Administrative data - introduction

Danilo Dolenc

GCC-STAT Regional Workshop

Muscat, Oman, 22-24 September 2019

Overview

- Basic metadata about administrative source
- Types of data sources possible to use in register-based statistics
- Examples of data sources
  - Stress on sources used for census and social statistics surveys
- Identifiers as the key for data linkage

Basic metadata (1)

- Content of data source
  - Population concerned
  - Topics (variables) available
    - Definition (methodology)
- Coverage
  - Whole population or part
    - Enrolment of students vs students receiving scholarship

Basic metadata (2)

- Obligatory metadata for every variable
  - Format
    - Number
    - Character (varchar)
  - Length
  - Classification used

Basic metadata (3)

- Timeliness
  - Time span between reference date / period and received data
- Periodicity of receiving data
- Type of reporting unit
  - Person
  - Enterprise for itself
  - Enterprise for employees

Basic metadata (4)

- Technical specifications of dataset
  - Data structure
    - List of variables
  - Data format
    - Most common .csv
  - Metadata to the dataset enclosed
  - Transfer of data to the NSI
    - Web service, File Transfer Protocol, DVD,…

Types of data sources

- Administrative registers
- Statistical registers
- Administrative databases
- Statistical databases (reuse of existing statistical surveys)
  - Full coverage
  - Sample surveys
  - Special statistical surveys

Administrative register (1)

- Maintained by relevant state institution
- Based on law, legal procedures and documentation
  - Rarely also statements are collected
    - Ethnicity
- Important for everyday life and/or business activities (official data)
- Updated on regular basis
  - Mostly daily

Administrative register (2)

- Several possibilities to extract data for statistical purposes
  - Current stock data
  - Stock data on specific point of time
  - Calendar stock data
  - Data on any change of stock (events)
  - Longitudinal (historic) data

Administrative register - examples

- Population Register
- Business Register
- Real Estate Register
- Employment Register
- Unemployment Register
- Vehicle Register

Statistical register

- Almost the same characteristics as administrative register
  - Based on administrative records
  - Statistical methods applied for updating
- Important difference - data cannot be:
  - Used for administrative purposes
  - Mediate to individuals (statistical confidentiality applied)

Statistical register - examples

- In theory, every administrative register could be converted into statistical one
  - Do we need a copy of administrative register supplemented with statistical editing?
- It is reasonable to set up statistical register if there is no administrative register
  - Statistical Register on Employment in our case

Administrative databases

- Extracted from administrative systems of different data provider (no events)
  - Changes could be tracked by comparing two consecutive databases
- Cross-sectional (stock) data at certain point of time
- Data for longer period of time (flow data) – calendar year

Administrative databases - examples

- Taxation data (calendar / fiscal year)
- Social security data (social benefits / transfers)
- Business accounts (annual, quarter)
- Foreign trade data (customs declarations)
- Enrolment of students in school / academic year (1 October = first day)
- Health Insurance data
- Issued visas
- Data on entry to / exit from country

Reuse of statistical surveys

- Regular surveys with full coverage
- Regular sample surveys
- Special surveys to
  - Improve quality / coverage of administrative data
  - Evaluate administrative data
  - Set up new statistical register

Reuse of statistical surveys - examples

- Demography surveys
  - New-born
  - Death
  - Migration
    - Internal (reliable data on usual residence)
    - International (over- and under-coverage
- Labour Force Survey
- Household Budget / Income Survey
- Irregular Immigration Survey

Identifiers

➤ Primary identifiers

   ➤ Should be unique

➤ Secondary identifiers

   ➤ Connected to the primary ones

      ➤ Could be duplicated

➤ Example

   ➤ PIN of person vs PIN of parents

More about in the editing session tomorrow

Basic identifiers (1)

- Personal identifiers
  - PIN
  - Tax number
  - Social security number
- Territorial identifiers
  - Addresses
  - Geo-codes (coordinates)

Basic identifiers (2)

- Business (enterprise, company) identifiers
- Real estate identifiers
  - Building ID
  - Part of building ID (Dwelling number and / or dwelling ID)
- Other identifiers
  - Household identifier
  - Electricity number
  - …

Personal administrative identifiers (1)

- Indicative
  - Slovenian PIN (DDMMYYYRRSNNNC)
- Non-indicative
  - Random defined
    - Should be numeric
      - Not begin with zero
    - With or without control number
      - Slovenian tax number (NNNNNNC)

