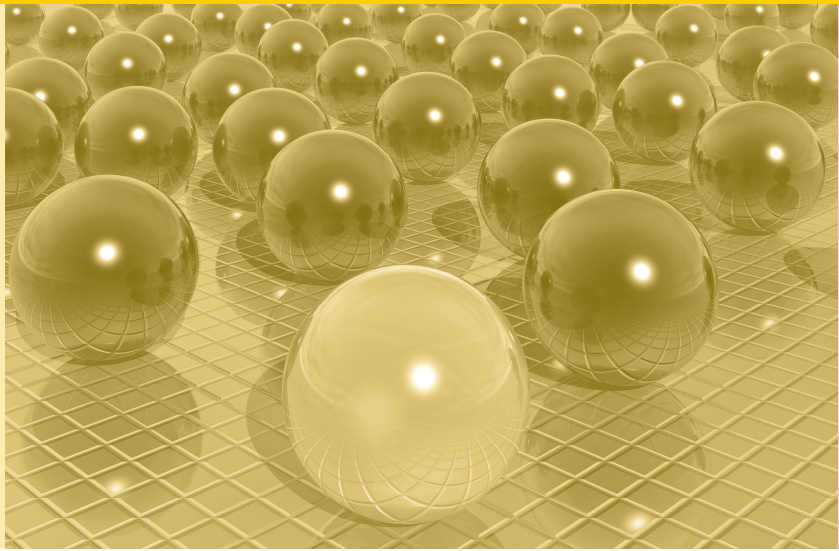


FDZ-Arbeitspapier Nr. 48



Bereitstellung der Daten des Zensus 2011 über die
Forschungsdatenzentren der Statistischen Ämter
des Bundes und der Länder

Stefanie Raab, Claudia Meisdrock

2015

Impressum

Herausgeber: Statistische Ämter des Bundes und der Länder
Herstellung: Information und Technik Nordrhein-Westfalen
Telefon 0211 9449-01 • Telefax 0211 9449-8000
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@it.nrw.de

Fachliche Informationen zu dieser Veröffentlichung:

Statistisches Bundesamt
Forschungsdatenzentrum

Tel.: 0611 75-4220
Fax: 0611 72-3915
forschungsdatenzentrum@destatis.de

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

Informationen zum Datenangebot:

Statistisches Bundesamt
Forschungsdatenzentrum

Tel.: 0611 75-4220
Fax: 0611 72-3915
forschungsdatenzentrum@destatis.de

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

Erscheinungsfolge: unregelmäßig
Erschienen im Oktober 2015

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

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2015
(im Auftrag der Herausbergemeinschaft)

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

Fotorechte Umschlag: ©artSILENCEcom – Fotolia.com

FDZ-Arbeitspapier Nr. 4,

Bereitstellung der Daten des Zensus 2011 über die
Forschungsdatenzentren der Statistischen Ämter
des Bundes und der Länder

Stefanie Raab, Claudia Meisdrock

2015

Bereitstellung der Daten des Zensus 2011 über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Stefanie Raab¹, Claudia Meisdrock²

Zusammenfassung:

Die hohe Komplexität der Methode und die Datenmenge des Zensus 2011 erforderten neue Wege der Datenbereitstellung in den Forschungsdatenzentren (FDZ). Anders als bei der Mehrzahl der Statistiken im Datenangebot der FDZ stehen für die Zensusdaten mehrere Produkte mit jeweils einem thematischen Schwerpunkt zur Verfügung. Für die On-Site Nutzung, d. h. am Gastwissenschaftlerarbeitsplatz und per kontrollierter Datenfernverarbeitung, können aktuell die Daten der Haushaltsstichprobe sowie die Daten der Gebäude- und Wohnungszählung für wissenschaftliche Zwecke ausgewertet werden.

In der vorliegenden Veröffentlichung werden neben dem Zensusmodell 2011 die verschiedenen Wege des Datenzugangs zu den Ergebnissen des Zensus 2011 aufgezeigt. Im Mittelpunkt steht die Möglichkeit, Einzeldaten über die FDZ auszuwerten.

Abstract:

The high complexity of the method and the huge amount of data of the 2011 census ask for new ways of data provision in the Research Data Centres (RDC). Unlike most of the statistics in the data offer of the RDC, there are several data products available for the census each of which has a different thematic priority. For the on-site use of data, that is safe centre and remote execution, there is currently data from the household survey and the census of buildings and housing available for scientific use.

In the present publication, the 2011 census model will be presented as well as the different ways of data access. The focus is on the possibility to analyse microdata via the RDC.

Schlüsselwörter: Zensus 2011, Datenzugang, Forschungsdatenzentren, Mikrodaten

¹ Forschungsdatenzentrum des Statistisches Bundesamts, E-Mail: stefanie.raab@destatis.de

² Forschungsdatenzentrum der Statistischen Ämter der Länder, Geschäftsstelle, E-Mail: forschungsdatenzentrum@it.nrw.de

Die Autorinnen bedanken sich bei den Mitgliedern der PG Zensus und den Vertretern der Fachseite für die wertvollen Hinweise und die Unterstützung, ohne die dieses Arbeitspapier nicht möglich gewesen wäre.

Inhalt

I. Der Zensus 2011	6
II. Die Methode des Zensus 2011	8
1. Das registergestützte Verfahren	8
2. Die Datenquellen des Zensus 2011	9
2.1 Gebäude- und Wohnungszählung	10
2.2 Haushaltsstichprobe.....	10
2.3 Befragungen in Wohnheimen/Gemeinschaftsunterkünften	10
2.4 Befragung zur Klärung von Unstimmigkeiten.....	11
2.5 Mehrfachfallprüfung	11
3. Zusammenführung der verschiedenen Datenquellen.....	11
III. Möglichkeiten des Datenzugangs	13
1. Zensusdatenbank.....	13
2. Zensus-Atlas	14
3. Download aggregierter Daten.....	14
4. Forschungsdatenzentren.....	14
4.1 Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder	14
4.2 Ablauf der Datennutzung für die Daten des Zensus 2011	15
4.2.1 Einarbeitung am Gastwissenschaftlerarbeitsplatz	16
4.2.2 Ergebniserstellung per kontrollierter Datenfernverarbeitung	17
4.3 Verfügbare Daten	18
4.3.1 Produkt 1 – „Haushaltsstichprobe“	19
a) Produkt 1: Enthaltene Merkmale.....	19
b) Produkt 1: Anonymisierung der Mikrodaten	19
c) Produkt 1: Geheimhaltung der Ergebnisse	20
d) Produkt 1: Datensatzgröße	21
e) Produkt 1: Schlüsselverzeichnis	21
4.3.2 Produkt 2 – „Gebäude- und Wohnungszählung“	21
a) Produkt 2: Enthaltene Merkmale.....	21
b) Produkt 2: Anonymisierung der Mikrodaten	22
c) Produkt 2: Geheimhaltung der Ergebnisse	22
d) Produkt 2: Datensatzgröße	24
e) Produkt 2: Schlüsselverzeichnis	24

f) Produkt 2: Besonderheiten bei der Auswertung.....	24
4.3.3 Weitere Produkte.....	24
4.4 Nutzungsbedingungen.....	24
4.4.1 Nutzungsberechtigte Einrichtungen.....	24
4.4.2 Beantragung der Datennutzung.....	25
4.4.3 Nutzungsdauer.....	25
4.4.4 Entstehende Kosten.....	25
IV. Literaturverzeichnis	27
V. Anhang	29
Anhang 1: Tipps für die Auswertung	30
Anhang 2: Regeln und Hinweise zur Syntaxerstellung für Zensus-Daten	32
Anhang 3: Rechtliche Grundlagen.....	34

Für einen besseren Lesefluss wurde in der vorliegenden Veröffentlichung das generische Maskulinum verwendet. Mit der männlichen Form eines Begriffs werden jeweils beide Geschlechter angesprochen.

I. Der Zensus 2011

Über das Datenangebot der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ) sind erste Daten des Zensus 2011 verfügbar. Damit stehen der Wissenschaft Einzeldaten der zum Stand 9. Mai 2011 durchgeführten Bevölkerungs-, Gebäude- und Wohnungszählung für eigene Auswertungen zur Verfügung.

Der Begriff „Zensus“ löst die früher genutzte Bezeichnung „Volkszählung“ ab. Beide Begriffe können synonym verwendet werden. Für den Zensus 2011 wurde abweichend von den vorhergehenden Erhebungen die Bezeichnung Zensus gewählt, um damit unter anderem auf die im Vergleich zu den „klassischen“ Volkszählungen modernere Erhebungsmethode hinzuweisen.

Eine Volkszählung dient u. a. dem Ziel, die Einwohnerzahlen für Bund, Länder und Gemeinden zu einem festgelegten Stichtag zu ermitteln. Die Durchführung von Volkszählungen ist allerdings sehr aufwendig und somit nur in größeren Zeitabständen möglich. Durch stetig ablaufende demografische Prozesse verändern sich jedoch Zahl und Zusammensetzung einer Bevölkerung zwischen zwei Volkszählungen nachhaltig. Zwischen zwei Volkszählungen ist daher eine regelmäßige Fortschreibung des Bevölkerungsbestandes nötig. Die hierbei entstehenden Fortschreibungsfehler bei der amtlichen Einwohnerzahl werden im Zuge einer erneuten Volkszählung korrigiert. Zusätzlich wurden beim Zensus 2011 u. a. Basisdaten zur Bevölkerung, zur Erwerbstätigkeit, zu Haushalten bzw. Familien und zur Wohnsituation erhoben, die Bund, Ländern, Kreisen und Kommunen als Planungsdaten dienen. Darüber hinaus werden die Daten des Zensus 2011 auch auf europäischer Ebene verwendet. Deshalb hat die Europäische Union (EU) festgelegt, dass ab dem Jahr 2011 alle Mitgliedstaaten im Abstand von 10 Jahren eine Volks-, Gebäude- und Wohnungszählung durchführen müssen.

Die letzte Volkszählung in der Bundesrepublik Deutschland fand im Jahr 1987 als Vollerhebung statt, d. h. dass damals alle Einwohner persönlich befragt wurden. Die persönliche Befragung der gesamten Bevölkerung Deutschlands ist jedoch mit einem hohen zeitlichen und personellen Aufwand und mit hohen Kosten verbunden. Um den Aufwand und die Kosten zu minimieren, wurde der Zensus 2011 registergestützt durchgeführt. Es wurden also bereits vorhandene Daten aus Verwaltungsregistern verwendet. Informationen, die in den Registern nicht vorlagen, wurden in der Regel durch zusätzliche Befragungen in Form von Voll- und Stichprobenerhebungen gewonnen.

