dieser Arbeiten gelöscht wurden. Aus den Erhebungsmerkmalen wurden verschiedene Aus-

wertungsmerkmale abgeleitet, die ebenfalls im Rahmen von Produkt 3 zur Verfügung gestellt werden. Um internationale Vergleiche zu vereinfachen, werden zusätzlich EU-Merkmale angeboten (vgl. Kapitel 7). Neben den Auswertungsmerkmalen werden in Produkt 3 auch verschiedene regionale Gebietsschlüssel und Merkmalskombinationen bereitgestellt. Einen genaueren Überblick über die zur Verfügung stehenden Merkmale gibt das [Schlüsselverzeichnis](#) zu Produkt 3.

Tabelle 2: Überblick über die Erhebungsmerkmale des Gesamtdatensatzes des Zensus 2011⁸

Teiltabelle	Merkmal	Merkmalsausprägung
Anschrift	Administrativer Orts- oder Gemeindename zum 09.05.2011	alphanumerisch
	Auswahlsatz Stadtteil	Ziffern (numerisch)
	Gemeindegrößenklasse	bis 1.999 Einwohner 2.000 bis 9.999 Einwohner 10.000 bis 29.999 Einwohner 30.000 bis 99.999 Einwohner 100.000 bis 399.999 Einwohner 400.000 bis 999.999 Einwohner 1.000.000 Einwohner und mehr
	Name des amtlichen Orts- oder Gemeindeteils	alphanumerisch
	Name des Orts- oder Gemeindeteils	alphanumerisch
	Postalischer Orts- oder Gemeindename	alphanumerisch
	Postleitzahl	5-stellige Nummer (alphanumerisch)
	Familie-EU	Größe der Kernfamilie EU
Typ der Kernfamilie EU		Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Familie-National	Größe der Kernfamilie National	Ziffern (numerisch)
	Typ der Kernfamilie National	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Generation Edge	Typ der Beziehung	Eltern Sonstige gesetzliche Vertreter Großeltern
Haushalt	Haushaltstyp Migrationshintergrund laut Melderegister	Alle aktiven Personen im Haushalt haben einen Migrationshintergrund. Es gibt sowohl aktive Personen mit, als auch ohne Migrationshintergrund im Haushalt. Alle aktiven Personen im Haushalt haben keinen Migrationshintergrund.
	Haushaltstyp National	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Partner Edge	Typ der Partner-Beziehung	Ehepaar Eingetragene Lebenspartnerschaft

⁸ Die Erhebungsmerkmale für die Teiltabellen Haushaltsstichprobe, Gebäude und Wohnung finden sich im Metadatenreport zu Produkt 1 bzw. Produkt 2.

		Nichteheliche Lebensgemeinschaft Nichteingetragene Lebenspartnerschaft
Person	Alter zum Stichtag aus HHGen	Ziffern (numerisch)
	Datum des Zuzugs in die Gemeinde	JJJJMMTT (numerisch)
	Einzugsdatum	JJJJMMTT (numerisch)
	In nichtehelicher Lebensgemeinschaft	Person lebt laut primärstatistischer Erhebung nicht in einer NELG. Person lebt laut primärstatistischer Erhebung in einer NELG. Für die Person wurde primärstatistisch nicht erhoben, ob sie in einer NELG lebt.
	Erwerbsstatus	Sozialversicherungspflichtig Beschäftigte Beamte/-innen, Richter/-innen, Soldaten/-innen Arbeitslos, in Umschulung Sonstige
	Familienstatus EU	Mann in einer Ehe Frau in einer Ehe Mann in einer eingetragenen Lebenspartnerschaft Frau in einer eingetragenen Lebenspartnerschaft Mann in einer nichtehelichen Lebensgemeinschaft Frau in einer nichtehelichen Lebensgemeinschaft Alleinerziehender Mann Alleinerziehende Frau Nachkomme
	Familienstatus National	Bezugsperson in einer Ehe Bezugsperson in einer eingetragenen Lebenspartnerschaft Bezugsperson in einer nicht-ehelichen Lebensgemeinschaft Bezugsperson Alleinerziehende/r Partner der Bezugsperson in einer Ehe Partner der Bezugsperson in einer eingetragenen Lebenspartnerschaft Partner der Bezugsperson in einer nichtehelichen Lebensgemeinschaft Nachkomme der Bezugsperson oder des Partners der Bezugsperson
	Familienstand laut RDB	ledig verheiratet verwitwet geschieden Lebenspartnerschaft Lebenspartner verstorben Lebenspartnerschaft aufgehoben Ehe aufgehoben Todeserklärung aufgelöste Lebenspartnerschaft
	Geburtsstaat referenziert EU	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
	Geburtsstaat referenziert national	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste

Geburtsjahr und –monat zum Stichtag aus HHGen	JJJMM (alphanumerisch)
Geschlecht	Mann Frau
Haushaltstatus laut EU-Typisierung laut HHGen	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Haushaltstatus laut nationaler Typisierung	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Ausführliches Bezugsland des Migrationshintergrunds der Person laut MR	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Bezugsland des Migrationshintergrunds der Person laut MR	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Migrationshintergrund der Person laut MR	Deutsche/r ohne Migrationshintergrund Deutsche/r mit Migrationshintergrund ohne eigene Migrationserfahrung Deutsche/r mit Migrationshintergrund mit eigener Migrationserfahrung Ausländer/in mit Migrationshintergrund ohne eigene Migrationserfahrung Ausländer/in mit Migrationshintergrund mit eigener Migrationserfahrung
Wohnung mit eigener Haushaltsführung	eigene Haushaltsführung keine eigene Haushaltsführung
Sonderbereich allgemein	Anschrift ohne Sonderbereich oder Karteileiche sensibler Bereich nicht sensibel oder Normalbereich reine Meldeanschrift Binnenschiffer, Seeleute Ausländische Kaserne, Botschaft Deutsche im Ausland
Herkunftsstaat bei Zuzug aus dem Ausland	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Staatsangehörigkeit 1 laut HHGen	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Staatsangehörigkeit 2 laut HHGen	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Staatsangehörigkeit 3 laut HHGen	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Staatsangehörigkeit 4 laut HHGen	Ausprägungen entsprechend der dem Schlüsselverzeichnis beiliegenden Kodeliste
Wohnstatus	Alleiniger Wohnsitz Hauptwohnsitz Nebenwohnsitz

