

Dipl.-Soziologin Claudia Weißer

# Auswertung der Geburtsorte im Zensus 2011

*Im Jahr 2011 wurde in Deutschland erstmals ein im Wesentlichen „registergestützter“ Zensus durchgeführt, mit dem für politische und wirtschaftliche Entscheidungen und Planungen verlässliche Basisdaten über die Bevölkerung ermittelt werden sollen. Hierzu werden Daten aus Melde- und anderen Verwaltungsregistern für den Zensus aufbereitet und ausgewertet. Den Angaben zum Geburtsort kommt eine besondere Bedeutung zu, da diese nicht nur dem internationalen Vergleich der Bevölkerung mit Migrationshintergrund dienen, sondern auch von den Statistischen Ämtern des Bundes und der Länder für die Aufbereitung und Durchführung des Zensus selbst genutzt werden. Um die Angaben zum Geburtsort aus den Melderegistern in eine für den Zensus nutzbare Form zu bringen, war der Aufbau eines Referenzverzeichnisses mit Geburtsort-Staat-Kombinationen notwendig. Im folgenden Beitrag werden die Anforderungen an dieses Verzeichnis sowie dessen Aufbau und Nutzung dargestellt.*

## 1 Ziele und Aufgaben

Die Europäische Union (EU) fordert im Rahmen der Verordnung (EG) Nr. 1201/2009<sup>1</sup> von den Mitgliedstaaten, bei den Volks- und Wohnungszählungen neben weiteren demografischen Daten auch die Angaben zu den Merkmalen „Geburtsort“ und „Geburtsstaat“ der Einwohnerinnen und Einwohner nach dem aktuellen, das heißt dem am 1. Januar 2011 geltenden Gebietsstand zu ermitteln. Ziel ist es dabei, die

Geburtsorte auf Staatenebene auszuweisen. Daraus abgeleitet können verschiedene Aggregationsstufen, beispielsweise nach EU/Nicht-EU oder nach Kontinenten, gebildet werden.

Im Zensusvorbereitungsgesetz 2011<sup>2</sup> ist geregelt, dass das Statistische Bundesamt ein vom Anschriften- und Gebäuderegister getrenntes Verzeichnis der Geburtsorte und Geburtsstaaten (Ortsverzeichnis) erstellt und führt. Dies wird von den Statistischen Ämtern des Bundes und der Länder für die Durchführung des Zensus genutzt, insbesondere um die Geburtsorte und deren Staatszugehörigkeit im Zensusdatensatz zu standardisieren und zu referenzieren.

Für den Aufbau dieses Ortsverzeichnisses stellen die in den Melderegistern gespeicherten Geburtsortsangaben die Datenbasis dar.<sup>3</sup> Die Referenzierung der Staatszugehörigkeit erfolgt nach zwei miteinander verzahnten Konzepten:

- › Um die Lieferverpflichtungen gegenüber der EU zu erfüllen, soll jeder Person in der Auswertungsdatenbank der ihr zugehörige (Geburts-)Staat auf der Grundlage der am 1. Januar 2011 bestehenden Grenzen zugewiesen sein. Die Ausweisung der Staatsangehörigkeit erfolgt nach dem vom Statistischen Bundesamt herausgegebenen „Staatsangehörigkeits- und Gebietschlüssel“. Staatschlüssel für nicht mehr existierende Staaten, wie beispielsweise Jugoslawien, werden in die Schlüssel der Nachfolgestaaten transferiert. Die Verschlüsselung

<sup>1</sup> Verordnung (EG) Nr. 1201/2009 der Kommission vom 30. November 2009 zur Durchführung der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates über Volks- und Wohnungszählungen in Bezug auf die technischen Spezifikationen für die Themen sowie für deren Untergliederung (Amtsblatt der EU Nr. L 329, Seite 29); in dieser Verordnung werden unter anderem die Spezifikationen für die Ausweisung des Geburtsstaates festgelegt.

<sup>2</sup> Gesetz zur Vorbereitung eines registergestützten Zensus einschließlich einer Gebäude- und Wohnungszählung 2011 (Zensusvorbereitungsgesetz 2011 – ZensusVorbG 2011) vom 8. Dezember 2007 (BGBl. I Seite 2808), geändert durch Artikel 3 des Gesetzes vom 8. Juli 2009 (BGBl. I Seite 1781).

<sup>3</sup> § 3 Zensusvorbereitungsgesetz 2011.

der von der EU geforderten Gliederungsebenen erfolgt im Zuge der Typisierung im Rahmen der Auswertung. Die Gliederungssystematik für das von der EU geforderte Pflichtmerkmal „Geburtsland/-ort“ ist in der EU-Verordnung festgelegt.<sup>4</sup>

- › Im Sinne des nationalen Auswertungskonzeptes sollen Personen, die bis zum 2. August 1945<sup>5</sup> 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 nach heutiger Staatsaufteilung im Ausland liegen.<sup>6</sup> Dies betrifft ehemalige deutsche Ostgebiete wie zum Beispiel Ostpreußen oder Schlesien. Die betreffenden Gemeinden in den ehemals deutschen Ostgebieten werden innerhalb des Ortsverzeichnisses gesondert ausgewiesen. Auf Personenebene wird der Geburtsstaat abhängig vom Geburtsdatum des Einwohners beziehungsweise der Einwohnerin bestimmt.

Neben diesen Konzepten zur Referenzierung der Geburtsstaaten soll mithilfe des unveränderlichen Merkmals „Geburtsort“ die Dublettensuche bei der Bereinigung der Melderegister unterstützt werden.<sup>7</sup> Hieraus ergibt sich eine weitere Anforderung an das Ortsverzeichnis:

- › Das Ortsverzeichnis dient dazu, die unterschiedlichen Einträge zum Geburtsort aus den Melderegistern vergleichbar zu machen. Ziel war es, die unterschiedlichen Schreibweisen des gleichen Ortes zu einer korrekten standardisierten Schreibweise zusammenzuführen. Zudem muss sichergestellt werden, dass Orte mit überholten Bezeichnungen (zum Beispiel früher selbstständige Gemeinden, die in einer anderen Gemeinde aufgegangen sind) unter ihrem neuen Namen gefunden werden. Bei der Standardisierung der Namen ausländischer Orte sind zudem die verschiedenen Transkriptionsweisen zu berücksichtigen.

<sup>4</sup> Gliederungsmerkmal "Place of Birth" (POB) nach der Verordnung (EG) Nr. 1201/2009 (siehe Fußnote 1).

<sup>5</sup> Das Datum 2. August 1945 stellt auf die Beschlüsse der Potsdamer Konferenz ab.

<sup>6</sup> Siehe Artikel 116 Grundgesetz.

<sup>7</sup> Bei der Mehrfachfallprüfung wird der Registerbestand auf Dubletten hin untersucht. Diese Untersuchung basiert auf der grundsätzlichen Annahme, dass es Personenmerkmale gibt, die eindeutig der gleichen Person zugewiesen werden können. Aus dem Datenkranz der Melderegister kann hierfür unter anderem das Merkmal Geburtsort herangezogen werden.

Übersicht 1

Angaben in den Melderegistern

Die nach Landesrecht für das Meldewesen zuständigen Stellen (Meldebehörden) haben die in ihrem Zuständigkeitsbereich wohnhaften Einwohnerinnen und Einwohner zu registrieren, um deren Identität und Wohnsitz feststellen und nachweisen zu können. Die Speicherung der gewonnenen Daten erfolgt bundesweit einheitlich auf der Grundlage des im Datensatz für das Meldewesen (DSMeld) hinterlegten Formats [Grundlage bildet das Melderechtsrahmengesetz – MRRG in der Fassung der Bekanntmachung vom 19. April 2002 (BGBl. I Seite 1342), zuletzt geändert durch Artikel 9 des Gesetzes vom 28. April 2011 (BGBl. I Seite 678)].

Darüber hinaus können die Bundesländer noch ergänzende Eingabefelder definieren. Die im Datensatz für das Meldewesen vorgesehenen Datenfelder müssen von den Meldebehörden nur in dem Umfang, wie dies zur Aufgabenerfüllung erforderlich ist, gespeichert werden. Die mindestens zu speichernden Daten ergeben sich aus dem Meldeschein.

2 Verwendete Datenquellen

2.1 Melderegisterdaten

Die Angaben in den Melderegistern (siehe Übersicht 1) bilden die Basis des Ortsverzeichnisses. Diese Angaben sind durch die Meldepflichtigen zu belegen, in der Regel durch amtliche Unterlagen wie Geburtsurkunde, Personalausweis, Reisepass und andere geeignete Dokumente. Im Unterschied zum aktuellen Wohnort können die Angaben zum Geburtsort auch historischen Charakter haben, wenn beispielsweise die aus der Geburtsurkunde übernommenen Angaben nicht mehr dem aktuellen Gebietsstand entsprechen. Für den Aufbau des Ortsverzeichnisses bildeten die Merkmale aus der Datenlieferung der Melderegister zum 1. April 2008 die Basis. Gemäß § 3 Zensusvorbereitungsgesetz 2011 lieferten die Meldebehörden hierzu die in der Übersicht 2 aufgeführten Angaben.

Im Feld „Geburtsort“ ist der Geburtsort des beziehungsweise der Meldepflichtigen als Klarschriftangabe vermerkt. Das Feld „Geburtsort – Staat“ enthält hingegen nur Angaben zum Geburtsstaat, wenn die Person im Ausland geboren wurde. In diesem Fall wird der 3-stellige Staatsangehö-

Übersicht 2

Für das Ortsverzeichnis relevante Felder aus dem Datensatz für das Meldewesen

Blatt Datensatz für das Meldewesen	Feldbezeichnung	Format des Feldinhaltes	Feldlänge	Bemerkungen
0602	Geburtsort	Klarschriftangabe	40	–
0603	Geburtsort – Staat	Schlüssel	3	nur bei im Ausland geborenen Personen
0604	Geburtsort – Standesamt	Klarschriftangabe	40	–
1223	Zuzug aus dem Ausland – Staat	Schlüssel	3	bei Zuzug aus dem Ausland ist der Zuzugsstaat anzugeben

rigkeitsschlüssel (Staatenschlüssel) der amtlichen Statistik eingetragen. Für Personen, die in Deutschland geboren wurden, soll dieses Feld leer bleiben. Das Feld „Geburtsort – Standesamt“ enthält als Klarschriftangabe das Standesamt, an dem die Geburt beurkundet wurde. Bei Zuzug aus dem Ausland wird im Feld „Zuzug aus dem Ausland – Staat“ der Staat, in dem die Person bisher gewohnt hat, in Form des Staatsangehörigkeitsschlüssels eingetragen.

Die in Übersicht 2 dargestellten Merkmale wurden in einer eigenständigen Datei „Ortsverzeichnis“ gespeichert und hatten keinen Bezug mehr zu anderen Registermerkmalen der Personen.

## 2.2 Referenzverzeichnisse

Bei der Erstellung des Ortsverzeichnisses dienten Referenzverzeichnisse dazu, die Zuweisung der Staatszugehörigkeit zu den Geburtsorten im Sinne der Konzepte zu gewährleisten und Geburtsortsangaben zu standardisieren und korrekt auszuweisen.

