

Statistisches Bundesamt (Destatis) (Ed.)

Periodical Part

WISTA – Wirtschaft und Statistik, Ausgabe 6/2024

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Statistisches Bundesamt (Destatis) (Ed.) (2024) : WISTA – Wirtschaft und Statistik, Ausgabe 6/2024, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 76, Iss. 6

This Version is available at:

<https://hdl.handle.net/10419/309505>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WISTA

Wirtschaft und Statistik

Themenausgabe Zensus 2022

C. Bretschi | S. Seibel |
I. Vorndran | M. Pfahl

Ermittlung der Einwohnerzahl im Zensus 2022

K. Dreschmitt | M. Pfahl

Die Personenerhebung im Zensus 2022

L. Boragk | A. Gaedke |
C. Gemmeke | C. Meyn

Die Ermittlung des Berichtskreises am Beispiel der
Sonderbereiche im Zensus 2022

H. Bretnütz | S. Estatico | S. Walther |
F. Hüsgen-Brodhäger | K. Lange |
B. Kleber | B. Habertzettl

Aufbereitung und Plausibilisierung der primärstatistischen
Erhebungsteile im Zensus 2022

D. Aude | J. Bolz | L. Boragk

Die Aktualisierung des Ortsverzeichnisses und
Plausibilisierung des Geburtsstaates im Zensus 2022

O. Kroppach | S. Rose

Umsetzung und Potenziale des Konzepts „Einwanderungs-
geschichte“ im Zensus 2022

T. Enderle | B. Kleber

Geheimhaltung mit der Cell-Key-Methode im Zensus 2022

A. Gaedke | M. Pfahl | A. Strohhalm

Die Online-First-Strategie und die Online-Fragebogen
im Zensus 2022

J. Hentschke | C.-A. Tobies |
S. Weber | D. Claus

Kern-Qualitätskennzahlen und Zielwerte im Zensus 2022

6 | 2024

ABKÜRZUNGEN

D	Durchschnitt (bei nicht addierfähigen Größen)
Vj	Vierteljahr
Hj	Halbjahr
a. n. g.	anderweitig nicht genannt
o. a. S.	ohne ausgeprägten Schwerpunkt
Mill.	Million
Mrd.	Milliarde

ZEICHENERKLÄRUNG

–	nichts vorhanden
0	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	Zahlenwert unbekannt oder geheim zu halten
...	Angabe fällt später an
X	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	keine Angaben, da Zahlenwert nicht sicher genug
()	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist
	Abweichungen in den Summen ergeben sich durch Runden der Zahlen.
	Tiefer gehende Internet-Verlinkungen sind hinterlegt.

INHALT

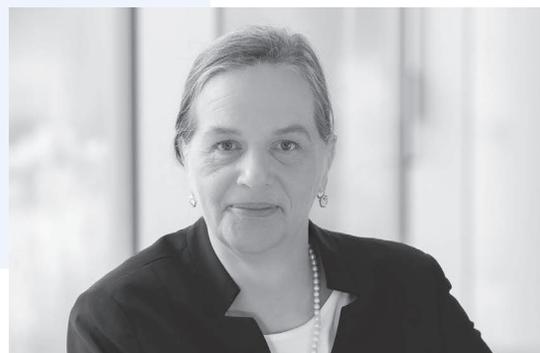
3	Editorial
4	Kennzahlen und Indikatoren
8	Aktuelle Informationsangebote
10	Kurznachrichten
17	Corinna Bretsch, Steffen Seibel, Ingeborg Vorndran, Miriam Pfahl Ermittlung der Einwohnerzahl im Zensus 2022 <i>Determining Germany's population in the 2022 Census</i>
29	Kai Dreschmitt, Miriam Pfahl Die Personenerhebung im Zensus 2022 <i>The survey of individuals in the 2022 Census</i>
41	Lisa Boragk, Annika Gaedke, Charlotte Gemmeke, Christoph Meyn Die Ermittlung des Berichtskreises am Beispiel der Sonderbereiche im Zensus 2022 <i>Determining the group of respondents based on the example of special facilities in the 2022 Census</i>
51	Hella Bretnütz, Sabrina Estatico, Sabrina Walther, Fabienne Hüsgen-Brodhacker, Kerstin Lange, Birgit Kleber, Benjamin Habertzettl Aufbereitung und Plausibilisierung der primärstatistischen Erhebungsteile im Zensus 2022 <i>Processing and editing of the primary statistical survey components of the 2022 Census</i>

INHALT

61	David Aude, Johannes Bolz, Lisa Boragk Die Aktualisierung des Ortsverzeichnisses und Plausibilisierung des Geburtsstaates im Zensus 2022 <i>Updating the directory of places and performing data editing for the country of birth in the 2022 Census</i>
73	Oliver Kroppach, Stefan Rose Umsetzung und Potenziale des Konzepts „Einwanderungsgeschichte“ im Zensus 2022 <i>Implementation and potential of the “immigration history” concept in the 2022 Census</i>
82	Tobias Enderle, Birgit Kleber Geheimhaltung mit der Cell-Key-Methode im Zensus 2022 <i>Ensuring confidentiality with the cell key method in the 2022 Census</i>
92	Annika Gaedke, Miriam Pfahl, Anna Strohaln Die Online-First-Strategie und die Online-Fragebogen im Zensus 2022 <i>The Online First strategy and online questionnaires in the 2022 Census</i>
106	Janine Hentschke, Cara-Aileen Tobies, Susanne Weber, Dennis Claus Kern-Qualitätskennzahlen und Zielwerte im Zensus 2022 <i>Core quality indicators and target values in the 2022 Census</i>

EDITORIAL

Dr. Ruth Brand



LIEBE LESERIN, LIEBER LESER,

die Beschäftigten des Statistischen Bundesamtes trauern um ihren ehemaligen Präsidenten Hans Günther Merk, der am 27. November 2024 im Alter von 94 Jahren verstarb. Sein Wirken ist eng verbunden mit dem Aufbau der kohärenten amtlichen Statistik im wiedervereinigten Deutschland.

Viele Entscheidungen in Bund, Ländern und Gemeinden beruhen auf Bevölkerungs- und Wohnungszahlen. Um verlässliche Basiszahlen für Planungen zu haben, ist eine regelmäßige Bestandsaufnahme dieser Daten notwendig. Dazu wurde zum Stichtag 15. Mai 2022 ein Zensus durchgeführt, mit dem Deutschland auch seine Teilnahmeverpflichtung an der EU-weiten Zensusrunde 2021 erfüllte. Im Juni 2024 hat die Veröffentlichungsphase der Ergebnisse des Zensus 2022 begonnen. Die [Webseite Zensus 2022](#) bietet Informationen und Tabellen zu den Themenfeldern Bevölkerung, Demografie, Gebäude und Wohnungen, Haushalte und Familien sowie Bildung und Erwerbstätigkeit. Die [Zensusdatenbank](#) und der [Zensus-Atlas](#) enthalten umfassende detaillierte Ergebnisse.

Die vorliegende Themenausgabe zum Zensus 2022 des Wissenschaftsmagazins WISTA dokumentiert Konzepte und Methoden, die bei den Erhebungen zum Einsatz kamen. Das Themenspektrum reicht dabei von den einzelnen Schritten der Einwohnerzahl-ermittlung und der erstmals verfolgten Online-First-Strategie über die Aufbereitung und Plausibilisierung der primärstatistischen Erhebungsteile, bis zur Qualitätsmessung mithilfe von Qualitätsindikatoren und der Geheimhaltung mittels Cell-Key-Methode.

In den kommenden Ausgaben werden Beiträge zur Hochrechnung der Zensusergebnisse, zur Mehrfachfallprüfung sowie zu gitterbasierten Auswertungen folgen.

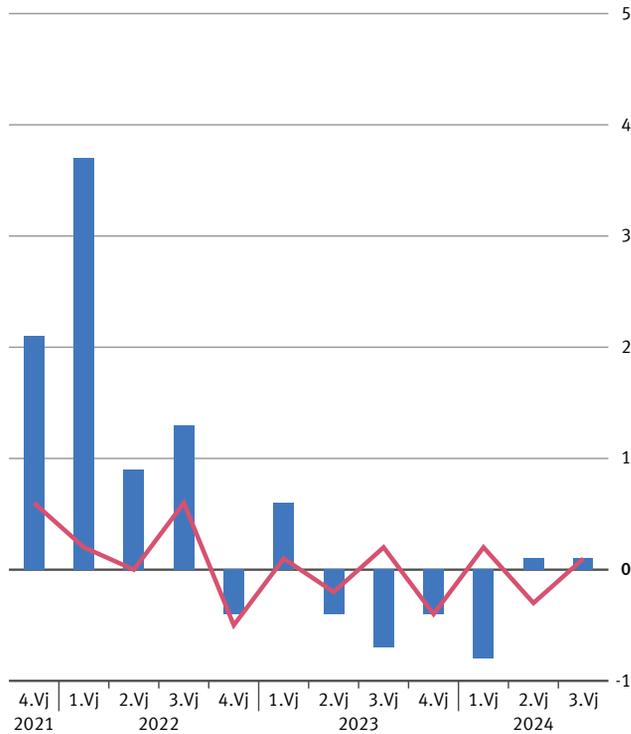
Ein ereignisreiches Jahr neigt sich dem Ende zu; ich wünsche Ihnen erholsame Festtage und alles Gute für das Jahr 2025.

Ruth Brand

Präsidentin des Statistischen Bundesamtes

Kennzahlen und Indikatoren

Bruttoinlandsprodukt
in %



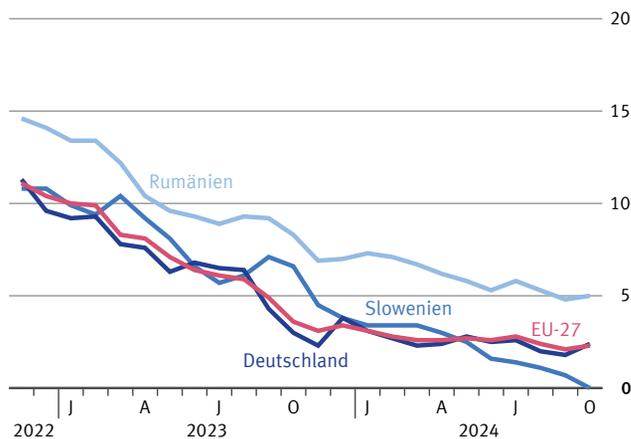
■ Veränderung gegenüber dem Vorjahresquartal (preisbereinigt)
 — Veränderung gegenüber dem Vorquartal (preis-, saison- und kalenderbereinigt)

Verbraucherpreisindex
2020 = 100

2023		2024	
Januar	114,3	Januar	117,6
Februar	115,2	Februar	118,1
März	116,1	März	118,6
April	116,6	April	119,2
Mai	116,5	Mai	119,3
Juni	116,8	Juni	119,4
Juli	117,1	Juli	119,8
August	117,5	August	119,7
September	117,8	September	119,7
Oktober	117,8	Oktober	120,2
November	117,3	November	119,9
Dezember	117,4		

119,9 ↑ **2,2 %**
 Veränderung zum Vorjahresmonat

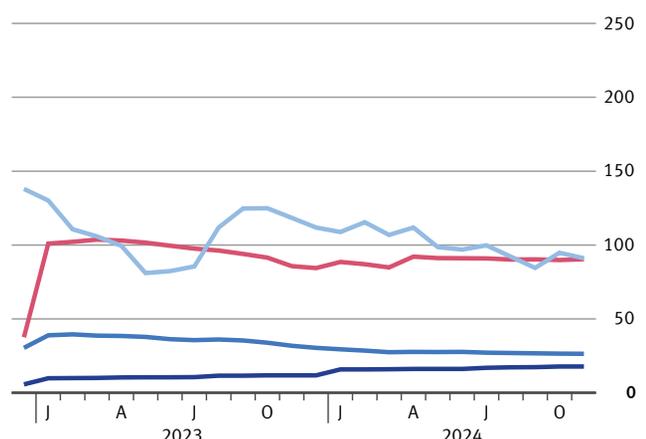
Harmonisierter Verbraucherpreisindex insgesamt
Veränderung gegenüber dem Vorjahresmonat in %



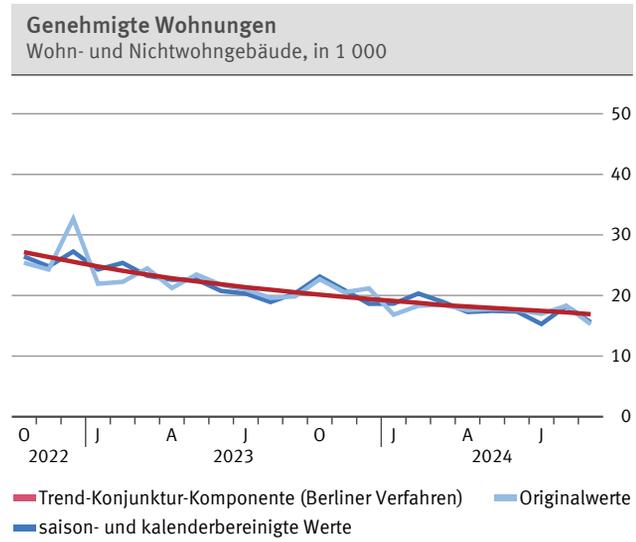
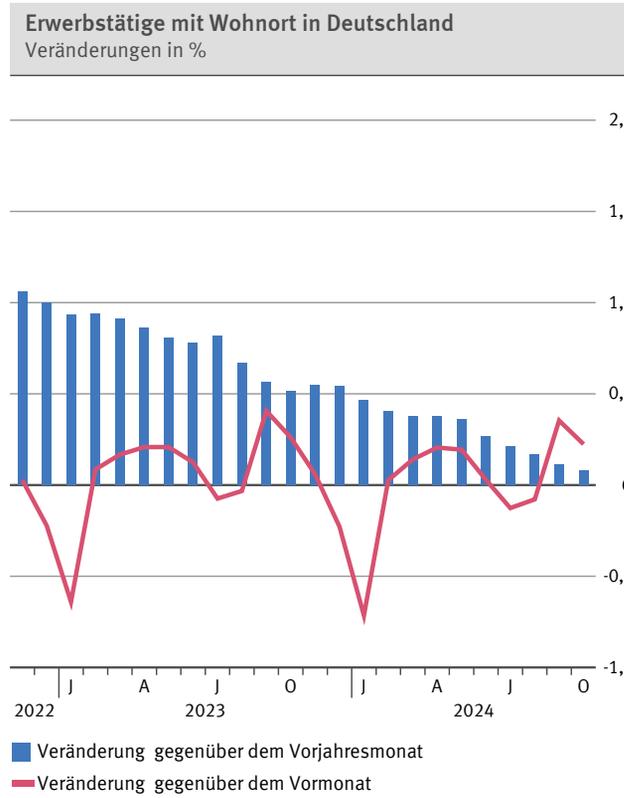
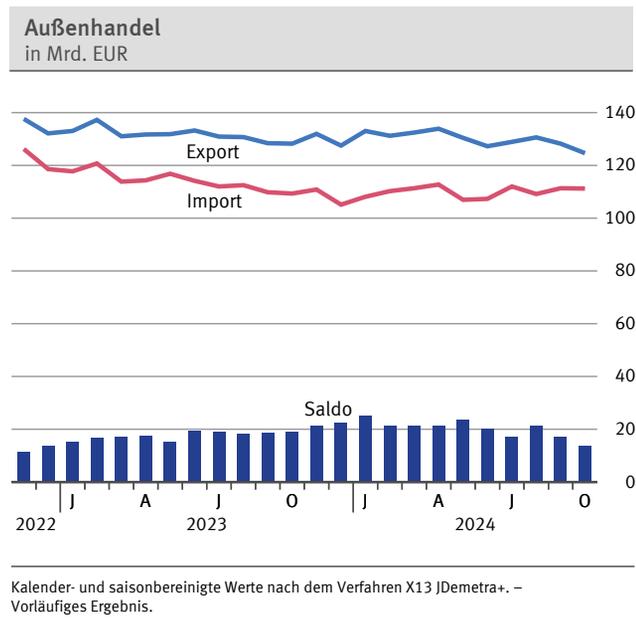
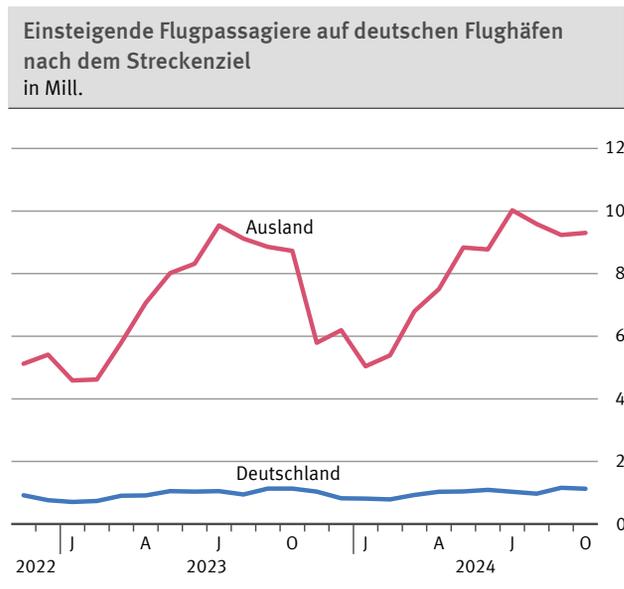
Dargestellt sind neben Deutschland und der Europäischen Union insgesamt (EU-27) die Länder mit der höchsten und der niedrigsten Veränderungsrate innerhalb der EU.

Stand: 13.12.2024

Entwicklung der Verbraucherpreise für Energie und Wasser
Preisabstand in % gegenüber dem Jahr 2020



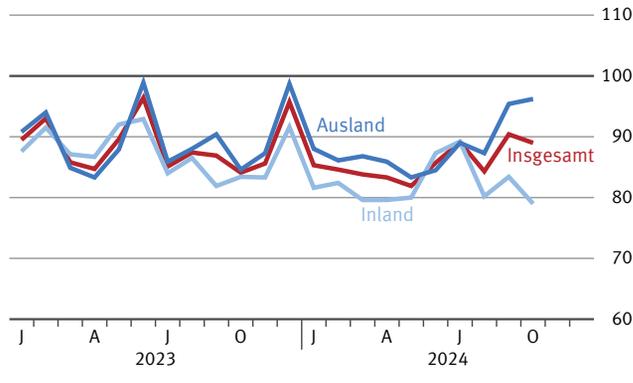
■ Frischwasser ■ Strom ■ Erdgas, ohne Betriebskosten ■ leichtes Heizöl



Stand: 13.12.2024

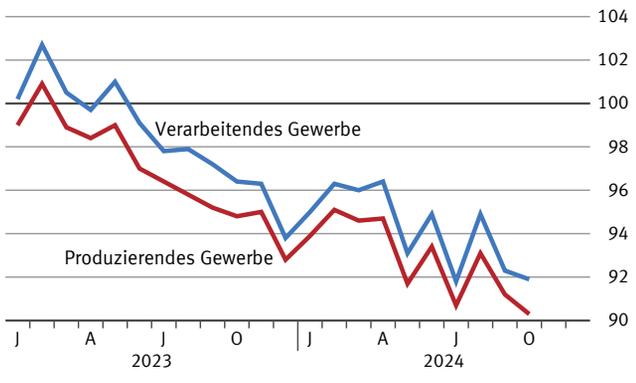
Kennzahlen und Indikatoren

Auftragseingang im Verarbeitenden Gewerbe
Volumenindex 2021 = 100



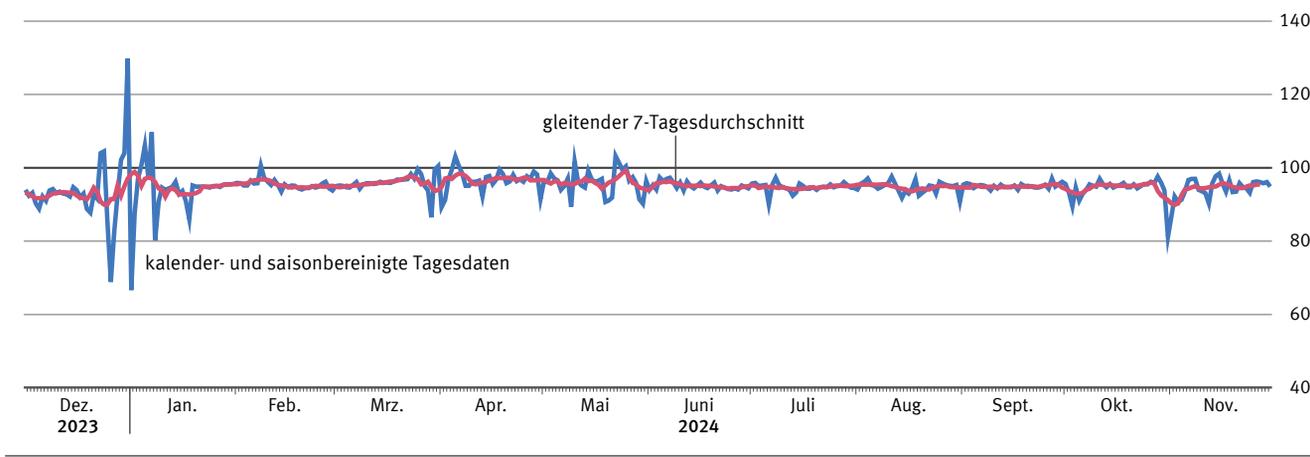
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+ - Vorläufiges Ergebnis.

Produktion im Produzierenden und Verarbeitenden Gewerbe
Index 2021 = 100



Kalender- und saisonbereinigte Werte nach dem Verfahren X13 JDemetra+ - Vorläufiges Ergebnis.

Lkw-Maut-Fahrleistungsindex
2021 = 100

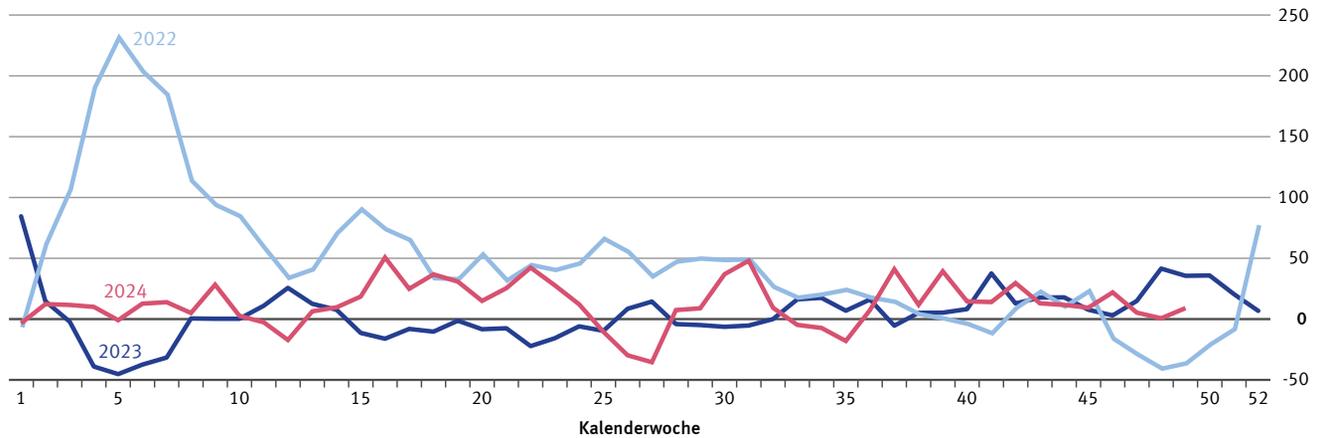


Quellen: Bundesamt für Logistik und Mobilität, Deutsche Bundesbank, Statistisches Bundesamt

Stand: 13.12.2024

Kennzahlen und Indikatoren

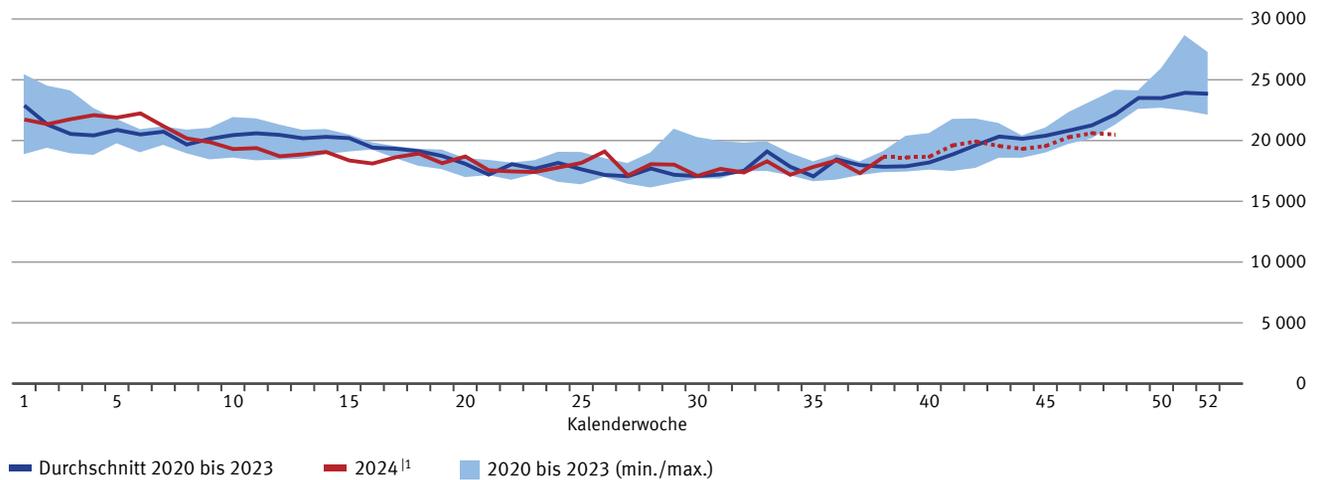
Neue Kreditverträge nach Kalenderwochen Veränderung gegenüber der entsprechenden Vorjahreswoche in %



Anfang 2022 zeigte sich mit Veränderungsraten von 100 % und mehr ein starker Anstieg im Vergleich zum Vorjahr; dabei handelt es sich um Sondereffekte, die seitens des Datenlieferanten nicht bereinigt werden konnten.

Quelle: SCHUFA Holding AG; Berechnung: Statistisches Bundesamt

Wöchentliche Sterbefallzahlen in Deutschland



Gestrichelte Werte enthalten Schätzanteil.

1 Sonderauswertung der vorläufigen Sterbefallzahlen.

Stand: 13.12.2024



Ukraine

Der Angriff Russlands auf die Ukraine und die damit verbundenen Sanktionen haben starke Auswirkungen auf Wirtschaft und Bevölkerung sowie den Energie-sektor. Auf einer Sonderseite zum Thema stellt das Statistische Bundesamt relevante Daten zur Verfügung. Über die Seite gelangt man auch zum [zentralen Hilfsportal](#) der Bundesregierung für Geflüchtete aus der Ukraine.

↳ www.destatis.de/Im-Fokus/Ukraine



Dashboard Deutschland

Das vom Statistischen Bundesamt entwickelte Datenportal bietet hochaktuelle und hochfrequente Zahlen, Daten und Fakten zu den Themen Arbeitsmarkt, Bauen und Wohnen, Energie, Finanzen, Konjunktur und Wirtschaft sowie Ukraine. Es trägt damit zu einem faktenbasierten demokratischen Diskurs der Öffentlichkeit und zur evidenzbasierten Entscheidungsfindung durch Politik und Verwaltung bei. Der integrierte Pulsmesser Wirtschaft bietet Einblicke in das aktuelle wirtschaftliche Geschehen, intuitives und einfaches Vergleichen von Daten sowie das Erkennen von konjunkturellen Entwicklungen und Zusammenhängen mithilfe täglicher, wöchentlicher, monatlicher und vierteljährlicher Indikatoren.

↳ www.dashboard-deutschland.de



EXSTAT – Experimentelle Statistiken

In der Rubrik „EXSTAT – Experimentelle Statistiken“ veröffentlicht das Statistische Bundesamt regelmäßig neue, innovative Projektergebnisse. Sie entstehen auf der Grundlage neuer Datenquellen und Methoden. Im Reifegrad und in der Qualität unterscheiden sie sich von amtlichen Statistiken, insbesondere in Bezug auf Harmonisierung, Erfassungsbereich und Methodik. Dennoch sind es Ergebnisse der Statistischen Ämter des Bundes und der Länder, die interessante, neue Perspektiven auf verschiedene Themenfelder der Statistik bieten.

↳ www.destatis.de/exstat



Ergebnisse des Zensus 2022

Wie viele Menschen leben in Deutschland? Wie wohnen und arbeiten sie? Diese Fragen beantworten die Ergebnisse des Zensus 2022. Im Wesentlichen geht es dabei um zwei Ziele: die Ermittlung aktueller Bevölkerungszahlen für Deutschland sowie Informationen zum Wohnraum mit der Gebäude- und Wohnungszählung.

↳ www.zensus2022.de



Klima

Der Klimawandel ist eine der größten Herausforderungen der heutigen Zeit, alle Bereiche der Gesellschaft sind betroffen. Wie beeinflusst unsere Lebens- und Wirtschaftsweise das Klima? Wie wirkt sich die Umstellung hin zu mehr Klimaschutz gesamtgesellschaftlich aus? Was bedeutet sie für unseren Alltag – vom Weg zur Arbeit bis zum aktuellen Strompreis? Wo zeigen sich die Folgen des Klimawandels? Daten und Fakten zum Thema Klima, Klimawandel und Klimaschutz sind gebündelt unter

↳ www.destatis.de/klima



Fachkräfte

Fachkräftemangel und Arbeitskräftebedarf sind zunehmend wichtige Faktoren für die wirtschaftliche Entwicklung in Deutschland. Daten und Fakten dazu bündelt das Statistische Bundesamt auf einer eigenen Sonderseite. Das Angebot umfasst die Bereiche Demografie, Erwerbstätigkeit, Bildung und Zuwanderung – und wird sukzessive erweitert.

↳ www.destatis.de/fachkraefte

KURZNACHRICHTEN

Trauer um Präsidenten a. D. Hans Günther Merk

Am 30. November 2024 ist Hans Günther Merk, Präsident des Statistischen Bundesamtes von 1992 bis 1995, im Alter von 94 Jahren verstorben.

Am 13. September 1930 im Kreis Bad Kreuznach geboren, war der Volljurist zunächst als Rechtsanwalt und danach als Richter beim Amtsgericht in Koblenz tätig. Ab 1959 wirkte er lange Jahre in verschiedenen Funktionen im Bundesministerium des Innern, unter anderem in den Bereichen Sport und Medien, Datenverarbeitung in der öffentlichen Verwaltung und Bevölkerungsfragen.

Im Juni 1992 wurde Hans Günther Merk zum Präsidenten des Statistischen Bundesamtes ernannt. In seiner Amtszeit wurde das System der Bundesstatistik vollständig in den neuen Ländern eingeführt. Damit war die amtliche Statistik in der Lage, erstmals nach der deutschen Vereinigung ein umfangreiches kohärentes Bild der wirtschaftlichen und sozialen Verhältnisse in ganz Deutschland zu zeichnen, auf Basis von nach einheitlichen Methoden erhobenen Daten.

Unter Hans Günther Merk erweiterte die amtliche Statistik in Deutschland ihr Informationsangebot um die Gesundheitsberichterstattung und die Umweltökonomischen Gesamtrechnungen. Nicht zuletzt setzte Hans Günther Merk auf die verstärkte Nutzung der Informationstechnologie bei der Statistikproduktion und -verbreitung.

Auch die internationale Zusammenarbeit lag Hans Günther Merk sehr am Herzen: Zum einen trieb er den Aufbau des Europäischen Statistischen Systems voran, zum anderen bemühte er sich noch während seiner im Jahr 1995 endenden Amtszeit darum, den 54. Weltkon-

gress des Internationalen Statistischen Instituts (ISI) in Deutschland zu veranstalten. Nach erfolgreicher Bewerbung übernahm er – schon als Pensionär – den Vorsitz des Nationalen Organisationskomitees des Treffens.

Hans Günther Merk blieb dem Statistischen Bundesamt stets verbunden. Die Beschäftigten des Statistischen Bundesamtes bedauern seinen Verlust. Unser besonderes Mitgefühl gilt seinen Angehörigen.

IN EIGENER SACHE

Verleihung des Wissenschafts- preises 2024

 **Statistical
Science
for the Society**
Wissenschaftlicher Nachwuchspreis

Das Statistische Bundesamt hat vier herausragende wissenschaftliche Arbeiten mit dem Wissenschaftlichen Nachwuchspreis „Statistical Science for the Society“ ausgezeichnet. Die Urkunden wurden im Rahmen des 33. Wissenschaftlichen Kolloquiums mit dem Thema „Energiewende und Energiepreiskrise – Zur Rolle der Daten“ verliehen, welches das Statistische Bundesamt zusammen mit der Deutschen Statistischen Gesellschaft am 28. und 29. November 2024 in Wiesbaden veranstaltete.

Der Vorsitzende des Gutachtendengremiums, Prof. Dr. Ralf Münnich, hielt die Laudationes; sie werden in der nächsten Ausgabe dieser Zeitschrift zu lesen sein.

Für das Statistische Bundesamt beglückwünschte Dr. Ruth Brand die Preisträgerinnen und Preisträger

und prämierte auf Empfehlung des unabhängigen Gutachtendengremiums die Arbeiten.

Den Wissenschaftspreis in der Kategorie „Dissertationen“ nahm Dr. Elisa Rottner für ihre Arbeit mit dem Thema „Electricity prices, emissions and firm performance – the case of German manufacturing“ entgegen. Die Arbeit entstand bei Prof. Dr. Beat Hintermann an der Universität Basel.

Einen weiteren Wissenschaftspreis in der Kategorie „Dissertationen“ erhielt Dr. Simon Schmaus für seine Arbeit zum Thema „Methoden regionalisierter dynamischer Mikrosimulationen“. Die Arbeit wurde betreut von Prof. Dr. Ralf Münnich an der Universität Trier.

Für die ebenfalls bei Prof. Dr. Ralf Münnich an der Universität Trier erstellte Masterarbeit zum Thema „Sampling for COVID-19 Statistics. A Comparison of Sampling Designs for Prevalence Estimates“ wurde Inken Siems mit einem Wissenschaftspreis in der Kategorie „Master-/Bachelorarbeiten“ ausgezeichnet.

Ebenfalls in der Kategorie „Master-/Bachelorarbeiten“ empfing Severin Weinand einen Wissenschaftspreis für die Bachelorarbeit zum Thema „Vergleich von Record-Linkage Methoden anhand der Mikro-Simulation eines bundesweiten Schülerregisters“. Die Arbeit wurde betreut von Prof. Dr. Rainer Schnell an der Universität Duisburg-Essen.

Die Preisträgerinnen und Preisträger werden ihre ausgezeichneten Arbeiten im Jahr 2025 in dieser Zeitschrift vorstellen.

Datenbank GENESIS-Online mit neuer Oberfläche

Seit Kurzem steht die Datenbank GENESIS-Online des Statistischen Bundesamtes mit neuer Oberfläche zur Verfügung. Sie bietet schnellere Datenabrufe sowie intuitive Recherche- und Anpassungsmöglichkeiten von Tabellen. Gleichzeitig ergeben sich Änderungen in der Struktur des Flatfile-CSV-Formats sowie im Datenausgabeformat. Hinweise dazu sind auf einer Informationsseite zusammengestellt.

↳ www-genesis.destatis.de

AUS EUROPA

109. DGINS-Konferenz

Die Konferenz der Generaldirektorinnen und Generaldirektoren europäischer Statistikämter (DGINS) fand am 23. und 24. Oktober 2024 in Tallinn (Estland) statt und widmete sich dem Thema „Neue Grenzen der amtlichen Statistik“. Damit ist erstens gemeint, die Aufgaben der amtlichen Statistik über das reguläre Statistikgeschäft hinaus zu erweitern, und zweitens, das vorhandene Portfolio der Statistik zu verbessern. Zu den Erweiterungen zählt die Umsetzung des Data Governance Acts¹. Ein professionelles Stakeholder-Management ist dabei ein wichtiger Hebel sowohl bei Innovationen als auch bei der amtlichen Statistik.

Session 1 „Effective Stakeholder Engagement: Foundation for Relevancy“: Das niederländische Statistikamt (Statistics Netherlands – CBS) erläuterte beispielhaft seine Kooperation mit der Wissenschaft. Schwerpunkte sind hierbei neue Formen der Zusammenarbeit durch geteilte Führungs- und Professorenstellen und Stellen für Promovierende, die Rekrutierung von EMOS²-Studierenden als neues Personal und die Umsetzung von Innovationen (zum Beispiel die Möglichkeit des Fernrechnens für Forschende). Das Bundesamt für Statistik der Schweiz stellte Kooperationen mit der Privatwirtschaft als Best Practice vor, wobei besonders herausfordernd ist, Anreize zu schaffen, damit Unternehmen ihre Daten der Öffentlichkeit zur Verfügung stellen. Deutschland legte am Beispiel der Umsetzung des Data Governance Acts dar, welche Vorteile eine intensivere Zusammenarbeit im öffentlichen, datenhaltenden Sektor hat.

Session 2 „Data Services Complementary to the Production of Official Statistics“: Hier lag der Fokus darauf, durch die Übernahme neuer Aufgaben Chancen zu eröffnen, um Innovationen zu implementieren. In einer anschließenden Plenardiskussion wurden Trends in der Zusammenarbeit und in der Koordination von neuen

1 Die Verordnung (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. Mai 2022 über europäische Daten-Governance und zur Änderung der Verordnung (EU) 2018/1724 (Daten-Governance-Rechtsakt – DGA) regelt den leichteren Zugang und die Nutzung von Verwaltungsdaten für Unternehmen und Wissenschaft und soll Innovationen fördern.

2 European Master in Official Statistics.

Serviceleistungen innerhalb der Europäischen Union eingehend diskutiert.

Session 3 „Towards the New Frontiers: Challenges and solutions“: Die Teilnehmenden befassten sich in einem Workshop im World-Café-Format mit den rechtlichen, finanziellen und organisatorischen Herausforderungen und Lösungen, dem Stakeholder-Management sowie den technologischen und ethischen Aspekten des Themas „Neue Grenzen der amtlichen Statistik“.

Thema der nächsten DGINS-Konferenz im Oktober 2025 in Split (Kroatien) sind die Tourismusstatistiken im Europäischen Statistischen System (ESS). Neben dem nachhaltigen Tourismus sollen neue Nutzungsmöglichkeiten innovativer Datenquellen sowie die Nutzung privat gehaltener Daten diskutiert werden. In den Diskussionen sollen die Grundzüge einer künftigen robusten und nutzerrelevanten Tourismusstatistik entwickelt und der Weg zu einer überarbeiteten Gesetzgebung in diesem Statistikbereich bereitet werden.

57. Sitzung des AESS

Die 57. Sitzung des Ausschusses für das Europäische Statistische System (AESS) fand am 24. und 25. Oktober 2024 in Tallinn (Estland) statt.

Dabei hat der AESS folgende Durchführungsverordnungen der Kommission verabschiedet:

- › Entwurf einer Durchführungsverordnung der Kommission zur Festlegung der technischen Eigenschaften der Datensätze der Stichprobenerhebung im Bereich Einkommen und Lebensbedingungen in Bezug auf Überschuldung, Verbrauch und Vermögen gemäß der Verordnung (EU) 2019/1700 des Europäischen Parlaments und des Rates und zur Änderung der Durchführungsverordnung (EU) 2021/2052 zur Festlegung der technischen Eigenschaften im Bereich Arbeitsmarkt und Wohnen.

Gemäß der Delegierten Verordnung (EU) 2020/256 der Kommission wird der Bereich „Einkommen und Lebensbedingungen“ im Jahr 2026 zusätzlich zu den jährlichen Variablen um einen Satz von sechsjährlichen Variablen zu „Überschuldung, Verbrauch und Vermögen“ sowie um einen Satz von dreijährlichen Variablen zu „Arbeitsmarkt und Wohnen“ ergänzt.

- › Durchführungsverordnung der Kommission über die technischen Angaben des Datensatzes für 2026 zum Ad-hoc-Thema „Beschäftigung auf digitalen Plattformen im Bereich Arbeitskräfte“ gemäß der Verordnung (EU) 2019/1700 des Europäischen Parlaments und des Rates.

Dadurch wird die Arbeitskräfteerhebung im Jahr 2026 um Variablen zum Thema „Beschäftigung auf digitalen Plattformen“ ergänzt.

- › Entwurf einer Durchführungsverordnung der Kommission zu den für das Referenzjahr 2026 gemäß der Verordnung (EU) 2018/1091 des Europäischen Parlaments und des Rates über integrierte Statistiken zu landwirtschaftlichen Betrieben zu liefernden Daten hinsichtlich der Liste der Variablen und ihrer Beschreibung.

Für die Aufnahme neuer Merkmale wurden zwei vorhandene Merkmale gestrichen. Deutschland bewertet die dadurch zu erreichende Kostenreduktion als gering, da die Streichung eine Anpassung der Erhebungsmethodik erfordert, welche national die Vorteile nicht kompensieren wird.

Im ersten Halbjahr 2025 beginnt mit Polen, Dänemark und Zypern eine neue Trio-Präsidentschaft. Das polnische Statistikamt stellt Maßnahmen ins Zentrum der statistischen Aktivitäten, die neue Datenquellen in bestehenden Erhebungen stärker berücksichtigen und Kosten bei der Statistikproduktion reduzieren sollen. Darüber hinaus werden EU-Verordnungen im Bereich Fischerei und Aquakultur (EFAS) und Gewerbeimmobilien (CREI) vorbereitet.

Von September 2022 bis Mai 2024 wurde mit Unterstützung eines externen Auftragnehmers und der nationalen Statistikämter aus Belgien, Deutschland, Spanien, Frankreich, Ungarn, den Niederlanden, Portugal und Norwegen ein „Review“ zu den Gesundheitsstatistiken durchgeführt. Die Überprüfung bestätigte die Relevanz der vorhandenen Daten für nationale und EU-Zwecke und deren Nutzung durch Stakeholder. Die Datennutzen formulierten einen klaren Bedarf an besseren Daten beispielsweise zur psychischen Gesundheit und zu den Determinanten der Gesundheit der alternden Bevölkerung. Der AESS begrüßte die erstellten Empfehlungen zur Weiterentwicklung im Grundsatz, die meisten Delegierten sahen jedoch bei fehlender Prioritätensetzung

die Gefahr einer Überlastung der personellen und finanziellen Ressourcen. Neu zu erhebende Merkmale in der Gesundheitsstatistik müssten einen direkten Bezug zur europäischen Politik haben und das ESS solle sich stärker in die Umsetzung und Ausgestaltung des europäischen Datenraums Gesundheit einbringen.

Das Statistische Amt der Europäischen Union, Eurostat, stellte erste Überlegungen einer ESS Vision 2030 vor. Wesentlicher Punkt dieser Strategie ist die Umsetzung von Innovationen bei gleichzeitiger Prioritätensetzung. Auch soll bei der inhaltlichen Ausgestaltung der ESS Vision 2030 eine stärkere Bindung an das europäische Mehrjahresprogramm erfolgen. Die ESS-Mitgliedstaaten sollen ihre ersten Einschätzungen zur ESS Vision 2030 an Eurostat weiterleiten.

10. EMOS-Workshop

Der European Master in Official Statistics (EMOS) ist ein Zertifikat, das vom European Statistical System Committee an Masterstudiengänge verliehen wird, die einen Schwerpunkt der Ausbildung im Bereich der amtlichen Statistik setzen. Es handelt sich um eine gemeinsame Initiative von Eurostat, dem Europäischen Statistischen System und dem Europäischen System der Zentralbanken. Sie zielt darauf ab, Studierende in amtlicher Statistik und Datenwissenschaft auszubilden, um hochqualifizierte Statistikerinnen und Statistiker mit Expertise in amtlicher Statistik hervorzubringen.

Der Einladung zum Workshop ins Statistische Bundesamt am 14. und 15. Oktober 2024 folgten etwa 95 Teilnehmende aus der Europäischen Union. Darunter waren Vertretungen der Universitäten mit EMOS-Studiengängen, der nationalen sowie europäischen Statistikbehörden, des europäischen Zentralbankensystems sowie aktuelle und ehemalige EMOS-Studierende. Ziele des Workshops waren, den Erfahrungsaustausch zu fördern, bewährte Verfahren aufzuzeigen und Wege für die künftige Entwicklung von EMOS zu erkunden.

Das EMOS-Board tagte ebenfalls während der Veranstaltung. Der EMOS-Vorstand erarbeitet derzeit ein Konzept, um das Label und die Studienprogramme bekannter und attraktiver zu machen. Ein Update der Zugangskriterien und der Lerninhalte sowie mögliche Weiterentwicklungspotenziale stehen ebenfalls auf der Agenda des EMOS-Boards.

Weitere Informationen zum Programm sowie die Präsentationen sind auf der EMOS-Website verfügbar:

➤ cros.ec.europa.eu

AUS DEM INLAND

20 Jahre RatSWD

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) ist ein unabhängiges Gremium von empirisch arbeitenden Wissenschaftlerinnen und Wissenschaftlern sowie Vertreterinnen und Vertretern der wichtigsten öffentlichen Einrichtungen zur Datenerhebung in Deutschland. Mit seiner Expertise berät der RatSWD die Politik zum Ausbau der Dateninfrastruktur.

Bei seiner 71. Sitzung am 14. und 15. November 2024 standen bedeutende Themen wie die Zukunft des Forschungsdatengesetzes, die Akkreditierungen von Forschungsdatenzentren (FDZ) sowie aktuelle Updates aus den verschiedenen Arbeitsgruppen auf der Tagesordnung.

Der RatSWD blickt mittlerweile auf zwei Jahrzehnte erfolgreichen Wirkens zurück und setzt sich auch in Zukunft für die Stärkung der datenbasierten Forschung in Deutschland ein.

➤ www.konsortswd.de

VERANSTALTUNGEN

Symposium zum Bundesstatistikgesetz

Unter der Leitung des Statistischen Bundesamtes sowie der Statistischen Ämter der Länder Bayern und Thüringen fand am 28. Oktober 2024 in Berlin ein Symposium über das Bundesstatistikgesetz (BStatG) statt. Organisiert wurde die Veranstaltung vom Hauptstadtbüro des Statistischen Bundesamtes.

Zu Beginn betonten die Präsidenten der drei beteiligten Statistikämter die Notwendigkeit einer Gesetzesreform, um verfassungskonforme Verbesserungen

insbesondere in Bezug auf die Verknüpfung von Daten (§13 BStatG) und den Zugang der Wissenschaft zu Daten (§16 BStatG) zu erzielen. Der erste Key-Note-Vortrag ging darauf ein, welche Grenzen und welche Möglichkeiten sich aus den nationalen und europäischen Datenschutzbestimmungen für eine Reform der Statistikgesetzgebung ergeben. Der Gesetzgeber habe, so Dr. Thomas Kienle, einen erheblichen Regelungsspielraum, der durch das datenschutzrechtliche Statistikprivileg noch erweitert werde.

Auch Prof. Dr. Jürgen Kühling sah im zweiten Key-Note-Vortrag gute Rahmenbedingungen für eine Reform des Bundesstatistikgesetzes. So gäbe es rechtliche Möglichkeiten für eine stärkere Verknüpfung von Daten, einen besseren Datenzugang für die Wissenschaft und für eine Flexibilisierung der Statistikgesetzgebung (weniger strenge Vorgaben, mehr Outputorientierung).

Bei der abschließenden Podiumsdiskussion unterstrich Prof. Dr. Kerstin Schneider (RatSWD) noch einmal die große Bedeutung eines besseren Datenzugangs für die Forschung. Hier hinke Deutschland im internationalen Vergleich derzeit hinterher.

↳ www.destatis.de

33. Wissenschaftliches Kolloquium

Am 28. und 29. November 2024 hat das Statistische Bundesamt gemeinsam mit der Deutschen Statistischen Gesellschaft das 33. Wissenschaftliche Kolloquium durchgeführt zum Thema „Energiewende und Energiepreiskrise – zur Rolle der Daten“. Programm und Abstracts stehen auf der Veranstaltungsseite bereits zur Verfügung, später wird dort auch die Tagungsdokumentation veröffentlicht:

↳ www.destatis.de

Jahrestagung des Landesamtes für Statistik Niedersachsen

Die Jahrestagung des Landesamtes für Statistik Niedersachsen (LSN) widmete sich am 19. September 2024 in Hannover dem Thema „Datenkompetenz in Gegenwart und Zukunft“. Sie hatte das Ziel, durch Vorträge und eine Podiumsdiskussion Expertinnen und Experten aus Wissenschaft, Bildung, Wirtschaft, Verwaltung und Politik zusammenzubringen und gemeinsam mit den rund 80 Teilnehmenden über das Thema zu diskutieren.

Daten bieten enormes Potenzial für vielfältige Anwendungsgebiete. Damit geht aber auch eine Diversifizierung der Datenquellen und -anbieter einher. Das Angebot ist schwer zu kanalisieren und zu kategorisieren, kurz: die Datenlage wird unübersichtlicher – vor allem für die interessierte Öffentlichkeit und Menschen, die für ihre Arbeit nur gelegentlich Daten nutzen. Die amtliche Statistik hat den Anspruch, in dieser Informationsflut als unabhängiger Datenproduzent mit neutralen und objektiven Daten Orientierung und Hilfestellung zu bieten.

Neben den eher auf die Rolle des Datenproduzenten gerichteten Aspekten ging es auch um die Frage, welche Maßnahmen zu ergreifen sind, um die Datenkompetenz in der Gesellschaft zu fördern. Denn nur wer Datenkompetenz besitzt, kann

- › die Zuverlässigkeit von Datenquellen beurteilen,
- › Daten zielgerichtet aufbereiten sowie einordnen und
- › sinnvolle Schlüsse ziehen.

Eine stärkere Nutzung von Daten kann zu besserer gesellschaftlicher Teilhabe, höherer wirtschaftlicher Wertschöpfung und fundierteren politischen Entscheidungen führen.

Die Präsentationen der verschiedenen Vorträge stehen auf der Website des Landesamtes für Statistik Niedersachsen zur Verfügung:

↳ magazin.statistik.niedersachsen.de

Fachtagung „Arbeitsmarkt und Demografischer Wandel“

Ein großer Kreis von Fachleuten aus der Arbeitsmarktforschung, der amtlichen Statistik und der Arbeitsmarktpolitik traf sich am 1. und 2. Oktober 2024 zur Fachtagung „Arbeitsmarkt und Demografischer Wandel – Empirische Befunde und Forschung zum Fachkräftebedarf“ im Statistischen Bundesamt.

Ein viel diskutiertes Thema zum Arbeitsmarktgeschehen ist der Fachkräftebedarf. Im Zentrum der Tagung stand die Frage, wie die derzeitige Situation und die künftige Entwicklung – gerade auch vor dem Hintergrund des demografischen Wandels in Deutschland – im Lichte der aktuellen Daten einzuschätzen sind. Knapp 50 Teilnehmerinnen und Teilnehmer diskutierten, wie sich der demografische Wandel auf das Arbeitsangebot auswirkt und welche Potenziale bestehen, um einem Mangel an Fachkräften zu begegnen. Darüber hinaus ging es auch um die Frage, welche Datenquellen und Indikatoren die aktuelle Entwicklung messen und abbilden können.

Die Bedeutung und Vielfalt des Themas zeigte sich in dem breiten Spektrum der Beiträge. Sie widmeten sich sehr unterschiedlichen Fragestellungen, wie den Folgen des demografischen Wandels, einer möglichen Aktivierung zusätzlicher Potenziale für den Arbeitsmarkt, dem Beitrag von Zuwanderung und Integration sowie neuen Strategien zum Umgang mit dem Arbeits- und Fachkräftemangel. Die Diskussionen verdeutlichten, dass die amtliche Statistik bereits ein umfangreiches Datenangebot zum Thema „Arbeitsmarkt und Demografischer Wandel“ zur Verfügung stellt. Gleichzeitig sind aber auch die permanenten Veränderungen auf dem Arbeitsmarkt zu beobachten, um die Berichterstattung kontinuierlich weiterentwickeln zu können.

↳ www.destatis.de

NEUERSCHEINUNGEN

Sozialbericht 2024

Wie leben die Menschen in Deutschland? Welche Trends und Entwicklungen sind erkennbar? Welche gesellschaftlichen Herausforderungen gibt es? Der Datenreport, die bewährte Verbindung aus amtlicher Statistik und sozialwissenschaftlicher Bewertung, ist in seiner 18. Ausgabe unter dem neuen Titel „Sozialbericht 2024. Ein Datenreport für Deutschland“ erschienen. Erstmals bündelt der verbesserte Online-Auftritt die Inhalte auf der neuen Website www.sozialbericht.de; der Webauftritt wird sukzessive noch um interaktive Grafiken erweitert.

Einige ausgewählte Ergebnisse des Berichts sind:

- › Die Vermögen in Deutschland sind ungleich verteilt, insbesondere zwischen Ost und West.
- › Menschen mit Einwanderungsgeschichte stellen mehr als ein Viertel der Erwerbspersonen und sind mit durchschnittlich 37,7 Jahren wesentlich jünger als Menschen ohne Einwanderungsgeschichte (47,2 Jahre).
- › Der Arbeitsmarkt könnte 645 000 Vollzeit-Arbeitskräfte dazugewinnen, wenn Mütter in dem Umfang erwerbstätig sein könnten, in dem sie es sich wünschen. Dazu braucht es unter anderem mehr Kindertagesbetreuung. Hier sind die Betreuungsquoten der Kinder unter drei Jahren in den letzten 15 Jahren zwar gestiegen, aber der Anstieg hat sich seit 2014 verlangsamt.

↳ www.sozialbericht.de

Jahresgutachten 2024/25: „Versäumnisse angehen, entschlossen modernisieren“

Der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung hat sein Jahresgutachten 2024/25 „Versäumnisse angehen, entschlossen modernisieren“ am 13. November 2024 der Bundesregierung überreicht. Neben dem Einleitungskapitel zur wirtschaftlichen Lage analysiert das Gutachten in vier weiteren Kapiteln Bereiche, die grundlegend modernisiert wer-

den sollten. Die Handlungsempfehlungen beziehen sich auf öffentliche Ausgaben, die Digitalisierung im Finanzsystem, den Wohnungsmarkt in Ballungsräumen und die Verkehrsinfrastruktur.

↳ www.sachverstaendigenrat-wirtschaft.de

Fachbericht „Zeitverwendung für Kultur und kulturelle Aktivitäten in Deutschland“

Menschen in Deutschland ab 10 Jahren verbringen je Tag im Durchschnitt eine Stunde und 18 Minuten mit Kulturaktivitäten wie dem Besuch kultureller Veranstaltungen und Einrichtungen oder künstlerischen Tätigkeiten. Auch Lesen, Musikhören oder Spielen – analog wie digital – zählen zu den Kulturaktivitäten.

Im Zehnjahresvergleich zeigt sich ein Anstieg der für Kultur und kulturelle Aktivitäten aufgewendeten Zeit um eine Stunde und 40 Minuten je Woche. Dieser ist hauptsächlich durch eine Zunahme beim Spielen am Computer, Tablet oder Smartphone bedingt (+ 1,5 Stunden).

Die Zeitverwendung für Kultur und kulturelle Aktivitäten unterscheidet sich nicht nur nach der Lebensphase von Personen, sondern auch nach dem Geschlecht, Alter und dem Bildungsstand. Dies zeigen weitere Analysen im Fachbericht „Zeitverwendung für Kultur und kulturelle Aktivitäten in Deutschland“ des Statistischen Bundesamtes:

↳ www.destatis.de

Health at a Glance: Europe 2024

Die durchschnittliche Lebenserwartung in Deutschland wird für das Jahr 2023 auf 81,2 Jahre geschätzt und lag damit zum ersten Mal unter dem EU-Durchschnitt (81,5 Jahre) und zwischen 2,6 und 3 Jahren unter der Lebenserwartung in Spanien, Italien und der Schweiz. Die Europäische Union steht vor einem Arbeitskräftemangel im Gesundheitswesen. Im Jahr 2022 haben EU-weit schätzungsweise 1,2 Millionen Ärztinnen und Ärzte, Krankenpflegerinnen und -pfleger und Hebammen gefehlt.

Deutschland hat im Vergleich zwar eine hohe Verfügbarkeit von ärztlichem und Pflegepersonal, aber auch ein hohes Aktivitätsniveau im Gesundheitswesen: Beispielsweise liegen die Krankenhausentlassungen je 1 000 Einwohnerinnen und Einwohner in Deutschland 40 % über dem EU-Durchschnitt.

Der OECD-Bericht „Health at a Glance: Europe 2024. State of Health in the EU Cycle“ widmet sich diesen Herausforderungen. Neben Vergleichen des Gesundheitszustands, der Gesundheitsrisiken, der Gesundheitsausgaben, des Zugangs und der Qualität der Versorgung in 40 europäischen Ländern enthält der Bericht Daten und Analysen zur Förderung einer gesunden Lebenserwartung und der Bekämpfung des Arbeitskräftemangels im Gesundheitswesen.

↳ www.oecd.org

ERMITTLUNG DER EINWOHNERZAHL IM ZENSUS 2022

Corinna Bretschi, Steffen Seibel, Ingeborg Vorndran, Miriam Pfahl

↳ **Schlüsselwörter:** Registergestützter Zensus – Melderegister – Mehrfachfallprüfung – Haushaltebefragung – Referenzdatenbestand

ZUSAMMENFASSUNG

Wesentliches Ziel eines Zensus ist die Ermittlung der Einwohnerzahlen für Bund, Länder und Gemeinden. Der Zensus 2022 wurde wie bereits der Zensus 2011 registergestützt durchgeführt. So basiert die Ermittlung der Einwohnerzahl auf den Melderegistern. Diese wurden jedoch nicht einfach ausgezählt, sondern es gab noch eine Reihe ergänzender und korrigierender Maßnahmen – teils innerhalb der Daten aus den Registern, teils durch zusätzliche primärstatistische Erhebungen. Der Beitrag stellt die einzelnen Schritte der Einwohnerzahlermittlung im Zensus 2022 vor.