Weiterführendes zum Merkmalskatalog

Das Schlüsselverzeichnis mit den entsprechenden Variablennamen, Ausprägungen und Merkmalsausprägungsdefinitionen finden Sie hier:

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/fdz_schluesselfverzeichnis_zensus_produk_t_3.pdf

5.3. Regionale Einheit

Die Ergebnisse können auf Bundes-, Landes-, Regierungsbezirks-, Kreis-, Gemeindeverbands- sowie auf Gemeindeebene dargestellt werden. Die Zuordnung von Anschriften zur Gitterzellengröße 10 km × 10 km sowie 1 km × 1 km ermöglicht Auswertungen des Zensusdatenbestandes nach Gitterzellen als kleinste Ebene der Datenauswertung. Gitterzellen mit der Seitenlänge 100 m können zur Erzeugung eigener räumlicher Aggregate (z. B. Lärmzonen um Flughäfen) genutzt werden. Die Gebietsstände entsprechen dem Stand vom 9. Mai 2011. Danach eingetretene Gebietsstandänderungen sind nicht berücksichtigt.

Zu beachten gilt jedoch, dass Ergebnisse auf Basis der Haushaltebefragung⁹ nur auf Bundes-, Landes-, Regierungsbezirks- und Kreisebene dargestellt werden können. Darüber hinaus liegt für große Gemeinden mit mindestens 10.000 Einwohnern die Gemeindeebene in Form des Amtlichen Gemeindefachschlüssels als tiefste regionale Gliederungsebene vor. Für kleine Gemeinden mit weniger als 10.000 Einwohnern sind Ergebnisse auf Gemeindeebene aufgrund der geringen Stichprobengröße nicht aussagekräftig. Aus diesem Grund ist die tiefste regionale Gliederungsebene für Befragungsdaten aus kleinen Gemeinden der Landkreis.

5.4. Mengengerüst

Die Tabellen der Familien- und Haushaltsmerkmale nach nationaler Definition umfassen rund 22,7 Millionen Familien- bzw. 37,6 Millionen Haushaltsdatensätze. Am GWAP steht hiervon eine 2 %-Stichprobe zur Verfügung. Die GWAP-Stichprobe enthält rund 420.000 Familien- und knapp 700.000 Haushaltsdatensätze (siehe Tabelle 3). Die Tabelle der Personendatensätze umfasst rund 80,2 Millionen Datensätze. Am GWAP steht hiervon eine 2 %-Stichprobe zur Verfügung. Die GWAP-Stichprobe enthält somit knapp 1,5 Millionen Personendatensätze (siehe Tabelle 4).

Die Teiltabelle Haushaltsstichprobe umfasst für die KDFV 7,6 Millionen Personendatensätze. Die am GWAP bereitgestellte Substichprobe umfasst 760.370 Datensätze. Nähere Informationen dazu können dem [Metadatenreport für Produkt 1](#) entnommen werden.

Die Daten der Gebäude- und Wohnungszählung umfassen rund 40,6 Millionen Wohnungsdatensätze und über 18,9 Millionen Gebäudedatensätze. Am GWAP steht hiervon eine 2 %-Stichprobe zur Verfügung. Die GWAP-Stichprobe enthält rund 340.000 Gebäude- und knapp 750.000 Wohnungsdatensätze. Weitere Informationen hierzu finden sich im [Metadatenreport zu Produkt 2](#).

⁹ Die Auswertung umfasst mindestens ein Merkmal, das ausschließlich in der Teiltabelle Haushaltsstichprobe (Produkt 1) erhoben wurde.

Tabelle 3: Anzahl der Familien- und Haushaltsdatensätze nach Bundesland

	KDFV (nach Geheimhaltung)		GWAP	
	Familien	Haushalte	Familien	Haushalte
Baden-Württemberg	2.937.763	4.709.228	57.338	92.922
Bayern	3.492.951	5.679.508	63.846	105.857
Berlin	837.296	1.794.936	10.906	23.593
Brandenburg	751.080	1.182.470	15.334	24.773
Bremen	168.732	331.078	2.373	4.566
Hamburg	430.150	876.502	5.898	12.397
Hessen	1.677.948	2.751.100	35.063	59.111
Mecklenburg-Vorpommern	484.158	803.002	8.363	14.114
Niedersachsen	2.215.300	3.517.840	41.988	68.910
Nordrhein-Westfalen	4.868.657	8.130.738	84.878	140.540
Rheinland-Pfalz	1.153.536	1.789.030	24.939	41.225
Saarland	287.195	459.853	6.043	10.369
Sachsen	1.206.196	2.048.482	21.825	37.317
Sachsen-Anhalt	686.345	1.127.332	13.400	22.086
Schleswig-Holstein	809.253	1.316.303	15.242	25.602
Thüringen	652.485	1.053.817	11.720	19.743
Deutschland	22.659.045	37.571.219	419.156	703.125

Tabelle 4: Anzahl der Personendatensätze nach Bundesland

	KDFV (nach Geheimhaltung)	GWAP
Baden-Württemberg	10.486.660	200.847
Bayern	12.397.614	225.058
Berlin	3.292.365	43.621
Brandenburg	2.455.780	50.087
Bremen	650.863	9.054
Hamburg	1.706.696	23.453
Hessen	5.971.816	123.905
Mecklenburg-Vorpommern	1.609.981	28.537
Niedersachsen	7.777.992	146.599
Nordrhein-Westfalen	17.538.252	301.455
Rheinland-Pfalz	3.989.808	86.226
Saarland	999.623	21.381
Sachsen	4.056.799	72.950
Sachsen-Anhalt	2.287.040	43.980
Schleswig-Holstein	2.800.119	52.462
Thüringen	2.188.589	39.694
Deutschland	80.219.695	1.469.309