### Staatenverzeichnis des Statistischen Bundesamtes

Zur Kodierung des Felds „Geburtsort – Staat“ im Melderegister wurde das vom Statistischen Bundesamt herausgegebene Verzeichnis der Staatsangehörigkeits- und Gebietschlüssel verwendet. Mithilfe einer aus diesem Verzeichnis erstellten Referenztabelle, dem Staatenverzeichnis, wurde beim Erstellen des Ortsverzeichnisses die Geburtsort-Staat-Kombination aufbereitet. Diese Referenztabelle enthielt demnach aktuelle und historische amtliche Staatsangehörigkeitsschlüssel und den ausgeschriebenen Staatsnamen in Klarschriftangabe einschließlich abhängiger Gebiete.

### Ortsnamensverzeichnis aus dem Anschriften- und Gebäuderegister

Um eine einheitliche Darstellung gleicher Orte trotz unterschiedlicher Einträge im Ausgangsdatenfeld „Geburtsort“ zu erreichen, lehnte sich die Form der Darstellung deutscher Orte an diejenige aus dem Anschriften- und Gebäuderegister<sup>8</sup> an. Eine hierfür generierte Referenztabelle aus dem Anschriften- und Gebäuderegister (AGR) enthielt für jeden Ort den standardisierten Ortsnamensstamm und bei Ortsnamen, die mehrfach in Deutschland vorkommen, deren Ortsnamenszusätze [zum Beispiel den Ortsnamensstamm FRANKFURT und die Ortsnamenszusätze „am Main“ und „(Oder)“].

### Ortschaftsverzeichnis des Bundesamtes für Kartographie und Geodäsie

Um dem nationalen Auswertungskonzept Rechnung zu tragen, waren Orte, die zu Deutschland nach dem Gebietsstand vom 31. Dezember 1937 gehörten, im Ortsverzeichnis zu kennzeichnen. Dies setzte die Erstellung eines „historischen Ortsverzeichnisses“ voraus, dessen Grundlage die Daten aus dem „Ortschaftsverzeichnis des Bundesamtes

für Kartographie und Geodäsie“ bildeten. Dieses Ortschaftsverzeichnis enthält alle ehemals zu Deutschland gehörenden Gebiete in ihrer Chronologie von 1900 bis 1993 und alle verwendeten Ortschaftsnamen, und zwar in den verschiedenen Schreibweisen, die aus wechselnden Staatszugehörigkeiten folgten. Mit dem historischen Ortschaftsverzeichnis können für einzelne Ortschaften Veränderungen bezüglich des Namens aufgrund von Veränderungen in der Verwaltungszuordnung über die Zeit abgebildet werden.

### Adressdateien der Deutschen Post AG

Zur Unterstützung ihrer Tätigkeit im Post- und Frachtpostbereich sowie für externe Dienstleistungen unterhält die Deutsche Post AG eine umfangreiche Datenbank mit Orts- und Adressangaben. Bei der Anlage der Datenbank hat die Deutsche Post AG immer das Ziel verfolgt, eine Adresse oder einen Ort eindeutig zu bezeichnen, auch wenn dieser mehrfach vorkommt. In der Datenbank werden dabei neben den aktuellen Bezeichnungen auch historische Angaben abgelegt und entsprechend bis auf Adressebene verlinkt.<sup>9</sup> Für den Zensus 2011 wurden nur die Dateien mit direktem Bezug zu aktuellen beziehungsweise historischen Ortsnamen und zu Ortsteilen betrachtet. Mit den Datensätzen der Ortsarchivdatei ist es möglich, die Veränderung der Ortsnamen, sei es durch Ortsumbenennungen oder durch Eingemeindungen, im Zeitraum zwischen 1970 und 2011 festzustellen.

## 3 Inhalte und Aufbereitung der originalen und der zu generierenden Felder des Ortsverzeichnisses

Zur Klassifizierung der Geburtsorte im Zensus 2011 wurden jedem Datensatz nach den oben aufgeführten Anforderungen ein referenzierter Geburtsort und ein referenzierter Geburtsstaat nach europäischem und nationalem Auswertungskonzept zugewiesen. Hierfür wurden die Felder des Meldesatzes genutzt. Die Zuweisung dieser referenzierten Variablen erfolgte innerhalb des Ortsverzeichnisses und wurde im Folgenden in den Referenzdatenbestand des Zensus 2011 übernommen. Der Referenzdatenbestand setzte sich aus den Datenlieferungen der Melderegister und der im Zensus genutzten Erwerbsregister zusammen und bildete die Grundlage für die Erstellung des Zensusdatensatzes.

### 3.1 Geburtsort

Im Ausgangsfeld „Geburtsort“ befanden sich eine Reihe unterschiedlicher Einträge, auch wenn sich die Datensätze auf den gleichen Geburtsort beziehen. Dies kann folgende Gründe haben: Bei Personen, die nicht mobil waren, also nicht über Gemeindegrenzen hinweg verzogen sind, bleibt der Geburtsortseintrag so bestehen, wie er eingetragen wurde; Eingemeindungen werden hier nicht nachvollzogen. Die Mobilität von Personen kann hingegen zu Neueinträgen des Geburtsortes im Ausgangsdatenfeld „Geburtsort“ nach dem dann aktuellen Gebietsstand führen. Durch Neuansmeldungen von Personen bei den Meldebehörden kann auch

<sup>8</sup> Näheres zum Aufbau des Anschriften- und Gebäuderegisters siehe Kleber, B./Maldonado, A./Scheuregger, D./Ziprik, K.: „Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011“ in WiSta 7/2009, Seite 629 ff.

<sup>9</sup> Die Daten werden regelmäßig zu Beginn eines neuen Quartals aktualisiert.

die Darstellung des gleichen Geburtsortes stark variieren. So können sich durch unterschiedliche Verwendung von Groß- und Kleinschreibung, diakritischen Zeichen<sup>10</sup>, durch orthografische Fehler oder Sonderzeichen Unterschiede ergeben, obwohl der gleiche Geburtsort dargestellt wird. Da es für die dezentrale Struktur des Meldesystems keine bundesweit einheitlichen Vorgaben gibt, variiert zudem die Form von Ortsnamensbestandteilen (ausgeschriebene/verkürzte Form) und von Ortsnamenszusätzen. Bei Geburtsortseinträgen von Personen, die im Ausland geboren wurden, treten zudem unterschiedliche Schreibweisen und Transkriptionen des gleichen Geburtsortes auf. Ummeldungen können aber auch zu Informationsverlusten führen. So können zum Beispiel wichtige Informationen im Ausgangsdatenfeld „Geburtsort – Standesamt“ bei einer Neuanschreibung entfallen. Da Angaben zum Geburtsort auch historischen Charakter besitzen können und einmal erfasste Geburtsorte und Staaten in den Registern der Meldebehörden in der Regel nicht evident gehalten werden, waren auch veraltete Ortsnamen und Staatenschlüssel sowie Hinweise auf Eingemeindungen, aktuelle und historische Regionalbezüge zu finden. In vielen Fällen befanden sich im Ausgangsdatenfeld „Geburtsort“ neben dem eigentlichen Ortsnamen auch zusätzliche Informationen und Erläuterungen, welche meist die Lage des Ortes genauer beschrieben. Hierzu zählen Kreis- und Gebietsangaben, der Geburtsstaat in Klarschrift und Zusatzinformationen zur Lagebestimmung wie „bei-Angaben“, das heißt die Ausweisung des nächstgelegenen Ortes oder weiterer geografischer Punkte. Weitere Eintragungen wiesen keinen Geburtsort aus, sondern nur höhere administrative Einheiten, oder ließen auf keinen Ortsnamen schließen.

Bei der Erstellung des Ortsverzeichnisses musste diesen Gegebenheiten Rechnung getragen sowie die Variation der Angaben vereinheitlicht, standardisiert und in einen referenzierten Geburtsort überführt werden. Es wurde ein Konzept entwickelt, das die Ausgangsdaten auf den aktuellen Gebietsstand der Gemeinde fokussierte und den Umgang mit Namensvariationen regelte.

### 3.2 Geburtsstaat

Der Geburtsstaat wurde in zwei Referenzvariablen ausgewiesen: Zum einen als referenzierter Geburtsstaat nach aktuellen Gebietsgrenzen, zum anderen wurde – wenn es sich um eine bis zum 2. August 1945 in den Grenzen Deutschlands von 1937 geborene Person handelte – ergänzend der deutsche Staatenschlüssel verwendet.

Grundlage für die Ausweisung aller gültigen Staatenschlüssel nach aktuellen Gebietsgrenzen bildete das Ausgangsdatenfeld „Geburtsort – Staat“. Die Meldebehörden befüllten das Feld auf der Basis der im Datensatz für das Meldewesen vorgegebenen Regelungen. In diesem Feld befanden sich auch inkorrekte Befüllungen. So gab es eine Reihe ausländischer Orte, denen kein oder ein undefinierter Staatenschlüssel zugewiesen war. Neben Staatenschlüsseln gab es auch Klarschriftangaben, deren Ausprägungen oft Rückschlüsse

auf den Staat zuließen, beispielsweise „POL“ für Polen. In vielen Fällen wurde der Ersatzschlüssel für „ohne Angabe“ („999“) eingetragen, obwohl es sich um einen deutschen Ort handelte.<sup>11</sup> Auch wurde häufig der deutsche Staatenschlüssel eingetragen, obwohl der Ort nicht in Deutschland lag. Zudem kamen veraltete Staatenschlüssel existierender Staaten und Schlüssel von ehemaligen Staaten wie der Tschechoslowakei vor. Bei ehemaligen Staaten blieben die veralteten Staatenschlüssel nur erhalten, wenn keiner ihrer Folgestaaten eindeutig zugeordnet werden konnte. Neben veralteten und nicht korrekt eingetragenen gültigen Staatenschlüsseln wurden Schlüssel gefunden, die weder zu einem Ordnungssystem gehörten noch kontinental zugeordnet werden konnten. Wenn kein Staat zugewiesen werden konnte, wurde der Schlüssel für „ohne Angabe“ gesetzt.

## 4 Aufbau des Ortsverzeichnisses

Die Ausweisung der Referenzvariablen basiert auf den Angaben zum Geburtsort in den Ausgangsdatenfeldern des Datensatzes für das Meldewesen. Um gleichgelagerte Arbeiten nicht mehrfach ausführen zu müssen, wurden die relevanten Daten aus der Ursprungsdatei aggregiert, gleiche Feldinhalte zusammengefasst und auf eine einheitliche Form standardisiert. Dies erfolgte im prozessualen Aufbau des Ortsverzeichnisses und unterteilte sich in die Phasen:

1. Generierung: Übertragen und Standardisieren der Daten aus den Melderegistern
2. Konsolidierung: Überprüfen und Bereinigen der Daten
3. Aufbereitung: Referenzierung von Geburtsort und Geburtsstaat

Der Aufbau des Ortsverzeichnisses wurde als iterativer Prozess angegangen, das heißt die bei der Aufbereitung gewonnenen Informationen flossen wiederum in die Aufbereitung noch zu bearbeitender Ausgangsdaten ein. Schaubild 1 veranschaulicht den Ablauf des Aufbaues und zeigt auf, welche Anwendungsmöglichkeiten das Ortsverzeichnis bietet.