↳ **Keywords:** register-assisted census – population register – check for multiple registration – household survey – stock of reference data

ABSTRACT

A primary objective of a census is to establish population counts for Germany, the Länder and the municipalities. The 2022 Census was conducted as a register-assisted census of the population, as was the 2011 Census. Population figures were therefore determined on the basis of the population registers. The entries in these registers were not simply counted, however; instead, a number of supplementary and corrective measures were also implemented, with some applying directly to the data from the registers and some involving additional primary data surveys. This article explains the individual steps involved in counting the population in the 2022 Census.

Corinna Bretschi

studierte Survey Methodology (MA) an der Universität Duisburg-Essen. Sie ist in der Gruppe „Registerbasierter Datenbestand des Zensus“ des Statistischen Bundesamtes Referentin im Referat „Methodik, Haushalte“ und war beim Zensus 2022 schwerpunktmäßig für die methodische Konzipierung der Haushaltsstichprobe zuständig.

Steffen Seibel

studierte Mathematik und Geographie an der Universität des Saarlandes. Er leitet in der Gruppe „Registerbasierter Datenbestand des Zensus“ des Statistischen Bundesamtes das Referat „Methodik, Haushalte“.

Ingeborg Vorndran

ist Diplom-Volkswirtin und leitete von November 2014 bis zu ihrem Ruhestand im Juli 2024 im Statistischen Bundesamt das Referat „Melderegister, Sonderbereiche, Mehrfachfallprüfung“ sowie ab 2023 auch die Haushaltegenerierung. Ihre Schwerpunkte lagen auf der konzeptionellen Vorbereitung und Umsetzung der Datenlieferungen aus den Melderegistern sowie der Weiterentwicklung der Methodik für die Erhebung an Wohnheimen und Gemeinschaftsunterkünften, den sogenannten Sonderbereichen, als Bestandteile des Zensus.

Miriam Pfahl

hat Politikwissenschaft, Soziologie und Öffentliches Recht studiert und ist als Referentin im Referat „Gebäude und Wohnungen“ in der Gruppe „Registerbasierter Datenbestand des Zensus“ des Statistischen Bundesamtes tätig. Für den Zensus 2022 war sie unter anderem für die Konzeption der Online-First-Strategie und die Entwicklung der Online-Fragebogen der Haushaltebefragung auf Stichprobenbasis zuständig.

1

Einleitung

Zur Umsetzung der EU-weiten Zensusrunde 2021/2022 hat sich Deutschland – wie bereits beim Zensus 2011 – für eine registergestützte Methode entschieden und bereits vorhandene Verwaltungsregister als Datenquellen genutzt. Für die Ermittlung der Einwohnerzahl sind dies die Melderegister, obwohl nicht alle Angaben aus den Registern präzise und aktuell sind. Manche Personen sind an ihrem Wohnort gar nicht gemeldet (Unterfassung, als Zugänge zu werten), andere stehen zwar im Register, sind aber schon längst umgezogen oder bereits verstorben (Überfassung, als Abgänge zu werten). Zur Ermittlung der Einwohnerzahlen erfolgte deshalb nicht eine einfache Auszählung aus den Melderegistern; vielmehr sah der Zensus 2022 eine Reihe von ergänzenden und korrigierenden Maßnahmen vor.

Das zweite Kapitel enthält die Definition der Einwohnerzahl, ein Überblick über das Vorgehen bei deren Ermittlung folgt in Kapitel 3. Kapitel 4 erläutert die Konsolidierung des Melderegisterbestands, Kapitel 5 geht kurz auf die Mehrfachfallprüfung ein. Ausführlich beschreibt Kapitel 6 die primärstatistischen Teile zur Ermittlung der Einwohnerzahl im Zensus 2022, bevor ein kurzes Fazit den Beitrag abschließt.

2

Definition Einwohnerzahl

Ziel des Zensus ist es festzustellen, wie viele Menschen in Deutschland, in den Bundesländern und in den Gemeinden wohnen. Im deutschen Zensus 2022 zählen die nach den melderechtlichen Vorschriften zum Zensusstichtag mit alleinigem oder Hauptwohnsitz meldepflichtigen Personen zur Bevölkerung. Hinzu kommen die im Ausland tätigen Angehörigen der Bundeswehr, der Polizeibehörden und des Auswärtigen Dienstes (§ 2 Gesetz über den Auswärtigen Dienst) sowie ihre dort ansässigen Familien (§ 2 Absatz 2 Zensusgesetz 2022).¹ Aus-

¹ Die sogenannten Deutschen im Ausland zählen, soweit sie nicht im Inland gemeldet sind, zur Bevölkerung Deutschlands. Sie werden bei der Gesamteinwohnerzahl Deutschlands berücksichtigt, können jedoch keinen einzelnen Gemeinden zugeordnet werden. Sie werden somit nur bei der Darstellung der Einwohnerzahl auf Bundesebene ausgewiesen, nicht jedoch auf Länder-, Kreis- und Gemeindeebene.

gangsbasis für die Ermittlung der Einwohnerzahl einer Gemeinde sind damit die Personendatensätze mit alleinigem oder Hauptwohnsitz, die von den Meldebehörden nach dem Zensusgesetz 2022 stichtagsrelevant aus den Melderegistern übermittelt wurden.

Nach diesem Konzept sind in Deutschland auch alle aus dem Ausland zuziehenden Personen, die sich bei den Meldebehörden anmelden, zählungsrelevant.² Zwar besteht keine Meldepflicht für kurzfristige Aufenthalte (seit November 2015 bundesweit weniger als drei Monate), freiwillige Anmeldungen für kurzfristige Aufenthalte sind aber möglich. Freiwillig Gemeldete sind nicht zählungsrelevant.³

Personen, die aus einer Wohnung ausziehen und keine neue Wohnung im Inland beziehen, müssen sich abmelden, unabhängig von der Dauer ihrer Abwesenheit. Wohnungslose, die an Anschriften gemeldet sind, von denen bekannt ist, dass sie dort nicht wohnen (zum Beispiel in Rathäusern), zählen nicht zur Einwohnerzahl.⁴ Wohnungslose, die sich in (Not-)Unterkünften aufhalten und von der jeweiligen Einrichtungsleitung gemeldet werden, zählen hingegen dazu.

Die gewählte Abgrenzung der Bevölkerung entspricht den Vorgaben der geltenden europäischen Rechtsgrundlage in Verbindung mit den Regelungen im Bundesmeldegesetz. Die europäische Definition für Bevölkerung⁵ legt einerseits den üblichen Aufenthaltsort, die sogenannte

- ² Ausnahme bilden die sogenannten freiwilligen Meldungen. Hierbei handelt es sich in der Regel um Personen, die entsprechend internationaler Abkommen bei einem Zensus nicht (im Gastland) zu zählen sind, sich aber freiwillig bei den Meldebehörden registrieren. Diese sind von der Zählung auszuschließen und sollten in der Haushalbefragung nicht befragt werden. Zu den Personen, die nicht zu zählen sind, gehören die in Deutschland stationierten Mitglieder ausländischer Streitkräfte und deren Angehörige sowie die in Deutschland tätigen Mitglieder diplomatischer und berufskonsularischer Vertretungen sowie deren Familienangehörige, falls die genannten Personen weder die deutsche Staatsangehörigkeit besitzen noch im Inland ständig ansässig sind, noch dort eine private Erwerbstätigkeit ausüben, sowie Personen, für die diese Befreiung in völkerrechtlichen Übereinkünften festgelegt ist.
- ³ Hierzu zählen zum Beispiel Saisonarbeiterinnen und Saisonarbeiter. Dieser Personenkreis ist aufgrund seines vorübergehenden Aufenthalts in Deutschland nicht meldepflichtig, kann sich jedoch auf eigenen Wunsch bei den Meldebehörden freiwillig registrieren lassen. Freiwillig Gemeldete werden nicht zur Bevölkerung gezählt.
- ⁴ Dies ergibt sich unmittelbar aus der Regelung in den §§ 17 und 19 Bundesmeldegesetz. Danach ist es verboten, einem Dritten eine Anschrift für eine Anmeldung zur Verfügung zu stellen, obwohl ein tatsächliches Wohnen gar nicht vorliegt (§ 19 Absatz 6 Bundesmeldegesetz).
- ⁵ Siehe Artikel 2 der Zensus-Rahmenverordnung (EG) Nr. 763/2008.

Zwölf-Monats-Regel, zugrunde: So werden Personen in einem Land nur dann gezählt, wenn ihre (tatsächliche oder beabsichtigte) Aufenthaltsdauer mindestens zwölf Monate beträgt. Andererseits zählen Personen, die für weniger als zwölf Monate fortgezogen sind, weiterhin zur Bevölkerung.

Da diese Abgrenzung im Wesentlichen nur im Rahmen einer primärstatistischen Erhebung zu erfassen ist, wurde eine Ausnahmeklausel in die EU-Rahmenverordnung eingefügt. Diese besagt, dass der übliche Aufenthaltsort auch als der Ort des rechtmäßigen oder eingetragenen Wohnsitzes aufgefasst werden kann, wenn die oben genannte Zwölf-Monats-Regel nicht festgestellt werden kann. Diese Ausnahmeregel legitimiert die Vorgehensweise der Mitgliedstaaten, die ihre Einwohnerzahlen basierend auf den Angaben ihrer Melderegister ermitteln.

Gegenüber dem Melderechtsrahmengesetz, das für den Zensus 2011 die Grundlage war, gelten nach dem Bundesmeldegesetz für die nachfolgenden Personengruppen im Zensus 2022 abweichende einwohnerzahle relevante Regelungen:

- › Personen in dienstlich bereitgestellten Unterkünften der Bundeswehr, der Polizei und des Bundesfreiwilligen-/Zivildienstes sowie Personen in Justizvollzugsanstalten (soweit die Dauer von zwölf Monaten nicht überschritten wird) oder Personen in Krankenhäusern, Pflegeheimen oder sonstigen Einrichtungen, die der Betreuung pflegebedürftiger oder behinderter Menschen oder der Heimerziehung dienen, müssen sich dort nicht anmelden, soweit sie anderweitig im Inland gemeldet sind.
- › Binnenschiffer und Seeleute müssen sich nur dann bei der Meldebehörde des Heimatortes ihres Schiffes oder Reeders anmelden, wenn sie in keiner anderen Wohnung im Inland gemeldet sind.
- › Für die Beherbergungsstätten gelten jetzt die gleichen Meldepflichten wie für die sonstigen Wohnungen, das heißt Personen, die im Inland gemeldet sind (bei einem Aufenthalt von länger als sechs Monaten), und Personen, die nicht im Inland gemeldet sind (bei einem Aufenthalt von länger als drei Monaten), sind meldepflichtig.

Das heißt Mitglieder der im ersten Aufzählungspunkt genannten Personengruppe werden im Zensus 2022 abweichend vom Zensus 2011 aufgrund der Änderungen des Bundesmelderechts an ihrer Meldeanschrift gezählt, selbst wenn sie am Stichtag an einer sogenannten Sonderanschrift⁶ wohnen.

3

Überblick über die Ermittlung der Einwohnerzahl

Die Ermittlung der Einwohnerzahl im Zensus 2022 fand in mehreren Schritten statt. Zum einen erfolgte eine Bereinigung innerhalb der von den Melderegistern übernommenen Daten durch eine Mehrfachfallprüfung. Bei der dezentralen Führung der Melderegister in Deutschland ist nicht auszuschließen, dass Personen in mehreren Gemeinden gleichzeitig mit mehr als einer alleinigen Wohnung oder Hauptwohnung (sogenannte Dublette) oder ausschließlich mit (einer oder mehreren) Nebenwohnung(en) gemeldet sind. Im Zuge einer Mehrfachfallprüfung werden solche unzulässigen Dubletten und Personen, die ausschließlich mit Nebenwohnungen gemeldet sind, ausgesteuert.

Zum anderen gab es zwei primärstatistische Erhebungen mit dem Ziel der statistischen Registerkorrektur: die Haushaltebefragung auf Stichprobenbasis (an sogenannten Normalanschriften) und die Befragung in Wohnheimen und Gemeinschaftsunterkünften (Boragk und andere, 2024; Dreschmitt/Pfahl, 2024).

Mit den Ergebnissen der Haushaltebefragung wurde an zufällig ausgewählten Stichprobenanschriften abgeglichen, wie viele Menschen laut Melderegister an der Anschrift leben müssten, aber faktisch nicht mehr dort wohnten, also Übererfassungen (Karteileichen) sind, beziehungsweise wie viele Menschen an der Anschrift

⁶ Zu den Sonderanschriften zählen alle Anschriften mit Einrichtungen oder Institutionen, deren Bewohnerinnen und Bewohner oder die dort gemeldeten Personen nicht über das normale Zensusverfahren erhoben werden dürfen oder können. Diese Einrichtungen beziehungsweise Institutionen werden im Zensus als Sonderbereiche bezeichnet. Sonderbereiche sind Anschriften mit Gemeinschaftsunterkünften oder Wohnheimen, in denen Personen in der Regel längerfristig untergebracht beziehungsweise wohnhaft sind. Dabei wird zwischen Gemeinschaftsunterkünften (zum Beispiel Einrichtungen für ältere und pflegebedürftige Menschen) und Wohnheimen (zum Beispiel Studierendenwohnheimen) unterschieden.

wohnten, aber zum Zensusstichtag (noch) nicht gemeldet waren, demnach Untererfassungen (Fehlbestände) sind. Diese Ergebnisse wurden später auf die Gesamtbevölkerung der Gemeinde hochgerechnet. Eine Befragung an allen Anschriften mit Wohnheimen und Gemeinschaftsunterkünften ergänzte die Haushaltebefragung. Diese gesonderte Befragung war notwendig, weil die melderechtlichen Angaben für die Bewohnerinnen und Bewohner von Wohnheimen und Gemeinschaftsunterkünften häufig ungenau sind, da es hier in der Regel eine hohe Fluktuation gibt.

Die gewonnenen Informationen wurden in einem zentralen Datenbestand, dem Referenzdatenbestand, zusammengefasst. Der Referenzdatenbestand besteht aus dem Steuerungsregister, das alle für den Zensus relevanten Anschriften enthält, und dem Personenbestand, der alle für den Zensus relevanten Personen enthält und sie einer Anschrift zuordnet. Der Referenzdatenbestand hat auch eine koordinierende Rolle: Er ist zugleich der Grundstein für die Verknüpfung zwischen den Erhebungsteilen und die Basis der Kommunikation zwischen den einzelnen Datenquellen. Die Schritte zur Ermittlung der amtlichen Einwohnerzahl einer Gemeinde zeigt [Grafik 1](#).

Nach §16 Zensusgesetz 2011 war in Gemeinden mit weniger als 10 000 Einwohnerinnen und Einwohnern (sogenannte kleine Gemeinden) beim Zensus 2011 die Befragung zur Klärung von Unstimmigkeiten durch-

geführt worden. Im Zensus 2022 wird sie nicht mehr als Instrument zur Korrektur der Melderegisterdaten verwendet.¹⁷ Dagegen erfolgt nun in allen Gemeinden eine Korrektur des um Dubletten bereinigten Melderegisterbestands über die Haushaltebefragung.

4

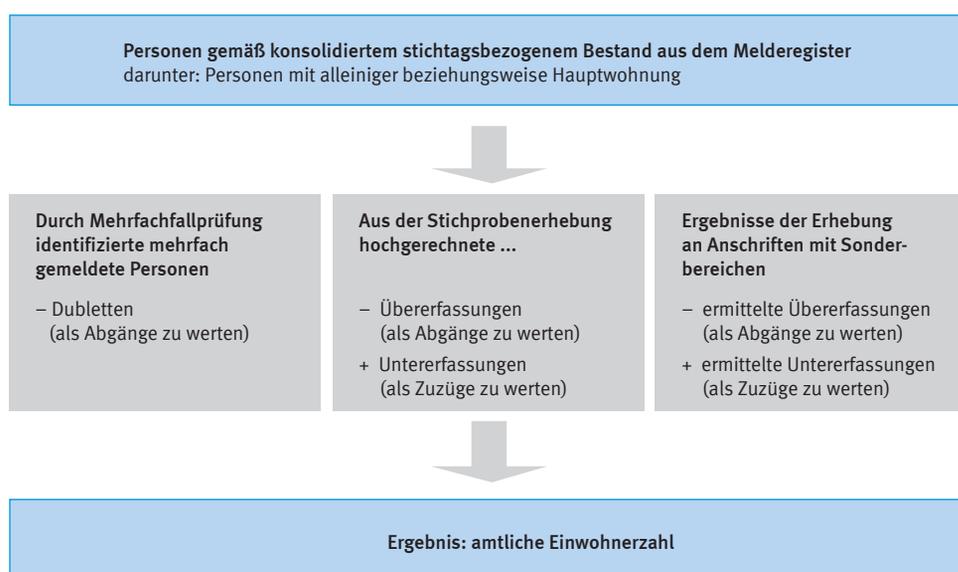
Konsolidierter Melderegisterbestand

Zur Ermittlung der Einwohnerzahl der Gemeinden fließen von allen Gemeinden zwei Datenlieferungen in die Berechnung ein. Die erste einwohnerzahlrelevante Lieferung des Melderegisters erfolgte zum Stichtag 15. Mai 2022 (MR.Z1), die zweite Lieferung drei Monate danach am 14. August 2022 (MR.Z2). Die zweite Datenlieferung enthält neben den zu diesem Zeitpunkt gemeldeten Personen zusätzlich den erweiterten Personenkreis. Hierzu zählen Datensätze zu Personen, die in der ersten Lieferung enthalten, aber bis zum Zensusstichtag bereits verstorben oder verzogen waren. Außerdem enthalten sind Angaben zu Personen, die bis zum Stichtag zugezogen oder geboren, aber zum Stichtag noch nicht

⁷ Eine Evaluierung der Ergebnisse des Zensus 2011 sowie weitere Untersuchungen der Statistischen Ämter des Bundes und der Länder haben gezeigt, dass für den Zensus 2022 eine Befragung zur Klärung von Unstimmigkeiten in kleinen Gemeinden keinen ausreichenden Korrektoreffekt liefern würde.

Grafik 1

Schritte zur Ermittlung der amtlichen Einwohnerzahl einer Gemeinde



angemeldet waren, auch wenn sie bis zum Zeitpunkt der zweiten Lieferung bereits wieder verzogen oder verstorben sind.

Damit wird die Feststellung der Stichtagsrelevanz für den betroffenen Personenkreis unter Berücksichtigung von Verzögerungen bei der Meldung an die Meldeämter sowie einer gewissen Bearbeitungsdauer in den Meldeämtern ermöglicht. Aus diesen beiden Datenlieferungen wird ein konsolidierter Datenbestand erstellt, welcher bereits um technische⁸ und temporäre⁹ Dubletten bereinigt ist und für die weiteren Prozessschritte – die Mehrfachfallprüfung, die Hochrechnung und die Haushalgenerierung – genutzt wird.

5

Mehrfachfallprüfung

Bei der Mehrfachfallprüfung wird bundesweit geprüft, ob Personen am Zensusstichtag mit mehr als einer alleinigen Wohnung oder Hauptwohnung (Mehrfachfälle) im konsolidierten Stichtags-Melderegisterbestand enthalten sind. Diese Mehrfachfälle werden maschinell den melderechtlichen Bestimmungen entsprechend anhand der Merkmale zur betroffenen Person, wie Name, Vorname, Geburtsdatum und -ort in Verbindung mit Informationen zu Anschriften und verzeigerten¹⁰ Familienangehörigen, geprüft. Es werden nur eindeutige Mehrfachfälle bereinigt: Handelt es sich bei zwei oder mehr Datensätzen um Dubletten beziehungsweise Mehrfachfälle, ist der aktuellere Datensatz zählungsrelevant; die älteren Datensätze werden mit einem Löschkennzeichen versehen und damit nicht gezählt. Zur Ermittlung des aktuellsten Datensatzes wird zum Beispiel das Datum der Anmeldung, eines Wohnungsstatuswechsels oder der letzten Familienstandsänderung betrachtet.

Im Zuge der Mehrfachfallprüfung wird zudem für Fehlbestände, die an Anschriften mit Sonderbereichen er-

8 Eine technische Dublette ist ein doppelter Personendatensatz innerhalb einer Anschrift(-engruppe), der beim Integrationsprozess der Dateneingänge entstehen kann. Es handelt sich jedoch nicht um eine tatsächliche Dublette in den Melderegistern.

9 Temporäre Dubletten enthalten stichtagsrelevante Datensätze an verschiedenen Anschriften(-gruppen). Sie können durch die Integration der Datenlieferungen MR.Z1 und MR.Z2 entstehen, sind aber keine wirklichen Dubletten der Melderegister am Zensusstichtag.

10 Verzeigerungen sind Querverweise im Melderegisterbestand zu Ehepartnern und Kindern.

mittelt werden, geprüft, ob die Person noch außerhalb der Sonderanschrift an einer Normalanschrift oder an einer weiteren Sonderanschrift gemeldet ist. An welcher Anschrift die Fehlbestände der Wohnheime und Gemeinschaftsunterkünfte als Einwohnerinnen und Einwohner gezählt werden, erfolgt dann aufgrund von im Vorfeld festgelegten Regeln für die Wohnungsstatusfestlegung (siehe auch Abschnitt 6.4).

Für die über die Haushaltebefragung festgestellten Register-Fehlbestände wird hingegen keine Mehrfachfallprüfung durchgeführt.¹¹

6

Primärstatistische Teile

Zu den primärstatistischen Teilen des Zensus 2022 gehören die Haushaltebefragung auf Stichprobenbasis an Normalanschriften sowie die Vollerhebung an Anschriften mit Wohnheimen und Gemeinschaftsunterkünften.

6.1 Erhebung in Wohnheimen und Gemeinschaftsunterkünften

Die Erhebung an Anschriften mit Wohnheimen und Gemeinschaftsunterkünften wurde als Vollerhebung durchgeführt. In Wohnheimen wurden die Befragten direkt befragt, während in Gemeinschaftsunterkünften eine Befragung über die Einrichtungsleitung erfolgte. Sogenannte Mischanschriften liegen vor, wenn an einer Anschrift gleichzeitig ein Wohnheim oder eine Gemeinschaftsunterkunft und/oder ein Normalbereich vorliegen. Die Fehlbestände der Sonderbereichserhebung wurden im Zuge der Mehrfachfallprüfung deutschlandweit mit den Melderegistern abgeglichen und danach der Wohnungsstatus (alleiniger Wohnsitz, Hauptwohnsitz, Nebenwohnsitz) festgelegt. Da an allen Anschriften mit Sonderbereichen die Angaben zu den Personen erhoben wurden, mussten die festgestellten Über- und Untererfassungen nicht hochgerechnet werden. Sie wurden ausgezählt und flossen direkt in die Einwohnerzahl ein.

11 Zur Begründung: Fehlbestände, die an einer Stichprobenanschrift festgestellt wurden, werden nicht einer Dublettenprüfung unterzogen, da der Dublettenpartner theoretisch bereits in den hochgerechneten Karteileichen enthalten ist.

6.2 Haushaltebefragung auf Stichprobenbasis

Im Zuge der Haushaltebefragung auf Stichprobenbasis wurden alle zum Stichtag an einer zufällig ausgewählten Anschrift lebenden Personen ermittelt und befragt. Interviewerinnen und Interviewer suchten die Stichprobenanschriften auf, um einen persönlichen Kontakt zu mindestens einer Person je Haushalt herzustellen. In einer persönlichen Befragung wurden die Anzahl der im Haushalt wohnhaften Personen sowie mindestens deren Kernmerkmale Name, Vorname, Geburtsdatum und Geschlecht erfasst. Die Befragungsergebnisse beziehungsweise die entsprechend ausgefüllten Erhebungsunterlagen wurden anschließend an die zuständige kommunale Erhebungsstelle zurückgegeben. Dort erfolgten die Bestätigung der festgestellten Existenzen unter zuvor festgelegten Voraussetzungen¹² sowie deren Übermittlung an den Referenzdatenbestand. Anschließend fand ein Abgleich mit dem konsolidierten stichtagsbezogenen Melderegisterbestand (nach Mehrfachfallprüfung) statt, um sowohl Karteileichen als auch Fehlbestände in den Melderegistern festzustellen. Die Einwohnerzahl ergibt sich aus der anschließenden statistikinternen Bereinigung des stichtagsbezogenen Melderegisterbestands um die aus der Haushaltebefragung hochgerechneten Über- und Untererfassungen und dem Ergebnis der Erhebung an Anschriften mit Wohnheimen und Gemeinschaftsunterkünften.

Die Auswahlgrundlage für die Hauptziehung der Haushaltsstichprobe stellten alle Anschriften mit potenziellem Wohnraum (mit Ausnahme von Anschriften mit Wohnheimen und Gemeinschaftsunterkünften) aus dem Steuerungsregister mit Stand der Melderegisterlieferung zum Stichtag 7. Februar 2021 dar. Für Anschriften mit Wohnraum, die zwischen der Hauptziehung der Stichprobe und der Melderegisterlieferung zum Stichtag 14. November 2021 neu in das Steuerungsregister aufgenommen wurden, oder die in diesem Zeitraum erstmals als mit Wohnraum markiert wurden, wurde

12 Voraussetzung für die Bestätigung einer Existenz ist das vollständige Vorliegen der Kernmerkmale Name, Vorname, Geburtsdatum, Geschlecht. Abhängig von der Quelle/Herkunft der Kernmerkmale sind gegebenenfalls weitere Maßnahmen, wie Postzustellauftrag oder Mahnverfahren, zur Bestätigung der Existenz erforderlich – beispielsweise, wenn Interviewerinnen und Interviewer in der persönlichen Befragung nicht alle Kernmerkmale erheben konnten.

eine ergänzende Stichprobe (Nachziehung) gezogen.¹³ Danach noch stichtagsrelevant ins Steuerungsregister aufgenommene Neuzugänge¹⁴ wurden bei der Ermittlung der Einwohnerzahl ohne interne Melderegister-Korrektur durch die Haushaltebefragung berücksichtigt (zum Beispiel Neubaugebiete), das heißt man ging davon aus, dass dort das Melderegister nach Mehrfachfallprüfung und -bereinigung korrekt ist (Klink/Lorentz, 2022).

Das Steuerungsregister enthält alle Anschriften mit Wohnraum und mit bewohnten Unterkünften. Es enthält auch unbemeldete Anschriften mit Wohnraum, sogenannte Nullanschriften. Als unbemeldet gilt eine Anschrift immer dann, wenn laut Melderegister niemand an dieser Anschrift gemeldet ist. Unbemeldete Anschriften mit Wohnraum gehören zur Auswahlgrundlage; sie werden bei der Stichprobenziehung bewusst berücksichtigt, um komplette Fehlbestandshaushalte aufzudecken und damit eine systematische Untererfassung der Einwohnerzahlen vermeiden zu können. Anschriften mit Wohnheimen und Gemeinschaftsunterkünften gehören hingegen nicht zur Auswahlgrundlage. An diesen Anschriften wird die Einwohnerzahl – wie oben beschrieben – mit der Erhebung an Anschriften mit Wohnheimen und Gemeinschaftsunterkünften als Voll-erhebung ermittelt.

Die Auswahl der Stichprobenanschriften erfolgt durch eine mathematisch-statistische Zufallsauswahl. Damit hat jede Anschrift mit Wohnraum an einer Normalanschrift in Deutschland eine Chance, für die Haushaltebefragung ausgewählt zu werden.

13 Anders als beim Zensus 2011 gab es im Zensus 2022 nur eine Stichprobennachziehung.

14 Es handelt sich hierbei meist um Anschriften, an denen bisher keine Personen gemeldet waren, zum Beispiel Neubaugebiete. Eine Stichprobe zur Bereinigung ist aufgrund des engen Zeitfensters zum Stichtag und der damit verbundenen Aktualität der Daten nicht notwendig. Die an diesen Anschriften gemeldeten Personen werden daher ohne weitere Prüfung als Einwohnerinnen und Einwohner gezählt. Das bedeutet, dass die Melderegister-Einträge (mit alleiniger oder Hauptwohnung) dieser Neuzugangsanschriften auf das Hochrechnungsergebnis (dessen Hochrechnungsrahmen die Neuzugänge nicht umfasst) addiert werden.

Präzisionsziel und Stichprobenumfang

Das Zensusgesetz 2022 gibt vor, dass grundsätzlich in allen Gemeinden eine statistikerinterne Korrektur des dublettenbereinigten Melderegisterbestands über die Haushaltebefragung erfolgen muss. Ferner ist festgelegt, welches Präzisionsziel für die Einwohnerzahl abhängig von der Gemeindegröße anzustreben ist.

Rückblick Zensus 2011: Beim Zensus 2011 wurden unterschiedliche Verfahren angewendet, um die Angaben aus den Melderegistern zu überprüfen und zu korrigieren. In allen Gemeinden gab es – unabhängig von ihrer Einwohnerzahl – eine Bereinigung um unzulässige Dubletten (Mehrfachfallprüfung) sowie eine Vollerhebung an Wohnheimen und Gemeinschaftsunterkünften. In großen Gemeinden mit mindestens 10 000 Einwohnerinnen und Einwohnern wurden darüber hinaus die aus Registern stammenden Daten mithilfe einer Stichprobe bereinigt. Dabei wurden etwa 10 % aller in Deutschland lebenden Personen persönlich befragt. Die so ermittelten Fehlbestände und Karteileichen wurden anschließend auf die gesamte Gemeinde hochgerechnet. In kleinen Gemeinden unter 10 000 Einwohnerinnen und Einwohnern gab es eine solche Stichprobe nicht. Hier fand stattdessen eine Befragung zur Klärung von Unstimmigkeiten statt: Es wurden Personen in der Regel nur an solchen Anschriften persönlich befragt, bei denen es Unstimmigkeiten zwischen den Angaben aus den Melderegistern und den Angaben der Gebäude- und Wohnungszählung gab. Mit den Ergebnissen des Zensus 2011 und deren Evaluierung liegt heute die Erkenntnis vor, dass in kleinen Gemeinden der tatsächliche, ex post festgestellte Korrekturbedarf 2011 höher war als vorab erwartet. Ein Festhalten am 2011er-Verfahren beim Zensus 2022 hätte dazu geführt, dass die Korrektur der Registerangaben bei kleinen Gemeinden vermutlich erneut nicht ausreichend ausfallen würde und ließ sich daher methodisch nicht rechtfertigen. Aus diesen Gründen gab es beim Zensus 2022 eine Haushaltebefragung auf Stichprobenbasis in allen Gemeinden.¹⁵

Zensus 2022: Wäre auch für alle kleinen Gemeinden die gleiche Präzision von 0,5 % einfachem relativem Standardfehler anzustreben wie in großen Gemeinden, müsste – rein rechnerisch – die Einwohnerzahl in Gemeinden mit weniger als 100 Einwohnerinnen und Einwohnern auf Bruchteile einer Person genau ermittelt werden. Eine solche Forderung ist aus fachlicher/wissenschaftlicher Sicht weder sinnvoll noch hinsichtlich der Belastung der Befragten vertretbar. Daher wurde das angestrebte prozentuale Präzisionsziel in kleinen Gemeinden im Vergleich zu großen Gemeinden gelockert. Im Gesetzgebungsverfahren zum Zensus 2022 wurde der wissenschaftlichen Empfehlung des Arbeitskreises „Mathematisch-statistische Methoden“ gefolgt und festgelegt, dass hier eine Präzisionszielfunktion zur Bestimmung des angestrebten Präzisionsziels einer Gemeinde verwendet werden sollte. Die Präzisionszielfunktion gibt für Gemeinden mit weniger als 10 000 Einwohnerinnen und Einwohnern das Präzisionsziel abhängig von der Gemeindegröße vor und geht bei 10 000 Einwohnerinnen und Einwohnern gleitend in das einheitliche Präzisionsziel von 0,5 % einfacher relativer Standardfehler wie im Zensus 2011 über. Die genaue Anwendung beziehungsweise Umsetzung der Präzisionszielfunktion erfolgte auf Basis der Empfehlungen von Professor Dr. Ralf Münnich (Universität Trier) und seinem Forschungsteam in Gemeinden zwischen 1 000 und 10 000 Einwohnerinnen und Einwohnern (Burgard und andere, 2019). Das Zensusgesetz 2022 legt fest, dass bei 1 000 Einwohnerinnen und Einwohnern ein gleitender Übergang zum einheitlichen absoluten Präzisionsziel von ± 15 Personen (einfacher absoluter Standardfehler) erfolgt. Ein Mindeststichprobenumfang von 100 Anschriften je Gemeinde konnte – zwecks belastbarer Hochrechnung – nicht unterschritten werden und somit kommt es in Gemeinden mit 100 oder weniger Anschriften zu einer Vollerhebung. Um vor diesem Hintergrund Länder mit überwiegend kleinteiliger Gemeindestruktur nicht zu benachteiligen, konnten die Länder im Gesetzgebungsverfahren zum Zensusgesetz entscheiden, ob sie als Gemeinden im Sinne dieser Vereinbarungen zum Präzisionsziel Gemeindeverbände (GV-Option) beziehungsweise sogenannte Gemeindeverbands-Reste (das heißt verbandsangehörige Gemeinden mit weniger als 2 000 Einwohnerinnen und Einwohnern; GV-Rest-Option) betrachten wollten.

Im Gesetzgebungsverfahren zum Zensusgesetz 2022 haben sich die Bundesländer Niedersachsen, Rhein-

¹⁵ Die Stichprobe zur Erfassung zusätzlicher Zensusmerkmale ist nicht Gegenstand dieses Beitrags. Sie wurde in den Gemeinden mit weniger als 10 000 Einwohnerinnen und Einwohnern als Unterstichprobe der Korrekturstichprobe realisiert.

land-Pfalz und Thüringen für die GV-Option und die Länder Mecklenburg-Vorpommern und Schleswig-Holstein für die GV-Rest-Option entschieden (§ 11 Absatz 1 Zensusgesetz 2022).

Bei der GV-(Rest-)Option erfolgt die Ermittlung der Einwohnerzahl in einem ersten Schritt für den Gemeindeverband beziehungsweise den Gemeindeverbands-Rest und wird in einem zweiten Schritt konsistent auf die verbandsangehörigen Einzelgemeinden heruntergebrochen. In dieses Herunterbrechen fließen sowohl Registerinformationen als auch Stichprobeninformationen ein, wobei den Stichprobeninformationen desto weniger Gewicht beigemessen wird, je geringer der Stichprobenumfang in dieser Gemeinde ist. Das Vorgehen zur Ermittlung der Einwohnerzahlen für verbandsangehörige Gemeinden wurde durch Professor Münnich begutachtet und als wissenschaftlich korrekt bewertet.

Unter Zugrundelegung der genannten Präzisionsziele beträgt der errechnete Umfang der Haushaltebefragung auf Stichprobenbasis bundesweit 10,3 Millionen Personen.

Auswahlverfahren/Schichtung

Um die Genauigkeit der hochgerechneten Stichprobenergebnisse zu erhöhen, wurde ein statistisches Verfahren – die Schichtung – angewandt. Dazu wurde vor der Stichprobenziehung ganz Deutschland in Erhebungsgebiete (Sampling Points, räumliche Schichtung) aufgeteilt. Ein Erhebungsgebiet entspricht entweder einer Gemeinde, einem Gemeindeverband im Fall der GV-Option, einem Gemeindeverbands-Rest im Fall der GV-Rest-Option oder einem zensuspezifischen Stadtteil. Zensuspezifische Stadtteile werden gemäß § 11 Absatz 1 Satz 1 Zensusgesetz 2022 in Großstädten mit mindestens 400 000 Einwohnerinnen und Einwohnern gebildet, sie weisen durchschnittlich 200 000 Einwohnerinnen und Einwohner auf. In Berlin beispielsweise entsprechen die zensuspezifischen Stadtteile den Berliner Bezirken. Dabei musste gewährleistet sein, dass die Einteilung sowohl lückenlos als auch überschneidungsfrei ist, dass also jede Anschrift genau einem Erhebungsgebiet angehört.

Um die Genauigkeit der Stichprobenergebnisse innerhalb der Erhebungsgebiete zu erhöhen, wurde eine zweite (fachliche) Schichtung eingeführt, bei der in allen

Erhebungsgebieten die Anschriften in verschiedene Größenklassen eingeteilt wurden. Diese Größenklassen beziehen sich darauf, wie viele Personen mit Haupt- und Nebenwohnsitz an einer Anschrift gemeldet sind (nach einer speziell dafür durchgeführten Mehrfachfallprüfung mit jeweiligem Stand des Steuerregisters für die Haupt- beziehungsweise Nachziehung). Eine Anstiftengrößenklasse in einem Erhebungsgebiet definiert eine Schicht. Anschriften mit Wohnheimen oder Gemeinschaftsunterkünften wurden für jedes Erhebungsgebiet in jeweils einer zusätzlichen Teilgesamtheit zusammengefasst, um sie gesondert berücksichtigen zu können. Die Festlegung der Anzahl der Anstiftengrößenklassen innerhalb eines Erhebungsgebiets sollte gemäß Empfehlung aus der Wissenschaft variabel abhängig von der Gemeindegröße erfolgen (1 bis 16 Schichten). Anders als im Zensus 2011 wurden im Zensus 2022 sogenannte Nullanschriften, das heißt Anschriften mit (potenziell) Wohnraum, aber ohne gemeldete Personen, in einer zusätzlichen separaten Schicht zusammengefasst.¹⁶

Hochrechnung aus der Haushaltebefragung für die Ermittlung der Einwohnerzahlen

Die über die Haushaltsstichprobe ermittelten Karteileichen und Fehlbestände werden nicht direkt auf die gesamte Gemeinde hochgerechnet. In der Regel überwiegen die paarigen Personen¹⁷ an einer Anschrift, sodass es mathematisch günstiger ist, zunächst jeweils die Zahl der mit alleinigem oder Hauptwohnsitz existenten Personen und die Zahl der mit alleinigem oder Hauptwohnsitz paarigen Personen hochzurechnen. Die hochgerechneten existenten Personen ergeben sich aus allen Personen, die in der Haushaltebefragung als Bewohnerschaft festgestellt wurden. Die hochgerechneten paarigen Personen ergeben sich aus allen Personen, die in der Haushaltebefragung in Übereinstimmung mit dem Melderegister festgestellt wurden.

16 Details zu Präzisionszielen, Schichtung, Stichprobenumfängen sowie weiteren Aspekten der Stichprobenziehung wie etwa dem Algorithmus zur Aufteilung des bundesweiten Gesamtstichprobenumfangs sind in Bretschi/Lorentz (2019) dargestellt.

17 Personen, die sowohl in der Erhebung als auch im Melderegister zählungsrelevant vorhanden sind.

Karteileichen und Fehlbestände ergeben sich dann durch Differenzbildung gemäß den Zusammenhängen:

- › Karteileichen gesamt = dublettenbereinigter Melderegisterbestand – hochgerechnete paarige Personen,
- › Fehlbestände gesamt = hochgerechnete existente Personen der Haushaltebefragung – hochgerechnete paarige Personen,
- › Einwohnerzahlbeitrag des Hochrechnungsteils = dublettenbereinigter Melderegisterbestand + Fehlbestände gesamt – Karteileichen gesamt.

Für die Schätzung der existenten und paarigen Personen wird – wie bereits im Zensus 2011 – ein verallgemeinerter Regressionsschätzer (GREG) genutzt. Die Regressionsschätzung erfolgt separat in jedem Stadtteil von Großstädten ab 400 000 Einwohnerinnen und Einwohnern sowie in allen übrigen Gemeinden beziehungsweise – bei Nutzung der GV-Option oder der GV-Rest-Option – für Gemeindeverbände und Gemeindeverbands-Reste.

Ermittlung der Einwohnerzahlen von verbandsangehörigen Gemeinden bei Nutzung der GV-Option oder der GV-Rest-Option

Eine verbandsangehörige Gemeinde im Sinne des Zensus 2022 ist eine Gemeinde, die eine echte Teilmenge eines Sampling Points ist, also eine Gemeinde, die Teil eines Sampling Points ist.

Die Stichprobenziehung erfolgte auf Sampling-Point-Ebene. Das heißt für jeden Sampling Point wurde eine Stichprobe gezogen, die eine verlässliche Schätzung der Anzahl an existenten Personen und der Anzahl an paarigen Personen im gesamten Sampling Point gewährleistet. Die Stichprobenumfänge der einzelnen Gemeinde eines Sampling Points sind im Allgemeinen aber sehr unterschiedlich. Vielfach gibt es einige verbandsangehörige Gemeinden mit mindestens 100 Stichprobenanschriften, einige Gemeinden mit 1 bis 99 Stichprobenanschriften und ein paar Gemeinden ohne Stichprobenanschriften.

Für Gemeinden mit mindestens 100 Stichprobenanschriften liefert eine Hochrechnung der Anzahl an existenten und paarigen Personen mit angepassten Hochrechnungsfaktoren der Ermittlung der Einwohnerzahl auf Sampling-Point-Ebene plausible Schätzer für die Anzahl an existenten und paarigen Personen auf Gemeinde-

ebene. Die Hochrechnungsfaktoren der Ermittlung der Einwohnerzahl auf Sampling-Point-Ebene werden dabei so angepasst, dass bestimmte gemeindespezifische Randwerte getroffen werden.

Für Gemeinden mit 1 bis 99 Stichprobenanschriften würde eine alleinige Hochrechnung der existenten und paarigen Personen mit den angepassten Hochrechnungsfaktoren der Sampling-Point-Schätzungen allerdings nicht zu zuverlässigen Ergebnissen führen, wobei die Zuverlässigkeit der Resultate der Hochrechnung mit sinkendem Stichprobenumfang abnimmt. Aus diesem Grund wird die gemeindespezifische Schätzung mit den angepassten Hochrechnungsfaktoren aus der Sampling-Point-Schätzung mithilfe des Registerwerts der Gemeinde stabilisiert, indem eine Konvexkombination aus dem Registerwert und dem Hochrechnungsergebnis verwendet wird. Hierbei wird das Hochrechnungsergebnis für größere Stichprobenumfänge stärker als für kleine Stichprobenumfänge berücksichtigt, weil die Verlässlichkeit der Hochrechnung mit wachsendem Stichprobenumfang zunimmt.

In Einzelfällen werden dadurch gemeindespezifische Schätzer für die Anzahl an existenten und paarigen Personen erstellt, die eine oder beide der folgenden Plausibilitätsbedingungen nicht erfüllen:

- › geschätzte Anzahl an paarigen Personen \leq Anzahl an gemeldeten Personen und
- › geschätzte Anzahl an paarigen Personen \leq geschätzte Anzahl an existenten Personen.

In diesen Fällen erfolgt eine algorithmische Anpassung der gemeindespezifischen Schätzer für die Anzahl an existenten und paarigen Personen, die sowohl die Konsistenz der gemeindespezifischen Schätzer zu den Schätzern auf Sampling-Point-Ebene als auch die Erfüllung der beiden aufgeführten Plausibilitätsbedingungen gewährleistet.

↳ Nachträgliche Differenzierung der Einwohnerzahlen von Gemeinden nach demografischen Bevölkerungskohorten

Für jeden Sampling Point werden 36 demografisch differenzierte Kohortenergebnisse geschätzt, und zwar die Kreuzkombination aus neun Altersklassen, deutsch/nicht deutsch, männlich/weiblich.

Die Hochrechnung schätzt sowohl die Anzahl an existenten Personen als auch die Anzahl an paarigen Personen auf Sampling-Point-Ebene. Der Einwohnerzahlbeitrag entspricht dabei der hochgerechneten Anzahl an existenten Personen.

Entsprechend ist nicht jede demografisch differenzierte Kohorte mit Stichprobeninformation hinreichend unterfüttert, sodass hier sogenannte synthetische Schätzverfahren zum Einsatz kommen.

Dabei handelt es sich um ein Verfahren des Herunterbrechens nach dem folgenden Grundprinzip: Zunächst werden verlässliche Schätzungen für Eckwerte auf Basis von Stichprobeninformationen auf übergeordneter Ebene erzeugt sowie Modelle geschätzt, welche einen Zusammenhang zwischen den Stichprobenexistenzen und den Registerexistenzen beschreiben. Mittels dieser Modelle werden die Einwohnerzahlen der Sampling Points auf die einzelnen Kohorten heruntergebrochen.

Das Verfahren kombiniert also stichprobenbasierte Eingangsschätzungen mit Registerauswertungen. Konkret werden die stichprobenbasierten Schätzungen für die Kohorten auf größerer Regionalebene (Kreisebene) sowie größeren Kohorten (fünf Altersklassen und Geschlecht) auf Sampling-Point-Ebene als Randwerte genutzt, die mithilfe der Registerwerte auf feinerer Regionalebene (unterhalb der Kreisebene) untergliedert werden.

Für die Anzahl der existenten Personen in den Kohorten und damit für die Einwohnerzahl der Kohorten erfolgt also stets eine Kombination von hochgerechneten Randwerten aus der Stichprobe und Feinuntergliederung mittels Melderegister.

6.3 Sonderfälle bei der Ermittlung der Einwohnerzahl

Neben den Normalanschriften für die Haushaltebefragung und den Anschriften an Wohnheimen und Gemeinschaftsunterkünften gibt es noch Sonderfallanschriften. An diesen Anschriften findet keine Befragung statt, sondern die Ermittlung der Bevölkerungszahl erfolgt anhand der mehrfachfallgeprüften Melderegisterdaten für die gesamte Anschrift.

Bei den Sonderfallanschriften mit Einrichtungen von Angehörigen ausländischer Streitkräfte sowie Diplomtinnen und Diplomaten werden die mehrfachfallgeprüften Melderegisterdaten abzüglich der freiwillig gemeldeten Personen gezählt, sofern sich an diesen Anschriften noch regulär gemeldete Personen befinden.

Darüber hinaus gibt es beispielsweise noch Meldeanschriften von Seeleuten und Binnenschiffern. Hier werden ebenfalls die mehrfachfallgeprüften Melderegisterdaten ohne primärstatistische Prüfung für die gesamte Anschrift übernommen.

6.4 Wohnungsstatusfestlegung

Ziel der Erhebungen ist es, mögliche Registerfehler – Über- und Untererfassungen (Karteileichen und Fehlbestände) – zu erkennen und anschließend, im Fall der Haushaltebefragung nach Hochrechnung, statistisch zu korrigieren. Bei Fehlbeständen steht die Information, ob eine Person an der erhobenen Anschrift zu zählen ist, ausschließlich aus der Erhebung zur Verfügung. Dagegen ist für die paarigen Personendatensätze im Vorfeld eine „Vorfahrtsregel“ zu definieren, ob die Information zum Wohnungsstatus im weiteren Verfahren aus der Primärerhebung oder aus dem Melderegister zu nutzen ist. Grund dafür ist, dass diese beiden Informationen nicht zwangsläufig übereinstimmen (zum Beispiel Hauptwohnsitz laut Register, Nebenwohnsitz laut Befragung oder umgekehrt). Für den Zensus 2022 wird der Wohnungsstatus bei allen primärstatistischen Personenerhebungen einheitlich aus der Registerinformation zugrunde gelegt. Das bedeutet beispielsweise, dass eine in der Stichprobe festgestellte existente Person, die laut Melderegister an einer bestimmten Anschrift mit Nebenwohnsitz gemeldet ist, beim Zensus an dieser Anschrift nicht in die Einwohnerzahl einfließt – unab-

hängig davon, welchen Wohnungsstatus sie bei der Befragung angibt.

Von der Meldepflicht befreite Personen, etwa in Pflegeheimen, Behindertenheimen oder Krankenhäusern, dürfen ihren Hauptwohnsitz an der Normalanschrift behalten. Sie werden – sofern für sie nach Mehrfachfallprüfung und Wohnungsstatusfestlegung nicht der Hauptwohnsitz an Wohnheimen oder Gemeinschaftsunterkünften festgestellt wurde – auch dort gezählt. Fehlbestände mit Hauptwohnsitz an einer Verwaltungsanschrift (zum Beispiel die eines Trägers von Wohnheimen oder Gemeinschaftsunterkünften), die nicht von der Meldepflicht befreit sind, werden mit ihrem Hauptwohnsitz an Wohnheimen oder Gemeinschaftsunterkünften gezählt, an dem sie sich aufhalten.

7

Fazit

Nach den am 25. Juni 2024 veröffentlichten Ergebnissen des Zensus lebten am 15. Mai 2022 rund 82,7 Millionen Einwohnerinnen und Einwohner in Deutschland. Der Zensus ist darauf ausgelegt, bundesweit vergleichbare Ergebnisse zu einem Stichtag zu liefern. Dabei sollen die Befragten so wenig wie möglich belastet werden. Daher kam beim Zensus 2022, wie schon beim Zensus 2011, ein registergestütztes Verfahren zum Einsatz. Das bedeutet, dass bereits vorhandene Verwaltungsdaten – insbesondere aus den Melderegistern – genutzt wurden. Wo Verwaltungsdaten nicht oder nicht in ausreichendem Umfang oder ausreichender Qualität vorhanden sind, erfolgten zusätzliche Befragungen, so auch wie in diesem Beitrag beschrieben bei der Ermittlung der Einwohnerzahl. 

LITERATURVERZEICHNIS

Boragk, Lisa/Gaedke, Annika/Gemmeke, Charlotte/Meyn, Christoph. [*Die Ermittlung des Berichtskreises am Beispiel der Sonderbereiche im Zensus 2022*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 41 ff.

Bretschi, Corinna/Lorentz, Kai. [*Präzisionsziele für die Ermittlung der Einwohnerzahl bei der Haushaltsstichprobe im Zensus 2021*](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 12 ff.

Burgard, Jan Pablo/Münnich, Ralf/Rupp, Martin. *Qualitätszielfunktionen für stark variierende Gemeindegroßen im Zensus 2021*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 14. Ausgabe 1/2020, Seite 5 ff. [Zugriff am 31. Oktober 2024]. Verfügbar unter: link.springer.com

Dreschmitt, Kai/Pfahl, Miriam. [*Die Personenerhebung im Zensus 2022*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 29 ff.

Klink, Steffen/Lorentz, Kai. [*Auswahlplan und Stichprobenhauptziehung für den Zensus 2022*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2022, Seite 13 ff.

RECHTSGRUNDLAGEN

Bundesmeldegesetz (BMG) vom 3. Mai 2013 (BGBl. I Seite 1084), das zuletzt durch Artikel 3 des Gesetzes vom 19. Juni 2024 (BGBl. 2024 I Nr. 206) geändert worden ist.

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Gesetz über den Auswärtigen Dienst (GAD) vom 30. August 1990 (BGBl. I Seite 1842), das zuletzt durch Artikel 14 Absatz 4 des Gesetzes vom 28. Juni 2021 (BGBl. I Seite 2250) geändert worden ist.

Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009 (BGBl. I Seite 1781).

Melderechtsrahmengesetz (MRRG) vom 16. August 1980 (BGBl. I Seite 1429), neu gefasst durch Bekanntmachung vom 19. April 2002 (BGBl. I Seite 1342), außer Kraft getreten am 1. November 2015 durch das Inkrafttreten des Bundesmeldegesetzes.

Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates vom 9. Juli 2008 über Volks- und Wohnungszählungen (Amtsblatt der EU Nr. L 218, Seite 14).

DIE PERSONENERHEBUNG IM ZENSUS 2022

Kai Dreschmitt, Miriam Pfahl

↳ **Schlüsselwörter:** Primärerhebung – Haushaltsstichprobe – Konzeption – Erhebungsinstrumente – Softwareentwicklung – Schulung

ZUSAMMENFASSUNG

Das registergestützte Zensusverfahren sieht vor, Verwaltungsregister als Datenquellen zu nutzen und diese durch zusätzlich primärstatistisch erhobene Angaben zu korrigieren und zu ergänzen. Die Personenerhebung ist eine tragende Säule des Zensus 2022 und zählt zu den ergänzenden Primärerhebungen. Dieser Beitrag gibt einen Überblick über die zentralen Vorbereitungen: Er beschreibt das Erhebungskonzept und stellt die Erhebungsinstrumente vor. Nach einem Überblick über den Aufbau der zur Personenerhebung genutzten IT-Systeme geht er auf das fachliche und IT-seitige Schulungskonzept ein und zeigt spezielle Herausforderungen auf. Auch die Datenaufbereitung und Plausibilisierung werden näher beleuchtet.

↳ **Keywords:** *primary survey – household sample – conceptual design – survey instruments – software development – training*

ABSTRACT

The register-assisted census procedure involves using administrative registers as data sources and correcting and supplementing the registers with additional primary statistical data. The survey of individuals is a key pillar of the 2022 Census and one of the supplementary primary surveys. This article provides an overview of the central preparations, explains the survey design and presents the survey instruments. After providing an overview of the structure of the IT systems used in the survey of individuals, the article proceeds to describe the approach for statistical and IT training and highlights specific challenges. It also examines data processing and data editing in greater detail.



Kai Dreschmitt

studierte Soziologie und Philosophie und ist wissenschaftlicher Mitarbeiter im Referat „Räumliche Bevölkerungsbewegung, Gebietsgliederungen“ des Statistischen Bundesamtes. Für den Zensus 2022 war er im Rahmen der Personenerhebung für die Betreuung des Erhebungsunterstützungssystems und die Schulung der Nutzenden zuständig.



Miriam Pfahl

studierte Politikwissenschaft, Soziologie und Öffentliches Recht und ist Referentin im Referat „Gebäude und Wohnungen“ des Statistischen Bundesamtes. Für den Zensus 2022 war sie unter anderem für die Konzeption der Online-First-Strategie und die Entwicklung der Online-Fragebogen der Haushaltsstichprobe zuständig.

1

Einleitung

Das in Deutschland sowohl 2011 als auch 2022 angewandte registergestützte Zensusverfahren sieht vor, vornehmlich vorhandene Verwaltungsregister als Datenquellen zu nutzen und diese durch zusätzlich primärstatistisch erhobene Angaben zu korrigieren und zu ergänzen. Dies ist immer dann erforderlich, wenn die geforderten Informationen nicht oder in nicht ausreichender Qualität in den Registern vorliegen. Neben der Gebäude- und Wohnungszählung und der Befragung an Gemeinschaftsunterkünften zählt die Personenerhebung zu den ergänzenden Primärerhebungen des Zensus.

Die Personenerhebung umfasst:

- › die Haushaltebefragung (Stichprobe),
- › die Befragung an Adressen mit Wohnheimen (Vollerhebung) und
- › die Wiederholungsbefragung (Unterstichprobe der Haushaltebefragung).

Während die Gebäude- und Wohnungszählung sowie die Befragungen an Adressen mit Wohnheimen und an Gemeinschaftsunterkünften als Vollerhebungen durchgeführt wurden, handelt es sich bei der Haushaltebefragung sowie der Wiederholungsbefragung¹ jeweils um eine Stichprobenbefragung. Um Informationen für ganz Deutschland zu erhalten, wurden die Befragungsergebnisse mithilfe mathematischer Verfahren auf die gesamte Bevölkerung hochgerechnet.

Für die Haushaltebefragung wurde eine zufällige Stichprobe aus deutschlandweit allen Adressen mit Wohnraum gezogen. Alle Personen, die zum Zensusstichtag am 15. Mai 2022 an den gezogenen Adressen wohnhaft waren, wurden für den Zensus befragt. Dies entspricht etwa 12% der Einwohnerinnen und Einwohner Deutschlands, rund 10,2 Millionen Menschen.

In Gemeinden mit weniger als 10 000 Einwohnerinnen und Einwohnern wurden die soziodemografischen Zusatzmerkmale zur Bildung und Erwerbstätigkeit nur bei einer weiteren Unterstichprobe von etwa 40% der Stichprobenanschriften erhoben.

¹ Auf die Wiederholungsbefragung geht dieser Beitrag nicht weiter ein. Der Fokus liegt auf der Haushaltebefragung auf Stichprobenbasis.

Die zur Personenerhebung zählende Wiederholungsbefragung fand ebenfalls als Unterstichprobe an lediglich 4% der Stichprobenanschriften statt. Im Zuge der Wiederholungsbefragung wurde somit ein kleiner Anteil der Auskunftsspflichtigen aus der Haupterhebung erneut zu einzelnen Merkmalen befragt.

Die Personenerhebung verfolgte zwei Ziele:

1. Durch die sogenannte Existenzfeststellung leistete sie einen wichtigen Beitrag bei der statistischen Korrektur der Melderegister zur Ermittlung der amtlichen Einwohnerzahlen von Bund, Ländern und Gemeinden. Sie schuf damit die Grundlage für die Fortschreibung der Einwohnerzahlen für die Zeit zwischen zwei Volkszählungen (Ziel 1).
2. Zudem ermittelte die Befragung weitere zentrale Strukturmerkmale der Bevölkerung zu den Themen Wohnsituation, Staatsangehörigkeit und Zuwanderung, Bildung und Ausbildung, Erwerbstätigkeit, Arbeitsort, Wirtschaftszweig und Beruf (Ziel 2).

Die nachgelagerte Wiederholungsbefragung diente dazu, die Qualität der ermittelten Einwohnerzahl zu bewerten.

Der Beitrag zeichnet die Stationen der konzeptionellen und technischen Vorbereitungen für die Erhebung 2022 nach. Auf die Durchführungsphase geht er dabei nur am Rande ein, da diese maßgeblich durch die Statistischen Ämter der Länder betreut wurde. Kapitel 2 beleuchtet die Vorbereitungen sowie die Neuerungen im Vergleich zum Zensus 2011. Die Kapitel 3 und 4 beschreiben das Erhebungskonzept und die Erhebungsinstrumente. Nach einem Überblick über die qualitativen und quantitativen Pretests in Kapitel 5 geht Kapitel 6 näher auf den Aufbau der zur Personenerhebung genutzten IT-Systeme ein. Darüber hinaus stellt Kapitel 7 das fachliche und IT-seitige Schulungskonzept vor und beschreibt spezielle Herausforderungen. Der auf die Erhebungsphase folgenden Datenaufbereitung und Plausibilisierung widmet sich das vorletzte Kapitel, bevor ein Fazit hinsichtlich der zentralen Erfolgsfaktoren gezogen wird.