Die hohe Komplexität der Methode und die Datenmenge erforderten neue Wege der Bereitstellung der Zensusdaten in den FDZ. Anders als bei der Mehrzahl der Statistiken im Datenangebot der FDZ stehen für die Zensusdaten mehrere Produkte mit jeweils einem thematischen Schwerpunkt zur Verfügung. Die Produkte werden zeitlich versetzt angeboten: Für die On-Site-Nutzung, d. h. am

Gastwissenschaftlerarbeitsplatz (GWAP) und per kontrollierter Datenfernverarbeitung (KDFV), können zunächst die Daten der Haushaltsstichprobe sowie die Daten der Gebäude- und Wohnungszählung ausgewertet werden. Die Bereitstellung weiterer Produkte erfolgt unter anderem in Abhängigkeit von der vorhandenen Nachfrage.

Die vorliegende Veröffentlichung bietet einen Überblick über die Methode des Zensus 2011 und zeigt die verschiedenen Wege des Datenzugangs zu den Ergebnissen des Zensus 2011 auf. Im Mittelpunkt die Frage, wie Einzeldaten über die FDZ ausgewertet werden können.

II. Die Methode des Zensus 2011

1. Das registergestützte Verfahren

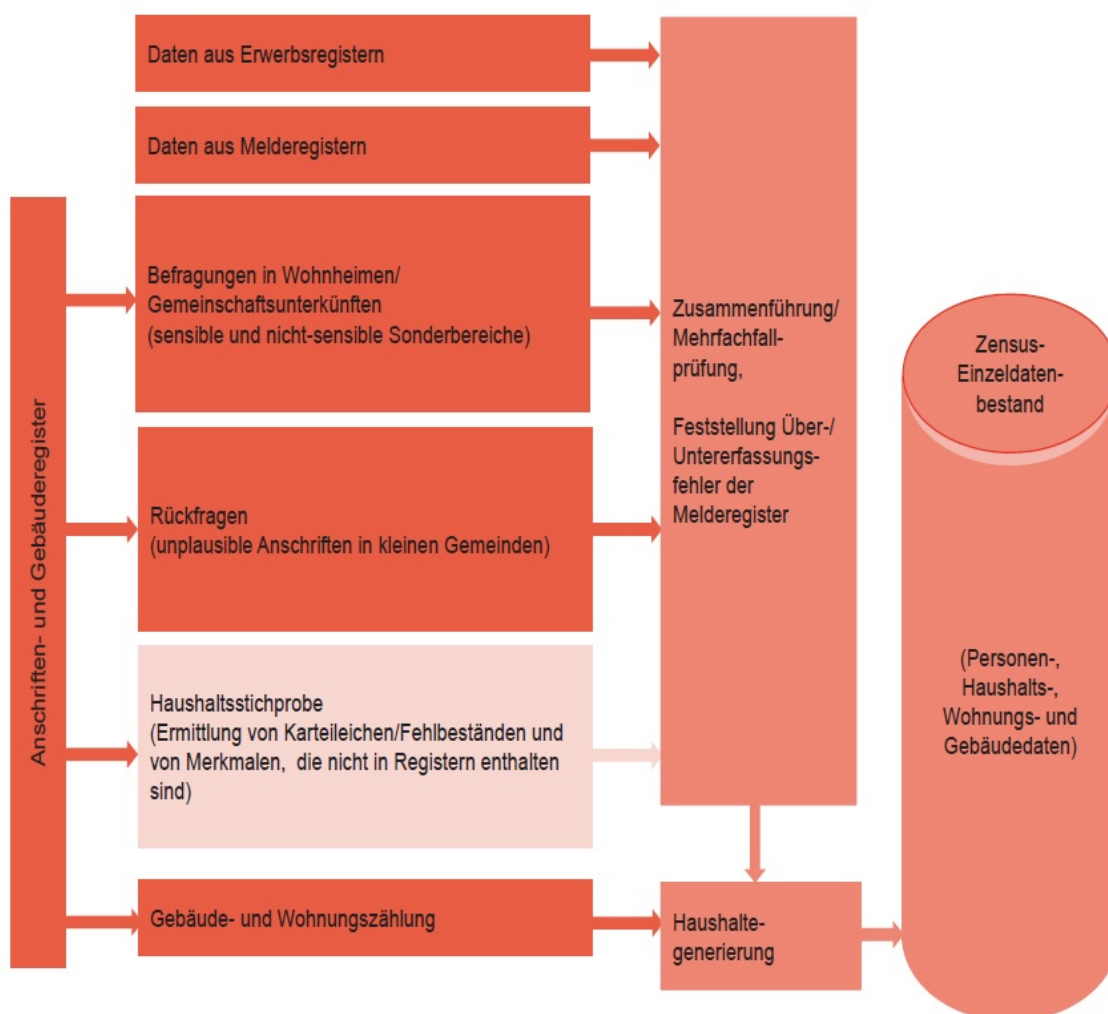
Beim Zensus 2011 wurde erstmals in Deutschland ein registergestütztes Verfahren eingesetzt, bei dem vorhandene Registerdaten der Verwaltungen mit den Ergebnissen unterschiedlicher Erhebungen ergänzt und Registerfehler statistisch korrigiert wurden. Diese Methode wurde in den Jahren 2001 bis 2003 von den Statistischen Ämtern des Bundes und der Länder umfassend getestet. Das Ergebnis dieses so genannten Zensustests war, dass die Voraussetzungen für die Durchführung eines registergestützten Zensus in Deutschland vorhanden sind und die Einwohnerzahlen sowie Daten zur Bevölkerungsstruktur registergestützt zuverlässig ermittelt werden können.

Resultierend aus den Erkenntnissen des Zensustests 2001 erfolgte für den Zensus 2011 der Aufbau eines Anschriften- und Gebäuderegisters. Dieses war über den gesamten Zensusablauf das Bindeglied zwischen allen Erhebungsteilen. Es diente dazu, die Erhebungen für den Zensus zu koordinieren, die aus verschiedenen Quellen stammenden Daten zusammenzuführen sowie die in den Zensus einzubeziehenden Gebäude, Wohnungen und Personen auf Vollständigkeit zu prüfen. Zudem bietet das Anschriften- und Gebäuderegister ein System für raumbezogene Analysen sowie die Schaffung einer Grundlage für kleinräumige Auswertungen des Zensus.

Für den Aufbau wurden ab dem Jahr 2008 u. a. verschiedene, bereits bestehende Verwaltungsregister genutzt: die Georeferenzierten Adressdaten Bund (GAB) der Vermessungsbehörden, die Daten der Meldebehörden und Daten der Personenregister der Bundesagentur für Arbeit.

Für die Durchführung des Zensus 2011 wurden Daten aus verschiedenen primären und sekundären Datenquellen verwendet. Die folgende Grafik gibt einen Überblick.

Das Zensusmodell im Überblick



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

2. Die Datenquellen des Zensus 2011

Die Basis des Zensus 2011 bilden die *Daten der Melderegister*, die alle gemeldeten Personen enthalten und die von den Meldebehörden zu drei Berichtszeitpunkten übermittelt wurden. Dabei dienten die zweite und die dritte Datenlieferung (zum Zensusstichtag sowie 3 Monate danach) der Ermittlung eines konsolidierten stichtagsbezogenen Bestands. Informationen zur Erwerbssituation der deutschen Bevölkerung lieferten u.a. die *Daten der Erwerbsregister*. Hierfür stellte die Bundesagentur für Arbeit die Daten für alle sozialversicherungspflichtigen Beschäftigten sowie zu allen arbeitslos und arbeitssuchend gemeldeten Personen zur Verfügung. Die öffentlichen Arbeitgeber lieferten entsprechende Daten für alle Beamten, Richter, Soldaten und Dienstordnungsangestellten.

Zur Gewinnung von noch nicht vorhandenen Informationen (z. B. zur Bildung oder zur Wohnsituation) sowie zur Sicherung der Qualität der Ergebnisse wurden ergänzende Erhebungen durchgeführt:

2.1 Gebäude- und Wohnungszählung

Zur Ermittlung der Grunddaten aller Gebäude und Wohnungen in Deutschland wurde eine Vollerhebung bei den Eigentümern bzw. den jeweiligen Verwaltern von Gebäuden und Wohnungen durchgeführt. Ziel der Erhebung war die Erfassung von Angaben zu allen Gebäuden mit Wohnraum und bewohnten Unterkünften in Deutschland.

2.2 Haushaltsstichprobe

Die Haushaltsstichprobe wurde als Stichprobenerhebung bei bundesweit knapp 10 % aller Adressen durchgeführt. Ziel der Haushaltsstichprobe war zum einen die Ermittlung der Über- und Untererfassungen der Melderegisterbestände. Durch eine Hochrechnung dieser sogenannten Karteileichen und Fehlbestände konnten die Daten der Melderegister statistisch korrigiert werden. Da der Zensustest ergeben hatte, dass die Melderegister in großen Gemeinden mit mindestens 10.000 Einwohnern deutlich ungenauer sind als in kleinen Gemeinden, erfolgte diese Korrektur nur in großen Gemeinden.

Weiteres Ziel der Haushaltsstichprobe war es, zusätzliche Angaben u. a. zur Bildung (z. B. dem Schulabschluss) sowie zur Erwerbstätigkeit (z. B. Informationen über Selbstständige und Angaben zum Erwerbsleben nach dem Konzept der Internationalen Arbeitsorganisation [ILO]) zu erfassen, die in den Registern nicht enthalten sind. Um die Informationen für ganz Deutschland hochrechnen zu können, wurde die Haushaltsstichprobe auch in Gemeinden mit weniger als 10.000 Einwohnern durchgeführt. Der Auswahlsatz der Stichprobe war hier in der Regel geringer und ermöglichte nur Rückschlüsse auf Kreisebene, nicht aber auf die einzelne Gemeinde.

Die Stichprobenziehung erfolgte auf Basis der Adressen mit Wohnraum und bewohnten Unterkünften des Adressen- und Gebäuderegisters. Für die Auswahl der Adressen wurde ein eigenes Stichprobendesign entwickelt. Befragt wurden jeweils alle an einer gezogenen Adresse mit Hauptwohnsitz gemeldeten Personen.

2.3 Befragungen in Wohnheimen/Gemeinschaftsunterkünften

Die Daten aller Bewohner von Wohnheimen und Gemeinschaftsunterkünften, den sogenannten Adressen mit Sonderbereichen, wurden ebenfalls mit dem Ziel gewonnen, die Karteileichen und Fehlbestände der Melderegisterdaten an diesen Adressen zu korrigieren. Darüber hinaus erfolgte eine Unterscheidung in sensible und nicht-sensible Sonderbereiche. Adressen mit nicht-sensiblen Sonderbereichen (z. B. Studentenwohnheime) konnten auch für die oben genannte Haushaltsstichprobe zufällig ausgewählt werden, um auch für den dort lebenden Personenkreis

zusätzliche Informationen zu gewinnen. An sensiblen Sonderbereichen (z. B. Justizvollzugsanstalten) wurden grundsätzlich keine zusätzlichen Merkmale erfragt.

