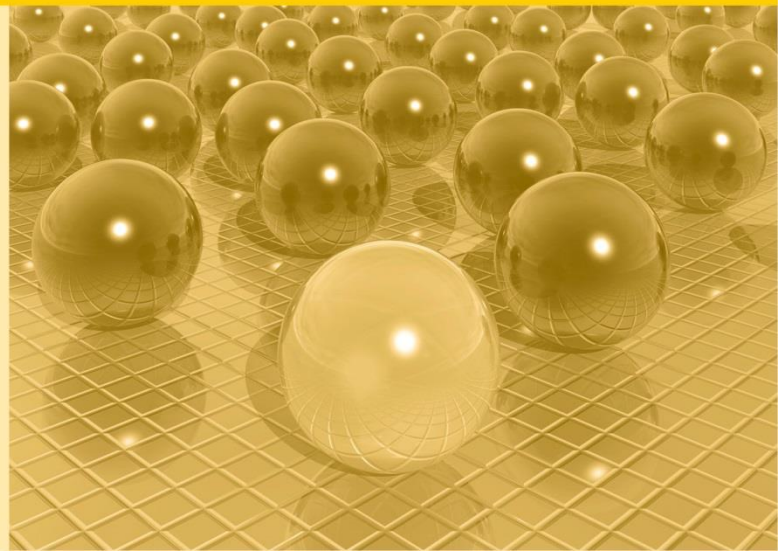


# Metadatenreport



Teil II: Produktspezifische Informationen zur Nutzung des Zensus 2022  
(Produkt 3 – Gesamtdatensatz) per On-Site-Nutzung

DOI: 10.21242/12111.2022.00.04.1.1.0

Version 1

## Impressum

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

### Fachliche Informationen zu dieser Veröffentlichung:

Forschungsdatenzentrum der  
Statistischen Ämter der Länder  
– Fürth –  
Tel.: 0911 98208-6857  
[fdz@statistik.bayern.de](mailto:fdz@statistik.bayern.de)

### Informationen zum Datenangebot:

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-2420  
Fax: 0611 75-3915  
[forschungsdatenzentrum@destatis.de](mailto:forschungsdatenzentrum@destatis.de)

Forschungsdatenzentrum der  
Statistischen Ämter der Länder  
– Geschäftsstelle –  
Tel.: 0211 9449-2883  
Fax: 0211 9449-8087  
[forschungsdatenzentrum@it.nrw.de](mailto:forschungsdatenzentrum@it.nrw.de)

Erscheinungsfolge: unregelmäßig  
Erschienen im Mai 2026

Diese Publikation wird kostenlos als PDF-Datei zum Download unter [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de) angeboten.

© Information und Technik Nordrhein-Westfalen, Düsseldorf, 2026  
(im Auftrag der Herausbergemeinschaft)

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

Fotorechte Umschlag: © artSILENCE – Fotolia.com

### Empfohlene Zitierung:

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder: Metadatenreport. Teil II: Produktspezifische Informationen zur Nutzung des Zensus 2022 am Gastwissenschafts Arbeitsplatz sowie per kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen). Version 1. DOI: 10.21242/12111.2022.00.04.1.1.0. Standort Fürth 2026.

# Metadatenreport

Teil II: Produktspezifische Informationen zur Nutzung des Zensus 2022  
(Produkt 3 – Gesamtdatensatz) am Gastwissenschaftsarbeitsplatz sowie per  
kontrollierter Datenfernverarbeitung (sog. On-Site-Zugangswege/-Nutzungen)

DOI: 10.21242/ 12111.2022.00.04.1.1.0

Version 1

## Inhalt

<b>1. Datenaufbereitung in den FDZ.....</b>	<b>2</b>
<b>1.1 Datenaufbereitung .....</b>	<b>2</b>
<b>1.2 Anonymisierungsmaßnahmen.....</b>	<b>3</b>
<b>1.3 Methodik der Verknüpfung.....</b>	<b>4</b>
<b>2 Produkt .....</b>	<b>5</b>
<b>2.1 Merkmale und Merkmalsbeschreibung .....</b>	<b>5</b>
<b>2.2 Vergleichbarkeit der Merkmale über die Zeit .....</b>	<b>7</b>
<b>2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen .....</b>	<b>8</b>
<b>2.4 Auswertbare regionale Ebene.....</b>	<b>8</b>
<b>3 Praktische Hinweise .....</b>	<b>9</b>
<b>3.1 Hinweise zur Geheimhaltung .....</b>	<b>9</b>
<b>3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung .....</b>	<b>9</b>
<b>3.1.2 Geheimhaltung von Ergebnissen .....</b>	<b>10</b>
<b>3.1.3 Cell-Key-Methode.....</b>	<b>10</b>
<b>3.2 FAQ .....</b>	<b>11</b>
<b>3.3 Verfügbare Tools .....</b>	<b>12</b>
<b>Abkürzungsverzeichnis.....</b>	<b>12</b>