2

Vorbereitungen und Neuerungen

2.1 Nach dem Zensus ist vor dem Zensus

Um die Personenerhebung 2022 vorzubereiten, wurden zunächst die Verfahren und Prozesse des Zensus 2011 umfassend geprüft und bewertet. Die Erfahrungsberichte der Statistischen Ämter der Länder, eine externe Evaluierung im Auftrag des Bundesministeriums des Innern und für Heimat, Empfehlungen des Rats für Sozial- und Wirtschaftsdaten sowie die Auswertung des Fragebogens 2011 lieferten allesamt wertvolle Erkenntnisse und Hinweise für die Neukonzeption der folgenden Zensusrunde. Das Grundmodell für die Personenerhebung 2022 sollte im Wesentlichen dem des Zensus 2011 entsprechen, jedoch mit Optimierungen unter Genauigkeits-, Zeit-, Effizienz- und Kostenaspekten. Neue Datenquellen sowie informationstechnische und methodische Neuerungen waren zu berücksichtigen, sofern diese einen zusätzlichen Nutzen stiften würden und für den Zensus eingesetzt werden könnten.

Zur konzeptionellen Vorbereitung der Personenerhebung 2022 wurde das Teilprojekt Personenerhebung (Konzeption & Aufbereitung) eingerichtet, welches sich aus Mitarbeitenden des Statistischen Bundesamtes und der Statistischen Landesämter zusammensetzte. Die Arbeitsgruppe Haushaltsstichprobe, die neben den Mitgliedern des Teilprojekts aus weiteren Vertreterinnen und Vertretern der Statistischen Ämter der Länder und zusätzlich der Kommunen bestand, nahm die im Teilprojekt erarbeiteten Konzepte ab.

Die Aufgaben des Teilprojekts Personenerhebung waren im Einzelnen:

- › Entwickeln und Bereitstellen eines bundeseinheitlichen Konzepts, um die Personenerhebung durchzuführen
- › Entwickeln und Bereitstellen bundeseinheitlicher Erhebungsinstrumente für die Existenzfeststellung, für die Erhebung der Ziel-1-Merkmale sowie für die Erhebung der Zusatzmerkmale (Ziel 2)

- › Formulieren und Bereitstellen von fachlichen Anforderungen an die IT-Entwicklung, um die Personenerhebung durchzuführen
- › Bereitstellen der betriebsbereiten IT-Anwendungen, um die Personenerhebung durchzuführen und aufzubereiten
- › Erstellen eines Schulungskonzepts und Bereitstellen von Schulungsunterlagen für die Mitarbeitenden der Statistischen Landesämter, der Erhebungsstellen und für die Erhebungsbeauftragten.

2.2 Neuerungen

Zu den wesentlichen Neuerungen im Zensus 2022 zählt unter anderem die zentrale Ansiedlung der IT-Systeme im Statistischen Bundesamt. Hier folgte man den Handlungsempfehlungen der Evaluationsergebnisse des Zensus 2011 und verzichtete auf eine dezentrale Aufteilung der IT-Systeme im Statistischen Verbund¹² zugunsten einer zentralen Lösung (Dittrich, 2019). So wurden insbesondere die Kommunikation zwischen den verschiedenen IT-Schnittstellen optimiert und Effizienzsteigerungen erreicht. Da alle Daten an einer zentralen Stelle gesammelt und weiterverarbeitet wurden, konnten sie zeitnah validiert und auf ihre Richtigkeit überprüft werden – beispielsweise in Bezug auf die sogenannte Erhebungsteilübergreifende Plausibilisierung. Mit dieser erfolgte bereits während der Erhebung ein Abgleich der Erhebungsdaten mit den Angaben aus den Melderegistern (Bretnütz und andere, 2024).

Um die Erhebung zu optimieren und die Erhebungskonzepte zeitgemäß anzupassen, wurde unter anderem der Einsatz mobiler Endgeräte bei den persönlichen Befragungen im Zensus 2022 geprüft. Gleichzeitig legten die Projektbeteiligten eine Online-First-Strategie für die primärstatistischen Erhebungen fest. Wenngleich die Auskunftgebenden auch im Zensus 2022 zwischen verschiedenen Optionen der Meldungsabgabe wählen konnten, stellte die Selbstauskunft über den Online-Fragebogen den präferierten Erhebungsweg¹³ dar. Das

2 Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

3 Bei der Personenerhebung wurde die Online-First-Strategie für die Erhebung der Zusatzmerkmale (Bildung und Erwerbstätigkeit) verfolgt. Für die vorangehende Existenzfeststellung war der Einsatz von Erhebungsbeauftragten vorgesehen (siehe hierzu auch das folgende Kapitel 3).

sollte durch eine Reihe gezielter Maßnahmen erreicht werden (Gaedke und andere, 2024; Freier/Mosel, 2019; Neutze und andere, 2017). Ziel war, bei der Personenerhebung eine Online-Selbstaussfüller-Quote von mehr als 50% zu erreichen. (Hentschke und andere, 2024, hier: Übersicht 3) Einige Bundesländer schufen zudem die Voraussetzungen, die interviewergestützten persönlichen Befragungen mithilfe von Tablets durchzuführen. Papierfragebogen sollten nur ersatzweise eingesetzt werden.

3

Erhebungskonzept

Die Personenerhebung war eine Kombination aus einer Vor-Ort-Befragung durch Interviewerinnen und Interviewer (Erhebungsbeauftragte) sowie – für weitere Fragen – einer sich anschließenden Selbstauskunft durch die Auskunftspflichtigen. Hierzu wurden kleinräumig (in der Regel in den Kommunalverwaltungen auf Kreisebene) Erhebungsstellen eingerichtet, die die Erhebung koordinierten und leiteten. Anders als bei der Gebäude- und Wohnungszählung war bei der Personenerhebung Interviewpersonal erforderlich, da nur die Anschriften festgelegt waren, an denen die Existenzfeststellung und die weiteren Befragungen durchgeführt werden sollten.

In einem ersten Schritt mussten die Interviewerinnen und Interviewer die ausgewählten Anschriften begehen und zunächst erhebungsorganisatorische Angaben erfassen (zum Beispiel Prüfung des Anschriftentyps, Anmerkungen zum Auffinden der Anschrift und zur Anzahl der Haushalte). Dies sollte zum einen sicherstellen, dass die korrekten Anschriften aufgesucht und beispielsweise keine Hinterhäuser, Seiteneingänge oder Ähnliches bei der späteren Befragung übersehen werden. Zum anderen hinterließen die Erhebungsbeauftragten Ankündigungsschreiben mit Informationen zum Zensus sowie eine Terminankündigung für die Interviews. Zum vereinbarten Termin ermittelten sie dann in einer persönlichen Befragung, welche Personen zum Zensusstichtag 15. Mai 2022 in den Haushalten wohnhaft und damit zählungsrelevant waren. Außerdem erhoben sie die für die Existenzfeststellung geforderten Merkmale⁴

4 Familienname, Vorname, Geburtsdatum, Geschlecht sowie an Anschriften mit Wohnheimen zusätzlich Geburtsname, Geburtsort und Geburtsstaat.

sowie weitere Angaben⁵. Für einen Großteil der Auskunftspflichtigen endete die Befragung an dieser Stelle. Lediglich bei einer Untermenge wurden darüber hinaus Informationen zur Bildung und Erwerbstätigkeit erfragt. Diese Angaben konnten die Auskunftspflichtigen selbstständig tätigen, bevorzugt online über den Fragebogen im Internet (Online First), auf Wunsch auch in Papierform oder bei Bedarf durch Interviewerunterstützung – in einigen Bundesländern tabletbasiert, gegebenenfalls auch telefonisch.

Bei Nichterreichbarkeit griff ein detailliertes Verfahren bestehend aus erneutem Kontaktversuch und Erinnerungs- und Mahnverfahren auf postalischem Weg, um alle Personen einer Anschrift zu erreichen.

Dieser kaskadenartig aufgebaute Katalog von Schritten war im Zuge der Existenzfeststellung zwingend und einheitlich zu durchlaufen. So wurde sichergestellt, dass für jede Person an einer Anschrift ihre Existenz beziehungsweise Nicht-Existenz im Hinblick auf die Einwohnerzahl-ermittlung eindeutig festgestellt werden konnte.

4

Erhebungsinstrumente

Bei der Personenerhebung kamen unterschiedliche standardisierte Erhebungsinstrumente zum Einsatz. Sie unterschieden sich insbesondere hinsichtlich der jeweiligen Nutzenden (Erhebungsbeauftragte oder Auskunftspflichtige), der Zielgruppe (Haushaltebefragung oder Erhebung an Wohnheimen) sowie des Merkmalskranzes, der jeweils zu erheben war. [↪ Übersicht 1](#)

4.1 Erhebungsliste für Erhebungsbeauftragte

Bei der Begehung der Erhebungsbezirke⁶ nutzten die Erhebungsbeauftragten eine Erhebungsliste, die sich aus einem Mantelbogen (je Erhebungsbezirk) sowie

5 Staatsangehörigkeit, Familienstand, Wohnungsstatus, gegebenenfalls Hauptwohnung.

6 Ein Erhebungsbezirk entsprach im Regelfall einer bestimmten (Stichproben-)Anschrift, wobei Anschriften mit vielen Haushalten gegebenenfalls in mehrere Erhebungsbezirke aufgeteilt wurden.

Übersicht 1

Erhebungsinstrumente der Personenerhebung

Erhebungsinstrumente für Erhebungsbeauftragte	
Anschriфтenebene	Mantelbogen: je Erhebungsbezirk
Haushaltsebene	Haushaltsbogen: je Haushalt
	Haushaltsbogen Wiederholungsbefragung: je Haushalt
Erhebungsinstrumente für befragte Personen (Fragebogen)	
Personenebene	Fragebogen zur Befragung an Ziel-1-Anschriften
	Fragebogen zur Befragung an Ziel-1- und Ziel-2-Anschriften
	Fragebogen zur Befragung an Ziel-1-Anschriften mit Wohnheimen
	Fragebogen zur Befragung an Ziel-1- und Ziel-2-Anschriften mit Wohnheimen
	Fragebogen zur Wiederholungsbefragung

einem oder mehreren Haushaltsbogen (je Haushalt) zusammensetzte.

Der Mantelbogen diente dazu, Ergebnisse und (Anschriфт-)Befunde⁷ der vor Ort an einem Erhebungsbezirk vorgefundenen Situation zu erfassen. Zudem enthielt er die Information, ob an der betreffenden Anschrift eine Ziel-1- oder eine Ziel-2-Erhebung vorgesehen war.

Im Haushaltsbogen wurden die erhebungsorganisatorischen Angaben zum Haushalt (zum Beispiel angekündigter Befragungstermin oder Auskunftsverweigerung), die allgemeinen Angaben zu den einzelnen Haushalten eines Erhebungsbezirks sowie die erfragten Untersuchungsergebnisse zu den einzelnen Personen erfasst.

Zur Unterstützung der Interviewerinnen und Interviewer standen für den Haushaltsbogen Übersetzungshilfen in 15 Sprachen zur Verfügung (Arabisch, Bulgarisch, Englisch, Französisch, Griechisch, Italienisch, Kroatisch, Persisch, Polnisch, Portugiesisch, Rumänisch, Russisch, Spanisch, Türkisch, Ukrainisch).

4.2 Fragebogen je befragte Person

Insgesamt gab es fünf verschiedene Fragebogen (siehe Übersicht 1): Da alle Haushalte zu den Ziel-1-Merkmalen, aber nur eine Unterstichprobe auch zu den Ziel-2-Merkmalen befragt wurden, wurden an Normalanschriften jeweils zwei unterschiedliche Fragebogen verwendet.

Der Kurzfragebogen deckte nur die Ziel-1-Merkmale ab, der längere Fragebogen sowohl die Ziel-1- als auch die Ziel-2-Merkmale. Der Ziel-1-Fragebogen kam nur zum Einsatz, falls die Erhebungsbeauftragten nicht alle Informationen zu einer Auskunftsperson über den Haushaltsbogen erfassen konnten. Da an Wohnheimanschriften zusätzliche Merkmale zu erheben waren, unterschieden sich die Fragebogen je nach Zielgruppe. Zudem gab es einen separaten Fragebogen für die Wiederholungsbefragung.

Im Gegensatz zu den Mantel- und Haushaltsbogen waren die fünf Fragebogen auch in elektronischer Form als Online-Formulare (IDEV⁸) zur Selbstauskunft hinsichtlich der weiterführenden Fragen (beispielsweise zur Erwerbstätigkeit) verfügbar. Alle Papierfragebogen enthielten zudem den Link zum Online-Formular und Zugangsdaten zu den Online-Fragebogen, um bei Interesse jederzeit eine Online-Teilnahme zu ermöglichen. Ein wesentlicher Vorteil der Online-Fragebogen war, dass diese außer in Deutsch in 15 weiteren Sprachen⁹ bereitgestellt wurden und somit von Personen mit schlechten Deutschkenntnissen in ihren jeweiligen Muttersprachen aufgerufen und ausgefüllt werden konnten (Gaedke und andere, 2024).

7 Die sogenannten Befunde spiegeln das Ergebnis der Begehung wider. Hier konnten unter anderem Auffälligkeiten vermerkt werden, zum Beispiel wenn es sich um eine rein gewerblich genutzte Anschrift handelt, die Anschrift nicht vorgefunden wurde oder eine Baulücke ohne Bewohnerinnen und Bewohner besteht.

8 IDEV steht für Internet Datenerhebung im Verbund und ist ein einheitliches Online-Meldeverfahren, das seit 2005 von den Statistischen Ämtern des Bundes und der Länder angeboten wird.

9 Arabisch, Bulgarisch, Englisch, Französisch, Griechisch, Italienisch, Kroatisch, Persisch, Polnisch, Portugiesisch, Rumänisch, Russisch, Spanisch, Türkisch, Ukrainisch.

5

Pretests

Um die für den Zensus 2022 angepassten Erhebungsverfahren und -instrumente vor ihrem Einsatz zu erproben, wurden im Vorfeld mehrere qualitative sowie ein quantitativer Pretest¹⁰ – jeweils mit unterschiedlichen Zielsetzungen – durchgeführt.

5.1 Qualitative Pretests

Zunächst wurden die für die Personenerhebung erstellten Erhebungsinstrumente im Pretestlabor des Statistischen Bundesamtes getestet. Hierfür hat eine kleine Gruppe Testpersonen (jeweils zwischen 20 und 25 Personen) sowohl die Papier- als auch die Online-Fragebogen unter Beobachtung ausgefüllt und anschließend in einem kognitiven Interview dazu Feedback gegeben. Von Interesse waren dabei vor allem die Verständlichkeit der Fragen, der Antworten und der Hinweistexte, der Filterführung (im Papierfragebogen) sowie der generellen Handhabbarkeit der Erhebungsinstrumente. Durch die Kombination von Beobachtung und Interview konnten potenzielle Stolperfallen erkannt und Nachbesserungen vorgenommen werden. Dies betraf sowohl gestalterische Aspekte als auch Formulierungen, um zum Beispiel relevante Hinweistexte besser wahrzunehmen und die Fragen und Antwortkategorien verständlicher zu machen.

Ebenfalls qualitativ und mit Fokus auf Verständlichkeit und Handhabbarkeit wurden diese Erhebungsinstrumente aus der Perspektive der Erhebungsbeauftragten getestet. Hierfür wurden zum einen Fokusgruppendifkussionen mit erfahrenen Interviewerinnen und Interviewern durchgeführt, zum anderen Interviewsituationen mit Testpersonen nachgestellt, beobachtet und im Anschluss ebenfalls kognitive Interviews durchgeführt. Wenngleich sich im Ergebnis zeigte, dass der Fragebogen grundsätzlich auch für ein Interview mit Erhebungsbeauftragten geeignet ist und die Konzeption als Selbstausfüller keine erkennbaren negativen Auswirkungen hat, konnten dennoch weitere Optimierungspotenziale identifiziert werden.

10 Ein Pretest ist ein Vorab-Testlauf mit wenigen Probandinnen und Probanden, um die Qualität der Erhebungsinstrumente und gegebenenfalls des Erhebungsablaufs zu prüfen.

5.2 Quantitativer Pretest

Neben den unter Laborbedingungen durchgeführten qualitativen Pretests fand im Zeitraum von November 2019 bis März 2020 auch eine praktische Pilotstudie unter realistischen Bedingungen der Haupterhebung im Feld statt. Befragt wurden insgesamt 8 632 Personen. Hierbei wurden nicht nur die Erhebungsinstrumente, sondern die komplette Erhebungsorganisation sowie die nachfolgende Aufbereitung erprobt. Zudem sollten Erkenntnisse zur Nutzung der verschiedenen Erhebungswege und insbesondere der Online-Fragebogen gewonnen werden (Estatico und andere, 2021). Methodik und Abläufe entsprachen dabei weitestmöglich denen der Haupterhebung: Es wurde eine Anschriftenstichprobe gezogen, um die Testpersonen zu gewinnen. Die Erhebungsbeauftragten erfassten zunächst erhebungsorganisatorische Angaben und stellten die Auskunftspflicht fest. Dann erhoben sie mindestens die für die Existenzfeststellung geforderten Informationen in einem persönlichen Interview und händigten denjenigen Personen, die auch für die Befragung der Zusatzmerkmale vorgesehen waren, Zugangsdaten für den Online-Fragebogen aus. Auf Wunsch erhoben sie diese Angaben im persönlichen Interview oder führten eine computergestützte Telefonbefragung durch.¹¹

Nicht nur die Feldarbeit und die Datenanalyse lieferten wichtige Informationen, sondern auch die im Online-Fragebogen protokollierten Paradata zur Nutzung, Handhabbarkeit und Verständlichkeit des elektronischen Fragebogens. Neben Informationen über die genutzte Geräteklasse (PC, Tablet, Smartphone) und das Betriebssystem wurden beispielsweise die Verweildauer je Seite, das Hin- und Herspringen im Fragebogen sowie die Dauer des Ausfüllvorgangs erfasst und ausgewertet. Die optionale Nachbefragung am Ende des Online-Formulars lieferte ebenfalls weitere Erkenntnisse zur Handhabung des Online-Fragebogens (Gaedke und andere, 2024).

Die Auswertungen des quantitativen Pretests haben bestätigt, dass sich sowohl die eingesetzten Erhebungsinstrumente als auch die erhebungsorganisatorischen Abläufe im Feld bewährt haben. Gleichwohl konnte auch hier weiterer Optimierungsbedarf identifiziert werden.

11 Abweichend zur Haupterhebung wurden keine weiteren persönlichen oder schriftlichen Kontaktversuche vorgenommen, da die vollständige Erreichung aller Zielpersonen beim Pretest nicht nötig war.

Zudem wurden wichtige Hinweise für das erfolgreiche Gelingen der Online-First-Strategie, die anstehenden Schulungen des Erhebungspersonals sowie eine zielgerichtete Öffentlichkeitsarbeit für den Zensus 2022 gewonnen.

6

Aufbau der IT-Unterstützung

Für die Personenerhebung wurden zwei IT-Anwendungen (sogenannte Fachanwendungen) maßgeschneidert entwickelt, um die Erhebung zu organisieren und durchzuführen und die Erhebungsdaten aufzubereiten: das Erhebungsunterstützungssystem sowie die Kodierung der Berufsbezeichnungen.

Die gesetzlichen Vorgaben zum Datenschutz und zur IT-Sicherheit erhöhten die Anforderungen sowohl an die zur Ergebniserfassung genutzte Software als auch an die dahinterliegende IT-Infrastruktur (Hardware). Hard- und Software hatten die millionenfachen Datenmeldungen und Nutzereingaben zuverlässig und sicher zu verarbeiten. Um die Zensus-Daten zu schützen wurde eine eigens dafür vorgesehene Sicherheitsumgebung aufgebaut, welche den notwendigen IT-Standards entspricht.

Gleichzeitig war für die Softwareentwicklung die Benutzerfreundlichkeit (Usability) des Produkts und damit eine intuitive Bedienung durch die Mitarbeitenden in den Kommunen und den Statistischen Ämtern der Länder wichtig.

6.1 Erhebungsunterstützungssystem

Das Erhebungsunterstützungssystem wurde als ein bundesländerübergreifendes Verwaltungs- und Steuerungsinstrument konzipiert, um die Aufgaben der Personenerhebung, das heißt die Befragung an Wohnheimen, die Befragung an Gemeinschaftsunterkünften, die Haushal- tebefragung auf Stichprobenbasis sowie die Wiederholungs- befragung, zu organisieren und zu bearbeiten. Zu den zentralen Funktionen zählten:

- › Verwaltung von Erhebungsstellen und Erhebungs- beauftragten,
- › Organisation der Schulungen der Erhebungsbeauf- tragten,

- › automatisierte Einteilung von Anschriften in Erhe- bungsbezirke,
- › Zuweisungen der Erhebungsbezirke zu Erhebungs- beauftragten,
- › Druck von Erhebungsunterlagen,
- › Protokollierung der Übergabe und Entgegennahme von Erhebungsunterlagen an Erhebungsbeauftragte,
- › automatisierter Empfang und Verarbeitung von Online-Meldungen,
- › manuelle Eingabe von Erhebungsergebnissen,
- › Organisation des Mahnwesens,
- › automatisierte Integration der Ergebnisse der Fach- anwendung Kodierung der Berufe,
- › verschiedene Möglichkeiten zur Arbeitsstandkon- trolle im Erhebungsprozess,
- › Ticketsystem für Kommunen und Statistische Landes- ämter, um Auffälligkeiten zu klären und Stichproben- anschriften zu pflegen,
- › verschiedene Exportfunktionalitäten zu Qualitäts- kennzahlen, Arbeitsstandkontrollen, Unterlagen der Erhebungsbeauftragten, Mahnwesen, Belegung und vieles mehr,
- › Validierung der Erhebungsergebnisse und automati- sierte Übertragung an den Referenzdatenbestand.

Dabei bestimmte ein Rollen- und Rechtekonzept mit zwei Hierarchieebenen (Mitarbeitende der Statistischen Landesämter und Mitarbeitende der Erhebungsstel- len), welche Funktionalitäten die jeweiligen Nutzenden sehen und bearbeiten konnten.

Der Statistische Verbund arbeitete die Anforderungen an die Software aus und formulierte sie in einem Las- tenheft. Die nachfolgende Softwareentwicklung durch einen IT-Dienstleister erfolgte inkrementell.¹² Dieses Vorgehen hatte den Vorteil, dass Fehler frühzeitig im Ent- wicklungsprozess erkannt und korrigiert werden konn- ten. Jeder Auslieferung eines Inkrements folgte eine mehrwöchige Qualitätssicherungsphase durch einen

12 Inkrementelle Softwareentwicklung ist ein iterativer Ansatz, bei dem Software in überschaubaren Teilmodulen entwickelt wird und regelmäßig funktionsfähige Zwischenversionen erstellt und an den Auftraggeber ausgeliefert werden. Jedes Inkrement bringt neue Funk- tionen oder Verbesserungen vorangegangener Inkremente mit sich, die sofort getestet und verwendet werden können.

externen Dienstleister mit anschließender Abnahme durch das Teilprojekt Personenerhebung.

In insgesamt elf Austauschterminen mit den künftigen Anwenderinnen und Anwendern wurden die aktuellen Inkremente präsentiert, um Feedback aus der Endnutzendenperspektive zu erhalten und gegebenenfalls Anpassungen vorzunehmen. Diese Termine waren mit Vertreterinnen und Vertretern der Statistischen Landesämter sowie der Kommunen besetzt, um deren Anforderungen und Erfahrungen bei der Programmierung berücksichtigen zu können. In einer mehrwöchigen Testphase vor Inbetriebnahme der Anwendung konnten darüber hinaus wertvolle Hinweise der Anwenderinnen und Anwender der Statistischen Landesämter umgesetzt werden.

6.2 Kodierung der Berufe

Die Fachanwendung Kodierung der Berufe diente der Nachkodierung von Berufsinformationen aus der Personenerhebung. Zentrale Funktionen der Fachanwendung waren die automatisierte und die manuelle Kodierung von Berufsangaben sowie die nachträgliche automatisierte Differenzierung von Meisterabschlüssen.¹³ Analog zur Fachanwendung Erhebungsunterstützungssystem wurde auch die Fachanwendung Kodierung der Berufe iterativ und inkrementell entwickelt.

7

Konzeption und Durchführung der Schulungen

Alle an der Personenerhebung im Zensus 2022 Beteiligten waren für die jeweils anstehenden Aufgaben zu schulen. Dies waren im Wesentlichen drei Gruppen:

- › Mitarbeitende der Statistischen Ämter der Länder in Fach- und Querschnittsbereichen (zum Beispiel Auskunftsdienst),

13 Die Differenzierung von Meisterabschlüssen war notwendig, um die spätere Typisierung der Berufsabschlüsse im internationalen Klassifikationssystem ISCED 2011 (Internationale Standardklassifikation des Bildungswesens) vorzunehmen. Sie erfolgte anhand eines Abgleichs der Angaben aus dem Fragebogen mit einer in der Fachanwendung hinterlegten Liste.

- › die kommunalen Erhebungsstellen, die sich zusammensetzen aus Mitarbeitenden der Kommunalverwaltungen, aber auch eigens für den Zensus neu eingestelltem Personal,
- › die Erhebungsbeauftragten, die zumeist von den Kommunen speziell für die Durchführung der Interviews eingestellt wurden.

Insgesamt wirkten mehr als 100 000 Personen an der Durchführung des Zensus 2022 mit. Diese verfügten über ein sehr heterogenes Vorwissen und sollten die unterschiedlichsten Rollen im Gesamtprozess wahrnehmen. Sie wurden zum einen fachlich (Konzepte, Erhebungsinstrumente, Erhebungsablauf, Rechtsgrundlagen und Datenschutz) und zum anderen technisch (in der Bedienung der IT-Anwendungen) auf die anstehenden Aufgaben vorbereitet.

7.1 Kaskadenprinzip

Um die unterschiedlichen Zielgruppen mit ihren jeweiligen Aufgabenfeldern angemessen zu schulen, war ein kaskadenartiger Wissenstransfer über zum Teil mehrtägige Multiplikatorenschulungen vorgesehen. Die Zuständigkeit für die Durchführung wurde dabei jeweils von einer Stufe zur nächsten weitergegeben und lag nicht bei einer zentralen Stelle. Die Schulungen waren ursprünglich meist in Präsenz geplant, wurden jedoch pandemiebedingt überwiegend online durchgeführt, sodass es möglich war, größere Schulungsgruppen zu bilden.

Zum Start der Schulungskaskade hat das Statistische Bundesamt Multiplikatoren der Statistischen Landesämter in den IT-Fachanwendungen in Form von Online-Schulungen sowie durch das Bereitstellen umfangreicher Materialien geschult.

Bei den IT-Schulungen handelte es sich um praktische Anwendungsschulungen in einer eigenen, geschützten Schulungsumgebung mit Schulungsdaten, in der die fachlichen Abläufe und der Umgang mit der Software erprobt wurden.¹⁴ Die Präsentation der verschiedenen Funktionen wurde dabei jeweils durch thematisch passende Übungen begleitet.

14 Die Schulungsumgebungen standen den Anwenderinnen und Anwendern auch im Nachgang der Schulungen weiterhin zur Verfügung, um beispielsweise neue Mitarbeitende einzuarbeiten.

Die Multiplikatoren der Statistischen Landesämter gaben die Inhalte sowohl an die Kolleginnen und Kollegen im jeweiligen Statistischen Landesamt als auch an die Leitungen der Erhebungsstellen weiter.¹⁵ Letztere wiederum schulten alle weiteren Mitarbeitenden der Erhebungsstellen sowie die Erhebungsbeauftragten in den jeweils für sie relevanten Themen. So war es möglich, innerhalb des vorgegebenen Zeitraums die große Anzahl aller in den Zensus involvierten Personen in den relevanten Bereichen zu schulen.

7.2 Schulungsmaterialien

Ergänzend zu den Schulungen sicherten umfangreiche, zentral erstellte Schulungsmaterialien die Qualität sowie ein bundesweit einheitliches Vorgehen. So konnten die Inhalte auch nach den Schulungen nachgeschlagen werden. Die unterschiedlichen Schulungsmaterialien waren rollenadäquat aufbereitet. Zu ihnen zählen:

- › klassische Handbücher – Nutzungs- und Schulungshandbücher, die als Nachschlagewerk dienten und alle Aufgaben, die vor, während und nach der Erhebung durch die jeweiligen Akteure zu erledigen waren, sowie die Funktionen der Fachanwendungen beschrieben,
- › Präsentationen – anschauliche Darstellung der Schulungsinhalte zur visuellen Unterstützung der Schulungen,
- › Kurzanleitungen für Erhebungsbeauftragte – komprimierte Form der Handbücher; kurze und primär grafische Anleitungen, maximal eine Doppelseite; Überblick über die wichtigsten Punkte/Abläufe,
- › kurze Video-Tutorials zum Erhebungsunterstützungssystem für die Erhebungsstellen – zeigten zentrale Funktionen/Abläufe in der Fachanwendung,
- › Übungsaufgaben zum Erproben der IT-Anwendungen.

15 Die Erhebungsstellen wurden nicht in der Fachanwendung Kodierung der Berufe geschult, da diese ausschließlich von den Statistischen Landesämtern genutzt wurde.

8

Datenaufbereitung

Bevor die im Zuge der Befragungen gewonnenen Erhebungsdaten ausgewertet und veröffentlicht werden konnten, waren verschiedene Aufbereitungsschritte zu durchlaufen.

8.1 Beleglesung

Die Rohdaten wurden zunächst in ein maschinell weiterverarbeitbares Format gebracht: Während die Antworten aus den elektronischen Erhebungsinstrumenten bereits in digitaler Form vorlagen, wurden die auf Papier getätigten Angaben zunächst digitalisiert. Im Rahmen der Beleglesung wurden die Fragebogen eingescannt, die Angaben mithilfe einer speziellen Software ausgelesen und gemäß zuvor definierten Vorgaben kodiert.¹⁶ Die Freitextangaben zum Beruf und zur Tätigkeitsbeschreibung wurden zusätzlich als digitales Bild gespeichert und sofern erforderlich während der manuellen Kodierung der Berufsangaben gesichtet und klassifiziert.

8.2 Kodierung der Berufsangaben

Beim Thema Erwerbstätigkeit wurden die Auskunftspersonen unter anderem nach ihrer Berufsbezeichnung gefragt. Hierbei handelte es sich um eine offen formulierte Frage.

Um die Angaben der Auskunftspflichtigen sowohl in der nationalen als auch in der internationalen Klassifikation der Berufe (Bundesagentur für Arbeit, 2021; International Labour Organization, 2012) berichten zu können, erfolgte eine Kodierung über drei verschiedene Vorgehensweisen:

1. Bereits durch die Auskunftspflichtigen selbst beim Ausfüllen des Onlinefragebogens: Nach Eingabe eines Stichwortes konnte in dem eingebundenen Klassifikationsserver nach der Berufsbezeichnung gesucht und eine zutreffende Angabe aus einer vorgegebenen Liste ausgewählt werden.

16 Für die Beleglesung waren ausschließlich die Fragebogen relevant, die für die Erhebung von Zusatzmerkmalen (Fragebogen HH2) eingesetzt wurden.

2. Durch eine automatische Kodierung, die anhand eines Abgleichs der eingegebenen Stichwörter mit der auf dem Klassifikationsserver hinterlegten Liste mit Berufen erfolgte.
3. Manuell, durch geschulte Mitarbeitende in den Statistischen Landesämtern, ebenfalls anhand der Klassifikationsserver-Liste.¹⁷

Ziel der elektronischen Unterstützung während des Erhebungsprozesses war, die Anzahl der manuell zu kodierenden Fälle im nachgelagerten Prozess der Aufbereitung so gering wie möglich zu halten. Die automatische und die manuelle Kodierung erfolgten in der Fachanwendung Kodierung der Berufe. Insgesamt konnten so 63 % der Fragebogen mit Berufsangaben (etwa 2,05 Millionen) entweder über den elektronischen IDEV-Fragebogen oder über die Fachanwendung Kodierung der Berufe automatisiert kodiert werden. Noch manuell kodiert wurden 37 % der Fragebogen mit Berufsangaben (etwa 1,2 Millionen).¹⁸

8.3 Integration, Plausibilisierung und Imputation

Demografische Angaben zu Personen lagen aus bis zu drei Quellen vor: aus dem Melderegister, der Erhebungsliste und dem Fragebogen. Während der Aufbereitung wurden diese Daten zusammengeführt und auf jeweils eine Angabe reduziert. Hierbei galt die Vorfahrtsregel: Melderegisterangaben vor Angaben aus der Erhebungsliste (Angaben aus den Interviews) vor Antworten aus dem Fragebogen. In einem zweiten Schritt wurden die Angaben zu den Zusatzmerkmalen Bildung und Erwerbstätigkeit auf Vollzähligkeit, Vollständigkeit, korrekte Befüllung und Widersprüchlichkeit geprüft und im Bedarfsfall maschinell bereinigt. Hierbei kam unter anderem die Imputations-Software CANCEIS¹⁹ zum Einsatz, die nach dem Nearest-Neighbour-Prinzip arbeitet. Dabei werden fehlende oder fehlerhafte Werte

-
- 17 Die manuellen Kodierungen wurden darüber hinaus nach einem Vieraugenprinzip qualitätsgesichert.
 - 18 Gründe für den verhältnismäßig hohen Anteil an manuellen Kodierungen liegen in der Qualitätssicherung. Die automatische Kodierung wurde nur bei sehr hoher Übereinstimmung verwendet.
 - 19 Die Software CANCEIS (Canadian Census Edit and Imputation System) wurde 1992 vom kanadischen Statistikamt Statistics Canada/Statistique Canada für die Aufbereitung und Imputation von Zensusdaten entwickelt.

eines Datensatzes (Empfänger) mit vollständigen und plausiblen Angaben eines anderen Datensatzes (Spender) derselben Erhebung ersetzt, sofern Spender- und Empfängerdatensätze sich in den übrigen vorhandenen Angaben überwiegend ähnlich sind (Bretnütz und andere, 2024).

8.4 Typisierung

Abschließend wurden aus dem nunmehr plausiblen Datenmaterial zusätzliche Merkmale abgeleitet, die nicht explizit durch die Datengewinnung bereitgestellt wurden. So errechnet sich beispielsweise das Merkmal Erwerbsstatus aus einer Vielzahl einzelner Erhebungsmerkmale rund um das Thema Erwerbstätigkeit. Diese sogenannten Auswertungsmerkmale wurden anhand der vorliegenden Erhebungsmerkmale (= Basismerkmale) im sogenannten Internen Auswertungssystem zum Zensus 2022²⁰ erstellt (typisiert) und schließlich über die [Zensusdatenbank](#) der interessierten Öffentlichkeit zur Verfügung gestellt.²¹

9

Fazit

Die Personenerhebung bildet eine der zentralen Säulen des registergestützten Zensusverfahrens in Deutschland. Dieser Beitrag zeichnet die Stationen der konzeptionellen und technischen Vorbereitungen für die Erhebung 2022 nach. Obwohl im Zensus 2011 bereits einmal eine Personenerhebung durchgeführt wurde, brachten die Vorbereitungen für den Zensus 2022 große organisatorische und technische Herausforderungen mit sich.

-
- 20 Das Interne Auswertungssystem zum Zensus 2022 ist eine zensus-spezifische Anwendung, um die statistischen Einzeldaten aus den verschiedenen Erhebungsteilen sowie weitere Metadaten des Zensus 2022 langfristig sicher zu halten. Zugleich bietet die Anwendung die Möglichkeit flexibler und individueller Auswertungen einschließlich statistischer Geheimhaltung der Ergebnisse.
 - 21 Bei der [Zensusdatenbank](#) handelt es sich um ein öffentlich zugängliches Recherchetooll, mit dem die Ergebnisse des Zensus 2022 im Detail und in Tabellenform für die Regionalebenen Bund, Länder, Kreise, Gemeinden und Bezirke (Berlin, Hamburg) abgerufen und mit anderen Merkmalen kombiniert werden können. Zu beachten ist, dass die aus der Stichprobenerhebung hochgerechneten Ergebnisse zu Bildung und Erwerbstätigkeit nur für Gemeinden und Gemeindeverbände mit einer Bevölkerungszahl von mindestens 10 000 sowie flächendeckend ab Kreisebene ausgewiesen werden.

Hierbei konnten drei maßgebliche Erfolgsfaktoren identifiziert werden:

1. Ein zentraler Erfolgsfaktor bei der Bewältigung der vielfältigen Aufgaben der Personenerhebung waren die kaskadenartigen Schulungen. Sie haben die bundeseinheitlichen Erhebungsstandards sowie die Nutzungshinweise für die IT-Anwendungen innerhalb kürzester Zeit an die immens große Zahl der Erhebungsbeteiligten weitergegeben.
2. Um die hohen Qualitätsstandards für amtliche Daten zu erreichen, war – neben den bundesweit einheitlichen Erhebungsinstrumenten – vor allem das einheitliche Vorgehen beim Aufsuchen der Auskunftspflichtigen durch die Interviewerinnen und Interviewer entscheidend. Hierzu trugen auch maßgeblich die rollenadäquat aufbereiteten Schulungsunterlagen bei.
3. Nicht zuletzt brachte die zentrale Ansiedlung der IT-Systeme erhebliche Vorteile bei der Datenverarbeitung mit sich. Ebenfalls bewährte sich die inkrementelle Softwareentwicklung, die gewährleistete, dass Fehler frühzeitig erkannt und Anregungen der Nutzenden – auch noch zu einem fortgeschrittenen Zeitpunkt der Entwicklung – umgesetzt wurden. 

LITERATURVERZEICHNIS

Bretnütz, Hella/Estatico, Sabrina/Walther, Sabrina/Hüsgen-Brodhäger, Fabienne/Lange, Kerstin/Kleber, Birgit/Haberzettl, Benjamin. [Aufbereitung und Plausibilisierung der primärstatistischen Erhebungsteile im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 51 ff.

Bundesagentur für Arbeit. *Klassifikation der Berufe – überarbeitete Fassung 2020*. Nürnberg, 2021. [Zugriff am 7. November 2024]. Verfügbar unter: statistik.arbeitsagentur.de

Dittrich, Stefan. [Der registergestützte Zensus 2021. Weiterentwicklung des Zensus 2011 und weitere Schritte hin zu einem registerbasierten Zensus](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. 2019, Seite 5 ff.

Estatico, Sabrina/Tornow, Tobias/Walther, Sabrina. [Generalprobe für den Zensus 2022: Erkenntnisse aus der Pilotstudie zur Haushaltebefragung](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2021, Seite 126 ff.

Freier, Benjamin/Mosel, Juliane. [Online First als Leitgedanke für effiziente Primärerhebungen beim Zensus 2021](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. 2019, Seite 46 ff.

Gaedke, Annika/Pfahl, Miriam/Strohalm, Anna. [Die Online-First-Strategie und die Online-Fragebogen im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 92 ff.

Hentschke, Janine/Tobies, Cara-Aileen/Weber, Susanne/Claus, Dennis. [Kern-Qualitätskennzahlen und Zielwerte im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 106 ff.

International Labour Organization. *International Standard Classification of Occupations 2008 (ISCO-08)*. Genf, 2012. [Zugriff am 7. November 2024]. Verfügbar unter: www.ilo.org

Neutze, Michael/Pfahl, Miriam/Schweinert-Albinus, Stefan. [Das Innovationspotenzial des Zensus 2021](#). In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 37 ff.

DIE ERMITTLUNG DES BERICHTSKREISES AM BEISPIEL DER SONDERBEREICHE IM ZENSUS 2022

Lisa Boragk, Annika Gaedke, Charlotte Gemmeke, Christoph Meyn

↳ **Schlüsselwörter:** Wohnheim – Gemeinschaftsunterkunft – Sonderanschrift – Anschriftenbestand – primärstatistische Vollerhebung

ZUSAMMENFASSUNG

Wie ist der Berichtskreis einer primärstatistischen Erhebung der amtlichen Statistik aufzubauen und zu pflegen, wenn dieser zu Beginn noch vollständig unbekannt ist? Diese Problematik beleuchtet der folgende Artikel am Beispiel der Erhebung an Sonderbereichen des Zensus 2022. Sie startete ohne Datengrundlage und hatte eine Vollerhebung der Personen in Wohnheimen und Gemeinschaftsunterkünften in Deutschland zum Ziel. Anhand der einzelnen Schritte von der Recherche, der Datenerhaltung und -pflege bis zur Qualitätssicherung und Aktualisierung sowie Verknüpfung mit anderen Datenbeständen wird gezeigt, wie der Berichtskreis der Erhebung aufgebaut und aktuell gehalten wurde.

↳ **Keywords:** residential establishment – collective living quarter – special-facility address – stock of addresses – primary statistical survey – complete enumeration

ABSTRACT

What approach should be taken to set up and maintain the group of respondents for a primary survey of official statistics if this group is completely unknown at the beginning? This article explores this challenge using the example of the survey conducted at special facilities as part of the 2022 Census. The survey started without a data basis and its aim was to carry out a complete enumeration of the people in residential establishments and collective living quarters in Germany. The article describes the individual steps involved, ranging from research, data storage and data maintenance to quality assurance, updating and linking with other data sets. On this basis, it explains how the group of survey respondents was set up and kept up to date.

Lisa Boragk

hat Volkswirtschaftslehre studiert und ist Referentin im Referat „Personenerhebungen, Anschriften“ des Statistischen Bundesamtes. Bis 2023 arbeitete sie im Projekt Zensus 2022 an der Weiterentwicklung des Ortsverzeichnisses sowie im Teilprojekt Sonderbereiche. Dort begleitete sie den Produktivbetrieb der IT-Anwendungen für die Durchführung der Sonderbereichserhebung.

Annika Gaedke

hat Soziologie studiert und ist als Referentin im Referat „Gebäude und Wohnungen“ des Statistischen Bundesamtes im Bereich des Einrichtungsregisters tätig. Bis 2023 war sie im Zensus 2022 für die fachliche Konzeption und Begleitung der IT-Entwicklung für die Erhebung an Sonderbereichen zuständig.

Charlotte Gemmeke

ist Politikwissenschaftlerin und leitet das Referat „IT-Querschnittsprozesse und -verfahren, IT-Wahlen“ des Statistischen Bundesamtes. Im Zensus-Referat „Melderegister, Sonderbereiche, Haushalgenerierung“ bereitete sie die Erhebung an Sonderbereichen konzeptionell vor und war hauptsächlich in der Anforderungsanalyse, der Qualitätssicherung und später im Support der Statistischen Ämter der Länder eingebunden.

Christoph Meyn

studierte Soziologie sowie Friedens- und Konfliktforschung und arbeitete von 2017 bis 2023 im Zensus-Referat „Melderegister, Sonderbereiche, Haushalgenerierung“ des Statistischen Bundesamtes. Seit 2020 war er als Teilprojektleiter insbesondere für die Konzeption, Koordination und den technischen und Nutzen-Support der IT-Entwicklung der Sonderbereiche verantwortlich.

1

Einleitung

Dieser Artikel behandelt mit der Sonderbereichserhebung einen spezifischen und in mehrfacher Hinsicht besonderen Erhebungsteil des Großprojekts Zensus 2022. In ihr erfolgte die Erhebung all jener Anschriften, die aus unterschiedlichen Gründen nicht durch das reguläre Verfahren des Zensus 2022 erfasst werden konnten. Welche Gründe konnten das sein und was bedeutete das für die Erhebung? Wie grenzten sich diese Anschriften von anderen ab und welche Herausforderungen gab es, die Struktur solcher Sonderanschriften konzeptionell und erhebungslogistisch zu erfassen? Der Artikel erläutert das Verfahren, um diese Anschriften in Deutschland vollzählig zu ermitteln, ihre Qualität zu sichern und für die weitere statistische Verarbeitung zugänglich zu machen. Das Ziel dieses Prozesses von Berichtskreis aufbau und -pflege war, valide Daten für die Durchführung der Erhebung an Sonderbereichen bereitzustellen.

Das Verfahren des Zensus 2022 war registergestützt – eine Kombination von Informationen aus Verwaltungsdaten und ergänzenden primärstatistischen Erhebungen (Haushaltsstichprobe und Sonderbereichserhebung).

Auf Grundlage der Ergebnisse des Zensus 2011 und des vorausgehenden Zensus tests war davon auszugehen, dass sich eine persönliche Befragung auf Stichprobenbasis nicht für alle Anschriften eignet. Dies galt zum Beispiel für studentische Wohnheime, da sich hier aufgrund der Fluktuation eine große Differenz zwischen den Meldedaten und den real aufgefundenen Personen gezeigt hat. Aus diesem Grund wurden Sonderbereiche, wie bereits im Zensus 2011, auch im Zensus 2022 vollständig und nicht über Stichproben erhoben. An einigen dieser Anschriften wären persönliche Befragungen durch Erhebungsbeauftragte (Interviewerinnen und Interviewer) zudem gar nicht oder nur schwer durchführbar gewesen, sodass die für die Einrichtungen verantwortlichen Stellen stellvertretend Auskunft gaben.¹ An bestimmten Anschriften wurde vollständig auf eine Erhebung verzichtet. Entweder erfolgte an ihnen die Ermittlung der Einwohnerzahl ausschließlich anhand der im Melderegister vorhandenen Angaben oder sie wurden

aus dem Zensusverfahren ausgesteuert. Ausgesteuert wurden zum Beispiel Anschriften von ausländischen Streitkräften oder Diplomatinen und Diplomaten und deren Angehörigen, die entsprechend internationaler Abkommen nicht zur Einwohnerzahl in Deutschland zu zählen sind und daher nicht erhoben wurden.

Diese Beispiele verdeutlichen, wie heterogen Sonderanschriften sind und dass für die Sonderbereichserhebung verschiedene konzeptionelle und erhebungstechnische Probleme zu lösen waren. Eine Herausforderung dabei war, dass es kein deutschlandweites für den Zensus nutzbares Register gab, das die unterschiedlichen Einrichtungsarten abbildet. Daher war die erste Aufgabe der Sonderbereichserhebung, diesen Bestand als Berichtskreis aufzubauen. Gleichzeitig war im Zuge der Berichtskreispflege sicherzustellen, dass der Berichtskreis über den mehrjährigen Verlauf der Vorbereitungs- und Erhebungsphase des Zensus aktuell blieb. Zudem war zu prüfen, dass die geführten Einrichtungen auch tatsächlich den konzeptionellen Festlegungen der Kategorien entsprachen.

Kapitel 2 erläutert, welche Sonderbereichsarten für die Erhebung der Sonderbereiche im Zensus 2022 unterschieden wurden und wie die technische Umsetzung erfolgte. Zur Ermittlung des Berichtskreises einer primärstatistischen Vollerhebung mit Vorrecherche, Vorhebung, Vorbefragung und Aktualisierung der Vorbefragung informiert Kapitel 3. Thema in Kapitel 4 ist die Berichtskreispflege mit der Vorbegehung der Sonderanschriften durch Erhebungsstellen sowie der Begehung und Befragung durch Erhebungsbeauftragte. Die Ergebnisse des Berichtskreis aufbaus schildert Kapitel 5. Das abschließende Kapitel 6 zieht ein Fazit zur Ermittlung des Berichtskreises der Sonderbereichserhebung im Zensus 2022.

1 Beispiele hierfür wären Gefängnisse und Kasernen, aber auch Klöster und Altenpflegeheime.

2

Die Erhebung der Sonderbereiche im Zensus

Die Sonderbereichserhebung wurde in die Befragung an Wohnheimen und in die Befragung an Gemeinschaftsunterkünften unterteilt. Darüber hinaus gab es Sonderfälle, welche nicht zu den zu befragenden Sonderbereichen gehörten, jedoch ebenfalls recherchiert und verwaltet werden mussten. Alle Anschriften, an denen sich Gemeinschaftsunterkünfte, Wohnheime oder Sonderfälle befanden, wurden als Sonderanschriften bezeichnet. Die Grundgesamtheit des Berichtskreises, also die zu befragenden Einheiten der Erhebung von Sonderbereichen, umfasste alle Wohnheime und Gemeinschaftsunterkünfte in Deutschland. Über sie wurden die Personen ermittelt, die an Anschriften mit Wohnheimen oder Gemeinschaftsunterkünften wohnten. Ihre Gesamtzahl festzustellen war das Ziel der Erhebung an Sonderanschriften.

2.1 Welche Sonderbereichsarten wurden unterschieden?

Die Statistischen Ämter der Länder und das Statistische Bundesamt haben – basierend auf der Regelung in § 2 Absatz 3 Zensusgesetz 2022 – verschiedene Arten von Sonderbereichen definiert, abgestimmt und festgelegt.

↳ Übersicht 1

In Deutschland existierte kein Datenbestand über alle Sonderbereiche und -anschriften, der als Grundlage für die Ermittlung der relevanten Grundgesamtheit dienen konnte: Weder lagen die Daten zentral vor, noch existierte eine entsprechende Software zu deren Aufbau, Zusammenführung, Aktualisierung, Befragung sowie Nutzbarmachung für die weiteren Teile des Zensus. Basis für die Durchführung der Erhebung an Anschriften mit Sonderbereichen war jedoch, dass die Daten als Sonderanschriftenbestand zentral vorlagen. Darüber hinaus waren Sonderanschriften zu markieren, die nicht befragt werden sollten oder durften. Auf Basis des Sonderanschriftenbestands sollten Steuerungs- und Klassifizierungsmerkmale auf Anschriftenebene erzeugt werden. Mit diesen wurde später bestimmt, welches

Übersicht 1

Unterscheidung der Sonderbereichsarten im Zensus 2022

Zu befragen	Wohnheime	Studierendenwohnheim
		Arbeitendenwohnheim
		Sonstiges Wohnheim
	Gemeinschaftsunterkünfte	Internat
		Kloster
		Mutter-/Vater-Kind-Einrichtung
		Gemeinschaftsunterkunft für Flüchtlinge
		(Not-)Unterkunft für Wohnungslose
		Sonstige sozialtherapeutische Einrichtung
		Einrichtung für ältere und/oder pflegebedürftige Menschen
		Einrichtung für Menschen mit Behinderung
		Einrichtung für Kinder und Jugendliche
		Krankenhaus, Palliativstation, Stationäres Hospiz, Psychiatrische Klinik und Maßregelvollzugs-einrichtung
		Justizvollzugsanstalt, auch Einrichtung für Abschiebehäftlinge
Kaserne (Bundeswehr) und Gemeinschaftsunterkunft (Polizei, Bundespolizei)		
Nicht zu befragen	Sonderfälle	Nichtsesshafte
		Ausländische Streitkräfte, Diplomaten
		Sonstige Sonderfälle (Seeleute, Binnenschiffer; Frauen-/Männerhaus) ¹

¹ Diese wurden aufgrund des Schutzbedarfs der dort lebenden Personen nicht aktiv recherchiert. Sofern eine solche Anschrift bekannt wurde, wurde sie den sonstigen Sonderfällen zugeordnet, um eine Reidentifizierung zu verhindern.

Verfahren zur Erhebung der Personendaten an der jeweiligen Anschrift anzuwenden und ob die Gebäude- und Wohnungszählung an dieser Anschrift durchzuführen war.

2.2 Technische Umsetzung

Um die im Folgenden beschriebenen Aufgaben zu unterstützen, wurde das IT-System SAPS (System zum Aufbau und zur Pflege des Sonderanschriftenbestands) entwickelt und betrieben. Das System haben sowohl Mitarbeitende der Statistischen Ämter der Länder als auch Mitarbeitende des Statistischen Bundesamtes für unterschiedliche Aufgaben genutzt. Als Teil der Anwendungslandschaft des Zensus 2022 tauschte das SAPS über integrierte Schnittstellen außerdem Daten mit weiteren Systemen aus. Hierbei zu nennen sind vor allem das Steuerungsregister als Teil des Referenzdatenbestands, dem für alle Erhebungsteile des Zensus zentralen Anschriftenbestand, sowie das Erhebungsunterstützungssystem, mit dem die Durchführung der Befragungen der Personenerhebung verwaltet wurde.

Die Integration der drei genannten Systeme und ihre wechselseitige Kommunikation waren dabei für die Prozesse und Ergebnisse der Erhebung von besonderer Relevanz. Ab dem Zeitpunkt der Inbetriebnahme des Referenzdatenbestands war es möglich, die Sonderanschriften im SAPS mit dem vorhandenen Anschriftenbestand des Steuerungsregisters maschinell oder manuell abzugleichen oder neu aufzunehmen. Hierdurch wurden die in den Systemen vorhandenen Anschriften miteinander verknüpft und übergreifend für den Zensus als Sonderanschriften gekennzeichnet. Zu Beginn der Erhebungsphase wurden die im SAPS vorliegenden Informationen zusätzlich an das Erhebungsunterstützungssystem übermittelt, um die Informationen über die recherchierten Sonderanschriften und -bereiche zur Durchführung der Erhebung in dieses zu integrieren. Die drei Systeme standen während der Erhebung in kontinuierlichem Austausch, um die während der Erhebungsphase entstandenen neuen Erkenntnisse zu synchronisieren.

3

Die Ermittlung des Berichtskreises einer primärstatistischen Vollerhebung

An Anschriften mit Sonderbereichen wurden alle dort wohnhaften Personen erhoben. Dies war notwendig, um die zwei wichtigsten Ziele des Zensus verfolgen zu können. Das war zum einen, die Melderegisterdaten zur Ermittlung der Einwohnerzahlen für Deutschland insgesamt sowie für verschiedene regionale Gebiets-einheiten zu korrigieren (Ziel 1). Zum anderen wurden soziodemografische Informationen zur Bevölkerung in Deutschland (Bildung, Erwerbstätigkeit, Beruf) erhoben, die nicht im Melderegister vorliegen (Ziel 2). Die Erhebung in Gemeinschaftsunterkünften verfolgte zum Schutz der dort wohnenden Personen nur Ziel 1. Das bedeutet, dass dort nur die Merkmale der Bewohnerinnen und Bewohner erfragt wurden, die benötigt wurden, um sie eindeutig identifizieren und mit dem Melderegister abgleichen zu können. An einer Unterstichprobe der Wohnheime wurden auch die weiteren Ziel-2-Merkmale erhoben. Der Befragungsmodus unterschied sich zwischen Wohnheimen und Gemeinschaftsunterkünften: In Wohnheimen füllten die dort wohnenden Personen die Fragebogen selbst aus, in Gemeinschaftsunterkünften gaben die Einrichtungsleitungen stellvertretend Auskunft über die Bewohnerinnen und Bewohner.

Für die Erhebung von Sonderanschriften wurde der Berichtskreis der Wohnheime und Gemeinschaftsunterkünfte gepflegt. Der Datenbestand im SAPS wurde zunächst anhand der Datenobjekte **Anschrift** und **Bereich** aufgebaut. Außerdem wurden **Träger** sowie **Ansprechpersonen** der Bereiche aufgenommen, um die Pflege des Datenbestands zu erleichtern. Hierfür nahmen die Statistischen Ämter der Länder zunächst eine Vorrecherche vor, dieser erste Datenbestand wurde anschließend in das SAPS eingepflegt.

In den chronologisch folgenden Arbeitsschritten der Vorerhebung, Vorbefragung und Aktualisierung der Vorbefragung wurde dieser Berichtskreis qualitätsgesichert und aktualisiert. Mit der Anbindung an das Steuerungsregister im Referenzdatenbestand erfolgte ein Abgleich auf Anschriftenebene mit dem dort vorhandenen Anschriftenbestand. Um die Haupterhebung vorzubereiten, wurden die Bereichsinformationen aus dem SAPS

Grafik 1

Vorbereitung und Ablauf der Erhebung der Sonderbereiche im Zensus 2022



an das Erhebungsunterstützungssystem übertragen. Im weiteren Verlauf wurden die Informationen kontinuierlich mithilfe von Schnittstellen in den drei Systemen gepflegt.

Neben den durchzuführenden Arbeitsschritten im SAPS war systemseitig auch im Erhebungsunterstützungssystem die Anlage weiterer Datenobjekte für die Durchführung der Haupterhebung erforderlich. So wurden beispielsweise sogenannte **Erhebungsbezirke** gebildet, dabei entsprach ein Erhebungsbezirk in der Regel einem Sonderbereich. Zusätzlich wurden bei Wohnheimen **Haushalte** und bei Gemeinschaftsunterkünften **Einrichtungen** unterhalb der Erhebungsbezirke angelegt. Innerhalb der Haushalte beziehungsweise der Einrichtungen wurden die erhobenen beziehungsweise gemeldeten **Personen** angelegt.

Die Vorbereitung der Erhebung erforderte verschiedene Schritte, um einen validen Datenbestand aufzubauen.

↘ Grafik 1

3.1 Vorrecherche der Statistischen Ämter der Länder

Die bereits für den Zensus 2011 recherchierten und erhobenen Informationen zu Sonderbereichen standen als Grundlage für die Erhebung nicht mehr zur Verfügung; sie waren aus Datenschutzgründen nach Beendigung der Arbeiten für den Zensus 2011 gelöscht worden. Zu Beginn der Erhebungsvorbereitungen für den Zensus 2022 und als Grundlage für den Aufbau des Sonderanschriftenbestands und die spätere Vorbefragung bei den Einrichtungsleitungen von Sonderbereichen waren daher zunächst potenzielle Einrichtungen durch die Statistischen Ämter der Länder zu identifizieren (Phase 1). Hierfür wurden unterschiedliche Recherchewege gewählt, wie

- › öffentlich zugängliche Quellen (zum Beispiel Internet oder Verzeichnisse),
- › statistikintern zugängliche Daten (beispielsweise die Krankenhausstatistik),
- › Anfragen bei (Landes-)Verwaltungen,
- › Direktansprache größerer Träger sozialer Einrichtungen.