2.4 Befragung zur Klärung von Unstimmigkeiten

In Gemeinden mit weniger als 10.000 Einwohnern erfolgte, wie bereits beschrieben, keine Korrektur der Melderegisterdaten durch die Ergebnisse der Haushaltsstichprobe. Traten an Adressen mit nur einer bewohnten Wohnung Unstimmigkeiten zwischen den Melderegisterdaten und den Informationen aus der Gebäude- und Wohnungszählung auf, wurden auch dort die Bewohner befragt.

2.5 Mehrfachfallprüfung

Hauptwohnsitz-Dubletten – also Personen, die laut Melderegister im Bundesgebiet mit mehr als einem Hauptwohnsitz gemeldet sind – oder alleinige Nebenwohnsitze – also Personen, die laut Melderegister lediglich mit einem oder mehreren Nebenwohnsitzen gemeldet sind – wurden mithilfe eines maschinellen Abgleichs bereinigt. Sobald allerdings einer der betroffenen Meldedatensätze in einer Gemeinde mit weniger als 10.000 Einwohnern war, erfolgte die Bereinigung mittels einer postalischen Erhebung (Befragung zur Klärung des Wohnsitzes), in der die Personen nach ihrem tatsächlichen Wohnstatus befragt wurden.

3. Zusammenführung der verschiedenen Datenquellen

Um Informationen über Haushaltszusammenhänge zu erhalten, wurden die Daten der unterschiedlichen Quellen zusammengeführt. Da die Haushaltszusammenhänge nicht vollständig in den Registern enthalten sind, wurden diese über das statistische Verfahren der Haushaltgenerierung automatisiert gebildet. Dabei wurden anschriftenweise Personen zu Haushalten zusammengefasst und den erfassten Wohnungen zugeordnet (Wohnhaushalte). Dadurch stehen neben den Informationen zu Personen, Gebäuden und Wohnungen auch Angaben über Haushalte und Familien zur Verfügung.

Weiterführende Literatur zur Methode des Zensus:

Ausführliche Beschreibungen der einzelnen Datenquellen, der verwendeten Verfahren sowie detaillierte Darstellungen zum Stichprobendesign, zum Hochrechnungserfahren und zur Geheimhaltung liefert der von den Statistischen Ämtern des Bundes und der Länder im Juni 2015 veröffentlichte Methodenband „[Zensus 2011: Methoden und Verfahren](#)“:

Statistische Ämter des Bundes und der Länder: Methoden und Verfahren beim Zensus 2011, 2015.
https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf

Darüber hinaus sind auf der Zensus-Seite www.zensus2011.de weitere Informationen und Artikel zu einzelnen Themengebieten rund um den Zensus 2011 zu finden. Über die Suchmaske kann gezielt nach Veröffentlichungen zu Schwerpunktthemen wie z. B. dem Anschriften- und Gebäuderegister oder Stichprobenverfahren recherchiert werden.

Statistische Ämter des Bundes und der Länder: Zensus 2011. Wissen, was morgen zählt.
<https://www.zensus2011.de>

III. Möglichkeiten des Datenzugangs

1. Zensusdatenbank

Für den Zensus 2011 gibt es die Möglichkeit, online mit nur wenigen Klicks eigene Ergebnisse zu erstellen. Möglich wird das durch die [Zensusdatenbank](#). Diese bietet drei Möglichkeiten für die Auswertung der Daten:

➤ Ergebnisse einfach und schnell

In diesem Auswertungsbereich können Einwohnerzahlen sowie vordefinierte Tabellen abgerufen werden. Das Angebot ist nach Themen geordnet und kann für unterschiedliche regionale Bezüge angezeigt werden. „Ergebnisse einfach und schnell“ bietet einen schnellen Überblick über die wichtigsten Kennzahlen einer Region.

➤ Ergebnisse individuell und dynamisch

In diesem Auswertungsbereich können Tabellen und Diagramme den eigenen Wünschen und Anforderungen entsprechend selbst zusammengestellt werden. Der Vergleich von Gemeinden oder Regionen ist möglich. „Ergebnisse individuell und dynamisch“ bietet einen einfachen Einstieg in individuelle Auswertungen.

➤ Ergebnisse kartografisch und visuell

In diesem Auswertungsbereich können verschiedene Indikatoren kartografisch dargestellt werden. Die Darstellung ist auf Landes-, Kreis- oder Gemeindeebene möglich. Die Klassenzahl sowie die Klassengrenzen können an die individuellen Bedürfnisse angepasst werden.

Um bei den Tabellen der Zensusdatenbank die Geheimhaltung der Ergebnisse und damit den Datenschutz gewährleisten zu können, wurden die Daten mit dem datenverändernden Geheimhaltungsverfahren SAFE („Sichere Anonymisierung für Einzeldaten“) anonymisiert. Anders als bei klassischen Geheimhaltungsmethoden wird ein Rückschluss auf statistische Einzelangaben verhindert, indem die Originaldaten leicht verändert wurden. Eine Beschreibung des SAFE-Verfahrens und dessen Auswirkungen werden in der [Nutzer-Kurzinformation zu SAFE](#) beschrieben.

Weiterführendes zur Zensusdatenbank:

Fischer, Christoph: Zensus 2011 – Die Zensusdatenbank nach Abschluss der Datenaufbereitung. In: Statistisches Monatsheft Baden-Württemberg 6/2014

Statistische Ämter des Bundes und der Länder: Zensusdatenbank. <https://ergebnisse.zensus2011.de/>

2. Zensus-Atlas

Über den [Zensus-Atlas](#) können online Rasterkarten zu verschiedenen Themen angezeigt werden. Die Darstellung der Karten erfolgt auf 1-km²-Rastern. Je nach Bedarf können einzelne Anzeigeebenen wie die Verwaltungsgrenzen, Gebietsnamen oder Gewässer ein- oder ausgeblendet werden. In der Karte kann die Ansicht für ganz Deutschland erfolgen, durch Hineinzoomen können aber auch kleinere Gebiete betrachtet werden.

Weiterführendes zum Zensus-Atlas:

Statistische Ämter des Bundes und der Länder: Zensus-Atlas. <https://atlas.zensus2011.de/>

3. Download aggregierter Daten

Zusätzlich zu den Angeboten der Zensusdatenbank und des Zensus-Atlas stehen CSV- und teilweise Excel-Tabellen mit umfassenden Personen-, Haushalts- und Familienmerkmalen sowie Angaben zu Gebäuden und Wohnungen zur Verfügung. Die Ergebnisse liegen auf Bundes-, Länder-, Kreis- und Gemeindeebene vor.

Auch die dem Zensus-Atlas zugrunde liegenden Daten können kostenfrei heruntergeladen werden. Diese liegen auf Gitterzellen-Ebene vor. Die Kantenlänge der Gitterzellen liegt dabei für die Bevölkerungszahl bei 100 Metern, für die weiteren Merkmale bei einem Kilometer.

Diese aggregierten Daten können [hier](#) kostenfrei heruntergeladen und individuell ausgewertet werden.

Weiterführendes zum Download aggregierter Zensus-Daten:

Statistische Ämter des Bundes und der Länder: Ergebnisse des Zensus 2011 zum Download. <https://www.zensus2011.de/SharedDocs/Aktuelles/Ergebnisse/DemografischeGrunddaten.html>

4. Forschungsdatenzentren

4.1 Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Das Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ der Länder) und das Forschungsdatenzentrum des Statistischen Bundesamts (FDZ des Bundes) ermöglichen der Wissenschaft den Zugang zu amtlichen Mikrodaten für wissenschaftliche Analysen.

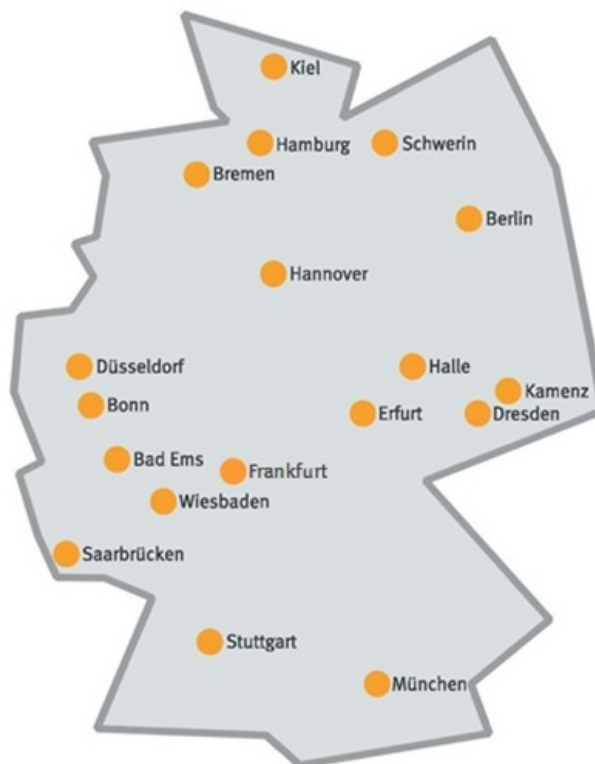
Die Bereitstellung von Mikrodaten wird durch § 16 (6) Bundesstatistikgesetz (BStatG) ermöglicht. Danach dürfen die FDZ der amtlichen Statistik Mikrodaten „an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung [übermitteln], wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft

zugeordnet werden können und die Empfänger Amtsträger, für den öffentlichen Dienst besonders Verpflichtete oder Verpflichtete nach Absatz 7 sind“³.

Auf dieser Grundlage stellen die FDZ gegenwärtig über 100 verschiedene Statistiken aus den Bereichen „Soziales“, „Wirtschaft, Finanzen und Steuern“, „Rechtspflege“ sowie „Agrar, Energie und Umwelt“ mit häufig mehreren Erhebungswellen bereit. Eine Übersicht über alle verfügbaren Statistiken finden Sie [hier](#).

Standorte des FDZ der Länder finden sich in allen Standorten der Statistischen Landesämter. Das FDZ des Bundes hat Standorte in Wiesbaden, Bonn und Berlin. Darüber hinaus betreibt es zusammen mit dem Hessischen Statistischen Landesamt einen Standort in Frankfurt am Main. Eine Übersicht über die Standorte der FDZ bietet nebenstehende Abbildung.

