

Canan, Coskun; Eberle, Jan

Article

Geburtsstaat und Geburtsort im Ausländerzentralregister – Nutzungsmöglichkeiten für die amtliche Statistik

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Canan, Coskun; Eberle, Jan (2022) : Geburtsstaat und Geburtsort im Ausländerzentralregister – Nutzungsmöglichkeiten für die amtliche Statistik, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 74, Iss. 2, pp. 53-67

This Version is available at:

<https://hdl.handle.net/10419/253431>

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.

GEBURTSSTAAT UND GEBURTSORT IM AUSLÄNDERZENTRALREGISTER – NUTZUNGSMÖGLICHKEITEN FÜR DIE AMTLICHE STATISTIK

Coşkun Canan, Jan Eberle

📌 **Schlüsselwörter:** Datenaufbereitung – administrative Daten – Geburtsstaat – Migration – Ausländerzentralregister (AZR) – Geburtsort

ZUSAMMENFASSUNG

Die amtliche Statistik verwendet die Merkmale Geburtsstaat und Geburtsort aus dem Ausländerzentralregister. Die Merkmale werden im Verwaltungskontext jedoch unvollständig (Geburtsstaat) beziehungsweise unstrukturiert als Freitextangabe (Geburtsort) und ohne Plausibilitätskontrollen erhoben.

Das Statistische Bundesamt entwickelt Konzepte zur Standardisierung, Prüfung und Ergänzung von Verwaltungsdaten, um diese für statistische Zwecke nutzbar zu machen. Der Fokus dieses Aufsatzes liegt zunächst auf der Aufbereitung der Angaben zum Geburtsstaat im Ausländerzentralregister zum Stichtag 31. Dezember 2020 und darauf basierenden ersten beispielhaften Auswertungen. Zudem wird ein Konzept für eine potenzielle Aufbereitung und Geokodierung des Geburtsorts vorgestellt.

📌 **Keywords:** data processing – administrative data – country of birth – migration – Central Register of Foreigners – place of birth

ABSTRACT

Official statistics use the variables “country of birth” and “place of birth” from the Central Register of Foreigners. In the administrative context, however, the variables are collected incompletely (country of birth) or unstructured as free text (place of birth) and without plausibility checks.

The Federal Statistical Office develops concepts for the standardisation, validation and supplementation of administrative data, so that these can be used for statistical purposes. The article focuses on the processing of the country of birth data from the Central Register of Foreigners for reference day 31 December 2020 and on first exemplary analyses based on that. Also, a concept for the potential processing and geocoding of the place of birth data is presented.



Dr. Coşkun Canan

hat an der Humboldt-Universität zu Berlin im Fach Soziologie promoviert. Er ist wissenschaftlicher Mitarbeiter im Referat „Bevölkerungstatistische Auswertungen und Analysen aus dem Mikrozensus“ des Statistischen Bundesamtes und befasst sich vor allem mit der Neudefinition des Migrationshintergrunds.



Jan Eberle

ist studierter Volkswirt und Referent im Referat „Bevölkerungsfortschreibung, Ausländer- und Integrationsstatistiken“ des Statistischen Bundesamtes. Zu seinen Arbeitsschwerpunkten gehören Auswertungen des Ausländerzentralregisters zu den Themengebieten Ausländerinnen und Ausländer sowie Schutzsuchende.

1

Einleitung

Die vermehrte Nutzung von Verwaltungsdaten und die damit einhergehende Entlastung der Auskunftgebenden ist ein Kernziel der Digitalen Agenda des Statistischen Bundesamtes (Statistisches Bundesamt, 2019). Die Nutzung des Ausländerzentralregisters (AZR) für die Erstellung der zentralen Ausländerstatistik auf der Basis des § 23 des Gesetzes über das Ausländerzentralregister (AZR-Gesetz) hat dabei schon eine lange Tradition in der Bevölkerungsstatistik. Bereits seit den 1970er-Jahren werden die Daten des AZR aufbereitet, um Informationen zur ausländischen Bevölkerung zu gewinnen (Fleischer, 1989).

Bis heute (Stand: März 2022) ist das AZR eines der größten Verwaltungsregister in Deutschland. Es enthält Informationen zu allen ausländischen Staatsangehörigen¹, die sich nicht nur vorübergehend – das heißt in der Regel länger als drei Monate – in Deutschland aufhalten. Das AZR führt die Datenbestände aller lokalen Ausländerbehörden zusammen und dient damit einer Vielzahl von mit asyl- und ausländerrechtlichen Verwaltungsaufgaben betrauten Behörden als zentrale Informationsplattform.

Wie bei Verwaltungsdaten üblich, ist das Statistische Bundesamt damit nur einer von zahlreichen Nutzern und die statistische Verwertbarkeit der Daten nur einer von vielen Verwendungszwecken. Entsprechend unterscheidet sich die Qualität der Rohdaten auch von jener von Daten aus Primärerhebungen. Vor einer Verwertung der Daten für statistische Zwecke steht daher in vielen Fällen eine sorgfältige Qualitätsprüfung und Aufbereitung.

Infolge der Fluchtmigration der Jahre 2015 und 2016 wurde das AZR kontinuierlich weiterentwickelt. Im Jahr 2019 hat der Gesetzgeber mit dem Zweiten Datenaustauschverbesserungsgesetz die Nutzungsmöglichkeiten für die amtliche Statistik weiter verbessert. Seitdem erhält das Statistische Bundesamt für die Erstellung der Ausländerstatistik auch die zusätzlichen Informationen zu Geburtsort und Geburtsstaat von ausländischen Staatsangehörigen.

1 Personen mit deutscher Staatsangehörigkeit, die zugleich eine ausländische Staatsangehörigkeit besitzen, sind im AZR nicht erfasst.

Angaben zum Geburtsort werden im AZR als Pflichtangabe zwar fast vollständig erfasst, allerdings erfolgt die Erfassung unstrukturiert als Freitextangabe. Angaben zum Geburtsstaat werden hingegen standardisiert als ISO-Code² erfasst, die Angaben sind jedoch in erheblichem Umfang unvollständig. Beide Merkmale werden im Verwaltungskontext ohne Plausibilitätskontrollen erhoben.

Das folgende Kapitel 2 informiert über die Rolle der Merkmale Geburtsstaat und Geburtsort in den Bevölkerungsstatistiken, Kapitel 3 untersucht die Qualität der dazu im AZR vorhandenen Daten. Die Aufbereitung der Angaben zum Geburtsstaat im AZR und erste Auswertungen zum Jahresende 2020 beschreiben die Kapitel 4 und 5. Ein mögliches Konzept zur Aufbereitung und Geokodierung des Merkmals Geburtsort wird in Kapitel 6 kurz vorgestellt. Der Beitrag schließt mit einem Fazit.

2

Bedeutung der Merkmale Geburtsstaat und Geburtsort in den Bevölkerungsstatistiken

Das Merkmal Geburtsstaat erweitert das Auswertungspotenzial der Ausländerstatistik und der gleichfalls auf dem AZR beruhenden Statistik über Schutzsuchende: Zuvor konnten zur Herkunft einer Person nur die Staatsangehörigkeit und ein binärer Indikator, ob eine Person in Deutschland geboren wurde, ausgewertet werden. Das Merkmal Geburtsstaat ermöglicht nun, neben der Staatsangehörigkeit die ausländische Bevölkerung nach ihren Herkunftsländern differenzierter zu betrachten.