# 1. Datenaufbereitung in den FDZ

*Hinweis:*

*Dieser Metadatenreport beschreibt Produkt 3 des Zensus 2022 („Gesamtdatensatz“). Daneben stellen die Forschungsdatenzentren (FDZ) die Produkte 1 („Haushaltsstichprobe“) und 2 („Gebäude- und Wohnungszählung“) bereit. Informationen zu den weiteren Produkten finden Sie auf der [Website der FDZ](#).*

## 1.1 Datenaufbereitung

Die Grundlage für die Erstellung von Produkt 3 bildete das Datenmaterial des Zensus 2022 nach der Zusammenführung der verschiedenen primärstatistischen und registergestützten Bestandteile im Rahmen der Haushaltgenerierung<sup>1</sup> und weiterer Aufbereitungsschritte. Die Originaldaten liegen als relationales Datenmodell in Form von verschiedenen Teiltabellen vor, die je nach Auswertungszweck über Identifikatoren miteinander verknüpft werden können. Aus der Teiltabelle, die die Daten der Haushaltsstichprobe enthält (PERSON\_STP), wurde das standardisierte Produkt 1 „Haushaltsstichprobe“ aufbereitet. Aus den beiden Teiltabellen, die die Gebäude- und Wohnungsdaten enthalten (GEBAEUDE, WOHNUNG), wurde das standardisierte Produkt 2 „Gebäude- und Wohnungszählung“ erstellt. Produkt 3 „Gesamtdatensatz“ beinhaltet ebenfalls die Daten der Haushaltsstichprobe und der Gebäude- und Wohnungszählung, enthält darüber hinaus jedoch weitere Angaben zu den regionalen Gliederungsebenen sowie den Bevölkerungs-, Haushalts-, und Familienstrukturen in Deutschland. Es ermöglicht somit Auswertungen aus dem Gesamtdatenbestand des Zensus 2022. Aufgrund seiner Größe und Komplexität kann Produkt 3 – anders als die Produkte 1 und 2 – nicht standardisiert zur Verfügung gestellt werden. Stattdessen wird für jedes Forschungsvorhaben eine individuelle Merkmalsauswahl aus den verfügbaren Teiltabellen getroffen und ein projektspezifischer Datensatz erstellt.

Im Zuge der Datenaufbereitung erfolgten zunächst Qualitätssicherungsmaßnahmen, um sicherzustellen, dass das Datenmaterial korrekt und vollständig an die FDZ übermittelt wurde. Für ausgewählte Merkmale und Merkmalskombinationen wurden Abgleiche mit den fachseitig veröffentlichten Zensusergebnissen durchgeführt. Um das Originalmaterial zur wissenschaftlichen Nutzung aufzubereiten, wurde es im Umfang deutlich reduziert. Bei mehrfach vorliegenden, nahezu

---

<sup>1</sup> Für genauere Informationen zur Methode des Zensus 2022 und der Haushaltgenerierung siehe Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, Metadatenreport, Teil I <https://www.forschungsdatenzentrum.de/de/haushalte/zensus-2022>

identischen Merkmalen (z.B. unterschiedlich klassierten Altersangaben) wurde möglichst nur eine Variante übernommen, wobei auf einen maximalen Informationsgehalt geachtet wurde<sup>2</sup>. Da sich die Nutzung der FDZ-Daten auf inhaltliche (und nicht methodische) Fragestellungen beschränkt, wurden Merkmale, die dahingehend nicht sinnvoll ausgewertet werden können, entfernt – dies betraf insbesondere Qualitätskennzeichen sowie Hilfs- und Prozessmerkmale. EU-Merkmale wurden beibehalten, um die internationale Vergleichbarkeit der Daten sicherzustellen. Weiterhin wurde die Anzahl an Beobachtungen reduziert, indem über den Aktivitätsstatus inaktive Personen, Haushalte, Wohnungen und Gebäude entfernt wurden.