Weitere Informationen zu den Forschungsdatenzentren bietet der gemeinsame [Internetauftritt der FDZ des Bundes und der Länder](#)



4.2 Ablauf der Datennutzung für die Daten des Zensus 2011

Die Nutzung der Mikrodaten des Zensus 2011 muss beantragt werden (s. Kapitel III.4.4.2). Die Auswertung der Daten kann nach der Bewilligung des Nutzungsantrags, der Unterzeichnung des Nutzungsvertrags und der Verpflichtung des Nutzers beginnen.

Die Auswertung der Daten des Zensus 2011 erfolgt durch eine kombinierte Nutzung von Gastwissenschaftlerarbeitsplatz (GWAP) und kontrollierter Datenfernverarbeitung (KDFV). Standardmäßig können die Daten mit den Auswertungsprogrammen SAS, SPSS und Stata ausgewertet werden.⁴

³ 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 worden ist (BStatG), §16 (6)

⁴ Ausnahme: Im Standort Saarbrücken ist SAS nicht verfügbar

Übersicht: Grade der Anonymität und Arten der Datenbereitstellung

In den FDZ werden folgende Grade der Anonymität unterschieden:

- Formale Anonymität
Schwächster Grad der Anonymität verbunden mit dem höchsten Analysepotenzial. Zum Erreichen der formalen Anonymität werden lediglich die direkten Identifikatoren aus dem Datenmaterial gelöscht. Formal anonyme Daten können nur über die kontrollierte Datenfernverarbeitung genutzt werden.
- Faktische Anonymität
Mittlerer Grad der Anonymität verbunden mit einem hohen Analysepotenzial. Zum Erreichen der faktischen Anonymität werden weitere Anonymisierungsmaßnahmen durchgeführt, z.B. die Verringerung der regionalen Gliederungstiefe. Eine Reidentifikation der Merkmalsträger ist dann nur noch mit einem unverhältnismäßig hohen Aufwand an Zeit, Kosten und Arbeitskraft möglich. Faktisch anonyme Daten können am Gastwissenschaftlerarbeitsplatz (On-Site) und über Scientific-Use-Files (Off-Site) genutzt werden. Die Daten der Scientific-Use-Files sind stärker anonymisiert, da sie außerhalb der geschützten Räume der amtlichen Statistik – aber nur in den Räumen der beantragenden wissenschaftlichen Einrichtung – genutzt werden dürfen. Scientific-Use-Files stehen für den Zensus 2011 derzeit nicht zur Verfügung.
- Absolute Anonymität
Stärkster Grad der Anonymität verbunden mit dem geringsten Analysepotenzial. Zum Erreichen der absoluten Anonymität werden weitere Anonymisierungsmaßnahmen durchgeführt, wie beispielsweise die Vergrößerung von Merkmalen zur Umsetzung von Mindestfallzahlen. Eine Reidentifikation kann dann nach menschlichem Ermessen ausgeschlossen werden. Absolut anonyme Daten können über Public-Use-Files oder CAMPUS-Files genutzt werden. Absolut anonyme Daten stehen für den Zensus 2011 derzeit nicht zur Verfügung.

4.2.1 Einarbeitung am Gastwissenschaftlerarbeitsplatz

Der GWAP ermöglicht es Nutzern, faktisch anonyme Daten in den Räumen der Statistischen Ämter des Bundes und der Länder auszuwerten. Da die Daten faktisch anonymisiert vorliegen, können die Nutzer am GWAP direkt mit den Daten arbeiten und diese einsehen. Da das Deanonymisierungsrisiko damit steigt, müssen die Daten in den sicheren Räumen der Statistischen Ämter des Bundes und der Länder verbleiben. Der Nutzer kommt daher in den für ihn am besten gelegenen Standort der FDZ, um dort mit den Daten zu arbeiten. Dafür steht ihm ein abgeschotteter Arbeitsplatz-PC zur Verfügung, der keinen Zugriff auf das Internet oder andere Laufwerke zulässt (Analyse-PC). Um Recherchen zu ermöglichen steht an vielen Standorten ein zweiter PC mit Internetverbindung bereit. Während dem Nutzungszeitraum kann ein Nutzer den GWAP mehrfach besuchen. Vor dem Besuch eines GWAP muss mit dem betreuenden Standort immer ein Termin vereinbart werden.

Für die Nutzung der Daten des Zensus 2011 besteht die Verpflichtung, sich zunächst am GWAP mit den Daten vertraut zu machen. Während diesem Besuch zur Einarbeitung am GWAP

können bereits Syntaxen⁵ erstellt werden, die anschließend per KDFV auf die Daten angewandt werden.

Für die Nutzung der Daten des Zensus 2011 am GWAP gelten folgende Einschränkungen:

- Um die faktische Anonymität der Daten und kürzere Programmlaufzeiten zu gewährleisten, arbeiten Nutzer am GWAP lediglich mit einer Stichprobe des Datenmaterials.
- Am GWAP erzeugte Ergebnisse werden nicht freigegeben. Stattdessen werden die am GWAP erstellten Syntaxen von den FDZ-Mitarbeitern auf das volle Datenmaterial per KDFV angewandt und erst dann – nach Prüfung auf Geheimhaltung – an den Nutzer übermittelt.

Für alle Zensus 2011-Produkte – ausgenommen die Haushaltsstichprobe – gilt außerdem Folgendes:⁶

- Häufigkeitstabellen und Grafiken werden auf Basis der Nach-SAFE-Daten erstellt. Große Abweichungen zwischen Vor- und Nach-SAFE-Daten werden vor der Freigabe der Ergebnisse gekennzeichnet.
- Der Nutzer erhält am GWAP eine Stichprobe der Vor-SAFE-Daten. Syntaxen für Häufigkeitstabellen und Grafiken müssen so geschrieben werden, dass sie ohne Anpassungen sowohl auf den Vor- als auch auf den Nach-SAFE-Datenbestand angewandt werden können (s. Anhang 2: Regeln und Hinweise zur Syntaxerstellung für Zensus-Daten).

4.2.2 Ergebniserstellung per kontrollierter Datenfernverarbeitung

Die KDFV ermöglicht die Analyse formal anonymisierter Mikrodaten. Da das Reidentifikationsrisiko bei formal anonymisierten Daten sehr hoch ist, erhält der Nutzer keinen direkten Zugang zu den Daten. Für die Daten des Zensus 2011 stehen zwei Möglichkeiten zur Datenauswertung per KDFV zur Verfügung:

- 1) Nutzung des GWAP zur Syntaxerstellung. Auch nach der Einarbeitungsphase kann der GWAP zur Erstellung von Syntaxen genutzt werden. Dieses Vorgehen hat den Vorteil, dass Ergebnisse direkt eingesehen und logische Fehler unter Umständen direkt behoben werden können. Dafür ist der Besuch eines GWAP nötig.
- 2) Nutzung eines Strukturdatensatzes. Dem Nutzer wird ein so genannter Strukturdatensatz übermittelt, der in seinem Aufbau dem formal anonymisierten Datensatz entspricht, aber keine inhaltlichen Informationen enthält. Mithilfe dieses Datensatzes und den Metadaten kann der Nutzer in seiner Institution Syntaxen erstellen, die er dann an die Mitarbeiter des betreuenden FDZ-Standorts sendet. Dabei ist es empfehlenswert, die am Arbeitsort erstellte Syntax am

⁵ Der Begriff Syntax wird im Folgenden für Syntax des Statistiksoftwareprogramms SPSS, für Stata-Do-files oder SAS-Programme verwendet.

⁶ Die Daten der Haushaltsstichprobe wurden nicht mit SAFE geheim gehalten. Eine Unterscheidung zwischen Vor- und Nach-SAFE-Daten entfällt für diese daher.

GWAP-Standort auf Fehler zu überprüfen. Es hat sich gezeigt, dass logische Fehler, die erst nach der Geheimhaltungsprüfung und Ergebnisübermittlung identifiziert wurden, den Zeitaufwand erhöhen.

In beiden Fällen werden die erstellten Syntaxen von den FDZ-Mitarbeitern auf das formal anonymisierte Datenmaterial angewandt. Die über die KDFV erstellten Ergebnisse werden von den FDZ-Mitarbeitern auf Geheimhaltung geprüft und dem Nutzer anschließend zugeschickt.

Während der Datennutzung (in der Regel drei Jahre) kann ein Nutzer mehrere Syntaxen einschicken. Von ihm erstellte und weiterhin benötigte Datensätze werden während der Nutzungsdauer gespeichert. Die Regeln zur Syntaxerstellung entsprechen den in Anhang 2 beschriebenen Vorgaben.

4.3 Verfügbare Daten

Um sowohl den Anforderungen der Geheimhaltung der Einzelangaben als auch dem Wunsch der Forschenden nach einem möglichst hohen Analysepotenzial nachzukommen, bieten die FDZ unterschiedlich stark anonymisierte Produkte über verschiedene Zugangswege an. Für den Zensus 2011 sind derzeit nur eine Kombination der beiden On-Site-Zugangswege (GWAP und KDFV), verfügbar, bei der die Daten in den geschützten Räumen der Statistischen Ämter des Bundes und der Länder verbleiben (s. Kapitel III.4.2).

Im Unterschied zum üblichen Vorgehen der FDZ wird für die Daten des Zensus 2011 nicht nur ein Produkt⁷ je Zugangsweg und Anonymisierungsgrad erstellt, sondern mehrere standardisierte Produkte bereitgestellt, die je einen thematischen Schwerpunkt haben. Hintergrund hierfür ist der große Datenumfang des Zensus, der Auswertungen über den gesamten Datensatz sehr aufwendig macht und der bei Analysen über das gesamte Datenmaterial untragbar lange Programmlaufzeiten verursachen würde. Ist die Untersuchung einer Forschungsfrage nicht mit diesen thematischen Produkten umsetzbar, wird nach einer Merkmalsauswahl auch die Nutzung des Gesamtmaterials ermöglicht.⁸

Die Bereitstellung der Zensus 2011-Produkte der FDZ erfolgt in mehreren zeitlich versetzten Stufen. Dies geschieht aus zwei Gründen: Zum einen wird für die Aufbereitung der Produkte unterschiedlich viel Zeit benötigt, weshalb die Mikrodaten zu unterschiedlichen Zeitpunkten bereitgestellt werden. Zum anderen können in späteren Stufen Erfahrungen mit den Wünschen und Anforderungen der Nutzer berücksichtigt werden, um diesen maßgeschneiderte Produkte bereitzustellen.

⁷ Ein Produkt bezeichnet die Daten einer Statistik, eines Erhebungsjahres und eines Zugangsweges.

