

Dipl.-Sozialwissenschaftlerin Katja Ziprik

Qualitätsaspekte des Anschriften- und Gebäude- registers im Zensus 2011

Für den registergestützten Zensus 2011 wurden verschiedene Datenquellen genutzt. Neben den Einwohnermelde-registern und den Personenregistern der Bundesagentur für Arbeit waren dies insbesondere eine schriftliche Befragung aller Eigentümer und Eigentümerinnen von Gebäuden mit Wohnraum sowie eine interviewer-gestützte Befragung von Haushalten auf Stichprobenbasis.

Da es in den verschiedenen Datenquellen kein einheitliches registerübergreifendes Personen-kennzeichen gab, mussten für die Zusammenführung der Daten auf Personenebene Hilfsmerkmale, wie zum Beispiel die Anschrift und Personenmerkmale, verwendet werden. Zu diesem Zweck wurde ein statistisches Register aller Anschriften und Gebäude mit Wohnraum und aller bewohnten Unterkünfte, das Anschriften- und Gebäuderegister (AGR), aufgebaut und geführt. Es bildet die Basis, Referenz und Auswahlgrundlage aller im Zensus durchgeführten Primärstatistiken, Vollerhebungen und Registerintegrationen und ermöglicht die vollständige Erfassung der Zielbevölkerung und Zielgebäude des Zensus 2011. Um diese Aufgaben erfüllen zu können, standen die Qualitätsanforderungen an das Anschriften- und Gebäuderegister und die getroffenen Maßnahmen zu deren Sicherung stets im Blickpunkt.

In diesem Aufsatz wird nach einer Einreihung des Anschriften- und Gebäuderegisters in das Verfahren des Zensus 2011 auf die Besonderheiten der deterministischen Zusammenführungsmethoden eingegangen. Im Anschluss werden die Aufgaben des Anschriften- und Gebäuderegisters selbst dargestellt. Aus diesen werden nachfolgend die durch das Register zu erfüllenden Qualitätsaspekte abgeleitet und beschrieben. Abschließend werden die Ergebnisse zusammengefasst und ein Ausblick gegeben.

1 Das Anschriften- und Gebäude- register im Zensus 2011

Die Europäische Union (EU) beschloss mit der Verordnung über Volks- und Wohnungszählungen vom 9. Juli 2008, dass in einem zehnjährlichen Turnus, beginnend 2011, umfassende Daten über die Bevölkerung und deren Wohnsituation für jeden Mitgliedstaat der EU bereitzustellen sind. Es wurden Pflichtmerkmale festgelegt, die europaweit durch die Zensusrunde 2011 zu gewinnen waren.¹ Die Europäische Union empfiehlt in ihrem Code of Practice, verstärkt administrative Register zur Statistikproduktion zu nutzen, um die Belastung der Auskunftspflichtigen zu reduzieren und die Kosteneffizienz zu steigern.² Ein weiterer Vorteil der Register-nutzung besteht darin, einen Großteil der Zensusmerkmale flächendeckend auch auf kleinräumiger Ebene auswerten zu können.³ In Deutschland wurde mit dem Zensusfest von 2001 der Grundstein für einen Methodenwechsel bei der deutschen Volkszählung gelegt. So wurde erkannt, dass in den Melde- und Erwerbsregistern ein Großteil der personen-

1 Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates vom 9. Juli 2008 über Volks- und Wohnungszählungen (Amtsblatt der EU Nr. L 218, Seite 15).

2 So lautet der Grundsatz 9.4 des Code of Practice: „Administrative Datenquellen werden – wann immer möglich – herangezogen, um doppelte Datenanforderungen zu vermeiden.“ Weiterhin wird der Gedanke der Datenzusammenführung bei der Statistikproduktion im Grundsatz 9.6 festgeschrieben: „Die statistischen Stellen fördern Maßnahmen, die die Verknüpfung von Datenquellen ermöglichen, um den Beantwortungsaufwand zu reduzieren.“ in Eurostat: „Verhaltenskodex für Europäische Statistiken“, angenommen vom Ausschuss für das Europäische Statistische System am 28. September 2011 (http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-32-11-955, abgerufen am 14. November 2012).

3 Zu den Vor- und Nachteilen von registerbasierten Statistiken siehe auch Wirtschaftskommission der Vereinten Nationen für Europa (UNECE): „Register-based statistics in the Nordic countries – Review of best practices with focus on population and social statistics“, New York/Genf 2007, Seite 11 f.

bezogenen Daten bereits flächendeckend in ausreichender Qualität vorliegt.⁴ In Deutschland hat sich deshalb die Politik auf Empfehlung der amtlichen Statistik für einen Methodenwechsel weg von einer flächendeckenden Befragung der gesamten Bevölkerung durch Erhebungsbeauftragte hin zu einem registergestützten Zensus 2011 entschieden.

Für den Zensus 2011 werden verschiedene Datenquellen genutzt, die in Vollerhebungen [Gebäude- und Wohnungszählung (GWZ), Erhebung an Adressen mit Sonderbereichen], administrative Register (Melde- und Erwerbsregister), Haushalbefragung auf Stichprobenbasis und Haushaltegenerierung unterteilt werden können. Sie werden nachfolgend als Erhebungsteile bezeichnet.⁵ Um alle Erhebungsteile miteinander zu verbinden, wurde das Adressen- und Gebäuderegister aufgebaut. Rechtliche Grundlage zum Aufbau und zur Verwendung des Adressen- und Gebäuderegisters im Zensus 2011 bilden das Zensusvorbereitungsgesetz 2011⁶ und das Gesetz über den registergestützten Zensus im Jahre 2011⁷. Die Zielpopulation des Adressen- und Gebäuderegisters wird hier definiert als „[...] die aktuelle Grundgesamtheit aller Gebäude mit Wohnraum einschließlich aller bewohnten Unterkünfte“⁸. Da alle Gebäude über die Adresse beschrieben werden, kann man die Definition dahingehend erweitern, dass die aktuelle Grundgesamtheit aller Adressen gesucht wird, an denen mindestens ein Gebäude mit Wohnraum oder eine bewohnte Unterkunft vorliegt, unabhängig davon, ob diese zum Erfassungszeitpunkt bewohnt waren oder nicht. Das Adressen- und Gebäuderegister ist ein statistisches Register, welches ausschließlich für statistische Zwecke, beispielsweise im Rahmen der Datenintegration und -koordination, zu nutzen ist.⁹ Es ist spätestens sechs Jahre nach dem Zensusstichtag, also zum 9. Mai 2017, zu löschen.¹⁰

Damit das Adressen- und Gebäuderegister für die Erhebungsvorbereitung genutzt werden konnte, wurde es bereits ab 2008 mit den in den §§ 4, 5 und 6 Zensusvorbereitungsgesetz 2011 verankerten Datenlieferungen der Melderegister, der Bundesagentur für Arbeit und den Georeferenzierten Adressdaten – Bund vom Bundesamt für Kartographie und Geodäsie aufgebaut.¹¹ Die Kombination der beiden personenbasierten Register (Melderegister und Register der Bundesagentur für Arbeit) in Verbindung mit den Daten der Vermessungsbehörden konnte Aufschluss

über Wohnraum geben. Jede Adresse, die in mindestens zwei der Datenquellen gefunden wurde, also mindestens in einem personenbasierten Register, wurde als Wohnraum klassifiziert. Alle Adressen, die nur in einer der drei Quellen gefunden werden konnten, wurden auf inhaltliche Korrektheit geprüft und nach § 14 Zensusgesetz 2011 durch die Statistischen Ämter der Länder hinsichtlich des Wohnraums klassifiziert beziehungsweise hinsichtlich ihrer Schreibweise korrigiert.¹² Zur weiteren Pflege des Adressen- und Gebäuderegisters wurden zu den Stichtagen 1. November 2010, 9. Mai 2011 und 9. August 2011 Lieferungen der etwa 86 Millionen personenbasierten Adressen der Meldebehörden in das Adressen- und Gebäuderegister integriert. Zudem wurden die Ergebnisse der Erhebungsteile genutzt, um die Adressen des Adressen- und Gebäuderegisters bis zum Zensusstichtag 9. Mai 2011 zu pflegen.