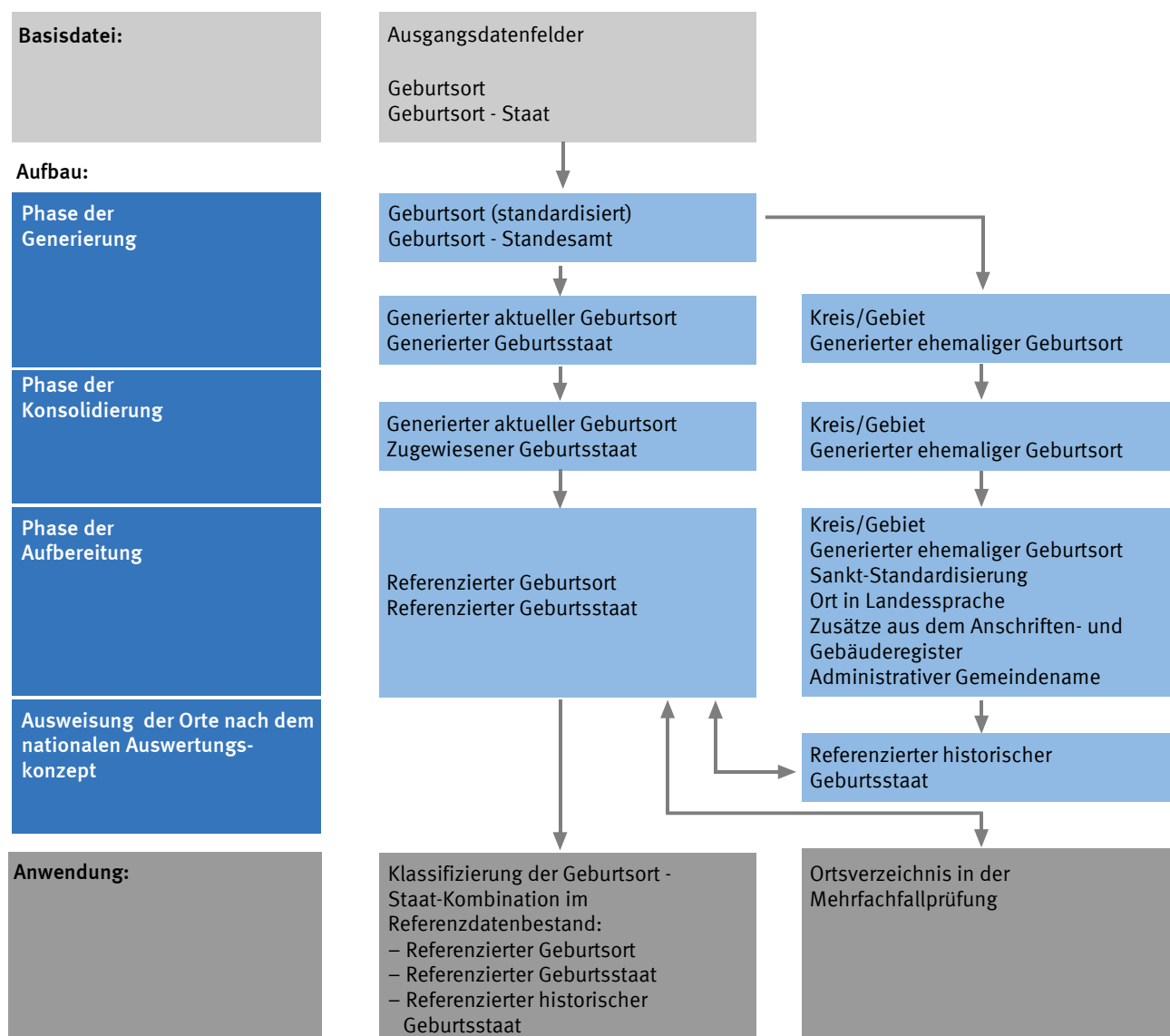
### 4.1 Phase der Generierung

In der ersten Phase wurden ein generierter Geburtsort und ein generierter Geburtsstaat aus dem vorliegenden Melderegistermaterial erstellt. Die erste Zuordnung von Staatenschlüsseln erfolgte aufgrund der Informationen im Ausgangsdatenfeld „Geburtsort – Staat“ unmittelbar über die Kodierung von Klarschriftangaben. Alle Zusatzinformationen aus den Ausgangsdatenfeldern, die für spätere Prozessschritte dienlich sein konnten, wurden in eigens dafür erstellten Feldern abgelegt. Zum Schutz der Originalangaben wurde der Inhalt des Ausgangsdatenfeldes „Geburtsort“ gedoppelt und standardisiert.

<sup>10</sup> Das sind grafische Zeichen, die in Verbindung mit bestimmten Buchstaben als Hinweis auf deren richtige Aussprache dienen (zum Beispiel Umlautpunkte).

<sup>11</sup> Die Ersatzschlüssel der Staatsangehörigkeit definieren die Angaben „staatenlos“ („997“), „ungeklärt“ („998“) und „ohne Angabe“ („999“).

Schaubild 1 Prozessualer Aufbau des Ortsverzeichnisses



2012 - 01 - 0840

#### 4.1.1 Generierung des Geburtsortes aus den Ausgangsdaten

Der referenzierte Geburtsort sollte der deutschsprachigen Ortsbezeichnung entsprechen. Bei ausländischen Orten blieben alle Ortsnamenszusätze erhalten, bei deutschen Orten wurden diese in Extraspalten gesondert abgelegt. In den Ausgangsdaten für den „Geburtsort“ befanden sich zudem Eintragungen, die keinen Geburtsort auswiesen. Höhere administrative Einheiten, wie Kreis- oder Gebietsangaben, wurden – um einen Informationsverlust zu verhindern – gesondert aufbereitet. Zudem wurden Eintragungen im Ausgangsdatenfeld, die auf keinen Ortsnamen schließen ließen, als nicht identifizierbare Orte definiert.

Um jedem Geburtsortseintrag einen referenzierten Geburtsort zuweisen zu können, musste der Geburtsort zunächst aus den Einträgen im Ausgangsdatenfeld „Geburtsort“ maschinell generiert werden. Hierzu wurden diese auf den

Stammnamen reduziert, das heißt der Feldinhalt wurde ohne Ortsnamenszusätze<sup>12</sup> als „generierter aktueller Geburtsort“ ausgewiesen. Zusätze wie „Hansestadt“ oder „Ostseebad“ wurden entfernt, hingegen blieben Bestandteile des Ortsnamens wie „Bad“ oder „Sankt“ erhalten.

Generell wurden alle Staatennamen aus der Ortsbezeichnung unter Zuhilfenahme des Staatenverzeichnisses entfernt.<sup>13</sup> Da das Staatenverzeichnis nur ausgeschriebene Staatennamen enthielt, wurde es um eine Liste erweitert, die neben gängigen Abkürzungen auch solche umfasste, die auf den gesammelten Erfahrungen bei der bisherigen Bearbeitung von Staats-, Gebiets- und Bundeslandsangaben basierten. So konnten auch Abkürzungen wie zum

<sup>12</sup> So wurde aus „Frankfurt am Main“ und „Frankfurt (Oder)“ nur „FRANKFURT“.

<sup>13</sup> Das Staatenverzeichnis enthielt zunächst alle ausgeschriebenen Staatennamen in standardisierter Form und deren abhängige Gebiete mit ihren Staatenschlüsseln, basierend auf der vom Statistischen Bundesamt herausgegebenen Liste zum Staatsangehörigkeits- und Gebietschlüssel von 2010.



Beispiel „RUSS. FOED“ oder solche, die Schreibfehler aufwiesen, berücksichtigt werden. Durch Schlüsselwörter wie „jetzt“ und „ehemals“ in dem Ausgangsdatenfeld „Geburtsort“ waren Eingemeindungen, Ortsumbenennungen oder Ähnliches erkennbar. Diese Früher-Jetzt-Bezeichnungen erlaubten eine Generierung des aktuellen und ehemaligen Geburtsortes, wobei die Darstellung dieser Schlüsselwörter stark variierte. Alle Bezeichnungen im Feld „Geburtsort“, die von der administrativen Einheit der Gemeinde abwichen, wie „Kreis“ und Varianten davon, wurden für den zu generierenden Geburtsort ebenso wie deren nachfolgende Textbestandteile abgetrennt.<sup>14</sup>

Da Informationen zu Kreisen und Gebieten für spätere Aufbereitungsschritte genutzt werden sollten, wurden sie in die Extraspalte „Kreis/Gebiet“ übernommen. Neben solchen, durch Schlagwörter zu identifizierenden Lageinformationen fanden sich auch häufig Bundesländerkürzel, Regionen oder Landstriche und Varianten davon in den Melderegistereinträgen.<sup>15</sup>

Zudem erfolgte bereits in der Phase der Generierung die erste Stufe der Standardisierung, das heißt das Großschreiben aller Buchstaben, das Ersetzen aller Umlaute durch Umschreibung mit „E“ („Ä“ wurde zu „AE“, „ß“ zu „SS“ und so weiter) und das Entfernen aller weiteren diakritischen Zeichen.

In einem weiteren Prozessschritt wurden alle nach dem letzten Wortteil (String) des Feldinhaltes stehenden Sonderzeichen sowie nachfolgende Textbestandteile entfernt, da davon ausgegangen wurde, dass nach diesem nur Zusatzinformationen stehen.

Im letzten Prozessschritt der Phase der Generierung erfolgte eine Abtrennung der generierten Geburtsorte ab der sechsten Stelle, sofern sich bis dahin ein Sonderzeichen oder ein Leerzeichen befand. In diesen Fällen wurde davon ausgegangen, dass vor dem Sonder- oder Leerzeichen der Ortsnamensstamm endet und dahinter nur zusätzliche Informationen angegeben sind.

#### 4.1.2 Erstellen einer Geburtsort – Staat-Kombination aus den Ausgangsdaten

Im Ausgangsdatenfeld „Geburtsort – Staat“ tragen die Meldebehörden den Staatenschlüssel des Geburtsortes ein. Nicht korrekte Einträge in diesem Feld, zum Beispiel fehlende Staatenschlüssel bei ausländischen Orten, falsche Staatenschlüssel, aber auch veraltete Staatenschlüssel, wurden durch Staateninformationen, die als Klarschriftangabe im Ausgangsdatenfeld „Geburtsort“ vorlagen, ergänzt oder gegebenenfalls korrigiert. Für den Abgleich der Klarschriftangabe wurde das erstellte und erweiterte Staatenverzeichnis genutzt. Mit dessen Hilfe konnte geprüft werden, ob im standardisierten Melderegisterfeld Zusatzinformationen vorhanden waren, welche die Zuordnung zu einem Staat ermöglichten. War dies der Fall, wurde für jeden Datensatz

der Staatenname und der zugehörige Staatenschlüssel in gesonderten Feldern ausgewiesen. Eine Überprüfung mit dem Feld „Geburtsort – Staat“ erfolgte erst in der Phase der Konsolidierung.