5.5. Datennutzung

Für Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung sind die Zensus 2011-Produkte der FDZ über eine kombinierte Nutzung von Kontrollierter Datenfernverarbeitung (KDFV) und Gastwissenschaftlerarbeitsplatz (GWAP) zugänglich. Konkret läuft die Nutzung der Zensusdaten folgendermaßen ab:

1. Erfüllung der formalen Voraussetzungen

Um mit der Datennutzung beginnen zu können, muss zunächst ein Antrag zur Datennutzung gestellt werden. Nach der positiven Prüfung des Antrags wird der Institution der Nutzenden ein Vertrag zugeschickt. Sobald dieser von einem zeichnungsberechtigten Vertreter der Institution unterschrieben und zurückgesandt wurde, werden noch nicht verpflichtete Nutzer entsprechend § 16 Abs. 7 BStatG zur Geheimhaltung verpflichtet. Sind diese formalen Voraussetzungen erfüllt, kann die eigentliche Datennutzung beginnen.

2. Einarbeitung am GWAP

Nutzer der Zensusdaten sind verpflichtet, sich zunächst am GWAP mit den Daten vertraut zu machen. Dafür vereinbaren sie mit dem für sie am besten gelegenen Standort der FDZ einen Termin, an dem sie in die geschützten Räume der amtlichen Statistik kommen und dort mit den Daten arbeiten. Aus Gründen der Geheimhaltung wie auch der Performance wird dabei am GWAP nur eine Stichprobe des Datenmaterials bereitgestellt, die für die Daten des Gesamtdatensatzes des Zensus 2011 in etwa 2 % und für die Daten der Haushaltebefragung circa 10 % der befragten Anschriften umfasst. Am GWAP kann der Nutzer einen Einblick in die Daten und Ergebnisse gewinnen und Auswertungsprogramme schreiben, die von den Mitarbeitern der FDZ per KDFV auf das Gesamtmaterial angewendet werden.

3. Datenauswertung per KDFV

Über die KDFV werden die Ergebnisse erzeugt, die den Nutzenden nach einer obligatorischen Prüfung auf Geheimhaltung übermittelt werden. Für den Zensus 2011 bieten die FDZ zwei Möglichkeiten für die Erstellung der Auswertungsprogramme an:

- Schreiben der Auswertungsprogramme am GWAP: Die Datensätze für den GWAP und die KDFV entsprechen sich in ihrem Aufbau, so dass am GWAP erstellte Programme ohne weitere Anpassungen auch für den KDFV-Datensatz nutzbar sind. Das Schreiben der Auswertungsprogramme am GWAP hat den Vorteil, dass Ergebnisse direkt gesichtet und syntaktische wie semantische Fehler leicht erkannt werden können.
- Schreiben der Auswertungsprogramme auf Basis eines Strukturdatensatzes (SDS): Nach der Einarbeitung am GWAP wird dem Nutzer ein SDS zugeschickt, der in seinem Aufbau dem Originalmaterial entspricht, der aber keinerlei inhaltliche Information bietet. Anhand dieses SDS kann der Nutzer an einem Ort seiner Wahl Programme schreiben, die er dann an seinen Ansprechpartner im FDZ übermittelt. Das Schreiben der Auswertungsprogramme auf Basis eines SDS hat den Vorteil, dass Nutzer zum Schreiben der Syntaxen nicht an einen GWAP kommen müssen. Dafür können semantische Fehler erst nach der Prüfung und Übermittlung der Ergebnisse erkannt werden.

Bei beiden Varianten werden die Syntaxen von den Mitarbeitern der FDZ auf das KDFV-Datenmaterial angewendet. Der Nutzer erhält seine Ergebnisse nach einer obligatorischen Geheimhaltungsprüfung.

Nutzern steht es frei, zwischen den beiden Möglichkeiten zum Schreiben der Syntaxen zu wählen und zu wechseln.

6. Geheimhaltung des Gesamtdatenbestandes des Zensus 2011

Die Einzeldaten des Zensus 2011 unterliegen – wie alle Einzelangaben der amtlichen Statistik – dem Statistikgeheimnis nach § 16 BStatG. Wegen der großen Bedeutung, die das Vertrauen der Befragten in die Wahrung des Datenschutzes ihrer vertraulichen Angaben hat, bezieht die amtliche Statistik die Verpflichtung zur statistischen Geheimhaltung auch auf den Schutz von in Tabellen zusammengefassten Angaben, wenn die Gefahr besteht, dass dadurch die Verhältnisse eines einzelnen Auskunftspflichtigen offenbart werden.

6.1. Tabellengeheimhaltung

Im Folgenden werden die von den FDZ eingesetzten Verfahren zur Ergebnisgeheimhaltung von Häufigkeitstabellen beschrieben. Die Tabellen werden hier entsprechend dem von der Fachseite genutzten Vorgehen behandelt, das beispielsweise in der [Zensusdatenbank](#)¹⁰ angewandt wird. Bei Produkt 3 kommen verschiedene Verfahren zur Geheimhaltung zum Einsatz. Dies ist davon abhängig, welche Teiltabellen genutzt werden.

Die Einzeldaten der statistischen Einheiten Familie, Haushalt und Personen sowie Gebäude und Wohnungen wurden dem datenverändernden Geheimhaltungsverfahren [SAFE](#)¹¹ unterzogen. Durch SAFE wird der Rückschluss auf Einzelangaben verhindert, indem die Einzeldaten selbst leicht verändert werden. Dabei erfolgt die Datenveränderung so, dass jede in den Originaldaten existierende Merkmalskombination (z. B. aus Geschlecht, Familienstand, Alter usw.) in den geschützten Daten mindestens dreimal oder gar nicht mehr auftritt. Diese Änderungen werden kontrolliert vorgenommen, mit dem Ziel, dass sie sich weitgehend untereinander ausgleichen – die Randsummen der Tabellen sich also vor und nach der Datenveränderung möglichst nicht unterscheiden. Durch diese Form der Mikroaggregation werden die Einzelangaben ausreichend geschützt, gleichzeitig bleibt das Analysepotenzial der Daten erhalten.