2 Deterministisches Record Linkage über räumlich und zeitlich instabile Zusammenführungsmerkmale

Die Qualität eines statistischen Registers, das aus mehreren Datenquellen gespeist wird, wird nicht nur durch die Qualität der Ausgangsdatenquellen bestimmt, sondern auch durch die genutzte Zusammenführungsmethode.¹³ Diese war entscheidend dafür, Maßnahmen zu bestimmen und zu bewerten, mit denen die Qualität des Registers verbessert werden konnte.

Die Adresse, die im Zensus 2011 als Zusammenführungsvariable genutzt wird, ist weder räumlich noch zeitlich stabil. Adressen können im Laufe der Zeit aufgrund von Neubauten neu aufgenommen werden oder bei Gebäudeabbrissen auch verschwinden. Durch Umbenennungen ihrer einzelnen Merkmale ist eine Adresse im Zeitverlauf veränderbar, sodass die gleiche Kombination von Merkmalen in einem Jahr ein anderes Objekt beschreiben kann als noch ein Jahr zuvor. Die im Zensus genutzten Register dienen administrativen Zwecken; daher lagen die für statistische Zwecke benötigten Variablen oft in anderen Definitionen und/oder unterschiedlichen Darstellungen vor, als sie nach den im Adressen- und Gebäuderegister festgelegten Definitionen benötigt wurden. Zwar waren die Qualitätsanforderungen an die in das Adressen- und Gebäuderegister eingearbeiteten Register im Hinblick auf ihre Vollständigkeit und ihre Relevanz erfüllt, jedoch variierte die Qualität der Zusammenführungsvariablen „Adresse“ in den Registern in hohem Maß. Durch die dezentrale Führung der Register wurden Adressen beispielsweise nach in den Gemeinden vorliegenden internen Regeln aufbereitet. So unterschieden sich die Register stark in der Darstellungsweise von Straßennamen, Ortsteilen und Hausnummernzusätzen.

4 Siehe Statistische Ämter des Bundes und der Länder: „Ergebnisse des Zensusstests“ in WiSta 8/2004, Seite 813 ff., hier: Seite 814 f.

5 Für ausführlichere Informationen zu dem Zensusmodell und den einzelnen Erhebungsteilen siehe Statistische Ämter des Bundes und der Länder: „Das registergestützte Verfahren beim Zensus 2011“, März 2011 (www.zensus2011.de im Bereich Zensus 2011).

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

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

8 Begründung zum Entwurf des Zensusvorbereitungsgesetzes 2011 (Bundestagsdrucksache Nr. 16/5525), B Besonderer Teil, zu § 2, zu Absatz 2, hier: Satz 2.

9 Siehe Vale, S.: „Using Administrative Sources for Official Statistics, Principles and Practices“, V1.2 (UNECE), August 2010, Kapitel 7.2. (www1.unece.org/stat/platform/download/attachments/8193015/Using+Administrative+Sources+v1.2.pdf?version=1, abgerufen am 14. November 2012).

10 Siehe § 15 Absatz 3 Zensusvorbereitungsgesetz 2011.

11 Die Datenlieferungen erfolgten zu den Stichtagen 1. April 2008 und 1. April 2010 für die Melderegister, 30. September 2007 beziehungsweise 13. März 2008 für die Daten der Bundesagentur für Arbeit sowie zum 1. April 2008, 1. April 2009 und 1. April 2010 für die Georeferenzierten Adressdaten – Bund.

12 Detailliertere Ausführungen zum Aufbau des Adressen- und Gebäuderegisters sind zu finden in Kleber, B./Maldonado, A./Scheuregger, D./Ziprik, K.: „Aufbau des Adressen- und Gebäuderegisters für den Zensus 2011“ in WiSta 7/2009, Seite 629 ff., und Maldonado, A./Scheuregger, D./Ziprik, K.: „Setting up the Central Register of Addresses and Buildings of the German Census 2011 – Data Quality Issues and Solutions“, 2010 (q2010.stat.fi/media/presentations/Maldonado_Scheuregger_Ziprik_Helsinki_Q2010_Final_paper.pdf, abgerufen am 14. November 2012).

13 Siehe Wallgren, A./Wallgren, B.: „Register-based Statistics: Administrative Data for Statistical Purposes“, New York 2007, Seite 175 f.

Erschwerend kam der unterschiedliche Zeitbezug der den Registern zugrunde liegenden Daten hinzu.¹⁴ Ein weiteres Problem ergab sich aus dem in Deutschland geltenden Rückspielverbot. Dieses besagt, dass Informationen aus administrativen Registern, die an die statistischen Ämter übermittelt wurden, wegen des Statistikgeheimnisses nicht mehr an die Administration zurückgegeben werden dürfen, beispielsweise um die Qualität der Ausgangsregister zu verbessern. Innerhalb der Statistik aufgedeckte Fehler in diesen Ausgangsregistern blieben daher auch bei jeder weiteren Lieferung an die statistischen Ämter erhalten. Daraus ergab sich für die Zusammenführung der Register die Aufgabe, nicht ausschließlich die jeweils singulären Fehler zu korrigieren, sondern vielmehr systematisch sicherzustellen, dass auch bei künftigen Datenlieferungen diese Fehler im Sinne der Effizienz und der Datenintegrität identisch und automatisiert zu korrigieren waren.

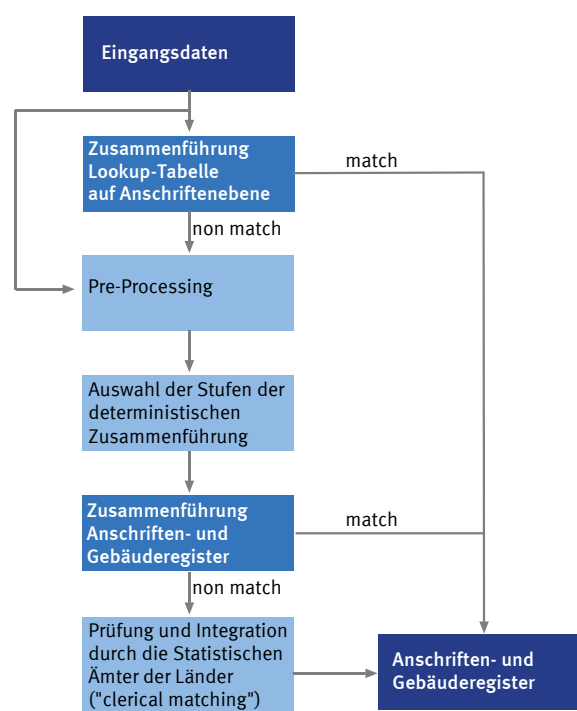
Zum Aufbau des Anschriften- und Gebäuderegisters sowie für dessen Aktualisierung und Vervollständigung wurden überwiegend deterministische Record-Linkage-Methoden genutzt. Deterministisches Record Linkage lässt sich beschreiben als Identitätsbestimmung eines Objekts in unterschiedlichen Dateien durch die exakte Übereinstimmung der Zusammenführungsvariablen.¹⁵ Diese wurde nach dem Aufbau des Anschriften- und Gebäuderegisters durch die Anschrift dargestellt, die als Kombination der Merkmale Amtlicher Gemeindeschlüssel, Postleitzahl, Ortsteil, Straße, Hausnummer, Hausnummernzusätze als eindeutige Zusammenführungsvariable verwendet wurde. Beim Record Linkage konnten verschiedene Teilmengen entstehen, die im Folgenden erläutert werden sollen. Durch ein „Match“ wird die Identität des gleichen Objekts in unterschiedlichen Dateien festgelegt. Kann ein Objekt nicht in anderen Dateien gefunden werden, so ist dies ein „Non-Match“. Ist das Objekt in beiden Dateien vorhanden und konnte fälschlicherweise nicht zugeordnet werden, so wird dieses als „False Non-Match“ klassifiziert. Ein „False Match“ bezeichnet ein Match zwischen zwei unterschiedlichen Objekten.