Nebenwohnsitzpersonen sind bei hochgerechneten Auswertungen nicht zu berücksichtigen und haben dementsprechend keinen Hochrechnungsfaktor. In der Tabelle PERSON\_STP wurden Nebenwohnsitzpersonen daher entfernt. Auf Basis der Tabelle PERSON können Nebenwohnsitzpersonen hingegen ausgezählt werden, weshalb sie hier weiterhin enthalten sind. Für Auswertungen ohne Nebenwohnsitzpersonen kann über das Merkmal WOHNSTATUS ein entsprechender Filter gesetzt werden.

Schließlich wurden Maßnahmen ergriffen um die Nutzungsfreundlichkeit des Datenmaterials zu erhöhen. Dies beinhaltete die Zuweisung von einheitlichen und übersichtlichen Variablen- und Wertelabels, z.B. die Angleichung an Merkmalsnamen des Zensus 2011 oder das Hinzufügen der Vorsilbe „EU\_“ für EU-Merkmale. Teilweise erfolgte eine aussagekräftige Rekodierung von Merkmalsausprägungen. Im gelieferten Datenmaterial waren lediglich Gitterzellen der Seitenlängen 100 m und 1 km enthalten. Da in fachseitigen Veröffentlichungen und in den FDZ-Daten des Zensus 2011 auch Gitterzellen der Seitenlänge 10 km enthalten sind, wurden diese im Zuge der Datenaufbereitung auch für den Zensus 2022 hinzugefügt.

Da die Verwendung von Wertelabels in Stata auf numerische Variablen beschränkt ist, erfolgte für alphanumerische Variablen – sofern sinnvoll – eine numerische Kodierung mit anschließender Vergabe von Wertelabels. Die finalen Teiltabellen liegen als gelabelte Stata-Datensätze vor, die sich z.B. über das Paket „haven“ auch mit Labels in R einlesen lassen.

## 1.2 Anonymisierungsmaßnahmen

Produkt 3 ist als formal anonym anzusehen, da die Daten keine direkten Identifikatoren und Hilfsmerkmale enthalten. Identifikationsnummern sind bereits in der Datenlieferung an die FDZ systemfrei umgesetzt. Der am Gastwissenschaftsarbeitsplatz (GWAP) bereitgestellte Datensatz

---

<sup>2</sup> Ausnahme sind Merkmale bei denen davon auszugehen ist, dass weniger detaillierte Merkmale häufiger verwendet werden, z.B. Berufsklassifikationen in der Teiltabelle PERSON\_STP. Diese Merkmale liegen zusätzlich auch gruppiert vor.

entspricht in seinem Aufbau dem Material für die Kontrollierte Datenfernverarbeitung (KDFV). Die tiefste regionale Gliederungsebene am GWAP ist jedoch die Gitterzelle der Seitenlänge 10 km. Via KDFV ist auch die Auswertung von 100-m-Gitterzellen möglich. Um zudem kürzere Programmlaufzeiten zu gewährleisten, wird für die Nutzung am GWAP eine Substichprobe der Teiltabellen auf Anschriftenebene erzeugt. Die GWAP-Datensätze von Produkt 3 dienen daher lediglich der Erstellung von Auswertungsprogrammen. Die finale Ergebniserstellung wird stets via KDFV am Vollmaterial vorgenommen.

### **1.3 Methodik der Verknüpfung**

Das Datenmaterial des Zensus 2022 liegt als relationales Datenmodell in Form von verschiedenen Teiltabellen vor. Im Rahmen von Produkt 3 werden folgende Teiltabellen zur Verfügung gestellt: ANSCHRIFT, FAMILIE, GEBAEUDE, HAUSHALT, KANTE, PERSON, PERSON\_STP, REGIONALINFORMATION und WOHNUNG. Die Teiltabellen können je nach Auswertungsinteresse über folgende Identifikatoren miteinander verknüpft werden: ANSCHRIFT\_ID, FAMILIE\_ID, GEBAEUDE\_ID, HAUSHALT\_ID, PERSON\_ID und WOHNUNG\_ID. Jede statistische Einheit (Familie, Haushalt, Person, Gebäude und Wohnung) ist mit genau einer Anschrift verknüpft, zu der wiederum über den 12-stelligen amtlichen Regionalschlüssel (AGS\_12) Regionalinformationen vorliegen. Die Verknüpfung erfolgt i.d.R. im Rahmen der projektspezifischen Aufbereitung durch den fachlich zuständigen FDZ-Standort in Bayern. In bestimmten Fällen kann es sinnvoller sein anstelle eines „flachen“ Datensatzes mehrere Teiltabellen bereitzustellen, die erst während der Datennutzung zur Ergebniserstellung miteinander verknüpft werden. Dies gilt beispielsweise für die Darstellung von Personenbeziehungen über die Teiltabelle KANTE.