Der Geburtsstaat ist ein unveränderlicher Indikator für das Kriterium Einwanderung. Die Staatsangehörigkeit einer Person kann sich dagegen durch eine Einbürgerung verändern. Ebenso erlaubt das Merkmal Geburtsstaat eine eindeutige Zuordnung von Personen zu einem Herkunftsstaat. Probleme wie bei der Zuordnung von Personen mit mehreren Staatsangehörigkeiten zu einem Staat treten hier nicht auf. Aus diesen Gründen zählt der Geburtsstaat auch zu den standarddemografischen

2 Die Internationale Organisation für Normung (International Organization for Standardization – ISO) legt Ländercodes in der Norm 3166-1 fest.

Merkmale im Bereich der internationalen Migrations- und Bevölkerungsstatistik. In Deutschland konnten Auswertungen zum Geburtsstaat bisher nur auf Basis der Befragungsdaten der Mikrozensusstichprobe erfolgen. Damit verbunden sind Einschränkungen, zum Beispiel kleinere Fallzahlen im Hinblick auf kleinräumige Auswertungen und auf die fachliche Gliederungstiefe. Auch in weiteren amtlichen Bevölkerungsstatistiken wird daher derzeit an der Erweiterung der Auswertungsmöglichkeiten des Merkmals Geburtsstaat gearbeitet, beispielsweise mithilfe von maschinellen Lernverfahren (Feuerhake und andere, 2020).

Die Angaben zum Geburtsort spielen für das Auswertungspotenzial der Ausländerstatistik und der Statistik über Schutzsuchende eine doppelte Rolle. Einerseits ist der Geburtsort von großer Bedeutung für die Qualitätsprüfung und die Aufbereitung der Angaben zum Geburtsstaat. Der Geburtsort einer Person ist eine Pflichtangabe bei der Registrierung einer Person im AZR und standardmäßig in vielen internationalen Ausweisdokumenten eingetragen. Angaben zum Geburtsstaat fehlen hingegen häufig in offiziellen Ausweisdokumenten, zudem besteht die Gefahr, dass sie bei der Erfassung mit der Staatsangehörigkeit oder dem Wohnort einer Person verwechselt werden. Über die Angaben zum Geburtsort ist es in vielen Fällen möglich, die Plausibilität von vorhandenen Angaben zum Geburtsstaat zu prüfen oder eine fehlende Angabe zu ergänzen.

Andererseits ist der Geburtsort selbst ein für die Beantwortung wichtiger Fragestellungen benötigtes Auswertungsmerkmal, zum Beispiel: Aus welchen indischen Regionen kommen die meisten Fachkräfte nach Deutschland? Kommen Schutzsuchende vermehrt aus Regionen, die stärker vom Klimawandel betroffen sind? Für die Beantwortung dieser und ähnlicher Fragen kann die Angabe zum Geburtsort als (Proxy-)Indikator für die Herkunft unterhalb der nationalen Grenzen herangezogen werden.

3

Datenqualität

Der Geburtsstaat wird als dreistelliger alphanumerischer ISO-Code erfasst (ISO 3166-1 alpha-3), zum Beispiel „DEU“ für Deutschland oder „FRA“ für Frankreich.

Die Angabe des Geburtsstaats ist kein Pflichtfeld bei der Registrierung im AZR. Vielmehr wird der Geburtsstaat im Zuge der Pflichtangabe zum Geburtsort häufig mit erfasst. Beim Geburtsort handelt es sich um eine Freitextangabe auf der Basis von amtlichen Dokumenten, sofern diese vorhanden sind. Andernfalls geschieht der Eintrag nach mündlicher Aussage.

Das Merkmal Geburtsstaat weist zum Stichtag 31. Dezember 2020 mit 33,2% einen hohen Anteil an fehlenden Einträgen auf. Die Angaben zum Geburtsort fehlen hingegen nur in 0,4% der Fälle. Zu rund 7,6 Millionen im AZR registrierten ausländischen Staatsangehörigen gibt es sowohl einen Eintrag zum Geburtsort als auch einen Eintrag zum Geburtsstaat (66,6%). Die vorhandenen Angaben können allerdings erhebliche Schreib- und Verknüpfungsfehler enthalten. Zum einen treten Rechtschreibfehler sowie unterschiedliche Darstellungen gleicher Orte auf (zum Beispiel „Paris“ und „Paris, Frankreich“), zum anderen können Geburtsorte mit falschen Geburtsstaaten verknüpft sein (zum Beispiel Paris mit dem Geburtsstaat Deutschland). Entsprechend sind hier qualitätssichernde Maßnahmen erforderlich. Bei knapp 3,8 Millionen registrierten ausländischen Staatsangehörigen ist zwar ein Geburtsort, aber kein Geburtsstaat registriert (33,1%). In der Zuweisung von bisher fehlenden Staaten zu vorhandenen Ortsangaben liegt vor diesem Hintergrund großes Potenzial für die Datenaufbereitung. [↘ Tabelle 1](#)

Um fehlerhafte Geburtsort-Geburtsstaat-Verknüpfungen zu identifizieren, muss ein Vergleichsdatenbestand an korrekten Verknüpfungen herangezogen werden. Damit ist es möglich, zum einen die Qualität des Geburtsstaatseintrags im AZR zu prüfen und zum anderen den Geburtsstaat zu ermitteln, wenn der Eintrag dazu fehlt.

Ähnliche Probleme und Zielsetzungen bestehen in mehreren amtlichen Statistiken (Feuerhake und andere, 2020). Daher hat das Statistische Bundesamt einen Vergleichsdatenbestand an weltweiten Ortsnamen und zugewiesenen Staaten als zentrale Leitdatei etabliert (Ortsdatei). Diese Ortsdatei besteht aus unterschiedlichen Referenzdateien: Die Grundlage bildet das Ortsverzeichnis des letzten Zensus. Dieses enthält die 2011 aus den Melderegistern erhobenen Angaben zu Geburtsorten und Geburtsstaaten der gesamten Bevölkerung. Diese Angaben werden zusätzlich durch das Gemeindeverzeichnis der Statistischen Ämter des Bundes und der Länder ergänzt und weiter um Geburtsorte aus den

Tabelle 1

Vorhandene und fehlende Angaben zum Geburtsort und Geburtsstaat

	Angabe Geburtsstaat		
	ja	nein	insgesamt
	Anzahl		
Angabe Geburtsort	7 639 070	3 793 385	11 432 460
ja	7 613 225	3 779 090	11 392 310
nein	25 850	14 300	40 145
	%		
Angabe Geburtsort ¹	66,8	33,2	100
ja	66,6	33,1	99,6
nein	0,2	0,1	0,4

1 Angaben, die offensichtlich ohne semantischen Inhalt sind (zum Beispiel „unbekannt“) wurden nicht als Angabe gewertet.

natürlichen Bevölkerungsstatistiken (Geburten und Sterbefälle) vervollständigt. Die Namen aller Gemeinden europäischer Staaten werden dem vom Statistischen Amt der Europäischen Union, Eurostat, geführten System der lokalen Verwaltungseinheiten entnommen. Ein historisches Staatenleitband deckt zusätzlich ehemalige Schreibweisen von Ortsnamen ab. Letztlich komplettiert ein Katalog von verschiedenen, auch fehlerhaften, Schreibweisen und von Synonymen für unbekannte oder fehlende Angaben die Leitdatei.