## 4.2 Phase der Konsolidierung

In der Phase der Konsolidierung erfolgten die Qualitätssicherung und das Erstellen eines konsolidierten Bestandes. Um später einen referenzierten Geburtsort und Geburtsstaat zu erhalten, mussten hierzu der bisher generierte aktuelle Geburtsort und -staat auf korrekte Ausweisung überprüft werden. Diese Überprüfung umfasste eine Kontrolle und Nachbereitung der internen Schritte der Generierung und der damit verbundenen angewandten maschinellen Generierungsroutinen.

### 4.2.1 Konsolidierung des generierten Geburtsortes

Es musste sichergestellt werden, dass der Stamm des Ortsnamens aus dem Melderegister korrekt in den generierten aktuellen Geburtsort übernommen wurde. Diese Qualitätssicherung geschah systematisch durch Plausibilitätskontrollen und diente der Kontrolle des zuvor für die Routinen gewählten Ablaufkonzepts und gegebenenfalls der Korrektur der entstandenen Felder.

So sollte beispielsweise der generierte ehemalige Geburtsort mit dem generierten aktuellen Geburtsort identisch sein, solange kein Hinweis auf „jetzt“ oder „früher“ oder andere Zeitbezüge im Ausgangsdatenfeld „Geburtsort“ vorlag.

Des Weiteren wurden alle Datensätze dahingehend überprüft, ob die im Zuge der Abtrennung genutzten Worte und Wortteile (Strings) wie zum Beispiel Staatskürzel auch Bestandteile des Namensstammes sind. Hierdurch wurden fehlerhafte Abtrennungen vermieden. Zudem erfolgte eine Überprüfung der generierten aktuellen und ehemaligen Geburtsorte auf Sonderzeichen wie Bindestriche. In einem aufwendigen Verfahren wurde zudem geprüft, ob es sich bei dem Stammmamen eines Ortes tatsächlich um die Zuweisung des Hauptteils der Gemeinde handelt. Da Eingemeindungen in der Regel nicht durch die Meldebehörden im Register nachgepflegt werden, konnten auch nur der ehemalige Ortsname oder Ortsteile als alleiniger String oder vorangestellter Teilstring im Ausgangsdatenfeld „Geburtsort“ stehen und somit in den generierten aktuellen Geburtsort übertragen werden. Durch verzeichnisinterne Abgleiche und mithilfe einer aus den Dateien der Deutschen Post AG erstellten Vergleichstabelle konnten über Stringabgleiche diese Orte identifiziert und der aktuelle Gemeindegemeinde zugeordnet werden.

### 4.2.2 Konsolidierung des generierten Geburtsstaates

Die in der ersten Phase aus den Klarschriftangaben im Ausgangsdatenfeld „Geburtsort“ erstellten Staatenschlüssel der ausgeschriebenen oder abgekürzten Staatenangaben wurden geprüft.<sup>16</sup> Dies erfolgte über einen Abgleich der

<sup>14</sup> Dies betraf Gebiete, Orts-, Gemeinde- und Stadtteile, Provinzen, Stadt- und Regierungsbezirke, Kantone, Landkreise sowie Bundesrepublik und alle Varianten davon.

<sup>15</sup> Beispiele hierfür sind Abkürzungen wie „ERZ“, welche in diesem Falle für Erzgebirge steht, oder „Schlesien“ und Varianten davon.

<sup>16</sup> Kritisch waren vor allem Staatenangaben in verkürzter Form wie „Niederl.“, das gleichermaßen für Niederlande als auch für Niederlausitz stehen kann.

Staatenschlüssel aus dem Ausgangsdatenfeld „Geburtsort – Staat“ mit denen, die aus diesen Zusatzinformationen gewonnen wurden. Sich gegenseitig einschließende Staatenschlüssel (zum Beispiel Tschechische Republik und Tschechoslowakei) fielen nicht in die Bearbeitungsmasse. Der Klarschriftangabe im Ausgangsdatenfeld „Geburtsort“ wurde in der Regel Vorrang vor dem im Ausgangsdatenfeld „Geburtsort – Staat“ stehenden Staatenschlüssel eingeräumt, da erstere meist dem aktuelleren Gebietsstand entsprach.<sup>17</sup> Die so qualitätsgesicherten Staatenschlüssel wurden in das Feld des generierten Geburtsstaates übernommen. Ebenfalls in dieses Feld übernommen wurden Staatenschlüssel, die aus der Klarschriftangabe resultierten und keinen Staatenschlüssel im Ausgangsdatenfeld „Geburtsort – Staat“ besaßen, sowie Staatenschlüssel aus dem Ausgangsdatenfeld „Geburtsort – Staat“, zu denen keine Klarschriftangabe vorlag.

Nach diesen bisherigen Ausweisungen stand noch eine Reihe nicht korrekter Geburtsort – Staat-Kombinationen im Ortsverzeichnis. Beispielsweise war der deutsche Geburtsort Koblenz im Melderegister in 99,74 % der Fälle mit dem Geburtsstaat Deutschland ausgewiesen, in 0,2 % beziehungsweise 0,06 % der Fälle jedoch mit den Staaten Türkei beziehungsweise Italien. Grund waren nicht korrekt eingetragene Staatenschlüssel im Ausgangsdatenfeld „Geburtsort – Staat“, die nicht mithilfe von Klarschriftangaben korrigiert werden konnten. Um eine höhere Qualität der Geburtsort – Staat-Kombination zu erzielen, wurden in diesen Fällen die Angaben zum Geburtsstaat referenziert. Dies erfolgte über die Festlegung eines Hauptstaates für jeden Ort. Zur Bestimmung des Hauptstaates wurde in einem ersten Schritt ermittelt, wie häufig ein Geburtsort mit den jeweiligen Staatsangaben vorlag und es wurde die entsprechende dahinter stehende Personenanzahl ausgezählt.<sup>18</sup> Dies wurde in Bezug zur Häufigkeit des Vorkommens des Geburtsortes beziehungsweise zur Gesamtzahl der Personen je Geburtsort gesetzt. Die häufigste Ausprägung legte den Hauptstaat für den Ort fest. Geburtsorten, deren Personen- und Ortsanteile die definierten Schwellenwerte unterschritten, wurde der jeweilige Hauptstaat zugewiesen. Die Schwellenwerte waren hierbei davon abhängig, wie häufig der Ort vorkam und ob es sich um einen ausländischen oder deutschen Hauptstaat handelte. Bei der Zuweisung der Staaten musste sehr vorsichtig vorgegangen werden, da Orte gleichen Namens auch in unterschiedlichen Staaten zu finden sind. Zudem lag hier die Prämisse zugrunde, den Eintragungen zum Geburtsstaat im Ausgangsmaterial der Meldebehörde eine höhere Präferenz einzuräumen als der Zuweisung eines Staates aufgrund der Ortsangabe.

Bei Orten, die noch mit „ohne Angabe“ (Staatenschlüssel „999“) im Verzeichnis standen, wurde der Staatenschlüssel „999“ nur dann auf den Hauptstaat umgesetzt, wenn

<sup>17</sup> Dies betraf Fälle, in denen ein Staatename im Originalmelderegisterfeld „Geburtsort“ stand und im Feld „Geburtsort – Staat“ „ohne Angabe“, oder solche, bei denen im Feld „Geburtsort – Staat“ der deutsche Staatenschlüssel stand und die Klarschriftangabe einen ausländischen Staat auswies, wie zum Beispiel „Ankara Türkei“. Darüber hinaus waren vor allem Orte aus ehemaligen Ostgebieten betroffen und Fälle, bei denen der Folgestaat nur als Klarschriftangabe vorlag.

<sup>18</sup> Da die Angaben des Ortsverzeichnisses in aggregierter Form vorlagen, wurde für die Auszählung der Personenzahl auf die Einzeldaten (Fallzahlen der Melderegisterdaten zum 1. April 2008) zurückgegriffen.

die Verteilung der Personenanteile dessen Bestimmung eindeutig zuließ und nochmals anhand von Klarschriftangaben qualitätsgesichert wurde.<sup>19</sup> Bei Gebietsangaben wie Ostpreußen, dessen Orte nach heutigen Staatsgrenzen sowohl zu Polen als auch zur Russischen Föderation zugeordnet werden könnten, wurde – um Fehlzuordnungen zu vermeiden – eine Liste von Orten erstellt, die schon eine laut Staatenschlüssel eindeutige Zuweisung zu Polen oder zur Russischen Föderation hatten, und über diese Liste der Staatenschlüssel zugewiesen.

Nach diesen Aufbereitungsschritten wird im Folgenden immer von einem zugewiesenen Geburtsstaat gesprochen. Alle weiteren Aufbereitungen basieren auf diesem zugewiesenen Geburtsstaat.

#### 4.2.3 Erarbeitung eines geeigneten Bestandes

Nach diesem Konsolidierungsschritt gab es in der Phase des Ortsverzeichnisaufbaues immer noch Geburtsorte, denen kein Staatenschlüssel zugeordnet werden konnte. Ein fehlender Staatenschlüssel im Feld „Geburtsort – Staat“ konnte nicht mit dem Geburtsland „Deutschland“ gleichgesetzt werden, diente jedoch als Indiz hierfür. Um einen gesicherten Bestand zu erhalten, in dem deutsche Orte als solche identifiziert wurden, erfolgte ein Abgleich des bisher generierten aktuellen Geburtsortes mit dem Verzeichnis der Deutschen Post AG. Konnte ein Ort durch dieses Verzeichnis bestätigt werden, erhielt er den deutschen Staatenschlüssel.

#### 4.2.4 Vervollständigung des konsolidierten Bestandes

Durch die bisherigen Prozessschritte konnte ein als konsolidiert geltender Bestand erstellt werden, der alle Datensätze beinhaltet, die eine Zuweisung des Geburtsstaates erhalten hatten. All diejenigen Orte, auf die dies nicht zutraf, unterlagen einer manuellen Überarbeitung, die vor allem Recherchen zu den Staaten im Internet unter Berücksichtigung von Zusatzinformationen umfasste. Darauf folgte eine Zusammenführung beider Bestände zu einem Gesamtbestand. Orte, denen bisher weder maschinell noch manuell ein Staatenschlüssel zugewiesen werden konnte, wurden nochmals mit diesem Gesamtbestand abgeglichen und gegebenenfalls einem Staat zugewiesen. Alle Orte, denen trotz dieses Abgleichs kein Staat eindeutig zugeordnet werden konnte, erfuhren einen Abgleich über die im Bestand enthaltenen Kreise mit einer Kreisliste. Diese Liste wurde aus den Kreisangaben des Ortsverzeichnisses generiert, qualitätsgesichert und enthielt somit staatenübergreifende Kreis-Staat-Kombinationen. Bei eindeutigem Treffer konnte so ebenfalls eine Schlüsselzuweisung erfolgen. Nach diesen Schritten existierte ein konsolidierter Gesamtbestand.