Um die Rate der False Matches auf ein Minimum zu reduzieren und damit dem amtlichen Qualitätsanspruch der Genauigkeit gerecht zu werden, wurden bei der Vervollständigung und Aktualisierung des Anschriften- und Gebäuderegisters ausschließlich deterministische Record-Linkage-Verfahren genutzt.¹⁶ Jede falsch zugeordnete Melderegisteranschrift hätte beispielsweise auf Personenebene zu Fehlern bei der Feststellung von Über- oder Untererfassungen der Melderegister durch die Haushaltestichprobe führen können.¹⁷

Akzeptierter Nebeneffekt des deterministischen Record-Linkage-Verfahrens war die hohe Rate an False Non-Matches.¹⁸ Deterministisches Record Linkage setzt generell eine sehr hohe Qualität der Zusammenführungsvariablen voraus, da bereits geringe Abweichungen in Schreibweisen sowie fehlende oder unkorrekte Angaben zu Non-Matches führen.

In Schaubild 1 wird das Schema der deterministischen Zusammenführung mit dem Anschriften- und Gebäuderegister dargestellt. Die einzelnen Schritte sollen im Folgenden erläutert werden.

Schaubild 1 Ablaufschema der deterministischen Zusammenführung (Record Linkage) mit dem Anschriften- und Gebäuderegister



2012 - 01 - 0913

Bei Registerintegrationen, wie beispielsweise der Melderegisterdaten 2011, war es im Vorfeld möglich, auf vorhandene anschriftenscharfe Zusammenführungsergebnisse vorhergegangener Registerintegrationen zurückzugreifen, die als Lookup-Tabelle fungieren konnten. Mit diesen Tabellen konnte vor dem Pre-Processing¹⁹ ein Match zwischen Eingangsdaten und Anschriften- und Gebäuderegister hergestellt werden. Standen keine Ergebnisse vorheriger Zusammenführungen zur Verfügung, so wurde dieser Schritt ausgelassen. Das Pre-Processing war die Voraussetzung für ein deterministisches Record Linkage über die Anschrift mit hohen Matching-Raten. Ein besonderes Augenmerk galt den Variablen, die vorwiegend heterogen vorlagen, wie der Straßennamen, die Ortsteile und die Hausnummernzusätze. Es folgten die Auswahl der Zusammenführungsstufen und

14 Bereits im Zensus 2001 wurde festgestellt, dass aufgrund der unterschiedlichen Darstellungsweisen von Hausnummernzusätzen in den geprüften Daten der Bundesagentur für Arbeit und den Melderegistern keine vollständige Deckungsgleichheit zu erzielen war. Probleme bereiteten ebenfalls die unterschiedlichen Zeitstände der Daten; siehe Statistische Ämter des Bundes und der Länder (Fußnote 4), hier: Seite 822 f.

15 Siehe Herzog, T./Scheuren, F./Winkler, W.: "Data Quality and Record Linkage Techniques", 2007, Seite 82.

16 Beim Aufbau des Anschriften- und Gebäuderegisters wurden auch nicht deterministische Record-Linkage-Verfahren genutzt. Die erzeugten Ergebnisse mussten aufgrund der Genauigkeitsanforderungen des Projekts aufwendig manuell nachbearbeitet beziehungsweise kontrolliert werden. Siehe Kleber, B./Maldonado, A./Scheuregger, D./Ziprik, K. (Fußnote 12), hier: Seite 633 ff.

17 Die Zuordnung von Personendaten wird über die Anschriften des Anschriften- und Gebäuderegisters gesteuert. Die Über- und Untererfassung der Melderegister wird über die Stichprobe erfasst, deren Erhebungseinheit die Anschrift ist. Ein Fehler bei der Anschriftenzuordnung kann damit zu einer falschen Betrachtung der Kartelleichen und Fehlbestände an dieser Anschrift führen.

18 Siehe Herzog, T./Scheuren, F./Winkler, W. (Fußnote 15), hier: Seite 82 f.

19 Pre-Processing wird im Folgenden zusammengefasst als Datenaufbereitung und Datenbereinigung; siehe Kleber, B./Maldonado, A./Scheuregger, D./Ziprik, K. (Fußnote 12), hier: Seite 635 ff.

die Zusammenführung mit dem Anschriften- und Gebäuderegister. Alle Datensätze, die nach den Zusammenführungskriterien als identisch angesehen wurden, wurden in das Anschriften- und Gebäuderegister übernommen. Aufgrund der heterogenen Ausgangsqualität, beispielsweise durch fehlende beziehungsweise falsche Werte in einzelnen Variablen, wurden alle Record-Linkage-Verfahren im Bereich des Anschriften- und Gebäuderegisters mehrstufig aufgebaut. Alle Non-Matches aus der ersten Stufe wurden mit reduzierten Zusammenführungsvariablen in einer weiteren Stufe mit dem Anschriften- und Gebäuderegister abgeglichen. Dieses Vorgehen wurde wiederholt, bis die letzte Stufe erreicht wurde. Durch die stufenabhängige Herabsetzung der Differenzierungskraft der Variablenkombination „Anschrift“ wurde angenommen, dass die Matching-Rate erhöht wird. Fehler, besondere Schreibweisen oder fehlende Werte in einzelnen Variablen, die bei der ersten Zusammenführung zu einem False Non-Match führten, konnten eindeutig zusammengeführt werden, wenn die qualitativ schlechte Variable der Kombination ausgelassen wurde. Andererseits konnte durch die gleichzeitig herbeigeführte schlechtere Differenzierungskraft der Zusammenführungsvariablen auch die Anzahl der False Matches erhöht und damit das Ergebnis der Anbindung qualitativ verschlechtert werden.²⁰ Die zunächst sehr strikt angewandten Kriterien führten in der ersten Stufe, abhängig von der Ausgangsdatenqualität, in der Regel zu sehr wenigen, aber qualitativ hochwertigen Matches. Die Reihenfolge der verwendeten Variablenkombination ergab sich aus den zuvor gesammelten Erfahrungen mit der Variablenqualität aus dem Zensusstest 2001 und dem Aufbau des Anschriften- und Gebäuderegisters.²¹ Um eine vollständige Integration aller False Non-Matches zu erreichen, wurden alle Non-Matches nach gesetzlicher Grundlage im Zensusgesetz 2011 und Zensusvorbereitungsgesetz 2011 durch die Statistischen Ämter der Länder geprüft beziehungsweise bearbeitet. Vale (siehe Fußnote 22) verweist darauf, dass die Vorteile einer solchen manuellen Integration beziehungsweise Prüfung („clerical matching“) darin bestehen, dass die zu bearbeitenden Datensätze mit intelligentem menschlichen Input betrachtet werden anstatt mit maschineller Logik. Die Nachteile der manuellen Bearbeitung lägen hingegen in den Bereichen der Reproduzierbarkeit, Kosteneffizienz und des Zeitaufwands.²² Weiterhin konnte eine unsystematische manuelle Bearbeitung zu Inkonsistenzen im Gesamtbestand führen. Demnach ist die Entscheidung Match oder Non-Match oftmals individuell, sofern keine weiteren Informationen zur Klärung der Daten herangezogen werden können.

3 Aufgaben des Anschriften- und Gebäuderegisters im Zensus 2011

Die qualitativen Anforderungen an ein statistisches Register sind andere als die, die an eine regelmäßige Primärstatistik gestellt werden.²³ Grundlegend für eine Qualitätsbewertung

des Anschriften- und Gebäuderegisters müssen deshalb die Aufgaben sein, die an das statistische Register gestellt werden.²⁴ Diese werden im Folgenden beschrieben.