Diese Recherchen umfassten insbesondere die Ermittlung der Einrichtungen, deren Bereichsart, Anschriften, Ansprechpersonen und Träger.

Beim Einpflegen dieser Informationen in das SAPS erfolgten automatisierte Abgleiche, um Dubletten zu vermeiden. Zusätzlich fanden Plausibilitätsprüfungen statt, welche verhinderten, dass unvollständige oder unplausible Angaben in das System aufgenommen wurden. Die im SAPS enthaltenen Informationen konnten über den gesamten Zeitraum der Erhebungsvorbereitung hinweg geändert und aktualisiert werden; somit war es möglich, bei Bedarf stets auf einen aktuellen Grundbestand an Sonderanschriften und zugehörigen Sonderbereichsangaben zuzugreifen.

3.2 Vorerhebung

Nach dem erstmaligen Aufbau des Sonderanschriftenbestands wurde 2019 die sogenannte Vorerhebung durchgeführt. Sie diente dazu, die Qualität zu sichern sowie die recherchierten Informationen zu Sonderbereichen (mit Ausnahme der Bereichsart Kaserne) und Sonderanschriften zu ergänzen und zu korrigieren (Phase 2). Die Statistischen Ämter der Länder führten die Vorerhebung durch, welche hauptsächlich bei den Trägern der Wohnheime und Gemeinschaftsunterkünfte beziehungsweise bei sogenannten selbstverwalteten Einrichtungen ohne Träger stattfand. Dazu erhielten die Träger beziehungsweise Einrichtungen eine Gesamtliste

aller ihnen vom Statistischen Landesamt bis zu diesem Zeitpunkt zugeordneten Bereiche sowie eine Möglichkeit, die Informationen zu den Bereichen und den Kontaktdaten von Ansprechpersonen zu vervollständigen beziehungsweise zu korrigieren. Darüber hinaus wurden die Träger beziehungsweise Einrichtungen gebeten, Einrichtungen zu ergänzen, die den Statistischen Ämtern der Länder bisher nicht bekannt waren. Anschließend konsolidierten die Statistischen Ämter der Länder diese Rückmeldungen und nahmen sie in das SAPS auf. Bei fehlenden Rückmeldungen oder fehlenden relevanten Angaben kontaktierten die Statistischen Ämter der Länder die Träger beziehungsweise Einrichtungen erneut.

Dieser Arbeitsschritt wurde systemseitig unterstützt. Die benötigten Listen konnten mit den den Trägern zugeordneten Bereichen und deren Informationen aus dem SAPS heraus für den Versand exportiert werden; auch wurden die Rückmeldungen der Träger und selbstverwalteten Einrichtungen ins SAPS eingepflegt und die Bearbeitungen abgeschlossen.

Als Ergebnis dieses Arbeitsschrittes lagen bis Ende Juni 2019 die Vorerhebungsdaten für alle Sonderbereiche vollständig, vollzählig und plausibel vor.

3.3 Vorbefragung

Als nächster Schritt wurde im Jahr 2020 die sogenannte Vorbefragung durchgeführt, hauptsächlich bei den Einrichtungsleitungen, vereinzelt auch bei den Trägern der Einrichtungen (Phase 3). Sie diente dazu, wesentliche Fragen zu den Anschriften der Wohnheime oder Gemeinschaftsunterkünfte, die nicht allein aus den Recherarbeiten und den Ergebnissen der Vorerhebung heraus beantwortet werden konnten, zu klären. Diese Informationen waren für die Durchführung der Befragungen zum Zensusstichtag durch die dafür eingesetzten Erhebungsbeauftragten von großer Relevanz. Die Vorbefragung war damit ein weiteres wichtiges Instrument, um die Qualität des Sonderanschriftenbestands zu verbessern. Die Vorbefragung erfolgte über einen Onlinefragebogen². In der Vorbefragung wurde erstmals im Rahmen

² Der Onlinefragebogen der Vorbefragung wurde anhand eines qualitativen Pretests Anfang 2019 getestet. Dieser wurde mit Einrichtungsleitungen oder Zuständigen der Träger unterschiedlicher Einrichtungsarten durchgeführt. Ziel des Pretests war, sowohl die Verständlichkeit des Frageinhalts als auch die Funktionalitäten des Fragebogens zu testen und den Fragebogen im Anschluss entsprechend benutzerfreundlicher gestalten zu können (Gaedke und andere, 2024).

der Berichtskreispflege ein Erinnerungsverfahren für die Auskunftgebenden eingesetzt. Die übermittelten Daten wurden in den Statistischen Ämtern der Länder geprüft und Änderungen in den Sonderanschriftenbestand eingepflegt. So konnten die Angaben zu den Bereichen im Sonderanschriftenbestand aktualisiert beziehungsweise ergänzt werden.

Die Vorbefragung enthielt unter anderem Fragen zu folgenden Merkmalen:

- › Art beziehungsweise Zweck der Einrichtung an der Anschrift,
- › Anzahl der Plätze der Einrichtung,
- › ob und wie sich die Einrichtung über mehrere Anschriften, zum Beispiel Nebenstraßen, Nebeneingänge, Anschriften in weiteren Orten, erstreckt,
- › ob und wo sich an der Anschrift neben der Einrichtung noch weitere Einrichtungen, die nicht in den Zuständigkeitsbereich des Trägers fallen, befinden (gegebenenfalls Gebäudeteil angeben),
- › ob und wo sich an der Anschrift der Einrichtung Wohnungen, die nicht zur Einrichtung gehören (zum Beispiel Hausmeisterwohnung oder Privatwohnungen), befinden,
- › ob geplant ist, die Einrichtung bis Mai 2022 aufzulösen, zu erweitern oder Ähnliches beziehungsweise ob es noch weitere für die Erhebung relevante Informationen/Bemerkungen gibt.

3.4 Aktualisierung der Vorbefragung

Aufgrund der Einschränkungen durch die Corona-Pandemie wurde der für 2021 geplante Zensus auf das Jahr 2022 verschoben. Um die Aktualität der bereits in der Vorbefragung erhobenen Informationen zu sichern und um mögliche Änderungen zu erfassen, wurden die Einrichtungsleitungen oder Träger der Einrichtungen im Jahr 2021 erneut kontaktiert (Phase 4). Änderungen konnten über ein vorbefülltes Onlineformular direkt korrigiert werden. Die Aktualisierung der Vorbefragung wurde im Zeitraum von März bis Ende 2021 durchgeführt.

4

Berichtskreispflege während der Erhebung

Die eigentliche Erhebung startete mit dem Zensusstichtag am 15. Mai 2022. Dies ist gleichzeitig der Zeitpunkt, auf den sich die Angaben durch die Auskunftspflichtigen selbst wie auch die Meldungen durch die Einrichtungsleitungen beziehen. Zu diesem Termin sollte der Sonderanschriftenbestand möglichst aktuell sein. Es war jedoch nicht auszuschließen, dass sich seit der Aktualisierung der Vorbefragung weitere Änderungen an Bereichen ergeben haben. Daher erfolgte durch die Erhebungsstellen, die Statistischen Ämter der Länder und die Erhebungsbeauftragten eine kontinuierliche Berichtskreispflege auch während der laufenden Erhebung.¹³

4.1 Vorbegehung der Sonderanschriften durch Erhebungsstellen

Mit dem Start der Vorbereitungsarbeiten der Haupterhebung sichteten in einem ersten Schritt die Erhebungsstellen die Anschriften mit Sonderbereichen ihres Zuständigkeitsbereiches vorab. Ziele waren, einen Überblick über die Anschriften mit Sonderbereichen zu erhalten, Fehler (zum Beispiel bei Adressen) oder zwischenzeitliche Änderungen zu ermitteln sowie die Vollständigkeit der Sonderbereiche im Zuständigkeitsbereich der jeweiligen Erhebungsstelle zu prüfen. Mitarbeitende der Erhebungsstellen suchten die Anschriften auf und prüften, ob die vorliegenden Angaben (Anschriften, Bereichsart) richtig und vollständig waren. Besonderes Augenmerk lag hierbei darauf, die realen Gegebenheiten an komplexen oder teilweise sehr großen Anschriftenarealen mit den vorliegenden Informationen abzugleichen. Daneben stellten insbesondere sogenannte Mischanschriften Herausforderungen dar: Da sich die Abgrenzung der einzelnen Sonderbereiche im

³ Für die Durchführung des Zensus wurden in den Kommunen Erhebungsstellen eingerichtet. Die Erhebungsstellen kümmerten sich eigenverantwortlich um die Anwerbung, Betreuung, Schulung und Koordination der Erhebungsbeauftragten. Sie bildeten Erhebungsbezirke und teilten die Erhebungsbeauftragten ein. Zudem lag die Kontrolle der Erhebungsunterlagen genauso in ihrem Verantwortungsbereich wie die Kontaktaufnahme mit säumigen Auskunftspflichtigen, sodass am Ende die Erhebungsunterlagen vollständig übermittelt werden konnten.

Datenbestand an konzeptionellen Kriterien des Zensus orientiert, kam es vor, dass an einer Anschrift mehrere Sonderbereiche mit teils unterschiedlicher Bereichsart im Sonderanschriftenbestand geführt wurden. Unter anderem weil sich dies teilweise nicht mit dem Eigenverständnis der Einrichtungen vor Ort deckte, galt es sowohl in diesem als auch in den weiteren Arbeitsschritten diese Abgrenzungen zu prüfen und wenn nötig zu aktualisieren.

Durch diesen Schritt der Berichtskreispflege erhielten die Erhebungsstellen und damit die Erhebungsbeauftragten gegebenenfalls zusätzliche Informationen, die trotz der vorgeschalteten Recherchearbeit der Statistischen Ämter der Länder nicht vorliegen konnten. Diese Informationen standen den Erhebungsbeauftragten für die später folgende Erhebung zur Verfügung.

Wurden Anpassungsbedarfe identifiziert, erfolgte die Dokumentation im Erhebungsunterstützungssystem. Die Meldung wurde an das zuständige Statistische Landesamt kommuniziert, welches nach Prüfung eine Aktualisierung im SAPS oder im Referenzdatenbestand vornahm. Die Änderung wurde im Anschluss den jeweils betroffenen Erhebungsstellen im Erhebungsunterstützungssystem angezeigt. Diese Bereichs- und Anschriftenaktualisierung war fortlaufend während der gesamten Erhebung möglich. Ebenso konnten die Erkenntnisse aus den Folgeschritten (Begehung und Erhebung) dokumentiert und danach im SAPS und im Referenzdatenbestand eingepflegt werden, sodass der Berichtskreis stets aktuell gehalten wurde.

4.2 Begehung durch Erhebungsbeauftragte

Die Erhebungsbeauftragten begingen ab oder bereits kurz vor dem Zensus-Stichtag die ihnen zugewiesenen Erhebungsbezirke. Bei jeder Begehung einer Anschrift führten die Erhebungsbeauftragten je Erhebungsbezirk die Erhebungsunterlagen mit sich und konnten darauf vermerken, wenn sie Abweichungen feststellten (zum Beispiel nicht mehr vorhandene Gemeinschaftsunterkünfte oder abweichende Bereichsart). Zusätzlich verteilten sie während der Begehung die Terminankündigungen für die Durchführung der persönlichen Befragung. Neben dem Hauptzweck der Erhebungsvorbereitung diente die Begehung somit auch der weiteren Berichtskreispflege

zusätzlich zu einer ersten Begehung durch die Erhebungsstellen.

4.3 Befragung durch Erhebungsbeauftragte

Den letzten Schritt der Erhebung stellte die Befragung durch die Erhebungsbeauftragten dar. Dies bildete im gleichen Zug die letzte Möglichkeit, den Berichtskreis zu aktualisieren. Nach Abschluss der Befragung überprüften die Erhebungsbeauftragten ihre Unterlagen und gaben diese in den Erhebungsstellen ab. Dort wurden die Ergebnisse in das Erhebungsunterstützungssystem eingepflegt und festgestellte Anpassungsbedarfe konnten weiterhin über das Erhebungsunterstützungssystem an das zuständige Statistische Landesamt kommuniziert werden. Bis zum Ende der Erhebungsphase 2022 wurden Anpassungen vorgenommen. Der somit entstandene Datenbestand bildet den finalen Berichtskreis der Sonderbereiche zum Stichtag des Zensus 2022 ab.

5

Ergebnisse des Berichtskreisaufbaus

Durch die beschriebenen Schritte war es möglich, sowohl den für die Vollerhebung an Anschriften mit Sonderbereichen benötigten Datenbestand aufzubauen als auch die Qualität der vorliegenden Informationen zu verbessern. Dabei lässt sich schrittweise eine Steigerung der Qualität des Datenbestands erkennen. Während bei der noch eher explorativen Vorerhebung im ersten Schritt etwa 15 % der Sonderbereiche geprüft wurden, stieg dieser Wert bei der konzeptionell und systemseitig deutlich weitergehenden Vorbefragung auf rund 75 %. Zu diesem Ergebnis trugen auch ungefähr 16 000 Neuaufnahmen von Sonderbereichen bei, die erst durch die Vorbefragung entdeckt und eingepflegt wurden.

Die so gewonnenen Angaben konnten unter anderem dafür genutzt werden, die Qualität der Anschrifteninformationen zu erhöhen. Damit war es möglich, bereits beim ersten maschinellen Abgleich der Anschriften aus dem Sonderanschriftenbestand mit den Referenzanschriften aus dem Steuerungsregister eine Trefferquote von 85 % zu erzielen. Durch die anschließenden

manuellen Arbeiten wurden schließlich fast alle Sonderanschriften an das Steuerungsregister angebunden und synchronisiert (99,9 %).

Zum Abschluss der Erhebung lagen für insgesamt 68 002 Sonderbereiche alle für die Erhebung relevanten Merkmale vor, sodass sie für die folgenden Verfahren des Zensus nutzbar waren. Davon waren 8 623 Wohnheime, von denen mit Abstand die meisten (5 518) Studierendenwohnheime waren. Von den insgesamt 53 360 Gemeinschaftsunterkünften waren mehr als 75 % Einrichtungen für ältere und/oder pflegebedürftige Menschen (11 753 Sonderbereiche), Gemeinschaftsunterkünfte für Geflüchtete (11 282 Sonderbereiche), Einrichtungen für Menschen mit Behinderungen (9 012 Sonderbereiche) oder Einrichtungen für Kinder und Jugendliche (8 593 Sonderbereiche). Bezogen auf die Anschriften zeigte sich, dass der Großteil dieser nur jeweils einen Sonderbereich aufwies: Nach Abschluss der Erhebung lagen 58 586 Anschriften im Bestand vor. Bei dem kleineren Teil davon handelte es sich um Mischanschriften (insgesamt 7 124 Anschriften), also Anschriften, an denen verschiedene Kombinationen aus Gemeinschaftsunterkünften und/oder Wohnheimen und/oder regulären Wohnmöglichkeiten vorlagen. Bei dem weitaus größeren Teil handelte es sich um Anschriften, an denen ausschließlich Wohnheime (7 543 Anschriften) oder ausschließlich Gemeinschaftsunterkünfte (43 919 Anschriften) lagen.

Bei der Interpretation der Zahlen ist es wichtig, zwischen Sonderbereichen im Sinne der Erfordernisse für die Vorbereitung und Durchführung des Zensus 2022 und den tatsächlich existierenden Einrichtungen, zum Beispiel Einrichtungen der Altenpflege, zu unterscheiden. Die Operationalisierung der in Kapitel 2 beschriebenen Definitionen der Bereichsarten, sprich deren Anwendung auf die vor Ort vorgefundenen Einrichtungen, diente einem erhebungspraktischen Ziel: Einheiten zu bilden, die voneinander für die Zwecke des Zensus abgrenzbar waren. Insbesondere bedeutete dies, dass Sonderbereiche sich von den vor Ort vorgefundenen Einrichtungen dadurch unterscheiden, dass sie immer auf eine Anschrift beschränkt sind, wohingegen sich die tatsächlichen Einrichtungen auch über mehrere Anschriften erstrecken können. Ebenfalls wurden Einrichtungen, die unterschiedlichen Bereichsarten zuzuordnende Leistungen anboten, als jeweils separate Sonderbereiche gezählt (beispielsweise wurde eine Einrichtung, die Altenpflege

und Betreuung von Menschen mit Behinderung anbietet, als zwei Sonderbereiche erfasst, selbst wenn die Einrichtungsleitung dieselbe war). Es ist daher zu beachten, dass die gewählte Abgrenzung der Bereichsarten nicht unbedingt dem Selbstverständnis der Einrichtungen oder dem allgemeinen Verständnis entsprechen muss.

Der Aufbau des Sonderanschriftenbestands hatte nicht zum Ziel, eine Statistik der Wohnheime und Gemeinschaftsunterkünfte in Deutschland zu erstellen. Im Sinne der Berichtskreispflege diente der Sonderanschriftenbestand vielmehr als Mittel zur Planung und Durchführung der Sonderbereichserhebung, also einer Statistik der in Sonderbereichen lebenden Personen. Daher sind bei der Interpretation der oben genannten Ergebnisse die zensuspezifischen Abgrenzungen zu berücksichtigen.

6

Fazit

Die etwa fünfjährige Zeitspanne, in der der Aufbau und die Berichtskreispflege der Sonderbereiche und der Sonderanschriften im Zensus 2022 stattgefunden hat, verdeutlicht, welchen Veränderungen dieser Datenbestand unterworfen ist. Die verschiedenen Schritte zur Qualitätssicherung der Sonderanschriften und Überprüfungen der vorliegenden Daten durch die Einrichtungsleitungen und Verantwortlichen bei den Trägern der Einrichtungen zeigten, dass sowohl das Hinzukommen neuer Einrichtungen als auch das Wegfallen bereits vorhandener Einrichtungen häufig auftritt – neue Einrichtungen werden gebaut, nicht mehr benötigte geschlossen. Umso wichtiger war es, diese Veränderungen zeitnah in den Datenbestand einzupflegen, um stets einen aktuellen, vollständigen und qualitätsgesicherten Bestand an Sonderbereichen für die unterschiedlichen Zwecke im Zensus zur Verfügung stellen zu können.

Neben dieser Dynamik ergab sich im Rahmen der Sonderbereichserhebung eine weitere bedeutsame Herausforderung, welche generelle Relevanz für eine Berichtskreisermittlung und -pflege hat. Die angelegten Definitionen zur Abgrenzung und Klassifikation der Daten stützen sich immer auf bestimmte konzeptionelle Festlegungen der jeweiligen Erhebung und unterscheiden sich möglicherweise von den tatsächlich existierenden Gegebenheiten. Als Idealtypen sind sie keine abschließenden Definitio-

nen des Berichtskreises, sondern dienen den jeweils konzeptionell vorgesehenen Zwecken. Hierbei handelt es sich um ein generelles Phänomen empirischer Datenerhebungen:

„Sowohl die Definition als auch die Verwendung von Begriffen sind weder eindeutig von einer Theorie ableitbar, noch ergeben sie sich aus der Wirklichkeit: Begriffe in der empirischen Sozialwissenschaft gründen immer auf einer Übereinkunft der Forscher, die in jedem einzelnen Fall darüber befinden müssen, ob diese Begriffe theoretisch sinnvoll und empirisch praktikabel sind.“ (Atteslander, 2000, hier: Seite 44)

Durch die oben beschriebenen Arbeiten wurde ein Gesamtdatenbestand von Sonderbereichen und ihren Anschriften des Zensus 2022 aufgebaut. Damit wurde zudem die Grundlage für ein deutschlandweites Register geschaffen, welches wichtige Informationen zu den unterschiedlichsten Arten von Einrichtungen liefern kann. Im Sinne einer nachhaltigen Datennutzung werden die Daten des Zensus 2022 verwendet, um ein stetiges Einrichtungsregister für einen künftig noch stärker registergestützten Zensus aufzubauen. Das Einrichtungsregister wird auf die Bedarfe dieses dynamischen Berichtskreises auszurichten sein und durch regelmäßig durchzuführende Aktualisierungen des darin liegenden Datenbestands auf die kontinuierlich stattfindenden Veränderungen in der Landschaft der Sonderbereiche stetig angepasst werden. 

LITERATURVERZEICHNIS

Atteslander, Peter. *Methoden der empirischen Sozialforschung*. 9. Auflage. Berlin 2000.

Gaedke, Annika/Pfahl, Miriam/Strohalm, Anna. [Die Online-First-Strategie und die Online-Fragebogen im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 92 ff.

RECHTSGRUNDLAGEN

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

AUFBEREITUNG UND PLAUSIBILISIERUNG DER PRIMÄRSTATISTISCHEN ERHEBUNGSTEILE IM ZENSUS 2022

Hella Bretnütz, Sabrina Estatico, Sabrina Walther, Fabienne Hüsgen-Brodhäger, Kerstin Lange, Birgit Kleber, Benjamin Habertzettl

↳ **Schlüsselwörter:** Datenaufbereitung – Plausibilitätskontrollen – Fehlerbereinigung – Imputation – Personenerhebung – Gebäude- und Wohnungszählung – Gemeinschaftsunterkünfte

ZUSAMMENFASSUNG

Fehlerhafte Datensätze können die Ergebnisse einer Statistik verzerren. Um dies zu vermeiden, fanden im Zensus 2022 verschiedene Aufbereitungs-, Plausibilisierungs- und Imputationsschritte statt. Für jeden Erhebungsteil des Zensus wurde dabei das beste Set an Plausibilisierungsmethoden genutzt, um eine valide Datenbasis für Auswertungen zu schaffen. Der Beitrag beschreibt die einzelnen Aufbereitungs- und Plausibilisierungsschritte für die unterschiedlichen Erhebungsteile des Zensus 2022 mit den jeweiligen Besonderheiten.

↳ **Keywords:** *data processing – plausibility checks – error elimination – imputation – survey of individuals – census of buildings and housing – collective living quarters*

ABSTRACT

Incorrect data sets may bias statistical results. To avoid this, various processing, editing and imputation steps were carried out in the 2022 Census. The best set of methods for plausibility checking was used for each survey component of the census in order to create a valid data basis for evaluations. This article describes the individual steps of processing and editing the different 2022 Census components and looks at special aspects of each component.

Hella Bretnütz

war beim Zensus 2022 bei der Gebäude- und Wohnungszählung mit der Datenaufbereitung und Imputation betraut.

Sabrina Estatico

war beim Zensus 2022 im Teilprojekt „Personenerhebung (Konzeption und Aufbereitung)“ unter anderem für die Plausibilisierung und Aufbereitung der Erhebungsdaten aus der Personenerhebung verantwortlich.

Sabrina Walther

war in der Personenerhebung des Zensus 2022 tätig und betreute hier unter anderem die Plausibilisierung und Imputation der Daten.

Fabienne Hüsgen-Brodhäger

war für den Zensus 2022 schwerpunktmäßig für die fachlichen Vorgaben und die Begleitung der technischen Umsetzung der erhebungsteilübergreifenden Plausibilisierung zuständig.

Kerstin Lange

war beim Zensus 2022 mit der Methodik und Umsetzung der Imputation betraut.

Birgit Kleber

betreut im Statistischen Bundesamt für den Zensus die Themen Imputation und Geheimhaltung.

Benjamin Habertzettl

war im Zensus 2022 im Bereich „Melderegister und Sonderbereiche“ für die Umsetzung der Fehlerprüfungen von Melderegister-Dateneingängen und der IDEV-Erhebung an Gemeinschaftsunterkünften zuständig.

Alle Autorinnen und der Autor dieses Artikels sind im Statistischen Bundesamt tätig.

1

Einleitung

Innerhalb einzelner Erhebungsteile und auch zwischen den verschiedenen Erhebungsteilen kam es im Zensus 2022 zu unplausiblen (fehlerhaften) oder fehlenden Werten. Unplausible oder fehlende Werte stellen vor allem bei der Interpretation der Ergebnisse eine Herausforderung dar, denn sie können die Ergebnisse verzerren (Preising und andere, 2021). Geeignete Aufbereitungsschritte sollen diesem Umstand Rechnung tragen und Ergebnisverzerrungen vermeiden. Neben der Aufbereitung der Daten (Kapitel 2) finden Plausibilisierungen und Imputationen in den Datensätzen (Kapitel 3) statt. Je nach Erhebungsteil des Zensus kommen dabei zum Teil unterschiedliche Aufbereitungsschritte zum Einsatz.

Der Zensus 2022 wurde als registergestützte Erhebung durchgeführt. Alle verwendeten Quellen waren hierbei in sich selbst und auch untereinander zu plausibilisieren. Neben dem Referenzdatenbestand, der alle Registerinformationen (insbesondere die Melderegisterdaten) und Primärstatistiken zusammenfasst, schloss der Zensus 2022 drei Primärstatistiken ein:

- › die Gebäude- und Wohnungszählung,
- › die Haushaltebefragung auf Stichprobenbasis (auch Haushaltsstichprobe genannt) und die Befragung an Wohnheimen (zusammengefasst unter dem Begriff Personenerhebung) sowie
- › die Befragung an Gemeinschaftsunterkünften (an Adressen mit sogenannten Sonderbereichen).

Der Beitrag beschreibt die vorgenommenen Plausibilisierungsarbeiten und nimmt dabei Bezug auf die zeitliche Zuordnung:

- › **Während der Erhebungsphase:**
Zwischen den Erhebungsteilen fanden erhebungsteilübergreifende Plausibilisierungen (Abschnitt 2.1) statt sowie verschiedene Aufbereitungsschritte und Plausibilisierungen innerhalb der Primärerhebungen (Abschnitt 2.2).
- › **Nach Abschluss der Erhebungsphase:**
Die Primärerhebungen wurden final plausibilisiert und fehlende oder unplausible Werte mittels eines Nächste-Nachbarn-Ansatzes imputiert (Kapitel 3).

Ein kurzes Fazit beschließt den Artikel.

2

Aufbereitung während der Erhebungsphase

2.1 Referenzdatenbestand

Der Referenzdatenbestand diente als zentrale Datenbank im Zensus 2022 dazu, die Erhebungen zu steuern und zu organisieren. Er enthielt Daten auf Personenebene (Personenbestand) und auf Adressenebene (Adressenbestand, Steuerregister). Die Ergebnisse aus den verschiedenen Erhebungsteilen liefen im Referenzdatenbestand zusammen. Dadurch konnte dieser genutzt werden, um durch automatisierte Abgleiche zwischen den Daten aus den verschiedenen Erhebungsteilen die Konsistenz des Gesamtdatenbestands des Zensus 2022 zu verbessern. Die Abgleiche erfolgten bereits im Aufbereitungsprozess und werden im Folgenden unter dem Begriff „erhebungsteilübergreifende Plausibilisierung“ (EÜPL) zusammengefasst. Zu unterscheiden sind hierbei die sogenannte EÜPL Existenzen während der Erhebungsphase und die Sonder-EÜPL (Abschnitt 3.1) zur Qualitätsverbesserung nach der Erhebungsphase.

Bei der EÜPL Existenzen handelte es sich um einen wichtigen methodischen Bestandteil des Zensus 2022 im Hinblick auf die korrekte Ermittlung der Einwohnerzahl. Die EÜPL Existenzen war relevant für Adressen, die im Zuge der Personenerhebung erfasst wurden. Sie sollte Unplausibilitäten zwischen den Angaben aus den Lieferungen der Melderegister (MR) und dem Ergebnis der primärstatistischen Erhebungen zur Existenzfeststellung auf Personenebene aufdecken sowie bei Bedarf bereinigen. Abweichungen zwischen Melderegister und einer Existenzfeststellung auf Personenebene an einer Adresse sollten überprüft werden, um mögliche Fehler bei der Erfassung der Personen oder der Abgrenzung der Adresse im Steuerregister auszuschließen. Den wesentlichen Input für die EÜPL Existenzen bildeten die gemeldeten Personen laut der aktuell vorliegenden MR-Lieferung. Zudem wurde das Ergebnis der Existenzfeststellung aus der Personenerhebung in Form von Strukturmerkmalen auf Adressenebene (Anzahl der vorläufigen Karteileichen, vorläufige Fehlbestände

und vorläufigen paarig existenten Personen)¹ herangezogen. Die EÜPL Existenzen wurde an einer Anschrift im Referenzdatenbestand automatisch gestartet, sobald die Integration des Rücklaufs aus der Personenerhebung abgeschlossen war.

Um potenziell unplausible Anschriften zu ermitteln, wurden verschiedene Prüfkategorien definiert, in die die Strukturmerkmale und bei Bedarf zusätzliche Informationen zum Beispiel aus der Gebäude- und Wohnungszählung eingingen. Beispielsweise sollten mit der Prüfkategorie „Deutlich mehr existente Personen, als laut MR zu erwarten“ Anschriften aufgedeckt werden, bei denen in der Erhebung möglicherweise nicht nur die jeweilige MR-Anschrift, sondern auch ein Nachbargebäude fälschlicherweise mit erhoben wurde. Die Prüfkategorie „Keine paarigen Datensätze und nur Karteileichen an einer Anschrift“ sollte dagegen Anschriften aufdecken, bei denen gar keine Existenzen in der Personenerhebung festgestellt wurden, weil die Anschrift gar nicht erhoben wurde, sondern stattdessen der Anschriftenbefund „Baulücke“ oder „komplett gewerblich genutzte Anschrift“ (fälschlicherweise) vergeben wurde. Um festzulegen, ab wann eine Anschrift als unplausibel gilt, wurden bundeseinheitliche Schwellenwerte definiert. Dabei wurden die Prüfbedingungen und Schwellenwerte leicht variiert, je nachdem, ob es sich um eine Stichprobenanschrift oder eine Anschrift mit Sonderbereichen (Wohnheime, Gemeinschaftsunterkünfte) handelte. Wurde bei einer Prüfung der Schwellenwert überschritten, konnten Mitarbeitende der Statistischen Ämter der Länder manuelle Prüfungen vornehmen. Diese Prüfungen führten entweder zur Auflösung der Unplausibilität oder gegebenenfalls zu einer erneuten Begehung der Anschrift.

2.2 Gebäude- und Wohnungszählung

Die Gebäude- und Wohnungszählung wurde als Vollerhebung aller Gebäude mit Wohnraum und darin befindlicher Wohnungen sowie bewohnten Unterkünften durchgeführt. Zu jeder zählungsrelevanten Einheit

1 Unter Karteileichen sind Personen zu verstehen, die laut Melderegister in einer Gemeinde leben, zum Stichtag der Erhebung aber nicht mehr an der im Melderegister geführten Anschrift wohnen. Personen, die an einer bestimmten Anschrift wohnen, jedoch nicht im Melderegister mit dieser Anschrift geführt werden, werden als Fehlbestände bezeichnet. Paarig existente Personen wohnen an der im Melderegister geführten Anschrift.

(Gebäude und/oder Wohnung) wurden die betreffenden Auskunftspflichtigen (in der Regel Eigentümerin, Eigentümer oder Verwaltung) über Verwaltungsdatenquellen ermittelt und angeschrieben. Dabei wurden zunächst Zugangsdaten für den Online-Fragebogen versendet und kein Papierfragebogen. Insgesamt betrug die Online-Quote bei der Gebäude- und Wohnungszählung 82 %. Einer der Vorteile der Online-Datenerhebung ist die mögliche Ad-hoc-Plausibilisierung bereits während der Eingabe der Daten durch die Auskunftspflichtigen (Freier/Mosel, 2019).

Sowohl die online erhobenen Daten als auch die Daten aus der Belegung von Papierfragebogen wurden in standardisierter Form in die automatisierte Datenaufbereitung der Gebäude- und Wohnungszählung übermittelt. Dieser Prozess erfolgte kontinuierlich während der Datenerhebungsphase. Sobald die Daten aller Auskunftspflichtigen eines Gebäudes vorlagen, wurden die Daten für dieses Gebäude einschließlich aller dazugehörigen Wohnungen maschinell aufbereitet. Am Ende der Erhebungsphase erfolgte die maschinelle Aufbereitung ein letztes Mal auch für die Daten aller derjenigen Gebäude, zu denen nicht alle erwarteten Antworten eingegangen waren.

Eine besondere Herausforderung der Gebäude- und Wohnungszählung waren Gebäude mit mehreren Eigentümerinnen und Eigentümern, und zwar insbesondere dann, wenn es sich um Gebäude mit Eigentumswohnungen gehandelt hat. Haben diese Eigentümerinnen beziehungsweise Eigentümer widersprüchliche Angaben zu den Gebäudemerkmalen² gemacht, war bei der Aufbereitung zu entscheiden, welche Angabe zu übernehmen war. Die von der jeweiligen Eigentümerin oder dem jeweiligen Eigentümer berichteten Angaben zu den einzelnen Wohnungen wurden immer als Ganzes übernommen. Um mit unterschiedlichen Angaben bei den Gebäudemerkmalen umzugehen, gab es eine Hierarchie nach Art der Eigentümer, zum Beispiel hatten Wohnungsunternehmen Vorrang gegenüber Privatpersonen. Gab es unterschiedliche Angaben in einer Gruppe von Eigentümerinnen und Eigentümern, wurde die meistgenannte Angabe übernommen.

2 Gebäudemerkmalen sind „Zahl der Wohnungen im Gebäude“, „Gebäudeart“ und „Gebäudetyp“, „Eigentumsverhältnisse“, „Baujahr“, „Heizung“ und „Energieträger“.

Bei widersprüchlichen Angaben zum Merkmal „Anzahl der Wohnungen im Gebäude“ wurde vor allem die Zahl der gelieferten Wohnungsdatensätze berücksichtigt.

Es kam vor, dass für einzelne Wohnungen eine Datenmeldung mehrfach abgegeben wurde, etwa weil eine Eigentümerin oder ein Eigentümer für ein und dieselbe Wohnung mehrfach gemeldet hatte oder weil mehrere Anschreiben für dieselbe Wohnung ergingen. Diese sogenannten Wohnungsdubletten zu erkennen und auszusortieren, war bei der Aufbereitung ebenfalls wichtig. Ergab die maschinelle Aufbereitung aufgrund der hinterlegten Regeln, dass die Anzahl der Wohnungen im Gebäude geringer war als die Anzahl der gelieferten Wohnungsdatensätze, wurden im nächsten Schritt die Wohnungsdubletten aussortiert. Um Wohnungsdubletten zu identifizieren, erfolgte unter anderem ein Vergleich der Bewohnernamen miteinander. Kamen Bewohnernamen mehrfach vor, wurde nur eine der Wohnungen übernommen.

3

Aufbereitung nach der Erhebungsphase

3.1 Referenzdatenbestand

Neben der erhebungsteilübergreifenden Plausibilisierung (EÜPL) Existenzen (Abschnitt 2.1) fand im Referenzdatenbestand noch eine Sonder-EÜPL nach Abschluss der Erhebungsphase statt, um die Qualität zu verbessern. Diese Überprüfung sollte gezielt noch einmal auffällige Klumpungen von (vorläufigen) Karteileichen im Zuständigkeitsbereich einzelner Erhebungsstellen kontrollieren. Bei einem Verdacht auf fehlerhafte Erhebung einer Anschrift konnten die Statistischen Landesämter diese prüfen und eine sogenannte Befundsetzung auf Anschriftenebene (Befund „Abgrenzungsproblem“ für Stichprobenanschriften beziehungsweise Befund „Anschrift konnte nicht abgeschlossen werden“ für Anschriften mit Sonderbereichen) durch das Statistische Bundesamt veranlassen. Eine nachträgliche Überprüfung und Übermittlung von Existenzen war aufgrund des großen zeitlichen Abstands zum Stichtag nicht sinnvoll. Diese Korrektur durch Befundsetzung (Befundkorrektur) hatte für Stichprobenanschriften eine Aussteuerung aus

der Hochrechnung zur Ermittlung von Über-/Untererfassung zur Folge. Für Anschriften mit Sonderbereichen wurden durch die Befundkorrektur Personen an diesen Anschriften gemäß dem Ergebnis der Mehrfachfallprüfung gezählt. Die Befundkorrektur führte dazu, dass sich die Nettostichprobe verkleinerte, wodurch sich der Standardfehler geringfügig erhöhte. Dieser Nachteil wog jedoch geringer als die Berücksichtigung von Anschriften mit festgestellten Ungenauigkeiten bei der Datenerhebung. Im Ergebnis wurden bei diesen Anschriften die Angaben der Melderegister ohne primärstatistische Korrektur für die Einwohnerzahlermittlung übernommen.

3.2 Allgemeine Aspekte bei Primärstatistiken

Ziel der Plausibilisierung und Imputation der Zensusdaten war es, wie bei anderen Primärstatistiken auch, plausible, valide und vollständige Datensätze für die nachgelagerten Aufbereitungsschritte bereitzustellen. Die Kernziele dieser Aufbereitungsphase sind:

1. Identifikation und Kennzeichnung von Fehlern in den Daten

Die Zensusdaten wurden auf folgende Fehler geprüft und entsprechend gekennzeichnet:

- › **Unzulässig fehlende Angabe:** Es liegen keine Daten vor, obwohl die Frage hätte beantwortet werden müssen (Item Nonresponse).
- › **Filterfehler:** Es liegen Daten vor, obwohl die Frage aufgrund der Filterführung nicht hätte beantwortet werden dürfen.
- › **Unzulässige Mehrfachangabe:** Es wurde mehr als eine Antwortmöglichkeit bei einer Frage ausgewählt, obwohl dies nicht zulässig war.
- › **Strukturunplausibilität:** Die Angabe entspricht nicht dem zulässigen Format oder liegt außerhalb des zulässigen Wertebereichs.
- › **Interunplausibilität:** Die Angabe ist widersprüchlich zu einer anderen Angabe dieser Befragungseinheit.

2. Fehlerbereinigungen

Zur Behebung wurden zwei Verfahren verwendet:

- › die deterministische Imputation (Abschnitte 3.3 und 3.4) und
- › das Hot-Deck-Nächste-Nachbarn-Verfahren (Abschnitt 3.6).

Während im Zensus 2011 der primäre Erhebungsweg die Befragung durch Interviewerinnen und Interviewer beziehungsweise die Befragung mit Papierfragebogen war, setzte der Zensus 2022 auf die Online-First-Strategie (Freier/Mosel, 2019; Gaedke und andere, 2024). Das erwies sich nicht zuletzt hinsichtlich der Datenqualität als ausgesprochen zielführend, da einige Plausibilitätsprüfungen bereits im Online-Erhebungssystem implementiert werden und erste Fehlerprüfungen und -bereinigungen schon bei der Datenerhebung erfolgen konnten.

3.3 Plausibilisierung der Gebäude- und Wohnungszählung

Die Plausibilisierung der Gebäude- und Wohnungszählung wurde im ersten Schritt als maschinelles Verfahren durchgeführt. Die zu plausibilisierende Einheit war das Gebäude einschließlich aller darin befindlichen Wohnungen. Stattgefunden hat die Plausibilisierung, sobald die Gebäudedaten eines Gebäudes aufbereitet worden waren.

Geprüft wurde auf Vollständigkeit, unzulässige Mehrfachcodierungen, Strukturplausibilität und logische Zusammenhänge zwischen den Merkmalen (Interplausibilität). Für jede Unplausibilität wurde ein Fehlerschlüssel gesetzt. In einigen Fällen konnte eine deterministische Fehlerkorrektur erfolgen. Wenn zum Beispiel bei einer Wohnung die Wohnfläche gefehlt hat, aber eine Raumanzahl angegeben war, wurde geprüft, ob es im Gebäude eine andere Wohnung mit der gleichen Anzahl der Räume gab. War dies der Fall, wurde die Fläche dieser Wohnung übernommen. Alle deterministischen Korrekturen wurden ebenfalls durch Fehlerschlüssel gekennzeichnet.

Mit Fehlerschlüsseln gekennzeichnete Datensätze konnten von den Mitarbeiterinnen und Mitarbeitern der Statistischen Landesämter bei Bedarf im zweiten Schritt

manuell geprüft und, wenn möglich, korrigiert werden. Besonders wichtig war dies bei dem Fehlerschlüssel, der eine Abweichung vom Gebäudemerkmal „Anzahl der Wohnungen im Gebäude“ und den tatsächlich vorhandenen Wohnungsdatensätzen aufzeigte.

Zusätzlich gab es eine Prüfung der erhebungsteilübergreifenden Plausibilität zwischen der Gebäude- und Wohnungszählung und dem Melderegister beziehungsweise der Personenerhebung. Diese fand beispielsweise statt, wenn an einer Anschrift mehr bewohnte Wohnungen in der Gebäude- und Wohnungszählung erhoben worden waren als Haushalte von der Personenerhebung. Bei der erhebungsteilübergreifenden Plausibilität wurden dabei lediglich sogenannte Prüffälle erstellt, die dazu dienten, auf mögliche Fehler hinzuweisen und gegebenenfalls einzelne Korrekturen vorzunehmen.

3.4 Plausibilisierung der Personenerhebung

Die Plausibilisierung (PL) der Daten aus der Personenerhebung wurde als maschinelles Verfahren konzipiert. Im ersten Schritt galt es zu berücksichtigen, dass einige Personenangaben aus bis zu drei Quellen vorliegen konnten:

1. Sogenannte elektronische Erhebungsliste der Personenerhebung, das heißt Kerndaten zur Person, die im Zuge der Existenzfeststellung durch die Interviewerinnen und Interviewer persönlich aufgenommen wurden,
2. Fragebogen der Personenerhebung und/oder
3. Melderegister.

Aus dem vorliegenden Datenmaterial wurde zunächst die für die weitere Datenaufbereitung herangezogene Datenquelle bestimmt. Lagen für Merkmale mehrere Datenquellen vor, kam folgende Vorfahrtsregel zur Anwendung: Melderegisterangaben³ vor Angaben aus den Interviews (elektronische Erhebungsliste) vor Antworten aus dem Fragebogen.

³ Gemäß § 11 Absatz 4 Zensusgesetz 2022 umfasst die Feststellung nach Absatz 1 (Haushaltebefragung auf Stichprobenbasis) nicht die Berichtigung der aus den Melderegistern übernommenen Daten zur Person. Die Melderegisterdaten waren somit – sofern vorhanden und plausibel – vorrangig zu verwenden. Eine Vorfahrt der Angaben aus der Erhebungsliste folgte hingegen auf Basis einer fachlichen Bewertung.

Des Weiteren wurden in der Personenerhebung unterschiedliche Fragebogen eingesetzt, die sich hinsichtlich des Merkmalskranzes und des Fragebogaufbaus unterschieden. Die Kurzbefragungen (Haushalte und Wohnheime) dienten ausschließlich dazu, Merkmale zur Ermittlung der Einwohnerzahlen zu erheben. Sie unterschieden sich dahingehend voneinander, dass die Befragung an Wohnheimen über die Haushalbefragung weitergehende Informationen wie etwa zum Geburtsnamen und zum Geburtsstaat einer Person erfasste. Mit den Zusatzbefragungen wurden weitere Merkmale – etwa zur Bildung und zur Erwerbstätigkeit – erhoben, die nicht (ausreichend) aus Verwaltungsregistern gewonnen werden konnten. Die meisten Fragen zur Bildung und Erwerbstätigkeit richteten sich nur an Personen im Alter von 15 Jahren und älter. Die Haushalbefragung und die Befragung an Wohnheimen unterschieden sich hierbei nicht.

Um dies bei den Fehlerprüfungen und -bereinigungen entsprechend berücksichtigen zu können, wurde das vorliegende Datenmaterial in die nachfolgenden Teilbestände (PL-Klassen) unterteilt:

1. Kurzbefragung an Haushalten
2. Kurzbefragung an Wohnheimen
3. Haushalbefragung und Befragung an Wohnheimen für Personen unter 15 Jahren (Zusatzbefragung)
4. Haushalbefragung und Befragung an Wohnheimen für Personen im Alter von 15 Jahren und älter (Zusatzbefragung)

Anschließend erfolgten nach PL-Klassen getrennt die Prüfungen der Daten auf Vollständigkeit, korrekte Beachtung der Filterführung, auf unzulässige Mehrfachangaben, auf Strukturplausibilität sowie auf Interplausibilität. Identifizierte Fehler wurden gekennzeichnet und gegebenenfalls bereits deterministisch bereinigt. Die deterministische Bereinigung wurde nur durchgeführt, wenn eine eindeutige Beziehung zwischen einem unplausiblen beziehungsweise fehlenden Merkmal und einem oder mehreren plausiblen Merkmalen vorlag. Wenn beispielsweise der Familienstand einer Person laut Melderegister „ledig“ ist, die Person im Fragebogen aber angegeben hat, dass sie mit einer Partnerin beziehungsweise einem Partner zusammenwohnt und mit dieser/diesem verheiratet ist, liegt eine Interplausibilität zwischen den vorliegenden Informationen

vor. In solchen Fällen wurde unter Berücksichtigung der Vorfahrtsregel die Angabe aus dem Fragebogen an die Angaben aus dem Melderegister angepasst und deterministisch dahingehend bereinigt, dass die Person zwar mit einer Partnerin beziehungsweise einem Partner zusammenwohnt, jedoch nicht mit dieser/diesem verheiratet ist.

Zur Fehlerdokumentation waren in dieser Aufbereitungsphase zwei Schritte vorgesehen: die bereits beschriebene Kennzeichnung der Fehler und das Anlegen sogenannter Qualitätskennzeichen. Die Kennzeichnung der Fehler lässt Rückschlüsse auf die Art des Fehlers zu, also ob beispielsweise eine Filterung missachtet wurde. Die Qualitätskennzeichen geben an, ob ein Merkmal plausibel ist, durch deterministische Imputation bereits verändert wurde oder unplausibel ist und durch das Hot-Deck-Verfahren (Abschnitt 3.6) korrigiert werden muss (Hentschke und andere, 2024). Bei plausiblen Merkmalen kann dem Qualitätskennzeichen zudem entnommen werden, aus welcher Quelle der plausible Wert ursprünglich stammt. Dies war wichtig, da Melderegisterangaben aufgrund von §11 Absatz 4 Zensusgesetz 2022 stets unverändert bleiben mussten. Somit durften nur Werte, die ursprünglich aus dem Fragebogen oder der elektronischen Erhebungsliste stammten, mithilfe deterministischer Imputation oder Hot-Deck-Verfahren im Nachhinein nochmals verändert werden.

3.5 Plausibilisierung der Gemeinschaftsunterkünfte

Die Plausibilisierung der Personendaten aus den Gemeinschaftsunterkünften erfolgte analog zur Vorgehensweise der Personenerhebung. Als Besonderheiten sind die nachstehenden Punkte hervorzuheben:

- › Personendaten konnten aus bis zu zwei Datenquellen vorliegen (Melderegister und elektronische Erhebungsliste).
- › Für die Gemeinschaftsunterkünfte gab es einen reduzierten Fragebogen, da die Einrichtungsleitungen stellvertretend für die Bewohnerinnen und Bewohner auskunftspflichtig waren. Daher war es für die Aufbereitung nicht notwendig, das vorliegende Datenmaterial wie in der Personenerhebung in Teilbestände aufzuteilen.

- › Die Gemeinschaftsunterkünfte bildeten eine eigene PL-Klasse, da der Umfang der erhobenen Merkmale geringer war.
- › Aufgrund des Befragungsaufbaus wurden lediglich vier Fehlerprüfungen vorgenommen. Die Prüfung auf falsche Filterführung entfiel, da die Befragungsstruktur kein Überspringen von Fragen vorsah.

3.6 Imputationssoftware CANCEIS

Im Anschluss an die beschriebenen Plausibilisierungsschritte und die deterministische Imputation erfolgte für alle primärstatistischen Erhebungsteile separat die sogenannte Hot-Deck-Imputation. Diese wurde mittels der Software CANCEIS⁴ durchgeführt, die einen Nächste-Nachbarn-Ansatz verwendet. Die Grundidee ist, alle fehlenden und fehlerhaften Werte eines Datensatzes (= Empfänger) mit den beobachteten Werten eines vollständigen und plausiblen Datensatzes (= Spender) aus derselben Erhebung zu ersetzen. Ziel dieses Imputationsverfahrens ist, einen vollständigen und plausiblen Datenbestand auszugeben.

CANCEIS wurde im Zensus 2022 für alle primärstatistischen Erhebungsteile (Gebäude- und Wohnungszählung, Personenerhebungen, Befragung an Anschriften mit Sonderbereichen) eingesetzt. Die Software kam bereits bei der Gebäude- und Wohnungszählung im Zensus 2011 zum Einsatz (Grunwald/Krause, 2014). Innerhalb der Erhebungsteile wurden die Daten vor der spenderbasierten Imputation analog zum Vorgehen bei der Plausibilisierung in verschiedene Imputationsklassen aufgeteilt, die die Daten in homogene Untergruppen unterteilten und einen einheitlichen Merkmalskranz umfassten (Abschnitte 3.4 und 3.5). Die Imputation fand innerhalb jeder Imputationsklasse separat statt, sodass als Spender immer nur ein Datensatz aus derselben Imputationsklasse wie der des Empfängers verwendet wurde.

Für jeden Empfänger wurde genau ein Spender verwendet. Das sollte sicherstellen, dass der imputierte Datensatz nicht nur plausibel wurde, sondern auch in der Erhebung selbst beobachtete Kombinationen von Merk-

malswerten enthielt. Es wurde ein Spender gesucht, der dem Empfänger in allen im Zensus erhobenen Merkmalen ähnlich war. Dem zugrunde liegt die Annahme, dass Zusammenhänge zwischen allen Bereichen und Merkmalen einer Erhebung bestehen. Ein und derselbe Spender kann für mehrere Empfänger spenden. Um die Werte des Spenders aber nicht zu häufig zu duplizieren, wurde für jeden Erhebungsteil eine Höchstgrenze festgelegt, wie häufig ein plausibler Datensatz maximal spenden durfte.

Die in CANCEIS implementierte Methodik bietet gegenüber anderen Verfahren den Vorteil, dass sie Plausibilitätsregeln bei der Imputation direkt berücksichtigen kann. Das heißt bei Inkonsistenzen zwischen Merkmalen ist nicht zunächst eine Identifizierung derjenigen Felder vorzunehmen, die imputiert werden müssen, und anschließend eine Imputation durchzuführen. Vielmehr entscheidet CANCEIS bei Inkonsistenzen zwischen Merkmalen anhand der Daten und der definierten und CANCEIS bereitgestellten Fehlerbeschreibungen, welcher Wert ersetzt werden soll (beziehungsweise welche Werte ersetzt werden sollen). Dabei wird der Grundsatz verfolgt, dass nur potenziell fehlerhafte Werte korrigiert und möglichst wenige der beobachteten Werte angepasst werden (Bankier, 2012).

Für die Imputation unterteilt CANCEIS einen Datenbestand zunächst in Empfänger- und Spenderdatensätze. Bei den Empfängern handelt es sich um all jene Datensätze, die mindestens eine unter Abschnitt 3.2 Nr. 1 beschriebene Fehlerart aufweisen. Die anderen, vollständig plausiblen Datensätze bilden den Pool der Spenderdatensätze. Anschließend wird für jeden Empfängerdatensatz ein Spenderdatensatz gesucht, der dem Empfänger in den vorhandenen Angaben möglichst ähnlich ist. Um die Ähnlichkeit zu bestimmen, wird die mathematische Distanz zwischen Empfänger und Spender mittels einer Distanzfunktion berechnet, die die in der Erhebung vorkommenden Merkmale berücksichtigt, die auch beim Empfänger erhoben werden. Von den Spendern mit den geringsten Distanzen zum Empfänger wird einer ausgewählt, dessen Werte für die zu ersetzenden Werte des Empfängers eingesetzt werden, sodass der imputierte Empfänger vervollständigt wird und alle Prüfregelein hält.

Um alle Imputationen auch im späteren Prozessverlauf nachvollziehen zu können, wurden beim Imputationsprozess, ergänzend zu den Qualitätskennzeichen der

⁴ Die Software CANCEIS (Canadian Census Edit and Imputation System) ersetzt fehlende oder unplausible Angaben der Empfängerdatensätze mit beobachteten Werten von Spenderdatensätzen aus derselben Erhebung.

Plausibilisierung (Abschnitt 3.4), weitere Qualitätskennzeichen angelegt. Sie gaben für jeden Wert an, ob dieser durch die spenderbasierte Imputation eingesetzt oder verändert wurde oder nicht.

3.7 Imputation der Gebäude- und Wohnungszählung

Die spenderbasierte Imputation mit CANCEIS wurde bei der Gebäude- und Wohnungszählung für zwei Arten von Antwortausfällen eingesetzt:

Unit Nonresponse: Kompletter Antwortausfall auf Gebäudeebene, es liegen keinerlei Daten zum Gebäude oder den darin befindlichen Wohnungen vor.

Item Nonresponse: Einzelne fehlende oder fehlerhafte Werte bei Gebäude- oder Wohnungsmerkmalen sowie fehlende Angaben zu ganzen Wohnungen, die sich innerhalb eines Gebäudes befinden, zu dem Rückmeldungen zu weiteren Wohnungen vorliegen.

Die im Datenmaterial vorhandenen vollständigen und plausiblen Gebäude können dabei sowohl bei der Unit- als auch bei der Item-Nonresponse-Imputation als Spender dienen.

Beim Unit Nonresponse lagen weder zum Gebäude noch zu den Wohnungen Erhebungsdaten vor, hier konnte die spenderbasierte Imputation nur unter Zuhilfenahme von Zusatzinformationen sinnvoll durchgeführt werden. Hierzu wurden Angaben aus vorliegenden Registern zur räumlichen Lage der Gebäude sowie Informationen aus externen Quellen zur Zahl der Haushalte und Personen an der Gebäudeanschrift in die Spendersuche einbezogen. Dieses Vorgehen stellte sicher, dass das Spendergebäude in derselben Region lag und die korrekte Gebäudegröße beziehungsweise Zahl an bewohnten Wohnungen aufwies. Für das zu imputierende Gebäude wurden dann alle Angaben zu dem Gebäude und den darin enthaltenen Wohnungen vom Spendergebäude übernommen.

Bei der Imputation des Item Nonresponse erfolgte die Imputation anhand zweier verschiedener Vorgehensweisen je nach Größe des Empfängergebäudes. Bei Gebäuden mit bis zu 15 Wohnungen wurden die Gebäude entsprechend der Anzahl der Wohnungen im Gebäude in 15 Imputationsklassen aufgeteilt. Um die Größen-

struktur der Gebäude zu berücksichtigen, kamen dann nur Spendergebäude infrage, die eine identische Anzahl Wohnungen aufwiesen. Dieses Vorgehen stellte sicher, dass beispielsweise Einfamilienhäuser auch nur Spender für andere Einfamilienhäuser sein konnten. Bei Gebäuden mit mehr als 15 Wohnungen konnte dieses Verfahren nicht beibehalten werden, da bei sehr großen Gebäuden gar keine oder nur sehr wenige (Spender-) Gebäude zur Verfügung standen. Diese Gebäude wurden in einer gemeinsamen Imputationsklasse für CANCEIS bereitgestellt. Dabei erfolgte die Imputation schrittweise, wobei zunächst die Wohnungen mit ihren Wohnungsmerkmalen und anschließend die Gebäudemerkmale imputiert wurden. Die Imputation der Wohnungen erfolgte bei beiden Vorgehensweisen unter Berücksichtigung der Anzahl der Haushalte an der Anschrift.

Einen Einblick in die Quoten der Unit- und Item-Nonresponse-Imputation bei der Gebäude- und Wohnungszählung des Zensus 2022 bietet der Aufsatz von Hentschke und anderen (2024).

3.8 Imputation der Personenerhebung

Bei der Personenerhebung fand ausschließlich eine Item-Nonresponse-Imputation statt, das heißt es wurden einzelne fehlende oder fehlerhafte Merkmale (zum Beispiel zur Klassenstufe) von Personen ergänzt. Eine Unit-Nonresponse-Imputation fand bei der Personenerhebung nicht statt. Fehlten komplette Rückmeldungen von Auskunftspflichtigen, wurde dies bei der Hochrechnung der Ergebnisse berücksichtigt.

Dabei galt es, die im Zensus 2022 definierte Vorfahrtsregel zu berücksichtigen, dass Personen, die sowohl im Melderegister als auch der primärstatistischen Erhebung enthalten sind, bei demografischen Angaben immer die Informationen aus dem Melderegister erhalten. So wurden die Angaben aus den Melderegistern, die in der Aufbereitungsphase bereits plausibilisiert wurden (Abschnitt 3.4), durch die spenderbasierte Imputation nicht mehr verändert.

Bei der Personenerhebung unterteilten sich die Imputationsklassen nach dem Merkmalskranz der Befragung und dem Alter der Befragten (PL-Klassen). Für die Hauptbefragung wurden die unter 15-Jährigen und die Personen, die 15 Jahre oder älter sind, getrennt voneinander imputiert, da den ab 15-Jährigen unter anderem auch

Fragen zur Berufstätigkeit gestellt wurden. Um die Spendersuche zu beschleunigen, wurden die Datensätze nach dem Alter sortiert in CANCEIS eingegeben, sodass Datensätze mit ähnlichem Alter zuerst als potenzielle Spender untersucht wurden.

Personen an Gemeinschaftsunterkünften wurden gemeinsam in einer Imputationsklasse plausibilisiert und imputiert.

4

Fazit

Mit der übergreifenden Plausibilisierung der verschiedenen Teilprojekte sowie den Plausibilisierungsroutinen innerhalb der elektronischen Erhebungsphase war es möglich, Erfassungs- und Abgrenzungsfehler bereits früh im Prozess der Aufbereitung des Zensus 2022 zu identifizieren und angemessene Maßnahmen zur Korrektur zu ergreifen. Die eingesetzten Plausibilisierungs- und Imputationsverfahren bei den primärstatistischen Erhebungsteilen konnten die nach Abschluss der Erhebungsphase noch vorhandenen fehlerhaften und fehlenden Werte identifizieren und korrigieren. Als Resultat der durchgeführten Aufbereitungsschritte steht ein stimmiger Zensusdatenbestand für Auswertungen zur Verfügung. 