Identifiers - administrative vs statistical

- Converting administrative personal identifiers into statistical ones
  - Better security
    - No links with outside world
  - Important for public acceptance
  - Necessary in case of administrative sources with different identifiers
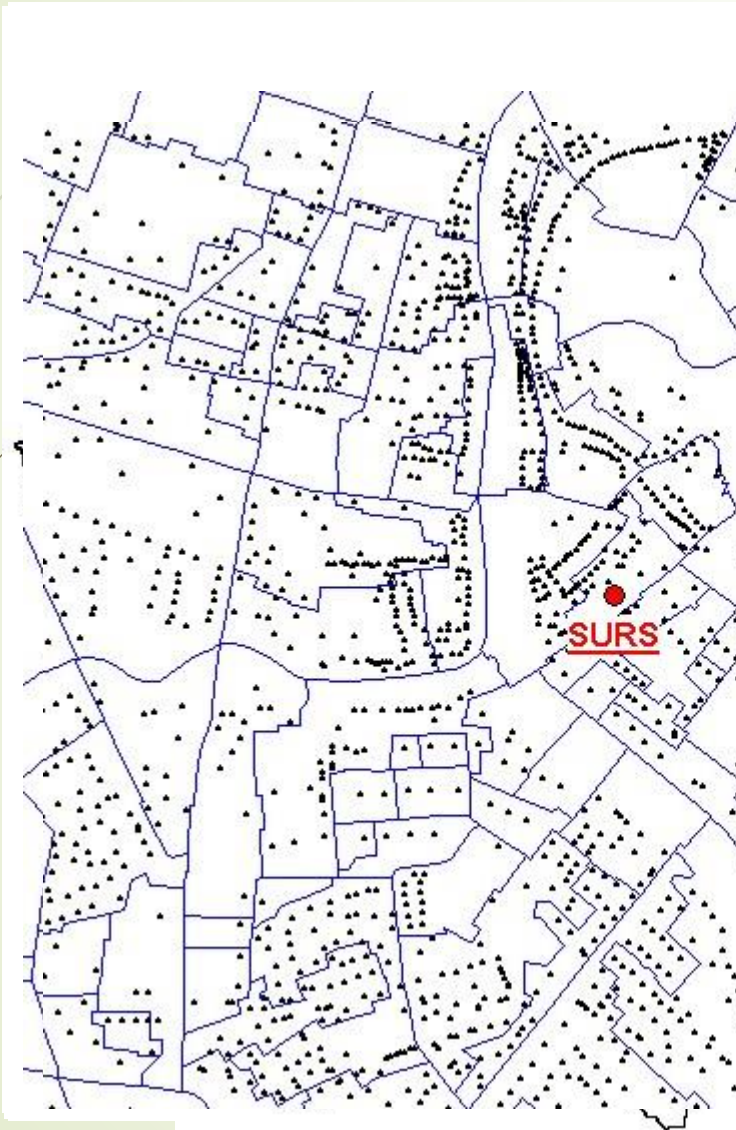 - Possible to follow changes of personal identifiers

Territorial identifiers (1)

- Indicative
  - Hierarchical structure
- Point based
  - Basic unit is a building address
  - Allow data dissemination on grid
    - Coordinates available
- Coverage must be complete
  - At least for inhabited buildings

Territorial identifiers (2)

- Indicative but must allow uniqueness
  - Additional elements could be included (entrance number)
  - Based on current addresses using national classification (publicly available)
    - 06101106200012
- Non-indicative
  - Special (non-changeable) ID of address
    - All other information are just metadata
    - Easy to follow history of address