Das Vorhaben, die Qualität der Angaben zu Geburtsorten und Geburtsstaaten in der Ausländerstatistik und in anderen Bevölkerungsstatistiken zu sichern, steht und fällt mit der Vollständigkeit und Qualität der Ortsdatei. Die Ortsdatei enthält mit insgesamt 2,2 Millionen Kombinationen eine sehr umfangreiche Liste aus Geburtsorten und -staaten. Im Hinblick auf die Qualität der Ortsdatei erfolgten sowohl automatisierte als auch manuelle Qualitätskontrollen und Verbesserungen. Beispielsweise wurden mittels WebScraping potenziell fehlerhafte Verknüpfungen identifiziert und manuell korrigiert (Feuerhake und andere, 2020). Derzeit wird geprüft, ob die Daten von OpenStreetMap die Qualität der Ortsdatei weiter verbessern können.

Jedoch können in der Ortsdatei auch weiterhin fehlerhafte Verknüpfungen auftreten oder Ortsnamen beziehungsweise relevante Schreibweisen von Ortsnamen fehlen. Die Ortsdatei befindet sich vor diesem Hintergrund in einem laufenden Qualitätssicherungsprozess und dürfte perspektivisch, vor allem durch den Zensus im Jahr 2022, weitere Verbesserungen im Hinblick auf Qualität und Vollständigkeit erfahren.

Durch den Vergleich mit der Ortsdatei kann eine erste Einschätzung der Qualität der Angaben zu Geburtsort und Geburtsstaat im AZR erfolgen. Zu rund 7,6 Millionen registrierten ausländischen Staatsangehörigen enthält das AZR sowohl einen Orts- als auch einen Staatseintrag (siehe Tabelle 1). Diese Kombinationen können mit den Angaben in der Ortsdatei abgeglichen werden. Für rund 92,7 % findet sich dabei eine entsprechende Geburtsort-Geburtsstaat-Verknüpfung in der Ortsdatei. Für die verbleibenden 7,3 % gibt es keine Entsprechung in der Ortsdatei.

Die Fälle ohne Entsprechung in der Ortsdatei zeigen, dass im AZR häufig deutsche Geburtsorte fälschlicherweise mit dem Ausland verknüpft sind. Hierbei könnte es sich um Verwechslungen zwischen Geburtsstaat und Staatsangehörigkeit einer Person bei der Erfassung handeln. Weiterhin sind Abweichungen zur Ortsdatei darauf zurückzuführen, dass im AZR Geburtsorte teilweise noch Staaten zugeordnet werden, die nicht mehr existieren oder denen sie heute nicht mehr angehören (zum Beispiel im ehemaligen Jugoslawien). Letztlich sind der Kreativität bei Eingaben in Freitextfeldern keine Grenzen gesetzt. So entstehen auch einige Fälle von eigentlich korrekten Geburtsortsangaben, die der Ortsdatei noch nicht bekannt sind. Die Qualitätsbewertung der Angaben zu Geburtsort und Geburtsstaat im AZR enthält [Übersicht 1](#).

Übersicht 1

Qualitätsbewertung der Merkmale Geburtsort und Geburtsstaat

Qualitätskriterium	Geburtsort	Geburtsstaat
Pflichtangabe	ja	nein
Vollständigkeit	vollständig (0,4 % missing)	unvollständig (33,2 % missing)
Standardisierung	unstrukturierte Erfassung als Freitext ohne Plausibilitätskontrolle	strukturierte Erfassung als ISO 3166 ALPHA-3-Code
Inhaltliche Korrektheit	teilweise Angaben ohne semantischen Inhalt, mit Schreibfehlern, in unterschiedlichen Schreibweisen oder unterschiedlichen Sprachen	teilweise Angaben von zweistelligen ISO-Codes
Eindeutigkeit	problematisch wegen mehrdeutiger Ortsangaben, zum Beispiel Neustadt	ja, eindeutige ISO-Codes
Konsistenz der Kombinationen	Inkonsistente Kombinationen (7,3 %): > deutsche Orte mit dem ausländischen Staat verknüpft > fehlende Schreibweisen in Ortsdatei > Zuweisungen zu ehemaligen Staaten	

4

Aufbereitung der Angaben zum Geburtsstaat

Mit den vorhandenen Angaben zum Geburtsort und -staat von ausländischen Staatsangehörigen im AZR ist es möglich, mithilfe geeigneter Signierungs- und Plausibilisierungsmethoden belastbare Daten zu produzieren. Dabei hängt die Qualität der aufbereiteten Daten maßgeblich von der Vollständigkeit und Qualität der Angaben in der Ortsdatei ab. Mit einer vollständigen und korrekten Ortsdatei als Leitdatei kann eine Integration der neuen Merkmale in die Datenaufbereitung der Ausländerstatistik erfolgen. Damit steht die Ortsdatei auch im Zentrum der Datenaufbereitung.

Um einen Vergleich mit der Ortsdatei zu ermöglichen, erfolgt zunächst eine Standardisierung. Hierbei wird der gelieferte ISO-Staatencode zu dem in der amtlichen Statistik verwendeten Staatencode der Staats- und Gebietssystematik umgewandelt (Statistisches Bundesamt, 2022). Die Freitextangaben zum Geburtsort werden standardisiert³ und Einträge ohne semantischen Inhalt aussortiert. Beim anschließenden Vergleich mit der Ortsdatei gelten nur jene Ort-Staat-Kombinationen als plausibel, die in der Ortsdatei enthalten sind. Ist einem Ort in der Ortsdatei ein abweichender Staat zugeordnet, wird die Zuweisung im AZR korrigiert. Im Sinne einer minimal-

³ Beispielsweise werden UTF-8-Sonderzeichen in Latin1-kompatible Zeichen umgewandelt und gebräuchliche Abkürzungen (zum Beispiel USA) aufgelöst.

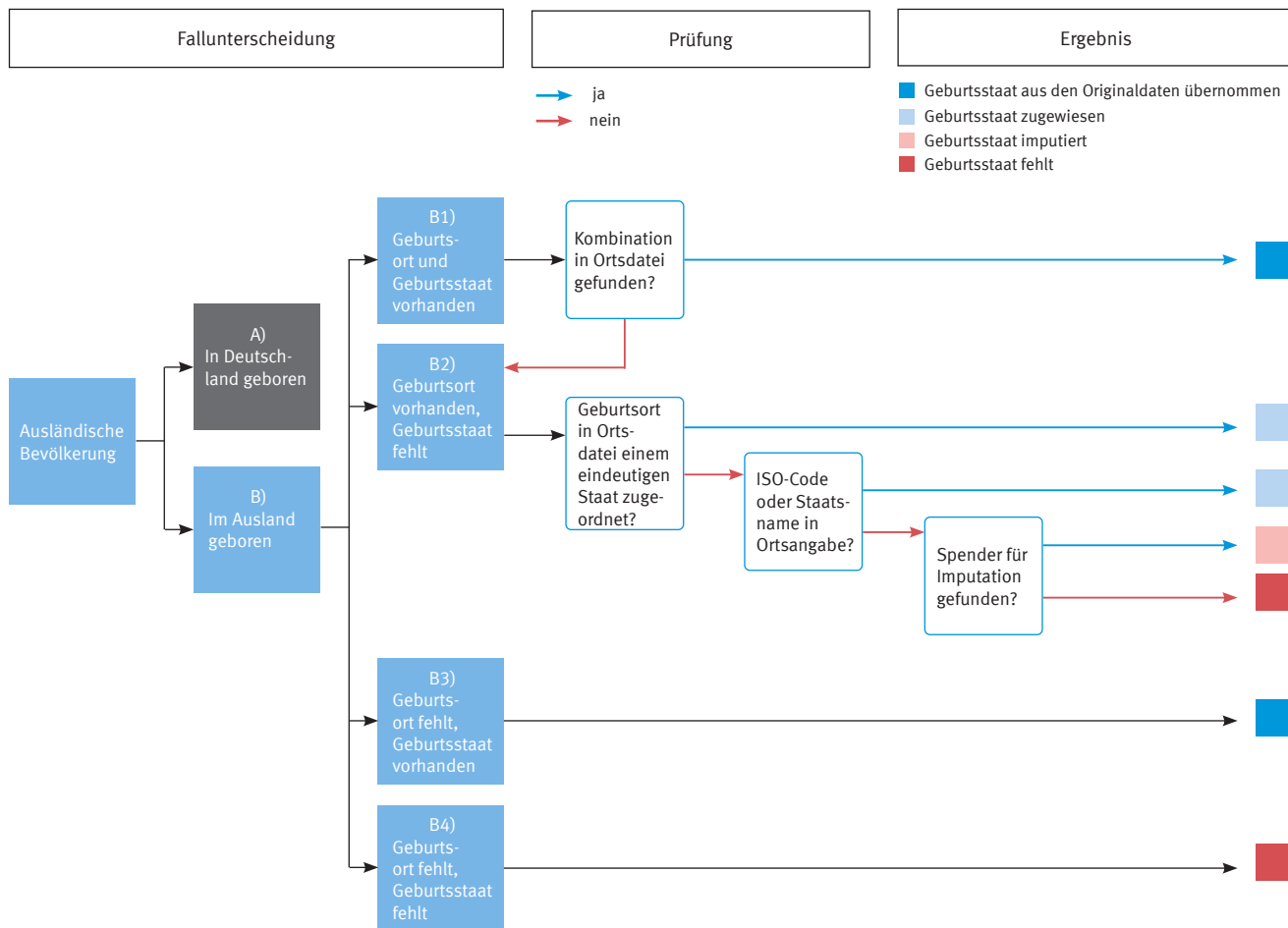
invasiven Aufbereitung beschränken sich Imputationen von Geburtsstaaten auf Fälle, in denen die Ortsangaben in der Ortsdatei mehr als einem Staat zugeordnet sind und keinen Hinweis auf einen konkreten Staat liefern. Den Ablauf der Datenaufbereitung, auf die im Folgenden genauer eingegangen wird, zeigt [Grafik 1](#).