LITERATURVERZEICHNIS

Bankier, Mike. *Imputing Numerical and Qualitative Variables Simultaneously*. Social Survey Methods Division, Statistics Canada (internes Dokument). 2012.

Freier, Benjamin/Mosel, Juliane. [Online First als Leitgedanke für effiziente Primärerhebungen beim Zensus 2021](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 46 ff.

Gaedke, Annika/Pfahl, Miriam/Strohalm, Anna. [Die Online-First-Strategie und die Online-Fragebogen im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 92 ff.

Grunwald, Sven/Krause, Anja. [Umgang mit fehlenden Angaben in der Gebäude- und Wohnungszählung 2011](#). In: Wirtschaft und Statistik. Ausgabe 8/2014, Seite 437 ff.

Hentschke, Janine/Tobies, Cara-Aileen/Weber, Susanne/Claus, Dennis. [Kern-Qualitätskennzahlen und Zielwerte im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 106 ff.

Preising, Marcel/Lange, Kerstin/Dumpert, Florian. [Imputation zur maschinellen Behandlung fehlender und unplausibler Werte in der amtlichen Statistik](#). In: WISTA Wirtschaft und Statistik. Ausgabe 5/2021, Seite 40 ff.

RECHTSGRUNDLAGEN

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

DIE AKTUALISIERUNG DES ORTS- VERZEICHNISSES UND PLAUSIBILI- SIERUNG DES GEBURTSSTAATES IM ZENSUS 2022

David Aude, Johannes Bolz, Lisa Boragk

📌 **Schlüsselwörter:** Geburtsort – Orts-Staat-Zuordnung – Melderegister –
Gebietsverzeichnis – Referenzierung

ZUSAMMENFASSUNG

Für den Zensus 2011 war mit dem Ortsverzeichnis eine Liste mit Ort-Staat-Kombinationen aufgebaut worden, welche für die Zuweisung von Geburtsstaaten auf Basis von gelieferten Angaben aus den Melderegistern verwendet wurde. Notwendig war der Aufbau eines Ortsverzeichnisses, weil das Merkmal „Geburtsort – Staat“ in den Melderegistern in vielen Fällen entweder nicht befüllt ist oder eine widersprüchliche Information zum gelieferten Geburtsort enthält. In der Begründung zu §6 Zensusgesetz 2022 ist festgelegt, dass die Daten aus den Melderegistern zur Aktualisierung des Ortsverzeichnisses genutzt werden. Der Artikel beschreibt das methodische Vorgehen bei der Aktualisierung des Ortsverzeichnisses und die Zuweisung von Geburtsstaaten auf Datensatzebene. Insgesamt wurde mithilfe des Ortsverzeichnisses für über 99% der Personen ein Geburtsstaat ermittelt.

📌 **Keywords:** *place of birth – allocation of place to country – population register – list of territories – referencing*

ABSTRACT

A directory of places was created for the 2011 Census, which provided a list of place and country combinations that was used to assign individuals a country of birth on the basis of data in the population registers. The development of a directory of this kind was necessary because in many cases the population registers either contain no information at all for “place of birth - country” or the information provided is inconsistent with the place of birth indicated. Referring to section 6 of the 2022 Census Act, the explanatory memorandum specifies that the data from the population registers are used to update the directory of places. This article describes the methodological approach taken to update the directory of places and assign countries of birth at the data record level. Overall, it was possible to determine a country of birth for over 99% of individuals with the help of the directory of places.

David Aude

hat Survey Methodology und Ostasienwissenschaften an der Universität Duisburg-Essen studiert und ist Wissenschaftlicher Mitarbeiter im Referat „Bevölkerungsstand“ der Gruppe „Registerbasierter Datenbestand des Zensus“ des Statistischen Bundesamtes. Er beschäftigt sich unter anderem mit der Weiterentwicklung des Ortsverzeichnisses und der Plausibilisierung des Geburtsstaates.

Johannes Bolz

studierte Sozial- und Wirtschaftswissenschaften sowie Soziologie an der Carl von Ossietzky Universität Oldenburg und der Universität Mannheim. Nach mehrjähriger Tätigkeit in der Kommunalstatistik arbeitet er seit 2020 im Statistischen Bundesamt. Im Zensus 2022 war er für die Weiterentwicklung des Ortsverzeichnisses verantwortlich. Mittlerweile ist er als Referent in der Gruppe „Registerbasierter Datenbestand des Zensus“ im Referat „Bevölkerungsstand“ unter anderem mit der Entwicklung ergänzender Bevölkerungsstatistiken befasst.

Lisa Boragk

hat Volkswirtschaftslehre an der Heinrich-Heine-Universität Düsseldorf und der Justus-Liebig-Universität Gießen studiert und ist Referentin im Referat „Personenerhebungen, Anschriften“ des Statistischen Bundesamtes. Bis 2023 arbeitete sie im Projekt Zensus 2022 an der Weiterentwicklung des Ortsverzeichnisses sowie im Teilprojekt Sonderbereiche. Aktuell ist sie als Technikkordinatorin für das Referat „Personenerhebungen, Anschriften“ tätig.

1

Einleitung

Die Europäische Union (EU) fordert von ihren Mitgliedstaaten, dass bei den Volks- und Wohnungszählungen neben weiteren Merkmalen auch das Merkmal „Geburtsstaat“ nach dem aktuell geltenden Gebietsstand zu ermitteln ist. Die Geburtsorts- sowie Geburtsstaatsangaben der deutschen Melderegister stellen jedoch keine ausreichende Basis dar, um für alle in den Melderegistern enthaltenen Personen den korrekten Geburtsstaat zu ermitteln. Aufgrund melderechtlicher Vorgaben sind die beiden Felder in unterschiedlicher Weise und Qualität befüllt.

Daher war bereits für den Zensus 2011 mit dem Ortsverzeichnis eine Liste mit Ort-Staat-Kombinationen aufgebaut worden, um auf Basis von gelieferten Geburtsorten aus den Melderegistern dem Geburtsort einen aktuellen Geburtsstaat zuzuweisen. Daraus abgeleitet können verschiedene Aggregationsstufen, beispielsweise nach EU/nicht aus der EU oder nach Kontinenten, gebildet werden. Die Zuordnung über das Ortsverzeichnis bildete zudem die Basis für die Erstellung von Ergebnissen zu Personen mit Einwanderungserfahrung im Zensus 2022. Die Leitdatei, Programmabläufe und Ergebnisse des Ortsverzeichnisses 2022 sind ferner die Grundlage für die Weiterentwicklung des registergestützten Zensus.

Das aus vielfältigen Quellen¹ zusammengestellte Ortsverzeichnis dient in der Datenaufbereitung dem Zweck, jeder Person einen korrekten Geburtsstaat zuzuweisen. Dabei werden Informationen zu Geburtsort, Geburtsstaat, Staatsangehörigkeiten und Zuzugsstaat verwendet, um den Geburtsstaat einer Person zu ermitteln. Außerdem wird die Information des Geburtsdatums verwendet, um Personen, die in ehemaligen deutschen Gebieten geboren wurden, zu identifizieren und ihnen Deutschland als Geburtsstaat im nationalen Auswertungskonzept zuzuweisen. Mithilfe des Ortsverzeichnisses kann für über 99 % der Personen ein Geburtsstaat ermittelt werden.

1 Unter anderem aus dem Gemeindeverzeichnis-Informationssystem (GV-ISys), aus LAU-Listen (LAU = Local administrative units [Listen lokaler Verwaltungseinheiten]), OpenStreetMap.

Nach einem kurzen Überblick über die beiden Auswertungskonzepte in Kapitel 2 stellt Kapitel 3 die für die Aktualisierung des Ortsverzeichnisses im Zensus 2022 verwendeten Datenquellen vor. Kapitel 4 behandelt die Aufbereitung der Daten und die dabei vorgenommene Qualitätssicherung. Die verschiedenen Prozesse der Zuordnung von Staaten zu Geburtsorten erläutert Kapitel 5. Der Artikel schließt mit einem kurzen Blick auf die Ergebnisse dieser Zuordnung und einem Fazit.

2

Auswertungskonzepte

Die Anforderungen der EU einerseits und die nationalen Anforderungen andererseits erforderten, dass die Zuordnung des Geburtsortes zu einem Geburtsstaat im Ortsverzeichnis nach zwei Auswertungskonzepten erfolgte.

2.1 Europäisches Auswertungskonzept

Für die Datenlieferungen an die EU war jeder Person der Geburtsstaat auf der Grundlage der am 1. Januar 2021 bestehenden Grenzen zuzuweisen. Dies erfolgte zunächst nach dem Schlüssel der vom Statistischen Bundesamt herausgegebenen [Staats- und Gebietssystematik](#). Für die Zuordnung nach dem europäischen Auswertungskonzept wurde die Gliederungssystematik gemäß der EU-Verordnung 2017/543 unter dem Merkmal „Geburtsland/-ort“ (beziehungsweise „Place of Birth“) verwendet.

2.2 Nationales Auswertungskonzept

Im Sinne des nationalen Auswertungskonzeptes sollen Personen, die bis zum 2. August 1945² in Orten geboren wurden, die zu Deutschland nach dem Gebietsstand vom 31. Dezember 1937 gehörten, zusätzlich als „in Deutschland Geborene“ ausgewiesen werden und zwar auch dann, wenn die Orte nicht innerhalb der aktuellen Grenzen Deutschlands liegen.³ Das betrifft insbesondere ehemalige deutsche Ostgebiete wie Ostpreußen

2 Das Datum 2. August 1945 stellt auf die Beschlüsse der Potsdamer Konferenz ab.

3 Siehe Artikel 116 Grundgesetz.

oder Schlesien. Auf der Ebene der Personen wird das Geburtsland in diesen Fällen abhängig vom Geburtsdatum bestimmt.

3

Verwendete Daten

Für die Weiterentwicklung wurden neben dem Ortsverzeichnis des Zensus 2011 und den Melderegisterangaben noch weitere Quellen verwendet. Nachfolgend werden alle Datenquellen kurz vorgestellt.

3.1 Ortsverzeichnis 2011

Das Ortsverzeichnis des Zensus 2011 diente als Ausgangsbasis für das Ortsverzeichnis des Zensus 2022. Insbesondere aufgrund der im Zuge der Migrationsbewegungen 2015/2016 neu aufgenommenen Geburtsorte in die Melderegister war dessen Stand veraltet und es musste aktualisiert und ergänzt werden. Im Zensus 2011 wurden für den Aufbau des Ortsverzeichnisses die Melderegisterinformationen zum „Geburtsort“, „Geburtsort – Staat“, „Geburtsort – Standesamt“ und der Zuzugsstaat bei Zuzug aus dem Ausland einer Person verwendet. Die etwa 87 Millionen für den Zensus 2011 übermittelten Melderegisterdatensätze (Haupt- und Nebenwohnsitze) wurden anhand dieser Merkmale zu rund 7,3 Millionen unterschiedlichen Fallkombinationen aggregiert (Weißer, 2012).

3.2 Melderegister

Für die Aktualisierung des Ortsverzeichnisses konnten die vier Datenlieferungen der Melderegister gemäß § 5 Zensusgesetz 2022 genutzt werden.

Für das Ortsverzeichnis 2022 wurden die folgenden Merkmale verwendet:

- › „Geburtsort“ (Freitextangabe)
- › „Geburtsort – Staat“ (dreistelliger Staatenschlüssel gemäß der Staats- und Gebietssystematik)
- › „Staatsangehörigkeiten 1 bis 4“ (Staatenschlüssel)
- › „Zuzug aus dem Ausland – Staat“ (Staatenschlüssel)
- › „Geburtsdatum“

Bei dem Merkmal „Geburtsort“ handelt es sich um ein Freitextfeld, das mit bis zu 70 Zeichen befüllt werden kann, daher können in diesem Feld die vielfältigsten Einträge vorhanden sein.

Beispielsweise

- › ist nicht nur die Stadt, in der diejenige Person geboren wurde, eingetragen, sondern auch der dazugehörige Stadtteil.
- › ist bei Eingemeindungen eine Kombination aus altem sowie neuem Namen des Ortes eingetragen.
- › sind Staatenschlüssel oder „unbekannt“ eingetragen.
- › gibt es historische Bezeichnungen aktueller Orte.
- › unterscheiden sich die Formen der Ortsnamensbestandteile sowie die Ortsnamenszusätze.
- › befinden sich in dem Feld „Geburtsort“ häufig Angaben, die die Lage des Ortes näher beschreiben.
- › sind sprachlich bedingte unterschiedliche Transkriptionen eines Geburtsortes zu finden.
- › finden sich ausländische statt der in Deutschland üblichen Schreibweise für die Orte.
- › existieren Schreibfehler.

Das Merkmal „Geburtsort – Staat“ enthält die eingetragenen Geburtsstaaten und sollte bestenfalls bei allen im Ausland geborenen Personen mit dem entsprechenden Gebiets- und Staatenschlüssel befüllt sein. Bei in Deutschland geborenen Personen sollte das Feld leer bleiben. Diese beiden Vorgaben sind allerdings nicht konsequent umgesetzt. Analysen des Merkmals ergaben, wie bereits 2011, dass teilweise widersprüchliche Angaben in den Merkmalen Geburtsort und Staatenschlüssel vorlagen. Zudem bekamen in Deutschland geborene ausländische Personen fälschlicherweise einen ausländischen Staatenschlüssel zugewiesen. Andere fehlerhafte Angaben kann das Feld „Geburtsort – Staat“ enthalten, wenn zum Beispiel der Ersatzschlüssel „999“ – keine Angabe – eingetragen wurde, obwohl sich der Geburtsort eindeutig einem Staat zuweisen lässt. Es können zudem veraltete Staatenschlüssel eingetragen worden sein (zum Beispiel „159“ Sowjetunion).

Genau wie das Merkmal „Geburtsort – Staat“ enthält die Staatsangehörigkeit den dreistelligen Staatenschlüssel und dient als weiterer Anhaltspunkt zur Referenzierung auf den korrekten Geburtsstaat, falls das Feld „Geburtsort – Staat“ nicht oder falsch befüllt ist. Für das Ortsverzeichnis wurden die ersten drei Staatsangehörigkeiten der Person verwendet.⁴

Das Merkmal „Zuzug aus dem Ausland – Staat“ diene ebenfalls als weiterer Anhaltspunkt zur Referenzierung auf den korrekten Geburtsstaat im Ortsverzeichnis. Dieses Feld enthält die Informationen zum Herkunftsgebiet bei Zuzug aus dem Ausland nach Deutschland. Ist eine Person mehrmals aus dem Ausland nach Deutschland zugezogen, ist das Gebiet des letzten Zuzugs eingetragen.

Das Merkmal „Geburtsdatum“ wurde benötigt, um die Anforderungen des nationalen Auswertungskonzeptes zu erfüllen. Dafür wurde aus der Angabe des Geburtsdatums eine Binärvariable konstruiert, welche lediglich unterscheidet, ob eine Person vor dem 2. August 1945 geboren wurde oder danach.

3.3 Referenzverzeichnisse

Referenzverzeichnisse bildeten zusammen mit den Einträgen des Ortsverzeichnisses 2011 die Basis für die Aktualisierung und Optimierung des neuen Ortsverzeichnisses 2022. Eine Vielzahl an potenziellen Datenquellen wurde recherchiert, um eine möglichst umfangreiche Datenbasis zur Zuordnung von Geburtsorten zu Staaten zu erhalten. Die verschiedenen Referenzverzeichnisse wurden vor der Verwendung eingehend geprüft – einige der Referenzverzeichnisse unterliegen keinen regelmäßigen Qualitätsprüfungen und Routinen – und je nach Qualitätseinschätzung für die Verwendung im Ortsverzeichnis priorisiert. Sie wurden nach erfolgter Aufbereitung, bei der auch aufgedeckte fehlerhafte Einträge entfernt wurden, in einer Leitdatei zusammengeführt, wobei jedem Eintrag eine Priorität, je nach Quelle, zugewiesen wurde.

⁴ Eine vierte Staatsangehörigkeit liegt nur in seltenen Fällen tatsächlich vor und hätte die Komplexität des Programmablaufs ohne nennenswerten Informationsgewinn unnötig erhöht.

Staaten- und Gebietsschlüssel (Statistisches Bundesamt)

Der „Staaten- und Gebietsschlüssel“ des Statistischen Bundesamtes (BEV-Code) enthält einen dreistelligen numerischen Schlüssel zur Kennzeichnung von Staaten, Staatsangehörigkeiten und Gebieten. Dieser gilt nur im nationalen Kontext und enthält aktuelle und historische Staatsangehörigkeiten sowie deren Staatenschlüssel und die Staatennamen in Klarschriftangabe einschließlich abhängiger Gebiete. Mit diesem Verzeichnis lässt sich jeder Geburtsstaat einem bestimmten Staatenschlüssel zuordnen, es wurde dementsprechend für die Befüllung der Ergebnismerkmale des Ortsverzeichnisses verwendet. Aufgrund der hohen Qualität des Staaten- und Gebietsschlüssels wurde ihm eine hohe Priorität zugewiesen.

Ergänzt wurde diese Liste durch das historische Staatenleitband und eine Liste von Staatenbezeichnungen. In diesen Listen befinden sich neben (aktuellen und historischen) Staatennamen und Angaben zu abhängigen Gebieten weitere Gebietsangaben, wie Bundesländer, Namen von Regionen, englische und rumänische Schreibweisen der Staaten, Falschschreibweisen und Abkürzungen, jeweils hinterlegt mit den dazugehörigen Staaten- und Gebietsschlüsseln.

GV-ISys (Statistisches Bundesamt)

Für die Leitdatei wurde weiterhin das Gemeindeverzeichnis-Informationssystem (GV-ISys) des Statistischen Bundesamtes verwendet. Dieses ist eine quartals- und jahresweise veröffentlichte Bestandsliste aller aktuell existierenden politischen Verwaltungsebenen und selbstständigen Gemeinden in Deutschland mit entsprechenden Merkmalen wie Name, amtlicher Gemeindegemeinschaftsschlüssel (AGS), Postleitzahl (PLZ), Fläche, Einwohnerzahl, siedlungsstrukturierter Typisierung. Zudem wird jährlich eine Übersicht über die Gebietsstandsänderungen veröffentlicht, in der zum Beispiel bei Eingemeindungen oder Gemeindezusammenschlüssen die vorher und anschließend gültigen Gemeindegemeinschaftenamen vermerkt sind. Die Listen gehen je nach Verfügbarkeit bis 1993 beziehungsweise bis 1970 zurück. Auch dieser Liste wurde aufgrund der hohen Qualität eine hohe Priorität zugewiesen.

LAU-Liste der Europäischen Union (Eurostat)

Ähnlich dem GV-ISys führt das Statistische Amt der Europäischen Union (Eurostat) die Verwaltungseinheiten der Mitgliedstaaten und einiger EU-Kandidatenländer in einer Liste. Die Verwaltungseinheiten werden dabei auf unterschiedlichen Ebenen vorgehalten, sogenannten NUTS (Nomenclature des unités territoriales statistiques). In Deutschland entspricht die NUTS-1-Ebene den Bundesländern, NUTS-2 den Regierungsbezirken, NUTS-3 den Kreisen und kreisfreien Städten. Unterhalb der NUTS folgt eine LAU-Ebene, die kleinste Verwaltungsebene, die in Deutschland den Gemeinden entspricht (Local Administrative Units). Die NUTS-Verordnung sieht vor, dass die EU-Mitgliedstaaten die Listen ihrer lokalen Verwaltungseinheiten an Eurostat senden. Für das Ortsverzeichnis wurden die LAU-Listen seit 2010 bis zum aktuellsten veröffentlichten Stand 2019 verwendet. Da es sich um die offiziellen Listen von Eurostat handelt, wurde die Qualität ebenfalls als hoch eingestuft.

Ortschaftsverzeichnis des Bundesamtes für Kartographie und Geodäsie

Um die Vorgaben des nationalen Auswertungskonzeptes zu erfüllen, müssen Orte, die zu Deutschland nach dem Gebietsstand vom 31. Dezember 1937 gehörten, Teil der Leitdatei sein. Somit ist sichergestellt, dass Personen, die bis zum 2. August 1945 in dem genannten Gebietsstand geboren worden sind, ein deutscher Geburtsort zugewiesen werden kann. Zu diesem Zweck wurde das Ortschaftsverzeichnis des Bundesamtes für Kartographie und Geodäsie genutzt, welches ehemals zu Deutschland gehörende Gebiete in ihrer Chronologie von 1900 bis 1993 und alle verwendeten Ortschaftsnamen in verschiedenen Schreibweisen enthält. Als amtlicher Liste des Bundesamtes für Kartographie und Geodäsie wurde ihr ebenfalls eine hohe Priorität zugewiesen.

Standesamtsregister ehemaliger deutscher Orte (Landesarchiv Berlin)

Durch das Ende des Zweiten Weltkrieges und das Potsdamer Abkommen, durch das ehemals dem Deutschen Reich zugehörige Gebiete von diesem abgetrennt und an Polen oder die damalige Sowjetunion angegliedert wurden (Ostpreußen, Westpreußen, Oberschlesien, Niederschlesien und Pommern), kam es zu großen Flucht-

bewegungen aus diesen Gebieten. Register aus den Standesämtern dieser Gebiete wurden auf der Flucht zum Teil mitgeführt. Dabei handelt es sich um Personenregister mit Informationen zu Geburten, Heirat und Tod. Die von dort mitgenommenen Standesamtsunterlagen blieben in Deutschland und wurden entsprechend der geltenden Zuständigkeit beim Standesamt I in Berlin verwahrt. Zusätzlich enthalten die dort liegenden Informationen Register zu während des Zweiten Weltkriegs in den besetzten Gebieten verstorbenen Deutschen ab 1940, die die damalige Tschechoslowakei, Polen, Litauen, Lettland, Estland, Belarus, die Ukraine, die Niederlande und Norwegen umfassen.

Im Jahr 2015 wurde diese Überlieferung an das Landesarchiv Berlin⁵ abgegeben. Diese Liste ergänzt das Ortschaftsverzeichnis des Bundesamtes für Kartographie und Geodäsie und wurde ebenfalls als von hoher Priorität eingeschätzt.

OpenStreetMap (OSM)

[OpenStreetMap](#) ist ein 2004 gegründetes Projekt, das Geodaten aus der ganzen Welt mithilfe Freiwilliger sammelt, strukturiert und allgemein frei zur Verfügung stellt (Open Data). Für die Aktualisierung des Ortsverzeichnisses wurde ein Auszug der OSM-Daten mit Orten („places“) erstellt, der nach einer Bereinigung um Duplikate und offensichtlich fehlerhafte Einträge etwa 3 Millionen Ortsangaben weltweit aufweist. Neben dem erheblichen Umfang und der Verfügbarkeit von Geokoordinaten ist ein weiterer Vorteil der OSM-Datenbank, dass Ortsbezeichnungen häufig in mehreren Sprachen und teilweise auch historischen Schreibweisen vorliegen. Aufgrund des großen Umfangs und der Tatsache, dass die OSM-Daten nutzergeneriert sind, ist eine vollständige Qualitätssicherung der Daten kaum möglich, die Priorität für das Ortsverzeichnis wurde daher als niedrig eingestuft.

Ortsverzeichnis der KOSIS-Gemeinschaft HHSTAT

Für die Ableitung des Migrationshintergrunds wird in weiten Teilen der Städtestatistik das von der [KOSIS-Gemeinschaft HHSTAT](#) entwickelte Programm MigraPro eingesetzt. MigraPro ermöglicht mit den im Melderegister enthaltenen Informationen, Aussagen über

5 [landesarchiv-berlin.de](https://www.landesarchiv-berlin.de)

den Migrationshintergrund zu treffen. Dafür wird auch ein Ortsverzeichnis bereitgestellt, welches den Angaben des Geburtsortes den dazugehörigen amtlichen Gemeindeschlüssel (bei Orten innerhalb Deutschlands) beziehungsweise den dreistelligen Staaten-/Gebietschlüssel des Statistischen Bundesamtes zuordnet. Dieses Ortsverzeichnis wird in den einzelnen Städten gepflegt und ergänzt. Die betreuende Stelle von HHSTAT, die dem Statistischen Amt der Landeshauptstadt Stuttgart angegliedert ist, hat dieses Verzeichnis dem Statistischen Bundesamt für das zu erstellende Ortsverzeichnis zur Verfügung gestellt. Da diese Liste allerdings kein amtliches Verzeichnis darstellt, wurde im Vergleich zu anderen Referenzverzeichnissen eine niedrigere Priorität zugewiesen.

Geschichtliches Ortsverzeichnis (MiniGOV)

Für in ehemals zu Deutschland gehörenden Gebieten geborene Personen ist zu erwarten, dass die Ortsbezeichnungen mehrheitlich in historischer Schreibweise vorliegen, die von der heute üblichen und daher in offiziellen Quellen geführten abweichen kann. Das Geschichtliche Ortsverzeichnis⁶ ist eine nutzergenerierte Datenbank, welche historische Ortsnamen besonders für die Familienforschung bereithält. Ein Auszug mit reduziertem Umfang wird von den Betreibenden unter dem Namen MiniGOV⁷ zur Verfügung gestellt und die Daten lassen sich frei herunterladen. Das MiniGOV enthält neben den Ortsnamen den (letzten) deutschen Namen und Geokoordinaten, was eine Verknüpfung mit anderen Datenquellen vereinfacht. Da es sich um eine nutzergenerierte Datenbank handelt, sind die Einträge nicht umfassend qualitätsgesichert und deshalb nur mit einer niedrigen Priorität eingeflossen. Das MiniGOV wurde ausschließlich für die Umsetzung des nationalen Auswertungskonzeptes verwendet.

⁶ Details finden sich unter: wiki.genealogy.net/GOV

⁷ Der Aufbau des MiniGOV ist unter wiki.genealogy.net/GOV/Mini-GOV beschrieben.

4

Datenaufbereitung und Qualitätssicherung der Referenzverzeichnisse und des Ortsverzeichnisses 2011

Damit eine möglichst effiziente Zuordnung zu einem Geburtsstaat möglich war, wurden die Angaben des Geburtsortes zuerst standardisiert. Dazu gehörten verschiedene Aufbereitungsschritte, wie die durchgängige Großschreibung aller Angaben oder die Umsetzung diakritischer Zeichen. Angaben zur näheren Beschreibung eines Geburtsortes, zum Beispiel zeitliche („Köln früher Porz“ beziehungsweise „Porz jetzt Köln“) oder geografische („Neuenkirchen bei Greifswald“) Verweise, wurden entfernt, sodass lediglich aktuelle Gemeindefür den Namen weiterverarbeitet wurden.

In einigen Fällen war als Geburtsort kein Ortsname eingetragen, sondern ein Vermerk, dass der Geburtsort unbekannt sei. Es wurde eine Referenzliste (Unbekannt-Liste) mit etwa 200 Einträgen zusammengestellt, die als Bezeichnungen für einen unbekanntes Geburtsort eingetragen wurden. Die Einträge zum Geburtsort wurden mit dieser Referenzliste abgeglichen und mit „UNBEKANNT“ kodiert. Zusätzlich zu diesen Unbekannt-Bezeichnungen wurden auch ungültige Angaben umgeschlüsselt, etwa wenn im Feld Geburtsort ein Geburtsdatum eingetragen wurde. Entsprechende Einträge wurden ebenfalls mit „UNBEKANNT“ kodiert.

War im Feld Geburtsort ein Hinweis auf einen Staat oder ein Bundesland beziehungsweise eine Region enthalten (zum Beispiel „Moskau/Russland“ oder „Hannover/Niedersachsen“), wurde der Geburtsort um diese Angaben bereinigt. Allerdings wurde diese Information mit den Angaben im Merkmal „Geburtsort – Staat“ abgeglichen und bei entsprechender Übereinstimmung als gesicherte Angabe bestätigt. Anderenfalls wurde die Angabe im Feld „Geburtsort“ mit den weiteren Angaben zu Staatsangehörigkeiten und Zuzugsstaat abgeglichen und bei Übereinstimmung als Geburtsstaat vermerkt.

Um die Anforderungen des deutschen Auswertungskonzeptes erfüllen zu können, war es notwendig, die in der Ortsdatei erfassten Orte aus den betreffenden Gebieten dem Gebietsstand vom 31. Dezember 1937 zuordnen zu können. Die verwendeten Listen sahen

eine solche Zuordnung nicht vor, verfügten jedoch häufig über Geokoordinaten. Daher erfolgte die Zuweisung mithilfe von Geokoordinaten und der Verwendung einer Shape-Datei (Format für vektorielle Geodaten), welche die Grenzen des Deutschen Reiches im Jahr 1937 abbildet.⁸ Über diese Zuweisung mithilfe von Geokoordinaten war es möglich, umfangreiche Quellen wie OpenStreetMap oder das MiniGOV zu verwenden, die häufig neben den aktuellen Ortsbezeichnungen auch historische deutsche Schreibweisen enthalten. Allerdings können durch die Zuweisung auf Basis von Geokoordinaten besonders in den Randgebieten Unschärfen entstehen, die manuell bearbeitet und überprüft wurden. Einer damit gegebenenfalls verbleibenden geringen Zahl an Fehlzuweisungen steht jedoch eine nahezu vollständige Liste aller Orte in den Grenzen von 1937 in aktuellen und historischen Schreibweisen gegenüber, die auf andere Weise nicht erstellt werden kann.

5

Zuordnung von Geburtsorten

5.1 Regeln und Prioritäten-Listen für Referenzverzeichnisse und mitgelieferte Merkmale

Einigen Geburtsorten wurde bereits im Zuge der Datenaufbereitung ein Staat zugewiesen (siehe Kapitel 4). Alle anderen Orte durchliefen nach der Datenaufbereitung eine Reihe von Prozessen mit dem Ziel, über unterschiedliche Abgleiche und Prozeduren einen Geburtsstaat zuzuweisen. Alle Zuordnungsschritte folgten dabei den für den jeweiligen Schritt entwickelten Regeln. Diese Regeln wurden nach Erstellung im Produktivlauf getestet, anhand der Echtdaten evaluiert und in einem fortwährenden Prozess weiterentwickelt beziehungsweise angepasst. Ziel war, einen standardisierten Ablauf für die Zuweisung eines Geburtsstaates zu schaffen.

Eine wichtige Rolle bei den Zuordnungen spielte die oben beschriebene, aus den Referenzverzeichnissen

und dem Ortsverzeichnis des Zensus 2011 zusammengeführte Leitdatei. Für die folgenden Zuordnungsschritte wurden die Orts-Staats-Kombinationen den Quellen nach in zwei Gruppen priorisiert. Die erste Gruppe von Quellen mit höherer Priorität umfasste qualitätsgeprüfte Listen öffentlicher Institutionen wie das GV-ISys oder die LAU-Liste. Die zweite Gruppe von Quellen mit niedrigerer Priorität umfasste Listen, die nicht vergleichbaren Qualitätsprüfungen oder Routinen unterlagen und daher auch nicht korrekte Orts-Staats-Kombinationen enthielten. Dazu zählten die Listen der Orte aus OpenStreetMap oder aus dem Ortsverzeichnis der KOSIS-Gemeinschaft HHSTAT.

5.2 Abgleich mit der Leitdatei

Nach den oben erläuterten Datenaufbereitungsschritten erfolgte ein Abgleich der Geburtsortsangaben aus den Melderegistern mit dem vorliegenden oder bereits aktualisierten Ortsverzeichnis. Hierbei wurden nur solche Geburtsorte einem Staat zugewiesen, bei denen Melderegisterangabe und Ortsverzeichnis exakt übereinstimmten. Diese Abstimmung fand in mehreren Schritten mit je zwei Dimensionen statt. Die erste war die zugeordnete Priorität der Quelle, aus der die Orts-Staats-Kombination in das Ortsverzeichnis übernommen wurde, die zweite war die Eindeutigkeit, in der ein Ort im Ortsverzeichnis vorhanden war. Es wurde unterschieden zwischen eindeutigen Übereinstimmungen (zum Beispiel Buxtehude) und mehrdeutigen Übereinstimmungen (zum Beispiel Rom, ein Ort, den es in mehreren Ländern gibt).

Bei eindeutiger Übereinstimmung eines Geburtsortes aus den Melderegistern mit einem Ort des Ortsverzeichnisses aus einer Quelle mit zugeordneter hoher Priorität wurde zumeist der aus dem Ortsverzeichnis eingetragene Staat als Geburtsstaat übernommen. Diese Regel galt unabhängig davon, ob der im Melderegister eingetragene „Geburtsort – Staat“ diesem entsprach. Ausnahmen davon bildeten lediglich Fälle, in denen ein im Melderegister eingetragener Geburtsstaat, eine Staatsangehörigkeit und ein Zugangsstaat übereinstimmten, dem vorliegenden Ortsverzeichnis aber „widersprachen“. Dann wurde der im Melderegister eingetragene Staat als Geburtsstaat beibehalten und als Orts-Staats-Kombination mit in das Ortsverzeichnis aufgenommen. Traf der eindeutige Abgleich des Ortes aus dem Melderegister einen Ort des Ortsverzeichnisses aus einer

⁸ Verwendet wurde eine Shape-Datei des Mosaic Projects, welches historische Geodaten für wissenschaftliche Zwecke zur Verfügung stellt, und zwar unter: mosaic.ipums.org

Übersicht 1

Regelwerk für Zuweisungen des Geburtsstaates bei eindeutiger Übereinstimmung des Geburtsortes mit Eintrag aus dem Ortsverzeichnis aus Quelle mit niedrigerer Priorität

Daten aus Melderegister					Ortsverzeichnis			Erfolgte Zuweisung Geburtsstaat
Geburtsort aus Melderegister	Geburtsort – Staat	1. Staatsangehörigkeit	2. und/oder 3. Staatsangehörigkeit	Zuzugsstaat	Ort aus Ortsverzeichnis	Priorität der Quelle des Eintrags	Staat	
Ortsname	xxx	beliebige Angaben			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	yyy	wenn mehr Angaben mit Ausprägung yyy als Angaben mit Ausprägung xxx			Ortsname	niedrigere Priorität	xxx	yyy
Ortsname	yyy	wenn gleich viele Angaben mit Ausprägung xxx/yyy			Ortsname	niedrigere Priorität	xxx	yyy
Ortsname	yyy	wenn zwei Angaben mit Ausprägung xxx und eine weitere Angabe mit Ausprägung yyy			Ortsname	niedrigere Priorität	xxx	yyy
Ortsname	yyy	wenn zwei Angaben mit Ausprägung xxx und keine weitere Angabe mit Ausprägung yyy			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	leer, ungültig	beliebige Angaben			Ortsname	niedrigere Priorität	xxx	xxx

xxx und yyy stehen hier für beliebige gültige Staatencodes (immer dreistellig), die entweder übereinstimmen oder sich unterscheiden. – Hervorgehoben sind die Fälle, in denen der zugewiesene Geburtsstaat (letzte Spalte) nicht dem Staat aus dem Ortsverzeichnis (vorletzte Spalte) entspricht.

Quelle mit niedrigerer Priorität, wurde der eingetragene Staat als Geburtsstaat in das Ortsverzeichnis übernommen, sofern beide Einträge denselben Geburtsstaat betrafen. War im Melderegister kein Geburtsort – Staat eingetragen, wurde der Staat aus dem Ortsverzeichnis als Geburtsstaat übernommen. War dagegen im Melderegister zu dem Geburtsort ein anderer Geburtsstaat eingetragen als im Ortsverzeichnis zu dem Ort vermerkt war, folgte die Zuweisung einem komplexeren Regelwerk,

das die Merkmale Geburtsort – Staat, Staatsangehörigkeit und Zuzugsstaat berücksichtigte. ➤ [Übersicht 1](#)

Bei Übereinstimmung eines Geburtsortes aus den Melderegistern mit mehreren Einträgen aus dem Ortsverzeichnis wurde Übereinstimmungen mit Orten, die aus einer Quelle mit höherer Priorität stammten, eine größere Bedeutung beigemessen. In beiden Fällen wurde bei Übereinstimmung der Angaben des Ortsverzeichnisses mit dem Geburtsort – Staat des Melderegisters

Übersicht 2

Regelwerk für Zuweisungen des Geburtsstaates bei mehrdeutiger Übereinstimmung des Geburtsortes mit Eintrag aus dem Ortsverzeichnis aus Quelle mit niedrigerer Priorität

Daten aus Melderegister					Ortsverzeichnis			Erfolgte Zuweisung Geburtsstaat
Geburtsort aus Melderegister	Geburtsort – Staat	1. Staatsangehörigkeit	2. und/oder 3. Staatsangehörigkeit	Zuzugsstaat	Ort aus Ortsverzeichnis	Priorität der Quelle des Eintrags	Staat	
Ortsname	xxx	beliebige Angaben			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	yyy	wenn mehr Angaben mit Ausprägung yyy als Angaben mit Ausprägung xxx			Ortsname	niedrigere Priorität	xxx	yyy
Ortsname	yyy	wenn gleich viele Angaben mit Ausprägung xxx/yyy			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	yyy	wenn zwei Angaben mit Ausprägung xxx und eine weitere Angabe mit Ausprägung yyy			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	yyy	wenn zwei Angaben mit Ausprägung xxx und keine weitere Angabe mit Ausprägung yyy			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	leer, ungültig	mindestens eine Angabe mit Ausprägung xxx			Ortsname	niedrigere Priorität	xxx	xxx
Ortsname	leer, ungültig	keine Ausprägung mit Angabe xxx (leer, ungültig, yyy)			Ortsname	niedrigere Priorität	xxx	999

xxx und yyy stehen hier für beliebige gültige Staatencodes (immer dreistellig), die entweder übereinstimmen oder sich unterscheiden. – Hervorgehoben sind die Fälle, in denen der zugewiesene Geburtsstaat (letzte Spalte) nicht dem Staat aus dem Ortsverzeichnis (vorletzte Spalte) entspricht.

dieser Staat auch zugewiesen. Gab es im Melderegister keinen Eintrag auf dem Merkmal Geburtsort – Staat oder deutete dieser auf einen anderen Staat hin als das Ortsverzeichnis, wurde kontrolliert, ob eine weitere Angabe des Melderegisters (Staatsangehörigkeit oder Zuzugsstaat) mit der Angabe des Ortsverzeichnisses übereinstimmte oder nicht. Auch hier wurde ein Regelwerk angewendet, das sich aber aufgrund der Mehrdeutigkeit der möglichen Zuweisungen von den oben dargestellten Regeln unterscheidet. [↘ Übersicht 2](#)

5.3 Fehlertolerante Suche in der Leitdatei

Bei diesem Schritt werden die Geburtsorte, denen in den vorherigen Schritten noch kein Staat zugewiesen werden konnte, in einem fehlertoleranten Verfahren mit der Leitdatei abgeglichen. Dies bedeutet, dass für einen Geburtsort aus dem Melderegister ein Staat aus einem Eintrag der Leitdatei als Zuordnung übernommen wird, obwohl keine exakte Übereinstimmung, jedoch eine hinreichende Ähnlichkeit zwischen dem Geburtsort und einem Eintrag in der Leitdatei existiert. Die Gründe für nicht übereinstimmende Angaben können sehr unterschiedlich sein. Eine erste Möglichkeit war, dass der Geburtsort des Melderegisters neben einem Ortsnamen aus der Leitdatei auch andere Angaben enthielt. Dies ist zum Beispiel der Fall, wenn der Geburtsort neben einem Gemeinamen weitere Angaben enthält, die nicht über die oben beschriebenen Verfahren identifiziert und gelöscht werden konnten. Weitere Gründe wie Falschschreibweisen von Gemeinden durch Vertausch von Buchstaben oder die Übernahme von Schreibweisen aus anderen Sprachen sind ebenfalls denkbar.

Bei fehlertoleranten Abgleichen kann die Anzahl der durchzuführenden Vergleiche schnell sehr groß werden, was zu sehr hohen Programmlaufzeiten führt. Um diese zu reduzieren, wurden die Suchräume eingegrenzt.⁹ Damit dieser Schritt möglichst viele Zuordnungen bewirkt, werden sukzessive verschiedene Blocking-Strategien angewandt. Die Ähnlichkeit des Geburtsortes aus

⁹ Beispielsweise durch die phonetischen Algorithmen Soundex und Kölner Phonetik (Postel, 1969) sowie zum Teil durch weitere Bedingungen zu Angaben mit Bezug zu Staaten, also Geburtsstaat, Staatsangehörigkeit beziehungsweise Zuzugsstaat. Da außerhalb der definierten Suchräume keine Datensätze verglichen werden, ist bei der Größe der Suchräume zwischen Programmlaufzeit und der Wahrscheinlichkeit, dass zusammengehörige Datensätze nicht dem gleichen Suchraum zugeordnet werden, abzuwägen.

dem Melderegister mit einem Eintrag aus der Leitdatei (wenn beide im gleichen Block liegen) wird über das Jaro-Winkler-Distanzmaß (Winkler, 1990) quantifiziert und eine Zuordnung nur durchgeführt, wenn der Jaro-Winkler-Wert über 0.96 lag. Sollten verschiedene Einträge aus der Leitdatei eine hohe Ähnlichkeit aufweisen, wurde der Geburtsstaat aus dem zugrunde liegenden Eintrag der Leitdatei, für den der Jaro-Winkler-Wert am höchsten ist, zugewiesen. Dieser Schritt wurde zudem in drei Unterschritten aufgeteilt. Zuerst wurden die standardisierten Geburtsorte miteinander verglichen. Im Anschluss erfolgte der Abgleich auf lexikografisch aufsteigend beziehungsweise absteigend sortierten, standardisierten Geburtsorten. Zuletzt wurden nur Teile der standardisierten Geburtsorte miteinander verglichen.

5.4 Fehlertolerante Suche innerhalb der Melderegister-Daten

Neben dem zuvor beschriebenen fehlertoleranten Abgleich mit der Leitdatei wurde eine fehlertolerante Suche innerhalb der Geburtsorte in den Melderegisterdaten, denen bereits ein Geburtsstaat zugewiesen werden konnte, durchgeführt. Auf diese Weise konnte Geburtsorten, die zwar nicht in der Leitdatei gefunden werden konnten, jedoch eine hohe Ähnlichkeit mit Orten innerhalb der Melderegister mit bereits plausibilisiertem Geburtsstaat aufweisen, ebenfalls ein Staat zugeordnet werden. Der Abgleich zwischen den verschiedenen Geburtsortseinträgen erfolgte dabei ähnlich wie im vorherigen Schritt beschrieben mithilfe von Blockingvarianten und dem Jaro-Winkler-Distanzmaß.

5.5 GeoNames

Die verbleibende Menge der durch vorherige Schritte nicht gefundenen Geburtsorte wurde zudem in einem separaten Schritt mit der GeoNames-Datenbank abgeglichen. Die geografische Datenbank von GeoNames umfasst alle Länder weltweit und enthält über zwölf Millionen Ortsnamen, die kostenlos heruntergeladen werden können. Bei dem Abgleich wurden auch alle in der Datenbank hinterlegten Alternativschreibweisen mit einbezogen. Da gerade bei den Alternativschreibweisen auch offensichtlich falsche, umgangssprachliche oder unplausible Einträge vorhanden sind, wurde ein Treffer nur bei Übereinstimmung mit dem gelieferten Geburts-

staat oder der ersten Staatsangehörigkeit akzeptiert. Aufgrund des confirmatorischen Charakters des Vorgehens sind Fehlzusweisungen nahezu ausgeschlossen, die Ergebnisse wurden dennoch stichprobenartig manuell überprüft.

5.6 Manuelle Zuordnung

Auch nach den vielfältigen Schritten und Routinen zur Zuordnung konnte nicht allen Orten ein standardisierter Geburtsstaat zugeschrieben werden. Diese Fälle wurden einzeln mittels verschiedener Quellen wie GeoNames, Wikipedia oder Google recherchiert. Dabei wurden die mitgelieferten Informationen im Melderegister mit einbezogen, um so Hinweise auf einen potenziellen Geburtsstaat zu erhalten. Die Reihenfolge der manuell zu recherchierenden Geburtsorte richtete sich nach der Häufigkeit ihres Vorkommens. Zunächst wurden also die am häufigsten vorkommenden Geburtsorte bearbeitet. Ab einem bestimmten Zeitpunkt standen hinter den zu recherchierenden Ortsnamen so wenige Personen (weniger als zehn), dass der Aufwand der manuellen Recherchen in keinem vertretbaren Verhältnis zur qualitativen Verbesserung des Ortsverzeichnisses stand. Mehrheitlich handelte es sich hier um Einzelfälle, bei denen (stichprobenartig) selbst die Recherche keine klare Zuordnung ergab. Die Ergebnisse der manuellen Liste wurden dann als ein Teil der Leitdatei aufgenommen, bildeten also ein eigenes Referenzverzeichnis. Manuelle Listen wurden nach jeder Datenlieferung erstellt und bearbeitet und standen so beim nächsten Durchlauf der Zuordnungsprozesse direkt zur Zuordnung zur Verfügung.

5.7 Imputation

Nicht für jeden Datensatz war es möglich, den Geburtsstaat über den Geburtsort zu bestimmen, da dieser zum Beispiel trotz manueller Recherche nicht gefunden werden konnte. Lag für diese Datensätze ein gültiger Geburtsstaat oder eine erste Staatsangehörigkeit vor, wurde dieser Eintrag übernommen.

Falls eine Person beispielsweise in Wladiwostok zur Zeit der Sowjetunion geboren wurde, wurde als Geburtsstaat nach aktuellem Gebietsstand der Nachfolgestaat Russische Föderation zugewiesen, da sich dieser zweifelsfrei zuordnen ließ. Dieses Vorgehen fand analog für alle nicht mehr existenten Staaten Anwendung. Ließ

sich der Geburtsort nicht eindeutig einem Nachfolgestaat zuordnen, wurde ein Geburtsstaat anhand einer Verteilung imputiert. Die Verteilung auf die Nachfolgestaaten wurde dabei aus den Daten selbst über die zuvor eindeutig zugeordneten Datensätze gewonnen.

Ungültige Staatenschlüssel, die auch durch Imputation nicht einem gültigen Staatenschlüssel zugeordnet werden konnten, wurden auf den Ersatzwert „999“ gesetzt.

6

Ergebnisse der Geburtsstaatsplausibilisierung im Zensus 2022

Für die Aktualisierung des Ortsverzeichnisses konnten insgesamt vier Datenlieferungen der Melderegister für den Zensus 2022 verwendet werden. Hierdurch war es bereits frühzeitig möglich, Programmabläufe zu testen und zu optimieren sowie rechtzeitig mit manuellen Recherchearbeiten zu beginnen.

Final wurden durch das Ortsverzeichnis im Zensus 2022 rund 94 Millionen Personendatensätze verarbeitet. Nach der Aggregation über die Merkmale Geburtsort*Geburtsort-Staat*Staatsangehörigkeiten*Zuzugsstaat*Geburtsdatum(binär) blieben 5 154 165 unterschiedliche Kombinationen übrig.¹⁰ [↘ Tabelle 1](#) bildet die prozentualen Zuordnungen der wichtigsten Zuordnungsschritte zu einem Geburtsstaat auf aggregierter Ebene ab.

Durch die manuellen Recherchen wurden 0,32% der Personendatensätze einem Geburtsstaat zugeordnet. Unter die Kategorie „übrige“ fallen Zuordnungen zum Beispiel über GeoNames oder Imputation durch Hinweise. Insgesamt wurden nach dem kompletten Durchlauf des Ortsverzeichnisses etwa 855 000 Personendatensätze dem Staatencode „999“ zugeordnet, bei ihnen konnte keine Angabe eines Geburtsstaates erfolgen.

¹⁰ Die ermittelten Kombinationen weisen eine extrem ungleiche Verteilung hinsichtlich der Besetzungszahlen auf. Etwa 70% der Kombinationen enthalten nur eine Person, während einige wenige Kombinationen sehr häufig vorkommen.

Die Aktualisierung des Ortsverzeichnisses und Plausibilisierung des Geburtsstaates im Zensus 2022

Tabelle 1

Zuordnungen der wichtigsten Zuordnungsschritte zu einem Geburtsstaat auf aggregierter Ebene im Ortsverzeichnis für den Zensus 2022

Geburtsort*Geburtsstaat*EF25*EF83* Geburtsdatum(binär)	Merkmal kombinationen	Personendatensätze
	%	
als „Unbekannt“ deklarierte Geburtsorte	0,27	0,96 ¹
Staaten- und Gebietsangaben innerhalb des Merkmals „Geburtsort“	6,78	4,71
Leitdatei – 1:1-Abgleich – einfacher Treffer	37,58	29,24
Leitdatei – 1:1-Abgleich – mehrfacher Treffer	29,18	62,19
Fehlertolerante Suche in Leitdatei – deutsche Orte	3,96	0,52
Fehlertolerante Suche in Leitdatei – ausländische Orte	6,86	0,57
Fehlertolerante Suche innerhalb der Melderegister-Daten	7,17	0,99
Manuelle Zuordnung	1,01	0,32
Übrige Zuordnungen	7,18	0,50

¹ Nicht alle als „Unbekannt“ deklarierten Geburtsorte führen zu einem unbekanntem Geburtsstaat (Staatencode „999“). Mitgelieferte Informationen zu Geburtsstaat und/oder Staatsangehörigkeit und Zuzugsstaat können Hinweise auf einen Geburtsstaat liefern.

7

Fazit

Die lücken- und in vielen Fällen fehlerhafte Befüllung des Merkmals Geburtsstaat in den Melderegisterangaben machte eine Plausibilisierung des Merkmals über den gelieferten Geburtsort notwendig. Da der Geburtsort als Freitextfeld mit schwankender Qualität befüllt ist, war zunächst eine umfangreiche Aufbereitung und danach eine schrittweise Zuordnung notwendig. Durch die beschriebene Methodik der Zuweisung und die Weiterentwicklung des Ortsverzeichnisses verblieb im Zensus 2022 nur eine geringe Restmenge an Datensätzen, denen kein Geburtsstaat zugewiesen werden konnte. Über die vier Datenlieferungen im Zeitraum von 1,5 Jahren ließ sich dabei eine stetige Verbesserung der Datenqualität der Melderegister erkennen. Die im Rahmen der Arbeiten entwickelte Zuweisungslogik weist ein hohes Nachnutzungspotenzial für Folgeprojekte wie den Registerzensus auf. Für die künftige Weiterentwicklung des Ortsverzeichnisses wird zur Steigerung der Datenqualität eine weitere Vereinheitlichung von Schreibweisen angestrebt. 

LITERATURVERZEICHNIS

Postel, Hans Joachim. *Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse*. In: IBM-Nachrichten. 19. Jahrgang. 1969. Seite 925 ff.

Statistisches Bundesamt. *Staats- und Gebietssystematik. Stand: August 2024*. [Zugriff am 24. Oktober 2024]. Verfügbar unter: www.destatis.de

Weißer, Claudia. *Auswertung der Geburtsorte im Zensus 2011*. In: Wirtschaft und Statistik. Ausgabe 10/2012, Seite 843 ff.

Winkler, William E. *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage*. 1990.

RECHTSGRUNDLAGEN

Durchführungsverordnung (EU) 2017/543 der Kommission vom 22. März 2017 zur Festlegung der Regeln für die Anwendung 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 (Amtsblatt der EU Nr. L 78, Seite 13).

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Grundgesetz (GG) für die Bundesrepublik Deutschland in der im Bundesgesetzblatt Teil III, Gliederungsnummer 100-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 1 des Gesetzes vom 19. Dezember 2022 (BGBl. I Seite 2478) geändert worden ist.

UMSETZUNG UND POTENZIALE DES KONZEPTS „EINWANDERUNGSGESCHICHTE“ IM ZENSUS 2022

Oliver Kroppach, Stefan Rose

↳ **Schlüsselwörter:** Registerdaten – Migration – Eingewanderte – kleinräumige Gliederung – Bevölkerungsstruktur

ZUSAMMENFASSUNG

Die Erfassung der Einwanderungsgeschichte im Zensus 2022 in Deutschland basiert erstmals vollständig auf Registerdaten. Dieser Ansatz ermöglicht eine kontinuierliche Erhebung ohne die Bevölkerung zu befragen. Dabei zeigte sich, dass die Registerdaten für die Typisierung der Einwanderungsgeschichte, insbesondere bei der Nachkommenschaft von Eingewanderten, unvollständig sind. Vergleiche mit dem Mikrozensus belegen, dass die Zensusergebnisse im Gegensatz zur Bevölkerungszahl und zur Nationalität das abgeleitete Merkmal „Nachkommen von Eingewanderten“ systematisch unterschätzen, was zu analytischen Herausforderungen führt. Der Zensus bietet im Gegenzug einen klaren Mehrwert durch seine kleinräumige Gliederung und ermöglicht so räumlich differenzierte Analysen der Verteilung von Einwanderergruppen.

↳ **Keywords:** register data – migration – immigrants – small-area breakdown – population structure

ABSTRACT

In Germany's 2022 Census, information on immigration history was captured entirely on the basis of register-based data for the first time. This approach enables continuous data collection without surveying the population. It has become clear, however, that the register data are incomplete for classifying immigration history, particularly with regard to the descendants of immigrants. Comparisons with the microcensus show that, in contrast to the population figure and nationality, the census results systematically underestimate the derived characteristic "descendants of immigrants", which leads to analytical challenges. Nevertheless, the census provides significant added value through its detailed geographical breakdown, thereby enabling nuanced analyses of the distribution of immigrant groups.



Oliver Kroppach

hat Soziologie und Management an der Johannes Gutenberg-Universität Mainz studiert. Er ist seit 2023 wissenschaftlicher Mitarbeiter des Statistischen Bundesamtes in der Gruppe „Registerbasierter Datenbestand des Zensus“ im Referat „Methodik, Haushalte“. Seine Haupttätigkeiten betreffen die Bereiche Migration, Demografie, Haushalte und Familien.



Stefan Rose

studierte Sozialwissenschaften in Erfurt und Siegen und ist seit 2018 im Statistischen Bundesamt tätig. Seine Aufgabenschwerpunkte als wissenschaftlicher Mitarbeiter in der Gruppe „Registerbasierter Datenbestand des Zensus“ innerhalb des Referates „Methodik, Haushalte“ liegen in den Bereichen Migration, Demografie, Haushalte und Familien.

1

Einleitung

Registerdaten zu nutzen wird zunehmend als zukunftsweisende Methode für Zensuserhebungen anerkannt. Diese Erhebungsmethode greift auf bereits existierende Verwaltungsdaten zurück, die in verschiedenen öffentlichen Registern gespeichert sind, und erfasst so detailliert und kontinuierlich gesellschaftliche Veränderungen. Mit dem Übergang zu registergestützten Daten können Zensusdaten in kürzeren Zeitintervallen aktualisiert werden. Dies erlaubt eine schnellere und präzisere Reaktion auf demografische Entwicklungen, ohne die Bevölkerung regelmäßig zu befragen (Söllner/Körner, 2022, hier: Seite 15 ff.). Dementsprechend ist es zunehmend wichtig, Informationen zur Bevölkerungsstruktur rein aus den Melderegistern zur Verfügung zu stellen (Baffour und andere, 2013, hier: Seite 417 ff.). So wurde das im Zensus 2022 erstmals angewandte Konzept „Einwanderungsgeschichte“ nur mit registergestützten Daten umgesetzt.

Informationen zur Migration wurden nicht im Rahmen der Haushaltsstichprobe des Zensus erhoben. Für die Berücksichtigung dieses Themenfeldes in der Befragung gab es weder durch die Europäische Union (EU) Vorgaben noch durch das Zensusgesetz rechtliche Grundlagen. Stattdessen wurden Merkmale aus den Melderegistern genutzt, die laut Zensusgesetz 2022 Teil der Datenlieferung an das Statistische Bundesamt sind. Bei der Nutzung von Registerdaten ist unter anderem zu beachten, dass diese für einen anderen Zweck als statistische Untersuchungen produziert werden und daher häufig die eigentliche Fragestellung nicht exakt abbilden können. Trotzdem war es möglich, für die Klassifizierung der Einwanderungsgeschichte auf das Geburtsland und das Jahr der Ankunft im Meldeland zurückzugreifen. Aufgrund des öffentlichen Interesses zum Thema Migration werden mit diesen Merkmalen Informationen zur Einwanderung bereitgestellt.

Kapitel 2 fasst die Entwicklung des Konzepts der Einwanderungsgeschichte zusammen, während Kapitel 3 dessen methodische Umsetzung im Zensus 2022 ausführt. Wie sich die Nutzung des Konzepts der Einwanderungsgeschichte im Zensus von der im Mikrozensus unterscheidet, ist in Kapitel 4 dargestellt. Die Vorteile der registerbasierten Erhebung erläutert Kapitel 5 anhand

der kleinräumigen Gliederung in Berlin. Abschließend enthält das letzte Kapitel Empfehlungen für die künftige Umsetzung des Konzepts Einwanderungsgeschichte.

2

Entwicklung des Konzepts der Einwanderungsgeschichte

Bis in die 1990er-Jahre erfasste die amtliche Statistik die Bevölkerung mit Migrationsbezug hauptsächlich über die Kategorie „ausländische Bevölkerung“. Diese Klassifizierung bezog sich ausschließlich auf Menschen ohne deutsche Staatsangehörigkeit. Das führte zu analytischen Problemen, da ein Teil der Migrantinnen und Migranten und deren Nachkommen nicht erfasst wurde, sowie zu normativen Problemen, da der Begriff „Ausländer“ eine rein rechtliche Zuschreibung ist und nicht die Migrationserfahrungen oder Identitäten der betroffenen Personen abbildet. Um diese Ungenauigkeit zu lösen, entwickelte das Statistische Bundesamt im Mikrozensus das Konzept des Migrationshintergrunds. Der Migrationshintergrund wurde nicht direkt erfasst, sondern anhand von Informationen zur Staatsangehörigkeit, Einbürgerung und dem Geburtsland der befragten Person sowie ihrer Eltern gebildet. Als Personen mit Migrationshintergrund gelten alle in Deutschland lebenden Personen, die selbst oder deren Eltern nicht durch Geburt die deutsche Staatsangehörigkeit besitzen. Erstmals wurden im Berichtsjahr 2005 entsprechende Zahlen veröffentlicht.

