

# Challenges in transition from field survey to the register- based statistical process - Slovenian census case

**Danilo Dolenc**

GCC-STAT Regional Workshop

Muscat, Oman, 22-24 September 2019

# Drivers

- Changing society
  - Regular updated data
    - 10 year periodicity – out of time
  - Timeliness
    - New data needed now
  - More detailed data
  - Geo-referencing
- Decreasing costs / increasing quality
- Participation - respondent burden – privacy

# Challenges (1)

- To keep all essential census features
  - **Individual enumeration**
  - Universality
  - Simultaneity
  - Periodicity
  - Small area data
- Converting administrative concepts into statistical ones
  - New methodology must be developed

## Challenges (2)

- ▶ Communication with data users
  - ▶ In preparation phase
    - ▶ Non-availability of data
      - ▶ Ethnicity data
    - ▶ New methodology
      - ▶ More statistical methods used
  - ▶ Dissemination
    - ▶ Break in series with previous censuses

# Challenges (3)

Household size	2002		2011		Change (%) 2011/2002
	Number	%	Number	%	
<b>TOTAL</b>	<b>684,847</b>	<b>100</b>	<b>813,531</b>	<b>100</b>	<b>18.8</b>
1 member	149,757	21.9	266,489	32.7	77.9
2 members	157,195	23.0	199,875	24.6	27.2
3 members	143,337	20.9	149,144	18.3	4.1
4 members	158,145	23.1	127,376	15.7	-19.5
5 members	49,575	7.2	44,320	5.4	-10.6
6 members	17,933	2.6	17,658	2.2	-1.5
7 members	5,808	0.8	5,689	0.7	-2.0
8 + members	3,097	0.5	2,980	0.4	-3.8

## Challenges (4)

- Close cooperation with data providers
  - Main stress on input quality of data
  - Pro-active role of NSI
- Standardization and harmonization of input data
  - Mostly inside NSI
  - Historical series used (till 1989)
- Pre-evaluation of statistical process, quality of inputs and outputs
  - Started at least 2 years before

## Challenges (5)

- ▶ To improve quality of the most essential variables in administrative sources
  - ▶ Under-coverage of dwelling identifiers in CPR
    - ▶ Solved partially in advance and partially in the process



# Challenges (6)

- Solutions for update missing dwelling identifiers (DI) in Central Population Register
  - Statistical
    - Automated procedures based on ownership of dwelling and registered residence
      - Not implemented in the CPR
  - Administrative
    - Common action by NSI and Ministry of the Interior
      - 49,000 letters, 75% response rate
        - Implemented in the CPR
- Improvement from 55% to 12% missing DI



# Novelties (1)

- ▶ New reference date (1 January)
  - ▶ Coherence with demographic and other statistics
    - ▶ In field enumeration 31 March
- ▶ Redefined statistical process according to the timeliness of input data
  - ▶ Four-stages productional scheme
    - ▶ No change of data from previous stage
  - ▶ Followed by dissemination after every stage
    - ▶ Final data only

## Novelties (2)

- ▶ Almost completely automated process
- ▶ Set up basic ORACLE tables with two shadow metadata tables
  - ▶ Status of record – any change recorded
    - ▶ All records vs highest value of the record
      - ▶ 9,413 mio records vs 2,113 records
  - ▶ Information which editing process changed data
    - ▶ Calculation of quality indicators for each variable
      - ▶ Identifiers automated correction rate – 4.2%
      - ▶ Identifiers manual correction rate – 1.1%

## Novelties (3)

- No Post Enumeration Survey
  - No need to verify / check / control administrative data
- Special survey on over-coverage conducted in 2016
  - Based on register-based labour force data

More in the afternoon session

# Some findings

- Better quality of outputs in the register-based census than in traditional one
  - No problems with
    - Under-coverage (2% in 2002 Census)
    - Double-counting (1% in 2002 Census)
  - Over-coverage around 0.8% at the individual level
- Quality of input data is improving
  - Labour force status imputation rates
    - 2011 – 1.81%   2015 – 1.32%   2018 – 1.13%

# Linkage of data – basic registers

CPR - Central Population Register

PIN	Address	Dwelling number	
-----	---------	-----------------	--

HR - Household Register

PIN	Address	Dwelling number	Household number
-----	---------	-----------------	------------------

