

Zensus 2011

Aktuelle methodische Fragen: Aufbau des Anschriften- und Gebäuderegisters

Gemeinsame Tagung DStatG/VDSt
am 18./19.01.2010, Düsseldorf
Johann Szenzenstein, Destatis

Aufbau des Anschriften- und Gebäuderegisters (AGR)

Gliederung des Vortrages

- Aufgaben des AGR
- Datenquellen des AGR
- Aufbereitung der Datenquellen
- Zusammenführung der Datenquellen
- Ergänzende Ermittlung von Gebäuden mit Wohnraum
- Die nächsten Schritte beim Aufbau des AGR



Aufbau des AGR

Die wichtigsten Aufgaben des AGR sind:

- Steuerung der Ablaufs der Gebäude- und Wohnungszählung
- Auswahlgrundlage für die Haushaltsstichprobe
- Prüfung der in den Zensus einzubeziehenden Gebäude, Wohnungen und Personen auf Vollzähligkeit
- Zusammenführung der Datenquellen zum kompletten Zensus-Datenkranz



Aufbau des AGR (1. Ausbaustufe)

Datenquellen des AGR

Datenquelle	Fallzahlen vor Aggregation (in Mill.)	Fallzahlen nach Aggregation (in Mill.)
Melderegister (MR)	86	18
Bundesagentur für Arbeit (BA)	35	15
Vermessungsverwaltung (GAB)	21	-



Aufbau des AGR

Was ist eine Anschrift?

- Amtlicher Gemeindeschlüssel (AGS)
- Postleitzahl (PLZ)
- Straßename
- Hausnummer
- Hausnummernzusatz



Aufbau des AGR

Aufbereitung der Datenquellen (1)

1. Plausibilisierung

z.B. Prüfung auf gültigen AGS und gültige PLZ

2. Parsing

z.B. Zerlegung der zusammengefasst gelieferten Merkmale „Straße, Hausnr.“ (BA-Datei) und Heraustrennen von Ortsteilnamen (GAB-Datei)

3. Standardisierung

z.B. Entfernen von Sonderzeichen und Zahlen aus dem Straßennamen, Großschreibung



Aufbau des AGR

Aufbereitung der Datenquellen (2)

4. Auflösung von Abkürzungen im Straßennamen

5. Bereinigung der Straßennamen

Stufe 1: Referenzierung der Straßennamen aus der GAB-Datei mit dem Straßenverzeichnis von Post Direkt

⇒ rd. 36 000 Prüffälle (3,1%)



Aufbau des AGR

Aufbereitung der Datenquellen (3)

5. Bereinigung der Straßennamen

Stufe 2: Referenzierung der Straßennamen aus der MR-Datei mit den bereinigten Straßennamen der GAB-Datei

⇒ rd. 30 000 Prüffälle (2,7%)

Stufe 3: Referenzierung der Straßennamen aus der BA-Datei mit den bereinigten Straßennamen der GAB- und MR-Datei (Ähnlichkeitsabgleich mit einer Kombination aus den Distanzmaßen „Jaro-Winkler“ und „Levenshtein“)

⇒ rd. 441 000 Prüffälle (20%)



Beispiel für die Aufbereitungsschritte

Originalanschrift

65203 Karl-Friedrich-Gauß-Str. 1 a

■ Plausibilisierung

65203 Karl-Friedrich-Gauß-Str. 1 a

■ Parsing

65203	Karl-Friedrich-Gauß-Str.	1	a
-------	--------------------------	---	---

■ Standardisierung

KARLFRIEDRICHGAUSSSTR

■ Referenzierung

CARLFRIEDRICHGAUSSSTR

Aufbereitete Anschrift

65203	Carl-Friedrich-Gauß-Str.	1	a
-------	--------------------------	---	---



Aufbau des AGR

Bereinigungseffekte der Datenaufbereitung

Daten- quelle	Zahl der unterschiedlichen Straßennamen		Bereini- gungs- effekt in %
	vor Datenaufbereitung	nach Datenaufbereitung	
GAB-Datei	1 159 465	1 147 123	- 1,06
MR-Datei	1 124 521	1 119 411	-0,45
BA-Datei	4 300 020	2 141 449	- 50,20



