

Population census micro data for research: the case of Slovenia

Danilo Dolenc

Statistical Office of the Republic of Slovenia

Ljubljana, First Regional Workshop, 25 April 2013



Outline:

- 1) Population and Housing Census 2002 in Slovenia
- 2) On-site tailor made access to census microdata for researcher in the Statistical Office
- 3) Census microdata in Social Science Data Archives (ADP - Slovenia)
- 4) Census microdata in Integrated Public Use Microdata Series (IPUMS International) at University of Minnesota (USA)
- 5) Conclusion

Census 2002 in Slovenia

- Last traditional census
 - But with pre-printed questionnaires (CPR data)
- Partly register-based
 - 10 topics entirely taken from registers
 - 6 partly (missing information collected in the field)
- Databases in Oracle (very limited access to data within SURS)
 - Population -150 variables
 - Households – 40 variables
 - Families – 40 variables
 - Housing -125 variables (90 for dwelling, 35 for building)

On-site access - SURS (1)

- Scientific use file
 - Available for registered researchers only
- On-site laboratory or remote access
- Anonymization
 - No names, no address, no direct identifiers
- Statistical identifiers included if needed
 - Possibility to link census data with other survey data

On-site access - SURS (2)

- Need to know principle – tailor made database for research
 - All variables available
 - Except data on ethnicity and religion (protected sensitive data)
 - Limitation of territory
 - Lowest territorial unit: enumeration district
 - Selected variables only - according to the request
- No additional statistical disclosure control applied in microdata sets

On-site access - SURS (3)

- Outputs – mostly aggregated tabular data
- Outputs checked by statisticians
 - To be in accordance with the dissemination and confidentiality policy of 2002 Census data
- Small number of requests
 - Less than 10
 - The first one in 2008

On-site access – ADP (1)

- Public use file
 - The main purpose – to learn how to use microdata and metadata
 - Predominantly for students
- No confidential information
 - 5 % sample of population living in private households and conventional dwellings
- Lowest territorial data – statistical region
- Selected variables only
 - 27 for population
 - 39 for housing

On-site access – ADP (2)

- Two steps of disclosure control
 - Aggregation of values (e.g. 5 years age group, first digit level of occupation classification only; country grouping)
 - Threshold implemented
- Three microdata sets available
 - Housing (buildings) – 9 variables – information loss 2%
 - Housing (dwellings) – 30 variables – information loss 8.5%
 - Population – 27 variables – information loss 10.2%

On-site access – ADP (3)

- Number of users
 - Academic year 2010/11 – 16
 - Academic year 2011/12 - 14
- Most wanted SURS microdata – Safety Survey 2001
 - Academic year 2010/11 – 48
 - Academic year 2011/12 – 38

Slovenian microdata and IPUMS International (1)

- Public use file
 - Slovenia is one of 68 countries included
- Main goals of [IPUMS International](#)
 - Collect and preserve data and documentation
 - Harmonize data
 - Disseminate the data absolutely free
- Sample frame
 - Private households with 1-6 members living in conventional dwellings
- Sample – 10 % (systematic sampling applied)

Slovenian microdata and IPUMS International (2)

- Similar microdata procedure as for ADP – done by SURS
 - Lowest territorial data – statistical region
 - Aggregation of values
 - Threshold implemented
- Three microdata sets available
 - Housing (buildings) – 5 variables / 2 protected – inf. loss 0.16%
 - Housing (dwellings) – 15 variables / 5 protected – inf. loss 0.81%
 - Population – 37 variables / 8 protected – information loss 23.0% (of which 19 % variable REGION)
- Identifiers for linking microdata sets included

Slovenian microdata and IPUMS International (3)

- 185 extracts in 2011 from Slovenian sample*
- Rank 45 of 55 countries
 - Rank 1 (Brasil) – 712 extracts

*Data source: Safety in numbers: the IPUMS/IECM trans-border, integrated census micro data dissemination system minimizes risks and maximizes benefits;
Working Group "Demography & Census, Luxembourg, September 2012

Conclusion

- 2011 Register-based Census
 - Same rules for SURS on-site access
 - Microdata set for IPUMS is already in the procedure
 - Methodology as for 2002 microdata-set
 - Deadline June 2013

Thanks for Listening

Contact: danilo.dolenc@gov.si

Website: <http://www.stat.si/popis2011/>