Dazu wird zunächst die Gruppe der in Deutschland geborenen ausländischen Staatsangehörigen von derjenigen der im Ausland geborenen ausländischen Staatsangehörigen getrennt. Diese Aufteilung erfolgt anhand eines binären Indikators in den AZR-Daten, der anzeigt, ob eine Person mit ausländischer Staatsangehörigkeit in Deutschland geboren wurde oder eingewandert ist.⁴ Damit ist diese Zuordnung unabhängig von den Angaben zum Geburtsort und Geburtsstaat. Ende 2020 waren von den 11,4 Millionen registrierten ausländischen Staatsangehörigen nach diesem Indikator rund 1,5 Millionen in Deutschland geboren (Tabelle 2: QS-Kennung 1). Die weiteren Aufbereitungsschritte beziehen sich demnach auf die verbleibenden 9,9 Millionen im Ausland geborenen ausländischen Staatsangehörigen. Der Ablauf dieser Schritte ist nach einer Fallunterscheidung strukturiert, die Datensätze danach unterscheidet, ob potenziell verwertbare Angaben zu Geburtsort und Geburtsstaat vorliegen.

⁴ Dieser Indikator beruht auf einem Abgleich zwischen Geburtsdatum und Ersteinreisdatum. Bei in Deutschland geborenen ausländischen Kindern wird bei der Registrierung im AZR das Geburtsdatum als Ersteinreisdatum eingetragen. Bei eingewanderten ausländischen Staatsangehörigen entspricht das Ersteinreisdatum nicht dem Geburtsdatum.

Grafik 1

Ablauf der Datenaufbereitung der Merkmale Geburtsort und Geburtsstaat aus dem Ausländerzentralregister



2022 - 0092

Eintrag zum Geburtsort und zum Geburtsstaat vorhanden

Ist für im Ausland geborene ausländische Staatsangehörige sowohl eine Angabe zum Geburtsort als auch eine Angabe zum Geburtsstaat vorhanden, wird die Kombination anhand eines Abgleichs mit der Ortsdatei zunächst auf Plausibilität überprüft. Entspricht die Kombination aus Geburtsort und Geburtsstaat einem Eintrag in der Ortsdatei, wird sie als plausibel eingestuft (Tabelle 2: QS-Kennung 2). Findet sich die Kombination nicht in der Ortsdatei wieder, erfolgen weitere Aufbereitungsschritte: Zunächst wird geprüft, ob sich die Angabe zum Geburtsort in Kombination mit einem anderen Geburtsstaat in der Ortsdatei wiederfindet. Dabei wird angenommen, dass ein vorhandener Eintrag zum Geburtsort zuverlässiger

ist als der dazu eingetragene Geburtsstaat, da bei der Erfassung des Geburtsstaats durch potenzielle Verwechslungen mit der Staatsangehörigkeit eine höhere Wahrscheinlichkeit für Fehlerfassungen besteht. Daher wird in diesen Fällen der Geburtsstaat aus der Ortsdatei übernommen (Tabelle 2: QS-Kennung 3).

Findet sich der registrierte Geburtsort nicht in der Ortsdatei oder existieren mehrere potenzielle Staaten zu einem Ortsnamen, wird überprüft, ob die Freitextangaben zum Geburtsort einen Hinweis auf einen konkreten Geburtsstaat enthalten. Häufig wird bei den Angaben zum Geburtsort im Freitext zusätzlich der Name oder der ISO-Code eines Staates erfasst. Finden sich in der Freitextangabe ein ISO-Code oder Staatenname aus der Staats- und Gebietssystematik (Statistisches Bundes-

Geburtsstaat und Geburtsort im Ausländerzentralregister – Nutzungsmöglichkeiten für die amtliche Statistik

Tabelle 2

Ergebnisse der Datenaufbereitung

QS-Kennung	Beschreibung	Häufigkeit	Prozent
1 bis 14	Ausländische Bevölkerung	11 432 460	100
1	A) In Deutschland geboren ¹	1 509 335	13,2
2 bis 14	B) Im Ausland geboren	9 923 125	86,8
2 bis 7	B1) Angabe zu Geburtsort: ja, Geburtsstaat: ja	6 819 120	59,6
2	Kombination plausibel	6 327 680	55,3
3 bis 7	Kombination nicht plausibel	491 440	4,3
3	Zuweisung über Geburtsort anhand Ortsdatei	80 415	0,7
4	Staatenname in Ortsangabe	6 130	0,1
5	Staatencode in Ortsangabe	10 870	0,1
6	Geburtsstaat konnte imputiert werden	30 205	0,3
7	Geburtsstaat konnte nicht imputiert werden	363 820	3,2
8 bis 12	B2) Angabe zu Geburtsort: ja, Geburtsstaat: nein	3 064 250	26,8
8	Zuweisung über Geburtsort anhand Ortsdatei	1 914 245	16,7
9	Staatenname in Ortsangabe	12 670	0,1
10	Staatencode in Ortsangabe	6 805	0,1
11	Geburtsstaat konnte imputiert werden	818 195	7,2
12	Geburtsstaat konnte nicht imputiert werden	312 335	2,7
13	B3) Angabe zu Geburtsort: nein, Geburtsstaat: ja	25 530	0,2
14	B4) Angabe zu Geburtsort: nein, Geburtsstaat: nein	14 225	0,1

68,8% **Angaben aus den Originaldaten**

17,8% **Zuweisungen anhand von Leitdateien**

7,4% **imputierte Angaben**

6,0% **fehlende Angaben**

¹ Die Zuordnung erfolgt anhand eines Abgleiches zwischen Geburtsdatum und Ersteinreisedatum. Bei der Registrierung von in Deutschland geborenen ausländischen Kindern wird im Ausländerzentralregister das Geburtsdatum als Ersteinreisedatum eingetragen.

amt, 2022) wird dieser als Geburtsstaat übernommen (Tabelle 2: QS-Kennungen 4 und 5).

