

Aude, David; Bolz, Johannes; Boragk, Lisa

Article

Die Aktualisierung des Ortsverzeichnisses und Plausibilisierung des Geburtsstaates im Zensus 2022

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Aude, David; Bolz, Johannes; Boragk, Lisa (2024) : Die Aktualisierung des Ortsverzeichnisses und Plausibilisierung des Geburtsstaates im Zensus 2022, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 76, Iss. 6, pp. 61-72

This Version is available at:

<https://hdl.handle.net/10419/309500>

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.

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 Gemeindennamen 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 unbekanntem 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 Zugugsstaat 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)	Merkmalskombinationen	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.

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.