⁸ Diese Daten stehen gegenwärtig noch nicht zur Verfügung.

4.3.1 Produkt 1 – „Haushaltsstichprobe“

a) Produkt 1: Enthaltene Merkmale

Die Haushaltsstichprobe des Zensus 2011 umfasst Daten auf Personenebene für etwa 10 % der Bevölkerung. Die Ergebnisse der Haushaltsstichprobe sind nur für Gemeinden mit mindestens 10.000 Einwohnern sowie flächendeckend ab Kreisebene repräsentativ (s. Kapitel II.2.2).

Folgende Inhalte sind in Produkt 1 – Haushaltsstichprobe enthalten:

- Wohnstatus (Haupt-/Nebenwohnung)
- Demografie (inkl. Migration)
- Berufliche Ausbildung, Bildung und Schulbesuch
- Erwerbstätigkeit und Arbeitssuche
- Glaube und Religion
- Tiefste regionale Gliederungsebene (KDFV und GWAP): Gemeinde

Einen genauen Überblick über die in Produkt 1 – „Haushaltsstichprobe“ enthaltenen Merkmale bietet das [Schlüsselverzeichnis](#).

b) Produkt 1: Anonymisierung der Mikrodaten

Folgende Maßnahmen wurden bei der Erstellung von Produkt 1 – „Haushaltsstichprobe“ zur On-Site-Nutzung durchgeführt, um die Anonymität der Daten zu gewährleisten:

Formale Anonymisierung des Datenmaterials (Nutzung per KDFV):

- Löschen der direkten Identifikatoren
- Bilden von systemfreien Nummern

Faktische Anonymisierung des Datenmaterials (Nutzung am GWAP):

- Ziehen einer 10 %-Unterstichprobe der Gesamt-Haushaltsstichprobe auf Basis der Anschriften

Die Stichprobenziehung am GWAP erfolgt zur Sicherstellung der faktischen Anonymität der Daten. Zusätzlich können dadurch die Programmlaufzeiten verkürzt werden. Finale Ergebnisse werden, wie in Kapitel III.4.2. beschrieben, auf Basis des KDFV-Datenmaterials erzeugt.

Nutzer können mit der Stichprobe am GWAP Ergebnisse sichten und auf ihre Verwertbarkeit prüfen. Die am GWAP erstellten Syntaxen werden ebenso wie die auf Basis des Strukturdatensatzes erstellten Syntaxen von den FDZ-Mitarbeitern per KDFV auf das Vollmaterial angewandt. Am GWAP anhand der Stichprobe erstellte Ergebnisse werden nicht an den Nutzer weitergegeben.

Eine genaue Übersicht über die für Produkt 1 – „Haushaltsstichprobe“ durchgeführten Aufbereitungsschritte und die Anonymisierungsmaßnahmen bietet der [Metadatenreport für Produkt 1](#).

c) Produkt 1: Geheimhaltung der Ergebnisse

Die Ergebnisgeheimhaltung erfolgt für Produkt 1 post-tabular.

Exkurs: Ansätze zu Geheimhaltung von Ergebnissen

Geheimhaltungsverfahren lassen sich in prä- und post-tabulare Verfahren unterscheiden. Gemeinsam ist beiden, dass die darauf abzielen, die Aufdeckung von Einzelangaben zu verhindern.

Prä-tabulare Verfahren setzen vor der Erstellung der Ergebnisse an. Bei ihrer Anwendung werden die Daten selbst so verändert, dass erzeugte Ergebnisse immer (sozusagen automatisch) geheim gehalten sind, ohne dass dafür weitere Schritte unternommen werden müssen. Ein Beispiel für eine prä-tabulare Geheimhaltung ist das oben vorgestellte SAFE-Verfahren.

Post-tabulare Verfahren setzen dagegen erst nach der Erstellung der Ergebnisse an. Durch sie werden die Ergebnisse derart verändert, dass Rückschlüsse auf Einzelangaben nicht mehr möglich sind. Bekannte Verfahren der post-tabularen Geheimhaltung sind beispielsweise das Sperren von kritischen Angaben oder die Rundung bzw. Überlagerung von Werten.

Die Geheimhaltung von Häufigkeitstabellen erfolgt in zwei Stufen: In einer ersten Stufe wird geprüft, ob die Tabellenzellen ausreichend stark besetzt sind. Übersteigt der geschätzte relative Standardfehler mehr als 15 %, wird das Ergebnis unterdrückt. Anschließend erfolgt eine kaufmännische Rundung der Tabellenwerte auf volle Zehner. Diese wird nicht nur aus Geheimhaltungsgründen durchgeführt, sondern auch, um Scheingenauigkeiten zu vermeiden.

Weiterführende multivariate Analysen werden entsprechend dem üblichen Verfahren der FDZ geheim gehalten (vgl. Anhang 1 – Tipps für die Auswertung).

Eine genaue Übersicht über die Geheimhaltung für das Produkt 1 – „Haushaltsstichprobe“ bietet der [Metadatenreport für Produkt 1](#).

d) Produkt 1: Datensatzgröße

Das über die FDZ bereitgestellte Zensus-Produkt 1 – „Haushaltsstichprobe“ hat folgende Größe:

	KDFV	GWAP
Anzahl der Merkmale	104	104
Anzahl der Fälle	7.620.007	760.370
Speicherplatz Datenmaterial:		
SAS	3,1 GB	300 MB
SPSS	6,0 GB	600 MB
Stata	2,4 GB	240 MB

e) Produkt 1: Schlüsselverzeichnis

Das Schlüsselverzeichnis für das Zensus-Produkt 1 – „Haushaltsstichprobe“ findet sich [hier](#).

Weitere Informationen zu den Merkmalen des Produkts 1 – „Haushaltsstichprobe“ enthält der [Metadatenreport für Produkt 1](#).

4.3.2 Produkt 2 – „Gebäude- und Wohnungszählung“

a) Produkt 2: Enthaltene Merkmale

Die Gebäude- und Wohnungszählung (GWZ) des Zensus 2011 ist eine Vollerhebung und umfasst damit Daten auf Gebäude- bzw. Wohnungsebene für alle bewohnten Gebäude und Wohnungen. Folgende Inhalte sind in Produkt 2 – Gebäude- und Wohnungszählung u. a. enthalten:

- Baujahr
- Eigentumsform
- Gebäudetyp
- Zahl der Wohnungen im Gebäude
- Nutzung der Wohnung (vermietet, leer stehend etc.)
- Wohnfläche
- Ausstattung der Wohnung mit Bad und WC
- Tiefste regionale Gliederungsebene (KDFV und GWAP): Gemeinde

Einen genauen Überblick über die in Produkt 2 – „Gebäude- und Wohnungszählung“ enthaltenen Merkmale bietet das [Schlüsselverzeichnis](#).

b) Produkt 2: Anonymisierung der Mikrodaten

Folgende Maßnahmen wurden bei der Erstellung von Produkt 2 – „Gebäude- und Wohnungszählung“ zur On-Site-Nutzung durchgeführt, um die Anonymität der Daten zu gewährleisten:

Formale Anonymisierung des Datenmaterials (Nutzung per KDFV):

- Löschen der direkten Identifikatoren.
- Bilden von systemfreien Nummern
- Erstellung von Datenbeständen mit und ohne Anwendung des SAFE-Verfahrens. (Grund: Häufigkeitsauszählungen können nur auf Basis des SAFE-gesicherten Datenbestands erstellt werden, weiterführende Analysen nur auf Basis der unveränderten originalen Daten.)

Faktische Anonymisierung des Datenmaterials (Nutzung am GWAP):

- Ziehen einer Stichprobe auf Basis der Anschriften⁹

Die Stichprobenziehung am GWAP erfolgt zur Sicherstellung der faktischen Anonymität der Daten. Zusätzlich können dadurch die Programmlaufzeiten verkürzt werden. Finale Ergebnisse werden, wie in Kapitel III.4.2. beschrieben, auf Basis des KDFV-Datenmaterials erzeugt.

Nutzer können mit der Stichprobe am GWAP Ergebnisse sichten und auf ihre Verwertbarkeit prüfen. Die am GWAP erstellten Syntaxen werden ebenso wie die auf Basis des Strukturdatensatzes erstellten Syntaxen von den FDZ-Mitarbeitern per KDFV auf das Vollmaterial angewandt. Am GWAP anhand der Stichprobe erstellte Ergebnisse werden nicht an den Nutzer weitergegeben.

Eine genaue Übersicht über die für Produkt 2 – „Gebäude- und Wohnungszählung“ durchgeführten Aufbereitungsschritte und die Anonymisierungsmaßnahmen bietet der [Metadatenreport für Produkt 2](#).

c) Produkt 2: Geheimhaltung der Ergebnisse

In Anlehnung an das für die Zensusdatenbank (s. Kapitel III.1) genutzte Geheimhaltungsverfahren werden Häufigkeitstabellen über die FDZ nur auf Basis der bereits mit SAFE geheim gehaltenen Daten weitergegeben.

⁹ Die Stichprobe umfasst 10 % der in der Haushaltsstichprobe befragten Anschriften und 1 % der nicht in der Haushaltsstichprobe befragten Anschriften

Exkurs: SAFE

Das prä-tabulare SAFE-Verfahren wurde für die Umsetzung der Tabellengeheimhaltung in der Zensusdatenbank genutzt. Bei SAFE handelt es sich um ein Mikroaggregationsverfahren. Dabei werden Fälle durch Gruppenbildung und Austausch einzelner Angaben so vereinheitlicht, 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

Jörg Höhne: SAFE – Ein Verfahren zur Geheimhaltung und Anonymisierung statistischer Einzelangaben.
In: Berliner Statistik: Monatsschrift 3/03. https://www.statistik-berlin-brandenburg.de/Publikationen/Aufsaetze/2003/MS-BE_200303-01.pdf

Höhne, Jörg: Das Geheimhaltungsverfahren SAFE. In: Zeitschrift für amtliche Statistik Berlin Brandenburg 2/2015 S. 16-33. https://www.statistik-berlin-brandenburg.de/publikationen/aufsaetze/2015/HZ_201502-01.pdf

Eine Erstellung von Häufigkeitstabellen auf Basis nicht mit SAFE geheim gehaltener Daten (den sogenannten „Vor-SAFE-Daten“) ist über die FDZ für Produkt 2 – „Gebäude- und Wohnungszählung“ aus Gründen der Geheimhaltung nicht möglich. Um dem Nutzer einen Anhaltspunkt für die Abweichungen zwischen Vor- und Nach-SAFE-Daten zu geben, werden starke Abweichungen in den freigegebenen Tabellen durch Klammern (relativ starke Abweichung) oder Punkte (Wert nicht belastbar) durch die FDZ-Mitarbeiter gekennzeichnet.

