

# TOWARDS A REGISTER-BASED CENSUS POST-2021 IN GERMANY

Objectives, requirements and challenges

Thomas Körner and Eva Grimm, FSO Germany

UNECE Group of Experts on Population and Housing Censuses

 Geneva, 18-20 September 2019

# OUTLINE

- I. Objectives, requirements and vision
- II. Selected challenges
  - a. Creation of a register of buildings and dwellings
  - b. Treatment of overcoverage and undercoverage in population register data
  - c. Linking register data without a unique person ID number
- III. Discussion

# I. OBJECTIVES, REQUIREMENTS AND VISION



# REQUIREMENTS OF A FUTURE PROOF CENSUS MODEL



## Current method: high burden, cost and efforts

- **Higher costs** compared to other European countries (2021: 1.4 billion Euros)
- **High response burden** (10 million persons in household survey plus 25 million property owners in housing survey)
- **Inconsistency** between census population figures and the number of persons in local population registers
  - **Controversy** at each census round
  - Regular **breaks in time series** of population figures



## New user requirements: more frequent, more timely and more detailed results



More **frequent** and more **timely** data while **reducing burden** for citizens and tax payers



More **frequent**, more **timely** and **geocoded** population figures from **2024** onwards, further changes from census round **2031** onwards

→ **Framework regulation** ESOP under preparation

# FROM REGISTER-ASSISTED TO REGISTER-BASED CENSUS IN GERMANY

## Register-assisted model (2021)

- Main data sources: registers **combined** with primary data collection
  - Decentralised population registers
  - Household sample **survey** (10 million persons)
  - Housing census (25 million property owners)
  - Complete enumeration of population in institutions (heads of institutions)
- Quality assurance for population register data
  - Statistical correction of over-/undercoverage of the data from local population registers based on the **household sample survey**
- Record linkage
  - Merging data from different sources in a central data stock without existing personal identification numbers