Das Anschriften- und Gebäuderegister diente der Vorbereitung sowie als Auswahlgrundlage der im Zensus 2011 durchgeführten Haushaltebefragung auf Stichprobenbasis. Im § 5 Absatz 4 Zensusvorbereitungsgesetz 2011 ist festgelegt, dass bestimmte Angaben aus dem Anschriften- und Gebäuderegister für die Erstellung einer separat zu führenden Stichprobenorganisationsdatei vorzuhalten waren. Diese Daten wurden genutzt, um die Methode der Stichprobenziehung zu optimieren und damit an den aktuellen Datenbestand anzupassen.²⁵ Als Auswahlgrundlage für die Stichprobenziehung legte das Anschriften- und Gebäuderegister den Grundstein für die Korrektur von Über- und Untererfassungsfehlern in den Melderegistern durch die Haushaltebefragung auf Stichprobenbasis. Fehlbestände konnten nur durch die Befragung sämtlicher Personen an einer Anschrift und den Abgleich mit dem Melderegister aufgedeckt werden. Die Auswahlgrundlage wurde über alle Anschriften mit Wohnraum im Anschriften- und Gebäuderegister gebildet.²⁶ Anschriften, die als „kein Wohnraum“ klassifiziert wurden, wurden über die im Register hierzu hinterlegten Merkmale ausgesteuert. Um die zeitliche Veränderung des Anschriftenbestands des Anschriften- und Gebäuderegisters (Neuaufnahmen) zwischen der Stichprobenziehung am 1. September 2010 und dem Zensusstichtag 9. Mai 2011 zu berücksichtigen, wurden ergänzende Stichproben gezogen, sodass alle Anschriften, die der Auswahlgrundlage der Haushaltebefragung zuzurechnen waren, zum Zensusstichtag eine Auswahlchance hatten.²⁷

Neben der Koordination der Haushaltstichprobe diente das Anschriften- und Gebäuderegister der Ablaufkontrolle aller im Zensus durchgeführten Primärstatistiken sowie der Steuerung der Gebäude- und Wohnungszählung.²⁸ Es stellte im Rahmen der Erhebungsvorbereitung die zu befragenden Anschriften bereit. Sonderanschriften wurden ebenso wie die Anschriften der Haushaltstichprobe im Anschriften- und Gebäuderegister ermittelt und markiert. Durch die direkte Gewinnung der Sonderanschriften und Stichprobenanschriften aus dem Anschriften- und Gebäuderegister waren diese bereits mit dem eindeutigen Anschriftenidentifikator der AGR-Datenbank versehen und konnten damit auf Anschriftenebene mit den anderen Erhebungsteilen in Bezug gesetzt werden. Mittels einer Kennzeichnung der Anschriften von Gebäuden im Wohnraum im Anschriften- und Gebäuderegister wurde auch die Auswahlgrundlage der Gebäude- und Wohnungszählung, die als postalische Befragung der Gebäudeeigentümer und Gebäudeeigentümerinnen durchgeführt wurde, festgelegt. Für alle entsprechend gekennzeichneten Anschriften mussten die jeweiligen

20 Siehe Herzog, T./Scheuren, F./Winkler, W. (Fußnote 15), hier: Seite 83.

21 Siehe Kleber, B./Maldonado, A./Scheuregger, D./Ziprik, K. (Fußnote 12), hier: Seite 635 ff., und Statistische Ämter des Bundes und der Länder (Fußnote 4), hier: Seite 822 f.

22 Siehe Vale, S. (Fußnote 9), hier: Kapitel 5.5.

23 Siehe Wallgren, A./Wallgren, B. (Fußnote 13).

24 Die Aufgaben des Anschriften- und Gebäuderegisters werden im § 2 Zensusvorbereitungsgesetz 2011 beschrieben.

25 Die Stichprobenorganisationsdatei wurde nach § 15 Absatz 2 Zensusvorbereitungsgesetz 2011 zum Zensusstichtag gelöscht.

26 Siehe § 2 Absatz 3 Ziffer 1 der Verordnung über Verfahren und Umfang der Haushaltebefragung auf Stichprobenbasis zum Zensusgesetz 2011 (Stichprobenverordnung Zensusgesetz 2011 – StichprobenV) vom 25. Juni 2010 (BGBl. I Seite 830).

27 Siehe Fußnote 26 sowie Berg, A./Bihler, W.: „Das Stichprobendesign der Haushaltstichprobe des Zensus 2011“ in WiSta 4/2011, Seite 317 ff., hier: Seite 327 ff.

28 Siehe § 2 Absatz 2 Ziffer 1 Zensusvorbereitungsgesetz 2011.

Gebäudeeigentümer und Gebäudeeigentümerinnen ermittelt werden.

Die Zusammenführung der Stichprobenergebnisse mit den Melde- und Erwerbsregistern, den Daten der Gebäude- und Wohnungszählung und dem Sonderanschriftenregister erfolgte im ersten Schritt anschriftenweise über die Angaben des Anschriften- und Gebäuderegisters. Angebunden und miteinander verknüpft wurden damit alle primärstatistischen Erhebungsteile (Haushaltebefragung auf Stichprobenbasis, Gebäude- und Wohnungszählung, Befragungen an Sonderbereichen) und alle Registerangaben (Erwerbs- und Melderegister). Die Kombination der Daten der Gebäude- und Wohnungszählung mit den Personenangaben erfolgt ebenfalls anschriftenweise im Rahmen einer Haushaltegenerierung. Die Anschrift beziehungsweise der Anschriftenidentifikator konnte bei Zusammenführungen auf Personenebene als Blockingvariable genutzt werden. Das bedeutet, dass die Menge der zu vergleichenden Personen zweier Erhebungsteile – beispielsweise der Melde- und der Haushaltebefragung – auf genau die Personen reduziert wurde, die an einer Anschrift gemeldet waren beziehungsweise die an dieser Anschrift befragt wurden. Da das Anschriften- und Gebäuderegister für die Erhebungsteile eine Referenz bildete, konnte durch Vergleiche mit dem Anschriften- und Gebäuderegister die Überprüfung der Gebäude- und Wohnungszählung, der Sonderbereiche und der Haushaltebefragung auf Vollständigkeit in ihren Ergebnissen beziehungsweise ihrer Zielpopulation prozessgestützt erfolgen. So mussten alle Datensätze, die bisher keine Angaben zu einem Eigentümer beziehungsweise einer Eigentümerin enthielten oder für die bisher keine Erhebungsbefunde vorlagen, nachrecherchiert beziehungsweise nachgereicht werden.

Eine weitere Aufgabe des Anschriften- und Gebäuderegisters war es, die Konsistenz zwischen den Daten aus den verschiedenen Erhebungsteilen zu sichern. Zum einen konnten durch neue Informationen aus den Erhebungsteilen Änderungen in Bezug auf die Anschrift und die Rolle der Anschrift im Zensus (zum Beispiel „Wohnraum“, „kein Wohnraum“) auftreten. Zum anderen konnten die Erhebungsergebnisse der Primärstatistik genutzt werden, um erhebungsteilübergreifende Konsistenzprüfungen durchzuführen. Die Datensynchronisation wurde primär dadurch gesichert, dass die relevanten Kennzeichen durch die Erhebungsteile in das Anschriften- und Gebäuderegister selbst eingespeist wurden. Wurde eine Anschrift beispielsweise zur Sonderanschrift mit besonderem Schutzbedarf, so durfte sie nicht mehr in der Haushaltebefragung erhoben werden. Die Auswahlgrundlage der Haushaltebefragung reduzierte sich damit. So konnte jeder Erhebungsteil über das Abrufen der Kennzeichen die sich ständig ändernde Referenzpopulation abfragen. Erhebungsteilübergreifende Konsistenzprüfungen im Anschriften- und Gebäuderegister wurden über seine Schnittstellenfunktion ermöglicht. Über die im Anschriften- und Gebäuderegister vorhandenen anschriftenbezogenen Ergebnisse der Erhebungsteile konnten Fehler aufgedeckt werden, die mit der alleinigen Betrachtung eines Registers oder einer Stichprobenerhebung nicht erkannt werden konnten. So war es die Referenz zu einer weiteren Datenquelle, die durch neue Informationen die Ergebnisse eines