Die Erstellung von Häufigkeitsauszählungen für Produkt 3 für die statistischen Einheiten Familie, Haushalt und Personen sowie Gebäude und Wohnungen aus den in Kapitel 5.1.1. genannten Teiltabellen erfolgt in den FDZ nur anhand der Nach-SAFE-Daten. Vor der Freigabe der Tabellen an den Nutzer werden die SAFE-gesicherten Ergebnisse noch mit den Vor-SAFE-Ergebnissen abgeglichen, wobei zu große Abweichungen von den Originalwerten analog dem Vorgehen der Zensus-Fachabteilungen gekennzeichnet werden (siehe Tabelle 5).

Tabelle 5: Kennzeichen der geheimhaltungsbedingten Abweichungen für den Nutzer

Symbol	Erläuterung
-	Nichts vorhanden (kann durch das Geheimhaltungsverfahren bedingt sein)
()	Aussagewert eingeschränkt, da der Zahlenwert durch das Geheimhaltungsverfahren relativ stark verändert wurde
.	Keine Angabe, da der Zahlenwert geheim zu halten ist oder durch das Geheimhaltungsverfahren zu stark verändert wurde

¹⁰ <https://ergebnisse.zensus2011.de/>

¹¹ SAFE steht für „Sichere Anonymisierung für Einzeldaten“. Weitere Hinweise finden Sie in der Nutzer-Kurzinformation zu SAFE unter: https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Nutzerhinweise_safe.pdf?blob=publicationFile&v=16

Die Daten der Haushaltebefragung wurden nicht mit SAFE geheim gehalten. Um sicherzustellen, dass Ergebnistabellen, die die Teiltabelle Haushaltsstichprobe beinhalten, belastbar sind und um Scheingenauigkeiten zu vermeiden, werden folgende Maßnahmen vor der Übermittlung von Ergebnistabellen an den Nutzer durchgeführt:

1. Einhaltung von zensuspezifischen Mindestfallzahlen

Damit Ergebnisse zuverlässig sind, muss eine von der regionalen Bezugsgröße abhängige Mindestfallzahl¹² eingehalten sein. Die Abhängigkeit von der regionalen Bezugsgröße ergibt sich durch unterschiedliche Stichprobengrößen. Die Sperrung von Tabellenfeldern, die auf zu geringe Besetzungszahlen zurückgehen, geschieht weniger aus Gründen der Geheimhaltung, sondern vor allem zur Vermeidung von statistisch nicht zuverlässigen Ergebnissen. Tabellen-Nullen werden gesperrt, da aufgrund der Stichprobeneigenschaft der Haushaltebefragung nicht sicher ist, ob eine Eigenschaft in der Grundgesamtheit tatsächlich nicht auftritt oder ob diese nur nicht in der gezogenen Stichprobe angetroffen wurde.

2. Rundung der Ergebnisse

Die Rundung der Tabellenwerte auf Basis der Haushaltsstichprobe auf volle Zehner geschieht zur Vermeidung von Scheingenauigkeiten. Die Rundung erfolgt kaufmännisch.

Da in der GWZ eine Totalerhebung durchgeführt wurde, ist in der Auswertung von GWZ-Daten dieses Vorgehen nicht notwendig. Die Sperrung entfällt ebenso bei Ergebnissen, die aus der Auszählung von Registern gewonnen wurden und zudem für Auswertungen von Haushalts- und/ oder Familienmerkmalen. Allerdings müssen die zensuspezifischen Mindestfallzahlen eingehalten und die Rundung durchgeführt werden sobald die Auswertung mit einem Merkmal aus der Haushaltsstichprobe kombiniert wird. Zur Auswertung von erwerbsstatistischen Merkmalen nach dem sogenannten Baukastenprinzip verweisen wir auf die Nutzerinformation zur Auswertung von Baukastenmerkmalen, die Sie gesondert zu Ihrer Nutzung erhalten können. Bei erwerbsstatistischen Auswertungen werden Auszählungen von Erwerbsregisterangaben mit Hochrechnungen der Ergebnisse aus der Haushaltebefragung auf Stichprobenbasis kombiniert. Je nach ausgewählter Merkmalskombination muss hierbei ein anderes Verfahren angewendet werden.

6.2. Geheimhaltung bei Verhältniszahlen

Verhältniszahlen (hierzu zählen auch Quoten, wie z. B. die Erwerbslosenquote) werden ebenfalls auf den Nach-SAFE-Daten erstellt. Der Nutzer hat hier die Möglichkeit, die Bestimmung von Schwellenwerten für eine adäquate Kennzeichnung von zu großen Abweichungen zu den Vor-SAFE-Daten selbst festzulegen. Hierzu werden dem Nutzer bei Bedarf ein Makro sowie ein Leitfaden zur Verfügung gestellt.

¹² Bei den hochgerechneten Zensusergebnissen aus der Haushaltsstichprobe werden die Ergebnisse mit zu geringen Besetzungszahlen nicht ausgewiesen, sondern durch einen Schrägstrich („/“) ersetzt.

Exkurs: SAFE

Das prä-tabulare SAFE-Verfahren wird für die Umsetzung der Tabellengeheimhaltung angewandt. Bei SAFE handelt es sich um ein Mikroaggregationsverfahren. Dabei werden Fälle durch Gruppenbildung und Austausch einzelner Angaben so verändert, dass jede Merkmalskombination in den Daten mindestens dreimal oder gar nicht vorkommt. Die Mikroaggregation wird dabei in der Regel so kontrolliert, dass die Veränderungen sich untereinander weitgehend ausgleichen, so dass die Abweichungen in den zentralen Auswertungstabellen minimiert werden.