Liefert auch dieser Abgleich keinen Hinweis auf einen konkreten Geburtsstaat, gilt der Datensatz als ein Datensatz mit Angabe zum Geburtsort, aber ohne Angabe zum Geburtsstaat. Es erfolgt (sofern möglich) wie im folgenden Abschnitt beschrieben, eine Imputation des Geburtsstaats (Tabelle 2: QS-Kennungen 6 und 7).

Eintrag zum Geburtsort, aber kein Eintrag zum Geburtsstaat vorhanden

Fehlt der Eintrag zum Geburtsstaat im AZR, während ein Geburtsort vorhanden ist, wird zunächst geprüft, ob für den Ortsnamen eine eindeutige Verknüpfung zu einem Staat in der Ortsdatei besteht. Ist dies der Fall, wird der entsprechende Geburtsstaat übernommen (Tabelle 2: QS-Kennung 8). Anderenfalls wird auch in diesen Fällen in Betracht gezogen, dass in der Freitextangabe zum Geburtsort der Name (Tabelle 2: QS-Kennung 9) oder ISO-Code (Tabelle 2: QS-Kennung 10) des Geburtsstaats

eingetragen worden sein könnte. Findet sich über diese regelbasierten Abgleiche kein eindeutiger Geburtsstaat, kommt ein probabilistisches Random Hot-Deck-Verfahren zum Einsatz.

Dieses Imputationsverfahren sucht für jeden Datensatz mit fehlendem Geburtsstaat (Empfänger) einen möglichst ähnlichen Datensatz mit Angabe zum Geburtsstaat (Spender). Das Verfahren macht sich zunutze, dass die fehlende Geburtsstaatszuordnung einer Person bei anderen vergleichbaren Personen potenziell vorliegen kann. Als vergleichbar gelten in diesem Kontext Personen mit gleichem Geburtsort und gleicher Staatsangehörigkeit. Dabei sind alle vollständigen und als plausibel bewerteten Datensätze potenzielle Spenderdatensätze.

Damit orientiert sich das Verfahren an den Verteilungen, die in den bereits plausibilisierten AZR-Daten beobachtet werden und erhält diese auch bei der Zuteilung von bisher unbekanntem Geburtsstaaten. Soll beispielsweise für eine Person mit französischer Staatsangehörigkeit (Person A in Tabelle 3) und Geburtsort Paris ein

Geburtsstaat imputiert werden, liefert die Verteilung der Geburtsstaaten von Französinnen und Franzosen mit Geburtsort Paris die benötigte Wahrscheinlichkeitsverteilung. So wurden im Beispiel in Tabelle 3 vier der fünf französischen Staatsangehörigen mit Geburtsort Paris und mit Angaben zum Geburtsstaat (Spender) in Frankreich geboren. Entsprechend würde Person A mit einer Wahrscheinlichkeit von 80 % Frankreich und mit 20 % die Vereinigten Staaten zugewiesen. Handelte es sich, wie bei Person B, um eine Person mit US-amerikanischer Staatsangehörigkeit, würden gemäß der beobachteten Verteilung mit 33 % Frankreich und mit 67 % die Vereinigten Staaten als Geburtsstaat zugewiesen.

↘ [Tabelle 3](#)

Tabelle 3

Probabilistische Imputation

	Staatsangehörigkeit	Geburtsort	Geburtsstaat
Empfänger			
Person A	französisch	Paris	N/A
Person B	US-amerikanisch	Paris	N/A
Spender			
Person C	französisch	Paris	USA
Person D	französisch	Paris	FRA
Person E	französisch	Paris	FRA
Person F	französisch	Paris	FRA
Person G	französisch	Paris	FRA
Person H	US-amerikanisch	Paris	USA
Person I	US-amerikanisch	Paris	USA
Person J	US-amerikanisch	Paris	FRA

Kann über passende Spenderdatensätze eine solche Wahrscheinlichkeitsverteilung abgeleitet werden, ist die Imputation erfolgreich (Tabelle 2: QS-Kennung 11). Finden sich keine passenden Spenderdatensätze, kann kein Geburtsstaat ergänzt werden und der Datensatz zählt zu den Antwortausfällen (Tabelle 2: QS-Kennung 12).

Kein Eintrag zum Geburtsort

Sofern kein Geburtsort, aber ein gültiger ISO-Code als Geburtsstaat eingetragen ist, wird dieser als plausibel betrachtet und verwendet (Tabelle 2: QS-Kennung 13).

Datensätze ohne Angabe zu Geburtsort und Geburtsstaat werden als komplette Antwortausfälle (Unit-Nonresponse) betrachtet und nicht weiter aufbereitet. Auf eine wahrscheinlichkeitsbasierte Zuordnung über die Staats-

angehörigkeit wird auch aufgrund der geringen Fallzahl (0,1 %) verzichtet (Tabelle 2: QS-Kennung 14).

Fallkonstellationen im Überblick

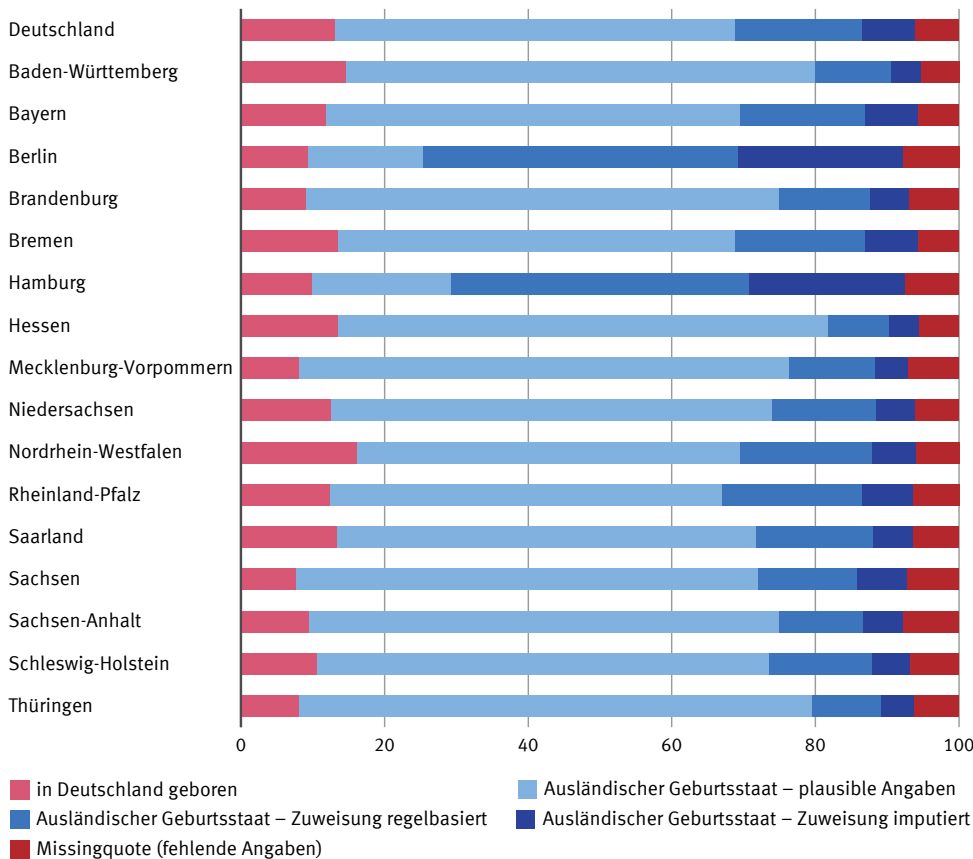
Tabelle 2 fasst die beschriebenen Fallkonstellationen zusammen und liefert damit eine Übersicht des Ergebnisses der Aufbereitung der AZR-Daten zum 31. Dezember 2020. Insgesamt wurden 68,8 % aller Geburtsstaatsangaben aus den Originaldaten übernommen.¹⁵ In 17,8 % der Fälle konnte ein fehlender Geburtsstaat anhand der Angabe zum Geburtsort über eine Leitdatei abgeleitet werden. Für 7,4 % der ausländischen Bevölkerung wurde der Geburtsstaat anhand der beobachteten Verteilung bei Personen mit gleichem Geburtsort und gleicher Staatsangehörigkeit imputiert (Imputationsquote).¹⁶ Mit letztlich rund 6,0 % verbleibenden fehlenden Angaben (Missingquote) hat die Aufbereitung insgesamt eine deutliche Qualitätssteigerung erwirkt.