Verknüpfungen mit externen Daten sind gemäß §13a BStatG erlaubt. Zum Schutz der Daten müssen diese im Vorfeld der Datennutzung explizit, möglichst bereits im Rahmen der Antragstellung, mit den FDZ abgestimmt werden. Der Zensus 2022 kann nicht auf Individualebene mit dem Zensus 2011 verknüpft werden. Zeitliche Vergleiche zwischen den beiden Erhebungen können lediglich über die Auswertung der regionalen Einheiten vorgenommen werden.

## 2 Produkt

### 2.1 Merkmale und Merkmalsbeschreibung

Das Produkt 3 „Gesamtdatensatz“ des Zensus 2022 umfasst neun Teiltabellen, in denen folgende Informationen enthalten sind:

➤ Tabelle ANSCHRIFT

Diese Tabelle enthält im Wesentlichen alle anschriftenbezogenen Informationen. Mit Hilfe der Merkmale dieser Tabelle ist es möglich, die Zensusergebnisse auf einer administrativen Ebene wie Bundesland, Landkreis oder Gemeinde sowie unabhängig von Verwaltungsgrenzen auf Gitterzellenebene auszuwerten. Die kleinste regionale Einheit sind Gitterzellen der Seitenlänge 100 m, die zur Erzeugung eigener räumlicher Aggregate verwendet werden können.

➤ Tabelle FAMILIE

Diese Tabelle enthält Familieninformationen aus der Haushaltegenerierung und bildet Familienzusammenhänge ab. Dazu gehören Informationen wie Familiengröße oder -typ sowie EU-Merkmale. Eine Familie besteht aus einer Menge von Personen, wobei jede Familie mindestens aus zwei Personen besteht. Es werden nur Personen mit eigener Haushaltsführung betrachtet. Personen in Gemeinschaftsunterkünften sind dementsprechend nicht enthalten. Nach nationaler Typisierung kann es nur eine Familie je Haushalt bzw. Wohnung geben. Personen mit Nebenwohnsitz werden bei nationalen Auswertungen berücksichtigt. Nach EU-Typisierung kann es mehrere Familien je Haushalt bzw. Wohnung geben. Hier werden Nebenwohnsitzpersonen nicht betrachtet. Bei der Mischform National EU kann es mehrere Familien je Haushalt bzw. Wohnung geben und es werden sowohl Haupt- als auch Nebenwohnsitzpersonen berücksichtigt.

➤ Tabelle GEBAEUDE

Diese Tabelle enthält alle Gebäudedatensätze nach der Haushaltegenerierung. Jedes Gebäude befindet sich an genau einer Anschrift und beinhaltet mindestens eine Wohnung. An einer Anschrift können sich mehrere Gebäude befinden. Die Gebäudemerkmale beinhalten z.B. Informationen zu Gebäudeart, Baujahr, Gebäudegröße und -typ, Heizungsart, Energieträger und Eigentumsform.

➤ Tabelle HAUSHALT

Diese Tabelle enthält die Haushalteinformationen aus der Haushaltegenerierung und bildet die Haushaltszusammenhänge ab. Hierzu gehören z.B. Merkmale wie Haushaltsgröße, Haushaltstyp und Wohndauer sowie EU-Merkmale. Es werden nur Personen mit eigener Haushaltsführung betrachtet. Personen in Gemeinschaftsunterkünften sind dementsprechend nicht enthalten. Nebenwohnsitzpersonen finden in der Auswertung von Haushalten Berücksichtigung. Ein Haushalt besteht aus einer Menge von Personen, wobei jeder Haushalt mindestens eine Person enthält.

➤ Tabelle KANTE

Diese Tabelle enthält Beziehungsinformationen von Person zu Person, die im Zuge der Haushaltegenerierung gebildet wurden. Die enthaltenen Merkmale ermöglichen die Darstellung von Partnerschaftsbeziehungen (z.B. Vormund und Mündel oder Elternteil und Kind) zwischen zwei Personen in einem Haushalt.