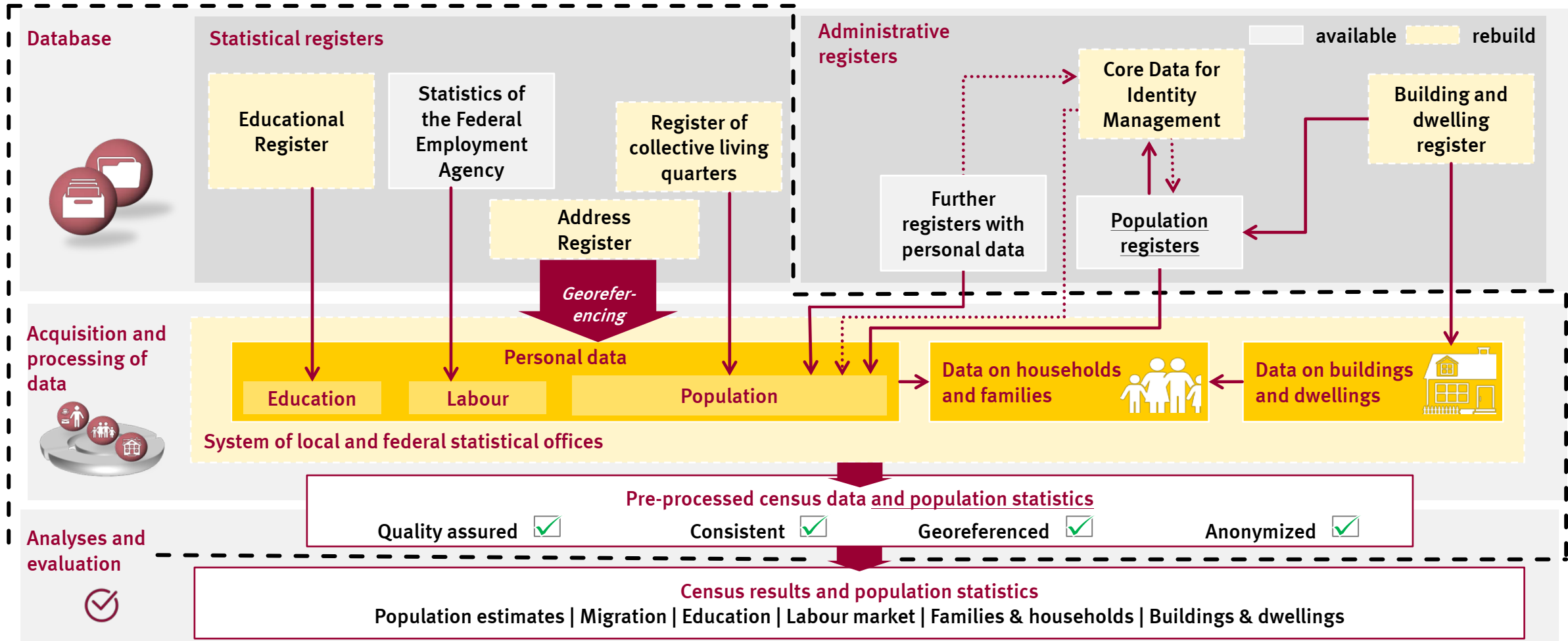


## Register-based model (post-2021)

- Main data sources: **only registers**
  - Decentralised population registers
  - Creation of **new registers**: Building and dwelling register, education register, register of institutions
  - Use of existing registers, e.g. for obtaining data on labour market participation
- Quality assurance for population register data
  - Using **signs of administrative life** to correct for overcoverage (and undercoverage) in local population registers
  - Use of **future core data system** for identity management
- Record linkage
  - Merging data based on personal characteristics without existing personal identification numbers
  - Possibly use of future personal ID numbers
  - Introduction of a dwelling identifier to link persons with dwellings and households

# POSSIBLE VISION OF A REGISTER-BASED CENSUS

— Statistical confidentiality —> Data transmission —> Quality assurance

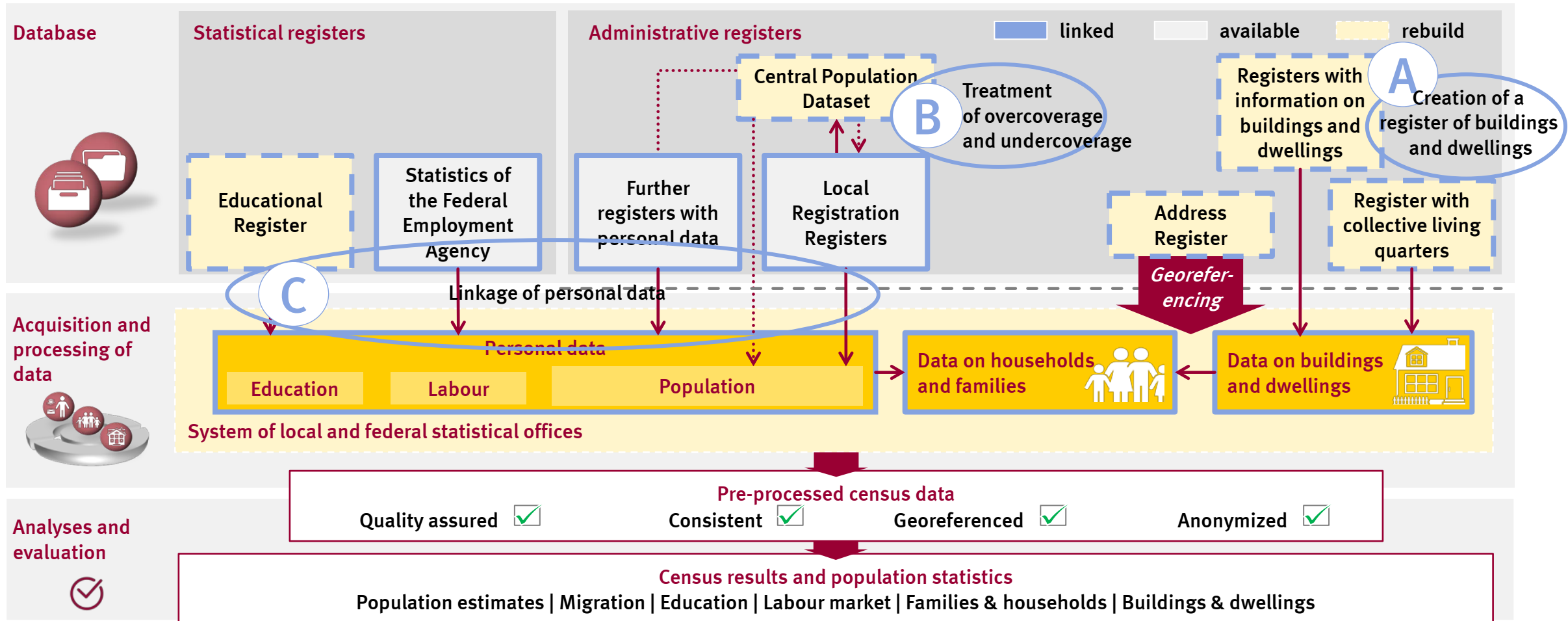


## II. SELECTED ISSUES



# SELECTED ISSUES

-- Statistical confidentiality    → Data transmission    ⋯→ Quality assurance





# A. CREATION OF A REGISTER OF BUILDINGS AND DWELLINGS



# REGISTER-BASED COLLECTION OF DATA ON BUILDINGS AND DWELLINGS

Existing administrative sources are not suitable for census purposes

Land registers	X
Data of land surveying offices	X
Local building registers	X
Planned property tax database	X

