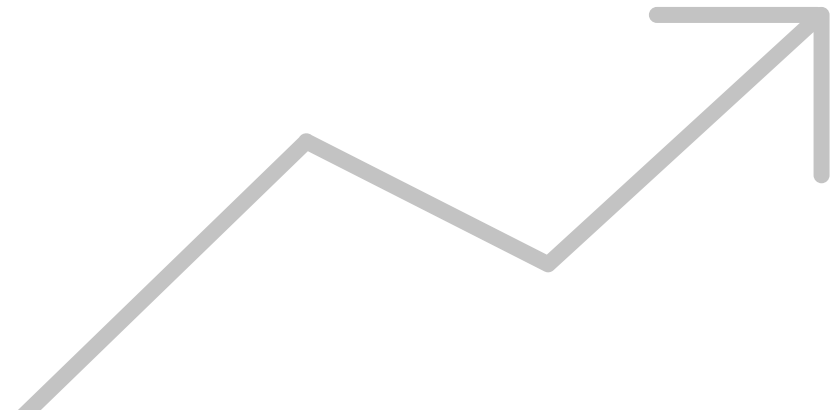


# Steps to Assess Data Quality of Registers



# Suggestions

1. Produce a register profile
2. Data fit concerning the overall concept of the census
3. Assessing data quality
4. Combining different data sources

# 1. Produce a register profile (1/2)

- Who owns the register?
- What is the primarily purpose of the register?
- Which population is (not) covered?
- Core variables

# 1. Produce a register profile (2/2)

- Profiling each (potential) register systematically
- On the right: Register profile as used by German Register Census to evaluate a register's general fit concerning it's use in the register census

## General information of the register

F-REB/0401  
Datum:

Register Steckbrief			
Allgemeine Angaben zum Register			
Name			
Kurzbeschreibung			
Ansprechpartner bei Nichterfüllung der Stelle			
Angaben zum Ansprechpartner bei Destatis			
Name			
Organisationseinheit			
Telefon			
Zentrale Merkmale im Datenbestand			
Datenbestand Statistik		Originaldatenbestand	
Name	Bund	Länder	
Vornamen			
Geburtsname			
Geburtsdatum			
Geburtsort und -staat			
Geschlecht			
Familienstand			
Staatsangehörigkeit			
Wohnanschrift			
Straße			
Postnummer			
PLZ			
Stadt			
Bundesland			
Identifikatoren		Weitere Merkmale	
z. B. Steuer-ID, Ausländerzentralregisternummer		z. B. Steuer-ID, Ausländerzentralregisternummer	
Sonstige Merkmale (die besondere für den LZB genutzt werden können)			
z. B. Angaben zu Ehepartnern und Kindern		z. B. Angaben zu Ehepartnern und Kindern	
Rechtsgrundlage(n)		Originaldatenbestand	
Statistik		Originaldatenbestand	
Fragenkatalog			
Übergeordnet		Originaldatenbestand	
Statistik		Originaldatenbestand	
Grundgesamtheit: Welche Datensätze sind vorhanden?			
Lebenslage / Personengruppe		Lebenslage / Personengruppe	
z. B. steuerpflichtige Rentner, Beamte		z. B. steuerpflichtige Rentner, Beamte	
Statistische Einheit (wer/was)		Statistische Einheit (wer/was)	
z. B. Führerschein, Person		z. B. Führerschein, Person	

Contact information of inhouse expert

Core variables

Legal basis

Information on data in register

## Information on transmission of data

F-REB/0401  
Datum:

Stichtag Datenerhebung Anzahl (ggf. Schätzung)	z. B. Beständige Prüfung	z. B. Beständige Prüfung
Weitere Datenquellen: Gibt es weitere Datenbestände mit Personeninformationen?		
Handelt es sich um ein einzelnes Register oder sind Daten über mehrere Bestände verteilt?		
Besteht Datenaustausch zu weiteren Datenquellen?		
Datenübermittlung		
Datenhaltung: Zentraler/Dezentraler Datenbestand		
Wer führt das Register technisch? (Registerführer)		
Wo liegen die Daten? (Registerbetreiber)		
Wer ist zuständig für die Datenübermittlung?		
Datenformat: Übermittlungsstandard vorhanden? Intern andere Datenformate?		
Vorlaufzeit: Wann müssen die Daten beantragt werden um Stichtag Testdaten Registerzensus abzurufen (16.05.2022)?		
Wo müssen die Daten beantragt werden?		
Qualitätsbestimmung		Originaldatenbestand
Qualität: Wie werden die Daten gepflegt?		z. B. gefüllter Staatenschlüssel
Erfolgt eine Aktualisierung?		
Wie oft werden die Daten aktualisiert?		
Welche Angaben werden regelmäßig genutzt?		
Wann werden individuelle Merkmale aktualisiert?		
Relevanz: Können nicht relevante Personen abgegrenzt werden?		
Gibt es einen Hinweis auf Auslandsaufenthalt?		
Gibt es einen Hinweis auf Verstorbene?		
Statistik: interne Prüfschritte		

F-REB/0401  
Datum:

Zusätzliche Informationen	
Sonstige Informationen: Lebenszeichenansatz	Freitext
Sonstige Informationen: Datenübermittlung	Freitext