➤ Tabelle PERSON

Diese Tabelle beinhaltet Personeninformationen nach Abschluss der Haushaltegenerierung. Eine Person befindet sich an genau einer Anschrift, ist genau einem oder – z.B. im Falle von Gemeinschaftsunterkünften – keinem Haushalt, einer oder keiner Familie und einer oder keiner Wohnung zugeordnet. Personen in Gemeinschaftsunterkünften und Deutsche im Ausland sind keiner Wohnung zugeordnet. Jeder Person ist genau ein oder kein Stichproben-Datensatz zugeordnet. Die Tabelle PERSON ist die zentrale Tabelle für Auszählungen aus dem Registerbestand und enthält demografische Merkmale und Auswertungsmerkmale nach nationaler und nach EU-Typisierung.

➤ Tabelle PERSON\_STP

Diese Tabelle ist die zentrale Tabelle für Hochrechnungen aus der Personenerhebung. Sie enthält alle Stichprobenmerkmale inklusive Hochrechnungsfaktoren sowie Auswertungsmerkmale nach nationaler und nach EU-Typisierung. Die Haushaltsstichprobe enthält z.B. Informationen zu Demografie und Wohnort, zu Staatsangehörigkeit und Einwanderungserfahrung, zu beruflicher Ausbildung, Bildung und Schulbesuch und zu Erwerbstätigkeit und Arbeitsort.

➤ Tabelle REGIONALINFORMATION

Diese Tabelle enthält die Kerninformationen aller regionalen Gliederungsebenen, wie etwa Flächen, Regionalschlüssel oder Namen administrativer Einheiten.

## ➤ Tabelle WOHNUNG

Diese Tabelle enthält alle Wohnungsdatensätze nach der Haushalgenerierung. Jede Wohnung ist genau einem Gebäude zugeordnet. Jeder Wohnung ist genau ein oder kein Haushalt zugeordnet. Die Wohnungsmerkmale beinhalten z.B. Informationen zu Nutzungsart, Wohnfläche und Raumanzahl. Für vermietete Wohnungen liegen zudem die Nettokaltmiete und für leerstehende Wohnungen Informationen zu Dauer und Gründen des Leerstands sowie zur Verfügbarkeit der Wohnung innerhalb der nächsten 3 Monate vor.

Für die Geheimhaltung mit der Cell-Key-Methode (CKM) wurden die statistischen Einheiten (Personen, Haushalte, Familien, Gebäude, Wohnungen) um die hierfür erforderlichen Merkmale, die sogenannten „Record Keys“<sup>3</sup>, erweitert.

Es wird zwischen Merkmalen unterschieden, die mittels Stichprobe erhoben wurden (Stichprobenmerkmale) und solchen, die aus Registern oder durch Vollerhebung gewonnen wurden (Auszählungsmerkmale). Stichprobenmerkmale werden (ohne Nebenwohnsitzpersonen) hochgerechnet, um Ergebnisse auf der gewünschten Aggregationsebene zu erzeugen. Auszählungsmerkmale werden hingegen ungewichtet ausgewertet. Auswertungen für Gebäude, Wohnungen, Haushalte und Familien erfolgen stets als Auszählung. Tabellenübergreifende Auswertungen, also Kombinationen von Personenmerkmalen aus der Stichprobe mit Registermerkmalen sind möglich.

Einen genauen Überblick über die in Produkt 3 enthaltenen Merkmale und die zugehörigen Merkmalskategorien bietet der Merkmalskatalog zu Produkt 3 auf unserer [FDZ-Website](#).

## **2.2 Vergleichbarkeit der Merkmale über die Zeit**

Mit Vorliegen des Zensus 2022 im Datenangebot der FDZ ist es erstmals möglich, vergleichende Auswertungen zwischen zwei Zensus-Erhebungen vorzunehmen. Die Vergleichbarkeit mit den Daten von 2011 wird im [Qualitätsbericht](#) des Zensus 2022 (vgl. S. 13 ff.) beschrieben. Merkmale, die nur in einer der beiden Zensusrunden erhoben wurden oder sich nicht aus inhaltlich tiefergegliederten Merkmalen identisch bilden lassen, eignen sich nicht für vergleichende Auswertungen. Das Gleiche gilt für Merkmale, die sich inhaltlich verändert haben. Inhaltliche Veränderungen traten insbesondere in der Haushaltsstichprobe auf. So sind insbesondere die Daten zum Beruf aufgrund von geänderten Definitionen und Abgrenzungen in der EU-

---

<sup>3</sup> Bei der Cell-Key-Methode wird jeder Beobachtungseinheiten (Record) im Originaldatenbestand eine gleichverteilte Zufallszahl aus dem Intervall [0,1) zugewiesen. Die Zufallsziehung der Record-Keys stellt die stochastische Komponente des Geheimhaltungsverfahrens dar.