### 4.3 Phase der Aufbereitung

Ziel der folgenden Phase „Aufbereitung“ war eine Qualitätsverbesserung. Während bisher die Informationen aus

<sup>19</sup> Durch Zusatzinformationen wie Landstriche, Regionen, Bundesländer, Staatenangaben und deren Abkürzungen wurden die Staatenschlüssel im zugewiesenen Staat gegebenenfalls korrigiert. Hauptaugenmerk lag auf Angaben zu Böhmen, zum Sudetenland und zum Riesengebirge sowie Varianten davon.

den Melderegistern genutzt wurden, um einen generierten aktuellen Geburtsort und zugewiesenen Geburtsstaat auszuweisen, wurden letztere im Folgenden selbst im Sinne der Anforderungen nochmals aufbereitet. Aufgrund der vielfältigen Einträge können nur die wichtigsten Bearbeitungsschritte dargelegt werden.

#### 4.3.1 Aufbereitung des Geburtsstaates

In den bisherigen Schritten wurden Klarschriftangaben zu Staaten im Ausgangsdatenfeld „Geburtsort“ genutzt, um den auszuweisenden Geburtsstaat zu generieren. Augenscheinliche Staatennamen in diesem Ausgangsdatenfeld durchliefen eine manuelle Kontrolle, da hier der Geburtsort der Person eingetragen sein sollte, einige Staatennamen aber gleichzeitig auch Ortsnamen sein können.<sup>20</sup> Stand ein vorangestellter Staatename und eine zusätzliche Angabe zu Ort oder Gebiet ebenfalls im Ausgangsdatenfeld „Geburtsort“, so wurden diese in den generierten aktuellen Geburtsort und -staat übernommen. War nur ein Staatename enthalten, wurden der generierte aktuelle Geburtsort sowie der Staatenschlüssel abhängig vom Eintrag im Ausgangsdatenfeld „Geburtsort – Staat“ ausgewiesen.

Um für veraltete Staatenschlüssel im zugewiesenen Staat einen aktuellen zuordnen zu können, wurde innerhalb des Ortsverzeichnisses ein Abgleich mit allen Orten durchgeführt, deren Staatenschlüssel für einen Folgestaat der ehemaligen Staatenschlüssel standen. Falls die Orte mit jeweils nur einem weiteren aktuellen Staatenschlüssel vorhanden waren, wurde für diese Orte der jeweilige aktuelle Staatenschlüssel übernommen.

In den Daten der Melderegisterlieferung von 2008, auf denen das Ortsverzeichnis basierte, wurden auch Schlüssel mitgeliefert, denen kein Staat zugeordnet werden konnte. Im Zuge einer Auswertung des Ausgangsdatenfeldes „Geburtsort – Staat“ wurden verschiedene unbestimmte Staatenschlüssel und Klarschriftangaben, welche zum Teil aus Staatskürzeln bestanden, identifiziert. Einem Großteil dieser nicht existenten Schlüssel konnte durch die bereits durchgeführte Staatenschlüsselzuweisung ein existierender Staat zugeordnet werden. Für die noch zu bereinigenden Datensätze wurde der zugewiesene Staat manuell korrigiert. Darauf folgte ein induktives Vorgehen, bei dem alle Ersatz- und unbekanntenen Staatenschlüssel im Feld zugewiesener Staat identifiziert wurden. Bei einem Teil dieser Schlüssel handelte es sich um ehemalige oder aktuelle Gebietsschlüssel, die in die äquivalenten Staatenschlüssel umgeschlüsselt wurden.

#### 4.3.2 Aufbereitung des Geburtsortes

Um eine korrekte Ausweisung der Ortsnamen zu realisieren, musste der in den bisherigen Phasen erstellte generierte aktuelle Geburtsort aufbereitet werden. Zunächst wurden Ortsnamen vereinheitlicht sowie alle Ortsnamensbestandteile korrekt ausgewiesen und alle Ortsnamenszusätze standardisiert. Dieses Verfahren garantierte eine einheitliche Darstellung gleicher Orte und stellte sicher, dass Orte, deren Namen mehrfach vorkommen und die nur durch ihre

Zusätze zu unterscheiden sind, gezielt als unterschiedlich ausgewiesen wurden. Des Weiteren wurden Melderegister-Einträge, die keinen Ortsnamen enthielten, sowie mögliche Transkriptionen für ausländische Geburtsorte gesondert ausgewiesen.

Hinter Datensätzen mit verschiedenen Schreibweisen kann sich ein und derselbe reale Ort verbergen. Hierzu mussten unterschiedliche Schreibweisen desselben Ortes identifiziert und zu einer homogenen Darstellung vereinigt werden. Eine Korrektur der Darstellung erfolgte nur, wenn eindeutig identifiziert werden konnte, dass es sich um denselben Ort handelte. Für ausländische Orte war dies nicht zu garantieren. Die Korrektur der Schreibweisen lehnte sich an das Grundprinzip der Dublettenerkennung in Form eines paarweisen Vergleiches ausgewählter Datensätze an. Um für die deutschen Orte so viele ähnliche Schreibweisen wie möglich zu erfassen, musste ein zweistufiger Prozessablauf erfolgen. Dieser bestand aus der maschinellen Ausweisung potenziell gleicher Orte über Ähnlichkeitssuchen und einer manuellen Korrektur der so identifizierten Orte. Orte, deren Name mehrfach in Deutschland vorkam, wurden hierbei nicht berücksichtigt. Ein Abgleich des generierten aktuellen Geburtsortes mit dem externen Verzeichnis der Deutschen Post AG garantierte die Festlegung der richtigen Schreibweise. Neben den Anfangsbuchstaben der Orte, deren Zeichenkettenlängen identisch sein mussten oder maximal in einer Stringstelle voneinander abweichen durften, diente auch die Personenanzahl je Schreibweise als Blockingvariable.<sup>21</sup> Die Auftrittshäufigkeit einer Schreibweise, gekennzeichnet durch die Personenanzahl je Schreibweise, diente als Blockingvariable, da davon auszugehen war, dass eine falsche Schreibweise seltener auftritt als die korrekte.<sup>22</sup> Für die Dublettenbestimmung wurde die Jaro-Winkler-Distanz<sup>23</sup> als Ähnlichkeitsmaß verwendet, welche die Ähnlichkeit zweier Zeichenketten zueinander misst und als Ergebnis einen Wert zwischen 0 und 1 ausgibt, wobei 1 einer 100%igen Übereinstimmung und 0 keiner Ähnlichkeit entspricht. Der Schwellenwert wurde auf 0,9 festgelegt. Für Vergleichspaare, deren Wert darüber lag und die somit als „ähnlich“ galten, erfolgte eine manuelle Recherche der korrekten Schreibweise sowie eine Prüfung des zugewiesenen Staates.

Um eine vereinheitlichte Schreibweise und die damit verbundene Ausweisung des Geburtsortes als „gleich“ zu garantieren, mussten zudem Abkürzungen der Ortsnamensbestandteile in eine einheitliche Form gebracht werden. Diese Vereinheitlichung erfolgte im Zuge mehrerer manueller Aufbereitungen. Allgemeingültige Abkürzungen (wie zum Beispiel von „Sankt“ und „Groß“) wurden anhand einer aus dem Verzeichnis selbst erstellten Abkürzungsübersicht

<sup>20</sup> Beispielsweise ist „Luxemburg“ nicht nur ein Ländername, sondern auch der Name der Hauptstadt Luxemburgs sowie einer Reihe von Orten vor allem in Osteuropa.

<sup>21</sup> Zur Steigerung der Effizienz wurde eine Partition der Tabelle, die alle deutschen Orte enthielt, basierend auf den Anfangsbuchstaben der Orte vorgenommen. Der Paarvergleich fand nur innerhalb einer Partition statt. Aufgrund urkundlicher Abschrift bei deutschen Orten war es unwahrscheinlich, dass mehrere Duplikate in verschiedenen Partitionen lagen.

<sup>22</sup> Es wurden innerhalb einer Partition nur Orte, deren Personenanzahl unter zehn lag, mit Orten, deren Personenanzahl über zehn lag, abgeglichen.

<sup>23</sup> Siehe hierzu Winkler, W. E.: „String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage“, Proceedings of the Section on Survey Research Methods, American Statistical Association/1990, Seite 354 ff., sowie Winkler, W. E.: „Overview of Record Linkage and Current Research Directions“, Research Report Series, 2006 ([www.census.gov/srd/papers/pdf/rss2006-02.pdf](http://www.census.gov/srd/papers/pdf/rss2006-02.pdf); abgerufen am 5. Oktober 2012).



identifiziert und nach festen Regeln standardisiert.<sup>24</sup> Starke Verkürzungen des gesamten Ortsnamens im generierten aktuellen Geburtsort wie zum Beispiel SBR (Saarbrücken) ließen sich nur durch Recherche und unter Berücksichtigung von Zusatzinformationen (Früher-Jetzt-Bezeichnung) erschließen.<sup>25</sup>

Bei der Überarbeitung des generierten aktuellen Geburtsortes mussten auch Ortsnamensbestandteile einbezogen werden, die in sich eine eigene Bedeutung trugen. So konnten diese zum Ortsnamen gehören, der Ortsnamen selbst sein oder nur Zusätze sein, die den Ortsnamen genauer ausweisen sollten. Ein Beispiel hierfür ist das Wort „ABBAU“, welches sowohl auf ein Abbaugelände für Bodenschätze hinweist als auch im Sinne von „Vorort“ verwendet werden kann. Über eine Schlagwortsuche wurden alle betroffenen Datensätze erfasst und dahingehend überprüft, ob sie Ortsnamensbestandteil sind.<sup>26</sup> Besonderheiten bildeten hierbei vorgestellte Schlagworte wie „Bahnhof“, „Wald“ oder auch „Försterei“, für welche im Zuge der manuellen Zuweisung danach gestrebt wurde, den Ort, in dem zum Beispiel der Bahnhof steht, auszuweisen.

Zur Verbesserung der Qualität des auszuweisenden Ortes unterlagen die Angaben im Ausgangsdatenfeld „Geburtsort“ in jedem Aufbereitungsschritt der Kontrolle auf Identifizierbarkeit eines Ortsnamens. Orte galten als nicht identifizierbar, wenn die Eintragungen im Ausgangsdatenfeld keine Rückschlüsse auf einen Ortsnamen zuließen. Aufgrund der unterschiedlichen Einträge konnte hierfür keine eindeutige Definition festgelegt werden.<sup>27</sup> Angaben, die keinen Ort bezeichneten, erhielten die Ausweisung „UNBEKANNT“. Dies galt auch bei jeglichen Arten von Angaben, welche die Bedeutung „unbekannt“ in sich trugen – unter dem Vorbehalt, dass im Ausgangsdatenfeld „Geburtsort – Standesamt“ kein Ortsname verzeichnet war. War hier ein Ortsname vorhanden, wurde dieser ausgewiesen.

Befanden sich in den Ausgangsdaten für den „Geburtsort“ Eintragungen, die keinen Geburtsort, sondern nur höhere administrative Einheiten wie Kreis- oder Gebietsangaben darstellten, wurden diese, um einen Informationsverlust auszugleichen, mit dem Hinweis auf die jeweilige administrative Einheit ausgewiesen.