Weiterführende Literatur

Statistische Ämter des Bundes und der Länder: Nutzer-Kurzinformation zu SAFE.

https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Nutzerhinweise_safe.pdf

Höhne, Jörg: Das Geheimhaltungsverfahren SAFE. In: Zeitschrift für amtliche Statistik Berlin Brandenburg. 2/2015, S.16–33.

https://www.statistik-berlin-brandenburg.de/produkte/zeitschrift/2015/HZ_201502.pdf

6.3. Geheimhaltung bei multivariaten Auswertungen

Für multivariate Analysen werden in den FDZ die Vor-SAFE-Daten verwendet. Die Geheimhaltung für derartige Analysen erfolgt entsprechend dem in den FDZ üblichen Vorgehen. Generell gilt dabei, dass Ergebnisse immer dann zu sperren sind, wenn sie dazu führen können, dass Einzelangaben einer Person zugeordnet werden können und damit das Risiko einer Re-Identifikation entsteht.

Die Geheimhaltung für weiterführende, multivariate Analysen, die die Haushaltsstichprobe beinhalten, erfolgt entsprechend dem in den FDZ üblichen Vorgehen. Weitere Informationen dazu finden sich im Metadatenreport zu Produkt 1.

Ausführlichere Informationen zur Geheimhaltung finden Sie an folgender Stelle:

- Statistische Ämter des Bundes und der Länder: Zensus 2011 – Methoden und Verfahren, S. 67-70

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethode_nVerfahren5121105119004.pdf?__blob=publicationFile

- Sarah Gießing et al.: Geheimhaltung beim Zensus 2011

https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2014_11_Destatis_Geheimhaltung.pdf

7. Datenqualität und Auffälligkeiten

Im Folgenden werden die Datenqualität sowie Auffälligkeiten der Daten des Zensus 2011 erläutert. Für Besonderheiten zu Produkt 1 und 2 sei an dieser Stelle auf die jeweiligen Metadatenreports der Produkte hingewiesen.

7.1 Abweichende Ergebnisse

Personen in sensiblen Sonderbereichen sowie im Ausland tätige Angehörige der Bundeswehr, der Polizeibehörden und des Auswärtigen Dienstes sowie ihre dort ansässigen Familien wurden bei der Ziehung der Haushaltsstichprobe für die Zusatzmerkmale nicht berücksichtigt. Die Ergebnisse von Auswertungen der Haushaltsstichprobe können daher von der Gesamtzahl der ausgewählten regionalen Einheit abweichen.

7.2 Abweichung zu Ergebnissen der Zensusdatenbank

Abweichungen der über die FDZ erstellten Merkmale zur Zensusdatenbank können je nach Merkmal aus verschiedenen Gründen entstehen:

1. Abweichungen bei demografischen Merkmalen: Demografische Merkmale werden in der Zensusdatenbank nicht über die Haushaltsstichprobe hochgerechnet, sondern aus den Registern ausgezählt. Registerauszählungen sind in der Zensusdatenbank am Hinweis „Auszählung aus dem bereinigten Registerbestand“ zu erkennen. Wurden Ergebnisse ausschließlich mit den Daten der Haushaltsstichprobe erzeugt, haben sie den Hinweis „Hochrechnung aus der Haushaltsstichprobe“. Ergebnisse, die sich sowohl aus Auszählungs- als auch als Hochrechnungsergebnissen zusammensetzen, sind mit dem Hinweis „Kombinierte Registerauszählung und Hochrechnung aus der Haushaltsstichprobe“ gekennzeichnet.
2. Abweichungen beim Merkmal „Beruf“: Für die Berechnung von Häufigkeiten zum Merkmal „Beruf“ sind nur die erwerbstätigen Personen relevant. Um bei der Auswertung des Berufs die Ergebnisse der Zensusdatenbank näherungsweise nachbilden zu können, muss daher der Filter ERWERBSTAT_KURZ = 11 gesetzt werden.
3. Abweichungen bei der Auswertung erwerbsstatistischer Merkmale: Die von der amtlichen Statistik veröffentlichten und auch in der Zensusdatenbank bereitgestellten Ergebnisse zu erwerbsstatistischen Merkmalen sind keine reinen Hochrechnungen der Stichprobe. Stattdessen setzen sich diese Ergebnisse aus einem Auszählungsteil und einem Hochrechnungsteil zusammen: Angaben aus den Erwerbsregistern werden, soweit sie vorliegen, ausgezählt. Da die Register aber keine Angaben für Selbstständige enthalten, werden diese aus den Daten der Haushalbefragung hochgerechnet. Werden beide Ergebnisse zusammengerechnet, ergibt sich die Erwerbsstruktur für die Gesamtbevölkerung. In der Zensusdatenbank sind diese Ergebnisse durch den Hinweis „Kombinierte Registerauszählung und Hochrechnung aus der Haushaltsstichprobe“ gekennzeichnet.

7.3 Fehlende Werte bei Identifikatoren

7.3.1 Fehlende Werte beim Gebäude- und Wohnungsidentifikator

Die Gebäude- und Wohnungs-ID wurde im Rahmen der HHGen vergeben. Im Rahmen der HHGen wurden die Daten der verschiedenen Datenquellen zusammengeführt. Dabei wurden auch statistische Wohnhaushalte generiert und Wohnungen zugeordnet. Bei den Fällen ohne zugewiesene Gebäude- und Wohnungs-ID handelt es sich um Personen, für die im Rahmen der HHGen keine Ver-

knüpfung mit einer Wohnung erzeugt werden konnte. Dies ist nicht ungewöhnlich und auf fehlenden/ nicht ausreichenden Wohnraum an der betreffenden Anschrift zurückzuführen.