Weiterführende Analysen, beispielsweise Regressionen, werden dagegen nur auf Basis der Vor-SAFE Daten durchgeführt. Hintergrund hierfür ist, dass SAFE als Geheimhaltungsverfahren vorrangig für die Optimierung von Häufigkeitstabellen entwickelt wurde. Die Geheimhaltung von Auswertungen, die über die Erstellung von Tabellen hinausgehenden, erfolgt daher bei Bedarf post-tabular entsprechend dem üblichen Verfahren der FDZ (vgl. Anhang 1 – Tipps für die Auswertung).

Eine genaue Übersicht über die Geheimhaltung für das Produkt 2 – „Gebäude- und Wohnungszählung“ bietet der [Metadatenreport für Produkt 2](#).

d) Produkt 2: Datensatzgröße

Das über die FDZ bereitgestellte Zensus-Produkt 2 – „Gebäude- und Wohnungszählung“ hat folgende Größe:

	KDFV	GWAP
Anzahl der Merkmale	44	44
Anzahl Wohnungsdatensätze	40,6 Mio.	750.000
Speicherplatz Datenmaterial		
SAS	7 GB	130 MB
SPSS	15 GB	280 MB
Stata	9 GB	160 MB

e) Produkt 2: Schlüsselverzeichnis

Das Schlüsselverzeichnis für das Zensus-Produkt 2 – „Gebäude- und Wohnungszählung“ findet sich [hier](#).

Weitere Informationen zu den Merkmalen des Produkts 2 – „Gebäude- und Wohnungszählung“ enthält der [Metadatenreport für Produkt 2](#).

f) Produkt 2: Besonderheiten bei der Auswertung

Da für die Erstellung von Häufigkeitstabellen ein Abgleich zwischen Vor- und Nach-SAFE-Ergebnissen stattfindet, müssen Syntaxen so geschrieben werden, dass sie ohne großen Änderungsaufwand sowohl auf den Vor- als auch auf den Nach-SAFE-Datenbestand angewendet werden können.

4.3.3 Weitere Produkte

Zu einem späteren Zeitpunkt können – je nach Bedarf der Nutzer und vorhandenen Ressourcen – weitere Produkte erstellt werden.

4.4 Nutzungsbedingungen

4.4.1 Nutzungsberechtigte Einrichtungen

Die Daten der FDZ können gemäß §16 Abs. 6 Bundesstatistikgesetz (BStatG) von „Hochschulen oder sonstige[n] Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung“ genutzt

werden. Hierbei findet aus rechtlichen Gründen eine Unterscheidung dahingehend statt, ob der Sitz der wissenschaftlichen Institution im In- oder Ausland liegt. Die Nationalität des Forschenden spielt dabei keine Rolle.

Die absolut anonymen Public-Use-Files und CAMPUS-Files können zusätzlich auch von der interessierten Allgemeinheit und von Mitarbeitern von kommerziellen Unternehmen und Institutionen, die nicht wissenschaftliche Institutionen im Sinne des §16 Abs. 6 BStatG sind, genutzt werden.

4.4.2 Beantragung der Datennutzung

Der Zugang zu den Mikrodaten der Forschungsdatenzentren muss beantragt werden.¹⁰ Die Vorlage für den Nutzungsantrag kann [hier](#) online ausgefüllt oder heruntergeladen werden.

Im Nutzungsantrag müssen Angaben zur antragstellenden Institution und zu allen Personen gemacht werden, die mit den Daten arbeiten sollen. Darüber hinaus muss das Projekt beschrieben werden, für das die Daten beantragt werden, sowie die geplanten Methoden. Zentral ist natürlich auch die Angabe, welche Daten genutzt werden sollen und welcher Zugangsweg gewünscht wird. Gegebenenfalls muss auch eine Variablenauswahl erfolgen.

Nach der Unterzeichnung des Nutzungsantrags durch eine für die beantragende Institution zeichnungsberechtigte Person wird der Antrag online, per Post oder Telefax an einen der Standorte der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder versandt.

4.4.3 Nutzungsdauer

Die Nutzung der Daten der FDZ der Statistischen Ämter ist auf eine Laufzeit von drei Jahren beschränkt. Es besteht die Möglichkeit der (kostenpflichtigen) Verlängerung für weitere drei Jahre.

4.4.4 Entstehende Kosten

Für die Nutzung der Daten über die FDZ werden Entgelte erhoben. Die Höhe ist abhängig von der Anzahl der beantragten Statistiken, der beantragten Jahre und der beantragten Zugangswege sowie davon, ob Daten des Standardangebotes oder projektspezifisch aufbereitete Daten nachgefragt werden.

Für die beiden standardisierten Produkte der ersten Bereitstellungsstufe, also Produkt 1 – „Haushaltsstichprobe“ und Produkt 2 – „Gebäude- und Wohnungszählung“ (Produkt 2), werden jeweils 250 EUR für die gemeinsame Nutzung von GWAP und KDFV erhoben. Die Nutzung der beiden Zugangswege wird – anders als üblich – nicht einzeln berechnet, da am GWAP keine

¹⁰ Ausnahme ist der CAMPUS-File, der für die Daten des Zensus 2011 aber nicht angeboten wird.

Ergebnisse erzeugt und für die Veröffentlichung genutzt werden dürfen und die KDFV grundsätzlich nicht ohne vorherige Nutzung eines GWAP erfolgen darf (s. Kapitel III.4.2).

Für den wissenschaftlichen Nachwuchs gibt es die Möglichkeit, Rabatte zu beantragen. Er kann das Angebot der FDZ für die Verwendung in Abschlussarbeiten zu einem reduzierten Preis nutzen. Studenten zahlen 25 EUR je Produkt. Dieses kann regulär bis zu einem Jahr lang ausgewertet werden. Doktoranden zahlen je Produkt 50 EUR und können dieses für bis zu drei Jahre nutzen.

IV. Literaturverzeichnis

Fischer, Christoph: Zensus 2011 – Die Zensusdatenbank nach Abschluss der Datenaufbereitung. In: Statistisches Monatsheft Baden-Württemberg 6/2014 http://www.statistik.baden-wuerttemberg.de/Veroeffentl/Monatshefte/PDF/Beitrag14_06_07.pdf, letzter Zugriff am 28.09.2015

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (Internetauftritt). <http://www.forschungsdatenzentrum.de/>, letzter Zugriff am 28.09.2015

Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009.

https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Zensusgesetz_2011.pdf, letzter Zugriff am 28.09.2015

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), das durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749) geändert worden ist

<https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Bundesstatistikgesetz.pdf>, letzter Zugriff am 28.09.2015

Gesetz zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 – ZensVorbG 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

<https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Zensusvorbereitungsgesetz.pdf>, letzter Zugriff am 28.09.2015

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

https://www.statistik-berlin-brandenburg.de/publikationen/aufsaeetze/2015/HZ_201502-01.pdf, letzter Zugriff am 28.09.2015

Höhne, Jörg: SAFE – Ein Verfahren zur Geheimhaltung und Anonymisierung statistischer Einzelangaben. In: Berliner Statistik: Monatsschrift 3/03. https://www.statistik-berlin-brandenburg.de/Publikationen/Aufsaeetze/2003/MS-BE_200303-01.pdf, letzter Zugriff am 28.09.2015

Statistische Ämter des Bundes und der Länder: Zensus 2011. Wissen, was morgen zählt (Internetauftritt). <https://www.zensus2011.de>, letzter Zugriff am 28.09.2015

Statistische Ämter des Bundes und der Länder: Methoden und Verfahren beim Zensus 2011, 2015. https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf, letzter Zugriff am 28.09.2015

Statistische Ämter des Bundes und der Länder: Nutzer-Kurzinformation zu SAFE. https://www.zensus2011.de/SharedDocs/Downloads/DE/Merkmale/Nutzerhinweise_safe.pdf, letzter Zugriff am 28.09.2015

Statistische Ämter des Bundes und der Länder: Zensusdatenbank Zensus 2011. <https://ergebnisse.zensus2011.de/>, letzter Zugriff am 28.09.2015

Verordnung (EG) Nr. Nr. 763/2008 des europäischen Parlaments und des Rates vom 9. Juli 2008 über Volks- und Wohnungszählungen
https://www.zensus2011.de/SharedDocs/Downloads/DE/Gesetze/Verordnung_des_Europaeischen_Parlaments_und_des_Rates_763_2008.pdf, letzter Zugriff am 28.09.2015

V. Anhang

Anhang 1: Tipps für die Auswertung

Anhang 2: Regeln und Hinweise zur Syntaxerstellung für Zensus-Daten

Anhang 3: Rechtliche Grundlage

Anhang 1: Tipps für die Auswertung

Vor Weitergabe der Ergebnisse müssen diese von den Mitarbeitern des FDZ auf geheim zu haltende Fälle geprüft werden. Um die Prüfung der Ergebnisse zu vereinfachen, sollten Sie bei der Erstellung der Ergebnisse zusätzlich folgende Punkte beachten.

Die folgenden Hinweise gelten vor allem für Häufigkeitsauswertungen, Kreuztabellen und andere deskriptive Auswertungen der Zensus 2011-Daten:

1. Bei Verwendung des Produkts 1 – „Haushaltsstichprobe“ wird für die Prüfung die Information benötigt, auf Basis welcher Fallzahlen die einzelnen Werte in einer Tabelle berechnet worden sind. Dies gilt beispielsweise für die Berechnung von Anteilswerten. Bei der Verwendung von Gewichtungsfaktoren müssen auch die ungewichteten Fallzahlen abgebildet werden.
2. Bei Verwendung des Produkts 2 – „Gebäude- und Wohnungszählung“ werden Häufigkeitstabellen auf Basis der SAFE-anonymisierten Daten veröffentlicht. Um Ihnen eine Einschätzung zu ermöglichen, wie groß die Abweichung zwischen den SAFE-Werten und den Originalwerten ist, werden große Abweichungen in den Ergebnistabellen gekennzeichnet. Bitte schreiben Sie Ihre Programmcodes daher zur Vermeidung langer Wartezeiten für die Ergebnisfreigabe so, dass zur Anwendung auf die Datenbestände vor und nach SAFE lediglich die Pfadangaben angepasst werden müssen. Schreiben Sie hierfür alle benötigten Pfadangaben an den Beginn Ihrer Programmcodes.
3. Geringe Fallzahlen (ungewichtet unter 3) sollten Sie durch geeignete Maßnahmen, z. B. Bildung von Klassen oder Zusammenfassungen von Merkmalsausprägungen, möglichst vermeiden. Informationen, welche Kategorien hiervon betroffen sind, erhalten Sie von den Mitarbeitern der FDZ.
4. Minima und Maxima werden im Allgemeinen nicht herausgegeben: Umgehen Sie dies z. B. durch Perzentile. Beachten Sie dabei, dass 1 % und 99 %-Perzentile mindestens eine Basis von $N = 300$ Fälle, 5 % und 95 %-Perzentile mindestens $N = 60$ Fälle (ungewichtet) erfordern.
5. Bedenken Sie die übergreifende Geheimhaltung. Vermeiden Sie z. B. bei der Verwendung von Filterbedingungen Restgrößen, die sich als Differenz zu „Insgesamt“ erschließen. Bitte erzeugen Sie in einem solchen Fall die Ergebnisse für die Restgröße explizit mit, da diese bei der Geheimhaltungsprüfung berücksichtigt werden muss.
6. Beachten Sie, dass bei der Geheimhaltungsprüfung auch der Output, den Sie bereits erhalten haben, berücksichtigt wird. Zum einen kann dies den Informationswert der von Ihnen erstellten Ergebnisse einschränken, da sich hieraus zusätzliche Sperrungen ergeben. Zum anderen steigt damit der Aufwand für die Prüfung und damit die Wartezeit bis zur Ergebnisfreigabe. Für

eine explorative Datenanalyse empfiehlt sich daher die Nutzung eines Gastwissenschaftlerarbeitsplatzes.