Die bundesweiten Ergebnisse überlagern dabei deutliche Unterschiede zwischen den Bundesländern. Besonders Berlin und Hamburg liegen mit Imputationsquoten von 22,9 beziehungsweise 21,8 % deutlich über dem Bundesdurchschnitt. Ohne diese beiden Ausreißer liegt die durchschnittliche Imputationsquote in den übrigen Ländern bei 5,7 %. Grund für diese große Abweichung ist, dass die Erfassung des Geburtsstaats bei der Registrierung von ausländischen Staatsangehörigen im AZR in Hamburg und Berlin vergleichsweise selten erfolgt: In Berlin weisen nur 16,0 % der registrierten ausländischen Staatsangehörigen eine plausible Angabe zum Geburtsstaat auf, in Hamburg liegt dieser Wert bei 19,3 %. Der entsprechende Durchschnitt der anderen Bundesländer liegt bei 59,9 %. Die dennoch geringen Missingquoten für Berlin und Hamburg lassen sich darauf zurückführen, dass im Zuge der Aufbereitung die Pflichtangaben zum Geburtsort genutzt werden können, um den fehlenden Geburtsstaat zuzuweisen. Die hierzu notwendigen Beispiele korrekter Ort-Staat-Kombinationen liefern die plausiblen Angaben aller Bundesländer. ↘ [Grafik 2](#)

- 5 Davon für 13,20 % der Geburtsstaat Deutschland auf Grundlage des Abgleichs von Geburtsdatum und Ersteinreisdatum, für 55,35 % ein ausländischer Geburtsstaat auf Grundlage plausibler Kombinationen der Merkmale Geburtsort und Geburtsstaat und für 0,22 % ein ausländischer Geburtsstaat auf der Grundlage einer Angabe zum Geburtsstaat ohne Angabe zum Geburtsort.
- 6 Davon 0,26 % Fälle mit nicht plausibler Kombination von Geburtsort und Geburtsstaat und 7,16 % mit fehlender Angabe zum Geburtsstaat.

Grafik 2

Ergebnisse der Aufbereitung der Daten aus dem Ausländerzentralregister zu Geburtsstaaten zum 31. Dezember 2020
in %



2022 - 0093

5

Auswertungen der Angaben zum Geburtsstaat

Auf Basis der aufbereiteten Daten lassen sich Auswertungen nach Staatsangehörigkeit und Auswertungen nach Geburtsstaat gegenüberstellen. Mit der Reform des Staatsangehörigkeitsgesetzes Anfang 2000 hat der deutsche Gesetzgeber neben dem Abstammungsprinzip (*ius sanguinis*) das Geburtsortsprinzip (*ius soli*) eingeführt. Seither können in Deutschland geborene Kinder ausländischer Eltern unter bestimmten Voraussetzungen die deutsche Staatsangehörigkeit erwerben. Zu den Voraussetzungen zählt, dass ein Elternteil seit mindestens acht Jahren rechtmäßig seinen gewöhnlichen Aufenthalt

in Deutschland hat und ein unbefristetes Aufenthaltsrecht besitzt (§ 4 Absatz 3 Staatsangehörigkeitsgesetz). Mit der Einführung des Geburtsortsprinzips ging die Zahl der in Deutschland ohne deutsche Staatsangehörigkeit geborenen Kinder zunächst deutlich zurück (Bundesamt für Migration und Flüchtlinge, 2020). Seit 2016 werden wieder deutlich mehr Kinder ohne deutsche Staatsangehörigkeit in Deutschland geboren. Eine Ursache dafür ist, dass die in den Jahren 2014 bis 2016 eingereisten Schutzsuchenden als Eltern die Mindestaufenthaltsdauer von acht Jahren noch nicht erfüllen können.

➤ **Tabelle 4** stellt den zehn häufigsten Staatsangehörigkeiten die zehn am häufigsten angegebenen Geburtsstaaten gegenüber. Dabei zeigt sich, dass Deutschland Ende 2020 der häufigste Geburtsstaat ausländischer Personen war (13 %). Außerdem fällt auf, dass sich die

Tabelle 4

Häufigste Geburtsstaaten und häufigste Staatsangehörigkeiten von im Ausländerzentralregister enthaltenen ausländischen Staatsangehörigen zum 31. Dezember 2020

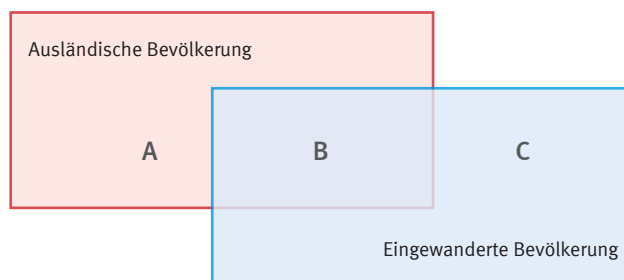
Rang	Geburtsstaat	Ausländische Staatsangehörige		Rang	Staatsangehörigkeit	Ausländische Staatsangehörige		Davon geboren in:		
		Anzahl	%			Anzahl	%	Deutschland	Staat der Staatsangehörigkeit	anderem Staat
Insgesamt		11 432 460	100	Insgesamt		11 432 460	100	X	X	X
1	Deutschland	1 509 335	13,2	1	Türkei	1 461 910	12,8	26,8	69,2	4,0
2	Türkei	1 028 560	9,0	2	Polen	866 690	7,6	6,1	90,6	3,3
3	Polen	792 765	6,9	3	Syrien	818 460	7,2	12,3	77,7	10,0
4	Syrien	687 430	6,0	4	Rumänien	799 180	7,0	7,3	77,3	15,4
5	Rumänien	628 240	5,5	5	Italien	648 360	5,7	24,1	62,4	13,5
6	Italien	427 860	3,7	6	Kroatien	426 845	3,7	12,9	51,3	35,7
7	Bulgarien	324 620	2,8	7	Bulgarien	388 700	3,4	7,3	82,7	10,0
8	Bosnien und Herzegowina	286 150	2,5	8	Griechenland	364 285	3,2	20,3	61,3	18,4
9	Griechenland	232 540	2,0	9	Afghanistan	271 805	2,4	9,9	74,3	15,8
10	Kroatien	231 585	2,0	10	Russische Föderation	263 300	2,3	5,3	73,6	21,1

Rangfolge der Staaten verändert, je nachdem, ob die Staatsangehörigkeit oder der Geburtsstaat als Indikator für die Herkunft einer Person herangezogen wird. Die kroatische Staatsangehörigkeit ist beispielsweise die am sechsthäufigsten vertretene ausländische Staatsangehörigkeit, liegt aber bei den häufigsten Geburtsstaaten nur auf dem zehnten Rang. Hintergrund ist, dass nur etwas mehr als die Hälfte der im AZR registrierten Kroatinnen und Kroaten auch in Kroatien geboren wurden. Knapp 13% der kroatischen Staatsangehörigen wurde in Deutschland geboren, rund 25% in Bosnien und Herzegowina und 11% in anderen Staaten.