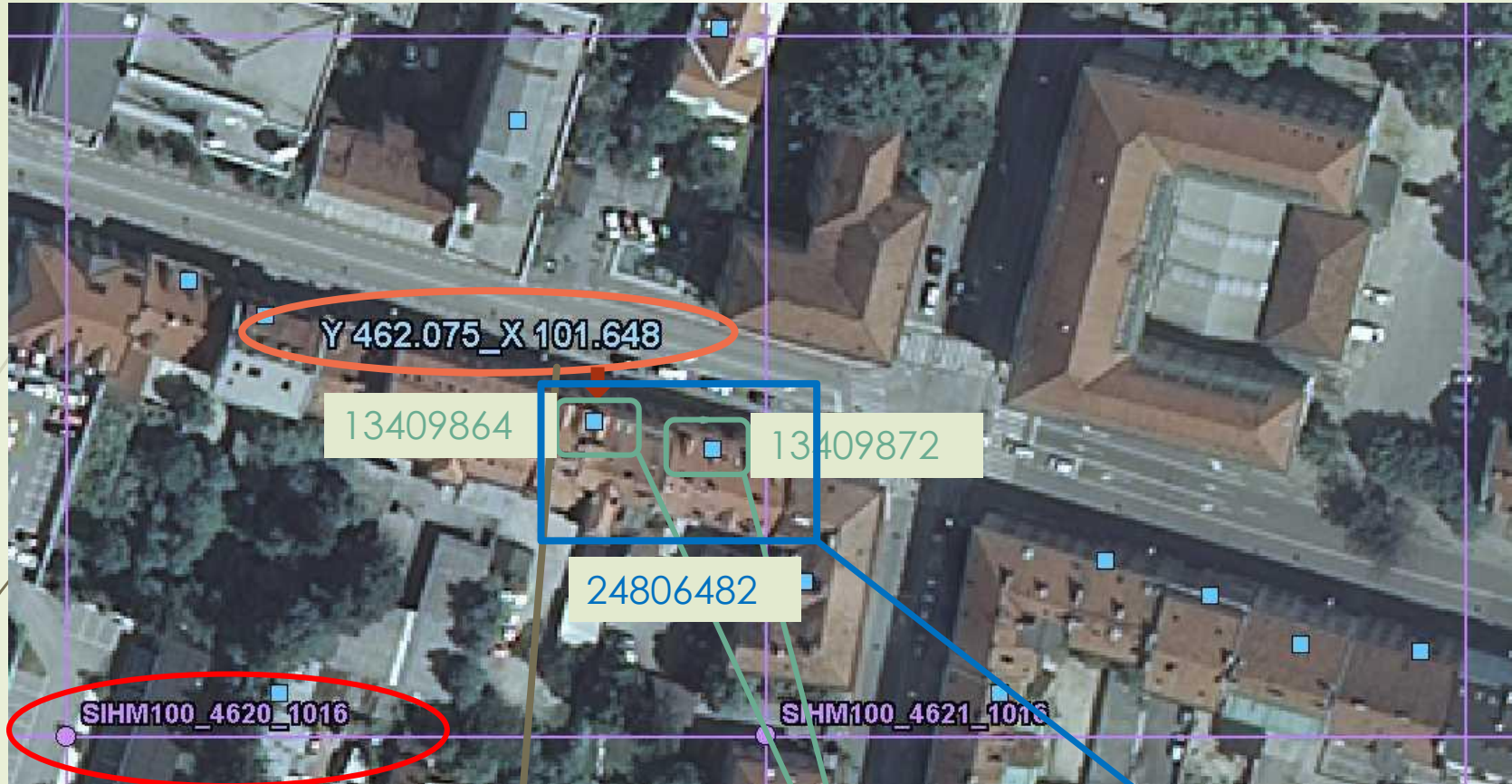# Address atributes (meta data)

- **HS_MID = 12596032**
- COUNTRY Slovenia
- REGION Osrednjeslovenska (07)
- ADM_UNIT Ljubljana (24)
- MUN Ljubljana (061)
- CITY Ljubljana (011)
- EA 28 (10305136)
- STREET Vožarski pot (0620)
- HOUSE NR 12 (012)
- X_H = 462283
- Y_H = 100165

Business identifier

- Important link between person and his employer
  - Several data could be extracted and assigned to the person's record
    - Industry
    - Location
    - Size of enterprise

Real estate identifiers

- Not the same as address identifiers
  - Building with several addresses
  - Address with several buildings
- Building part identifier with metadata
  - Type of use of unit is crucial for census purposes
    - To distinguishing conventional dwellings, institutional / collective living quarters and other housing units

Y 462.075_X 101.648

13409864

13409872

24806482

SIHM100_4620_1016

SIHM100_4621_1016

Grid code

X, Y coordinate of the address

Address identifiers

Building identifier

Household identifier

➤ Linking together persons living at the same dwelling

  ➤ In census context important for determine private household concept used

    ➤ Housekeeping (one dwelling = one household)

    ➤ Household-dwelling (one dwelling = one or more households)

Conclusion

- Register-based census or traditional census
  - It is all about identifiers and linkage of
    - Individuals to households
    - Households to dwelling
    - Dwellings to building