Anhang 2: Regeln und Hinweise zur Syntaxerstellung für Zensus-Daten

Der Aufbau eines Programmcodes für die Nutzung über einen On-Site-Zugangsweg der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder muss einigen formalen Vorgaben genügen. Damit wird den Mitarbeitern des FDZ die Einarbeitung in den Programminhalt vereinfacht und die zur Prüfung der Ergebnisse notwendige Zeitspanne verkürzt sich erheblich. Ein Programmcode muss es den Mitarbeitern des FDZ ermöglichen, Ziel und Inhalt der Analysen schnell zu erfassen und zu beurteilen. Dazu geben wir Ihnen hier die notwendigen Konventionen für die Syntaxerstellung an die Hand.

Regeln für den Programmcode für die Auswertung der Daten des Zensus 2011:

1. Geben Sie **Projekttitel**, das verwendete **Datenmaterial** und Ihre **Kontaktinformation** bitte im Programmkopf an.
2. Beschreiben Sie **Ziel** und **Zweck** des Programms in der Programmerläuterung. Gegebenenfalls sind auch Informationen über die **Einordnung** der aktuellen Auswertungen in den Stand und Fortschritt des Projektes, **Bezüge** zu vorhergehenden und zukünftigen Auswertungen bzw. eventuelle Änderungen der Auswertungsstrategie nötig. Die Erläuterung enthält auch eine **Liste der verwendeten Variablen** des Originaldatensatzes und der von dem Programmcode neu erzeugten Variablen einschließlich deren Labels.
3. Jedes Output-Element (Auswertungsschritte bzw. Ergebnistabellen) muss mit einem Titel versehen und fortlaufend **nummeriert** werden. Tabellen müssen selbsterklärend sein, d. h. aus den Tabellenvorspalten und -zeilen muss der Tabelleninhalt eindeutig hervorgehen. Tabellen müssen handhabbar sein, sie dürfen eine sinnvoll darstellbare Größe nicht überschreiten. Die Spaltenanzahl sollte kleiner sein als die Zeilenanzahl. Die Auswertungen sind mit einer Spalte für die Anzahl (n) zu versehen.
4. Bitte setzen Sie **Pfadreferenzen** auf verwendete Dateien an den Programmanfang.
5. Für die Nutzung des Zensus-Produkts 2 – „Gebäude- und Wohnungszählung“ ist zu beachten, dass **Häufigkeitstabellen** nur auf Grundlage der **SAFE**-anonymisierten Daten bereitgestellt werden. Zur **Kennzeichnung großer Abweichungen** zwischen Vor- und Nach-SAFE-Ergebnissen müssen Häufigkeitstabellen auf Basis beider Datenbestände erstellt werden. Gleiches gilt, wenn Häufigkeiten als „Nebenprodukte“ z. B. von Regressionen ausgegeben werden. Achten Sie bei der Erstellung von Programmcodes darauf, dass für die Anwendung eines Programmcodes auf die Vor- und Nach-SAFE-Daten nur die Pfadangaben geändert werden müssen und damit die Vor- und Nach-SAFE-Outputs exakt identisch sind. Nur so kann

ein automatisierter Abgleich durchgeführt und die Wartezeit für die Ergebnisfreigabe erheblich reduziert werden.

6. Der Abgleich der Vor- und Nach-SAFE-Ergebnisse erfolgt automatisiert in **Excel**. Entsprechende Ergebnisse sollten also direkt als Excel-Datei erzeugt werden.
7. **Strukturieren** Sie den Programmcode optisch klar. Schleifen sind einzurücken und Abstände zwischen den Programmblöcken einzufügen. Die Groß- und Kleinschreibung sollte einheitlich verwendet werden. Bitte nummerieren Sie die Programmabschnitte und teilen Sie lange Programme in Module auf.
8. Die Programme müssen von Ihnen **kommentiert** werden. Alle Programmblöcke und einzelne erklärungsbedürftige Befehle bzw. Auswertungen sind mit verständlichen und ausreichend detaillierten Kommentaren zu versehen, damit die Programmschritte leichter nachvollzogen werden können. Insbesondere sind Makros ausführlich zu dokumentieren. Variablen sind im Kommentar mit ihrem Label anzugeben.
9. Vergeben Sie **sprechende Namen** für neu gebildete Variablen und versehen Sie diese mit verständlichen Variablen- und Wertelabeln.
10. Bitte nutzen Sie **einheitliche Kennzeichnungen** für im Code benutzte Objekte (Beispiel: Relationen [entweder einheitlich „>“ oder „GT“], Missing Values [„.“, „0“ oder „-x“], etc.).
11. Die **Protokollfunktion** der Statistik-Software (das Anlegen von Logfiles) muss von Ihrem Programmcode eingeschaltet bzw. darf nicht ausgeschaltet werden (abhängig von der Statistik-Software).
12. Bei geringfügigen Änderungen der Syntax lassen Sie bitte nur die **betroffenen Programmelemente** noch einmal rechnen.

Entspricht ein Programmcode nicht diesen Vorgaben, sind die Mitarbeiter des FDZ angehalten, diesen zur Überarbeitung an Sie zurückzuschicken.

Bitte testen Sie Ihren Programmcode ausführlich an den am Gastwissenschaftlerarbeitsplatz bereitgestellten Daten oder an dem zur Verfügung gestellten Strukturdatensatz. Syntaktische Fehler werden von den Mitarbeitern des FDZ nicht bearbeitet und führen dazu, dass Sie Ihren Code zur Verbesserung zurückerhalten. Die inhaltliche Sinnhaftigkeit Ihres Programmcodes wird im FDZ nicht geprüft. Auf Wunsch können Sie sich jedoch von den Mitarbeitern beraten lassen.

Anhang 3: Rechtliche Grundlagen

Für die Durchführung des Zensus 2011 waren vor allem folgende Gesetze und Verordnungen relevant:

- **„Gesetz zur Erprobung eines registergestützten Zensus (Zensusertestgesetz – ZensTeG)“** vom 27. Juli 2001: Prüfung der Methode des registergestützten Verfahrens.
- **„Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011)“** vom 8. Juli 2009: Umsetzung der Verordnung Nr. 763/2008 für die in Deutschland genutzte Methode des registergestützten Zensus; Beschluss, welche Informationen aus den Registern entnommen und welche über die Stichprobe erhoben werden sollten; Klärung von Auskunftspflicht und Datenschutz.
- **„Gesetz zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 – ZensVorbG 2011)“**, das durch Artikel 3 des Gesetzes vom 8. Juli 2009 (BGBl. I S. 1781) geändert worden ist: Festlegung des Aufbaus des Anschriften- und Gebäuderegisters (AGR).¹¹
- **„Verordnung über Verfahren und Umfang der Haushaltebefragung auf Stichprobenbasis zum Zensusgesetz 2011 (Stichprobenverordnung Zensusgesetz 2011 – StichprobenV)“** vom 25. Juni 2010: Festlegung der Rahmenbedingungen für die Stichprobenziehung
- **Zensusausführungsgesetze** auf Länderebene: Klärung der Umsetzung der Vorgaben des Zensusgesetzes für jedes Bundesland.
 - **Baden-Württemberg:** „Gesetz zur Ausführung des Zensusgesetzes 2011 (AGZensG 2011)“ vom 29. Juli 2010
 - **Bayern:** „Bayerisches Statistikgesetz, Abschnitt V (Sonderregelungen für die Durchführung des Zensus 2011)“ vom 23. Juli 2010
 - **Berlin:** „Gesetz zur Ausführung des Zensusgesetzes 2011 im Land Berlin (Zensusausführungsgesetz Berlin – ZensusAGBln)“ vom 1. Dezember 2010
 - **Brandenburg:** „Gesetz zur Ausführung des Zensusgesetzes 2011 im Land Brandenburg (Brandenburgisches Zensusausführungsgesetz – ZensusAGBbg)“ vom 22. September 2010
 - **Bremen:** „Gesetz zur Ausführung des Zensusgesetzes 2011 (Zensusausführungsgesetz – ZensAG)“ vom 5. Oktober 2010

¹¹ Das AGR steuerte unter anderem den Ablauf der Gebäude- und Wohnungszählung und war Auswahlgrundlage für die Ziehung der Haushaltsstichprobe.