Bei Ortseinträgen, in denen Zahlen einzelne Buchstaben ersetzten, wie beispielsweise „K1ew“, wurde der äquivalente Buchstabe eingesetzt. Durch das Herausfinden der Position der Zahl innerhalb des Ausgangsdatenfeldes wurde eine Masse identifiziert und darauf geprüft, ob die Zahl als

Ortsnamensbestandteil zu sehen ist, der erhalten bleiben muss, oder ob sie als Zusatzinformation beziehungsweise meldebehördlich intern für ein Ordnungssystem genutzt wird und somit entfernt werden musste.

Die Angaben zum Geburtsort in den Melderegistern stammen in der Regel aus dem vorgelegten deutschen Personaldokument – dies ist bei der Erstanmeldung die Geburtsurkunde, später dann der Personalausweis – und sind weitestgehend zuverlässig. Bei Zuwanderungen aus dem Ausland werden Dokumente oder mündliche Angaben genutzt und möglicherweise unterschiedliche Transkriptionsverfahren bei der „Eindeutschung“ ausländischer Ortsbezeichnungen angewendet. Um der unterschiedlichen Qualität der Ausgangsdaten Rechnung zu tragen, blieben Ortsnamenszusätze bei ausländischen Orten im referenzierten Geburtsort erhalten. Bei deutschen Orten entfielen die Ortsnamenszusätze und wurden nur für Ortsnamen, die mehrfach vorkommen, in einer Extraspalte ausgewiesen. Die Standardisierung lehnte sich an die durch die Referenztafel des Anschriften- und Gebäuderegisters vorgegebene Form an. Zudem wurden für diese Orte auch der administrative Gemeindegemeindegemeinde- und -schlüssel separat ausgewiesen. Hilfreich erwiesen sich für deutsche und ausländische Orte Zusatzinformationen zur Lagebestimmung in Form von „bei“-Angaben. Zwar konnten die nächstgelegenen Orte nicht immer gefunden werden, die Angaben besaßen jedoch eine hohe Relevanz sowohl für manuelle Prüfungen – vor allem bei mehrfach vorkommenden Ortsnamen – als auch in allen Ausweisungsprozessen zur Bestimmung des Geburtsortes auf Gemeindeebene.

Zudem kann derselbe Ort je nach Sprache verschiedene Namen besitzen. Beispiele hierfür sind die tschechische Stadt „Karlovy Vary“ (deutsch: Karlsbad) sowie die polnische Stadt „Bytom“ (deutsch: Beuthen) oder „Munich“ (deutsch: München). Die deutschsprachige Ortsbezeichnung wurde in den referenzierten Geburtsort und der Ort in Landessprache in eine Hilfsspalte übernommen, um vor allem für die Mehrfachfallprüfung sicherzustellen, dass ein Geburtsort trotz unterschiedlicher Transkriptionen als gleicher Ort ausgewiesen wurde. Hier waren neben dem Einsatz einer Transkriptionsliste aufwendige Recherchearbeiten erforderlich.

Nach den bisherigen Aufbereitungen kann im Folgenden von einem referenzierten Geburtsort und einem referenzierten Geburtsstaat gesprochen werden.

#### 4.4 Ausweisung der Orte nach dem nationalen Auswertungskonzept

Nach den Anforderungen des nationalen Auswertungskonzeptes sollen Personen, die bis zum 2. August 1945 in den Grenzen Deutschlands nach dem Gebietsstand vom 31. Dezember 1937 geboren wurden, als „in Deutschland geborene“ Personen ausgewiesen werden. Dazu wurden Informationen zum Geburtsdatum genutzt, die aus den Melderegistern zur Verfügung standen. Zuvor mussten innerhalb des Ortsverzeichnisses die potenziell für diesen Abgleich relevanten Orte als historische Orte ausgewiesen werden. Ein Ort wurde innerhalb des Ortsverzeichnisses als historischer Ort gekennzeichnet, indem neben dem Staatenschlüssel nach aktuellem Gebietsstand im referenzierten Geburts-

<sup>24</sup> Es musste lediglich sichergestellt sein, dass es sich tatsächlich um einen Ortsnamensbestandteil handelte, da einige Meldebehörden einzelne Buchstaben als internes Ordnungssystem verwenden und so zum Beispiel den Buchstaben „A“ (für „Austria“) dem Ortsnamen voranstellen, um diesen Ort als österreichisch zu kennzeichnen.

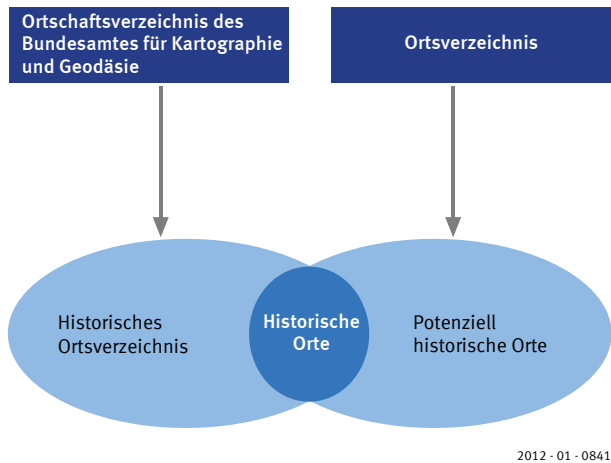
<sup>25</sup> Genutzt wurden hierfür neben Zusatzinformationen aus dem Originalmelderegisterfeld „Geburtsort“ („Altenkessel j. SBR“) auch Informationen aus dem Feld „Geburtsort – Standesamt“.

<sup>26</sup> Diese Schlagworte umfassen beispielsweise Toponyme (Dorf, Stadt, Burg, Villa, Kirchspiel), Adverbien („deutsch“, „schlesisch“, „bohmisch“, „adlig“), Lagebezeichnungen („unter“, „hinten“, „vor der“) oder Himmelsrichtungen („north“, „west“).

<sup>27</sup> So können diese aus Zahlen oder einzelnen Buchstaben bestehen. Alle Informationen wie zum Beispiel Postleitzahlen oder geografische Koordinaten wurden genutzt, um eine Gemeinde auszuweisen und entsprechend einen zugehörigen Staatenschlüssel ermitteln zu können.

staat der deutsche Staatenschlüssel im Feld referenzierter historischer Geburtsstaat hinterlegt wurde. Für alle in diesem Zusammenhang nicht relevanten Orte wurden Staatenschlüssel nach aktuellem Gebietsstand ausgewiesen. Die Kennzeichnung der Orte als historisch erfolgte durch einen stufenweisen Abgleich des „historischen Ortsverzeichnisses“ mit den potenziellen historischen Orten des Ortsverzeichnisses (siehe Schaubild 2).

Schaubild 2 Ausweis historischer Orte im Ortsverzeichnis



4.4.1 Erstellen des historischen Ortsverzeichnisses

Aus der vom Bundesamt für Kartographie und Geodäsie in elektronischer Form vorliegenden Reihe historischer Ortschaftsverzeichnisse wurden auf Kreisebene sämtliche nach den Grenzen vom 31. Dezember 1937 zu Deutschland gehörenden Ortsangaben herausgefiltert. Diese 17 263 Datensätze umfassende Datei „historisches Ortsverzeichnis“ enthält die für die gefilterten Orte chronologische Veränderung der Ortsnamen in standardisierter Form und deren Staatszugehörigkeiten. Um zu vermeiden, dass Orte, die aktuell in Deutschland liegen, aufgrund eines gleichen Ortsnamens als historische Grenzorte ausgewiesen würden, erfolgten nacheinander durchgeführte Abgleiche dieser Datei mit Ortsnamen und Ortsteilen sowohl aus dem Anschriften- und Gebäuderegister als auch aus dem Verzeichnis der Deutschen Post AG. Orte, deren alte oder aktuelle Ortsbezeichnung heute noch in Deutschland zu finden ist, wurden in dieser Menge nicht mehr berücksichtigt.

4.4.2 Identifikation der potenziell historischen Orte im Ortsverzeichnis

Um nicht alle Datensätze des Ortsverzeichnisses mit dem historischen Ortsverzeichnis abgleichen zu müssen, wurde eine Teilmenge mit potenziell historischen Orten aus dem Ortsverzeichnis erstellt. Diese enthielt nur Orte, die referenzierten Staat die Staatenschlüssel Polens, der Russischen Föderation, Deutschlands, der ehemaligen Tschechoslowakei, der Tschechischen Republik oder der ehemaligen Sowjetunion enthielten und/oder in den Zusatzinformationen im Ausgangsdatenfeld Hinweise auf Hinterpommern, Ostbrandenburg, Niederschlesien, Oberschlesien, Ostpreußen und Varianten davon. Ausgeschlossen aus dieser

Menge wurden im Gegenzug alle Orte, die zwar die oben aufgeführten Staatenschlüssel besaßen, aber im Ausgangsdatenfeld Hinweise auf Wartheland, Westpreußen, Sudetenland, Böhmen und Mähren, Memelland, Sudauen, Bialystok oder Generalgouvernement enthielten. Die Begründung dieser Vorgaben liegt im Grenzverlauf vom 31. Dezember 1937, der bestimmte Teile der heutigen Staaten Polen, Russische Föderation und Tschechische Republik und deren Vorgängerstaaten einschließt, andere wiederum nicht.

4.4.3 Ausweisung des Geburtsstaates

Durch einen dreistufigen Abgleich der beiden so erstellten Teilmengen wurden die historischen Orte ausgewiesen. In der ersten Stufe erfolgte ein Abgleich des referenzierten Geburtsortes der potenziell historischen Orte des Ortsverzeichnisses mit den standardisierten Orten aus dem historischen Ortsverzeichnis. Aus dem historischen Ortsverzeichnis wurden hierfür alle relevanten Ortsumbenennungen<sup>28</sup> einbezogen, um auch Melderegistereinträge als historische Orte zu klassifizieren, in denen ein aktuellerer Ortsname des historischen Ortes stand. Neben diesem Abgleich der Ortsnamen war der Staatenschlüssel im referenzierten Geburtsstaat für die Ausweisung als historischer Ort unabhängig, da – sollte auch der Staatenschlüssel im historischen Ortsverzeichnis identisch sein – diese Kombination einen Ort eindeutig innerhalb der Grenzen vom 31. Dezember 1937 zuordnet. Die gleiche Ausweisung erhielten Orte, die als referenzierten Geburtsstaat einen deutschen Staatenschlüssel, „ohne Angabe“ oder einen veralteten Staatenschlüssel besaßen, wenn der historische Ort in einem Folgestaat davon lag.

In der zweiten Stufe wurden Kreisangaben genutzt, um weitere Zuordnungen zu erhalten. Hierzu wurden Kreisangaben aus der Teilmenge des Ortsverzeichnisses mit den standardisierten Kreisangaben des historischen Ortsverzeichnisses abgeglichen. Bei Übereinstimmung wurde der Ort nach den gleichen Regeln als historisch ausgewiesen wie bei dem direkten Ortsnamensabgleich, auch wenn der Ort an sich in der ersten Stufe nicht als historisch erkannt wurde.