Das Konzept des Migrationshintergrunds ist in der Wissenschaft und der Öffentlichkeit sowohl analytisch als auch normativ umstritten. Auf analytischer Ebene wird kritisiert, dass es die Staatsangehörigkeit mit Migrationserfahrungen vermischt und eine sehr heterogene Bevölkerungsgruppe erfasst. Es differenziert nicht ausreichend zwischen den verschiedenen Generationen von Eingewanderten und deren Nachkommen, was es erschwert, soziostrukturelle Prozesse der Integration zu analysieren. Auch wird bemängelt, dass das Konzept komplex konstruiert ist, da es verschiedene Personengruppen und Sachverhalte kombiniert, was zu Unschärfen führen kann. Auf normativer Ebene gibt es Kritik daran, dass der Begriff Migrationshintergrund häufig nicht der Selbstwahrnehmung der betroffenen Personen

entspricht. Insbesondere Nachkommen von Eingewanderten mit deutscher Staatsangehörigkeit betrachten sich häufig nicht als Menschen mit Migrationshintergrund. Der Begriff wird als stigmatisierend empfunden und übernimmt oft die negative Konnotation des früheren „Ausländer“-Begriffs (Will, 2020; Fachkommission Integrationsfähigkeit, 2021, hier: Seite 220 ff.).

Im Januar 2021 legte die Fachkommission zu den Rahmenbedingungen der Integrationsfähigkeit (im Weiteren kurz: Fachkommission Integrationsfähigkeit) ihren Bericht „Gemeinsam die Einwanderungsgesellschaft gestalten“ vor. Darin empfiehlt die Kommission, den Begriff Migrationshintergrund durch das Konzept der „Eingewanderten und ihrer (direkten) Nachkommen“ zu ersetzen. Die Neudefinition umfasst Menschen, die entweder selbst nach 1950 in das heutige Deutschland eingewandert sind (Eingewanderte) oder deren beide Elternteile ab dieser Zeit zugezogen sind (Nachkommen von Eingewanderten) (Fachkommission Integrationsfähigkeit, 2021, hier: Seite 222 f.). Damit unterscheidet die Definition klar zwischen der ersten Generation, die selbst eingewandert ist, und der zweiten Generation, die in Deutschland geboren wurde. Ab der dritten Generation wird diese Einordnung nicht mehr vorgenommen. Zudem wurde das Kriterium der Einwanderung seit 1950 zuvor nicht auf Vertriebene und Aussiedler angewendet, während es im neuen Konzept auf alle Gruppen ausgedehnt wird. Der neue Begriff dient dazu, den bisherigen Diskurs zu vereinfachen und besser mit internationalen Definitionen vergleichen zu können. Weiterhin will die Kommission mit dieser Neudefinition Stigmatisierung vermeiden, die im Zusammenhang mit dem Begriff Migrationshintergrund entstehen kann. Der Fokus liegt nicht mehr auf der Staatsangehörigkeit, sondern auf der direkten Wanderungserfahrung einer Person.

Das Statistische Bundesamt hat sich mit dem Vorschlag der Fachkommission Integrationsfähigkeit auseinandergesetzt, um den Begriff „Eingewanderte und ihre (direkten) Nachkommen“ für die statistische Erfassung zu verwenden. Da der Vorschlag jedoch sprachlich sperrig ist und es keine geeignete Komplementärkategorie gibt, hat sich das Statistische Bundesamt für den Begriff „Einwanderungsgeschichte“ entschieden. Begrifflich werden nun klar abgegrenzte Kategorien eingeführt: „Eingewanderte“, „Nachkommen von Eingewanderten“ und die zusammenfassende Bezeichnung „Personen mit Einwanderungsgeschichte“. Personen mit nur einem

seit 1950 eingewanderten Elternteil gelten als „Personen mit einseitiger Einwanderungsgeschichte“, gehören jedoch nicht zur Gruppe der „Personen mit Einwanderungsgeschichte“. Personen, die vor 1950 eingewandert sind oder deren Elternteile nicht beide seit 1950 zugezogen sind, werden als „Personen ohne Einwanderungsgeschichte“ klassifiziert (Canan/Petschel, 2023, hier: Seite 65 ff.).

3

Methodische Umsetzung im Zensus 2022

Der Empfehlung der Fachkommission Integrationsfähigkeit und der Umsetzung im Mikrozensus folgend, übernimmt der Zensus 2022 das Konzept der Einwanderungsgeschichte und ersetzt damit das Konzept des Migrationshintergrunds aus dem Zensus 2011. Um die Einwanderungsgeschichte zu erfassen, werden Informationen zum Geburtsland, zum Zuzugsdatum sowie die Verknüpfung von Eltern mit ihren Kindern benötigt. Die Herausforderung bei der Umsetzung des Konzepts Einwanderungsgeschichte gemäß den Empfehlungen der Fachkommission Integrationsfähigkeit auf Basis registrierter Daten besteht darin, dass die erforderlichen Merkmale und Beziehungsinformationen unvollständig vorliegen.

Die Befüllungsrate des Zuzugsdatums in den Meldeeregistern ist gering, für die Geburtsjahrgänge von 1945 bis 1950 beträgt sie lediglich 14,6%. Daher ist eine zeitliche Nachverfolgung der Einwanderung für diese Kohorte nicht möglich. Dies hat zur Folge, dass es nicht möglich ist, die von der Fachkommission Integrationsfähigkeit definierte Datumsgrenze konsequent anzuwenden. Die unzureichende Datenlage macht auch eine präzise Imputation unmöglich und würde zu fehlerhaften Schätzungen und Verzerrungen in den Ergebnissen führen (Rubin, 1976). Vor 1950 geborene Personen können daher nicht verlässlich in Bezug auf ihre Einwanderungsgeschichte klassifiziert werden, da nicht nachvollzogen werden kann, ob sie vor oder nach 1950 eingewandert sind. Die Fachkommission Integrationsfähigkeit hat diese Datumsgrenze gezielt eingesetzt, um Einwanderungsbewegungen infolge des Zweiten Weltkriegs von der Erfassung auszu-

nehmen. Insbesondere Vertriebene aus ehemaligen Gebieten des Deutschen Reiches sollen dadurch nicht als Personen mit Einwanderungserfahrung klassifiziert werden. Allerdings werden durch diese Regelung auch Personen, die vor 1950 aus anderen Regionen eingewandert sind, nicht erfasst. Der Zensus 2022 verwendet daher den Geburtsstaat einer Person als Kriterium für die Typisierung der Einwanderungserfahrung. Personen, die vor dem 2. August 1945 geboren wurden, werden gemäß den Grenzen des Deutschen Reiches vom 31. Dezember 1937 typisiert. Personen, die einschließlich und nach dem 2. August 1945 geboren wurden, werden anhand der heutigen Grenzen der Bundesrepublik Deutschland klassifiziert. Dies ermöglicht es, dass Vertriebene, die nach dem 2. August 1945 geboren wurden, den Status „Person mit Einwanderungserfahrung beziehungsweise -geschichte“ erhalten können. Diese Vorgehensweise steht im Einklang mit der Definition der Fachkommission Integrationsfähigkeit, die für die Typisierung der Einwanderungsgeschichte ausschließlich die am Geburtsstaat erkennbare Einwanderungserfahrung und nicht die Staatsangehörigkeit heranzieht.

Die direkte Nachkommenschaft (ein- und zweiseitig) kann nur über die sogenannten Verzeigerungen, die Eltern-Kind-Beziehungen aus dem Melderegister, gewonnen werden. Allerdings liegen Informationen für Eltern-Kind-Beziehungen im Melderegister nur für Personen unter 19 Jahren vor. Verwandtschaftsbeziehungen und ein sich aus diesem Verhältnis abzuleitender Status kann nur für Personen gleich und unter 18 Jahren erfasst werden. Innerhalb der Haushaltegenerierung des Zensus 2022 werden anhand der Haushaltszusammenhänge auch Personenbeziehungen, wie die Eltern-Kind-Beziehung, plausibilisiert und neu erstellt. Nach der Haushaltegenerierung liegen deshalb neben den Registerinformationen auch statistisch generierte Eltern-Kind-Beziehungen vor. Mit diesen statistischen Beziehungen kann die direkte Nachkommenschaft für Personen über 18 Jahren gefunden werden. Diese Zuordnung gelingt allerdings nur, wenn diese Personen noch bei ihren Eltern im gleichen Haushalt leben und selbst keine Familie durch Heirat und/oder eigene Kinder gegründet haben. Würden auch diese Personen in die Auswertung aufgenommen, wären mehr Nachkommen von Eingewanderten zu identifizieren. Allerdings wäre gleichzeitig die Auswertung der Gruppe „Nachkommen von Eingewanderten“ verzerrt. So zeigt sich, dass die Bereitschaft oder die Notwendigkeit, nach der Volljährigkeit weiterhin bei

den Eltern zu leben, kulturell geprägt ist. Mittels eines Bonferroni-Gruppenvergleichs-Tests wurde zwischen den fünf größten Herkunftsstaaten jeweils ein signifikanter Unterschied des Auszugsalters der über 18-Jährigen mit den statistisch hergeleiteten Eltern gefunden. Beispielsweise leben Nachkommen von eingewanderten Personen aus Kasachstan im Durchschnitt acht Jahre länger bei ihren Eltern im Haushalt als die Nachkommen von Eingewanderten, die in Polen geboren wurden. Bei Berücksichtigung der umfangreichen Grundgesamtheit liegt hier eine mittlere bis große Effektstärke ($d=0,78$) nach Cohens D vor. Zwischen den Geschlechtern gibt es eine Differenz zwischen den durchschnittlich zuhause lebenden Frauen ($M=26,26$) und jener von Männern ($M=26,52$).¹ Daraus ergibt sich, dass die Auswertung von Nachkommen von Eingewanderten mit anderen Auswertungsmerkmalen verzerrte Ergebnisse liefern würde, da aus dem Zensusdatensatz keine Angaben zu den Nachkommen, die nicht mehr bei ihren Eltern leben, getroffen werden können. Wie bereits ausgeführt, werden deshalb nur Personen unter 19 Jahren ausgewertet.

Zusammenfassend lässt sich zur Umsetzung der Definition der Einwanderungsgeschichte im Zensus 2022 festhalten:

Eine Einwanderungsgeschichte haben Personen, die entweder selbst oder deren beide Elternteile nach Deutschland eingewandert sind, das heißt deren Geburtsstaat nicht Deutschland ist. Personen, die vor dem 2. August 1945 geboren wurden, werden anhand der Grenzen des Deutschen Reiches von 31. Dezember 1937 klassifiziert. Personen, die einschließlich und nach dem 2. August 1945 geboren wurden, werden anhand der Grenzen der heutigen Bundesrepublik Deutschland typisiert. Die Informationen zur Nachkommenschaft von Eingewanderten, unabhängig von beid- oder einseitiger Einwanderungsgeschichte, kann nur für Personen gleich und unter 18 Jahren ausgewiesen werden.

1 Die Abkürzung M steht für arithmetisches Mittel.

4

Vergleich mit der direkten Befragung im Mikrozensus

Beim Vergleich der auf Registerdaten basierenden Ergebnisse der Einwanderungsgeschichte des Zensus 2022 und den auf einer Primärerhebung beruhenden Zahlen des Mikrozensus 2022 wird deutlich, dass sich die Daten zur Bevölkerung mit Einwanderungsgeschichte voneinander unterscheiden. Während die Anzahl der Personen, die selbst eingewandert sind, in beiden Erhebungen weitgehend übereinstimmt, zeigen sich Abweichungen bei der Zuordnung des Merkmals „Nachkommen von Eingewanderten“.

Ein wesentlicher Faktor für diese Diskrepanz liegt darin, dass der Zensus lediglich Nachkommen ausweist, die zum Erhebungszeitpunkt höchstens 18 Jahre alt sind. Ältere Nachkommen werden hingegen in der Zenserhebung den Personen ohne Einwanderungsgeschichte zugeordnet. Die Merkmalsausprägung „Nachkommen von Eingewanderten“ wird somit systematisch unterschätzt. Da ein erheblicher Anteil der Nachkommen von Zuwanderern älter als 18 Jahre ist, bilden die Statistiken des Zensus diese Bevölkerungsgruppe nicht ab.

↘ **Tabelle 1**

Die Berechnung der Personen mit Einwanderungsgeschichte mittels Registerdaten ist also klar eingeschränkt verglichen mit einer Primärerhebung. So ermöglicht der Mikrozensus durch die korrekte Abbildung von Familien zusätzliche Auswertungen bezüglich der Haushalte und Familien mit Einwanderungsgeschichte (Keller/Körner, 2024, hier: Seite 41 ff.), während dies für den Zensus schwieriger ist. Der Mehrwert der Zensusdaten ergibt

Tabelle 1

Vergleich der Ergebnisse zur Einwanderungsgeschichte aus dem Zensus 2022 und dem Mikrozensus 2022

	Zensus 2022		Mikrozensus 2022	
	1 000	%	1 000	%
Bevölkerung insgesamt	82 720	100	83 097	100
Ohne Einwanderungsgeschichte	63 139	76,3	59 004	71,0
Mit Einwanderungsgeschichte	18 142	21,9	20 224	24,3
Eingewanderte	15 590	18,8	15 311	18,4
Nachkommen	2 552	3,1	4 913	5,9
Mit einseitiger Einwanderungsgeschichte	1 439	1,7	3 868	4,7

sich erst bei der kleinräumigen Differenzierung der Ergebnisse.

5

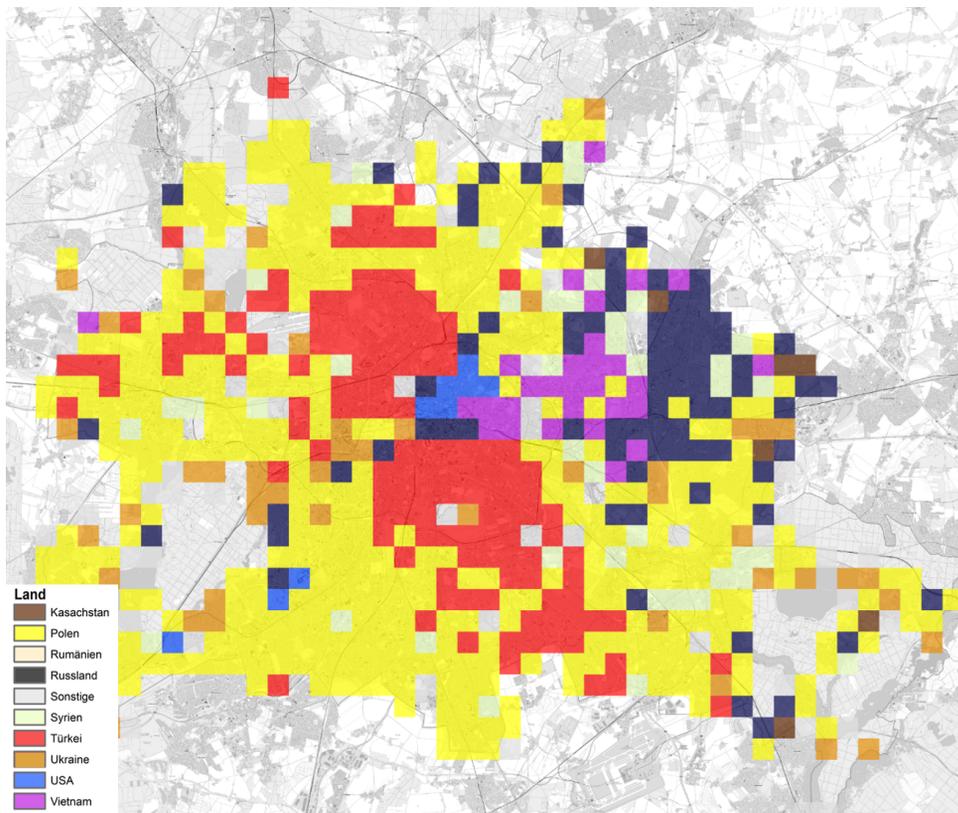
Die Relevanz des Zensus am Beispiel der Einwanderungsgeschichte in Berlin

Beispielhaft für die Vorteile, die die Erfassung der Einwanderungsgeschichte im Zensus 2022 bietet, steht die kleinräumige Gliederung in Berlin. Im Gegensatz zu einer Häufigkeitstabelle geben Gitterzellenkarten tiefere Einblicke in die Verteilungsstruktur von Personen mit Einwanderungsgeschichte.

↘ **Grafik 1** auf Seite 78 stellt die Stadt Berlin eingeteilt in 1 Quadratkilometer große Quadrate (Gitterzellen) dar. Jedes Quadrat zeigt das dort häufigste Geburtsland für alle Personen mit Einwanderungsgeschichte in der jeweiligen Gitterzelle. Die Grafik veranschaulicht die Verteilung der häufigsten Nationalitäten von Einwanderern in Berlin, wobei die verschiedenen Farben für unterschiedliche Geburtsländer stehen. Die Farben sind nicht gleichmäßig über die Karte verteilt, sondern bilden Cluster. Dies deutet auf eine räumliche Konzentration bestimmter Bevölkerungsgruppen hin. Gelb, das weite Teile Berlins dominiert, repräsentiert aus Polen eingewanderte Personen. Diese Gruppe ist insbesondere im westlichen und südlichen Teil der Stadt stark vertreten. Die türkischen Einwanderer (rot) sind ebenfalls in vielen Bereichen Berlins stark präsent, insbesondere in den zentralen und südöstlichen Bezirken. Es ist deutlich erkennbar, dass sich große türkische Gemeinschaften in diesen Gebieten etabliert haben, was die historische Zuwanderung von Gastarbeitern in die Stadt wider-

Grafik 1

Verteilung der häufigsten Nationalitäten von Einwanderern in Berlin in 1 Quadratkilometer großen Gitterzellen



Ergebnis des Zensus 2022.

spiegelt. Einwanderer aus Vietnam (violett) konzentrieren sich ebenfalls auf bestimmte Bezirke, vor allem im zentralen und östlichen Teil Berlins. Diese Verteilung könnte auf historische Migrationen zurückzuführen sein, insbesondere aus der Zeit vor 1990, als vietnamesische Vertragsarbeiter in der ehemaligen DDR und Berlin (Ost) lebten.

Bereits mit diesem Blick auf die Karte lässt sich ein guter Eindruck über die regionale Verteilung von Personen mit Einwanderungsgeschichte gewinnen. Für eine genauere Untersuchung des Zentrums Berlins wird die kleinräumige Gliederung auf 100 Quadratmeter in [Grafik 2](#) verkleinert. Auch hier fällt die Konzentration türkischer Einwanderer, dargestellt in Rot, vor allem im westlichen und südlichen Teil des Zentrums (Tiergarten, Kreuzberg) auf.

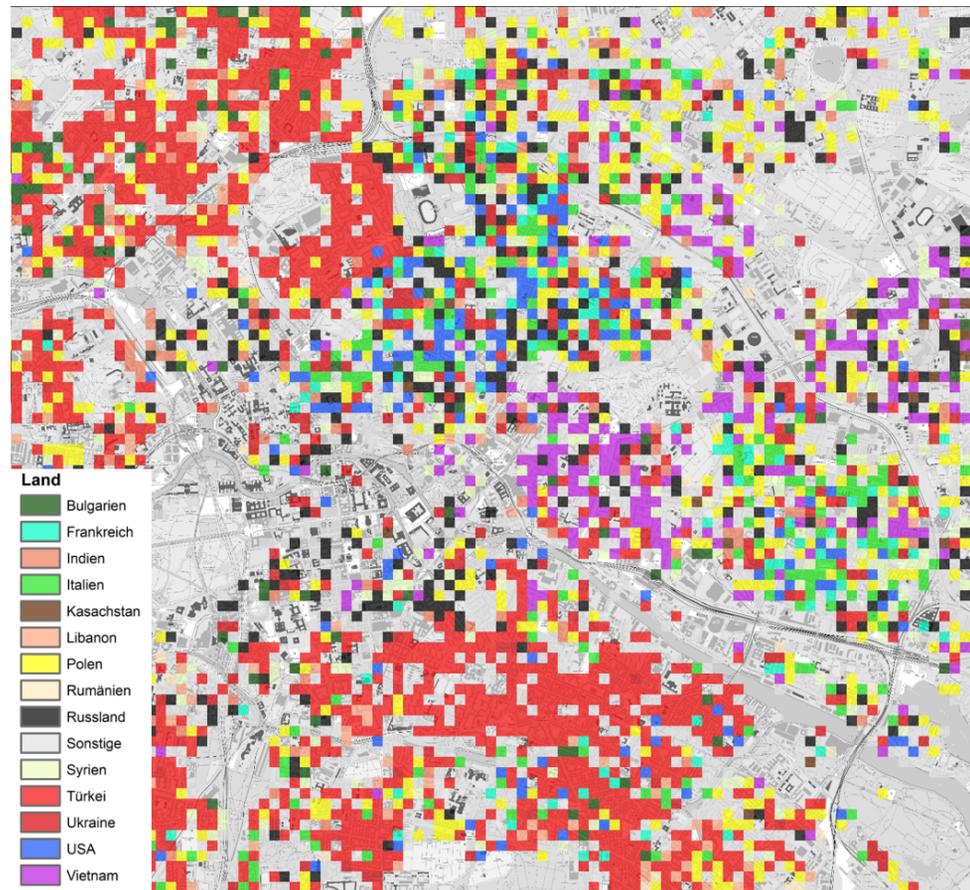
Vietnamesische Einwanderer, markiert durch lila Flächen, haben sich besonders im östlichen Teil des Zen-

trums (Friedrichshain) angesiedelt. Eine weitere stark vertretene Gruppe sind italienische Einwanderer, die durch grüne Flächen gekennzeichnet sind. Ihre Präsenz ist in einigen verstreuten, aber klar erkennbaren Bereichen des Stadtzentrums (Mitte) zu sehen. Vor allem die Gebiete in Mitte sind besonders vielfältig und weisen darauf hin, dass dort eine große Mischung verschiedener Einwanderergruppen lebt.

Die Karten verdeutlichen die Vorteile, die das Konzept der Einwanderungsgeschichte im Zensus bietet. Solche Analysen können zeigen, dass bestimmte Einwanderergruppen in bestimmten Gebieten besonders stark vertreten sind – beispielsweise durch die starke Konzentration der türkischen Gemeinschaft im westlichen und südlichen Zentrum sowie der russischen und vietnamesischen Gemeinschaften im östlichen Teil der Stadt. Derartige Unterschiede innerhalb einer einzigen Stadt lassen sich nur durch eine feinmaschige, kleinräu-

Grafik 2

Verteilung der häufigsten Nationalitäten von Einwanderern im Zentrum Berlins in 100 Quadratmeter großen Gitterzellen



Ergebnis des Zensus 2022.

mige Analyse erkennen. Diese kulturelle Vielfalt ist oft unsichtbar, wenn nur eine grobe Gliederung oder große administrative Einheiten betrachtet werden.

Eine kleinräumige Gliederung ermöglicht es Stadtplanern und politischen Entscheidungsträgern, spezifische Maßnahmen zur Integration und Förderung der unterschiedlichen Bevölkerungsgruppen zu entwickeln. In Stadtvierteln mit einer hohen Konzentration bestimmter Nationalitäten können beispielsweise gezielte Bildungs- und Sozialprogramme eingerichtet werden, die auf die Bedürfnisse dieser Gemeinschaften zugeschnitten sind. Ohne eine solch kleinteilige Betrachtung wäre es schwer, präzise Lösungen zu entwickeln, die sich wirklich an den Bedürfnissen der jeweiligen Nachbarschaften ausrichten. Beispielsweise zeigt die starke Konzentration der türkischen und vietnamesischen Gemeinschaften in

manchen Vierteln, dass diese Gruppen möglicherweise eine Art „ethnisches Cluster“ bilden, mit spezifischen Bedürfnissen und Herausforderungen.

Eine kleinräumige Gliederung ermöglicht zudem, Veränderungen in der Bevölkerungszusammensetzung besser nachverfolgen zu können. Migrationstrends, wie die zunehmende Zahl von Einwanderern aus bestimmten Ländern (zum Beispiel aus der Ukraine oder aus Syrien), lassen sich in kleineren Einheiten schneller erkennen und analysieren. Diese Informationen sind für Politik und Verwaltung notwendig, um auf aktuelle Herausforderungen zu reagieren und entsprechende Integrationsmaßnahmen rechtzeitig einzuleiten.

6

Empfehlungen für die künftige Umsetzung

Die im Zensus 2022 implementierte Klassifizierung zur Einwanderungsgeschichte stellt die einzige bundesweit einheitliche Datenquelle dar, die eine präzise Erfassung und Darstellung von Informationen zur Einwanderungsgeschichte in kleinräumig gegliederten Regionen ermöglicht. Es wird deswegen weiterhin wichtig bleiben, die Einwanderung in künftigen Zensus zu typisieren. Gleichzeitig führt die Umsetzung mit reinen Registerdaten dazu, dass die Nachkommenschaft von Personen mit Einwanderungserfahrung nicht für Personen über 18 Jahre erfasst werden kann. Die Vorgehensweise und die Auswertungsmöglichkeiten des Zensus 2022 sind von der Qualität der Melderegister und den rechtlichen Rahmenbedingungen abhängig. Die Melderegister erfassen die Informationen zur Nachkommenschaft für die Typisierung der Personen mit Einwanderungsgeschichte nur unzureichend. Dadurch wird das Merkmal Einwanderungsgeschichte systematisch unterschätzt. Auch konnte die von der Fachkommission Integrationsfähigkeit vorgeschlagene Datumsgrenze nicht berücksichtigt werden, weil das Zuzugsdatum bei älteren Personen nicht einheitlich befüllt ist.

Lösungen für diese Problematik wären immer damit verbunden, die Methodik im Zensus zu verändern. Die Einwanderungsgeschichte in einer Haushaltsstichprobe abzufragen wäre die sicherste Möglichkeit, sie im Zensus zu erfassen. Durch die direkte Befragung könnten zusätzliche Informationen zur Nachkommenschaft gewonnen und dann hochgerechnet werden. Diese Vorgehensweise wäre mit der Methodik des Mikrozensus vergleichbar, nur dass die Haushaltsstichprobe größer wäre und somit räumlich detailliertere Ergebnisse möglich wären – nicht vergleichbar jedoch mit den oben dargestellten räumlichen Auflösungen, die eine Registernutzung erlaubt. Ein tiefergehender Lösungsansatz besteht darin, die Melderegister anzupassen. Aktuell werden die sogenannten Verzeigerungen – also die Eltern-Kind-Verknüpfungen – mit dem Erreichen des 19. Lebensjahres gelöscht. Würde dies entfallen, könnten künftig deutlich mehr Nachkommen mit Einwanderungsgeschichte erfasst werden. Allerdings wäre auch dann ein Nachweis

verstorbener Eltern nicht möglich, da diese Information nicht in den Melderegistern vorhanden ist.

Mit den vorliegenden Daten des Zensus kann ausgewertet werden, inwiefern sich die Lebensumstände der Personen mit Einwanderungsgeschichte von denen der übrigen Bevölkerung unterscheiden, beispielsweise bezüglich der Familienform oder der Haushaltsgröße. Ein zentrales Motiv für die Einführung des Konzepts der Einwanderungsgeschichte ist die Empfehlung der Fachkommission Integrationsfähigkeit. Es wird ein einfaches und aktuelles Konzept eingeführt, das innerhalb der amtlichen Statistik bereits existiert und die Einwanderungsbiografie der Menschen in Deutschland berücksichtigt. [\[1\]](#)

LITERATURVERZEICHNIS

Baffour, Bernard/King, Thomas/Valente, Paolo. *The Modern Census: Evolution, Examples and Evaluation*. In: International Statistical Review. Band 81, Ausgabe 3/2013, Seite 407 ff. DOI: [10.1111/insr.12036](https://doi.org/10.1111/insr.12036)

Will, Anne-Kathrin. *Migrationshintergrund – wieso, woher, wohin?* Bundeszentrale für politische Bildung. 2020. [Zugriff am 24. Oktober 2024]. Verfügbar unter: www.bpb.de

Canan, Coşkun/Petschel, Anja. *Die Umsetzung des Konzepts „Einwanderungsgeschichte“ im Mikrozensus 2022*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2023, Seite 61 ff.

Fachkommission Integrationsfähigkeit. *Gemeinsam die Einwanderungsgesellschaft gestalten. Bericht der Fachkommission der Bundesregierung zu den Rahmenbedingungen der Integrationsfähigkeit*. 2021. [Zugriff am 22. Oktober 2024]. Verfügbar unter: www.fachkommission-integrationsfaehigkeit.de

Keller, Matthias/Körner, Thomas. *Haushalte, Familien und Lebensformen mit Einwanderungsgeschichte – Auswirkungen auf die Erwerbsbeteiligung von Eltern*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2024, Seite 39 ff.

Rubin, Donald B. *Inference and missing data*. In: Biometrika. Band 63, Ausgabe 3/1976, Seite 581 ff. DOI: [10.1093/biomet/63.3.581](https://doi.org/10.1093/biomet/63.3.581)

Söllner, René/Körner, Thomas. *Der Registerzensus: Ziele, Anforderungen und Umsetzungsansätze*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2022, Seite 13 ff.

RECHTSGRUNDLAGEN

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Dr. Tobias Enderle

ist Volkswirt und hat an der Universität Trier zum Thema Imputation promoviert. Im Referat „Statistische Geheimhaltung“ des Statistischen Bundesamtes befasst er sich als Referent mit Fragen der mathematisch-statistischen Geheimhaltungsmethodik und deren Software-Implementierung.

Birgit Kleber

ist Diplom-Soziologin und arbeitet seit 2008 im Statistischen Bundesamt. Für den Zensus betreut sie die Themen Imputation und Geheimhaltung, zunächst für den Zensus 2011 und aktuell im Referat Methodik, Haushalte als Referentin für den Zensus 2022.

GEHEIMHALTUNG MIT DER CELL-KEY-METHODE IM ZENSUS 2022

Tobias Enderle, Birgit Kleber

↘ **Schlüsselwörter:** Volkszählung – informationelle Selbstbestimmung – stochastische Überlagerung – Informationsverlust – Geheimhaltungswerkzeug

ZUSAMMENFASSUNG

Beim Zensus 2022 hat die amtliche Statistik in Deutschland das datenverändernde Geheimhaltungsverfahren der stochastischen Überlagerung nach der Cell-Key-Methode (CKM) für Veröffentlichungen produktiv eingeführt. Das generische Geheimhaltungswerkzeug wurde entwickelt, da mit dem Wunsch, Nutzenden flexible Auswertungsoptionen anzubieten, eine Automatisierung der Geheimhaltung zwingend einhergehen musste. Der Beitrag erläutert zunächst das Verfahren der Cell-Key-Methode und geht auf die Herausforderungen ein, die damit verbunden sind. Anschließend wird die praktisch-technische Umsetzung der Cell-Key-Methode im Kontext des Zensus 2022 dargelegt.

↘ **Keywords:** population census – informational self-determination – stochastic perturbation – loss of information – SDC tool

ABSTRACT

In the 2022 Census, official statistics institutions in Germany introduced the cell key method (CKM) – a perturbative disclosure control method – for data publications. The generic statistical disclosure control (SDC) tool was developed as the desire to offer users flexible evaluation options made the automation of confidentiality essential. This article first describes the cell key method and examines the challenges associated with it. It then explains the practical and technical implementation of the cell key method in the context of the 2022 Census.

1

Einleitung

Um das Grundrecht aller Bürgerinnen und Bürger auf informationelle Selbstbestimmung zu schützen und den Vorgaben des § 16 Bundesstatistikgesetz Rechnung zu tragen, dürfen die Veröffentlichungen des Zensus keine Rückschlüsse auf Einzelfälle ermöglichen. Gleichzeitig sollen die Daten so viele Informationen und Erkenntnisse wie möglich bereitstellen.

Bei einem komplexen und umfangreichen Tabellenprogramm, wie es für den Zensus 2022 veröffentlicht wird, wären die Geheimhaltungsarbeiten mit einem traditionellen Verfahren wie der Zellspernung äußerst zeit- und kostenintensiv: Sperrmuster müssten aufwendig erstellt und abgestimmt werden. Zudem ist Zellspernung mit einem hohen Informations- und Flexibilitätsverlust verbunden. Daher wird im Zensus 2022 – wie beim Zensus 2011¹ – ein datenveränderndes Geheimhaltungsverfahren eingesetzt. Es findet ein post-tabulares stochastisches Überlagerungsverfahren Anwendung, die sogenannte Cell-Key-Methode. Das neue Verfahren ermöglicht eine hohe Flexibilität bei der Erstellung von Auswertungen und eine hohe – vorab festgelegte – Genauigkeit der Ergebnisse.²

Exkurs

Streng genommen findet mit der Cell-Key-Methode im Zensus eine Tabellengeheimhaltung statt. Eine Tabelle wird von Tabellengliederungen aufgespannt. Zur Gliederung von Tabellen werden üblicherweise kategoriale Merkmale verwendet, die die Merkmalsträger selbst kategorisieren (zum Beispiel Region, Alter, Geschlecht, Gebäudeart). In einer Häufigkeitstabelle gibt ein Tabellenfeld die Anzahl der betreffenden Merkmalsträger an (Einwohnerzahl, Anzahl leerstehender Gebäude). In

einer Wertetabelle (siehe hierzu Abschnitt 3.3 „Überlagerung stetiger Merkmale“) enthält ein Tabellenfeld als Zellwert die Merkmalssumme (oder eine daraus abgeleitete Kennzahl) eines quantitativen Wertmerkmals für die betreffenden Merkmalsträger (zum Beispiel Miete) – im Folgenden Wertsomme genannt. Ein Veröffentlichungsergebnis kann demnach auch eine einfache Kennzahl umfassen (zum Beispiel Leerstandsquote oder durchschnittliche Miete in Wiesbaden) und muss nicht als Tabelle dargestellt werden.

Nach den Rahmenbedingungen und Anforderungen an die Geheimhaltung im Zensus in Kapitel 2 beschreibt Kapitel 3 ausführlich die einzelnen Schritte, wie die stochastische Überlagerung nach der Cell-Key-Methode abläuft. Kapitel 4 informiert über die Qualität der Fallzahl-Ergebnisse im Zensus, die technischen Umsetzungen der Cell-Key-Methode erläutert Kapitel 5. Ein kurzes Fazit schließt den Beitrag ab.

2

Rahmenbedingungen und Anforderungen an die Geheimhaltung

Die Cell-Key-Methode wird im Zensus für alle Veröffentlichungen, die auf demografischen Daten, Gebäude- und Wohnungsdaten, Haushaltsdaten und Familiendaten basieren, angewendet. Bei Stichprobenergebnissen, also Tabellen, deren Werte ausschließlich Resultat der Hochrechnung sind, ist die statistische Geheimhaltung durch die Hochrechnung und anschließende Rundung auf ein Vielfaches von 10 sichergestellt.

Aufgrund ihrer rechtlichen Relevanz werden Einwohnerzahlen³ für administrative Gebietseinheiten (Gemeinden, Verbandsgemeinden, Kreise, Regierungsbezirke, Länder und Bund) im Original veröffentlicht. Hierunter fallen auch die Einwohnerzahlen der Bundestagswahlkreise. Die entsprechenden Ergebnisse werden nicht mit der Cell-Key-Methode geheim gehalten und daher in den Erstellungsprozessen von der Überlagerung ausgeschlossen.

1 Beim Zensus 2011 wurde das prä-tabulare Verfahren SAFE (Sichere Anonymisierung für Einzeldaten) verwendet (Statistische Ämter des Bundes und der Länder, 2014).

2 Das Verfahren der stochastischen Überlagerung geht auf das australischen Statistikamt entwickelte Überlagerungsverfahren zurück (Fraser/Wooton, 2016; Thompson und andere, 2013). Das für die amtliche Statistik angepasste Verfahren mit Cell-Keys ist in Giessing (2016) und Enderle/Vollmar (2019) beschrieben. Höhne/Höhninger (2018) geben ebenfalls eine Einführung in die Methode. Die Anwendung der Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder stellen Setzer und andere (2024) sowie Rothe und andere (2024) dar.

3 Informationen zur Ermittlung der Einwohnerzahl im Zensus 2022 siehe Bretschki und andere (2024).

Eine weitere Besonderheit bei der Nutzung der Cell-Key-Methode als Geheimhaltungsverfahren stellen im Zensus Fallzahlen von 0, 1 und 2 dar. Die 0 ist von der Überlagerung ausgeschlossen, sodass eine nicht vorhandene Beobachtung auch nicht künstlich erzeugt wird. Kleine Fallzahlen können hingegen auf eine 0 verändert werden.¹⁴ Fallzahlen von 1 oder 2 im Original werden immer überlagert und sind in den Zensus-Veröffentlichungen nie enthalten.

3

Stochastische Überlagerung nach der Cell-Key-Methode

Bei der Cell-Key-Methode erfolgt die Geheimhaltung bei der Erzeugung der Ergebnisse (post-tabular). Dabei wird für jedes Originalergebnis (beziehungsweise Tabellenfeld, englisch: cell) ein „zufälliger Überlagerungswert“ deterministisch (das heißt nach vorgegebenen Regeln) ermittelt. Anstelle des Originalergebnisses wird jeweils die Summe aus Originalergebnis und Überlagerungswert veröffentlicht.

3.1 Vom zufälligen Record-Key zur konsistenten Geheimhaltung

Man spricht von einer stochastischen, also zufälligen Überlagerung der Originalergebnisse, da der Überlagerungswert mithilfe einer gleichverteilten Zufallszahl ermittelt wird. Auch wenn es sich um ein post-tabulares Verfahren handelt, müssen die Mikrodaten zunächst erweitert werden. Jedem Datensatz wird eine gleichverteilte Zufallszahl – der sogenannte Record-Key – angefügt. Er liegt im Intervall $[0,1)$ und hat 15 Nachkommastellen. Beim Zensus 2022 erhalten die Einzeldatensätze aller statistischen Einheiten (Personen, Haushalte, Familien, Wohnungen und Gebäude) Record-Keys.

Für die Erstellung von Ergebnissen, die den Bereich der amtlichen Statistik verlassen, werden – parallel zur Bildung eines Originalergebnisses (zum Beispiel die Häufigkeitsauszählung bei Fallzahltabellen) – anhand der Record-Keys die sogenannten Cell-Keys ermittelt. Für alle Einzeldatensätze beziehungsweise Merkmals-träger, die zum jeweiligen Originalergebnis beitragen (das heißt mit identischen Ausprägungskombinationen) wird die Summe der Record-Keys gebildet. Die aufsummierten Record-Keys werden auf das Ausgangsintervall $[0,1)$ rücktransformiert, das heißt auf die Nachkommastellen beschränkt. Diese Cell-Keys liegen selbst wiederum gleichverteilt im Intervall $[0,1)$. Jeder Ausprägungskombination einer Tabelle ist damit neben ihrem Originalergebnis auch ein Cell-Key zugeordnet. Durch die Vorgehensweise des Aufsummierens der den Merkmalsträgern fest zugewiesenen Record-Keys erhalten logisch identische Ausprägungskombinationen automatisch immer denselben konsistenten Cell-Key.

Um eine ganzheitliche Konsistenz im Zensus zu gewährleisten, wurden die Record-Keys der einzelnen Datenbestände von den Personen ausgehend zudem logisch „vererbt“. Besteht zum Beispiel ein Gebäude aus zwei Wohnungen, so entspricht der Record-Key des Gebäudes dem Cell-Key der beiden Wohnungen. Der Record-Key eines Einpersonenhaushalts entspricht dem Record-Key dieser einen Person.

Im letzten Verfahrensschritt – dem deterministischen „Lookup“ – wird für jedes Originalergebnis anhand des Cell-Keys der jeweilige „Überlagerungswert“ aus einer Überlagerungstabelle (englisch: perturbation table, kurz: ptable) ausgelesen.

Die verwendete Überlagerungstabelle¹⁵ beschreibt die statistikspezifischen Regeln des deterministischen Prozesses (siehe Kapitel 2 „Rahmenbedingungen und Anforderungen an die Geheimhaltung“). Dazu gehören die statistischen Eigenschaften der Überlagerungen wie Unverzerrtheit und konstante Varianz sowie zensuspezifische Parameter des Verfahrens, wie die maxi-

⁴ Um die Schutzwirkung der Geheimhaltung zu gewährleisten, darf keine Unterscheidung zwischen originalen Nullen und auf 0 veränderten Fallzahlen möglich sein. Zudem sollen auf 0 veränderte Werte nicht fälschlich als nicht vorhanden gelesen werden. In den Veröffentlichungen des Zensus werden sowohl im Original fehlende Beobachtungen als auch auf 0 veränderte Werte nicht mit dem Zahlenwert „0“, sondern einheitlich durch das wertersetzende Zeichen Dash „-“ (Legende: „Genau Null oder auf Null geändert“) dargestellt.

⁵ Die Überlagerungstabelle ist eine für IT-Verfahren geeignete Form der sogenannten Übergangsmatrix. Diese enthält die bedingten Übergangswahrscheinlichkeiten für das Eintreten einer bestimmten Überlagerung abhängig von der Größe der Originalhäufigkeit (Enderle/Vollmar, 2019, hier: Seite 92 ff., Abschnitt 4.2). Für die Berechnung wird das für die Statistiksoftware R entwickelte „ptable“-Paket (Enderle, 2023) verwendet. Die Parameter der Überlagerungsmatrix werden dabei statistikspezifisch festgelegt.

male Höhe der Überlagerungen. Die Auswirkungen der festgelegten Überlagerungstabelle auf die Qualität der Veröffentlichungsergebnisse werden im Kapitel 4 „Qualitätsaussagen zur Überlagerung“ aufgegriffen.

3.2 Fallzahlüberlagerung

In Fallzahltabellen werden absolute Häufigkeiten dargestellt. Der Überlagerungswert, der zur jeweiligen Originalhäufigkeit hinzuaddiert wird, stellt die Differenz aus Ziel- und Originalhäufigkeit dar und ist wegen der konsistenten Cell-Keys für logisch identische Ausprägungskombinationen immer identisch.

Die Überlagerungstabelle kann grafisch als Überlagerungstableau dargestellt werden. [Grafik 1](#) Jeder Balken entspricht einer Originalhäufigkeit. Unterschiedliche Farben entsprechen unterschiedlichen Überlagerungswerten und die jeweilige Summe der farbigen Teilbalken entspricht der in der Übergangsmatrix vorgegebenen Wahrscheinlichkeit, mit der es zu der betreffenden Überlagerung der entsprechenden Originalhäufigkeit kommt. Der Lookup-Schritt „liest“ die Überlagerung im Überlagerungstableau in der durch den Originalwert i gegebenen Zeile an der Stelle $p = \text{Cell-Key}$ ab. Die Überlagerungswerte für alle nicht explizit genannten Originalhäufigkeiten (hier: $i > 4$) werden aus der „obersten“ Verteilung (4+) abgelesen.

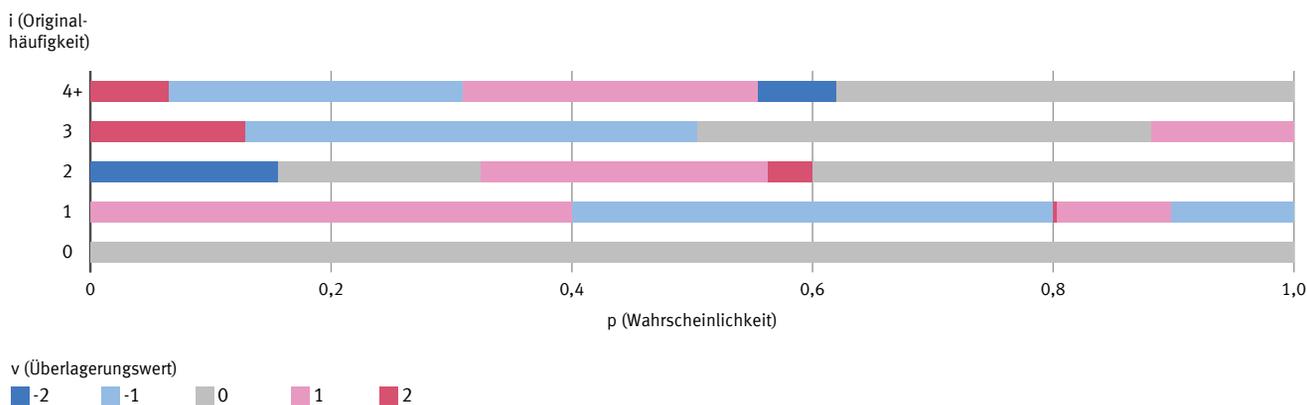
Das folgende Fallzahl-Beispiel illustriert den Lookup-Schritt:

Der Cell-Key für drei leerstehende Wohnungen der Baujahr-Kategorie 1950 bis 1960 sei die Summe der einzelnen Record-Keys $0,95 + 0,40 + 0,75 : 0,1$ (das heißt die Summe der Record-Keys ohne Vorkommastelle). Anhand des in Grafik 1 gegebenen Überlagerungstableaus wird für die Originalhäufigkeit $i = 3$ (Wohnungen) und $p = 0,1$ (Cell-Key) ein Überlagerungswert von $+2$ (dunkelblauer Balken) abgelesen. In diesem konstruierten Beispiel würden in einer Veröffentlichung fünf Wohnungen ausgewiesen werden.

Genauigkeit gegenüber Nicht-Additivität

Durch die konsistenten Cell-Keys liefert das Verfahren konsistente Ergebnisse. Um dies zu gewährleisten, müssen grundsätzlich alle Teilergebnisse gleichbehandelt werden. Dies bedeutet, dass alle Tabellenfelder separat überlagert werden müssen. Das Verfahren unterscheidet damit nicht zwischen unterschiedlichen Tabellenfeldern – wie Innenfeldern und dem Insgesamt. Die jeweiligen Tabellenfelder einer Tabellenzeile oder -spalte nach Geheimhaltung addieren sich daher nicht notwendigerweise zur ausgewiesenen Ingesamt-Summe. Die Summe aus Männern und Frauen zum Beispiel wird separat überlagert und nicht aus den überlagerten Häu-

Grafik 1
Modifiziertes Überlagerungstableau mit permutierten Intervallen



Die Parameter des Beispiels sind so gewählt, dass nach Geheimhaltung keine Fallzahlen von 1 auftreten und die maximale Abweichung ± 2 beträgt. Das Beispiel gibt dabei nicht die im Zensus 2022 verwendete Parametrisierung wieder und dient hier nur zu Veranschaulichungszwecken. Zur Erhöhung der Schutzwirkung sind die Intervalle pro Originalhäufigkeit zufällig angeordnet (permutiert) und teilweise in Subintervalle aufgeteilt.

figkeiten von Männern und Frauen gebildet. Man spricht von der sogenannten Nicht-Additivität⁶.

Das folgende Fallzahl-Beispiel in [Tabelle 1](#) veranschaulicht dies:

Das in der ersten Tabellenzeile ausgewiesene Insgesamt über alle Altersklassen für die Ausprägung „Männlich“ beträgt 175. Werden die zugehörigen Tabellenfelder der ersten Zeile jedoch separat addiert, ergibt deren Summe 173 (= 20 + 31 + 32 + 40 + 50). Das Insgesamt über die komplette Tabelle ist mit 371 ausgewiesen. Werden die Fallzahlen für die Ausprägungen „Männlich“ und „Weiblich“ jedoch separat addiert, ergibt deren Summe 372 (= 175 + 197). Auch bei einer Addition über die Insgesamthäufigkeiten aller Altersklassen (47 + 56 + 71 + 86 + 109 = 369) sowie bei einer Addition über alle Innenfelder der Tabelle (20 + 31 + 32 + 40 + 50 + 25 + 25 + 40 + 45 + 60 = 368) ergeben sich jeweils leichte Abweichungen zum ausgewiesenen Insgesamt-Wert.

Die Nicht-Additivität widerspricht zunächst der Intuition. Sie wird jedoch in Kauf genommen, da hierdurch zwei wichtige Eigenschaften des Verfahrens erst möglich sind:

1. Tabellenübergreifende Konsistenz: Für ein bestimmtes Ergebnis (zum Beispiel Anzahl der unter 18-Jährigen) ist der hinzuaddierte Überlagerungswert und somit das dargestellte Ergebnis immer identisch, unabhängig davon, in welcher anderen Kombination dieses Ergebnis dargestellt wird (zum Beispiel Personen nach Altersklassen und Geschlecht oder Personen nach Altersklassen und Nationalität).

⁶ Untersuchungen auf Basis von regulären Ergebnistabellen des Zensus 2011 haben gezeigt, dass das nachträgliche Wiederherstellen der Additivität bereits bei vergleichsweise simplen Tabellen (bis zu zwei Sachmerkmale bei einem ausgewählten Tabellenset) mit einem erheblichen Genauigkeitsverlust verbunden wäre.

2. Genauigkeit (geringer Informationsverlust): Da alle Ergebnisse getrennt voneinander überlagert werden, wird vermieden, dass sich eine Reihe zufällig gleich gerichteter Überlagerungen in Summen kumulieren und dann im Einzelfall größere Veränderungen zwischen Originalhäufigkeiten und überlagerten Häufigkeiten hervorrufen. Um im Beispiel zu bleiben: Wenn die „unter 18-Jährigen“ in einer Tabelle als Summe der entsprechenden 18 einzelnen Altersjahre (0, 1, 2, ...) dargestellt werden und in allen 18 Altersjahren der Überlagerungswert bei einer Maximalabweichung von $D = 2$ zufällig negativ ausfällt (zum Beispiel immer -2), würde ein als Summe der 18 Abweichungen gebildeter Überlagerungswert (im Beispiel: -36) das Gesamtergebnis deutlich stärker verändern als dies bei einer separaten Überlagerung des Gesamtergebnisses der Fall wäre. Dafür ist lediglich die vorab festgelegte Abweichung von ± 2 möglich.

Das bis hier beschriebene Verfahren zeigt die Geheimhaltung von Fallzahlauswertungen. Der Zensus bietet daneben viele spannende Kennzahlen, deren überlagertes Ergebnis nicht ausschließlich aus der Überlagerung von Fallzahlen resultiert.

Kennzahlen auf Basis von Fallzahlen

Kennzahlen wie Anteile, Verhältnisse, Veränderungsraten oder Salden werden auf Basis der überlagerten Fallzahlen berechnet. Hierbei können zwei unerwünschte Effekte auftreten:

1. Ungenauigkeit (hoher Informationsverlust): Bei kleinen Fallzahlen kann es zu großen absoluten Abweichungen zwischen originaler und der Kennzahl nach Geheimhaltung kommen.
2. Inkonsistenz/Unplausibilität: Da die Originalhäufigkeiten der Terme (zum Beispiel Zähler und Nenner) wie alle anderen Häufigkeiten unabhängig voneinander

Tabelle 1

Kreuztabelle mit Fallzahlen nach Geschlecht und fünf Altersklassen

	Alter					Insgesamt
	unter 18 Jahre	18 bis 29 Jahre	30 bis 49 Jahre	50 bis 64 Jahre	65 Jahre und älter	
	Anzahl					
Männlich	20	31	32	40	50	175
Weiblich	25	25	40	45	60	197
Insgesamt	47	56	71	86	109	371

der überlagert werden, kann es bei Kennzahlen, bei denen die Original-Terme nahe beieinander liegen, zu unplausiblen Werten kommen. Beispiele wären Anteile größer 100 % oder Salden, die statt einer Steigerung einen Rückgang ausweisen.

Zu beiden Effekten wurden geeignete Maßnahmen ergriffen:

- › Kennzahlen, die aufgrund der möglichen – und nicht tatsächlichen – Überlagerung eine hohe Ungenauigkeit⁷ aufweisen können, werden geklammert dargestellt.
- › Kennzahlen behalten auch nach Überlagerung ihre grundsätzliche Aussagekraft durch eine entsprechende Behandlung (zum Beispiel Skalierung) im Zuge der Überlagerung. So behalten Salden beispielsweise ihr Vorzeichen oder Anteile werden nach Überlagerung mit maximal 100 % ausgewiesen.

3.3 Überlagerung stetiger Merkmale

Aus dem Zensus werden auch metrisch skalierte (sowohl stetige als auch quasi stetige) Merkmale in Wertetabellen veröffentlicht. Bei stetigen Merkmalen bestehen bei Anwendung der Cell-Key-Methode grundsätzlich zwei Umsetzungsoptionen: Mit der ersten Option wird eine höhere Genauigkeit der gebildeten Wertsumme selbst erreicht, die zweite Option erzielt demgegenüber eine höhere Genauigkeit bei der Ermittlung von Mittelwerten (Quotient mit Wertsumme als Zähler und Anzahl beitragender Merkmalsträger im Nenner). Im Zensus wurde der Veröffentlichung von Durchschnittswerten eine höhere Priorität eingeräumt als der Veröffentlichung aufsummierter stetiger Merkmale und daher Option 2 umgesetzt.⁸

Die Überlagerung von Wertetabellen erfolgt vom Grundprinzip her zunächst wie die Überlagerung von Fallzahltabellen. Während bei Fallzahltabellen vorrangig kleine Fallzahlen geheim gehalten werden, sind bei Wertetabellen (zumeist) die größten Beitragswerte (zum Bei-

- 7 Übersteigt der aufgrund einer möglichen Überlagerung statistisch zu erwartende relative Fehler der Kennzahl 15 %, so wird das Ergebnis geklammert (Legende: „Aussagewert eingeschränkt, da der Zahlenwert durch das Geheimhaltungsverfahren möglicherweise relativ stark verändert wurde“).
- 8 Beispielsweise ist die durchschnittliche Miete einer Gemeinde eine deutlich stärker nachgefragte Information als die Summe aller Mieten einer Gemeinde zusammen.

spiel größte Miete) einer Wertsumme zu schützen. Bei der Überlagerung stetiger Merkmale müssen jedoch einige grundlegende Anpassungen vorgenommen werden:⁹

1. Die Stärke (Varianz) der Überlagerung korreliert mit der Größe der größten Einzelwerte und wird durch wählbare Parameter optimal an die Daten angepasst.
2. Es wird gewährleistet, dass auch für nicht ganzzahlige Werte Überlagerungswerte bereitgestellt werden. Der Überlagerungswert selbst ist daher metrisch skaliert und wird anhand einer für stetige Merkmale konzipierten Überlagerungstabelle ermittelt.

Mittelwerte werden wie Anteile durch eine Klammerung qualitätsgesichert dargestellt, falls von einem eingeschränkten Interpretationsgehalt nach Geheimhaltung ausgegangen werden kann.

3.4 Sperrung von Ergebnissen

Neben der bereits skizzierten Klammerung von Kennzahlen wurde auch eine Sperrung der Auswertungen umgesetzt, falls Ergebnisse ein zu hohes Aufdeckungsrisiko und/oder einen zu hohen Informationsverlust bergen. Mögliche Sperrungen werden durch das wertersetze Zeichen „,“ (Legende: „Zahlenwert unbekannt oder geheim“) dargestellt. Die Prüfung erfolgt für einzelne Tabellenteile separat, wobei die zu prüfenden Tabellenteile anhand ihrer Aggregationsebene voneinander unterschieden werden. Im Falle einer einfachen Kreuztabelle und zwei Merkmalen sind das zum Beispiel vier unterschiedliche Tabellenteile: Innenfelder, Ingesamt-Felder der Zeilen, Ingesamt-Felder der Spalten und Ingesamt-Eckfeld. [↘ Tabelle 2](#) Die Prüfung auf mögliche Sperrungen wird immer mit den Innenfeldern auf der feinsten Ebene begonnen. Weisen diese weder ein hohes Aufdeckungsrisiko noch einen hohen Informationsverlust aus, wird die gesamte Tabelle freigegeben. Werden die Innenfelder bereits als zu sperren bewertet, werden die nächsthöheren Aggregate auf eine mögliche Sperrung hin geprüft. So wird sukzessive bis zum Eckwert (höchste Aggregationsebene) fortgefahren. Die Sperrentscheidung betrifft in der Regel wenige Tabellenteile – meist nur die Innenfelder – einer Abfrage.

9 Giessing/Tent (2019) zeigen Erweiterungen der Cell-Key-Methode für stetige Merkmale.

Tabelle 2

Beispieltable mit 4 Tabellenteilen

		Art des Gebäudes		Insgesamt
		Wohngebäude	Sonstige Gebäude mit Wohnraum	
Gebäudetyp	Freistehendes Haus	.	.	.
	Doppelhaushälfte	.	.	.
	Gereihtes Haus	.	.	.
	Anderer Gebäudetyp	.	.	.
Insgesamt		1 034	206	1 242

Jede Farbe kennzeichnet einen Tabellenteil. Die Innenfelder (dunkelblau) und Ingesamt-Felder des Merkmals „Gebäudetyp“ (hellblau) können gesperrt sein, während die Ingesamt-Felder des Merkmals „Art des Gebäudes“ (hellrot) und das Ingesamt-Eckfeld (dunkelrot) freigegeben werden.

Um mögliche Risiken und zu hohe Informationsverluste zu vermeiden, gleichzeitig aber das Veröffentlichungsprogramm des Zensus nicht zu stark einzuschränken, wurden verschiedene Sperrregeln anhand der Ergebnisse des Zensus 2011 untersucht.¹⁰ Die Einstellungen wurden schließlich so gewählt, dass je Bundesland 98 % der getesteten Tabellen auf Gemeindeebene vollständig angezeigt werden. Deutschlandweit erhalten lediglich 0,1 % der getesteten Tabellen auf Gemeindeebene eine Sperrung und auf Kreisebene können alle Tabellen vollständig veröffentlicht werden.