Erhebungsteils in eine andere Relation setzen konnte.²⁹ Zu Konsistenzproblemen konnte es im Anschriften- und Gebäuderegister kommen, sobald zwei Erhebungsteile für ein und dieselbe Anschrift sich widersprechende Angaben enthielten, beispielsweise wenn durch die Haushaltebefragung ein Ausfall (etwa durch eine Baulücke) einer Anschrift und durch die Gebäude- und Wohnungszählung ein Gebäude mit Wohnraum ermittelt wurde. Für die meisten Fallkonstellationen, die zu Inkonsistenzen führten, wurden festgelegte Regeln im Anschriften- und Gebäuderegister oder der nachfolgenden Haushaltegenerierung implementiert. Für weitere Fallkonstellationen wurden entweder Überprüfungen durch die Statistischen Ämter der Länder oder auch primärstatistische Erhebungen durchgeführt, wie zum Beispiel die Befragung zur Klärung von Unstimmigkeiten nach § 16 Zensusgesetz 2011. Hierbei wurden Anschriften betrachtet, die nur eine bewohnte Wohnung aufwiesen und in Gemeinden mit weniger als 10 000 Einwohnern lagen. Traten stark abweichende Personenanzahlen je Anschrift zwischen Melderegister und der Gebäude- und Wohnungszählung auf, wurde die Anschrift durch einen Interviewer oder eine Interviewerin erhoben. Das anschriftenbezogene Ergebnis der Befragung wurde anschließend in das Anschriften- und Gebäuderegister aufgenommen und konnte somit weitere Indizien für die Rolle der Anschrift liefern, zum Beispiel ob „Wohnraum“ oder „kein Wohnraum“ vorlag.

Eine weitere Aufgabe des Anschriften- und Gebäuderegisters ist die Schaffung einer Grundlage für Auswertungen auf kleinräumiger Ebene.³⁰ Unter kleinräumigen Gliederungen sind geografisch spezifizierte Gebiete unterhalb beziehungsweise losgelöst von der Gemeindeebene zu verstehen. Im Anschriften- und Gebäuderegister können unterschiedliche Gliederungssysteme gespeichert werden, wie die Bundestagswahlkreise, Stadtteile, Gliederungssysteme der Kommunen oder Gliederungssysteme für Sonderauswertungen. Die Vorhaltung jedweder Daten im Anschriften- und Gebäuderegister setzt deren Integration voraus. Für Gemeinden bilden kleinräumige Gliederungssysteme eine Möglichkeit, eine bedarfsgerechte Datengrundlage aus dem Zensus 2011 zu gewinnen. Diese Zahlen können vor allem im Rahmen der Stadtentwicklung, beispielsweise für Bedarfsplanungen oder Prognosemodelle, genutzt werden.

4 Qualitätsanforderungen und Qualitätssicherungsmaßnahmen

Im Mittelpunkt der zu betrachtenden Qualitätsanforderungen an das Anschriften- und Gebäuderegister und der getroffenen Maßnahmen zur Qualitätssicherung müssen diejenigen Aspekte stehen, die sich aus den oben beschriebenen Aufgaben ableiten lassen. Um als Referenzpopulation der Erhebungsteile des Zensus 2011 genutzt werden zu können, musste das Anschriften- und Gebäuderegister vollständig in Bezug auf seine Zielpopulation und korrekt in Bezug auf das Merkmal „Wohnraum“ sein. Der hohe Qualitätsanspruch an das Merkmal „Wohnraum“ veranlasste den Gesetzgeber dazu, in § 14 Zensusgesetz 2011 festzulegen,

²⁹ Siehe Wallgren, A./Wallgren, B. (Fußnote 13), hier: Seite 100 ff.

³⁰ Siehe § 2 Absatz 2 Ziffer 4 Zensusvorbereitungsgesetz 2011.

dass Adressen, deren Wohnraumstatus unklar war, in den Statistischen Ämtern der Länder zu überprüfen waren. Weiterhin musste das Adressen- und Gebäuderegister im Hinblick auf die getätigten Datenzusammenführungen aktuell sein, um keine False Matches oder False Non-Matches zu verursachen. Zudem wurden die aktuellen Adressen-schreibweisen für die Vorbereitung der primärstatistischen Erhebungen benötigt.

4.1 Homogenität und Redundanzfreiheit

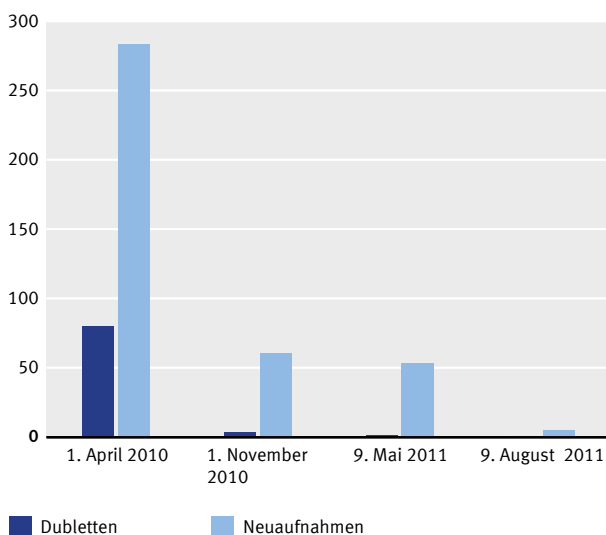
Durch die dezentrale Organisation und die Unabhängigkeit der administrativen Register voneinander werden entstehende Änderungen an Adressen nicht zum gleichen Zeitpunkt in allen Registern festgehalten. Weiterhin gibt es keine verbindliche Form der Adressendarstellung in den Register führenden Stellen mit der Konsequenz, dass Adressen innerhalb und zwischen den Registern nicht einheitlich dargestellt werden. Die beschriebene deterministische Methode der Datenzusammenführung basiert auf der Annahme, dass die Ausgangsqualität der Zusammenführungsvariablen sehr hoch ist. Um die Heterogenität der Adressenschreibweisen zu harmonisieren, erfolgte eine intensive Arbeit im Bereich des Pre-Processings. In der Phase des AGR-Aufbaus wurden die Adressen aus den Registern anhand einer einfachen maschinellen Standardisierung sowie durch mehrjährige manuelle Korrektur- und Prüfarbeiten von etwa 2,7 Millionen Adressen in den Statistischen Ämtern der Länder harmonisiert.³¹ Um die Standardisierung im Bereich der Straßennamen kosteneffizienter, schneller und konsistent zu gestalten, wurde eine Lookup-Tabelle als Teil des Straßenthesaurus aufgebaut, die das gesammelte Wissen der Vorarbeiten enthielt. Diese Lookup-Tabelle konnte bei nachfolgenden Datenanbindungen an das Adressen- und Gebäuderegister genutzt werden.