7.3.2 Fehlende Werte bei der Haushaltsnummer

Die Datensätze, für die das Merkmal „HAUSHALTSNUMMER“ nicht gefüllt ist, sind allesamt nicht auswertungsrelevant. Es handelt sich durchgehend um Personen an Adressen mit neutralem Stichprobenausfall, i. d. R. um nicht klar abgrenzbare Adressen (z. B. Gebäude, die zu mehreren Adressen gehören (sog. „Eckhäuser“)) Der neutrale Stichprobenausfall wird dadurch ausgeglichen, dass die für die Hochrechnung relevanten Adressen einen höheren Hochrechnungsfaktor erhalten.

7.4 EU-Merkmale

In den Zensusdaten sind Merkmale enthalten, die von der EU vorgegeben wurden und die in allen Mitgliedstaaten inhaltsgleich erhoben wurden. Die Methode und die Stichtage im Jahr 2011 waren in den einzelnen Ländern der EU unterschiedlich. In der öffentlichen Datenbank des EU-Zensus („Census Hub“) können diese Zensusdaten für alle EU-Mitgliedstaaten abgerufen werden. Die dort abzurufenden Tabellen basieren auf voraggregierte Daten in vorher festgelegten Datenquadern. Für international vergleichende Forschung werden somit harmonisierte und vergleichbare Informationen bereitgestellt.

Weiterführendes zu Eurostat

Einen Überblick über die Zensusdaten der Mitgliedstaaten der Europäischen Union finden Sie hier:

<http://ec.europa.eu/eurostat/web/population-and-housing-census/overview>

<https://ec.europa.eu/CensusHub2>

In den FDZ-Produkten besteht zusätzlich die Möglichkeit, die deutschen Einzeldaten zu nutzen. Die harmonisierten EU-Merkmale erkennt man am Präfix „EU_“ vor dem eigentlichen Variablennamen (siehe Merkmalskatalog). Zu beachten ist, dass sich die Definitionen der EU-Merkmale teilweise von den nationalen Definitionen unterscheiden. Informationen hierzu enthalten die Schlüsselverzeichnisse.

7.5 Merkmal Glaubensrichtung

Die Beantwortung der Frage zur Glaubensrichtung war im Zensus 2011 freiwillig. Bei der Interpretation von Auswertungsergebnissen sollte berücksichtigt werden, dass dieses Merkmal nicht immer befüllt vorliegt.

7.6 Angaben zu Studierenden

Das Merkmal TAETIGKEIT_AKTUELL weist Studierende in der Ausprägung 8 "Student/-in" separat aus, im Merkmal ERWERBSTAT_AUSF sind sie in der Ausprägung 23 "Schüler/-innen und Studierende (nicht erwerbsaktiv)" enthalten. In diesen Kategorien werden nur Studierende ausgewiesen, die nicht erwerbsaktiv sind. Erwerbsaktive Studierende, die beispielsweise einen Nebenjob haben, sind darin nicht enthalten. Die Gesamtzahl der Studierenden in Deutschland lässt sich daher aus den oben genannten Merkmalen nicht ableiten.

7.7 Ergebnisse zu Haushalten und Familien

Im Zensus 2011 wurden im Unterschied zu anderen Haushaltsstatistiken, wie beispielsweise dem Mikrozensus, keine Wirtschaftshaushalte, sondern Wohnhaushalte erfasst. Alle Personen, die in einer Wohnung leben, gelten somit als Mitglieder desselben Haushalts.

Um die Wohnsituation eines Haushalts möglichst realitätsnah abbilden zu können, werden bei der Betrachtung von Haushalten und Familien neben den Personen am Hauptwohnsitz auch Personen am Nebenwohnsitz berücksichtigt, während sie bei anderen Auswertungen, z. B. der Einwohnerzahl, der Bevölkerung nach Erwerbsbeteiligung oder Bildung, nicht enthalten sind. Betrachtet man beispielsweise ein Paar mit zwei Kindern, von dem ein Partner aus beruflichen Gründen noch eine weitere Wohnung in einer anderen Gemeinde bewohnt, so ist es wichtig, einerseits die vierköpfige Familie realitätsgetreu widerzuspiegeln und andererseits nicht außer Acht zu lassen, dass zusätzlich in einer anderen Wohnung ein Einpersonenhaushalt besteht und diese Wohnung nicht leer steht. Zu beachten ist allerdings, dass Auswertungen von Personen nach Haushalts- bzw. Familienmerkmalen nur Personen mit Hauptwohnsitz beinhalten. Hier werden – ebenso wie zur Feststellung der Einwohnerzahlen – nur Personen mit alleinigem bzw. Hauptwohnsitz gezählt. In Deutschland wurden nicht meldepflichtige Personen bei der Bildung von Haushalten nicht berücksichtigt.

8. FAQ¹³

Warum gibt das EU-Merkmal „Größe des Ortes“ (EU-Merkmal "Locality" – „EU_LOC“) für Personen innerhalb einer Gemeinde verschiedene Größenklassen aus?

Das Merkmal „EU_LOC“ bildet die regionalen Einheiten nicht nach bestehenden Verwaltungsgrenzen ab, sondern mit Hilfe einer Zuordnung über Distanzen. Anschriften, die weniger als 200 m voneinander entfernt liegen, bilden eine gemeinsame LOC-Einheit. Alle Personen einer LOC-Einheit tragen dann zur Ausprägung der Größenklasse LOC bei.

Was gibt es bei der Einarbeitung am GWAP zu beachten?