4

Qualitätsaussagen zur Überlagerung

Der Überlagerungswert, der zu den Originalergebnissen addiert wird und im Mittel 0 ist, nimmt bei Fallzahlentabellen ganzzahlige Werte überwiegend zwischen -2 und +2 an und hat damit auf größere Ergebnisse quasi keinen Einfluss. Der Überlagerungswert kann auch 0 betragen, sodass das Veröffentlichungsergebnis mit dem Originalergebnis übereinstimmt. Die Parameter des Verfahrens (zum Beispiel die maximale Überlagerung, die Varianz der Überlagerung oder die Wahrscheinlichkeiten, dass mit 0 überlagert wird) selbst können aus Gründen der Schutzwirkung nicht veröffentlicht werden.

10 Insgesamt wurden 574 reguläre Tabellen aus dem Veröffentlichungsprogramm von 2011 untersucht. Davon wurden 13 vom Fach- und Veröffentlichungsbereich als zentrale Tabellen definiert und prioritär betrachtet.

Stattdessen können die folgenden Qualitätsaussagen herangezogen werden, um die Qualität der Fallzahl-Ergebnisse des Zensus zu bewerten:

1. Der Absolutbetrag der Abweichungen zwischen überlagerten und originalen Fallzahlen liegt für reguläre Tabellen im Mittel unter 1.
2. Mindestens 65 % der Fallzahlen in den Tabellen bleiben unverändert oder weichen um maximal 1 vom Originalwert ab.
3. Bei höchstens 20 % der Fallzahlen in den Tabellen liegt die Abweichung bei 3 oder mehr.
4. Bei höchstens 10 % der Fallzahlen in den Tabellen liegt die Abweichung bei 4 oder mehr.

Selbstständige Weiterverarbeitungen von überlagerten Ergebnissen (zum Beispiel eigenständige Summenbildung aus überlagerten Fallzahlen statt auf das bereitgestellte überlagerte Ingesamt direkt zuzugreifen) können zu höheren Informationsverlusten beziehungsweise Ungenauigkeiten in der Ergebnisdarstellung führen. Das Veröffentlichungsprogramm des Zensus 2022 ist entsprechend umfangreich gestaltet.

5

Technische Umsetzungen der Cell-Key-Methode

Zur Automatisierung und einheitlichen Durchführung der Geheimhaltung mit der stochastischen Überlagerung nach der Cell-Key-Methode hat das Statistische Bundesamt im Rahmen des Zensus ein flexibles und generisches Geheimhaltungswerkzeug, die Fachanwendung CKM-Geheimhaltung¹¹ entwickelt. Die Verwendung und Anbindung der modularen Fachanwendung an bestehende Werkzeuge der amtlichen Statistik (Analyse- und Auswertungswerkzeuge, Datenbanken und landeseigene Informationssysteme) ermöglichen eine harmonisierte Geheimhaltung und konsistente Veröffentlichung einschließlich Qualitätskennzeichnungen (Klammerungen und Sperrungen). Das Geheimhaltungswerkzeug wird in den Statistischen Ämtern des Bundes und der Länder (als Backend-System) bereits für diverse

11 Die Fachanwendung basiert grundsätzlich auf einer JAVA-Architektur und bietet als Webservice eine REST-Schnittstelle.

Veröffentlichungsprodukte des Zensus 2011 und des Zensus 2022 eingesetzt:

- › Zensusdatenbank,
- › Regionaltabellen,
- › Gitterdaten,
- › Zensus-Atlas und
- › Sonderauswertungen für Dritte.

Geheimhaltung durch Dritte

Neben den Statistischen Ämtern des Bundes und der Länder erhalten auch weitere Einrichtungen – wie die Wissenschaft über die Forschungsdatenzentren (Setzer und andere, 2024; Rothe und andere, 2024) oder die Kommunen mit abgeschotteter Statistikstelle – Einzeldaten aus dem Zensus (§ 16 Absätze 5 und 6 Bundesstatistikgesetz sowie § 32 Absatz 2 Zensusgesetz 2022). Sollen hieraus Ergebnisse veröffentlicht werden, muss sichergestellt sein, dass die Anwendung der Cell-Key-Methode unter Verwendung der Zensus-Parametrisierung erfolgt und zu identischen und konsistenten Ergebnissen führt. Hierzu wird das im Statistischen Bundesamt entwickelte Geheimhaltungswerkzeug auch an Auswertungssysteme der jeweiligen Einrichtungen (zum Beispiel Werkzeuge der KOSIS-Arbeitsgemeinschaften KO.R und DUVA für Kommunen mit abgeschotteter Statistikstelle) angebunden.

Die Geheimhaltung der vom Statistischen Amt der Europäischen Union (Eurostat) veröffentlichten Ergebnisse zum deutschen Zensus 2022 wurde ebenfalls mittels des Geheimhaltungswerkzeuges sichergestellt.

6

Fazit

Die Cell-Key-Methode zeigt sich – mit einigem Vorlauf – als geeignetes Verfahren zur Geheimhaltung der deutschen Zensusergebnisse. So konnte der Umfang des Veröffentlichungsprogramms – insbesondere hinsichtlich der regionalen Auswertungstiefe – deutlich erhöht werden. Zudem konnte mit der Entwicklung der Fachanwendung CKM-Geheimhaltung ein Auswertungssystem-übergreifendes Instrument geschaffen und be-

reitgestellt werden, welches eine Automatisierung der Geheimhaltung der Ergebnisse ermöglicht. Diese Umsetzung weist jedoch auch Herausforderungen auf, die allerdings bereits in der Konzeptionsphase erkannt und entsprechend vorbereitet wurden. Im Ergebnis stellt das Verfahren konsistente Ergebnisse mit einer hohen, vorab festgelegten Genauigkeit zur Verfügung. 

LITERATURVERZEICHNIS

Bretschi, Corinna/Seibel, Steffen/Vorndran, Ingeborg/Meyn, Christoph. *Ermittlung der Einwohnerzahl im Zensus 2022*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 17 ff.

Enderle, Tobias. *pstable: Generation of Perturbation Tables for the Cell-Key Method*. R package version 1.0.0. 2023. [Zugriff am 1. November 2024]. Verfügbar unter: CRAN.R-project.org

Enderle, Tobias/Vollmar, Meike. *Geheimhaltung in der Hochschulstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 87 ff.

Fraser, Bruce/Wooton, Janice. *A proposed method for confidentialising tabular output to protect against differencing*. Work session on Statistical Data Confidentiality. Supporting paper. Genf 2005. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

Giessing, Sarah. *Computational Issues in the Design of Transition Probabilities and Disclosure Risk Estimation for Additive Noise*. In: Domingo-Ferrer, Josep/Peji-Bach, Mirjana (Herausgeber). *Privacy in Statistical Databases*. LNCS (Lecture Notes in Computer Science). 2016. Ausgabe 9867, Seite 237 ff.
DOI: [10.1007/978-3-319-45381-1_18](https://doi.org/10.1007/978-3-319-45381-1_18)

Giessing, Sarah/Tent, Reinhard. *Concepts for generalising tools implementing the cell key method to the case of continuous variables*. In: Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Den Haag 2019. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

Höhne, Jörg/Höninger, Julia. *Die Cell-Key-Methode – ein Geheimhaltungsverfahren*. In: Zeitschrift für amtliche Statistik Berlin Brandenburg. Ausgabe 3+4/2018, Seite 14 ff. [Zugriff am 9. Dezember 2024]. Verfügbar unter: www.statistischebibliothek.de

Rothe, Patrick/Güttgemanns, Volker/Rohde, Johannes/Setzer, Stefanie. *Die Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Teil 2: Auswirkungen des neuen Geheimhaltungsverfahrens*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2024, Seite 45 ff.

Setzer, Stefanie/Rohde, Johannes/Güttgemanns, Volker/Rothe, Patrick. *Die Cell-Key-Methode in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Teil 1: Vorstellung des neuen Geheimhaltungsverfahrens*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2024, Seite 31 ff.

Statistische Ämter des Bundes und der Länder. *Nutzer-Kurzinformation zu SAFE*. 2014. [Zugriff am 1. November 2024]. Verfügbar unter: www.zensus2011.de

Thompson, Gwenda/Broadfoot, Stephen/Elazar, Daniel. *Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics*. Paper presented at the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Ottawa 2013. [Zugriff am 1. November 2024]. Verfügbar unter: unece.org

RECHTSGRUNDLAGEN

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 14 des Gesetzes vom 8. Mai 2024 (BGBl. 2024 I Nr. 152) geändert worden ist.

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Annika Gaedke

hat Soziologie studiert und ist Referentin im Referat „Gebäude und Wohnungen“ der Gruppe „Registerbasierter Datenbestand des Zensus“ des Statistischen Bundesamtes im Bereich des Einrichtungsregisters. Bis 2023 war sie im Zensus 2022 für die fachliche Konzeption und Begleitung der IT-Entwicklung für die Erhebung an Sonderbereichen zuständig.

Miriam Pfahl

hat Politikwissenschaft, Soziologie und Öffentliches Recht studiert und ist als Referentin im Referat „Gebäude und Wohnungen“ des Statistischen Bundesamtes tätig. Für den Zensus 2022 war sie unter anderem für die Konzeption der Online-First-Strategie und die Entwicklung der Online-Fragebogen der Haushalbefragung auf Stichprobenbasis zuständig.

Anna Strohm

hat Soziologie studiert und ist Referentin im Referat „Basisregister für Unternehmen: Grundsatz- und Querschnittsaufgaben“ des Statistischen Bundesamtes. Von 2019 bis 2024 war sie im Zensus 2022 für die fachliche Konzeption und Begleitung der IT-Entwicklung für die Gebäude- und Wohnungszählung zuständig.

DIE ONLINE-FIRST-STRATEGIE UND DIE ONLINE-FRAGEBOGEN IM ZENSUS 2022

Annika Gaedke, Miriam Pfahl, Anna Strohm

📌 **Schlüsselwörter:** Volkszählung – Primärerhebung – digitaler Wandel – Haushalbefragung – Gebäude- und Wohnungszählung – Gemeinschaftsunterkünfte

ZUSAMMENFASSUNG

Im Zensus 2022 wurde erstmals auf eine Online-First-Strategie gesetzt, um die Erhebung für die insgesamt 34 Millionen Auskunftspflichtigen der verschiedenen Erhebungsteile effizienter und anwendungsfreundlicher zu gestalten. Unter anderem führten Erfahrungen aus dem Zensus 2011 dazu, die Online-Fragebogen zu verbessern, an die aktuellen Bedürfnisse der Befragten anzupassen und die unterschiedlichen Anforderungen der drei ergänzenden Befragungen des registergestützten Verfahrens zu berücksichtigen. Erkenntnisse aus den durchgeführten Pretests flossen ebenfalls in die Optimierungen ein. Verschiedene Maßnahmen erleichterten die Teilnahme für die Befragten, was zu sehr hohen Online-Quoten führte, dem Ziel der Online-First-Strategie.

📌 **Keywords:** population census – primary survey – digital transformation – household survey – census of buildings and housing – collective living quarters

ABSTRACT

The 2022 Census was the first census in Germany to adopt an Online First strategy. The aim was to make the survey more efficient and user-friendly for the 34 million respondents to the individual survey components. Insights gained from the 2011 Census, for example, resulted in improvements to the online questionnaires, adapting them to the current needs of respondents and addressing the different requirements of the three supplementary surveys of the register-assisted procedure. Findings from the completed pretests were also incorporated into the optimisations. A variety of measures made participation easier for the respondents, resulting in very high online rates, which is the goal of the Online First strategy.

1

Einleitung

„Wenn Sie schon einmal einen Online-Fragebogen ausgefüllt haben, gehen Sie bitte weiter zu Frage 2. Ansonsten überspringen Sie bitte Frage 2 und fahren mit Frage 3 fort.“ Anweisungen dieser Art kennen vermutlich alle, die schon einmal einen Papierfragebogen ausgefüllt haben und wissen, wie verwirrend solche Filterführungen teilweise sein können. Dass Befragte nur die Fragen sehen, die sie aufgrund ihrer bisherigen Angaben in einem Fragebogen betreffen, ist einer von vielen Vorteilen, die Online-Fragebogen im Vergleich zu Papierfragebogen bieten.

Um die Umwelt zu schonen, war ein Ziel beim Zensus 2022, den Anteil an gedruckten Papierfragebogen im Vergleich zum Zensus 2011 deutlich zu verringern. Hätte man allein die ausgefüllten Fragebogen der Gebäude- und Wohnungszählung beim Zensus 2011 übereinandergestapelt, wäre dieser Stapel höher als der Mount Everest mit seinen 8848 Metern Höhe gewesen (Freier/Mosel, 2019, hier: Seite 52).

Diese Vorteile des Verzichts auf Papierfragebogen bildeten die Grundlage für eine wesentliche Zielsetzung des Zensus 2022 gegenüber der Vorgängerrunde im Jahr 2011: das konsequente Verfolgen einer Online-First-Strategie bei den drei ergänzenden Befragungen des registergestützten Verfahrens. Zwar gab es weiterhin Papierfragebogen und persönliche Interviews, um online-ferne Bevölkerungsgruppen zu erreichen. Präferierter Erhebungsmodus im Statistischen Verbund¹ war jedoch die selbstadministrierte Auskunft über die jeweiligen Online-Fragebogen. So wurde unter anderem die individuelle Belastung der insgesamt etwa 34 Millionen Auskunftspflichtigen der Gebäude- und Wohnungszählung, der Erhebung an Sonderbereichen (Wohnheime und Gemeinschaftsunterkünfte) sowie der Haushaltebefragung auf Stichprobenbasis² minimiert. Aus diesem Grund wurden für den Zensus 2022 alle Erhebungsprozesse so konzipiert, dass die Fragebogen von den Befragten standardmäßig online ausgefüllt werden sollten.

1 Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

2 Für Informationen zur Anzahl der Auskunftspflichtigen nach Erhebungsteil siehe www.zensus2022.de

Internet-Befragungen bieten zahlreiche Vorteile: Online-Meldungen sind schnell, kostengünstig, ressourcenschonend, zeitsouverän und belastungsarm. Durch die Möglichkeit von technisch unterstützten Dialogverfahren liefern sie Ergebnisse von vergleichsweise hoher Genauigkeit. [↪ Übersicht 1](#)

Übersicht 1

Vorteile selbstadministrierter Online-Meldungen

Komfort für die Befragten: Zeitsouveränes Ausfüllen in einer bevorzugten Sprache sowie flexible Wahl des Endgeräts

Effiziente Filterführung: Befragte werden entlastet, da sie nur die sie betreffenden Fragen sehen

Geringere Fehlerrate und damit höhere Ergebnissenauigkeit durch Sofort-Plausibilisierung der Erhebungsdaten bei der Eingabe

Schnelle Verarbeitung ohne Medienbrüche: Ergebnisse liegen sofort nach Befragungsende in digitalisierter Form vor, keine Beleglesung und Nachkodierung erforderlich und dadurch Vermeidung von Verarbeitungsfehlern

Ressourcenschonung: Es wird deutlich weniger Papier bedruckt, verschickt, digitalisiert und vernichtet

Kosteneinsparung, wenn wenig oder kein Interviewpersonal benötigt wird und wenn im Fall von postalischen Erhebungen Porto gespart werden kann

Bereits im Zensus 2011, bei dem ebenfalls verschiedene Erhebungswege angeboten wurden, wählte ein Drittel der Befragten der Gebäude- und Wohnungszählung die Möglichkeit, die Meldung über das Web-Formular abzugeben, ohne dass diese explizit beworben worden war. Auch bei der Haushaltebefragung auf Stichprobenbasis im Zensus 2011 nutzte ein Drittel der Selbstausfüller die Möglichkeit der Online-Meldung.³

In den elf Jahren zwischen den beiden Zensusrunden hat die Internetnutzung rasant zugenommen: Hatten beim letzten Zensus 2011 erst 77 % der privaten Haushalte einen Internetzugang, waren es 2020 bereits 92 % (Statistisches Bundesamt, 2020). Im Jahr 2022 sprach der Verbreitungsgrad in der Bevölkerung – einen Internetanschluss besaßen 95 % der Bevölkerung ab 14 Jahre (ARD/ZDF-Onlinestudie, 2023) – dafür, dass die Online-Meldung als das Standardverfahren vorausgesetzt werden konnte.

Die Statistischen Ämter des Bundes und der Länder entschieden daher, den Zugang zum Online-Fragebogen attraktiv und nutzerfreundlich zu gestalten, um die vorrangige Nutzung des Online-Meldewegs bei den Befra-

3 Allerdings erfolgte bei 79 % der Auskunftspflichtigen die Befragung durch Interviewerinnen und Interviewer, sodass die Online-Quote – bezogen auf alle Befragten – nur 7 % betrug.

gungen zum Zensus zu ermöglichen. Gleichzeitig waren potenzielle technische Hürden zu beseitigen, um einen niedrighschwelligigen Zugang für alle Bevölkerungsgruppen zu gewährleisten. Dies erforderte neben konzeptionellen Anpassungen in der Erhebungsdurchführung eine grundlegende Überarbeitung des im Statistischen Verbund standardmäßig zur Durchführung von Online-Meldungen genutzten IDEV-Verfahrens⁴. Das Layout und die Bedienbarkeit von IDEV wurden speziell für den Zensus 2022 grundlegend überarbeitet und modernisiert mit dem Ziel, für alle angebotenen Fragebogen eine einheitliche Bedienoberfläche zu bieten.

Dieser Artikel erläutert die zentralen Anpassungen, die für die Online-Meldungen zum Zensus 2022 umgesetzt wurden. Dabei geht er auf die unterschiedlichen Anforderungen der Erhebungsteile ein, die zu einigen Besonderheiten der verschiedenen Fragebogen führten. Kapitel 2 informiert über verschiedene Anpassungen, die sich aus dem digitalen Wandel und den Erkenntnissen aus dem Zensus 2011 ergaben. Die einzelnen Online-Fragebogen und ihre speziellen Merkmale sind in Kapitel 3 beschrieben. Den Ablauf der Pretests zu den Fragebogen schildert Kapitel 4, während Kapitel 5 den Beitrag mit einem Fazit abschließt.

2

Digitaler Wandel und Erkenntnisse aus dem Zensus 2011

Ausgangspunkt für die Modernisierung des Online-Meldevorgangs IDEV für den Zensus 2022 waren zum einen die Möglichkeiten und die damit verbundenen Anforderungen des digitalen Wandels, zum anderen Erkenntnisse aus dem Zensus 2011.

4 Für die Durchführung von Online-Meldungen und Online-Befragungen steht in der amtlichen Statistik das Online-Meldevorgang IDEV (Internet Datenerhebung im Verbund) zur Verfügung. Es handelt sich hierbei um einen einheitlichen Meldeweg, der seit 2005 von den Statistischen Ämtern des Bundes und der Länder angeboten wird und mit dem über das Internet Meldungen zu verschiedenen amtlichen Statistiken übermittelt werden können.

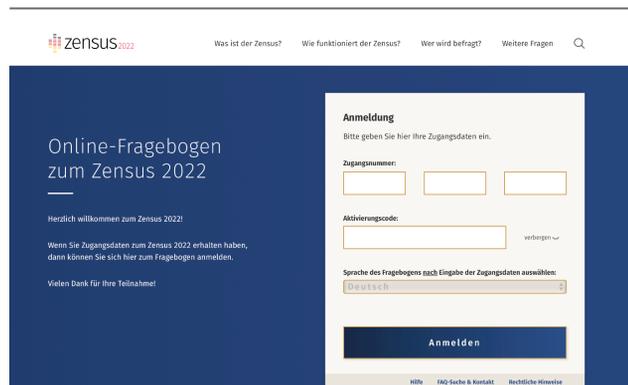
2.1 Login und Zugangsdaten

Eine wesentliche Voraussetzung für die vorzugsweise Nutzung des Online-Meldewegs durch die auskunftspflichtigen Personen ist, dass der elektronische Fragebogen einfach, schnell und über viele verschiedene internetfähige Geräte aufgerufen und bedient werden kann.

Die Erfahrungen aus dem Zensus 2011 haben gezeigt, dass bereits der Zugang zum Online-Fragebogen eine hohe Hürde für die Nutzung darstellen kann. Das Auffinden der korrekten Webseite sowie die Auswahl der korrekten erhebungsspezifischen Anmeldemaske stellte für einige Auskunftspflichtige 2011 eine erhebliche Herausforderung dar. Um nicht bereits an dieser Stelle erste online-meldewillige Personen zu verlieren, wurde für den Zensus 2022 ein einheitliches Anmeldefenster für alle Zensusbefragungen an zentral sichtbarer Stelle auf der damaligen Startseite der Zensus-Webseite www.zensus2022.de platziert. Durch die Eingabe der individuellen Zugangsdaten erfolgte die automatische Weiterleitung auf das zutreffende Fragebogenformular (Gebäude- und Wohnungszählung, Haushaltebefragung, Erhebung an Wohnheimen oder Erhebung an Gemeinschaftsunterkünften). ↘ Grafik 1

Grafik 1

Anmeldemaske auf der Zensus-2022-Homepage



Eine weitere zentrale Anpassung betraf an dieser Stelle die Gestaltung der Anmeldedaten. Auch hier haben die Erfahrungen aus dem Zensus 2011 gezeigt, dass die damals wenig nutzerfreundliche Gestaltung der Login-Daten für den elektronischen Fragebogen viele Auskunftspflichtige bereits am Anmeldevorgang zum Online-Fragebogen scheitern ließ. Daher wurden die

Zugangsdaten für den Zensus 2022 unter Berücksichtigung der Vorgaben zur Datensicherheit weitestmöglich vereinfacht. Ähnlich wie beim Zensus 2011 waren eine Zugangsnummer und ein Aktivierungscode für das Login in den Online-Fragebogen vorgesehen. Die 12-stellige Zugangsnummer – bestehend aus numerischen Zeichen – wurde zur erleichterten Eingabe in drei Blöcke zu je vier Ziffern gegliedert. Beim ebenfalls 12-stelligen Aktivierungscode – bestehend aus numerischen Zeichen und ausschließlich Kleinbuchstaben – wurden anders als 2011 stark ähnelnde Ziffern oder Buchstaben („l, i, o, r, n, v“ und die Ziffer „0“) sowie Sonderzeichen zugunsten einer besseren Lesbarkeit ausgeschlossen. Außerdem sorgten die ausschließliche Kleinschreibung und die Wahl einer gut leserlichen Schriftart für eine leichtere Eingabe, ohne Sicherheitsverluste hinnehmen zu müssen.

Für die Meldung zur Gebäude- und Wohnungszählung wurde den Auskunftgebenden darüber hinaus neben den klassischen Anmeldedaten ein QR-Code angeboten. Zusätzlich zur URL der Login-Seite des Online-Fragebogens enthielt der QR-Code die Zugangsnummer, die beim Scannen automatisch in die Anmeldemaske übernommen wurde. Nutzten Auskunftgebende diesen Weg der Anmeldung, war nur noch der 12-stellige Aktivierungscode einzugeben. Dieses Vorgehen vereinfachte die Anmeldung und minimierte weitere Fehlerquellen, beispielsweise durch Vertippen. [↘ Grafik 2](#)

Grafik 2

Anmeldedaten mit QR-Code auf einem Musteranschreiben der Gebäude- und Wohnungszählung im Zensus 2022

2.2 Responsives Webdesign und mobil optimierte Fragebogen

Waren die Online-Fragebogen für den Zensus 2011 noch ausschließlich für die Nutzung auf Desktop-Computern ausgerichtet, machten es der digitale Wandel und die damit einhergehende verbreitete Nutzung von Smartphones und Tabletcomputern zwingend erforderlich, für den Zensus 2022 ein sogenanntes responsives Design

umzusetzen. Responsives Webdesign bedeutet, dass die Bildschirmansicht auf die Eigenschaften des verwendeten Endgeräts reagiert und die aufgerufenen Seiten in einer optimierten Variante anzeigt – jeweils angepasst an Betriebssystem, Browser, Endgerät und Bildschirmgröße. So werden die Fragebogen sowohl auf einem großen PC-Bildschirm als auch auf einem deutlich kleineren Smartphone-Bildschirm übersichtlich und anwendungsfreundlich dargestellt. Auch deshalb wurde – anders als noch beim Zensus 2011 – nur eine Frage je Seite angezeigt, sodass eine Bedienung ohne Scrollen möglich war. [↘ Grafiken 3 und 4](#)

Grafik 3

Frage nach Gebäudetyp und Baujahr im Fragebogen der Gebäude- und Wohnungszählung 2011

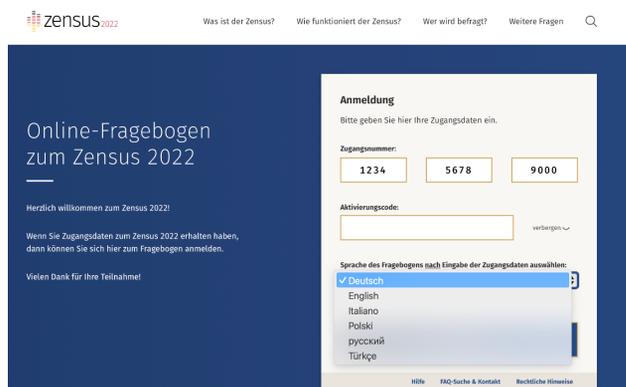
Grafik 4

Frage nach dem Gebäudetyp im Fragebogen der Gebäude- und Wohnungszählung 2022

2.3 Mehrsprachigkeit

Um Befragten, deren Muttersprache nicht Deutsch ist, die Meldung zum Zensus zu vereinfachen, wurden die Fragebogen zur Gebäude- und Wohnungszählung in fünf Sprachen (Englisch, Italienisch, Polnisch, Russisch und Türkisch) angeboten. Die Fragebogen zur Haushaltebefragung auf Stichprobenbasis und zur Befragung in Wohnheimen waren in 15 Sprachen (Arabisch, Bulgarisch, Englisch, Französisch, Griechisch, Italienisch, Kroatisch, Persisch, Polnisch, Portugiesisch, Rumänisch, Russisch, Spanisch, Türkisch, Ukrainisch) verfügbar.⁵ Die Bereitstellung der fremdsprachigen Fragebogen ausschließlich in der Online-Version sollte die Online-Quote zusätzlich positiv beeinflussen. Die gewünschte Sprache konnte in der Anmeldemaske auf der Homepage des Zensus 2022 ausgewählt werden und der Fragebogen wurde nach Eingabe der Zugangsdaten in der entsprechenden Sprache angezeigt. [↗ Grafik 5](#)

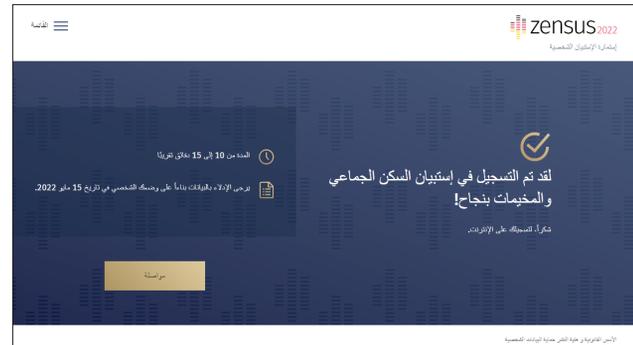
Grafik 5
Anmeldemaske mit Sprachenauswahl für den Fragebogen der Gebäude- und Wohnungszählung im Zensus 2022



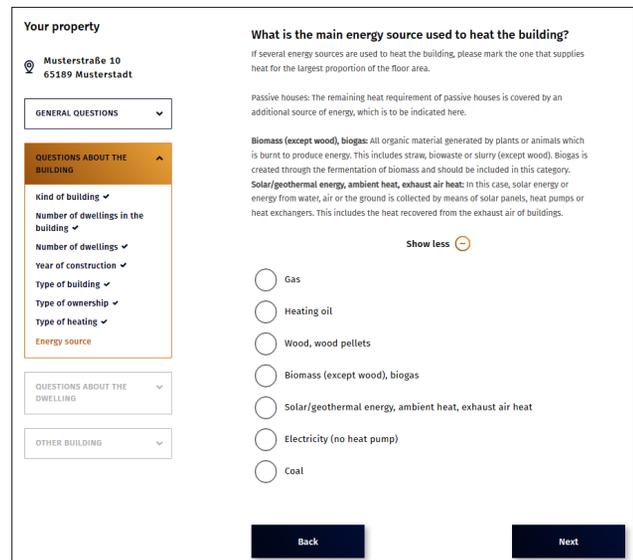
Beispiele für Fragebogenteile in verschiedenen Sprachen zeigen die [↗ Grafiken 6, 7 und 8](#).

⁵ Im Zensus 2011 wurden Übersetzungshilfen als PDF-Download angeboten. Über Auswertungen der Zugriffe auf die Übersetzungshilfen im Zensus 2011 sowie auf Grundlage von Mikrozensusdaten zur im Haushalt gesprochenen Sprache beziehungsweise zum Migrationshintergrund von Personen mit Wohneigentum erfolgte die Auswahl der Sprachen für den Zensus 2022.

Grafik 6
Startseite des Fragebogens der Befragung an Wohnheimen in arabischer Sprache im Zensus 2022



Grafik 7
Frage nach dem Energieträger in englischer Sprache im Fragebogen der Gebäude- und Wohnungszählung im Zensus 2022



Grafik 8
Frage zum Geburtsdatum in französischer Sprache im Fragebogen der Haushaltebefragung im Zensus 2022



2.4 Weitere Anpassungen

Außerdem wurde eine **Offline-Funktionalität** implementiert. Diese sorgte dafür, dass die leeren Fragebogenformulare nach dem Aufrufen in den Gerätespeicher (Cache) geladen wurden, und ermöglichte ein Ausfüllen auch bei schlechter oder unterbrochener Internetverbindung: Jeder Klick auf die jeweils nächste Fragebogen-seite sicherte den aktuellen Stand der eingegebenen Daten im Cache. So konnte nach einem unvorhergesehenen Verbindungsabbruch die Eingabe dort fortgesetzt werden, wo sie unterbrochen worden war. Beim Abmelden und beim erfolgreichen Absenden des Fragebogens wurde die Sicherung der Eingaben automatisch aus dem Cache gelöscht.

Neben der im Hintergrund ablaufenden Sicherung der Daten im Cache konnten Auskunftgebende zusätzlich eine **lokale Sicherung** aktiv auswählen. Diese konnte zu einem späteren Zeitpunkt erneut in das Formular geladen werden – beispielsweise wenn sich während des Ausfüllvorgangs herausstellte, dass erforderliche Informationen erst recherchiert werden mussten und der Fragebogen deshalb nicht vollständig ausgefüllt werden konnte.

Um den für Selbstausfüller entwickelten Online-Fragebogen auch für die **Erhebung durch Interviewpersonal** nutzbar zu machen, wurden die Formulare dahingehend angepasst, dass bei Nutzung einer speziellen Interviewerkennung die Eingabe mehrerer Meldungen ohne erneutes Login möglich war.

Darüber hinaus erfolgte eine **Vielzahl weiterer Anpassungen**, die allesamt darauf ausgerichtet waren, den Online-Meldeweg attraktiv und anwendungsfreundlich zu gestalten und so die Grundlage für eine hohe Online-Teilnahmebereitschaft zu schaffen, wie

- › die Optimierung der Seitenaufzugsgeschwindigkeit,
- › die Verwendung gängiger Design-Symbole (Beispiele dafür sind das sogenannte Burger-Symbol, wodurch eine Menüleiste für weitere Punkte aufgeklappt wird, das Plus-Zeichen, um Text aufzuklappen, beziehungsweise das Minus-Zeichen, um Text zuzuklappen),
- › der Verzicht auf Bedienelemente oder Menüs, die für die Beantwortung des Fragebogens nicht zwingend erforderlich sind (beispielsweise von der Darstellung auf verschiedenen Reitern abzusehen),

- › barrierefreie Fragebogen (große Schriftgröße, Bedienung und Eingabe allein durch Nutzung der Tastatur möglich).

3

Besonderheiten der einzelnen Online-Fragebogen im Zensus 2022

Neben den bisher beschriebenen allgemeinen Anforderungen an die Online-Fragebogen des Zensus 2022 gab es einige erhebungsteilspezifische Anforderungen, die die folgenden Abschnitte erläutern.

3.1 Gebäude- und Wohnungszählung

Die Besonderheit des Fragebogens der Gebäude- und Wohnungszählung im Vergleich zu den anderen Erhebungen im Zensus 2022 bestand vor allem in der Funktionalität, zusätzliche Gebäude anzulegen.

Stimmten Befragte im Zensus 2011 der Frage, ob weitere Gebäude an der Anschrift – zum Beispiel Hinterhäuser ohne eigene Hausnummer – vorhanden sind, zu, mussten sie das zuständige Statistische Landesamt kontaktieren und neue Zugangsdaten zum Online-Fragebogen für das weitere Gebäude anfordern. Die Befragten mussten das weitere Gebäude über einen neuen Online-Fragebogen erfassen. Dies bedeutete sowohl für die Befragten als auch für die Statistischen Ämter der Länder Mehraufwand. Zudem setzte dieses Vorgehen voraus, dass die Befragten die Statistischen Ämter der Länder auch tatsächlich kontaktierten, um die neuen

Grafik 9

Weitere Gebäude an der Anschrift im Online-Fragebogen der Gebäude- und Wohnungszählung im Zensus 2022

The screenshot shows a web interface for a questionnaire. At the top, it says 'Ihr Objekt' followed by the address 'Musterstraße 10, 65189 Musterstadt'. Below this are three dropdown menus: 'ALLGEMEINE FRAGEN', 'GEBÄUDEFRAGEN', and 'WOHNUNGSFRAGEN'. The 'GEBÄUDEFRAGEN' menu is currently expanded, showing a sub-menu 'WEITERES GEBÄUDE' which is highlighted in orange. To the right of the form, there is a text box with the following message: 'Sie haben angegeben, dass sich ein weiteres Gebäude an dieser Anschrift (Musterstraße 10, 65189 Musterstadt) in Ihrem Eigentum oder Ihrer Verwaltung befindet, für das Sie keine Zugangsnummer erhalten haben. Die Fragen zu diesem Gebäude folgen jetzt.' Below this message are two buttons: 'Zurück' and 'Weiter'.

Zugangsdaten anzufordern. Im Zensus 2022 war es hingegen möglich, direkt nach der Frage zum weiteren Gebäude im selben Online-Fragebogen dieses weitere Gebäude selbst zu erfassen. Insgesamt bot der Online-Fragebogen die Möglichkeit, bis zu drei weitere Gebäude zu erfassen. Dies erhöhte den Komfort der Befragten und schonte Ressourcen der Statistischen Ämter der Länder, weil auf die Anforderung eines weiteren Fragebogens beziehungsweise neuer Zugangsdaten verzichtet werden konnte. [↪ Grafik 9](#)

3.2 Haushaltebefragung und Erhebung an Wohnheimen

Neben den allgemeinen übergreifenden Anpassungen wie Responsivität, Mehrsprachigkeit oder dem vereinfachten Login-Prozess gab es für die Formulare der Haushaltebefragung und Erhebung an Wohnheimen weitere spezifische Anforderungen. So wurden bei den Themenbereichen Ausbildung und Erwerbstätigkeit die Auskunftgebenden nach ihrer Meisterausbildung, der Berufsbezeichnung sowie ihrem Arbeitsort befragt. Hierbei handelte es sich um offene Fragen, deren Beantwortung für gewöhnlich als Freitext erfolgt. Diese Freitextangaben müssen anschließend im Zuge der Datenaufbereitung kodiert werden – was einen erheblichen manuellen Aufwand in den Statistischen Landesämtern bedeutet. Um dies zu umgehen beziehungsweise den Aufwand zu minimieren, wurden an dieser Stelle in den Fragebogenformularen Datenbanken hinterlegt (Klassifi-

Grafik 10

Einbindung des Klassifikationservers zur Meistersuche im Fragebogen der Haushaltebefragung im Zensus 2022

kationsserver mit Listen aller Berufe und Meisterausbildungen beziehungsweise eine Liste aller Postleitzahlen mit den dazugehörigen Gemeinden). Aus diesen konnten die Befragten nach der Eingabe eines Schlagwortes aus den daraufhin unterbreiteten Vorschlägen die zutreffende Angabe auswählen. [↪ Grafik 10](#)

Auf diese Weise konnte bei einem Großteil der Meldungen die nachgelagerte Kodierung entfallen.

3.3 Erhebung an Gemeinschaftsunterkünften

Die Besonderheit der Erhebung an Gemeinschaftsunterkünften bestand darin, dass die Einrichtungsleitungen auskunftspflichtig waren und die benötigten Angaben zu allen Bewohnerinnen und Bewohnern der Einrichtung im Online-Fragebogen (dem sogenannten Erfassungsbogen) tätigen mussten. Die persönliche Erhebung durch Erhebungsbeauftragte in einer Gemeinschaftsunterkunft könnte möglicherweise die Bewohnerinnen und Bewohner sozial benachteiligen (zum Beispiel in Justizvollzugsanstalten oder in Gemeinschaftsunterkünften für Geflüchtete). Ebenfalls wäre sie bei einigen Personen aufgrund deren gesundheitlicher Verfassung nicht oder nur schwer durchzuführen (zum Beispiel in Alten- und Pflegeheimen).

Je nach Größe der Einrichtung konnte es also sein, dass die auskunftspflichtige Einrichtungsleitung für mehrere hundert Personen Angaben übermitteln musste. Daher wurde der Erfassungsbogen als tabellarische Liste angelegt, in der die Einrichtungsleitungen am PC die Daten zu jeder Bewohnerin und jedem Bewohner in je eine Zeile eintragen konnten. [↪ Grafik 11](#) Dies war eine Verbesserung gegenüber dem Fragebogen des Zensus 2011, in dem die Angaben zu jeder Person auf einer eigenen Formulareseite eingetragen werden mussten, wodurch der Ausfüllprozess weniger übersichtlich war. [↪ Grafik 12](#)

Grafik 11

Tabellarisches Formular zur Eintragung der Daten der Personen in Gemeinschaftsunterkünften im Erfassungsbogen der Erhebung an Gemeinschaftsunterkünften im Zensus 2022

Bitte machen Sie für jede Bewohnerin und jeden Bewohner Ihrer Einrichtung nachfolgende Angaben zum Stand 15. Mai 2022.

Falls es Ihnen nicht möglich ist, alle Angaben zu einer Person zu machen bzw. der Geburtsname identisch mit dem Nachnamen ist, lassen Sie das betreffende Feld leer.

Über das Papierkorb-Symbol können Sie eine Zeile löschen. Eine neue Zeile können Sie unterhalb der Tabelle über das Plus-Symbol hinzufügen.

Nachname <small>(i)</small>	Geburtsname <small>(falls abweichend vom Nachnamen) <small>(i)</small></small>	Vorname/-n <small>(i)</small>	Geschlecht <small>(nach Geburtenregister) <small>(i)</small></small>	Familienstand <small>(i)</small>	Geburtsdatum <small>(TT.MM.JJJJ) <small>(i)</small></small>	Geburtsort	Geburtsstaat <small>(i)</small>	Staatsangehörigkeit	
Mustermann	Gabler	Erika	Weiblich <input type="checkbox"/>	Verheiratet <input type="checkbox"/>	18.08.1964	Musterstadt	Deutschland <input type="checkbox"/>	Nur deutsch <input type="checkbox"/>	
Mustermann		Max Otto	Männlich <input type="checkbox"/>	Verheiratet <input type="checkbox"/>	24.02.1976	Musterstadt	Deutschland <input type="checkbox"/>	Nur deutsch <input type="checkbox"/>	

Sie haben Zeilen für 2 Personen angelegt.

Neue Zeile hinzufügen

Zurück **Weiter**

Grafik 12

Formular zur Eintragung der Daten der Personen in Gemeinschaftsunterkünften aus dem Zensus 2011

Angaben zu den Personen

Person 1

Person 1 von 1

Familienname

Geburtsname (falls abweichend)

Vorname/-n info

Geschlecht

Familienstand info

Geburtsdatum (TTMMJJJJ)

Geburtsort info

Geburtsstaat info

Erste Staatsangehörigkeit info

Zweite Staatsangehörigkeit info

Datum des Bezugs (TTMMJJJJ) info

Status der Wohnung

[Weitere Person bearbeiten](#)

Grafik 13

Auswahl zwischen Import und manueller Eingabe im Erfassungsbogen der Erhebung an Gemeinschaftsunterkünften im Zensus 2022

Bitte übermitteln Sie die erforderlichen Angaben zu allen Bewohnerinnen und Bewohnern der Einrichtung. Bitte beziehen Sie Ihre Angaben auf die Situation der Einrichtung am 15. Mai 2022. Sie haben zwei Möglichkeiten Ihre Daten im Formular zu erfassen:

<p>Import ⓘ</p> <p>Sie können die Daten direkt in das Formular importieren. Beachten Sie dabei bitte die Anleitung zum Import von Daten und gehen Sie folgendermaßen vor:</p> <ol style="list-style-type: none">1. Erstellen Sie eine Datei im CSV-Format. Dazu können Sie unsere Vorlage nutzen.2. Fügen Sie die erforderlichen Angaben zu allen Bewohnerinnen und Bewohnern der Einrichtung in die Datei ein.3. Speichern Sie die Datei im CSV-Format.4. Laden Sie die Datei über das Menü oben rechts in das Formular. Nachdem Sie Ihre Daten importiert haben, können Sie diese direkt im Formular noch einmal überprüfen und gegebenenfalls weiter bearbeiten.	<p>Manuelle Eingabe ⓘ</p> <p>Sie können die Daten auch hier im Formular eingeben. Dazu tragen Sie bitte die Anzahl der Personen in untenstehendes Feld ein und bestätigen Sie diese mit "Weiter". Auf der nächsten Seite erscheint eine Tabelle, in die Sie Ihre Angaben eintragen können.</p> <p>Bitte geben Sie die Anzahl der Bewohnerinnen und Bewohner dieser Einrichtung an und nutzen danach die Schaltfläche "Weiter".</p> <input type="text"/>
---	--

Zurück Weiter

Um den Erfassungsbogen bestmöglich an die Lebensrealität der Einrichtungsleitungen anzupassen und ihnen die Meldung zu erleichtern, war es neben der manuellen Eingabe der benötigten Informationen auch möglich, eine CSV-Datei mit den Daten der Bewohnerinnen und Bewohner zu importieren. ➤ Grafik 13

Dies war besonders für Einrichtungsleitungen größerer Einrichtungen relevant, da sie bereits vorhandene Daten aus ihren Verwaltungssystemen in diese Vorlagen integrieren konnten und damit das manuelle Ausfüllen des Formulars entfiel.

Auch die lokale Sicherung war für die Einrichtungsleitungen sehr hilfreich, da dadurch das Ausfüllen des Formulars nicht zwingend ohne Unterbrechung erfolgen musste.

Es war davon auszugehen, dass die Einrichtungsleitungen Deutsch sprechen und verstehen, daher war es nicht notwendig, dieses Formular in verschiedenen Sprachen anzubieten. Ebenso wurde auf ein responsives Webdesign verzichtet, da wegen der angenommenen Nutzung von Desktop-PCs oder Laptops keine Anpassung des Designs an Tablets oder Smartphones erfolgen musste. Des Weiteren wurde die Ausstattung der Einrichtungen mit PCs mit Internetanschlüssen vorausgesetzt, weshalb eine Papier-Version des Erfassungsbogens nur als Notfall-Alternative erstellt wurde.⁶ Diese kam nur in

⁶ Diese Annahmen konnte der durchgeführte qualitative Pretest bestätigen.

Ausnahmefällen zum Einsatz, falls Einrichtungsleitungen zum Beispiel aufgrund technischer Probleme nicht über das Online-Formular melden konnten. Die über diesen Weg vereinzelt eingegangenen Papier-Fragebogen wurden dann in den Online-Fragebogen für das Interviewpersonal übertragen und konnten somit auch als Online-Dateneingänge weiterverarbeitet werden.

4

Pretests

Vor der Durchführung einer Erhebung wird in der Regel ein Pretest (Vortest) durchgeführt, um beispielsweise die Erhebungsinstrumente auf ihre Verwendbarkeit zu überprüfen. Dabei testen Personen, die zur Zielgruppe der Erhebung gehören, das Erhebungsinstrument. Auch im Zensus 2022 wurden unterschiedliche Pretests durchgeführt, mit Fokus unter anderem darauf, die Online-Fragebogen (zum Beispiel Login-Vorgang, Navigation) zu testen sowie Verständnisschwierigkeiten und gegebenenfalls technische Fehler zu ermitteln und zu beseitigen. Dabei standen vor allem die umfangreichen Änderungen am Online-Fragebogen im Vergleich zu 2011 im Mittelpunkt. Denn nur wenn alles reibungslos funktioniert kommen die bereits genannten Vorteile einer Online-Befragung optimal zum Tragen.

Eine Besonderheit des Pretests der Haushaltebefragung, der Erhebung an Wohnheimen und der Gebäude- und Wohnungszählung im Zensus 2022 war die Auswertung der Paradata. Paradata sind Daten, die während der Befragung quasi nebenbei, das heißt ohne Zutun der Befragten, entstehen. So wurden zum Beispiel das Betriebssystem, der benutzte Browser, die verwendete Browserversion, die Auflösung des Bildschirms sowie Startdatum und Startuhrzeit erfasst. Zusätzlich wurde die Uhrzeit aufgezeichnet, bei der auf welcher Seite des Fragebogens welche Aktion ausgeführt wurde. Paradata wurden auch dann ermittelt, wenn Befragte den Fragebogen abbrachen und nicht beendeten.

Diese Paradata ermöglichen Aussagen zum Verhalten der Befragten. So konnten zum Beispiel die Dauer der Befragung, die durchschnittliche Verweildauer auf einer Fragebogenseite und die Aktionen der Befragten, wie das Bedienen der Zurück-Taste, gespeichert und ausgewertet werden. Des Weiteren konnten die Paradata in Verbindung mit den Befragungsdaten gebracht werden.

4.1 Pretest des Fragebogens der Gebäude- und Wohnungszählung

Der Online-Fragebogen der Gebäude- und Wohnungszählung wurde mit qualitativen und quantitativen Pretests auf seine Verwendbarkeit untersucht. Im Folgenden stehen die Ergebnisse der Paradata-Auswertung des quantitativen Pretests im Fokus, weil dort der Schwerpunkt unter anderem auf der Bedienbarkeit des Online-Fragebogens lag. Im Oktober 2019 wurden 49 500 Personen für den quantitativen Pretest angeschrieben, von denen 6 583 den Online-Fragebogen ausfüllten. Neben der Auswertung der Paradata wurden Feedback-Fragen am Ende des Online-Fragebogens gestellt sowie mit einer Stichprobe ein strukturiertes Telefoninterview durchgeführt.

Bei der Auswertung lag ein besonderer Schwerpunkt auf den technischen Neuerungen des Online-Fragebogens:

Etwa die Hälfte der Befragten hat das Menü – dargestellt als Burger-Symbol – aufgerufen. Auch die eingeklappten Erklärungstexte haben etwa 40 % der Befragten über das Plus-Zeichen ausgeklappt.

Die Auswertung des Betriebssystems in Kombination mit der Bildschirmauflösung ergab, dass 19 % der Befragten

den Fragebogen auf einem Smartphone oder Tablet ausgefüllt und die restlichen Befragten dazu den Computer genutzt haben. Den QR-Code auf dem Anschreiben verwendeten etwa 35 % der Smartphone-Nutzenden gegenüber nur etwa 4 % aller Befragten.

Um zu ermitteln, ob die Befragten Schwierigkeiten bei der Eingabe der Zugangsnummer und des Aktivierungscodes hatten, wurden die Fälle herausgefiltert, die das Feld der Zugangsnummer innerhalb von 60 Sekunden mehrfach angeklickt haben. Ergebnis war, dass etwa 5 % der Befragten eine Korrektur ihrer Zugangsdaten beim Login vorgenommen haben. Zum Zeitpunkt des Pretests war der Gestaltungsspielraum beim Design der Zugangsdaten bereits erschöpft, sodass hier vor allem der QR-Code als Verbesserung vorgesehen wurde: Durch den personalisierten QR-Code mussten die Befragten nur noch 12 statt 24 Zeichen eingeben, was die Fehleranfälligkeit weiter verringerte.

4.2 Pretest des Erfassungsbogens für Gemeinschaftsunterkünfte

Anfang 2021 wurden 18 Einrichtungsleitungen oder weitere verantwortliche Beschäftigte unterschiedlicher Bereichsarten und unterschiedlicher Einrichtungsgrößen um ihre Teilnahme an einem qualitativen Pretest gebeten.

Die Testpersonen wurden zufällig in zwei Testgruppen eingeteilt. Gruppe 1 hatte zunächst als „Hausaufgabe“ die zur Verfügung gestellte CSV-Importdatei vorab ausgefüllt. Diese wurde dann in einem Einzelinterview mit Mitarbeitenden des Pretestbereichs des Statistischen Bundesamtes besprochen. Anschließend wurden dieselben Testpersonen gebeten, den Erfassungsbogen zu testen. Der Fokus lag hierbei auf der Verständlichkeit der CSV-Importanleitung und der Nutzung der Importfunktion. Bei Gruppe 2 lag das Augenmerk auf der Überprüfung der Befragtenführung des Erfassungsbogens und dem spontanen Eindruck des Formulars. Die Testpersonen wurden beim Umgang mit den verschiedenen Bestandteilen des Formulars beobachtet und später wurden ihre Eindrücke in webbasierten Einzelinterviews besprochen. Der Pretest erfolgte unter realen Bedingungen in der gewohnten Arbeitsumgebung der Testpersonen. Insbesondere die Upload-Funktion von CSV-Dateien führte zu Unklarheiten, sodass der Umgang

damit und ihre Vorteile noch genauer und verständlicher in den Erläuterungstexten dargestellt werden mussten. Der Erfassungsbogen selbst in seiner Tabellenform wurde überwiegend positiv und der Umgang damit als intuitiv verständlich wahrgenommen. Durch den Pretest konnten Probleme oder Unklarheiten bei der Bearbeitung des Erfassungsbogens ermittelt und dieser anhand der Erkenntnisse vor Produktivgang nochmals optimiert werden.

4.3 Pretest der Fragebogen der Haushaltebefragung und Befragung an Wohnheimen

Auch die Fragebogen der Haushaltebefragung sowie der Befragung an Wohnheimen wurden vor ihrem Einsatz sowohl einem qualitativen als auch einem quantitativen Pretest unterzogen. Ziel war unter anderem, Erkenntnisse zur Nutzung des Online-Meldeweges zu gewinnen und die Praxistauglichkeit der Erhebungsinstrumente zu erproben.

Beim qualitativen Pretest füllte eine Gruppe von 25 Testpersonen im Januar 2019 die Online-Fragebogen im Pretestlabor des Statistischen Bundesamtes aus, wurde dabei beobachtet und anschließend hinsichtlich Ver-

ständlichkeit und Nutzerfreundlichkeit befragt. Insgesamt kamen die Testpersonen gut mit dem Online-Erhebungsinstrument zurecht, sowohl bei der Bearbeitung am PC als auch am Smartphone. Die Beobachtungen und die anschließenden Gespräche zeigten dennoch einige Handhabungsprobleme und damit Verbesserungspotenzial auf. Insbesondere bei Aufklappenüsil fiel es den Testpersonen bei sehr umfangreichen Antwortkategorien nicht immer leicht, die unterschiedliche Darstellung von Ober- und Unterkategorien zu erkennen.

➤ Grafik 14

Beim quantitativen Pretest im Zeitraum von November 2019 bis März 2020 wurde der komplette Erhebungsablauf unter weitgehend realistischen Bedingungen und mithilfe zufällig ausgewählter Testpersonen erprobt. Hierzu haben die teilnehmenden Statistischen Landesämter⁷ insgesamt 8 632 Personen befragt. Darunter befanden sich knapp 7 000 Personen, die auch Auskunft zu den erweiterten Zensusmerkmalen Bildung und Erwerbstätigkeit gaben und somit entsprechend der Erhebungskonzeption die wesentliche Zielgruppe der Online-Fragebogen darstellten. Das methodische Vorgehen entsprach dabei dem der Haupterhebung, das heißt

⁷ Teilgenommen haben alle Statistischen Landesämter mit Ausnahme des Hessischen Statistischen Landesamtes und des Statistischen Amtes Mecklenburg-Vorpommern.

Grafik 14

Pretest der Fragebogen der Haushaltebefragung und Befragung an Wohnheimen im Zensus 2022: Beispiele für umfangreiche Antwortkategorien mit Ober- und Unterkategorien und sehr umfangreichen Ausprägungslisten

The image shows two side-by-side screenshots of survey forms. The left screenshot displays a list of education levels with checkboxes. The right screenshot displays a list of occupation categories with radio buttons.

Left Screenshot (Education):

- Zwei- oder dreijährige Ausbildung an einer Ausbildungsstätte/Schule für Gesundheits- und Sozialberufe (z. B. Krankenpflege, PTA, MTA)
- Meisterin/Meister
- Technikerin/Techniker oder gleichwertiger Fachschulabschluss
- Abschluss einer Fachschule der DDR
- Abschluss einer Fachakademie (nur in Bayern)
- Keinen der genannten beruflichen Ausbildungsabschlüsse
- Hochschulabschluss
 - Bachelor
 - Master
 - Diplom (FH)
 - Diplom an Berufsakademien
 - Diplom (Universität), Lehramtsprüfung, Staatsexamen, Magistra/Magister, künstlerischer Abschluss oder vergleichbarer Abschluss an einer Universität
 - Promotion
 - Keinen der genannten Hochschulabschlüsse

Right Screenshot (Occupation):

- Baugewerbe, Hoch- und Tiefbau
- Handel, Verkehr und Lagerei, Gastgewerbe/Beherbergung und Gastronomie
- Information und Kommunikation z. B. Telekommunikation, Dienstleistungen der Informationstechnologie, Medien und Verlagswesen
- Banken/Finanz- und Versicherungsdienstleister
- Grundstücks- und Wohnungswesen
- Öffentliche Verwaltung, Verteidigung, Sozialversicherung, Erziehung und Unterricht, Gesundheits- und Sozialwesen
 - Öffentliche Verwaltung, Gerichte, Öffentliche Sicherheit und Ordnung, Verteidigung, Sozialversicherung
 - Erziehung und Unterricht z. B. Hochschulen, Schulen, sonstige Schulen (auch Fahrschulen), Kindergärten
 - Gesundheits- und Sozialwesen z. B. Krankenhäuser, Arztpraxen, Alten- und Pflegeheime
- Freiberufliche, wissenschaftliche und technische Dienstleistungen sowie sonstige wirtschaftliche Dienstleistungen
- Sonstige Dienstleistungen

es wurden Personen an zufällig ausgewählten Anschriften befragt; nur so konnten zuverlässige Erkenntnisse – auch über das Funktionieren der Online-First-Strategie – gewonnen werden.

Neben der Auswertung von Paradata fand eine Nachbefragung am Ende des Online-Fragebogens statt, mit der ergänzend Feedback der Befragten zur Handhabbarkeit und Bedienung des Online-Fragebogens sowie zum Anmeldeprozess eingeholt wurde. Sofern die Teilnehmenden nicht online gemeldet hatten, sondern eine computergestützte Telefonbefragung erfolgte, wurden abschließend die Gründe für den Verzicht auf die Online-Teilnahme erfragt.

Bereits im Zuge des quantitativen Pretests zeigte sich eine hohe Online-Teilnahmebereitschaft (57%). Generell wurde die Handhabbarkeit des Online-Fragebogens überwiegend als sehr gut bis eher gut bewertet. Knapp drei Viertel der Befragten hatten mit dem Anmeldevorgang keinerlei Schwierigkeiten. Ebenfalls drei Viertel der Befragten füllten den Fragebogen am PC aus, 17% nutzten ein Smartphone und 8% ein Tablet. Je nach Fragebogen dauerte der Ausfüllvorgang zwischen fünf und zehn bis zwölf Minuten – eine wichtige Erkenntnis für die spätere Bewerbung der Online-Fragebogen in der Haupterhebung.

Die Auswertung der Ergebnisse zeigte Optimierungspotenzial an unterschiedlichen Stellen. Neben inhaltlichen Komponenten wie der Modifizierung von Frageelementen (wie etwa Formulierungen, Hinweistexte, Antwortoptionen) gab es auch Anpassungsbedarf in der technischen Umsetzung. So wurde beispielsweise die elektronisch unterstützte Abfrage des Berufs mittels Suchfunktion in einer hinterlegten Datenbank nicht von allen teilnehmenden Personen wahrgenommen.

Basierend auf den Erkenntnissen der Pretests wurden alle Fragebogen im Nachgang überarbeitet und somit für die Auskunftspflichtigen bestmöglich optimiert.

5

Fazit und Ausblick

Auch wenn im Zensus 2022 nicht komplett auf den Einsatz von Papierfragebogen verzichtet werden konnte, haben die hier beschriebenen Maßnahmen dazu beigetragen, dass eine sehr hohe Online-Quote erreicht werden konnte: Rund 82% der Meldungen zur Gebäude- und Wohnungszählung sowie 73% der Meldungen zu den Zusatzmerkmalen Bildung und Erwerbstätigkeit bei der Haushaltebefragung und der Befragung an Wohnheimen⁸ haben die Auskunftspflichtigen selbst über die jeweiligen Online-Fragebogen übermittelt. Die ausgefüllten Papierfragebogen der Gebäude- und Wohnungszählung 2022 wären somit übereinandergestapelt niedriger gewesen als die Zugspitze mit 2 962 Metern Höhe.