Aufbau des AGR

Aufbau eines Straßenthesaurus

- Die in den Datenquellen (GAB, MR, Post) vorgefundenen Schreibweisen eines Straßennamens werden in einem Straßenthesaurus aufbewahrt.
- Dieser Straßenthesaurus kann für die späteren Datenlieferungen zum Ausbau bzw. zur Aktualisierung des AGR genutzt werden; d.h. die aufwändigen Standardisierungsarbeiten müssen bei späteren Datenlieferungen nicht wiederholt werden.
- Der Straßenthesaurus enthält derzeit:
4 348 335 nicht standardisierte Straßennamen und
1 002 112 standardisierte Straßennamen.



Aufbau des AGR

Ergebnis der Zusammenführung der „verbesserten“ Datenbestände

Datei	Anschriften	
	Anzahl	in %
AGR.vorläufig	21 583 934	100,0
in GAB+MR+BA	13 864 019	64,2
nur in GAB+MR	4 711 949	21,8
nur in GAB+BA	201 710	0,9
nur in MR+BA	191 366	0,9
nur in GAB	2 452 234	11,4
nur in MR	118 902	0,6
nur in BA	(43 754)	0,2



Ergänzende Ermittlung von Adressen von Gebäuden mit Wohnraum

- Die Statistischen Ämter der Länder müssen nach § 14 ZensG 2011 die offenen GAB-, MR- und BA-Adressen überprüfen, ob es sich um Adressen von Gebäuden mit Wohnraum oder bewohnten Unterkünften handelt.
- Offene GAB-Adressen können z. B. sein:
 - ausschließlich gewerblich genutzte Gebäude
 - Wohngebäude mit (komplettem) Leerstand
 - Wohngebäude mit bewohntem Wohnraum, Personen aber nicht gemeldet
 - nicht existente Adressen



Ergänzende Ermittlung von Anschriften von Gebäuden mit Wohnraum

- Offene MR-Anschriften können sein:
 - neu errichtete Gebäude, die in der GAB-Datei 2008 noch nicht enthalten waren. Von den rund 120 Tsd. offenen MR-Anschriften (gegenüber GAB 2008) konnten in der aktuellen GAB-Datei 2009 cirka 54 Tsd. Anschriften (= rd. 45 %) gefunden werden.
 - sogenannte „fiktive“ Anschriften, denen im Melderegister Personen zugeordnet sind, die dort aber nicht wohnen (z.B. obdachlose Personen).



Ergänzende Ermittlung von Anschriften von Gebäuden mit Wohnraum

Die Prüfung der offenen Anschriften erfolgt in einem dreistufigen Verfahren:

- Stufe 1: anhand vorliegender interner oder externer Datenquellen (z.B. Unternehmensregister, Telefonverzeichnisse und die Gebäudedatei von Post Direkt)
- Stufe 2 (nur Restfälle aus Stufe 1): anhand weiterer, nicht personenbezogener Unterlagen z.B. der Vermessungsverwaltung (Automatisiertes Liegenschaftsbuch - ALB) und vom Meldewesen genutzte Adressdateien
- Stufe 3 (Restfälle aus Stufe 2): durch schriftliche Befragung der Gebäudeeigentümer oder Prüfung vor Ort