Lagen die Adressen unerkannt in heterogenen Schreibweisen vor, konnte dies zur redundanten Erfassung des gleichen Objekts im Adressen- und Gebäuderegister führen. Als Folge für die Erhebungsteile wäre hierdurch ihre Referenzpopulation nicht korrekt erfasst worden. Zusammenführungen mit einem Datenbestand, der ein identisches Objekt mehrfach in unterschiedlichen Darstellungen enthält (Dublette), hätten zu False Matches führen können und beispielsweise im Bereich der Haushalbefragung zu einem Erhebungsausfall der Adresse. Das Qualitätsproblem der Redundanz konnte durch eine mangelnde Datenqualität und eine gleichzeitig nicht greifende Standardisierung auftreten. Um die Qualität des Adressen- und Gebäuderegisters zu sichern, musste es um alle Dubletten bereinigt werden. Um die Redundanzfreiheit des Adressen- und Gebäuderegisters zu sichern, wurden daher alle Adressenvariablen systematisch im Statistischen Bundesamt und in den Statistischen Ämtern der Länder auf Dubletten hin untersucht. Die durchgeführte Dublettenprüfung im Statistischen Bundesamt berücksichtigte dabei insbesondere die Variablen, die den Erfahrungen nach am häufigsten heterogenen Darstellungsformen unterlagen, wie Hausnummernzusätze, Ortsteile und Straßennamen. Ziel war es, alle

potenziell mehrfach vorhandenen Adressen an die Statistischen Ämter der Länder zur manuellen Überprüfung und Korrektur zu übermitteln. Um die zu prüfende Menge handhabbar zu machen, wurden sehr strikte Matching-Kriterien angesetzt, die insgesamt wenige, aber dafür wahrscheinlich identische Adressenpaare verknüpften. Der AGR-Datenbestand wurde für jede der Prüfungen mit sich selbst über die Adresse zusammengeführt, und zwar jeweils mit Anpassungen in den untersuchten Variablen. Für die Prüfung der Straßennamen wurde beispielsweise ein Teilstringvergleich durchgeführt. Für die Prüfung des Ortsteils wurde unter anderem untersucht, ob dieser Teil des Straßennamens ist oder in anderer Schreibweise, zum Beispiel in abgekürzter Form, vorlag. Die Hausnummernzusätze wurden sowohl auf unterschiedliche Darstellungsweisen, wie zum Beispiel 1a und 1/1, hin untersucht als auch auf die Abbildung der Zusätze in falschen Feldern, zum Beispiel innerhalb der Variablen „Hausnummer“ anstatt in der Variablen „Hausnummernzusatz“.

Schaubild 2 zeigt die identifizierten Dubletten aus den Melderegisterintegrationen 2010 bis 2011. Hier werden die Neuaufnahmen aufgrund der Melderegisterintegrationen in Kombination mit den hiervon später als Dubletten identifizierten Fällen dargestellt. Insbesondere bei der Integration der Melderegisterlieferung zum 1. April 2010 trat ein in Relation zu den weiteren Registerintegrationen hoher Anteil an später identifizierten Dubletten auf.

Schaubild 2 Erkannte Dubletten aus den Melderegisterintegrationen 2010 und 2011
1 000



2012 - 01 - 0914

Dieser hohe Anteil ist damit zu erklären, dass diese Registerlieferung die erste registerbasierte Aktualisierung des Adressen- und Gebäuderegisters seit seinem Aufbau mit den Registerlieferungen vom 1. April 2008 darstellte. Die Unbeständigkeit der Adresse als Merkmalskombination lässt darauf schließen, dass viele Neuaufnahmen der Registerlieferung 1. April 2010 bereits im Adressen- und Gebäuderegister vorhanden waren, allerdings mit einer

³¹ Siehe Maldonado, A./Scheuregger, D./Ziprik, K. (Fußnote 12), hier: Seite 5.

Tabelle 1 Anbindungsquoten der kommunalen kleinräumigen Gliederung nach Anbindungsstufen

| | Qualitätskennzeichen | Angebundene Anschriften ¹ | |
|------------------------|----------------------|--------------------------------------|------------------------------|
| | | Anzahl | Anbindungsquote ¹ |
| Anbindungsstufe 1 ... | AAAAAA | 828 910 | 15,9 |
| Anbindungsstufe 2 ... | AAABAA | 1 627 760 | 31,3 |
| Anbindungsstufe 3 ... | AAOAAA | 1 413 237 | 27,2 |
| Anbindungsstufe 4 ... | AAOBAA | 958 700 | 18,4 |
| Anbindungsstufe 5 ... | AAAAAO | 15 480 | 0,3 |
| Anbindungsstufe 6 ... | AAABAO | 75 550 | 1,5 |
| Anbindungsstufe 7 ... | AAOAAO | 17 992 | 0,3 |
| Anbindungsstufe 8 ... | AAOBAO | 45 818 | 0,9 |
| Anbindungsstufe 9 ... | AAAAOO | 61 818 | 1,2 |
| Anbindungsstufe 10 ... | AAABOO | 6 847 | 0,1 |
| Anbindungsstufe 11 ... | AAOAOO | 52 232 | 1,0 |
| Anbindungsstufe 12 ... | AAOBOO | 6 512 | 0,1 |

¹ Auswertung zum 30. August 2012. Anzahl der insgesamt zu berücksichtigenden Anschriften in den zu befüllenden Gemeinden = 5 198 063. Anbindung auf Anschriftenebene.

hat, kann aufgrund der Vermischung von Harmonisierung und Aktualisierung der Daten aber nicht eindeutig ermittelt werden.

4.3 Vollständigkeit

Um als Referenz der Vollständigkeit für die Erhebungsteilfunktionen zu können, muss das Anschriften- und Gebäuderegister auch selbst diesem Kriterium entsprechen. Das Anschriften- und Gebäuderegister basiert auf administrativen Registern, die originär anderen Zwecken als der amtlichen Statistik dienen. Als Nachteil der administrativen Registernutzung wird aus diesem Grund häufig die Untererfassung der zu ermittelnden Zielpopulation genannt.³⁷ Die zum Aufbau und zur Aktualisierung des Anschriften- und Gebäuderegisters genutzten Register konnten – jeweils für sich betrachtet – die Definition der Vollständigkeit des Anschriften- und Gebäuderegisters nicht erfüllen. Erst durch die Kombination der einzelnen Register konnte diese erreicht werden. Die Untererfassung der Register bezog sich dabei auf ihre Variablenvollständigkeit und die Populationsvollständigkeit. Um die Variablenvollständigkeit herzustellen, wurden die Register miteinander verschnitten und die neben der Anschrift nach § 2 Absatz 3 Zensusvorbereitungsgesetz 2011 zu speichernden Merkmale in der AGR-Datenbank abgelegt. So konnten beispielsweise Strukturmerkmale zur Beschäftigungsstruktur aus den Daten der Bundesagentur für Arbeit gewonnen werden und Koordinaten zur Georeferenzierung der Anschriften aus den Georeferenzierten Adressdaten – Bund. Weitere Merkmale wurden in der Zensusdurchführungsphase durch die Erhebungsteile, zum Beispiel die „Art der Einrichtung“ aus der Erhebung an Anschriften mit Sonderbereichen, gewonnen.

Die Populationsvollständigkeit bezieht sich auf die Vollständigkeit der zu erfassenden Anschriften insgesamt. Sie ist ausgerichtet an der Instabilität der Anschrift selbst. Die räumliche Instabilität der Anschrift (Neubauten, Abrisse

und Umnutzungen³⁸) konnte aufgrund des Mangels an Erfahrungen in diesem Bereich beziehungsweise des Mangels an einem flächendeckenden Register kaum abgeschätzt werden. Die Vervollständigung des Anschriften- und Gebäuderegisters zum Stichtag 9. Mai 2011 wurde durch die Integration dreier unterschiedlicher Datenquellen durchgeführt: administrative Registerlieferungen (Melderegister und Daten der Vermessungsbehörden), primärstatische Erhebungsteile und Erkenntnisse aus weiteren Quellen. Zum einen waren also bisher im Anschriften- und Gebäuderegister nicht erfasste Gebäude mit Wohnraum durch die Melderegisterintegrationen 2010 und 2011 neu aufzunehmen (Neuaufnahme), zum anderen wurden bereits im Register vorhandene Anschriften zu Anschriften „ohne Wohnraum“ beziehungsweise Anschriften „mit Wohnraum“ umklassifiziert. Umklassifizierungen von Gebäuden konnten ausschließlich aus Erkenntnissen aus den primärstatischen Erhebungsteilen, anderen Quellen oder als Ergebnis einer Prüfung (zum Beispiel als Ergebnis der Anschriftenprüfung nach § 14 Zensusgesetz 2011) gewonnen werden, da in den gesetzlich geregelten Lieferungen der administrativen Register keine weiteren Informationen zur Gebäudenutzungsart vorgesehen waren. Aus dem Anschriften- und Gebäuderegister kann entnommen werden, dass insgesamt 2 410 380 Anschriften in ihrer Wohnraumklassifizierung auf „kein Wohnraum“ verändert wurden. Das sind zum derzeitigen Stand etwa 11,1 % der AGR-Anschriften insgesamt. Die Klassifizierung der Anschrift auf „mit Wohnraum“ war bereits im Vorhinein durch die Melderegisterintegration abgedeckt, da alle zu Wohnraum umgenutzten Gebäude, die bewohnt wurden, in den Melderegister-Daten enthalten sein mussten.