Currently, a total enumeration of all buildings and dwellings by a survey of all 25 million owners is necessary

Currently, there is no nationwide register with information on buildings and dwellings → a new register is necessary

## Statistical requirements for housing data



### EU requirements

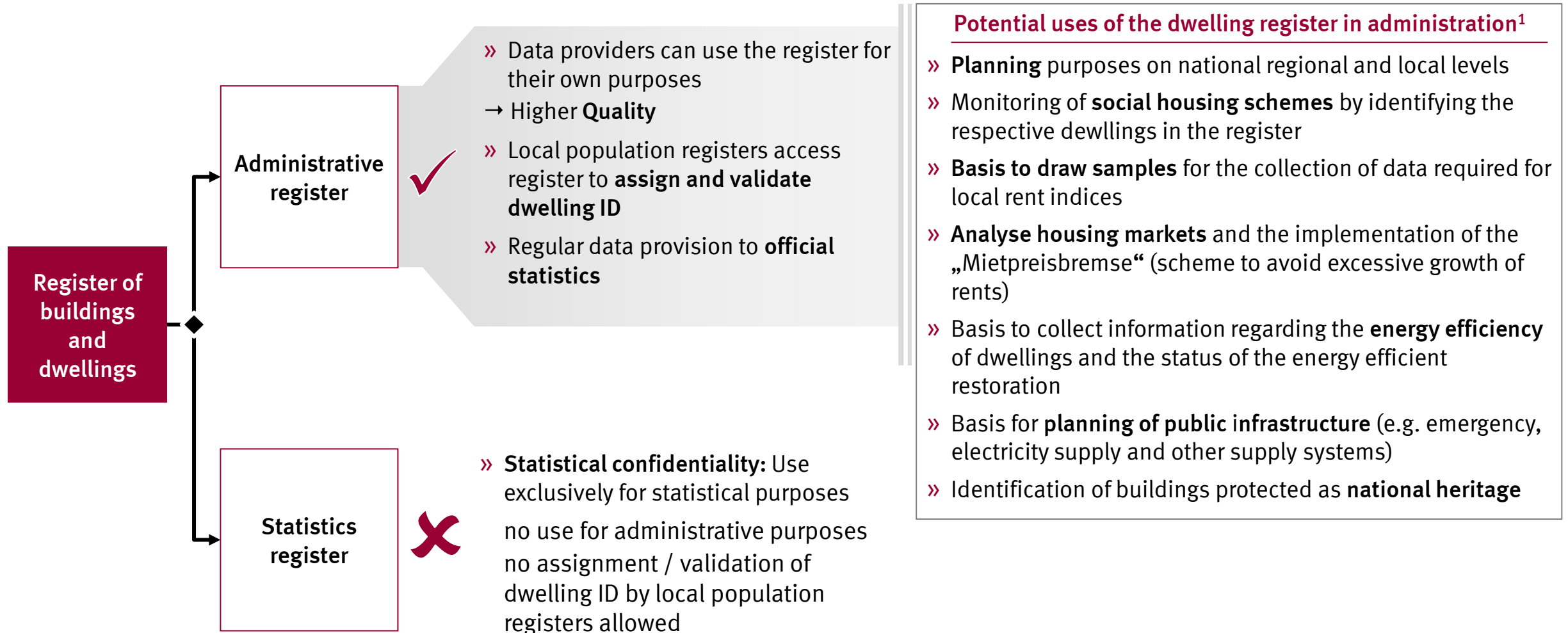
- » Collection of data about dwellings
- » Introduction of an identifier for buildings and dwellings is necessary to get information about households and housing conditions
- » Geocoding of housing data



### National requirements

- » Generally more detailed breakdowns of variables
- » Also information about buildings
- » Possible future requirements for administrative purposes (e.g. energy efficiency of buildings; commercial buildings)