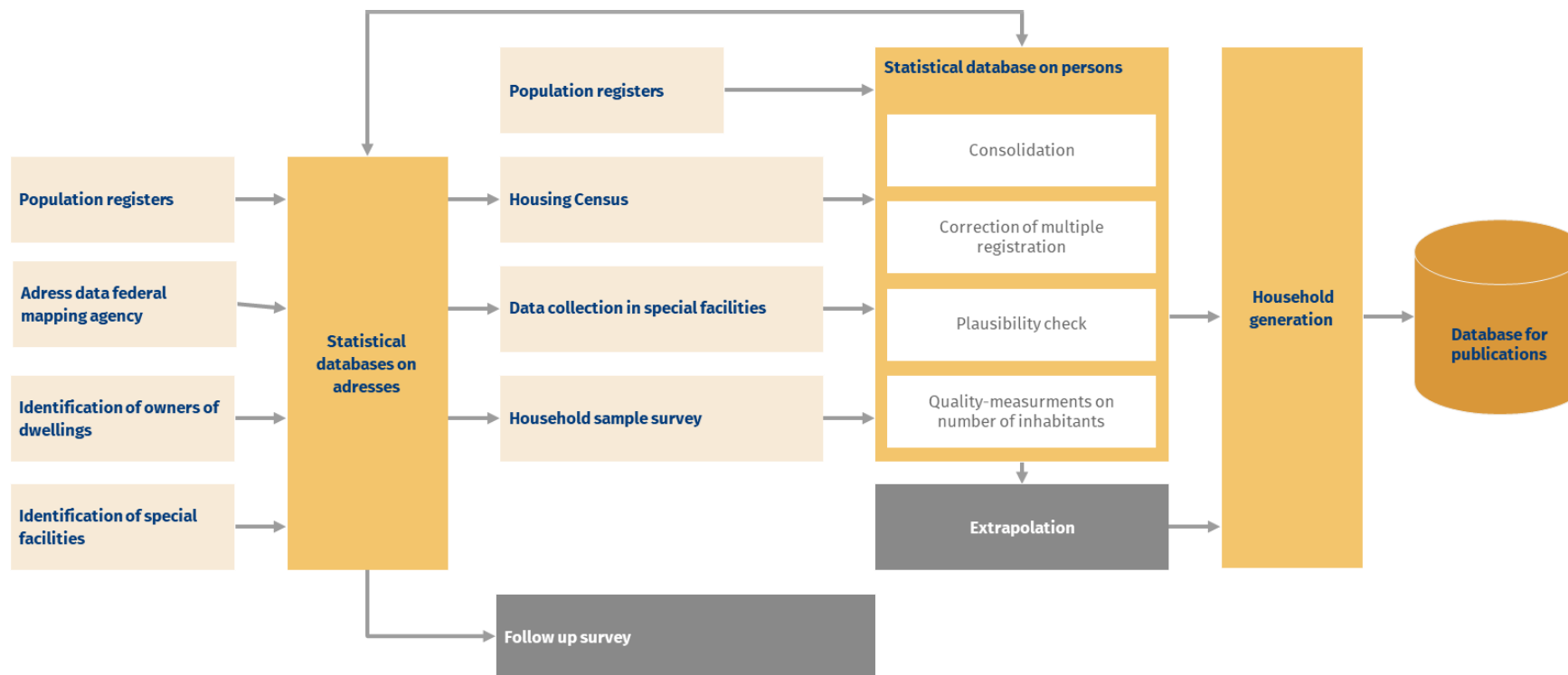
Further information / notes

Quality assessment

## 2. Data fit concerning the overall concept of the census (1/2)

- Benefit for census model with respect to different units
  - Regional level
  - Population
  - Buildings
  - Households
  - Families

## 2. Data fit concerning the overall concept of the census (2/2)



# 3. Assessing data quality

- Number of records (counted or estimation)
- Comparison of data agreed on to be delivered and the actual data received
  - » Completeness of variables
  - » Variable formats as expected
- Variable characteristics
  - » Overall:
    - » Number of entries for each variable
    - » Number of missings for each variable
    - » Missing codes
    - » Content-related usability of the variables
  - » Numerical variables:
    - » Spread of numerical variables (e.g. min, max, average)
  - » Alphanumerical variables:
    - » Quality of alphabetical variables
      - Compliance to filling regulations or rules for spelling
      - Frequently used abbreviations
      - Check for undisplayable characters
  - » Categorical variables:
    - » Values and their meaning
  - » Date variables:
    - » Date format
    - » Plausibility of dates
  - » Timeliness of the territorial/local keys
  - » Interplausibility of the variable values

## 4. Combining different data sources

- Aim of combination of the data source
  - Add coverage
  - Add variables
- Identifier in both sources
  - ID-number
  - quasi-identifier (e.g. combination of sex, birthday, birth place, names /name similarities)



# Literature

UNECE (2021): Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses. [https://unece.org/sites/default/files/2021-10/ECECESSTAT20214\\_WEB.pdf](https://unece.org/sites/default/files/2021-10/ECECESSTAT20214_WEB.pdf)

# Contact

Statistisches Bundesamt  
65180 Wiesbaden  
Germany

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

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

Contact Person  
Stefan Dittrich  
[Stefan.dittrich@destatis.de](mailto:Stefan.dittrich@destatis.de)  
Phone +49 611 75-2237