Nutzer der Zensusdaten sind verpflichtet, sich zunächst am GWAP mit den Daten vertraut zu machen. Dafür vereinbaren sie mit dem für sie am besten gelegenen Standort der FDZ einen Termin, an dem sie in die geschützten Räume der amtlichen Statistik kommen und dort mit den Daten arbeiten. Aus Gründen der Geheimhaltung wie auch der Performance wird dabei am GWAP nur eine Stichprobe des Datenmaterials bereitgestellt, die für die Daten des Gesamtdatensatzes des Zensus 2011 in etwa 2% und für die Daten der Haushaltebefragung circa 10% der befragten Anschriften umfasst. Am GWAP kann der Nutzer somit einen Einblick in die Daten und Ergebnisse gewinnen und Auswertungsprogramme schreiben, die von den Mitarbeitern der FDZ per KDFV auf das Gesamtmaterial angewendet werden. Für die am GWAP zur Verfügung gestellten Substichproben wurden die Hochrechnungsfaktoren in den FDZ mittels einer einfachen gebundenen Hochrechnung an das Vollmaterial angepasst. Damit kann auch mit der Unterstichprobe auf die Gesamtbevölkerung geschlossen werden. Die durch die gebundene Hochrechnung entstehenden Ergebnis-Ungenauigkeiten werden in Kauf genommen, da am GWAP erzeugte Ergebnisse nicht veröffentlicht werden. Der GWAP dient aufgrund des großen Datenumfanges des Zensus nur zur Erstellung der Syntaxen und zur ersten Sichtung von Ergebnissen.

Welche Besonderheiten gibt es bei den Verknüpfungstabellen zu beachten?

Die Teiltabelle Generation Edge enthält die im Rahmen der HHGen gewonnenen Informationen zu Nachkommensbeziehungen bzw. gesetzlichen Vertretungen zwischen jeweils zwei Personen. Eine Beziehung zwischen zwei Personen wird über die Merkmale VORFAHRE_ID und NACHFAHRE_ID dargestellt. Das Merkmal GENERATIONTYP zeigt an, ob es sich bei der Beziehung um eine Eltern- oder Großeltern-Kind-Beziehung oder um eine Beziehung zwischen gesetzlichem Vertreter und dessen Mündel handelt.

Die Teiltabelle Partner Edge enthält die im Rahmen der HHGen gewonnenen Informationen zu Partnerschaftsbeziehungen zwischen jeweils zwei Personen, wobei jede Person lediglich an einer einzigen Partnerschaftsbeziehung teilnehmen kann. Eine Beziehung zwischen zwei Personen wird über die Merkmale LEFT_PARTNER_ID und RIGHT_PARTNER_ID dargestellt. Das Merkmal PARTNERTYP zeigt an, ob es sich bei der Beziehung um eine Ehe, eine eingetragene Lebenspartnerschaft, eine nichtehe-

¹³ Die Metadatenreports zu Produkt 1 und 2 enthalten weitere produktspezifische FAQs.

liche Lebensgemeinschaft oder eine nichteingetragene Lebenspartnerschaft (nur Stichprobenanschriften) handelt.

Grundsätzlich gilt für beide Beziehungsarten zu beachten, dass Beziehungen nur zwischen zwei Personen desselben Haushalts angezeigt werden können.

Welche Besonderheiten gibt es bei der Teiltabelle Person zu beachten?

In dieser Teiltabelle sind die endgültigen Personeninformationen nach HHGen abgelegt. Diese Tabelle ist die zentrale Auswertungstabelle für die Erstellung von Personenergebnissen. Die Summe der Datensätze je regionaler Einheit entspricht der Zahl der Einwohner je regionaler Einheit.

Folgende Besonderheiten sind zu beachten:

- Eine Person befindet sich an genau einer Anschrift.
- Jede Person ist genau einem oder keinem Haushalt zugeordnet.
- Eine Person ist Wohnungsnutzer genau einer oder keiner Wohnung.
- Jeder Person ist genau ein oder kein Stichproben-Einzeldatensatz zugeordnet.

Die Teiltabelle enthält zudem nur auswertungsrelevante Datensätze. Beispielsweise sind inaktive Personen oder Nebenwohnsitzpersonen aufgrund fehlender typisierter Auswertungsmerkmale nicht im Datensatz vorhanden.

Welche Besonderheiten gibt es bei der Teiltabelle Familie-National zu beachten?

Diese Teiltabelle enthält die Familienmerkmale nach nationaler Definition.

Es gilt:

- In einem Haushalt kann keine oder genau eine Kernfamilie National bestehen.
- Eine Kernfamilie National besteht aus mindestens zwei Personen, die durch Partner- oder Generationenbeziehungen (nur Eltern und Großeltern, keine sonstigen gesetzlichen Vertreter) verbunden sind.
- Inaktive Personen sind nie Teil einer Kernfamilie National.

Die Teiltabelle enthält zudem nur auswertungsrelevante Datensätze.

Welche Besonderheiten gibt es bei der Teiltabelle Familie-EU zu beachten?

Diese Teiltabelle enthält die Familienmerkmale nach Definition für die EU.

Es gilt:

- In einem Haushalt können beliebig viele Kernfamilien National EU gemäß den Verordnungen der EU Nr. 1201/2009 und Nr. 763/2008 inklusive Nebenwohnsitzbevölkerung bestehen.

- Eine Kernfamilie besteht aus mindestens zwei Personen, die durch Partner- oder Generationenbeziehungen (nur Eltern und Großeltern, keine sonstigen gesetzlichen Vertreter) verbunden sind.
- Inaktive Personen sind nie Teil einer Kernfamilie EU.

Die Teiltabelle enthält zudem nur auswertungsrelevante Datensätze.

Welche Besonderheiten gibt es bei der Teiltabelle Haushalt zu beachten?

Die Teiltabelle enthält die Haushaltsinformationen aus der HHGen und bildet damit die Haushaltszusammenhänge ab.

Es gilt:

- Ein Haushalt besteht aus einer Menge von Personen, wobei jeder Haushalt mindestens eine Person enthält.
- Haushalte, die aus ausschließlich freiwillig gemeldeten Personen bestehen, sind nicht im Datensatz enthalten.

Die Teiltabelle enthält zudem nur auswertungsrelevante Datensätze.