# THE DWELLING REGISTER NEEDS TO BE AN ADMINISTRATIVE REGISTER



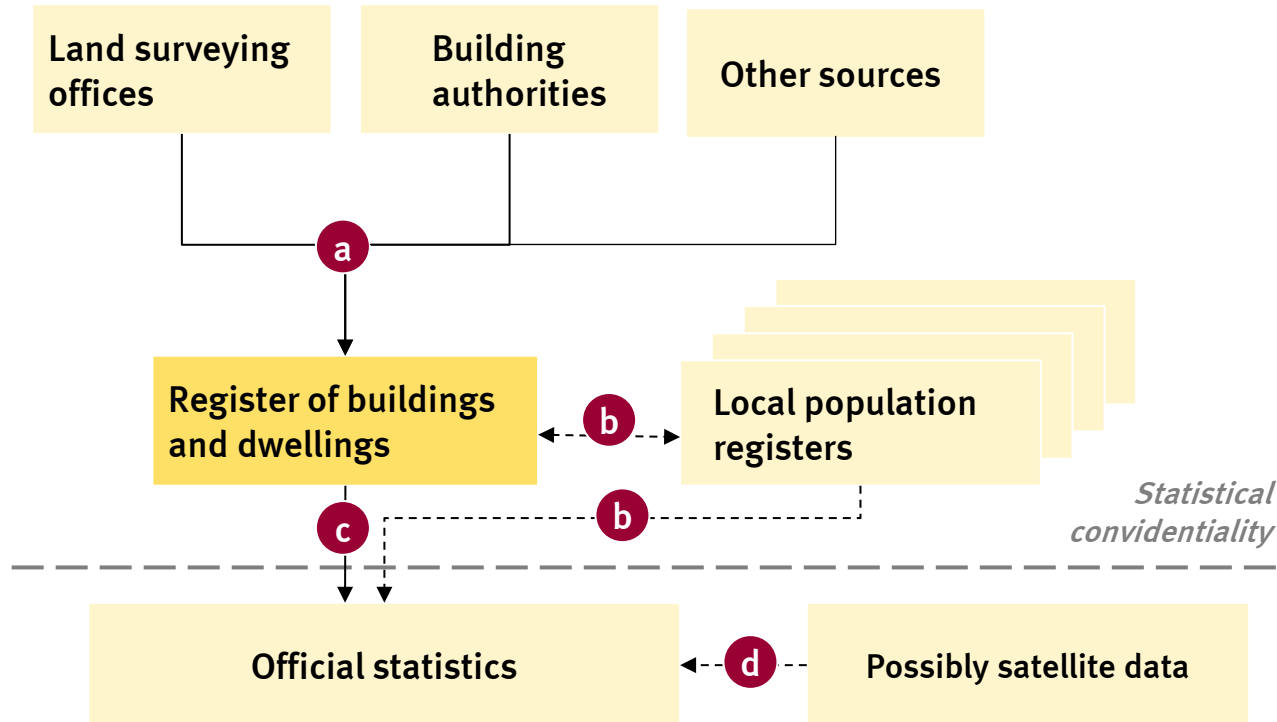
<sup>1</sup> To be validated

# CONSTRUCTION AND MAINTENANCE OF A REGISTER OF BUILDINGS AND DWELLINGS (GWR)

## Sources of construction and maintenance of the GWR

---▶ Quality assurance

—▶ Data transmission



First of all, collecting data for the construction of the register (stock of buildings and dwellings)

- a** Updating by the use of administrative data (building authorities and land surveying offices)
- b** Linkage and comparison with data in the local population registers
  - » Record linkage by identifiers of buildings and dwellings
  - » Editing of data on dwellings by using data from the population registers inside the statistical offices
- c** Provision of data to the statistical offices at the reporting date
- d** Quality assurance by comparison with satellite data and possibly other sources
  - » Possibly automatic evaluation of constructional changes by satellite images (technical review is still necessary)

# **B. TREATMENT OF OVERCOVERAGE AND UNDERCOVERAGE IN POPULATION REGISTER DATA**



# TYPES OF COVERAGE ERRORS IN THE POPULATION REGISTERS

Type of coverage error	Description	Consequence if not corrected
<b>Overcoverage in the population registers („pure“ cases of overcoverage)</b>	Omitted deregistration after death	Persons would be counted although they do not live in the country
	Omitted deregistration when emigrating to another country	Persons would be counted in Germany as well as in another country
	Omitted deregistration when moving the center of life (main residence) to another country	
	Registration at more than one sole or main place of residence.	Persons would be counted several times within the country (duplicates)
<b>Complementary cases of over- and undercoverage in the population registers</b>	Omitted change of registration when moving within the country	Persons would be counted in the wrong municipality
	Omitted change of registration when moving the center of life within the country (e.g. exchange of main residence with secondary residence)	
<b>Undercoverage in the population registers („pure“ cases of undercoverage)</b>	Omitted registration after birth	Persons would not be counted although they live in the country
	Omitted registration when immigrating from another country	
	Omitted registration of persons who are deregistered „to unknown“ although still living in the country	
	Registration only with secondary residence	