Neben den regelmäßigen Integrationen der Register der Meldebehörden und der Vermessungsbehörden wurden auch die Primärstatistiken des Zensus genutzt, um das Anschriften- und Gebäuderegister in seiner Zielpopulation zu vervollständigen. Seit April 2009 wurde im Rahmen der Gebäude- und Wohnungszählung mit der Ermittlung der Auskunftspflichtigen begonnen.³⁹ Dazu wurden die Objektschriften der Gebäude- und Wohnungszählung mit den Anschriften des Anschriften- und Gebäuderegisters abgeglichen. Recherchierte Eigentümer und Eigentümerinnen, für die es im Anschriften- und Gebäuderegister noch keine Gebäudeanschrift gab, konnten durch die Statistischen Ämter der Länder neu aufgenommen werden. Neben Informationen, die durch die Datenbestände der Gebäude- und Wohnungszählung hinzukamen, wurden auch Daten aus der Ermittlung von Sonderbereichen an das Anschriften- und Gebäuderegister angebunden und trugen mit gleichem Ablauf zur Vervollständigung der Datenbank bei. Eine weitere Möglichkeit, das Anschriften- und Gebäuderegister zu vervollständigen, bot sich durch das Einbeziehen weiterer Informationsquellen, wie zum Beispiel von Daten der für die Führung des Liegenschaftskatasters zuständigen Stellen. Diese Stellen können frühzeitig Informationen über Neubaugebiete und Abrissgebäude haben, was insbesondere

³⁸ Eine Umnutzung beschreibt nach Landesbauordnung den Vorgang der Veränderung der Nutzungsart eines Gebäudes. Im Zensus ist dies relevant, wenn ein Gebäude an einer Anschrift von einem Gebäude mit Wohnraum zu einem ohne Wohnraum wird (zum Beispiel Gebäudeabriss, Umnutzung zu einem Bürogebäude) oder umgekehrt (zum Beispiel Umnutzung eines Bürogebäudes zu einem Wohngebäude).

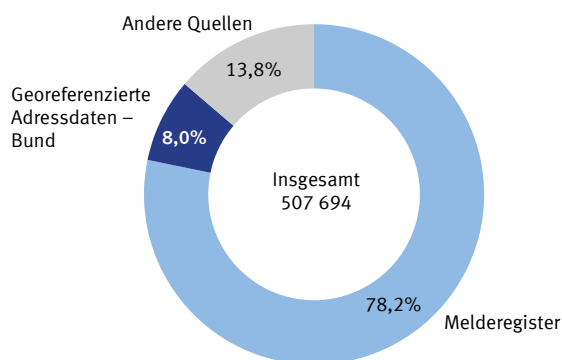
³⁹ Siehe § 10 Zensusvorbereitungsgesetz 2011.

³⁷ Siehe Fußnote 3.

die Möglichkeit eröffnete, anfallende Umnutzungen in das Anschriften- und Gebäuderegister einzuarbeiten.

Schaubild 3 zeigt die Neuaufnahmen im Anschriften- und Gebäuderegister unterteilt nach Ausgangsdatenquellen (Registerdaten, Primärstatistik und andere Quellen).

Schaubild 3 Neuaufnahmen im Anschriften- und Gebäuderegister nach ihrer Herkunft September 2010 bis Juli 2011



2012 - 01 - 0915

Der Großteil der Neuaufnahmen ist auf die Integration der administrativen Register zurückzuführen. Nur ein geringer Anteil von 13,8% der Neuaufnahmen im Zeitraum September 2010 bis Juli 2011 erfolgte aufgrund von Erkenntnissen aus anderen Erhebungsteilen oder Erkenntnissen aus anderen Informationsquellen.

4.4 Korrektheit

Die Eigenschaft „Wohnraum“ oder „kein Wohnraum“ beeinflusst maßgeblich die Rolle der Anschrift im Zensus. Die Erhebungsteile bildeten ihre Auswahlgrundlage im ersten Schritt basierend auf der Wohnraumklassifikation. Damit das Anschriften- und Gebäuderegister sowohl als Steuerungsinstrument als auch als Auswahlgrundlage der Primärstatistik fungieren konnte, musste seine Korrektheit in Bezug auf das Merkmal „Wohnraum“ gesichert sein.

Die maschinelle Wohnraumklassifizierung erfolgte bei der Registerintegration immer dann, wenn mindestens zwei Register (Daten der Bundesagentur für Arbeit, Georeferenzierte Adressdaten – Bund und Melderegister) dieselbe Anschrift in ihrem Bestand aufwiesen. Der Großteil der Anschriften des Anschriften- und Gebäuderegisters wurde auf diesem Wege maschinell hinsichtlich des Merkmals Wohnraum klassifiziert. Basierend auf manuellen Prüfungen der False Non-Matches und Non-Matches der Registerzusammenführungen als auch aufgrund weiterer Erkenntnisse (zum Beispiel Umnutzungen) wurden ländersseitig 15,9% der AGR-Anschriften klassifiziert oder gegebenenfalls umklassifiziert. False Matches der deterministischen Registerzusammenführung oder auch falsche Klassifikationen der manuellen Bearbeitung konnten zu falsch klassifizierten Anschriften im Anschriften- und Gebäuderegister führen. Klumpungen von falschen Klassifikationen konnten bei der

maschinellen Zusammenführung in den Matches zwischen Anschriften aus den Georeferenzierten Adressdaten – Bund und dem Register der Bundesagentur für Arbeit gefunden werden. In den Daten der Bundesagentur für Arbeit wurden auch Anschriften des Arbeitgebers (zum Beispiel Bürogebäude, landwirtschaftliche Gebäude) verwendet, sodass ein technisch korrektes, inhaltlich aber falsches Ergebnis bei der Datenzusammenführung mit den Georeferenzierten Adressdaten – Bund erzielt wurde. Da keine Referenzdatei zu den im Zensus geforderten Definitionen der Anschriften mit Wohnraum in Deutschland vorlag, konnten während des Aufbaus des Anschriften- und Gebäuderegisters und dessen Aktualisierung die Ergebnisse nicht maschinell kontrolliert werden.⁴⁰

Mithilfe der primärstatistischen Erhebungsteile kann ein Teil der möglicherweise falsch als „Wohnraum“ klassifizierten Anschriften im Anschriften- und Gebäuderegister identifiziert werden. Anschriften, die als Ausfall in einer Erhebung klassifiziert wurden, wurden im Anschriften- und Gebäuderegister gekennzeichnet. Eine falsche Klassifikation als Anschrift „ohne Wohnraum“ geht in der Regel in keine weitere Primärstatistik ein und kann daher auch nicht durch diese geprüft werden. Daher wurde diese Klassifikation im Rahmen der Prüfung nach § 14 Zensusgesetz 2011 auch sehr restriktiv vergeben. Anschriftenbezogene Ergebnisse bezüglich einer falschen Klassifikation der Anschrift können aus allen Anschriften des Anschriften- und Gebäuderegisters ausgewertet werden, die primärstatistisch erhoben wurden. Das sind 9,0% der im Rahmen der Haushaltstestichprobe gezogenen AGR-Anschriften (einschließlich der Wiederholungsbefragung), 0,3% der AGR-Anschriften zur Befragung an Anschriften mit Sonderbereichen, 2,1% der AGR-Anschriften zur Befragung zur Klärung von Unstimmigkeiten und 89,8% der AGR-Anschriften, die im Rahmen der Gebäude- und Wohnungszählung zu bearbeiten waren. Zum Vergleich der AGR-Klassifikation sollen die beiden umfangreichsten Erhebungen, die Gebäude- und Wohnungszählung und die Befragung der Haushalte auf Stichprobenbasis, betrachtet werden.