RER - Real Estate Register

	Address	Dwelling number	
--	---------	-----------------	--

# Linkage of data – other sources

## Employment Register

PIN	Business Number
-----	-----------------

## Business Register

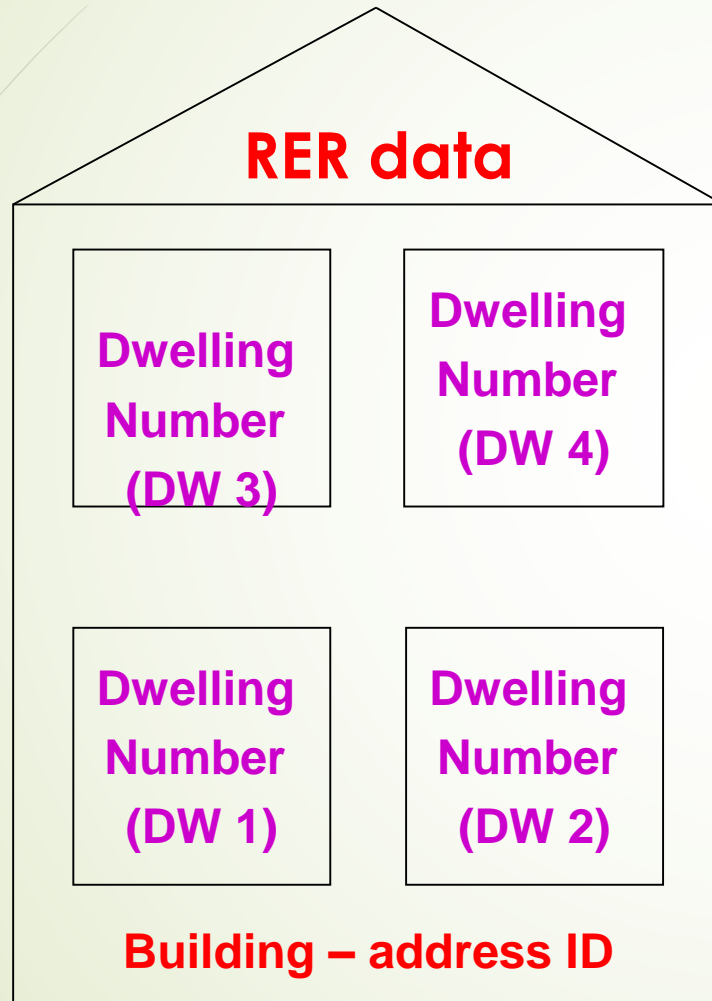
	Business Number
--	-----------------

## All other population data sources

PIN
-----



# Linkage of data – input data



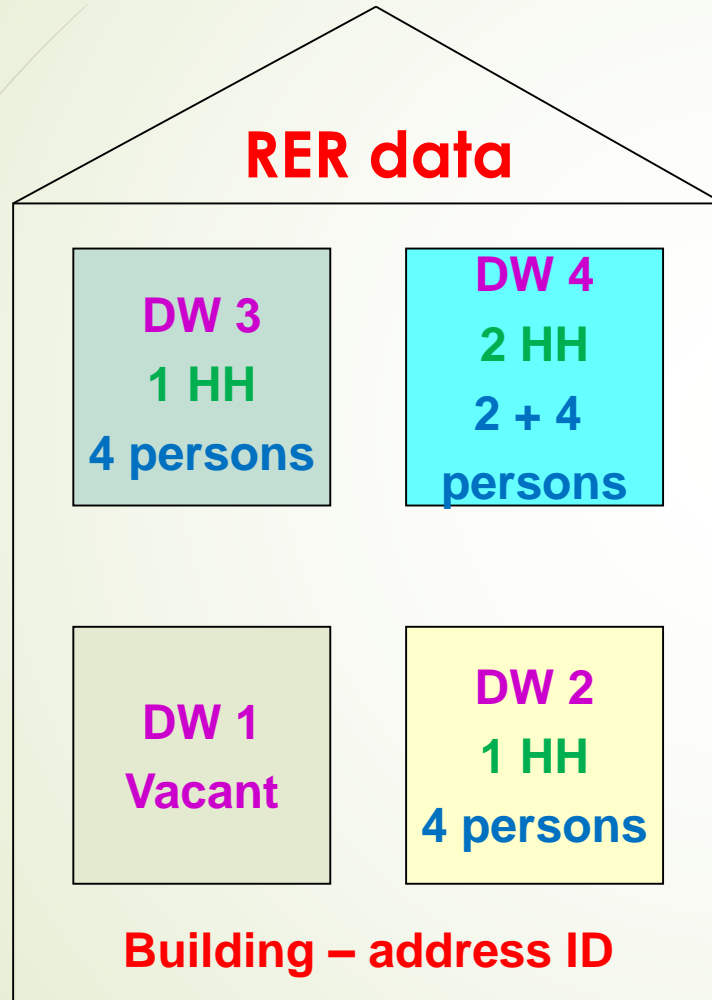
PIN	Address ID	DW		PIN	Address ID	HH
100089700	23470898	4		100089700	23470898	4
108979529	23470898	3		108979529	23470898	1
123457805	23470898	3		123457805	23470898	1
135790740	23470898	4		135790740	23470898	4
145092232	23470898	4		145092232	23470898	3
250789532	23470898	3		250789532	23470898	1
340090023	23470898	2		340090023	23470898	2
345678149	23470898	2		345678149	23470898	2
498230857	23470898	3		498230857	23470898	1
567725951	23470898	4		567725951	23470898	3
658735773	23470898	4		658735773	23470898	4
789568391	23470898	4		789568391	23470898	4
897600036	23470898	2		897600036	23470898	2
987650128	23470898	2		987650128	23470898	2

**CRP data**

**HR data**



# Data integration – output data



PIN	Address ID	DW		PIN	Address ID	HH
108979529	23470898	3		108979529	23470898	1
123457805	23470898	3		123457805	23470898	1
250789532	23470898	3		250789532	23470898	1
498230857	23470898	3		498230857	23470898	1
897600036	23470898	2		897600036	23470898	2
345678149	23470898	2		345678149	23470898	2
340090023	23470898	2		340090023	23470898	2
987650128	23470898	2		987650128	23470898	2
145092232	23470898	4		145092232	23470898	3
567725951	23470898	4		567725951	23470898	3
658735773	23470898	4		658735773	23470898	4
100089700	23470898	4		100089700	23470898	4
789568391	23470898	4		789568391	23470898	4
135790740	23470898	4		135790740	23470898	4

**CRP data**

**HR data**

# Traditional vs. register-based census

---

	<b>2002</b>	<b>2011</b>
Number of NSI employees	100 +	6
Number of field work staff	10,000	-
Budget	10 mio EUR	-
First final results time span	1 year	4 months
Improved quality	Sub(o)bjective	Systematic
Security of personal data	Under risk	No risk
Frequency	Every 10 years	Q, A, 3-4 Y