# CHALLENGES

## „Pure“ cases of overcoverage



- **Marker:** Cases with a record in a population register but no record in any comparison register
- **Prerequisites:**
  - Timeliness of the register entry in comparison registers
  - Comparison registers include personal identification numbers or personal variables that enable linkage of records

## Complementary cases



- **Potential Marker:** Cases in which information on the residence status of a person differ between population register and comparison register
- **Prerequisites:**
  - Address data or local area codes in comparison registers
- **Challenges:**
  - Possibilities to link address data must be assessed in a feasibility study

## „Pure“ cases of undercoverage



- **Potential Marker:** Cases with a record in a comparison register but no record in any population register
- **Challenges:**
  - Records in comparison registers do not necessarily go along with main residence status
  - Personal variables have to be imputed

# CHALLENGES

- Development of methods for assessing undercoverage and complementary cases
- Identification of suitable comparison registers
  - Each person residing in Germany needs to have a chance to be included in at least one comparison register
  - Registers must include personal identification numbers or personal variables that enable linkage of records
  - The timeliness of register entries has to be determinable
  - The Quality of the register data has to be assured by constant maintenance
  - Register entries need to be independent from entries in population registers
  - Address data or local area codes need to be available in sufficient quality (if implemented for assessment of undercoverage and complementary cases)



# C. LINKING REGISTER DATA WITHOUT UNIQUE ID NUMBER



# POSSIBILITIES OF LINKING PERSONAL DATA: IDENTIFIERS

Data basis	Specification	Possibilities of record linkage through all relevant registers
<b>Constant identifier</b> (unique string; → e. g. encrypted personal identifier in Austria)	<b>Unencrypted</b> → 12 345 678 911 <b>Encrypted</b> → 2ffd9b5809145c0e150ed538aaa45a84299f36fb3c85c5e66ce6f43bfee6e5e078dc6249962e4b1e777c07d01ee6d95ce5a83281c7d452f568447055f5f57d00	1) A uniform constant un-/encrypted identifier is provided by all relevant registers 2) Different constant un-/encrypted identifiers are contained in multiple registers and at least one register has an overlap of those identifiers
<b>Identifying personal characteristics</b> (→ e. g. name, first name, date of birth)	<b>Hybrid types</b>  <b>Non-anonymized</b> → Smith, John, 1965-12-18  <b>Anonymized</b> → smi, joh, 196-12-18	3) Uniform constant un-/encrypted identifiers are provided by at least two registers and non-anonymized identifying personal attributes by the other registers; needs at least one register with overlap 4) As 3), but identifying personal attributes are provided anonymized
		5) Only non-anonymized identifying personal characteristics through all relevant registers 6) Only anonymized identifying personal characteristics through all relevant registers

# Challenges of establishing a record linkage procedure without unique ID

## Possible variables for record linkage

- Names
  - Name
  - Given name
  - Name at birth
- Date and place of birth
  - Dates of birth
  - Place of birth
  - Country of birth
- Place of residence
  - Main residence
  - Secondary residences
- Sex
- Marital status



## Challenges and limitations

- Biunique identification of persons
  - Problematic population groups include
    - Twins
    - Very common names (e.g. Peter Müller)
    - Phonetically similar names
    - Imputed (or socially desirable) birth dates
- Linking different registers
  - Availability of (harmonised) variables across different registers
  - Transcription of foreign names
  - Differences in data quality according to administrative requirements
- Data protection requirements
  - Separating identifying information

Which are the requirements for a biunique identification of residents in Germany?  
 Which additional requirements exist due to differential practices in keeping registries?  
 How can record linkage be assessed in simulation studies?  
 Which record linkage procedures can be applied?

**Expert evaluation  
ongoing**

# III. CONCLUSIONS AND QUESTIONS



---

## CONCLUSIONS AND QUESTIONS

- **Do you have experiences with...**
  - ... creating a building and dwelling register?
  - ... setting up the signs of life approach for quality assurance?
  - ... assessing undercoverage in population registers using administrative signs of life?
  - .... linking data without personal identification numbers?
  
- **Would you like to cooperate with us in the development of such approaches?**

# THANK YOU.

[thomas.koerner@destatis.de](mailto:thomas.koerner@destatis.de)