Um die Qualität der aufbereiteten Daten extern zu validieren, können die Ergebnisse mit Auswertungen aus dem Mikrozensus verglichen werden. Hierbei ist zu beachten, dass bei den Auswertungen zum Thema Einwanderung und Geburtsstaat aus dem AZR nur ausländische Staatsangehörige betrachtet werden können. Deutsche Staatsangehörige (dazu zählen auch Personen mit einer deutschen und einer ausländischen Staatsangehörigkeit) sind in den AZR-Daten nicht enthalten. [↘ Grafik 3](#) Damit berücksichtigen die Auswertungen aus dem AZR beispielsweise keine Personen, die im Ausland geboren wurden und nach ihrer Einwanderung eingebürgert wurden (Gruppe C). Der Mikrozensus als repräsentative Befragung der gesamten Bevölkerung hingegen enthält auch diese Gruppe.

Grafik 3

Bevölkerungsgruppen nach Geburtsstaat im Ausländerzentralregister und im Mikrozensus



Gruppe A: Im Inland geborene ausländische Bevölkerung
 Gruppe B: Eingewanderte ausländische Staatsangehörige
 Gruppe C: Eingewanderte deutsche Staatsangehörige

2022 - 0094

Für einen Vergleich der Ergebnisse zum Geburtsstaat im AZR und im Mikrozensus können demnach im Inland geborene (Gruppe A) und im Ausland geborene (Gruppe B) Ausländerinnen und Ausländer herangezogen werden. Wegen der Corona-Pandemie bestehen für den Mikrozensus 2020 Qualitätseinschränkungen im Hinblick auf die gewohnte fachliche Auswertungstiefe (Statistisches Bundesamt, 2021a). Aus diesem Grund wurden Daten des Mikrozensus 2019 für den Vergleich herangezogen. [↘ Tabelle 5](#) stellt die Anteile der im Inland und im Ausland geborenen ausländischen Bevölkerung sowie die relative Verteilung einzelner Geburtsstaaten aus beiden Datenquellen gegenüber.

Tabelle 5

Ergebnisse des AZR-Mikrozensus-Vergleichs

	Auswertung aus dem Ausländerzentralregister zum 31. Dezember 2020	Ergebnis des Mikrozensus 2019	Differenz
	%		Prozentpunkte
Ausländische Bevölkerung insgesamt	100	100	0,0
Im Inland geboren	13,3	15,5	- 2,2
Im Ausland geboren	86,7	84,5	2,2
darunter:			
Mit Angabe zum Geburtsstaat	93,0	99,8	- 6,8
Türkei	11,1	11,4	- 0,3
Polen	8,6	8,3	0,3
Syrien	7,4	7,9	- 0,5
Rumänien	6,8	5,6	1,2
Italien	4,6	5,7	- 1,0
Bulgarien	3,5	2,9	0,6
Bosnien und Herzegowina	3,1	3,1	0,0
Griechenland	2,5	3,2	- 0,7
Kroatien	2,5	3,0	- 0,5
Russische Föderation	2,5	2,8	- 0,3
Kosovo	2,4	2,5	0,0
Irak	2,4	2,3	0,0
Afghanistan	2,2	2,1	0,1
Serbien	2,0	2,1	- 0,1
Ungarn	1,8	1,8	0,0
Österreich	1,6	1,9	- 0,3
China	1,5	1,4	0,1
Iran, Islamische Republik	1,5	1,2	0,2
Ukraine	1,4	1,6	- 0,1
Spanien	1,3	1,4	- 0,1
Sonstige Staaten	29,1	27,7	1,4

↳ Nachweis des Geburtsstaats im AZR und im Mikrozensus

Bei einem Vergleich von Ergebnissen aus Mikrozensus und AZR sind generell folgende Einschränkungen zu beachten: Der Mikrozensus ist eine repräsentative Bevölkerungsbefragung auf Basis einer Stichprobe. Die Zahlen aus dem AZR hingegen entsprechen einer Vollerhebung. Die Hochrechnung des Mikrozensus erfolgt mit den Ergebnissen der Bevölkerungsfortschreibung. Zwischen Bevölkerungsfortschreibung und AZR bestehen deutliche Abweichungen im Hinblick auf die absolute Größe der ausländischen Bevölkerung (Statistisches Bundesamt, 2021b). Darüber hinaus wird im Mikrozensus nur die Bevölkerung in Privathaushalten zum Geburtsstaat befragt, während im AZR auch Personen in Gemeinschaftsunterkünften enthalten sind.

Der Anteil an Eingewanderten unter der ausländischen Bevölkerung liegt im Mikrozensus mit 84,5% um 2,2 Prozentpunkte unter dem Ergebnis der AZR-Daten. Im Hinblick auf die häufigsten Geburtsstaaten identifiziert der Mikrozensus nahezu die gleichen Staaten wie das AZR. Auch die Rangfolgen und die Verteilungen zeigen ein vergleichbares Bild. Diese Übereinstimmung – trotz der großen Unterschiede in der Erhebungsmethodik – ist ein starker Indikator für die Validität der aus dem AZR aufbereiteten Daten.

6

Ausblick: Mögliche Geokodierung von Geburtsorten mit OpenStreetMap

Angaben zum Geburtsort von ausländischen Staatsangehörigen im AZR bieten einen Indikator für die Herkunft einer Person innerhalb der nationalen Grenzen. Sie liefern damit vielfältige Möglichkeiten zur Erweiterung des Auswertungspotenzials der Ausländerstatistik und der Statistik über Schutzsuchende. Die zuverlässige Bestimmung des korrekten Geburtsstaats ist eine notwendige Voraussetzung zur eindeutigen Identifikation eines Geburtsortes. Für die Geokodierung von Ortsangaben müssen konzeptionell drei weitere Schritte erfolgen: Gazetteer Matching, Toponym Resolution und Geokodierung. Einen guten Überblick bietet Leidner (2008).

Im Rahmen des Gazetteer Matchings wird festgestellt, welche potenziellen geografischen Orte mit einem Ortsnamen gemeint sein könnten (Goldberg und andere, 2007). Damit ist das Gazetteer Matching vergleichbar mit dem Suchen von Ortsnamen in einem Ortsverzeichnis (englisch: Gazetteer). Als weltweites Ortsverzeichnis können die Geodaten von OpenStreetMap (OSM) direkt oder aus einer vorher mittels OSM-Daten geokodierten Ortsdatei herangezogen werden.

Die OpenStreetMap-Community sammelt seit 2004 weltweite Geoinformationen und stellt diese als Open Data zur Verfügung. Vergleichbar mit der freien Enzyklopädie Wikipedia werden die Informationen freiwillig und ohne kommerzielle Interessen von einer offenen Community gesammelt und laufend aktualisiert.