In der dritten Stufe wurden alle Orte zugeordnet, deren Abgleich in der ersten und zweiten Stufe zu mehreren Treffern im historischen Ortsverzeichnis führte und denen somit kein eindeutiger aktueller Staat zugewiesen werden konnte. Befand sich ein Ort in mehreren Staaten, so wurde die Kreisangabe zur Staatszuweisung genutzt. Eindeutigkeit war gegeben, wenn die Kombination aus Ort und Kreis in nur einem Staat zu finden war.

4.5 Inhalte des Ortsverzeichnisses

Nach dem Ausweisen der historischen Orte liegt zu jedem Melderegistereintrag ein qualitätsgesicherter referenzierter Geburtsort mit zugehörigem, durch die Auswertungskonzepte definiertem referenziertem Geburtsstaat nach aktuellem Gebietsstand und in den Grenzen vom 31. Dezember 1937 vor. Neben diesen für die Auswertungsda-

<sup>28</sup> Die relevanten Ortsumbenennungen waren der standardisierte Ortsname des historischen Ortes bis 1937, der standardisierte Ortsname des Ortes bis 1945 und der aktuelle Ortsname.

tenbank relevanten Ausweisungen enthält das Ortsverzeichnis eine Reihe aus der Aufbereitung resultierender zusätzlicher Informationen zu den jeweiligen Orten, wie den ehemaligen Ortsnamen oder eine äquivalente Ortsbezeichnung in Landessprache. Mehrfach in Deutschland vorkommende Ortsnamen sind nicht nur durch Ortsnamenssätze gekennzeichnet, sondern erhielten eine Ausweisung des administrativen Gemeindegamens einschließlich des Gemeindegamenschlüssels. Weitere im Zuge der Aufbereitung entstandene Informationen, wie Kreisangaben zu ausländischen und deutschen Kreisen oder die Ausweisung der Staatenschlüssel nach Klarschriftangaben, befinden sich ebenfalls im Ortsverzeichnis.

## 5 Anwendung des Ortsverzeichnisses

### 5.1 Rolle des Ortsverzeichnisses in der Mehrfachfallprüfung

Die Ermittlung der Einwohnerzahlen im Rahmen des Zensus 2011 basiert auf den Melderegisterdaten. Da jede Person in Deutschland bei der Ermittlung der Einwohnerzahlen nur einmal mit alleinigem oder Hauptwohnsitz gezählt werden darf, mussten durch die Mehrfachfallprüfung unzulässige Mehrfachanmeldungen ermittelt und bereinigt werden. Um doppelte Datensätze, welche die gleiche Person repräsentieren, zu kennzeichnen, wurde im Zuge des Teilschrittes „Dublettensuche“ ein Vergleich des Datenbestandes mit sich selbst durchgeführt mit dem Ziel, identische oder ähnliche Dublettenelemente an unterschiedlichen Anschriften zu finden. Dem Vergleich lag die Annahme zugrunde, dass es bestimmte (im Regelfall) unveränderliche Grundmerkmale einer Person gibt. Dies sind der Geburtsname, das Geburtsdatum, die Vornamen, das Geschlecht und der Geburtsort. Diese Merkmale sollten sich nur in Ausnahmefällen nach der Erstaufnahme in das Melderegister ändern. Bei der Erfassung der unveränderlichen Merkmale wurde durch die Meldebehörden unterschiedlich vorgegangen.<sup>29</sup> Um Personen als Mehrfachfälle zu identifizieren, war es notwendig, Standardisierungen durchzuführen.

Jedem Melderegistereintrag musste eine standardisierte Form des Geburtsort zugeordnet werden. Diese lag als referenzierter Geburtsort vor. Zudem musste sichergestellt werden, dass trotz unterschiedlicher Einträge in den Melderegistern – hervorgerufen durch Eingemeindungen oder Ortsumbenennungen des Geburtsortes – die Geburtsorte dennoch als „gleich“ identifiziert werden konnten. Hierzu dienten alle Aufbereitungen, die Ortsteile einer aktuellen Gemeinde zuordneten. Der generierte ehemalige Ortsname wurde genutzt, um Orte trotz Umbenennung als gleich auszuweisen. Stand nun im generierten Ort noch ein ehemaliger Ortsname, konnte dieser, sofern er auf sein aktuelles Pendant traf, als „gleicher“ Ort identifiziert werden. Orte, die sich durch abweichende Übersetzungen unterschieden, wurden mithilfe der Ausweisung des Ortes in Lan-

dessprache für die Mehrfachfallprüfung auf eine einheitliche „Übersetzung“ standardisiert. Des Weiteren mussten Orte, deren Namen mehrfach in Deutschland vorkamen, aber gleich geschrieben wurden, als unterschiedliche Orte gekennzeichnet werden. Hierzu wurden der administrative Gemeindegamensname sowie der administrative Gemeindegamenschlüssel ergänzend verwendet.

Durch den referenzierten Geburtsort konnten in einem maschinellen Abgleich Geburtsorte verschiedener Datensätze als gleich oder nicht gleich identifiziert werden. Eine sich anschließende maschinell unterstützte Prüfung der verbleibenden noch nicht abschließend bestimmten Dubletten im Rahmen der Mehrfachfallprüfung basierte ebenfalls auf dem referenzierten Geburtsort.

### 5.2 Klassifizierung der Geburtsort – Staat-Kombination im Referenzdatenbestand

Die im Zuge der Aufbereitung des Ortsverzeichnisses erstellten referenzierten Geburtsorte, Geburtsstaaten und historischen Geburtsstaaten müssen zur Erfüllung der Anforderungen der Auswertungskonzepte in den Referenzdatenbestand<sup>30</sup> integriert werden, um später in der Auswertungsdatenbank vorzuliegen. Diese Anbindung erfolgte in mehreren Stufen, wobei zunächst ein direkter Abgleich über die Ausgangsdatenfelder und dann ein Abgleich unter Einbeziehen der Hilfsmerkmale stattfand. Zum Schluss wurden die Staaten nach nationalem und EU-Staatenkonzept anhand des Geburtsdatums zugewiesen und die Felder nachbereitet.

#### 5.2.1 Erste Stufe: Direkte Anbindung über die Originalfelder

Zur Vorbereitung der Anbindung erhielten alle Geburtsortfelder, die im aktuellen Melderegisterbestand als leer identifiziert wurden, im referenzierten Geburtsort die Ausweisung „UNBEKANNT“. Als leer wurde ein Geburtsortfeld dann angesehen, wenn im Ausgangsdatenfeld keine Buchstaben vorhanden waren.<sup>31</sup> Analog hierzu wurde mit Einträgen verfahren, die im Ausgangsdatenfeld „unbekannt“ oder Synonyme hiervon stehen hatten. Gültige Staatenschlüssel wurden in den referenzierten Geburtsstaat übernommen. Bei fehlender Angabe wurde auf „ohne Angabe“ gesetzt. Nach dieser vorbereitenden Maßnahme erfolgte die direkte Anbindung.

In einem ersten Schritt wurden die Ausgangsdatenfelder „Geburtsort“, „Geburtsort – Standesamt“ und „Geburtsort – Staat“ in unveränderter Form genutzt. Durch einen Abgleich dieser Originalfelder aus dem Referenzdatenbestand mit denen aus dem Ortsverzeichnis (1:1-Anbindung) wurden bei eindeutiger Zuordnung die referenzierten Felder des Ortsverzeichnisses in den Referenzdatenbestand übernommen.

<sup>29</sup> Näheres zum Vorgehen in der Mehrfachfallprüfung siehe Diehl, E.-M.: „Methode der Mehrfachfallprüfung im Zensus 2011“ in WiSta 6/2012, Seite 473 ff.

<sup>30</sup> Siehe hierzu §§ 9 und 12 des Gesetzes über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009 (BGBl. I Seite 1781).

<sup>31</sup> Dies betrifft auch Einträge, die nur aus Sonderzeichen oder Geburtsdaten bestehen.

Nach dieser 1:1-Anbindung wurde in einem zweiten Schritt über die zuvor standardisierten Ausgangsdatenfelder „Geburtsort“ und „Geburtsort – Standesamt“ angebunden. Durch die Standardisierung ergaben sich eine Vereinheitlichung von unterschiedlichen Formen der Darstellung der Orts- und Zusatzangaben und damit verbunden weitere Übereinstimmungen der Einträge im Ortsverzeichnis mit denen im Referenzdatenbestand.

Die Standardisierung der Originalfelder („Geburtsort“ und „Geburtsort – Standesamt“) erfolgte zunächst in einer leichten Form.<sup>32</sup> Diese entsprach der schon im Ortsverzeichnis vorhandenen Standardisierung des Ausgangsdatenfeldes „Geburtsort“. Es wurden nur eindeutige Treffer in den Referenzdatenbestand übernommen. In den ersten beiden Anbindungsschritten traten keine Mehrfachtreffer auf. Erst bei den Anbindungen mit der stärkeren Form der Standardisierung<sup>33</sup> kam es vor, dass zu einem Datensatz des Referenzdatenbestandes mehrere Datensätze aus dem Ortsverzeichnis passten (Mehrfachtreffer).<sup>34</sup> Dieser Fallkonstellation wurde in der zweiten Stufe der Anbindung nachgegangen.

Im dritten Schritt, der direkten Anbindung, wurden nicht mehr alle Ausgangsdatenfelder genutzt, sondern beim Abgleich nacheinander einzelne Felder weggelassen. Zunächst wurde der Geburtsstaat, dann das Standesamt und anschließend wurden beide Felder nicht mehr berücksichtigt. Im letzten Schritt erfolgte eine Anbindung ausschließlich aufgrund des leicht standardisierten Ausgangsdatenfeldes „Geburtsort“. Auch hier wurden nur eindeutige Treffer angebunden.

Sobald bei der Anbindung mindestens einer der Mehrfachtreffer ein ehemals deutscher Ort war, wurde mit den Staatenschlüsseln des ehemals deutschen Ortes angebunden, sofern es nur eine Möglichkeit für einen ehemals deutschen Ort gab und für die anderen Fälle der Staat nach nationalem und EU-Staatenkonzept übereinstimmte. Alle Meldesätze, die hierdurch keine Klassifizierung durch das Ortsverzeichnis erhalten hatten oder Mehrfachtreffer waren, wurden in der nächsten Stufe bearbeitet.

### 5.2.2 Zweite Stufe: Anbindung unter Einbeziehung der Hilfsmerkmale

Im Folgenden wurden bei Datensätzen, die in der ersten Stufe nicht eindeutig angebunden werden konnten, weitere im Referenzdatenbestand vorliegende Merkmale der Person wie deren Staatsangehörigkeiten und/oder der Zuzug aus dem Ausland mit einbezogen, um Mehrfachanbindungen zu vermeiden. Um in diesen Fällen eine plausible Klassifizierung des Geburtsstaates der Person zu erreichen, wurde geprüft, ob eine oder mehrere Staatsangehörigkeiten der Person identisch zu den referenzierten Geburtsstaaten der Mehrfachtreffer aus dem Ortsverzeichnis waren. Bei einem

eindeutigen Treffer wurde dieser verwendet und die referenzierten Felder in den Referenzdatenbestand übernommen. Stimmten dagegen mehrere oder keine Staatsangehörigkeiten der Person mit den referenzierten Geburtsstaaten aus dem Ortsverzeichnis überein, wurde überprüft, ob Informationen über den Zuzug aus dem Ausland vorlagen, die einem der referenzierten Geburtsstaaten entsprachen.