Beide Primärstatistiken kommen in Bezug auf die Wohnraumklassifikation nicht immer zum gleichen Ergebnis. Bundesweit sind etwa 2,28% der GWZ-relevanten Stichprobenanschriften in der Gebäude- und Wohnungszählung als Wohnraum und stichprobenseitig als „kein Wohnraum“ (Ausfall) gekennzeichnet oder umgekehrt.⁴¹ Eine eindeutige Aussage über die Anteile der falsch klassifizierten Anschriften mit Wohnraum im Anschriften- und Gebäuderegister kann daher weder eindeutig mit den Ergebnissen der Gebäude- und Wohnungszählung noch mit den Ergebnissen der Haushaltbefragung getroffen werden. Für eine Bewertung des Anschriften- und Gebäuderegisters soll daher die Schnittmenge der übereinstimmenden Ergebnisse der

⁴⁰ So ist es beispielsweise im Rahmen der Gebäude- und Wohnungszählung möglich, für die Konsistenzprüfung auf Referenzdateien der Wohnungsbestandsfortschreibung zurückzugreifen.

⁴¹ Eine falsche Klassifikation des Wohnraummerkmals im Anschriften- und Gebäuderegister wird hier aus den Ausfallgründen der Haushaltbefragung „ausländische Streitkräfte“, „komplett gewerblich genutzte Anschrift“ und „Baulücke“ ermittelt. Die gewonnenen Angaben aus der Gebäude- und Wohnungszählung stammen aus dem Ergebnis der Gebäude- und Wohnungszählung zu „Wohnraum“ und „kein Wohnraum“. Zusätzliche Bedingung ist, dass im Anschriften- und Gebäuderegister „Wohnraum“ hinterlegt wurde.

Gebäude- und Wohnungszählung und der Haushaltebefragung genutzt werden. Die doppelte Bestätigung durch zwei verschiedene Primärstatistiken kann ein verifiziertes Abbild der potenziell falschen Klassifikation im Anschriften- und Gebäuderegister bieten. Die Anschriften mit falscher Klassifikation auf „Wohnraum“ haben bundesweit einen Anteil von etwa 0,40% an den GWZ-relevanten Stichprobenanschriften insgesamt. Wertet man diese Menge hinsichtlich des Grundes ihrer Klassifikation aus, so kommt man auf eine Quote von etwa 0,17% False Matches innerhalb der GWZ-relevanten Stichprobenanschriften, also falsch klassifizierter Anschriften aufgrund der maschinellen Zusammenführung.

5 Fazit

Die wohl größte Schwierigkeit bei der Qualitätsmessung des Anschriften- und Gebäuderegisters ist das Fehlen jeglicher Referenzregister. Einzelne Qualitätsmerkmale, wie die Korrektheit und die Redundanzfreiheit, konnten rückblickend über die Primärstatistiken beziehungsweise das Anschriften- und Gebäuderegister selbst näherungsweise bemessen werden. Im Falle der Korrektheit musste als Referenz die Schnittmenge der identischen Ergebnisse der Gebäude- und Wohnungszählung und der Haushaltebefragung herangezogen werden, da beide Erhebungsteile keine vollständig identischen Ergebnisse auswiesen. Die hieraus abgeleiteten Ergebnisse zeigen, dass bundesweit etwa 0,40% der GWZ-relevanten Stichprobenanschriften von der Wohnraumklassifizierung des Anschriften- und Gebäuderegisters abweichen. Bestimmt man in dieser Menge den Anteil der falschen Klassifikationen, die aufgrund maschineller Anbindungen (False Match) getroffen wurden, kommt man auf einen Anteil von etwa 0,17%.

Kritischer sehen die Werte für die im Anschriften- und Gebäuderegister erkannten Dubletten aus. So wären im Durchschnitt 21,2% aller Melderegisterneuaufnahmen mehrfach im Anschriften- und Gebäuderegister vorhanden, wenn keine Dublettenbereinigung durchgeführt worden wäre. Hieraus folgt, dass eine sehr strikt durchgeführte Dublettenprüfung des Anschriften- und Gebäuderegisters auch für ein mögliches künftiges Anschriften- und Gebäuderegister im Zensus 2021 eine zwingende und auszuweitende Maßnahme der Qualitätssicherung sein muss. Weitere Qualitätskriterien, wie die Vollständigkeit und Aktualität, können ohne eine Referenz nicht eindeutig bemessen werden. Die Aufgliederung der Anbindungsquoten der kommunalen kleinräumigen Gliederungssysteme an das Anschriften- und Gebäuderegister zeigt, dass mehr als die Hälfte aller Matches aufgrund der getroffenen Maßnahmen zur Harmonisierung und Aktualisierung der Datenbestände getroffen wurde. Die hier durchgeführten Aufbereitungsarbeiten zum Beispiel mit dem Straßentheseus leisten damit einen sehr hohen Beitrag zur Verknüpfung des Anschriften- und Gebäuderegisters mit weiteren Datenquellen. Ein Indiz für die Vollständigkeit des Anschriften- und Gebäuderegisters liefert der Vergleich der Neuaufnahmen nach Datenquellen. Ersichtlich wird, dass nur ein geringer Teil der Neuaufnahmen auf die Daten der Primärstatistik oder auf andere Quellen zurückzuführen ist. Die Kombination der genutzten

Register kann damit den Bestand an Anschriften mit Wohnraum nahezu vollständig abbilden.

Auch wenn die Qualität des Anschriften- und Gebäuderegisters nicht eindeutig zu bewerten ist, kann zumindest ein sehr positiver Blick auf das dargestellte Ergebnis geworfen werden. Trotz der nicht optimalen Voraussetzungen zum Aufbau und zur Führung eines solchen Registers sind die nachweisbaren Fehlerraten sehr gering.

Das Anschriften- und Gebäuderegister hat qualitativ sehr davon profitiert, dass es durch die Statistischen Landesämter und das Statistische Bundesamt aktualisiert und vervollständigt wurde. Eine weitere Pflege der Anschriften im Anschriften- und Gebäuderegister, wie sie in vielen Kommunen bereits mit den eigenen Anschriftenregistern seit Jahren betrieben wird, könnte die Vergleichbarkeit der Daten im Zeitverlauf und eine bessere Bewertung der Qualität ermöglichen. Bei entsprechenden gesetzlichen Regelungen wäre darüber hinaus auch die Nutzung des Anschriften- und Gebäuderegisters in weiteren Statistikbereichen vorstellbar, wie zum Beispiel bei der Bautätigkeitsstatistik. Die Vorteile der Registernutzung, wie zum Beispiel die Kosteneffizienz bei der Statistikproduktion, würden auch hier zum Tragen kommen.

Vor dem Hintergrund, dass auch zum nächsten Zensus im Jahr 2021 von weitgehend identischen Problemstellungen für einen registergestützten Zensus in Bundes-, Länder- und Kommunalstrukturen auszugehen ist, könnten die bisher gewonnenen Erkenntnisse zur automatisierten und manuellen Anschriftenbereinigung unter Wahrung der Datenschutzbestimmungen genutzt werden. Im Hinblick auf die dargestellten arbeitsintensiven Prozesse ist eine solche Vorgehensweise aus Gründen der zeitlichen und monetären Effizienz zu empfehlen. [U](#)

Auszug aus Wirtschaft und Statistik

Herausgeber

Statistisches Bundesamt, Wiesbaden

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

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 |
| 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.