Die Datenqualität ist dabei regional unterschiedlich und hängt stark davon ab, wie aktiv die jeweiligen Editorinnen und Editoren sind. Für den Zeitraum 2013 bis 2019 identifizieren Seto und andere (2020) die aktivsten Communities in Deutschland, in der Russischen Föderation und in den Vereinigten Staaten. Seit dem verheerenden Erdbeben in Haiti im Jahr 2010 sind die regionalen Unterschiede im Hinblick auf den Zugang zu Geodaten vermehrt in den Fokus gerückt. Seitdem schließt die OSM-Community, zu der auch zahlreiche Hilfsorganisationen wie das Rote Kreuz zählen, im Rahmen des sogenannten humanitären Mapping gezielt Lücken in den Geodaten in weniger entwickelten Regionen der Welt (Herfort und andere, 2020).

Ergebnis des Gazetteer Matchings ist eine Liste potenzieller geografischer Orte, die mit einem Ortsnamen (Toponym) gemeint sein könnten. Im Rahmen der Toponym Resolution (Ortsnamenauflösung) wird für diese Kandidaten eine Wahrscheinlichkeitsverteilung bestimmt (DeLozier und andere, 2015). Hierfür können die Häufigkeitsverteilungen in den vorhandenen Daten herangezogen werden. Dabei wird die Mehrdeutigkeit von Ortsangaben durch identifizierende Namenszusätze, wie der zusätzlichen Nennung des Bundesstaats oder eines Flusses, genutzt. Beispielsweise würde die Wahrscheinlichkeit, dass mit der Angabe „Frankfurt“ als Geburtsort „Frankfurt am Main“ gemeint ist, über die relative Häufigkeit der Angabe von „Frankfurt am Main“ geschätzt.

$$P(\text{Ort}=\text{Frankfurt am Main, DE} \mid \text{Name}=\text{Frankfurt, Staat}=\text{DE}) \\ = h(\text{Frankfurt am Main, DE} \mid \text{Staat}=\text{DE})$$

Die anschließende Geokodierung kann online über eine Programmierschnittstelle (englisch: Application Programming Interface – API) erfolgen. Ein wesentlicher Vorteil der Nutzung einer API ist, dass diese teilweise Funktionen zur automatisierten Vervollständigung (Autocomplete) anbieten, die automatisch kleinere Schreibfehler korrigieren und unterschiedliche Schreibweisen eines Orts erkennen. Damit ersetzt die Autocomplete-Funktion aufwendige nutzerseitige Vorverarbeitung (Preprocessing). Die Erprobung dieses Konzepts in einer Machbarkeitsstudie steht noch aus.

7


Fazit

In diesem Beitrag wurde die Aufbereitung der Merkmale Geburtsstaat und Geburtsort aus dem Ausländerzentralregister für deren Verwendung in der amtlichen Statistik vorgestellt. Angaben zum Geburtsort sind im Ausländerzentralregister fast vollständig erfasst, allerdings erfolgt die Erfassung unstrukturiert als Freitextangabe. Angaben zum Geburtsstaat hingegen werden standardisiert als ISO-Code erfasst, diese Angaben sind allerdings in erheblichem Umfang unvollständig.

In einem ersten Schritt hat das Statistische Bundesamt ein Konzept zur Aufbereitung der Angaben zum Geburtsstaat erarbeitet. In diesem Konzept spielt der im Statis-

tischen Bundesamt geführte Vergleichsdatenbestand an weltweiten Ortsnamen und zugewiesenen Staaten eine zentrale Rolle. Die Qualität der aufbereiteten Daten hängt maßgeblich von der Qualität dieser Leitdatei ab.

Die Ergebnisse der Aufbereitung der Daten des Ausländerzentralregisters zum Jahresende 2020 sind vielversprechend. Für im Ausland geborene ausländische Staatsangehörige konnte der Anteil an fehlenden (26,8%) oder nicht plausiblen (4,3%) Angaben zum Geburtsstaat von insgesamt 31,1% in den Rohdaten auf 6,0% in den aufbereiteten Daten verringert werden. Damit wurde die Datenqualität der Verwaltungsdaten für die statistische Nutzung wesentlich verbessert. Die gute Vergleichbarkeit mit Auswertungen zum Geburtsstaat aus dem Mikrozensus zeigt die Belastbarkeit der aufbereiteten Daten.

In einem zweiten Schritt könnten die Angaben zu den Geburtsorten geokodiert werden. Durch die Geokodierung kann die Herkunft einer Person besser lokal definiert werden, was neue Auswertungsmöglichkeiten auf kleinräumigerer Ebene eröffnet. Ein potenzielles Aufbereitungskonzept könnte auf den Geodaten von OpenStreetMap basieren. 

LITERATURVERZEICHNIS

Bundesamt für Migration und Flüchtlinge. *Migrationsbericht 2020*. [Zugriff am 1. März 2022]. Verfügbar unter: www.bamf.de

DeLozier, Grant/Baldrige, Jason/London, Loretta. *Gazetteer-independent toponym resolution using geographic word profiles*. In: Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015. Seite 2382 ff.

Feuerhake, Jörg/Lange, Kerstin/Siegismund, Annelen/Vigneau, Elsa. *Kodierung des Geburtsstaats in der Wanderungsstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 98 ff.

Fleischer, Henning. *Entwicklung der Ausländerzahl seit 1987*. In: Wirtschaft und Statistik. Ausgabe 9/1989, Seite 594 ff.

Goldberg, Daniel W./Wilson, John P./Knoblock, Craig A. *From text to geographic coordinates: the current state of geocoding*. In: URISA Journal. Jahrgang 19. Ausgabe 1/2007, Seite 33 ff.

Herfort, Benjamin/Lautenbach, Sven/Porto de Albuquerque, João/Anderson, Jennings/Zipf, Alexander. *The evolution of humanitarian mapping within the OpenStreetMap community*. In: Scientific Reports. Jahrgang 11. Artikel 3037/2021.

Leidner, Jochen L. *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. 2007.

Seto, Toshikazu/Kanasugi, Hiroshi/Nishimura, Yuichiro. *Quality verification of volunteered geographic information using osm notes data in a global context*. 2020. In: International Journal of Geo-Information. Jahrgang 9. Ausgabe 6/2020, Seite 372 ff.

Statistische Ämter des Bundes und der Länder. *Gemeindeverzeichnis-Online*. [Zugriff am 10. März 2022]. Verfügbar unter: www.statistikportal.de

Statistisches Bundesamt. *Digitale Agenda des Statistischen Bundesamtes. 03/2019*. [Zugriff am 8. März 2022]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Ausländerstatistik*. 2021a. [Zugriff am 1. März 2022]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Ausländische Bevölkerung*. 2021b. [Zugriff am 1. März 2022]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Staats- und Gebietssystematik*. [Zugriff am 1. März 2022]. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

Gesetz über das Ausländerzentralregister (AZR-Gesetz) vom 2. September 1994 (BGBl. I Seite 2265), das zuletzt durch Artikel 8 des Gesetzes vom 9. Juli 2021 (BGBl. I Seite 2467) geändert worden ist.

Staatsangehörigkeitsgesetz (StAG) in der im Bundesgesetzblatt Teil III, Gliederungsnummer 102-1, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 1 des Gesetzes vom 12. August 2021 (BGBl. I Seite 3538) geändert worden ist.

Zweites Gesetz zur Verbesserung der Registrierung und des Datenaustausches zu aufenthalts- und asylrechtlichen Zwecken (Zweites Datenaustauschverbesserungsgesetz – 2. DAVG) vom 4. August 2019 (BGBl. I Seite 1131).

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im April 2022
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-22002-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2022
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.