### 5.2.3 Dritte Stufe: Anbindung an den standardisierten Ortsnamen

*Anbindung des Originalgeburtsortes an den standardisierten Geburtsort*

Das Ausgangsdatenfeld „Geburtsort“ wurde zunächst mit, dann ohne Verwendung des Geburtsstaates direkt an den referenzierten Geburtsort angebunden. Hierbei wurde sowohl der Staat nach nationalem Staatenkonzept als auch der Staat nach EU-Staatenkonzept verwendet. Eindeutige Treffer wurden angebunden und in den Referenzdatenbestand übernommen. Ergaben sich Mehrfachtreffer lediglich aus dem Geburtsstaat, so wurde der referenzierte Geburtsort aus dem Ortsverzeichnis in den Referenzdatenbestand übernommen und der Staatenschlüssel mit „ohne Angabe“ ausgewiesen. Anschließend wurde das Verfahren analog mit dem Feld des generierten ehemaligen Geburtsortes wiederholt, wobei wiederum der referenzierte Geburtsort in den Referenzdatenbestand übernommen wurde.<sup>35</sup>

*Teilanbindung des Originalgeburtsortes an den standardisierten Geburtsort*

In einem weiteren Schritt wurde für die verbleibenden Fälle nur ein Teil des Geburtsortsfeldes untersucht. Im Ortsverzeichnis stehen im Ausgangsdatenfeld „Geburtsort“ neben dem Geburtsort auch Zusatzinformationen. Es lag daher nahe, einen Teil des Melderegisterfeldes unbeachtet zu lassen. Mehr als 99% der referenzierten Geburtsorte bestehen aus höchstens drei Wörtern. Das Geburtsortsfeld wurde daher zerlegt und der Reihe nach die ersten drei Wörter, die ersten zwei Wörter und das erste Wort des Geburtsortes für die Anbindung an den referenzierten Geburtsort verwendet. Auch hier wurde zunächst mit Staatenschlüssel angebunden und anschließend ohne.<sup>36</sup> Bei mehreren getroffenen Staatenschlüsseln wurde der Geburtsstaat „999“ zugewiesen.

### 5.2.4 Ausweisung ehemals deutscher Ostgebiete

In Artikel 116 Absatz 1 des Grundgesetzes für die Bundesrepublik Deutschland ist Folgendes geregelt: „Deutscher im Sinne dieses Grundgesetzes ist vorbehaltlich anderweitiger gesetzlicher Regelung, wer die deutsche Staatsangehörigkeit besitzt oder als Flüchtling oder Vertriebener deutscher Volkszugehörigkeit oder als dessen Ehegatte oder Abkömmling in dem Gebiete des Deutschen Reiches nach dem Stande vom 31. Dezember 1937 Aufnahme gefunden hat.“ In diesem Sinne wurden bereits durch das „historische Ortsverzeichnis“ alle Gemeinden gekennzeichnet, die nach

<sup>32</sup> Die leichte Standardisierung ersetzt alle Kleinbuchstaben durch Großbuchstaben und löst die Umlaute auf (aus Ä wird AE und so weiter).

<sup>33</sup> In der zweiten, stärkeren Standardisierung der Originalmelderegisterfelder wurden Sonder- durch Leerzeichen ersetzt und Ziffern und Satzzeichen entfernt. Einträge im Referenzdatenbestand, die sich noch nicht im Ortsverzeichnis befanden, würden ohne die zweite Standardisierung keine direkte Anbindung erzielen.

<sup>34</sup> Dies betraf nach der zweiten Standardisierung vor allem ortsähnliche Bezeichnungen wie Kolchosen, da sich hier nur eine Eindeutigkeit über die Zahlen in den Ortsnamensbestandteilen ergibt.

<sup>35</sup> Für Mehrfachtreffer, die sich aufgrund des referenzierten Geburtsortes unterschieden, wurde der ehemalige Geburtsort übernommen. Ursache hierfür war, dass verschiedene aktuelle Gemeinden gleichnamige Ortsteile besitzen können.

<sup>36</sup> Bei der Anbindung mit einem Wort wurde nicht angebunden, wenn der referenzierte Geburtsort kürzer als vier Zeichen und kürzer als das Originalfeld war.



den Grenzen vom 31. Dezember 1937 in Deutschland lagen. Bei der Klassifizierung der Geburtsorte für den Referenzdatenbestand wurde für diejenigen Personen, die bis zum 2. August 1945 in diesen gekennzeichneten Gemeinden geboren wurden, entsprechend der Auswertungskonzepte neben dem Staatenschlüssel nach aktuellem Gebietsstand auch der deutsche Staatenschlüssel ausgewiesen. Für alle übrigen Personen wurde in dem für das nationale Auswertungskonzept vorgesehenen Feld der Staatenschlüssel nach aktuellem Gebietsstand ausgewiesen.

### 5.2.5 Qualitätssicherung

Nach der maschinellen Anbindung wurden diejenigen Datensätze aufbereitet, die im letzten Schritt der Anbindung die Zuweisung „999“ erhalten hatten. Durch einen Abgleich mit Ortsnamen aus dem Anschriften- und Gebäuderegister wurde diesen unter der Bedingung, dass das Originalfeld „Geburtsort – Staat“ leer war oder den deutschen Staatenschlüssel enthielt und keine Schlagworte im Ausgangsdatenfeld „Geburtsort“ auf einen ausländischen Staat hinwiesen, der deutsche Staatenschlüssel zugewiesen.<sup>37</sup>

Wurde vom Ortsverzeichnis ein ehemaliger Staatenschlüssel im referenzierten Geburtsstaat zugewiesen und im Originalfeld „Geburtsort – Staat“ stand ein aktueller Staatenschlüssel, so wurde dieser übernommen. Gleiches galt für Klarschriftangaben im Ausgangsdatenfeld „Geburtsort“.

### 5.2.6 Manuelle Klassifikation

Zwischen dem Stand der Melderegister, der zum Aufbau des Ortsverzeichnisses verwendet wurde, und dem Stand der Melderegister zum Zensusstichtag lagen drei Jahre. Durch das mehrstufige Vorgehen bei der Anbindung konnten nicht nur die seit 2008 bestehenden Geburtsortseinträge der Melderegisterdaten im Referenzdatenbestand durch das Ortsverzeichnis klassifiziert werden, sondern auch fast alle in diesem Zeitraum hinzugekommenen Neuaufnahmen von Geburtsortseinträgen im Ausgangsdatenfeld „Geburtsort“. Aufgrund der Vielfalt der Geburtsortsangaben im Melderegister konnte es jedoch vorkommen, dass sich Neueintragen im Register nicht durch das Ortsverzeichnis klassifizieren ließen. Seltene Schreibfehler, ungewöhnliche kryptische Abkürzungen oder bisher nicht vorgekommene Ortsangaben im Originalfeld des Melderegisters führten dazu, dass eine Anbindung des Ortsverzeichnisses nicht erfolgen konnte. Um diese Personenangaben im Zensus 2011 nicht ungeprüft auszuweisen, mussten die offenen Fälle manuell klassifiziert werden. Eine manuelle Ausweisung wurde notwendig, da ein Großteil des noch nicht klassifizierten Bestandes ohne Staatenschlüsseleintrag im Ausgangsdatenfeld „Geburtsort – Staat“ vorlag. Diese Datensätze maschinell mit dem deutschen Staatenschlüssel auszuweisen oder auf „ohne Angabe“ zu setzen, wäre ein hoher Qualitätsverlust gewesen. Die Vorgaben für die manuelle Bearbeitung lehnten sich an die Regeln für die Ausweisung des referenzierten Geburtsortes und Geburtsstaates an. Die Ausweisung der historischen Orte erfolgte erst nach dieser manuellen Klassifikation und entspricht

dem Vorgehen bei der Ausweisung der Orte nach dem nationalen Auswertungskonzept.

## 6 Fazit

Nach den ersten Erfahrungen beim Aufbau eines Ortsverzeichnisses auf Basis von Melderegisterdaten im Rahmen des Zensus 2011 zeigte sich, dass eine Referenzierung des Datenmaterials notwendig war, um Auswertungen auf eine valide Basis zu stellen. Zur Steigerung der Qualität in einem künftigen Zensus müssten weitere Verfahren entwickelt werden, um jedem Geburtsort eine äquivalente Ortsbezeichnung zuzuordnen sowie vereinheitlichte Schreibweisen von ausländischen Orten zu erzeugen. Die Nutzung der klassifizierten Geburtsort – Staat-Kombination im Zensus 2011 und in anderen Erhebungen ermöglicht es, bei Auswertungen beispielsweise zum aktuellen Wohnort sowohl innerdeutsche als auch internationale Migrationsbewegungen abhängig vom Alter und weiteren Personenmerkmalen abzubilden. Ebenso lässt sich darstellen, wie sich Zuwanderer auf bestimmte Städte und aus bestimmten Gebieten verteilen. Eine reine Auswertung der originalen Einträge der Melderegisterfelder würde aufgrund inkorrekt gefüllter Verzeichnungen der Ergebnisse führen. [\[1\]](#)

<sup>37</sup> Die Schlagworte lauteten „Schlesien“, „Preußen“, „Schles.“, „Sudeten.“ und so weiter.

## Auszug aus Wirtschaft und Statistik

### Herausgeber

Statistisches Bundesamt, Wiesbaden

[www.destatis.de](http://www.destatis.de)

### Schriftleitung

Roderich Egeler, Präsident des Statistischen Bundesamtes

Brigitte Reimann (verantwortlich für den Inhalt)

Telefon: + 49 (0) 6 11 / 75 20 86

### Ihr Kontakt zu uns

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

### Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

Telefax: + 49 (0) 6 11 / 75 33 30

### Abkürzungen

WiSta	=	Wirtschaft und Statistik
MD	=	Monatsdurchschnitt
VjD	=	Vierteljahresdurchschnitt
HjD	=	Halbjahresdurchschnitt
JD	=	Jahresdurchschnitt
D	=	Durchschnitt (bei nicht addierfähigen Größen)
Vj	=	Vierteljahr
Hj	=	Halbjahr
a. n. g.	=	anderweitig nicht genannt
o. a. S.	=	ohne ausgeprägten Schwerpunkt
St	=	Stück
Mill.	=	Million
Mrd.	=	Milliarde

### Zeichenerklärung

p	=	vorläufige Zahl
r	=	berichtigte Zahl
s	=	geschätzte Zahl
–	=	nichts vorhanden
0	=	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	=	Zahlenwert unbekannt oder geheim zu halten
...	=	Angabe fällt später an
X	=	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	=	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	=	keine Angaben, da Zahlenwert nicht sicher genug
()	=	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.