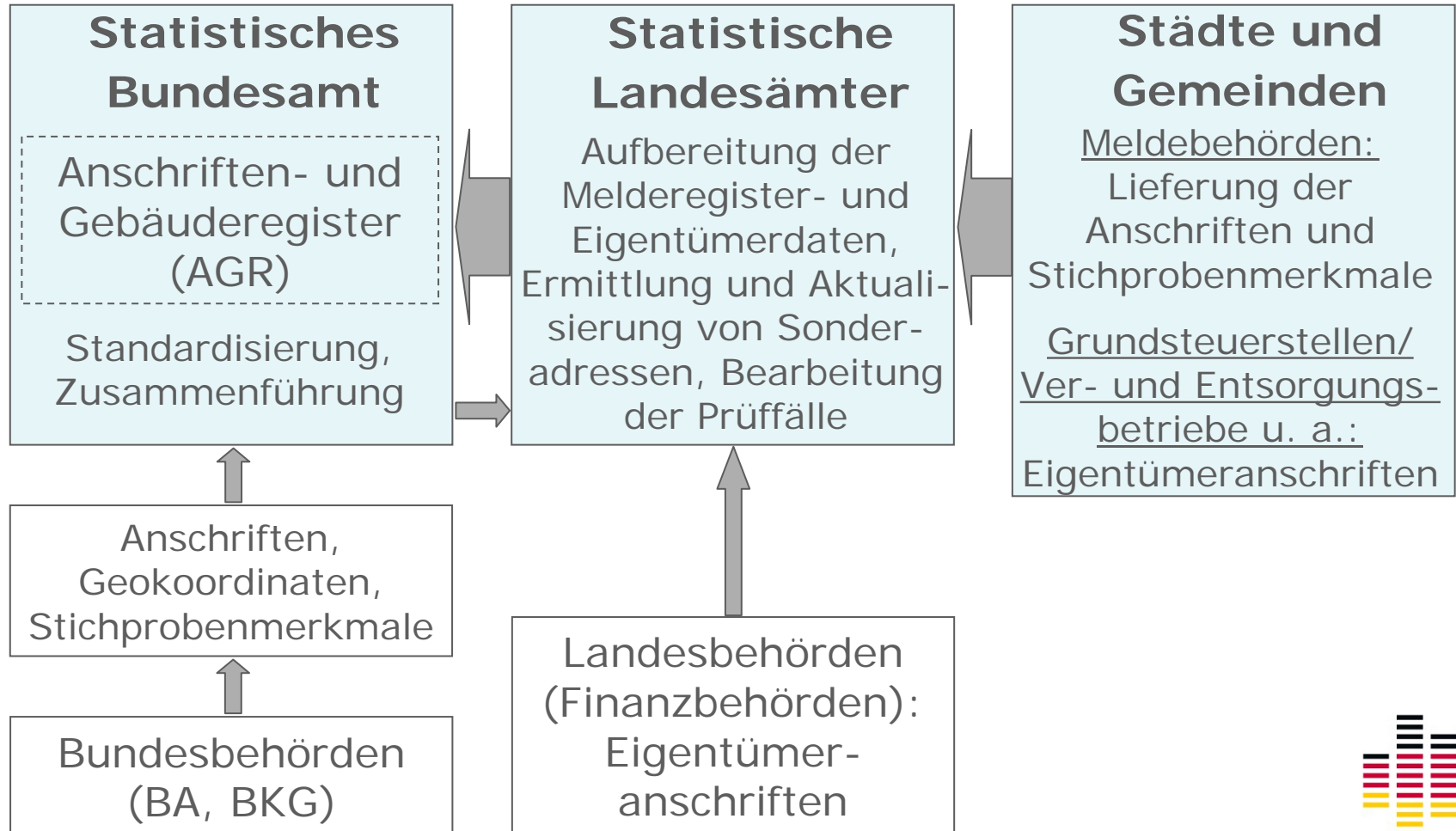


Die nächsten Schritte beim Aufbau des AGR

- Anbindung des Sonderanschriftenregisters: die Daten zu den Sonderanschriften kommen aus einer im Zeitraum Aug. bis Dez. 2009 durchgeführten Befragung bei den Trägern/Einrichtungen (rd. 50 000 Befragte)
- Integration der Daten zu den Gebäudeeigentümern: vorrangig genutzte Datenquellen zur Ermittlung der Gebäudeeigentümer: „Grundsteuer-Daten“ und „Daten der Ver- bzw. Entsorgungsbetriebe“
- Aktualisierung des AGR mit den Meldedaten April 2010 und den GAB-Daten 2010



Aufgabenteilung beim Aufbau AGR



Vielen Dank für Ihre Aufmerksamkeit!

Johann Szenzenstein

Statistisches Bundesamt

Tel: 0611 – 75 22 37

E-Mail: johann.szenzenstein@destatis.de