- **Hamburg:** „Hamburgisches Gesetz zur Ausführung des Zensusgesetzes 2011“ vom 6. April 2010
 - **Hessen:** „Hessisches Ausführungsgesetz zum Zensusgesetz 2011“ vom 23. Juni 2010
 - **Mecklenburg-Vorpommern:** „Gesetz zur Ausführung des Zensusgesetzes 2011 in Mecklenburg-Vorpommern (Zensusausführungsgesetz – ZensAG M-V)“ vom 17. Mai 2010
 - **Niedersachsen:** „Niedersächsisches Ausführungsgesetz zum Zensusgesetz 2011 (Nds. AG ZensG 2011)“ vom 6. Oktober 2010
 - **Nordrhein-Westfalen:** „Ausführungsgesetz des Landes Nordrhein-Westfalen zum Zensusgesetz 2011 (Zensusgesetz 2011-Ausführungsgesetz NRW – ZensG 2011 AG NRW)“ vom 16. November 2010
 - **Rheinland-Pfalz:** „Landesgesetz zur Ausführung des Zensusgesetzes 2011“ vom 28. September 2010
 - **Saarland:** „Gesetz Nr. 1713 zur Ausführung des Zensusgesetzes 2011“ vom 16. Juni 2010
 - **Sachsen:** „Gesetz zur Ausführung des Zensusgesetzes 2011 im Freistaat Sachsen (Sächsisches Zensusausführungsgesetz – SächsZensGAG)“ vom 23. September 2010
 - **Sachsen-Anhalt:** „Ausführungsgesetz des Landes Sachsen-Anhalt zum Zensusgesetz 2011 (Zensusausführungsgesetz Sachsen-Anhalt – ZensAG LSA)“ vom 8. Juli 2010
 - **Schleswig-Holstein:** „Gesetz zur Ausführung des Zensusgesetzes 2011 (Zensusausführungsgesetz – ZensGAG)“ vom 12. Oktober 2010
 - **Thüringen:** „Thüringer Gesetz zur Ausführung des Zensusgesetzes 2011 (ThürAGZensG 2011)“ vom 26. Juni 2010
- **„Verordnung (EG) Nr. 763/2008 des europäischen Parlaments und des Rates über Volks- und Wohnungszählungen“** vom 9. Juli 2008: Begründung der Durchführung des Zensus und Festlegung des Merkmalskranzes für die Datenlieferung an die EU.
 - **„Verordnung (EG) Nr. 1201/2009 der Kommission zur Durchführung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die technischen Spezifikationen für die Themen sowie für deren Untergliederungen“** vom 30. November 2009: Festlegung der Merkmalspezifikationen und Definition der verwendeten Begriffe.

- **„Verordnung (EU) Nr. 519/2010 der Kommission zur Annahme des Programms der statistischen Daten und der Metadaten für Volks- und Wohnungszählungen gemäß der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates“** vom 16. Juni 2010: Festlegung des Programms zur Daten- und Metadatenlieferung an Eurostat.
- **„Verordnung (EU) Nr. 1151/2010 der Kommission zur Durchführung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die Modalitäten und die Struktur der Qualitätsberichte sowie das technische Format der Datenübermittlung“** vom 8. Dezember 2010: Festlegung der der Struktur der Qualitätsberichte und der Modalitäten für die Übermittlung an Eurostat.

Bisher sind in der Reihe folgende FDZ-Arbeitspapiere erschienen:

Arbeitspapier Nr. 47: FiRe – Ein Schritt zur Teilautomatisierung der Geheimhaltungsprüfung, J. Pohlisch, J. Höninger, R., Voshage, Dezember 2014

Arbeitspapier Nr. 46: Entwicklung des Mikrosimulationsmodells EITDsim, G. Struch, April 2013

Arbeitspapier Nr. 45: Ein Mikrosimulationsmodell zur Berechnung der einkommensteuerlichen Bemessungsgrundlage und der Steuerzahlung auf Basis des Taxpayer-Panels, T. P., Schmidt, April 2013

Arbeitspapier Nr. 44: Das SAS-Makro newvar. Entwicklung und Anwendung eines Hilfsinstruments zur effizienten Erstellung neuer Variablen in der DRG-Statistik, T. Hochgürtel, T. Lösch, März 2012

Arbeitspapier Nr. 43: Average wage, qualification of the workforce and export performance in German enterprises: Evidence from KombiFiD data, J. Wagner, Januar 2012

Arbeitspapier Nr. 42: The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study, J. Wagner, Dezember 2011

Arbeitspapier Nr. 41: How to define an enterprise and assign trade declarations to the right one: Exploration of German traders' micro transaction data, C. Stirböck, August 2011

Arbeitspapier Nr. 40: Definition von nutzerseitigen Kriterien für Datenstrukturfiles, J. Höninger/M. Rosemann/R. Voshage, Juli 2011

Arbeitspapier Nr. 39: Improvement of data access – The long way to remote data access in Germany, M. Brandt/M. Zwick, Juni 2011

Arbeitspapier Nr. 38: Decentralised Access to Confidential Microdata in Europe, M. Brandt/P. Eilsberger/M. Zwick, Juni 2011

Arbeitspapier Nr. 37: Masking Micro Data with Stochastic Noise, J. Höhne/J. Höninger, Mai 2011

Arbeitspapier Nr. 36: Enthüllungsrisiko beim Remote Access: Die Schwerpunkteigenschaft der Regressionsgerade, A. Vogel, April 2011

Arbeitspapier Nr. 35: Temporary agency work and firm performance, S. Nielen/A. Schiersch, April 2011

Arbeitspapier Nr. 34: Harmonisation of statistical confidentiality in the Federal Republic of Germany, M. Brandt/A. Crößmann/C. Gürke, März 2011

Arbeitspapier Nr. 33: Remote Access. Eine Welt ohne Mikrodaten ??, G. Ronning/P. Bleninger/ J. Drechsler/C. Gürke, Februar 2011

Arbeitspapier Nr. 32: Compiling a Harmonized Database from Germany's 1978 to 2003 Sample Surveys of Income and Expenditure. T. Bönke/C. Schröder/C. Werdt, Mai 2010

Arbeitspapier Nr. 31: The Research Potential of New Types of Enterprise Data based on Surveys from Official Statistics in Germany., J. Wagner, Oktober 2009

Arbeitspapier Nr. 30: Geschlechterspezifische Einkommensunterschiede bei Selbstständigen im Vergleich zu abhängig Beschäftigten – Ein empirischer Vergleich auf der Grundlage steuerstatistischer Mikrodaten, P. Eilsberger/M. Zwick, Januar 2008

- Arbeitspapier Nr. 29:* Reichtum in Niedersachsen und anderen Bundesländern –Ergebnisse der Steuergeschäftsstatistik 2003 für Selbstständige (Freie Berufe und Unternehmer) und abhängig Beschäftigte, P. Böhm/J. Merz, November 2008
- Arbeitspapier Nr. 28:* Exports and Productivity in the German Business Services Sector. First Evidence from the Turnover Tax Statistics Panel, A. Vogel, Juli 2009
- Arbeitspapier Nr. 27:* Künstler in den Daten der amtlichen Statistik, C. Haak, August 2008
- Arbeitspapier Nr. 26:* Union Density and Varieties of Coverage: The Anatomy of Union Wage Effects in Germany, B. Fitzenberger/K. Kohn/A. C. Lembcke, August 2008
- Arbeitspapier Nr. 25:* German engineering firms during the 1990's. How efficient are export champions?, A. Schiersch, Juli 2008
- Arbeitspapier Nr. 24:* Zum Einkommensreichtum Älterer in Deutschland – Neue Reichtumskennzahlen und Ergebnisse aus der Lohn- und Einkommensteuerstatistik (FAST 2001), P. Böhm/J. Merz, Februar 2008
- Arbeitspapier Nr. 23:* Neue Datenangebote in den Forschungsdatenzentren. Betriebs- und Unternehmensdaten im Längsschnitt, M. Brandt/D. Oberschachtsiek/R. Pohl, November 2007
- Arbeitspapier Nr. 22:* Stichprobendaten von Versicherten der gesetzlichen Krankenversicherung – Grundlage und Struktur des Datenmaterials, P. Lugert, Dezember 2007
- Arbeitspapier Nr. 21:* KombiFid – Kombinierte Firmendaten für Deutschland, S. Bender/J. Wagner/ M. Zwick, November 2007
- Arbeitspapier Nr. 20:* Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, U. Kaiser/ J. Wagner, Juni 2007
- Arbeitspapier Nr. 18:* Die Gehalts- und Lohnstrukturerhebung: Methodik, Datenzugang und Forschungspotential, H.-P. Hafner/ R. Lenz, Mai 2007
- Arbeitspapier Nr. 17:* Anonymisation of Linked Employer Employee Datasets. Theoretical Thoughts and an Application to the German Structure of Earnings Survey, H.-P. Hafner/R. Lenz, Dezember 2006
- Arbeitspapier Nr. 16:* Die europäische Union – Integration von unten oder Eliteprojekt? Eine Sekundäranalyse von Mikrodaten der amtlichen Statistik, R. Nauenburg, November 2006
- Arbeitspapier Nr. 15:* Keeping in Touch – A Benefit of Public Holidays Using German Time Use diary Data, J. Merz/L. Osberg, November 2006
- Arbeitspapier Nr. 14:* Zur Konzeption eines Taxpayer-Panels für Deutschland, D. Vorgrimler/C. Gräß/S. Kriete-Dodds, November 2006
- Arbeitspapier Nr. 13:* Anonymisierte Daten der amtlichen Steuerstatistik, D. Vorgrimler, September 2006
- Arbeitspapier Nr. 12:* Mikrosimulation in der Betriebswirtschaftlichen Steuerlehre, R. Maiterth, August 2006
- Arbeitspapier Nr. 11:* Der Anteil der freien Berufe und der Gewerbetreibenden an der Gemeindefinanzierung, M. Zwick, September 2006

- Arbeitspapier Nr. 10:* Konstruktion und Bewertung eines ökonomischen Einkommens aus der Faktisch Anonymisierten Lohn- und Einkommensteuerstatistik, T. Bönke/F. Neher/C. Schröder, August 2006
- Arbeitspapier Nr. 9:* Anonymising business micro data – results of a German project, R. Lenz/M. Rosemann/D. Vorgrimler/R. Sturm, Juni 2006
- Arbeitspapier Nr. 8:* Scientific analyses using the Continuing Vocational Training Survey 2000, R. Lenz/H.-P. Hafner/D. Schmidt, Juni 2006
- Arbeitspapier Nr. 7:* A standard for the release of microdata, R. Lenz/ D. Vorgrimler/M. Scheffler, Juni 2006
- Arbeitspapier Nr. 6:* Measuring the disclosure protection of micro aggregated business microdata, R. Lenz, Juni 2006
- Arbeitspapier Nr. 5:* De facto anonymised microdata file on income tax statistics 1998, J. Merz/ D. Vorgrimler/M. Zwick, Oktober 2005
- Arbeitspapier Nr. 4:* Matching German turnover tax statistics, R. Lenz/D. Vorgrimler, Juni 2005
- Arbeitspapier Nr. 3:* The research data centres of the Federal Statistical Office and the statistical offices of the Länder, S. Zühlke/M. Zwick/S. Scharnhorst/T. Wende, März 2005
- Arbeitspapier Nr. 2:* Eine kommunale Einkommen- und Körperschaftsteuer als Alternative zur deutschen Gewerbesteuer: Eine empirische Analyse für ausgewählte Gemeinden, R. Maiterth/M. Zwick, April 2005
- Arbeitspapier Nr. 1:* Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik, H. Müller, März 2005