Anhang

Anhang 1: Abkürzungsverzeichnis

Anhang 2: Weiterführende Quellen und Literatur

Anhang 3: Fragebogen der Erhebungsteile

Anhang 1: Abkürzungsverzeichnis

AGR	Anschriften- und Gebäuderegister
AGS	Amtlicher Gemeindeschlüssel
BKU	Befragung zur Klärung von Unstimmigkeiten
BRRG	Beamtenrechtsrahmengesetz
BStatG	Bundesstatistikgesetz vom 22. Januar 1987 (BGBl. I S. 462, 565), das zuletzt durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749) geändert wurde
EG	Europäische Gemeinschaft
EU	Europäische Union
Eurostat	Statistisches Amt der Europäischen Union
FDZ	Forschungsdatenzentrum/Forschungsdatenzentren
GAB	Georeferenzierte Adressdaten Bund
GWAP	Gastwissenschaftlerarbeitsplatz
GWZ	Gebäude- und Wohnungszählung
HHGen	Haushaltegenerierung
ILO	Internationale Arbeitsorganisation
KDFV	Kontrollierte Datenfernverarbeitung
NELG	Nichteheliche Lebensgemeinschaft
RDB	Referenzdatenbestand
SDS	Strukturdatensatz
StichprobenV	Verordnung über Verfahren und Umfang der Haushaltebefragung auf Stichprobenbasis zum Zensusgesetz 2011
ZensG 2011	Zensusgesetz 2011 vom 8. Juli 2009 (BGBl. I S. 1781)
ZensVorbG 2011	Zensusvorbereitungsgesetz 2011 vom 8. Dezember 2007 (BGBl. I S. 2808), das durch Artikel 3 des Gesetzes vom 8. Juli 2009 (BGBl. I S. 1781) geändert worden ist

Anhang 2: Weiterführende Quellen und Literatur

Berg, Andreas und Bihler, Wolf (2011): Das Stichprobendesign der Haushaltsstichprobe des Zensus 2011. In: Wirtschaft und Statistik 4/2011, S. 317-328.

https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2011_04_Destatis_Das_Stichprobendesign_der_Haushaltsstichprobe_des_Zensus_2011.pdf?__blob=publicationFile&v=22, letzter Zugriff am 12.02.2016

Hofmeister, Katrin und Dr. Fürnrohr, Michael: Das Korrekturverfahren beim Zensus 2011. In: Bayern in Zahlen 6/2014, S. 310-318.

<https://www.statistik.bayern.de/veroeffentlichungen/epaper.php?pid=41876&t=1>, letzter Zugriff am 12.02.2016

Höhne, Jörg: Das Geheimhaltungsverfahren SAFE. In: Zeitschrift für amtliche Statistik Berlin Brandenburg. 2/2015, S.16-33.

https://www.statistik-berlin-brandenburg.de/produkte/zeitschrift/2015/HZ_201502.pdf, letzter Zugriff am 11.02.2016

Kleber, Birgit; Maldonado, Andrea; Scheuregger, Daniel und Ziprik, Katja: Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011. In: Wirtschaft und Statistik 7/2009, S.629-640.

https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Zensus/Zensus.pdf?__blob=publicationFile, letzter Zugriff am 11.02.2016

Kreuzmair, Ingrid und Reisch, Marco: ZENSUS 2011: Ablauf der Haushaltegenerierung. In: Bayern in Zahlen 9/2012, S. 615-624.

<https://www.statistik.bayern.de/veroeffentlichungen/epaper.php?pid=40830&t=1>, letzter Zugriff am 11.02.2016

Statistische Ämter des Bundes und der Länder (2015): Zensus 2011 Methoden und Verfahren

https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Zensus/ZensusBuLaMethodenVerfahren5121105119004.pdf?__blob=publicationFile, letzter Zugriff am 11.02.2016

Statistische Ämter des Bundes und der Länder (2014): Nutzer-Kurzinformation zu SAFE

https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Nutzerhinweise_safe.pdf?__blob=publicationFile&v=16, letzter Zugriff am 11.02.2016

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 1 – Haushaltsstichprobe Zensus 2011

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_1.pdf, letzter Zugriff am 11.02.2016

Statistische Ämter des Bundes und der Länder Forschungsdatenzentren (2015): Metadatenreport Produkt 2 – Gebäude- und Wohnungszählung Zensus 2011

http://www.forschungsdatenzentrum.de/bestand/zensus_2011/metadaten_produk_2.pdf, letzter Zugriff am 11.02.2016

Anhang 3: Fragebogen der Erhebungsteile

Den Fragebogen zur **Haushaltebefragung auf Stichprobenbasis** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Fragebogen_Haushaltebefragung.pdf?__blob=publicationFile&v=13

Den Fragebogen zur **GWZ** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Fragebogen_Gebaeude_und_Wohnungszaehlung.pdf?__blob=publicationFile&v=8

Den Fragebogen zur Erhebung über die Bevölkerung an **Anschriften mit sensiblen Sonderbereichen** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Sensible_Sonderbereiche.pdf?__blob=publicationFile&v=6

Den Fragebogen zur Erhebung über die Bevölkerung an **Anschriften mit Wohnheimen/ Gemeinschaftsunterkünften** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Wohnheime_und_Gemeinschaftsunterkuenfte.pdf?__blob=publicationFile&v=4

Den **kombinierten Fragebogen** zur Erhebung über die Bevölkerung an **Anschriften mit Wohnheimen/ Gemeinschaftsunterkünften und zur Haushaltebefragung auf Stichprobenbasis** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Kombinierte_Befragung.pdf?__blob=publicationFile&v=5

Den Fragebogen zur **Wiederholungsbefragung** zur Bewertung der Qualität der Zensusergebnisse finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Wiederholungsbefragung.pdf?__blob=publicationFile&v=5

Den Fragebogen zur **Klärung von Unstimmigkeiten** in den verschiedenen Datenquellen finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Klaerung_von_Unstimmigkeiten.pdf?__blob=publicationFile&v=6

Den Fragebogen zur **Klärung des Wohnsitzes** finden Sie unter:

https://www.zensus2011.de/SharedDocs/Downloads/DE/Fragebogen/Befragung_zur_Klaerung_des_Wohnsitzes.pdf?__blob=publicationFile&v=6