Durchführungsverordnung (EU) Nummer 543/2017 nur eingeschränkt miteinander vergleichbar. Aufgrund der geänderten Methodik bei der Erhebung ist zudem die Vergleichbarkeit von Bildungsmerkmalen stark eingeschränkt und stellenweise nicht gegeben. Das Migrationskonzept des Zensus 2011 wurde im Zensus 2022 durch das Konzept der Einwanderungsgeschichte ersetzt. Aussagen zu Migrationshintergrund und Migrationserfahrung sind daher nicht vergleichbar.

Durch Gebietsreformen in den Jahren zwischen 2011 und 2022 ist die Vergleichbarkeit für regionale Einheiten wie dem amtlichen Gemeindegemeinschaftsschlüssel nicht mehr vollständig gegeben. Bei Datennutzung des Zensus 2011 wird daher ein Merkmal zur Verfügung gestellt, dass die Umschlüsselung auf die Gebietsgrenzen des 15. Mai 2022 ermöglicht.

Eine detaillierte Gegenüberstellung kann auf Basis der Schlüsselverzeichnisse von Produkt 3 zum [Zensus 2011](#) und [Zensus 2022](#) vorgenommen werden, in denen die Merkmale und ihre Definitionen beschrieben werden.

## **2.3 Eckwerte relevanter Merkmale und Merkmalskombinationen**

Die Fallzahlen der FDZ-Daten für Produkt 3 entsprechen nach Anwendung der CKM-Geheimhaltung (siehe Kapitel 3.1.3. „Cell-Key-Methode“) denen des Zensus-Veröffentlichungsprogrammes. Diese Informationen lassen sich z.B. in der [Zensus-Datenbank](#) abrufen. Weitere Veröffentlichungen zur GWZ auf Gitterzellen-Ebene sind in den [Publikationen](#) zum Zensus 2022 enthalten oder als Kartenanwendung im [Zensus-Atlas](#) zu finden.

## **2.4 Auswertbare regionale Ebene**

Die Ergebnisse können auf Bundes-, Landes-, Regierungsbezirks-, Kreis-, Gemeindeverbands- sowie auf Gemeindeebene dargestellt werden. Zudem erlaubt Produkt 3 – anders als Produkt 1 und Produkt 2 – kleinräumige Auswertungen. Die Zuordnung von Anschriften zu Gitterzellengröße 10 km x 10 km sowie 1 km x 1 km ermöglicht Auswertungen nach Gitterzellen. Gitterzellen mit der Seitenlänge 100 m können zur Erzeugung eigener räumlicher Aggregate (z. B. Lärmzonen um Flughäfen) genutzt werden. Die Gebietsstände entsprechen dem Stand vom 15. Mai 2022. Danach eingetretene Gebietsstandsänderungen sind nicht berücksichtigt. Zu beachten ist, dass die tiefste regionale Ebene für Ergebnisse auf Basis der Haushaltsstichprobe<sup>4</sup> der amtliche Gemeindegemeinschaftsschlüssel (AGS\_12) ist. Da bei dieser Ebene mit sehr vielen Geheimhaltungsfällen zu rechnen ist, sollte die Auswertung nach Möglichkeit auf Bundes-, Landes-, Regierungsbezirks- oder

---

<sup>4</sup> Dies betrifft Auswertungen, die mindestens ein Merkmal umfassen, welches ausschließlich in der Teiltabelle PERSON\_STP (Produkt 1) erhoben wurde.

Kreisebene erfolgen. Unterhalb dieser Ebene liegen aufgrund der geringen Stichprobengröße nur für Gemeinden und zensuspezifische Stadtteile mit einer Einwohnerzahl von mindestens 10 000 Einwohnerinnen und Einwohnern verlässliche Angaben vor.

## **3 Praktische Hinweise**

### **3.1 Hinweise zur Geheimhaltung**

#### **3.1.1 Gesetzliche Grundlagen der statistischen Geheimhaltung**

Unter Geheimhaltung versteht man das Herstellen der absoluten Anonymität der Ergebnisse statistischer Analysen. Konkret bedeutet das, dass im Rahmen der Geheimhaltung sichergestellt wird, dass mit den veröffentlichten Ergebnissen keine Rückschlüsse auf einen Einzelfall (z.B. Person, Betrieb, Einrichtung) gezogen werden können. Statistische Geheimhaltung wird überall dort angewendet, wo statistische Ergebnisse oder Einzeldaten die geschützten Räume der amtlichen Statistik verlassen.