Über die Auswertung der Online-Quote hinaus sind Aussagen darüber, wie die Befragten mit den Online-Fragebogen zurechtgekommen sind, nur schwer zu treffen. Wie die verschiedenen Funktionalitäten des Online-Fragebogens angenommen wurden, hätte – wie in den Pretests – über die Analyse der Paradata herausgefunden werden können, ebenso wie weitergehende Erkenntnisse zum Nutzungsverhalten. Allerdings hätten die Befragten aus Datenschutzgründen dazu aktiv ihre Einwilligung geben müssen. Um die Teilnahmebereitschaft der 34 Millionen Befragten nicht einzuschränken, wurde im Zensus 2022 daher auf die Erhebung der Paradata verzichtet. Für künftig durchzuführende Statistiken – insbesondere, wenn Online-Fragebogen neu oder in überarbeiteter Form eingesetzt werden – scheint eine individuelle Prüfung sinnvoll, ob die Nutzung der Paradata gewinnbringende Erkenntnisse erwarten lässt.

Die Anpassung des IDEV-Designs hat auch über den Zensus 2022 hinaus Auswirkungen: Seit August 2024 ist eine neue IDEV-Version in der amtlichen Statistik im Einsatz, für die der Zensus als Pilotprojekt diente. So werden in Zukunft alle Online-Meldungen der amtlichen Statistik im modernen und zeitgemäßen Design umgesetzt.

⁸ Die berechnete Quote bezieht sich auf zwölf Bundesländer. Für vier Bundesländer wurde auf eine Auswertung verzichtet, da dort die Interviewerinnen und Interviewer vorrangig eine Erhebung per Tablet durchführten. Dieser zusätzliche Befragungsmodus in diesen vier Ländern hat das Online-Selbstaussfüllerformular zum Großteil ersetzt.

Die Optimierung der Online-Fragebogen, aber auch das konsequente Bewerben des Online-Meldewegs auf der Zensus-Webseite, in den Anschreiben an die Befragten sowie durch das Erhebungspersonal haben dazu beigetragen, dass die Online-Quote im Vergleich zum Zensus 2011 deutlich gestiegen ist. Während 2011 ein Drittel der Befragten den Online-Meldeweg wählte, waren es 2022 bereits etwa drei Viertel. Mit Blick auf den nächsten Zensus im Jahr 2031 stellt sich allerdings die Frage, ob nicht gänzlich auf Papierfragebogen verzichtet werden kann.

Aus heutiger Sicht mag die Fokussierung auf die Online-Meldewege naheliegend sein. Während der Vorbereitungen zum Zensus 2022 war deren Akzeptanz jedoch nicht einfach zu erreichen. Dies lässt sich am besten an der Ausgestaltung von § 23 Zensusgesetz 2022 ablesen: Einerseits lautet der zweite Satz in Absatz 1 *„Die Auskunftserteilung erfolgt grundsätzlich elektronisch.“*, andererseits steht im selben Absatz *„Im Fall der schriftlichen Auskunftserteilung können die ausgefüllten Erhebungsvordrucke gebührenfrei übersendet werden, ...“*. Die zweite Formulierung kannte noch nicht einmal das Zensusgesetz 2011 – damals mussten die Auskunftgebenden das Porto selbst tragen. Die in diesem Artikel berichteten Vorteile zeigen jedoch deutlich, dass sich der Mut, sich auf die Online-First-Strategie zu verständigen, mehr als gelohnt hat und auch zwingend erforderlich war, um eine zeitgemäße Erhebung durchführen zu können. Im Hinblick auf den kommenden Zensus sollten der Mut und die Offenheit für Neuerungen beibehalten werden. 

LITERATURVERZEICHNIS

ARD/ZDF-Onlinestudie (im Auftrag der ARD/ZDF-Forschungskommission). *ARD/ZDF-Onlinestudie 2023*. 2023. [Zugriff am 5. November 2024]. Verfügbar unter: www.ard-zdf-onlinestudie.de

Estatico, Sabrina/Tornow, Tobias/Walther, Sabrina. *Generalprobe für den Zensus 2022: Erkenntnisse aus der Pilotstudie zur Haushaltebefragung*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2021, Seite 126 ff.

Freier, Benjamin/Mosel, Juliane. *Online First als Leitgedanke für effiziente Primärerhebungen beim Zensus 2021*. In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 46 ff.

Kleinhubbert, Guido. *Truthuhn, unzerteilt*. In: Der Spiegel. Ausgabe 31/2017, Seite 22 ff.

Neutze, Michael/Pfahl, Miriam/Schweinert-Albinus, Stefan. *Das Innovationspotenzial des Zensus 2021*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 37 ff.

Statistisches Bundesamt. *Ausstattung privater Haushalte mit Internet und Breitbandanschluss im Zeitvergleich*. 2020. [Zugriff am 29. Oktober 2024]. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Dr. Janine Hentschke

studierte Publizistik- und Kommunikationswissenschaft sowie Betriebswirtschaftslehre mit Fokus auf sozialwissenschaftliche empirische Forschung. Sie ist seit 2020 im Statistischen Bundesamt tätig, zunächst im Zensus-Referat „Haushaltsstichprobe (Konzeption und Aufbereitung)“. Seit 2022 unterstützt sie die Projektleitung Qualitätsmanagement Zensus und befasst sich mit der Weiterentwicklung der Qualitätsmessung und -bewertung im Statistischen Verbund.

Cara-Aileen Tobies

studierte Politikwissenschaft mit Schwerpunkt sozialwissenschaftliche Datenanalyse und Volkswirtschaftslehre. Sie ist seit 2021 im Statistischen Bundesamt tätig, zunächst in der Vergabestatistik und später im Qualitätsmanagement Zensus. Hier betreut sie schwerpunktmäßig die Festlegung und Bereitstellung der Kern-Qualitätskennzahlen für die Erhebungsphase der primärstatistischen Teilprojekte sowie für die Haushaltegenerierung.

Susanne Weber und Dennis Claus

sind im Qualitätsmanagement Zensus im Statistischen Bundesamt tätig und haben durch ihre Mitarbeit bei der Festlegung und Bereitstellung der teilprojektspezifischen Kern-Qualitätskennzahlen im Zensus 2022 die Inhalte und Entstehung dieses Artikels unterstützt. Susanne Weber ist seit 2023 für die Evaluierung des Qualitätsmanagements Zensus zuständig, Dennis Claus seit 2019 für die Bestandsaufnahmen zur Umsetzung der Qualitätsrichtlinien im Zensus.

KERN-QUALITÄTSKENNZAHLEN UND ZIELWERTE IM ZENSUS 2022

Ausgewählte Ergebnisse der Qualitätsmessung

Janine Hentschke, Cara-Aileen Tobies, Susanne Weber,
Dennis Claus

📌 **Schlüsselwörter:** Rücklaufquoten – Antwortausfallquoten – Imputationsquoten – Qualitätsindikatoren – Qualitätsdatenblätter im Verbund

ZUSAMMENFASSUNG

Der Artikel stellt ausgewählte Ergebnisse zu den teilprojektspezifischen Kern-Qualitätskennzahlen und Zielwerten vor, die im Zensus 2022 veröffentlicht wurden. Die fünf relevanten Teilprojekte sind Melderegister, Gebäude- und Wohnungszählung, Erhebung an Sonderbereichen, Personenerhebung und Haushaltegenerierung. Die Ergebnisse ihrer Qualitätsmessung werden teilprojektübergreifend verglichen und anhand von Vergleichswerten der amtlichen Statistik und freiwilligen Befragungen eingeordnet. Aus den Gegenüberstellungen der Qualitätsindikatoren lassen sich wertvolle Erkenntnisse für die fachstatistischen Erhebungs- und Aufbereitungsprozesse sowie für die Weiterentwicklung des Qualitätsmanagements ableiten.

📌 **Keywords:** response rate – non-response rate – imputation rate – quality indicators – quality data sheets in the network of official statistics

ABSTRACT

The article presents selected results for core quality indicators and target values in five census subprojects, which were published in the 2022 Census. The subprojects in question are the population register, the census of buildings and housing, the survey conducted at special facilities, the survey of individuals and the household generating procedure. The quality measurement results for the subprojects are directly compared and assessed in relation to reference values from official statistics and voluntary surveys. By comparing the quality indicators, valuable insights can be drawn for statistical data collection and processing in the respective fields and for the further development of quality management procedures.

1

Einleitung

Nach Abschluss der Arbeiten im Zensus 2022 stehen die Ergebnisse zu Einwohnerzahlen und zur Gebäude- und Wohnungssituation in Deutschland bereit. Sie dienen als Planungs- und Entscheidungsgrundlage in verschiedenen Gesellschaftsbereichen wie Politik, Verwaltung und Wirtschaft. Neben der nationalen Bedeutung der Daten besteht auch ein internationales Interesse an den Zensusergebnissen, was sich beispielsweise in der Lieferverpflichtung an das Statistische Amt der Europäischen Union (Eurostat) widerspiegelt (Dittrich und andere, 2022).

Um eine hohe Datenqualität sicherzustellen, wurde der Zensus 2022 von einem systematischen Qualitätsmanagement unterstützt (Tümmler/Meinke, 2019). Zur systematischen Qualitätsmessung und -bewertung wurde unter anderem das Instrument Qualitätsdatenblätter im Verbund¹ eingesetzt. Mit diesem Instrument des Qualitätsmanagements der deutschen amtlichen Statistik lassen sich wichtige Kenngrößen der Erhebungs- und Aufbereitungsprozesse messen und bewerten. Die ableitbaren Qualitätsindikatoren (zum Beispiel Antwortausfallquote) basieren auf den ESS Quality and Performance Indicators (Eurostat, 2014). Das Europäische Statistische System (ESS) empfiehlt sie, um die Zuverlässigkeit und Genauigkeit der Ergebnisse zu bewerten (Meinke/Hentschke, 2022).

↳ Relevante Qualitätsbegriffe

Qualität in der amtlichen Statistik: Die Qualitätsanforderungen in der amtlichen Statistik entsprechen dem Bedarf der Nutzerinnen und Nutzer („fitness for use“), zum Beispiel Genauigkeit und Zuverlässigkeit der Statistikergebnisse. Qualität umfasst den gesamten Geschäftsprozess von der Bedarfsbestimmung über die Datengewinnung und -aufbereitung sowie Ergebnisverbreitung bis hin zur Evaluierung.

Qualitätsmanagement: Die Statistischen Ämter des Bundes und der Länder haben das Ziel, das erreichte Qualitätsniveau der Statistiken dauerhaft zu gewährleisten

und mit einer ganzheitlichen Perspektive die Qualität aller Produkte, Dienstleistungen und Prozesse kontinuierlich zu verbessern. Dazu setzt der Statistische Verbund eine Vielzahl von Instrumenten zum Management der Datenqualität ein.

Qualitätsdatenblätter im Verbund sind ein Instrument des Qualitätsmanagements im Statistischen Verbund, mit welchem die Qualität der Statistikerstellungsprozesse und ihrer Produkte gemessen, bewertet und dokumentiert wird. Das Produktionsdatenblatt umfasst variable quantitative Informationen zur Prozess- und Produktqualität, die die Statistikproduzierenden bei der Steuerung der Statistikerstellung unterstützen (zum Beispiel Mahnwesen). Das Indikatorenblatt enthält die Qualitätsindikatoren zur Genauigkeit der Ergebnisse, die mittels der Daten im Produktionsdatenblatt berechnet werden (zum Beispiel Antwortausfälle der Einheiten). Das Stammdatenblatt beinhaltet ergänzende, eher stabile Qualitätsinformationen zur Statistik (zum Beispiel Tage bis zur ersten Mahnung).

Kern-Qualitätskennzahlen: Eine Qualitätskennzahl ist ein quantitatives Maß beziehungsweise ein quantitativer Indikator, um die Qualität von Prozessen und Produkten zu messen und zu bewerten. Hierfür wurden im Zensus 2022 die teilprojektspezifischen Kern-Qualitätskennzahlen in Anlehnung an das Instrument Qualitätsdatenblätter im Verbund entwickelt.

Die Kern-Qualitätskennzahlen im Zensus 2022 sind eine zensuspezifische Ausgestaltung des Instruments Qualitätsdatenblätter im Verbund. Dies beinhaltet auch den erstmaligen Einsatz von bundeseinheitlichen Zielwerten in der amtlichen Statistik. Dabei ermöglicht ein Soll-Ist-Abgleich, die angestrebten Qualitätsziele systematisch und transparent zu messen und zu bewerten. An die Ausführungen hierzu von Meinke/Hentschke (2022) knüpft der vorliegende Artikel an und legt den Schwerpunkt auf die Ergebnisbetrachtung der Kern-Qualitätskennzahlen und Zielwerte. Die Qualität des Zensus ist nicht anhand einer einzelnen Kennzahl messbar. Daher sind verschiedene Kern-Qualitätskennzahlen zu berücksichtigen. Der Beitrag betrachtet nur Zielwerte, die nach Erhebungsabschluss zu erreichen sind und somit primär einen Evaluierungszweck verfolgen – zum Beispiel um über einen Ländervergleich gute Praxisbeispiele abzuleiten. Eine Kernbotschaft dieses Artikels ist es, die Vergleichbarkeit von Statistiken zu stärken und zur statis-

¹ Mit „Verbund“ ist der Statistische Verbund gemeint, den die Statistischen Ämter des Bundes und der Länder bilden.

tikübergreifenden Analyse zu motivieren. Der Zensus ist ein gutes Praxisbeispiel für den Einsatz des Instruments Qualitätsdatenblätter im Verbund, an dem sich andere Fachstatistiken orientieren können, um kennzahlenbasierte Vergleiche anzustreben. Denn wie die folgenden Ausführungen zeigen, sind nur wenige Qualitätskennzahlen tatsächlich statistikspezifisch und einzigartig. Die meisten Qualitätskennzahlen lassen sich – unter Berücksichtigung der methodischen Besonderheiten – durchaus über Teilprojekte beziehungsweise Statistiken hinweg vergleichen.

Kapitel 2 gibt einen Überblick über die relevanten Teilprojekte des Zensus 2022, die Qualitätsmessungen mittels Kern-Qualitätskennzahlen und Zielwerten durchgeführt und ihre Ergebnisse veröffentlicht haben. Die folgenden Kapitel 3 und 4 stellen die erzielten Ergebnisse für ausgewählte Kern-Qualitätskennzahlen und Zielwerte vor. Diese werden, soweit es möglich ist, teilprojektübergreifend verglichen und öffentlich verfügbaren Vergleichswerten aus dem Zensus 2011, aus dem Mikrozensus² und dem European Social Survey³ gegenübergestellt. Um die Imputationsquoten (Abschnitt 4.2) einzuordnen, wird der Zensus 2021 für England und Wales herangezogen. Das Ziel ist es, Anhaltspunkte zu geben, um die Ergebnisse der Kern-Qualitätskennzahlen zu interpretieren und einzuordnen. Es ist nicht beabsichtigt, eine vollumfängliche methodische Diskussion zu führen und für alle Kern-Qualitätskennzahlen entsprechende Vergleichswerte zugrunde zu legen. Dies ist auch nicht ohne Weiteres möglich, da Veröffentlichungen vergleichbarer Kennzahlen fehlen. Kapitel 5 fasst die resultierenden Erkenntnisse und Weiterentwicklungspotenziale für das Qualitätsmanagement zusammen.

2 Der Mikrozensus ist eine jährlich und dezentral durchgeführte statistische Erhebung auf Stichprobenbasis, die teils auskunftspflichtige, teils freiwillige Fragen enthält (Statistisches Bundesamt, 2022; Statistisches Bundesamt, 2023).

3 Der European Social Survey ist eine länderübergreifende freiwillige Befragung auf Stichprobenbasis zu gesellschaftlichen Strukturen in Europa. Sie wird seit 2001 zweijährlich durchgeführt. Fester Bestandteil sind unter anderem Fragen zu Haushaltszusammensetzung, Geschlecht und Alter der befragten Personen (European Social Survey, 2022).

2

Teilprojekte mit Kern-Qualitätskennzahlen

Die Teilprojekte im Zensus 2022 dienen unterschiedlichen Zwecken und greifen ineinander, um die Gesamtprojektziele des Zensus – Ermittlung der Einwohnerzahlen sowie Gebäude- und Wohnungszählung – zu erreichen. Beispielsweise stellt das Teilprojekt Melderegister personenbezogene Daten der Meldebehörden für die Vorbereitung und Durchführung anderer Teilprojekte bereit. Das Teilprojekt Haushaltgenerierung wiederum nutzt zur Ergebnisermittlung die Daten des Melderegisters und der Primärerhebungen (siehe auch Übersicht 1).

In fünf Teilprojekten des Zensus 2022 wurden teilprojektspezifische Kern-Qualitätskennzahlen für die Qualitätsmessung und -bewertung im Statistischen Verbund abgestimmt:

- › Melderegister,
- › Gebäude- und Wohnungszählung,
- › Erhebung an Sonderbereichen⁴,
- › Personenerhebung⁵,
- › Haushaltgenerierung.

Bei der Festlegung und Interpretation der Kern-Qualitätskennzahlen sind die teilprojektspezifischen Besonderheiten zu berücksichtigen. Beispielsweise ist zu beachten, ob es sich um eine primär- oder sekundärstatistische Erhebung handelt und welche Einheiten den Kennzahlen zugrunde liegen. So ist etwa die Quote der festgestellten Existenzen durch Erhebungsbeauftragte eine Besonderheit und ein Alleinstellungsmerkmal der Personenerhebung. Demgegenüber können

4 Zum Teilprojekt Erhebung an Sonderbereichen wird hier nur der Erhebungsteil zu Gemeinschaftsunterkünften betrachtet. Die Befragung an Wohnheimen ist im Teilprojekt Personenerhebung eingeschlossen.

Die Erhebung an Sonderbereichen definiert Gemeinschaftsunterkünfte als Einrichtungen, deren Bewohnerinnen und Bewohner keinen eigenen Haushalt führen, sondern durch die Betreibenden versorgt und/oder betreut werden. Dies sind zum Beispiel Heime für pflegebedürftige Menschen oder Justizvollzugsanstalten.

5 Zum Teilprojekt Personenerhebung werden hier die Haushaltbefragung auf Stichprobenbasis und die Befragung an Wohnheimen betrachtet.

Übersicht 1

Eckdaten und veröffentlichte Kern-Qualitätskennzahlen der fünf Teilprojekte des Zensus 2022

	Melderegister	Gebäude- und Wohnungszählung	Erhebung an Sonderbereichen	Personenerhebung	Haushaltgenerierung
Zweck	Datenübermittlung der Meldebehörden zur Vorbereitung und Durchführung des Zensus	Zählung der Gebäude mit Wohnraum, bewohnten Unterküften und Wohnungen zum Zensusstichtag	Feststellung der an Adressen mit Sonderbereichen wohnenden Personen	Feststellung der an Adressen mit Wohnraum wohnhaften Personen zur Ermittlung der Einwohnerzahl sowie Erhebung weiterer soziodemografischer Merkmale (Bildung und Erwerbstätigkeit)	Ermittlung von Zensusergebnissen zu Haushalten und Familien unter Nutzung von Daten aus Melderegistern, Gebäude- und Wohnungszählung sowie Haushaltebefragung
Primär- oder Sekundärerhebung	Sekundärerhebung	Primärerhebung	Primärerhebung	Primärerhebung	keine Erhebung
Vollerhebung oder Stichprobe	Vollerhebung	Vollerhebung	Vollerhebung	Haushaltebefragung auf Stichprobenbasis, Vollerhebung an Wohnheimen	keine Erhebung
Gesetzliche Auskunftspflicht (§ 23 Zensusgesetz 2022)	Lieferverpflichtung (§ 5 Zensusgesetz 2022)	ja (§ 24 Zensusgesetz 2022)	ja (§ 26 Zensusgesetz 2022)	ja (§§ 25, 26 Zensusgesetz 2022)	nicht relevant, da keine Erhebung
Einheiten	Meldepflichtige Personen, Gemeinden	Gebäude, Wohnungen	Gemeinschaftsunterkünfte	Personen	Haushalte, Wohnungen
Veröffentlichte Kern-Qualitätskennzahlen (Quoten)	Antwortausfälle bei Einheiten, Abgearbeitete Mussfehler ¹ , Zurückgewiesene Pakete beim Datenempfang ¹	Rückläufe, Onlinemeldungen, Antwortausfälle bei Einheiten, Imputationen bei Einheiten, Imputationen bei Merkmalen	Rückläufe	Onlinemeldungen, Festgestellte Existenzen durch Erhebungsbeauftragte, Antwortausfälle bei Einheiten	Verknüpfung Haushalt mit Wohnung
Zielwerte zu Kern-Qualitätskennzahlen	0% Antwortausfälle bei Einheiten, 100% abgearbeitete Mussfehler ¹	85 bis 95% Rückläufe, 60% Onlinemeldungen	95% Rückläufe	50% Onlinemeldungen	100% Verknüpfung Haushalt mit Wohnung

¹ Diese Kern-Qualitätskennzahl wird in diesem Artikel nicht behandelt, sie kann im Qualitätsbericht Zensus (Statistisches Bundesamt, 2024) nachgeschlagen werden.

Quoten etwa zu Rückläufen, Onlinemeldungen und Antwortausfällen in mehreren Teilprojekten relevant sein. [↪ Übersicht 1](#) stellt die Eckdaten und veröffentlichten Kern-Qualitätskennzahlen der Teilprojekte dar.

Zudem vereinbarte der Statistische Verbund für geeignete Kennzahlen vor Beginn der Arbeitsphase zu erreichende Zielwerte (zum Beispiel 95%-Rücklaufquote in der Erhebung an Sonderbereichen). Zielwerte steuern den Erhebungsprozess, beispielsweise ermöglichen sie mit einem Hochlaufplan, prozessbegleitend qualitätssichernde Maßnahmen zu veranlassen.

3

Kern-Qualitätskennzahlen zur Datengewinnung

Dieses Kapitel stellt die Ergebnisse der veröffentlichten Kern-Qualitätskennzahlen zur Datengewinnungsphase der Gebäude- und Wohnungszählung, der Erhebung an Sonderbereichen sowie der Personenerhebung vor und ordnet sie übergreifend ein. Betrachtet werden die Quoten für Rückläufe und für Onlinemeldungen sowie die Quote zur Existenzfeststellung durch Erhebungsbeauftragte.

3.1 Rückläufe

Die Rückläufe in den primärstatistischen Zensuserhebungen bilden die beantworteten Fragebogen von Wohnungs- und Gebäudeeigentümerinnen und -eigentümern, von Betreibenden der Gemeinschaftsunterkünfte oder von Personen, die in Haushalten oder Wohnheimen leben. Für die Rückmeldungen der Auskunftspflichtigen stehen alternative Meldewege – zum Beispiel Online-Selbstausfüllerformular oder Papierfragebogen – zur Verfügung (Dittrich und andere, 2022).

Ein wichtiges Bewertungskriterium für Rücklaufquoten in statistischen Erhebungen ist die Auskunftspflicht. Im Vergleich zu freiwilligen Befragungen waren aufgrund der geltenden gesetzlichen Auskunftspflicht in den Zensuserhebungen (§ 23 Zensusgesetz 2022) hohe Rücklaufquoten zu erwarten.⁶ Dementsprechend wurden vor

⁶ Im Mikrozensus wirkte sich die Einführung der Auskunftspflicht für zuvor freiwillige Fragen positiv auf das Antwortverhalten aus (Hochgürtel/Rammelt, 2018).

Übersicht 2

Rücklaufquoten¹ der Gebäude- und Wohnungszählung und der Erhebung an Sonderbereichen im Zensus 2022

	Gebäude- und Wohnungszählung	Erhebung an Sonderbereichen
Berechnung der Rücklaufquote	Anteil der eingegangenen Fragebogen von Einzel- und Mehrfacheigentümerinnen/-eigentümern an allen versendeten und zustellbaren Fragebogen	Anteil der Gemeinschaftsunterkünfte, für die eine Rückmeldung verzeichnet werden konnte, an allen Gemeinschaftsunterkünften
Zielwert(korridor) für Rücklaufquote	85 bis 95 %	95 %
Erreichter Bundeswert	95 %	95 %
Erreichte Länderwerte	90 bis 99 %	84 bis 99 %

¹ Mehrfacheigentümerinnen und -eigentümer erhielten entsprechend ihrer Eigentumsverhältnisse mehrere Fragebogen. Die Quote enthält keine Rückläufe von Großeigentümerinnen und -eigentümern und wurde bereinigt um Mehrfachrückläufe (zum Beispiel Doppelmeldungen).

Beginn der Erhebungsphase in den beiden hier betrachteten Vollerhebungen hohe Zielwerte festgesetzt.

Der festgelegte Zielwertkorridor von 85 bis 95 % in der Gebäude- und Wohnungszählung wurde sowohl auf Bundesebene mit 95 % als auch auf Bundesländerebene mit 90 bis 99 % erreicht. Die erzielten Ergebnisse übersteigen die Mindesterwartung von 85 % deutlich, sodass sich ein möglicher Zielwertkorridor für künftige Befragungen durchaus bei 90 bis 99 % ansetzen ließe.

↳ Übersicht 2

In der Erhebung an Sonderbereichen wurde der Zielwert von 95 % im Bund erfüllt. Die Länderwerte bewegen sich hier zwischen 84 und 99 %. Verschiedene Gründe sind dafür maßgeblich, dass einzelne Länder den Zielwert nicht erreicht haben, beispielsweise fehlender Rücklauf von nicht mehr existierenden Gemeinschaftsunterkünften. Die Gegenüberstellung der beiden Teilprojekte legt aus Sicht des Qualitätsmanagements den Schluss nahe, dass bei dezentralen Statistiken ein Zielwertkorridor beziehungsweise Toleranzbereich eine sinnvolle Alternative zum Einzelzielwert ist. Dies ermöglicht es, die zu erwartenden Schwankungen der Länderergebnisse in den Zielgrößen zu berücksichtigen.

Im Vergleich zu den auskunftspflichtigen Erhebungen werden in freiwilligen Befragungen deutlich geringere Rücklaufquoten erreicht. Zum Beispiel verzeichnet Deutschland lediglich 26,7 % Rückläufe in einer länderübergreifenden Befragung mit Rücklaufquoten zwischen 26,7 und 54,2 % in den beteiligten europäischen Ländern (European Social Survey, 2024). Insbesondere für freiwillige Befragungen mit repräsentativen Stichproben spielen Rücklaufquoten eine maßgebliche Rolle. Maßnahmen, die hierbei positiv auf Rückläufe wirken kön-

nen, sind beispielsweise Mixed-Mode-Befragungen⁷ und Belohnungen⁸.

Zusammenfassend ist für die erzielten hohen Rücklaufquoten in den Zensususerhebungen festzuhalten, dass die Eigentümerinnen und Eigentümer von Gebäuden und Wohnungen sowie die Betreiberinnen und Betreiber von Gemeinschaftsunterkünften mit ihren Rückmeldungen zu einer umfangreichen Datenbasis im Zensus beigetragen haben. Diese Kennzahl erlaubt jedoch noch keine Aussage darüber, wie hoch die Datenqualität ist. Dazu sind die eingegangenen Daten noch zu überprüfen, wie verwertbar sie sind (siehe Abschnitt 4.1 Antwortausfälle).

3.2 Onlinemeldungen

In den Primärerhebungen des Zensus umfassen Onlinemeldungen die beantworteten Fragebogen, die über das Online-Selbstaussfüllerformular eingegangen sind. Neben der gesetzlichen Verpflichtung⁹ wird mit diesem Meldeweg auch die Online-First-Strategie verfolgt, die gemäß Grundsatz 9 des Verhaltenskodex für europäische Statistiken (Europäische Union, 2018) auf eine möglichst belastungsarme Auskunftserteilung der Befragten abzielt. So können diese das Online-Selbstaussfüllerformular über verschiedene Endgeräte, wie Smartphone oder Tablet, beantworten (Gaedke und andere, 2024).

- 7 In Mixed-Mode-Befragungen können die Befragten zwischen verschiedenen Erhebungsmodi wählen, zum Beispiel Face-to-Face-Interview oder Online-Selbstaussfüllerfragebogen. Dies trägt zu höheren Rückläufen bei (Stadtmüller und andere, 2021).
- 8 Belohnungen können zum Beispiel Geldzahlungen für vollendete Interviews sein, wie in Deutschland im European Social Survey (2024) eingesetzt.
- 9 Die Auskunftserteilung erfolgt grundsätzlich elektronisch (§ 23 Absatz 1 Satz 2 Zensusgesetz 2022).

Aufgrund der bereitgestellten alternativen Meldewege war zu erwarten, dass nur ein Teil der Auskünfte als Onlinemeldung eingehen würde. Das Ergebnis der Pilotierung zur Haushaltebefragung mit 57% Meldungen über das Online-Selbstaussfüllerformular¹⁰ ließ auf eine ähnliche Quote in der Haupterhebung schließen (Estatico und andere, 2021). Die Zielwerte für die Onlinemeldequoten wurden auf 50% für die Personenerhebung und 60% für die Gebäude- und Wohnungszählung vor Beginn der Erhebungsphase festgelegt.

In beiden Primärerhebungen wurden die Zielwerte zu den Onlinemeldequoten erfüllt. In der Personenerhebung wurde ein Bundesergebnis von 73% und in den Bundesländern Onlinemeldequoten von 57 bis 88% erzielt (↪ Übersicht 3). Die länderindividuelle Durchführung des Erhebungsgeschäfts beeinflusst maßgeblich, wie hoch die jeweiligen Quoten in den Bundesländern ausfallen. Zum Beispiel führte ein verstärkter Einsatz interviewgestützter Befragungen zu geringeren Onlinemeldequoten.¹¹ Im Vergleich zum Zensus 2011 ist der Anteil der Onlinemeldungen deutlich gestiegen. Damals nutzten nur 7% der Auskunftspflichtigen in der Haushaltsstichprobe den Online-Fragebogen (Statistisches Bundesamt, 2011).

In der Gebäude- und Wohnungszählung erreichten die Onlinemeldequoten 82% auf Bundesebene und zwischen 74 und 91% in den Ländern (siehe Übersicht 3).

- 10 Bezogen auf die Ziel-2-Befragung, mit welcher weitere Zensusmerkmale wie Bildung und Erwerbstätigkeit erhoben wurden.
- 11 Die Onlinemeldequote ist auf die Ziel-2-Befragung beschränkt, mit welcher weitere Merkmale zu Bildung und Erwerbstätigkeit erfasst wurden. Die Auswertung umfasst nur zwölf Bundesländer. Für vier Bundesländer wurde auf eine Auswertung verzichtet, da dort die Interviewerinnen und Interviewer vorrangig ein Tablet mit der Zensus-App verwendeten. Somit wurde in diesen vier Ländern ein zusätzlicher Befragungsmodus genutzt, der das Online-Selbstaussfüllerformular zum Großteil ersetzte.

Übersicht 3

Onlinemeldequoten der Gebäude- und Wohnungszählung und der Personenerhebung

	Gebäude- und Wohnungszählung	Personenerhebung
Berechnung der Onlinemeldequote	Anteil der über das Online-Selbstaussfüllerformular zurückgelaufenen Meldungen an allen zurückgelaufenen Meldungen der Einzel- und Mehrfacheigentümerinnen/-eigentümer	Anteil der Personen, die über das Online-Selbstaussfüllerformular die Befragung beantwortet haben, an allen Personen, die die Befragung beantwortet haben
Zielwert für Onlinemeldequote	60%	50%
Erreichter Bundeswert	82%	73%
Erreichte Länderwerte	74 bis 91%	57 bis 88%

Hier wirkte sich beispielsweise die länderindividuelle Maßnahme, im Erinnerungsverfahren verstärkt Papierfragebogen beizulegen, mindernd auf die Onlinemeldequoten aus. Im Zensus 2011 lag die Onlinequote bei nur etwa einem Drittel (Statistisches Bundesamt, 2011).

In den letzten beiden Erhebungen des Mikrozensus nutzten die Befragten den Online-Fragebogen mit Anteilen von 50,7 beziehungsweise 50,4% (Statistisches Bundesamt, 2022; Statistisches Bundesamt, 2023).

Insgesamt ist festzuhalten, dass die erreichten hohen Onlinemeldequoten der Gebäude- und Wohnungszählung und der Personenerhebung einen positiven Beitrag zur Online-First-Strategie geleistet haben. Andere Meldewege wie Papierfragebogen wurden deutlich weniger genutzt – auf Bundesebene etwa bei jeder fünften Rückmeldung in der Gebäude- und Wohnungszählung und bei jeder vierten in der Personenerhebung.

3.3 Festgestellte Existenzen

Die Quote der festgestellten Existenzen durch Erhebungsbeauftragte ist eine einzigartige, teilprojektspezifische Kern-Qualitätskennzahl der Personenerhebung. Die Existenz einer Person gilt als festgestellt, wenn ihr Name, Vorname, Geburtsdatum und Geschlecht vollständig erhoben wurden. Diese Personen sind zählungsrelevant für die Ermittlung der Einwohnerzahlen.¹² Um die Situation vor Ort in Augenschein zu nehmen und damit eine hohe Datenqualität sicherzustellen, erfolgten persönliche Befragungen durch geschulte Interviewerinnen und Interviewer (sogenannte Erhebungsbeauftragte) (Dreschmitt/Pfahl, 2024).

Die Anzahl der festgestellten Existenzen wird in der primärstatistischen Personenerhebung ermittelt und ist vorab nicht bekannt.¹³ Ebenso kann nicht prognostiziert werden, wie viele Existenzen in persönlichen Interviews von den Erhebungsbeauftragten tatsächlich festgestellt werden können. Dementsprechend wurde für diese Quote kein Zielwert festgelegt. Erwartet wurde ein recht hoher Anteil der festgestellten Existenzen durch Erhebungsbeauftragte. Andere Maßnahmen beziehungsweise Erhebungsmodi – zum Beispiel schriftliche Befragungen bei Nichterreichbarkeit der Auskunftspflichtigen – würden den geringeren Anteil ausmachen.

Im Ergebnis fiel die Quote der festgestellten Existenzen durch Erhebungsbeauftragte mit einem Bundeswert von 97% und Werten zwischen 94 und 98% in den Bundesländern sehr hoch aus (↗ Übersicht 4). Dies spricht für diesen Erhebungsmodus, denn über andere Erhebungsmodi wie schriftliche Befragungen wurde nur ein sehr geringer Teil der Existenzen festgestellt – lediglich 2 bis 6%. Somit wurde – gegenüber dem Zensus 2011 mit 79% Vor-Ort-Befragungen¹⁴ – im Zensus 2022 die Quote der persönlichen Interviews deutlich erhöht (Statistisches Bundesamt, 2011).

12 Die Anzahl der festgestellten Existenzen ist nicht gleichzusetzen mit der Einwohnerzahl, die über verschiedene Datenaufbereitungsschritte ermittelt wird (zum Beispiel Hochrechnungsverfahren; siehe auch Bretschki und andere, 2024).

13 Aufgrund des unbekanntes Nenners ist keine Rücklaufquote berechenbar.

14 Bezieht sich auf die Haushaltsstichprobe und enthält auch Ziel-2-Befragungen.

Übersicht 4

Quote der festgestellten Existenzen durch Erhebungsbeauftragte in der Personenerhebung im Zensus 2022

	Personenerhebung
Berechnung der Quote der festgestellten Existenzen durch Erhebungsbeauftragte	Anteil der Personen, deren Existenz durch Erhebungsbeauftragte festgestellt wurde, an der Gesamtzahl der als existent festgestellten Personen
Zielwert	kein Zielwert
Erreichter Bundeswert	97%
Erreichte Länderwerte	94 bis 98%

4

Kern-Qualitätskennzahlen zur Datenaufbereitung

Dieses Kapitel präsentiert die Ergebnisse ausgewählter Kern-Qualitätskennzahlen¹⁵ zur Datenaufbereitungsphase für die Teilprojekte Gebäude- und Wohnungszählung, Personenerhebung, Melderegister und Haushalgenerierung und nimmt eine übergreifende Einordnung vor. Betrachtet werden die Quoten zu Antwortausfällen und Imputationen sowie zu Verknüpfungen von Haushalten mit Wohnungen.

4.1 Antwortausfälle

Liegen keine oder keine verwertbaren Daten vor, spricht man von Antwortausfällen bei Einheiten (Unit Nonresponse) oder Antwortausfällen bei Merkmalen (Item Nonresponse). Diese Kern-Qualitätskennzahlen werden zum Abschluss der Arbeiten zur Datenaufbereitung ermittelt. Beispielsweise können sich bei den Datenprüfungen (zum Beispiel auf Vollständigkeit und Schlüssigkeit der beantworteten Fragebogen) Antwortausfälle ergeben. Es ist jedoch auch möglich, dass mittels Korrekturverfahren die nicht verwertbaren Daten einzelner Merkmale plau-

sibilisiert und somit Antwortausfälle reduziert werden (Bretnütz und andere, 2024).¹⁶

Die folgende Betrachtung untersucht nur die Antwortausfälle bei Einheiten (Unit Nonresponse). So beziehen sich Antwortausfälle in der primärstatistischen Personenerhebung auf die Einheit „Personen“ und im sekundärstatistischen Teilprojekt Melderegister auf die Einheit „Gemeinden“ (siehe Übersicht 1). Die Antwortausfälle lassen sich bei Personen unter anderem auf ihr individuelles Antwortverhalten beziehungsweise ihre mangelnde Antwortbereitschaft zurückführen.¹⁷ Ähnliches gilt für das Antwortverhalten von Eigentümerinnen und Eigentümern in der Gebäude- und Wohnungszählung, die den beziehungsweise die versendeten Fragebogen zu Gebäuden nicht beantwortet haben (hier: Antwortausfall auf Gebäudeebene, siehe hierzu Bretnütz und andere, 2024). Im Teilprojekt Melderegister hingegen ist es ausgeschlossen, dass Gemeinden als staatliche Verwaltungseinheiten wegen mangelnder Bereitschaft nicht antworten, denn ihrer gesetzlichen Lieferverpflichtung für den Zensus müssen Meldebehörden nachkommen (§ 5 Zensusgesetz 2022). Als Grund für Antwortausfälle wären beispielsweise technisch verursachte Lieferverzögerungen denkbar. Dementsprechend war es nur im Teilprojekt Melderegister sinnvoll, einen Zielwert von 0% Antwortausfällen festzusetzen (↗ Übersicht 5). In den

15 Alle veröffentlichten Kern-Qualitätskennzahlen mit Erläuterungen sind in den Qualitätsberichten des Zensus 2022 und der Gebäude- und Wohnungszählung zu finden (Statistisches Bundesamt, 2024).

16 Antwortausfälle und Rückläufe summieren sich somit nicht unbedingt auf 100%.

17 Neben der gesetzlichen Auskunftspflicht tragen auch die eingerichteten Erinnerungs- und Mahnverfahren dazu bei, Antwortausfälle bei Einheiten zu reduzieren.

Übersicht 5

Antwortausfallquoten bei Einheiten in Gebäude- und Wohnungszählung, Personenerhebung und im Melderegister im Zensus 2022

	Gebäude- und Wohnungszählung	Personenerhebung	Melderegister
Berechnung der Antwortausfallquote bei Einheiten	Anteil der versendeten Fragebogen, für die keine oder keine verwertbaren Daten gesammelt wurden, an allen versendeten Fragebogen, für die Daten gewonnen werden sollten	Anteil der Personen aus der Ziel-2-Stichprobe, für die keine Angaben zur Ziel-2-Befragung vorliegen, an allen Personen, für die Angaben vorliegen sollten	Anteil der Gemeinden, die keine Melderegisterdaten geliefert haben, an allen heranzuziehenden Gemeinden
Zielwert für Antwortausfallquote bei Einheiten	kein Zielwert	kein Zielwert	0%
Erreichter Bundeswert	5%	10%	0%
Erreichte Länderwerte	1 bis 13%	4 bis 17%	keine Länderwerte ausgewiesen

Primärerhebungen werden so wenig Antwortausfälle wie möglich angestrebt.

Im sekundärstatistischen Erhebungsteil Melderegister wurde der Zielwert von 0% Antwortausfällen erreicht, denn alle herangezogenen Gemeinden haben ihre Melderegisterdaten geliefert. Ein Vergleichswert aus dem Zensus 2011 ist nicht verfügbar. In der Gebäude- und Wohnungszählung sind auf Bundesebene für 5% (Länderwerte 1 bis 13%) der Fragebogen keine Antworten vorhanden. Dieser Wert gleicht dem im Zensus 2011 mit 5,3% (Statistisches Bundesamt, 2011). In der Personenerhebung liegen für 10% (Länderwerte 4 bis 17%) der Personen in der Ziel-2-Befragung¹⁸ keine Angaben vor. Ähnlich hoch war die Antwortausfallquote der letzten beiden Haushaltebefragungen des Mikrozensus mit Bundesergebnissen von 10,6 und 11,6% (Statistisches Bundesamt, 2022; Statistisches Bundesamt 2023). Aus dem Zensus 2011 steht kein Vergleichswert für die Personenerhebung zur Verfügung.

Die Antwortausfallquoten in freiwilligen Befragungen fallen deutlich höher aus. In der bereits erwähnten länderübergreifenden Befragung in Europa liegen diese zwischen 47,6 und 74,4% der befragten Personen und in Deutschland bei 73,7% (European Social Survey, 2024).

Zusammenfassend sprechen die geringen Antwortausfallquoten der beiden Primärerhebungen im Zensus 2022 für eine gute Datenqualität, denn der Großteil der zu erhebenden Daten liegt vor und ist verwertbar. Die

18 In der Ziel-2-Befragung werden weitere Merkmale wie Bildung und Erwerbstätigkeit erhoben.

Antwortausfallquote bei Einheiten des sekundärstatistischen Erhebungsteils Melderegister ist dabei anders zu interpretieren, denn sie stellt streng genommen eine Rücklaufquote von 100% dar. Inwiefern die gelieferten Melderegisterdaten eine gute Datenqualität im Sinne verwertbarer Daten aufweisen, ergibt sich erst, nachdem sie überprüft und aufbereitet wurden.

4.2 Imputationen

Die Imputation ersetzt unzulässig fehlende beziehungsweise unplausible Angaben in den erhobenen Daten, um vollständige und plausible Daten zu erhalten (Bretnütz und andere, 2024).

➤ **Übersicht 6** zeigt die Imputationsquoten bei Einheiten und bei Merkmalen für die Gebäude- und Wohnungszählung. Für die Imputation von Einheiten (hier: Gebäude) wird ein automatisiertes Imputationsverfahren mit der Software CANCEIS¹⁹ verwendet. Die Imputationen von Merkmalen (zum Beispiel Wohnungsmiete) umfassen alle verwendeten Verfahren (zum Beispiel manuelle Imputationen; siehe hierzu Bretnütz und andere, 2024). Eine Zielwertsetzung für diese beiden Quoten ist nicht erfolgt.

Die Imputationsquote bei Einheiten liegt in der Gebäude- und Wohnungszählung bei 8% auf Bundesebene und bezogen auf die Länderebene zwischen 0 und 14%. Das

19 Die Software CANCEIS (Canadian Census Edit and Imputation System) ersetzt fehlende oder unplausible Angaben der Empfänger-Datensätze mit beobachteten Werten von Spender-Datensätzen aus derselben Erhebung.

Übersicht 6

Imputationsquoten bei Einheiten und bei Merkmalen in der Gebäude- und Wohnungszählung im Zensus 2022

	Imputation bei Einheiten	Imputation bei Merkmalen
Berechnung der Imputationsquote	Anteil der Datensätze, in denen alle Angaben automatisch durch imputierte Werte ersetzt wurden, an allen Datensätzen, für die Daten bei Erhebungsabschluss gewonnen werden sollten	Anteil der Datensätze, bei denen Angaben für eine bestimmte Frage (Merkmal) aufgrund von Antwortausfall ¹ durch imputierte Werte ersetzt wurden, an allen Datensätzen, für die Daten bei Erhebungsabschluss gewonnen werden sollten
Zielwert für Imputationsquote	kein Zielwert	kein Zielwert
Erreichter Bundeswert	8%	Wohnungsmiete: 15% Baujahr: 10% Energieträger: 11%
Erreichte Länderwerte	0 bis 14%	keine Länderwerte ausgewiesen

1 Antwortausfall umfasst fehlende und fehlerhafte Merkmalsangaben.

heißt auf Bundesebene: Bei 8% der Gebäude-Datensätze wurden alle Angaben automatisch imputiert und die ermittelten Antwortausfälle bei Gebäuden (Unit Non-response) mit verwertbaren Daten befüllt. Bei der merkmalsbezogenen Imputation werden einzelne Merkmale betrachtet. Die Imputationsquoten beispielsweise für die Merkmale Wohnungsmiete, Baujahr und Energieträger in der Gebäude- und Wohnungszählung 2022 liegen zwischen 10 und 15%. Somit wurden beispielsweise in 15% der Datensätze Angaben zum Merkmal Wohnungsmiete mit imputierten Werten ersetzt.

Als Anhaltspunkt werden hier die CANCEIS-Imputationsquoten für Merkmale auf Datensatzebene im Zensus 2021 für England und Wales herangezogen. Dort wurde bei 9% der Haushalts- und bei 26% der Personendatensätze mindestens ein Merkmal imputiert (Office for National Statistics, 2022). Auch wenn hier Durchschnittswerte auf Ebene der Einheiten betrachtet werden, liegen die Imputationsquoten bei Merkmalen der Gebäude- und Wohnungszählung in vergleichbarer – gegenüber den Personendatensätzen sogar etwas geringerer – Größenordnung.

Schlussfolgernd aus dem Vergleich können die merkmalsbezogenen Imputationsquoten der Gebäude- und Wohnungszählung als positiv bewertet werden.

4.3 Verknüpfungen Haushalt mit Wohnung

Die Verknüpfungsquote von Haushalten mit Wohnungen ist eine einzigartige, teilprojektspezifische Kern-Qualitätskennzahl der Haushaltegenerierung. Ein Haushalt gilt als verknüpft, wenn er einer Wohnung zugeordnet ist. Da es in den Registern, die als Datenquellen für den Zensus 2022 herangezogen wurden, keine Informationen über Haushaltezusammenhänge gibt (zum Beispiel Anzahl und Struktur von Haushalten), wurden diese im Zuge der Datenaufbereitung generiert. Im Rahmen der Haushaltegenerierung werden mittels automatisierter Verfahren Daten aus Melderegister, Haushaltsstichprobe, Befragung an Wohnheimen sowie Gebäude- und Wohnungszählung zu Haushalten zusammengeführt und mit konkreten Wohnungen verknüpft (Massih-Tehrani/Reichert, 2019). Im Zensus werden Wohnhaushalte gebildet.

Da jedem generierten Haushalt eine Wohnung zugeordnet werden sollte, wurde der Zielwert auf 100% für den Zensus 2022 festgelegt. Das erzielte Bundesergebnis von 100% belegt, dass jeder Haushalt mit einer bewohnten Wohnung verknüpft ist (↗ Übersicht 7). Somit wurde das gesetzte Ziel erreicht und die Verknüpfungsquote von 99,5% im Zensus 2011 übertroffen.

Übersicht 7

Verknüpfungsquote für Haushalte mit Wohnungen in der Haushaltegenerierung im Zensus 2022

	Haushaltegenerierung
Berechnung der Verknüpfungsquote	Anteil der Haushalte, die mit einer bewohnten Wohnung verknüpft sind, an der Gesamtzahl aller bewohnten Wohnungen
Zielwert für Verknüpfungsquote	100%
Erreichter Bundeswert	100%
Erreichte Länderwerte	keine Länderwerte ausgewiesen

Festzuhalten ist: Das Ziel der Haushaltegenerierung im Zensus 2022, dass alle zählungsrelevanten Haushalte mit einer bewohnten Wohnung verknüpft sind, konnte erreicht werden.

5

Fazit und Ausblick

Der Zensus 2022 hat mit dem Einsatz von Zielwerten für Kern-Qualitätskennzahlen in den Erhebungs- und Aufbereitungsprozessen sowie mit der Veröffentlichung der hiermit erreichten Qualitätsergebnisse Neuland in der deutschen amtlichen Statistik betreten. Aus den in diesem Artikel diskutierten Ergebnissen der Qualitätsmessung des Zensus 2022 lassen sich für das Qualitätsmanagement folgende Erkenntnisse und Ziele zusammenfassen:

- › Bei der statistikspezifischen Ausgestaltung der Qualitätsdatenblätter im Verbund sind neben statistikübergreifenden, vergleichbaren Qualitätsindikatoren (zum Beispiel Rücklaufquote) auch einzigartige, statistikspezifische Qualitätsindikatoren (zum Beispiel Verknüpfungsquote der Haushaltegenerierung) zu berücksichtigen.

- › Mit vorab festgelegten Zielwerten und Soll-Ist-Prüfungen lassen sich die Qualitätsmessungen und -bewertungen und damit auch die Erhebungs- und Aufbereitungsprozesse sinnvoll unterstützen.
- › Die Veröffentlichung von Qualitätsindikatoren im nationalen Qualitätsbericht wie im Zensus 2022 dient nicht nur der Dokumentation, sondern macht die Ergebnisse der Qualitätsmessungen auch für externe Nutzende transparent und zugänglich.

Aus (statistik-)übergreifenden Vergleichen von Qualitätsindikatoren und Erhebungsspezifika lassen sich zum einen gute Praxisbeispiele für die fachstatistischen Erhebungs- und Aufbereitungsprozesse ableiten. Zum Beispiel können – unabhängig von der Auskunftspflicht – alternative Erhebungsmodi für höhere Rücklaufquoten eingesetzt werden. Zum anderen kann das übergreifende Qualitätsmanagement wertvolle Erkenntnisse gewinnen, wie etwa Toleranzbereiche beziehungsweise Zielwertkorridore bei künftigen Erhebungen von dezentral durchgeführten Statistiken. Darüber hinaus sollte das Qualitätsmanagement die Recherche und komparative Betrachtung nationaler und internationaler Qualitätskennzahlen vorantreiben.

Ein Vergleich von einzigartigen, statistikspezifischen Qualitätsindikatoren ist vorrangig auf vorherige Berichtszeiträume beschränkt, sofern Zahlen verfügbar sind (hier: Zensus 2011). Innerhalb eines Berichtszeitraums

können vorab festgelegte Zielwerte als Vergleichswert beziehungsweise Maßstab für die Zielerreichung der Arbeiten dienen, zum Beispiel 100% für die Verknüpfungsquote der Haushaltegenerierung im Zensus 2022.

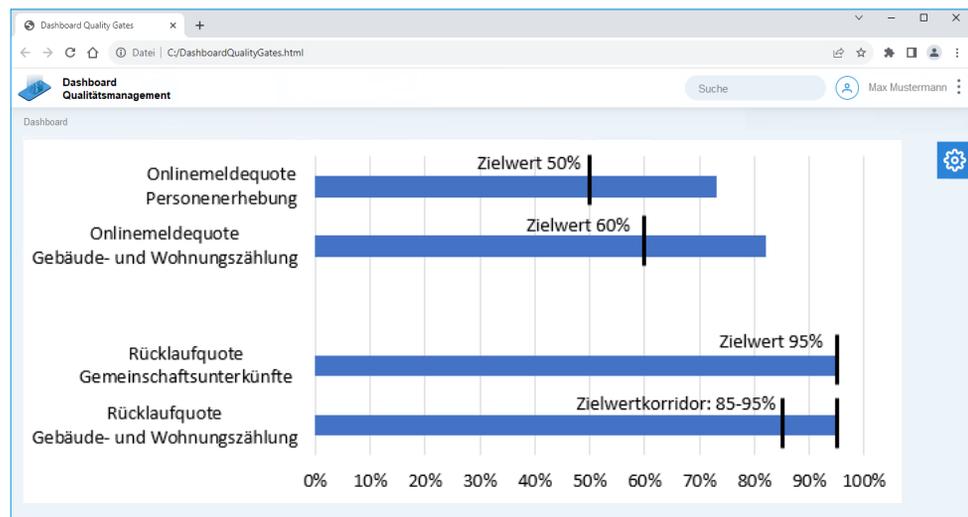
Von den Erfahrungen im Zensus 2022 können unter anderem Sekundärstatistiken und Rechenwerke profitieren und etwa aus den spezifischen Kern-Qualitätskennzahlen und Zielwerten des Melderegisters und der Haushaltegenerierung eigene Ausgestaltungen der Qualitätsdatenblätter erarbeiten und erproben.

Das Qualitätsmanagement sollte das Ziel verfolgen, ausgehend von der Veröffentlichung des Zensus 2022 adäquate Qualitätsstandards und -vorgaben für die nationale Qualitätsberichterstattung in der amtlichen Statistik zu erarbeiten. Beispielsweise könnten nach diesem Vorbild neben Antwortausfall- und Imputationsquoten zur Datenaufbereitung auch Quoten zur Datenerhebung wie etwa Rückläufe und Onlinemeldungen systematisiert in die nationale Qualitätsberichterstattung einfließen.²⁰ Außerdem könnte mit einem nutzungsfreundlichen, medienbruchfreien Zugang zu den veröffentlichten Qualitätsindikatoren der Fachstatistiken – etwa mit einem Dashboard wie in [Grafik 1](#) – ein

20 Seitens des Statistischen Amtes der Europäischen Union (Eurostat) sind unter anderem Quoten zu Antwortausfällen und Imputationen für die europäische Qualitätsberichterstattung vorgesehen, nicht jedoch zu Rückläufen und Onlinemeldungen (Eurostat, 2014).

Grafik 1

Mögliche Darstellung der Onlinemelde- und Rücklaufquoten mit Zielwerten (Bundesergebnisse) in einem Dashboard



zentraler Anlaufpunkt und eine bedarfsgerechte Darstellung der Vergleichsmöglichkeiten für die interessierten Nutzerinnen und Nutzer bereitgestellt werden.

Diese Maßnahmen im Rahmen der Qualitätsdatenblätter im Verbund tragen zur Erfüllung der Qualitätsgrundsätze Zugänglichkeit und Klarheit (Nr. 15) sowie Kohärenz und Vergleichbarkeit (Nr. 14) des Verhaltenskodex für europäische Statistiken bei. 

LITERATURVERZEICHNIS

Bretnütz, Hella/Estatico, Sabrina/Walther, Sabrina/Hüsgen-Brodhäger, Fabienne/Lange, Kerstin/Kleber, Birgit/Haberzettl, Benjamin. [Aufbereitung und Plausibilisierung der primärstatistischen Erhebungssteile im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 51 ff.

Bretschi, Corinna/Seibel, Steffen/Vorndran, Ingeborg/Pfahl, Miriam. [Ermittlung der Einwohnerzahl im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 17 ff.

Dittrich, Stefan/Bretschi, Corinna/Stepien, Halina Danuta/Vorndran, Ingeborg/Michel, Bernd/Kleber, Birgit/Timm, Ulrike/Pfahl, Miriam. [Der Zensus 2022 – Mit Online First an der Schwelle zu einem Registerzensus](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2022, Seite 90 ff.

Dreschmitt, Kai/Pfahl, Miriam. [Die Personenerhebung im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 29 ff.

Estatico, Sabrina/Tornow, Tobias/Walther, Sabrina. [Generalprobe für den Zensus 2022: Erkenntnisse aus der Pilotstudie zur Haushaltebefragung](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2021, Seite 126 ff.

Europäische Kommission. *ESS Guidelines for the Implementation of the ESS Quality and Performance Indicators (QPI)*. 2014. [Zugriff am 31. Oktober 2024]. Verfügbar unter: ec.europa.eu

Europäische Union. *Verhaltenskodex für europäische Statistiken*. Angenommen vom Ausschuss für das Europäische Statistische System am 16. November 2017. Luxemburg 2018. [Zugriff am 31. Oktober 2024]. Verfügbar unter: ec.europa.eu

European Social Survey European Research Infrastructure Consortium (ESS ERIC). *ESS round 11 - 2023. Social inequalities in health, Gender in contemporary Europe*. Sikt – Norwegian Agency for Shared Services in Education and Research. 2024. DOI: [10.21338/ess11-2023](https://doi.org/10.21338/ess11-2023)

European Social Survey. *ESS Round 11 Source Questionnaire*. 2022. London: ESS ERIC Headquarters c/o City, University of London. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.europeansocialsurvey.org

Gaedke, Annika/Pfahl, Miriam/Strohalm, Anna. [Die Online-First-Strategie und die Online-Fragebogen im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2024, Seite 92 ff.

Hochgürtel, Tim/Rammelt, Sabine. [Die auskunftspflichtige Erfassung von Lebensgemeinschaften im Mikrozensus ab 2017](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2018, Seite 47 ff.

LITERATURVERZEICHNIS

Massih-Tehrani, Nilgun/Reichert, Adrian. [Die Haushaltegenerierung im Zensus 2021](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 36 ff.

Meinke, Irina/Hentschke, Janine. [Kern-Qualitätskennzahlen im Zensus 2022](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2022, Seite 25 ff.

Office for National Statistics. *Item editing and imputation process for Census 2021, England and Wales*. 2022. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.ons.gov.uk

Stadtmüller, Sven/Beuthner, Christoph/Silber, Henning. *Mixed-Mode Surveys*. Mannheim 2021. GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines). DOI: [10.15465/gesis-sg_en_038](https://doi.org/10.15465/gesis-sg_en_038)

Statistisches Bundesamt. *Qualitätsbericht Gebäude- und Wohnungszählung 2011*. 2016. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Gebäude- und Wohnungszählung 2024*. [Im Erscheinen].

Statistisches Bundesamt. *Qualitätsbericht Mikrozensus 2022*. 2023. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Mikrozensus 2023*. 2024. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Zensus 2011*. 2016. [Zugriff am 31. Oktober 2024]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Zensus 2024*. [Im Erscheinen].

Tümmler, Thorsten/Meinke, Irina. [Aufbau des Qualitätsmanagements für den Zensus 2021](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 59 ff.

RECHTSGRUNDLAGEN

Gesetz zur Durchführung des Zensus im Jahr 2022 (Zensusgesetz 2022 – ZensG 2022) vom 26. November 2019 (BGBl. I Seite 1851), das durch Artikel 2 des Gesetzes vom 3. Dezember 2020 (BGBl. I Seite 2675) geändert worden ist.

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Dezember 2024
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-24006-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2024
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.