Die Geheimhaltung in der amtlichen Statistik ist in § 16 Bundesstatistikgesetz (BStatG) geregelt und beinhaltet, dass Einzelangaben über persönliche und sachliche Verhältnisse, die für eine Bundesstatistik angegeben werden, von den jeweils durchführenden statistischen Stellen geheim zu halten sind, soweit es keine anderslautenden Bestimmungen gibt. Dies wird auch als Statistikgeheimnis bezeichnet. Das Statistikgeheimnis verpflichtet die amtliche Statistik, die erhaltenen Informationen zu schützen, d.h. sie in einer Form zu anonymisieren, die keine Rückschlüsse mehr auf die betreffende Person und den dargelegten Sachverhalt enthält. Die Geheimhaltung ist auch im Hinblick auf die informationelle Selbstbestimmung von besonderem Interesse: Viele Erhebungen der amtlichen Statistik unterliegen der Auskunftspflicht, somit steht es den Befragten nicht frei, selbst zu entscheiden, ob sie eine Information weitergeben möchten. Die amtliche Statistik muss deshalb sicherstellen, dass die erhobenen Daten keinem Befragten zugeordnet werden können.

Das BStatG sieht jedoch auch Fälle vor, in denen das Statistikgeheimnis nicht gilt. In § 16 BStatG sind die Ausnahmen von der Geheimhaltungspflicht dargelegt. Unter anderem wird dort festgelegt, unter welchen Umständen die Daten der amtlichen Statistik für die Wissenschaft zugänglich gemacht werden dürfen und welche Regeln dabei einzuhalten sind.

### 3.1.2 Geheimhaltung von Ergebnissen

Um die gesetzlich vorgeschriebene Geheimhaltung von Einzelfällen in den Daten sicherzustellen, müssen alle Ergebnisse, die am Gastwissenschaftsarbeitsplatz oder per Kontrollierter Datenfernverarbeitung erzeugt werden, vor ihrer Freigabe an die Nutzenden von den FDZ einer Geheimhaltungsprüfung unterzogen werden. Dabei stellen die FDZ sicher, dass die Ergebnisse absolut anonym sind und eine Reidentifikation einzelner Befragter nach menschlichem Ermessen ausgeschlossen werden kann. Entsprechend handeln auch die Fachabteilungen der Statistischen Ämter vor der Veröffentlichung von Ergebnissen.

Zur Sicherstellung der Geheimhaltung wenden die FDZ verschiedene Geheimhaltungsregeln an, die jeweils individuell auf die jeweilige Statistik zugeschnitten sind. In der Broschüre „[Regelungen zur Auswertung von Mikrodaten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#)“ werden die gebräuchlichsten Regeln zur primären Geheimhaltung dargestellt. Diese Regeln werden in den FDZ im Grunde auf alle Statistiken angewendet. Die Anlage der Broschüre „[Übersicht Geheimhaltungsregeln](#)“ enthält Informationen darüber, welche Geheimhaltungsregeln auf welche Statistiken anzuwenden sind. Abweichend von den üblichen Regelungen der FDZ gelten für Ergebnisse, die aus der Haushaltsstichprobe des Zensus 2022 hochgerechnet werden, die folgenden Regeln:

- Mindestfallzahl

Es gilt eine Mindestfallzahl von 211 Fällen in den hochgerechneten Ergebnissen. Bitte geben Sie daher immer auch die hochgerechneten Fallzahlen aus. Diese Mindestfallzahl gilt auch für weiterführende Auswertungen wie z. B. Regressionen. Bitte geben Sie auch hier die dahinterstehenden hochgerechneten Fallzahlen aus.

- Ergebnisrundung

Nach der Prüfung auf Einhaltung der Mindestfallzahl werden alle verbleibenden Ergebnisse auf ein Vielfaches von 10 gerundet (kaufmännische Rundung).

### 3.1.3 Cell-Key-Methode

Die Daten des Zensus 2022 werden für die Produkte 2 („Gebäude- und Wohnungszählung“) und 3 („Gesamtdatensatz“) mit dem datenverändernden Geheimhaltungsverfahren der stochastischen Überlagerung nach der Cell-Key-Methode (CKM) geheim gehalten. Die CKM löst das für den Zensus 2011 verwendete prä-tabulare Geheimhaltungsverfahren SAFE (Sichere Anonymisierung für

Einzeldaten) ab. Auch die FDZ-Produkte des Zensus 2011 werden ab sofort mit CKM geheim gehalten.

Nähere Informationen zur Cell-Key-Methode können dem Artikel „Geheimhaltung mit der Cell-Key-Methode im Zensus 2022“ ([Enderle und Kleber 2024](#)) entnommen werden. Eine Übersicht über Literatur zur CKM findet sich zudem im [Statistikportal der Statistischen Ämter des Bundes und der Länder](#). Bitte beachten Sie auch die Hinweise zur Arbeit mit Daten, bei denen die CKM angewendet wird und der damit verbundenen Programmierregeln im FDZ. Sie finden eine entsprechende Handreichung auf unserer [FDZ-Website](#).

## 3.2 FAQ

### **Wie erhalte ich Zugang zu den Zensus-Daten?**

Die Zensus-Daten können bei den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder beantragt werden. Die Datennutzung ist entgeltpflichtig. Auf den Seiten der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder finden Sie nähere Informationen zu [Datenangebot](#), [Datenzugang](#), [Beantragung](#) und [Bedingungen](#).

Der Datenzugang erfolgt über die Zugangswege Gastwissenschaftsarbetsplatz (GWAP) und Kontrollierte Datenfernverarbeitung (KDFV). Am GWAP stehen die formal anonymisierte Einzeldaten aufgrund des Datenumfangs lediglich als Stichprobe zur Verfügung. Die finale Ergebniserstellung erfolgt für Produkt 3 ausschließlich per KDFV. Der GWAP dient häufig der Erprobung von Auswertungsprogrammen und Modellen, da die Mikrodaten direkt eingesehen werden können. Kostenpflichtig ist bei einer kombinierten Nutzung nur der Zugangsweg KDFV, die GWAP-Nutzung erfolgt kostenneutral.

### **Wer erhält Datenzugang?**

Über die Forschungsdatenzentren bieten die Statistischen Ämter des Bundes und der Länder ausschließlich Hochschulen und sonstigen Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung verschiedene Formen des Zugangs zu ausgesuchten Datenbeständen der amtlichen Statistik für wissenschaftliche Zwecke an. Auch Studierenden, die Einzelangaben der amtlichen Statistik für Seminar- oder Abschlussarbeiten nutzen möchten, stehen die genannten Nutzungswege offen.

### **Wo erhalte ich Auskunft, wenn ich Fragen zum Datenangebot habe?**

Informationen zum Datenangebot und zum Datenzugang sind auf den [Seiten der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#) abrufbar. Bitte

wenden Sie sich bei weiteren auftretenden Fragen zu Produkt 3 über die im Impressum genannten Kontaktinformationen an das fachlich zuständige FDZ der Statistischen Ämter der Länder am Standort Fürth.

### **In welcher Form werden die Daten bereitgestellt?**

Sofern es Umfang und Struktur der ausgewählten Merkmale erlauben, werden die Teiltabellen vom fachlich zuständigen FDZ-Standort (Bayern) über Identifikatoren miteinander verknüpft und als ein sogenannter „flacher“ Datensatz bereitgestellt, der aus einer Tabelle mit nur einer Analyseebene besteht. In bestimmten Fällen kann es sinnvoller sein mehrere Teiltabellen bereitzustellen, die im Rahmen der FDZ-Datennutzung eigenständig für bestimmte Analysezwecke verknüpft werden, beispielsweise wenn Auswertungen der Teiltabelle KANTE gewünscht sind. Im Vorfeld muss eine projektspezifische Variablenauswahl getroffen werden, um den Umfang des Datensatzes zu reduzieren.

### **3.3 Verfügbare Tools**

Die FDZ stellen auf ihrer Website [Codes in Stata und R](#) zur Verfügung, mit denen die Mehrfachangabe im Merkmal BERUFABS\_ALLE in mehrere Merkmale zu den einzelnen Ausbildungsabschlüssen aufgeteilt wird.

### **Abkürzungsverzeichnis**

BStatG	Bundesstatistikgesetz vom 2. Januar 1987
CKM	Cell-Key-Methode
FDZ	Forschungsdatenzentrum/Forschungsdatenzentren
GWZ	Gebäude- und Wohnungszählung
HHSt	Haushaltsstichprobe
SAFE	Sichere Anonymisierung für Einzeldaten

Statistische Ämter des Bundes und der Länder,  
Metadatenreport – Teil II: Produktspezifische Informationen zur Nutzung des Zensus 2022 (Produkt 3 –  
Gesamtdatensatz) per On-Site-Nutzung

Fotorechte Umschlag: © artSILENCE – Fotolia.com