Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Sixtieth plenary session
Paris, 6-8 June 2012
Item 3(a) of the provisional agenda
2010 round of censuses – innovations and lessons learned

Challenges in the transition from traditional to register–based census in Austria

Note by Statistics Austria

Summary

The paper presents the milestones in the transition from conventional to register–based census in Austria. Furthermore it describes a structural approach for the quality assessment of administrative data. The process based on three stages (raw data, combined data, imputed data) derives quality indicators that should cover all the available quality information for each attribute. To guarantee the applicability of the quality framework for the register–based census 2011, the procedure was tested by the use of data from the register–based labour market statistics 2009.

An advantage of the approach is that the quality assessment and data processing are independent. This offers the possibility to apply the methods on other register–based data sets. Moreover, the cooperation between the national statistical organization and the register authorities is intensified because the data holder is integrated in the quality assessment process.

Please recycle

I. The evolution from traditional to register–based census in Austria

1.   In May 2001, the last traditional population census was carried out in Austria, accompanied by a building and housing census as well as a census of local units of employment. This combined census covered approximately 2.0 million buildings, 3.8 million dwellings and 3.3 million households containing 8.1 million persons by using paper questionnaires. Thus traditional census was a sophisticated and costly effort.

2.   Given the challenging requirements for traditional censuses, the importance of administrative data sources for statistical purposes has recently been rising. The processing of data which has already been recorded by administrative authorities offers numerous advantages compared to survey data such as diminishing costs, removed burden for respondents and the prompt availability of the data. Using data from existing registers should ensure an optimal reflection of reality at reasonable expenditures by combining registers via unique linking variables, enhancing data quality and harmonizing definitions. An obvious advantage is exhibited by the regular updates of register information in order to keep track of any changes in the data describing the units and their attributes (Statistics Finland, 2004, p. 10).

3.   Thus, an increasing number of National Statistics Offices (NSOs) promote register–based censuses as a replacement for costly conventional censuses. In 2000, the Austrian council of ministers decided to establish the new method for the census 2011 and in 2006, the regulatory framework came in force by the juridical authorities. Austria is among six European countries (besides Sweden, Finland, Norway, Denmark and Slovenia) that carried out a register–based census in 2011.

4.   Based on the number of population from this census, the monetary amounts of fiscal equalization between municipalities and the federal financial authorities as well as the number of eligible voters in the prospective elections are determined. Furthermore, information on commuters, education and employment offer important insights for economic and social policies. However, administrative data may apply definitions that differ from the needs of the NSO even though the data are of good quality (see United Nations, 2007, p. 3). Therefore, the NSO has to decide whether the data is adequate for the issue of interest.

A.   The principle of redundancy

5.   A key problem arises with selecting appropriate data sources for supplying the required information. The register–based census aims at covering all relevant variables that were formerly provided by a traditional census. In this respect, the census in 2001 was the initial spark for the creation of some data sources, e.g. the Central Population Register (CPR), the Housing Register of Buildings and Dwellings (HR) or the Register of Educational Attainment (EAR). Prior to 2001, an interconnected network of population records did not exist in Austria. Each municipality administrated its own records and usually the data was not even entered into electronic systems. Introducing the housing register in 2001, a centralized population register evolved and the municipalities had to provide their records. The last traditional Austrian census was based on information from this newly created population register. In 2004, the Housing Register of Buildings and Dwellings was synchronized with the population register for the first time. Moreover, the Register of Educational Attainment was founded during the census process of 2001.

6.   Figure 1 illustrates all base registers of the census and their links to the respective topics. The red-shaded data sources are maintained by Statistics Austria, the remaining information is provided by external data holders, like the Public Employment Service Austria (AMS) or the Austrian Social Security (SVA). The Central Population Register (CPR) forms the backbone of the census, since the units of analysis are individuals with main residence in Austria. To assure the quality of the census results, the base registers are backed up by seven comparison registers. These seven fields of administrative units are provided by 35 data holders and are mainly used for cross checks as well as the supply with information that is not available or only partly available in the base registers (Berka et al., 2010, p. 300).

7.   Given the independence of the various registers as well as the autonomous process of data collection, the sources sometimes contain contradictory values for the same variable. Therefore, the principle of redundancy is used to ensure sufficient quality by acquiring information on sex, nationality or age from as many registers as possible. A particular method developed by Statistics Austria aims at identifying one particular base register to provide the information for a certain variable, whereas the comparison registers are used to confirm the values in the base registers (Lenk, 2008, p. 3). While the registers provide sufficient information for most of the characteristics of the conventional census, some variables could not be included in the register–based census. For instance, the duration of the daily commute, colloquial language or religion are not captured by any of the registers at hand.
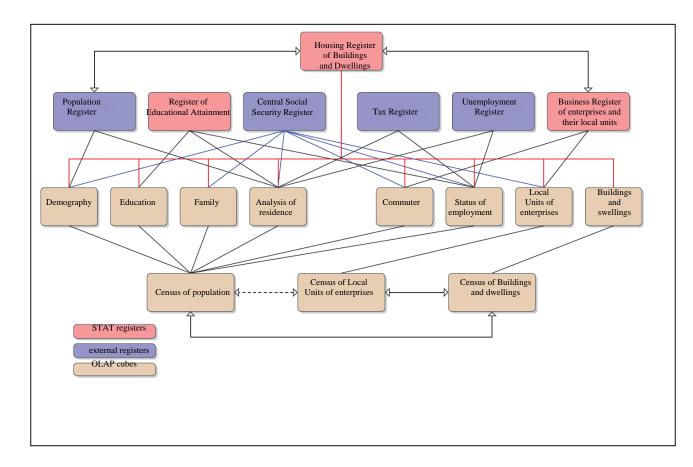
B.   Data privacy protection

8.   Since the names and social security identification of individuals are not part of the data delivery, a unique identification number is needed for merging information from different data holders. To ensure data privacy the introduction of a branch-specific personal identification number for official statistics (bPIN OS) is required before delivering the data to Statistics Austria. Each administrative branch in Austria, like "social security", "taxes" or "social welfare", has its own bPIN. These 172-digit PINs, which should serve for privacy protected communication between public authorities via e-government, are derived by the Austrian Data Protection Commission (DPC) from the SourcePIN Register (Stammzahlregister).

9.   Figure 2 illustrates the achievement of data privacy protection through the derivation of branch-specific as well as PINs for official statistics. The data holders demand these PINs from the DPC for each person by delivering the name, sex, date of birth, birth place and address of an individual. The branch-specific PIN (bPIN) as well as the PIN for official statistics (bPIN OS) are derived from the SourcePIN Register, using a special and very complex algorithm developed by the DPC. The bPIN OS for register owners other than Statistics Austria is only provided in an encrypted form. The data holder has to send the data together with the encrypted bPIN OS to Statistics Austria. Moreover, the register owner delivers its own encrypted branch-specific PIN for each individual in order to identify the respective record in case of further inquiries by Statistics Austria (Lenk, 2008, p. 5).

10.   Each data holder is able to decrypt its own bPIN into a 28-digit number, hence only the NSO is able to decipher the bPIN OS and to use it as a common linking variable for the incoming register data. The described procedure of encrypting personal identification was first applied in a test census in 2006.
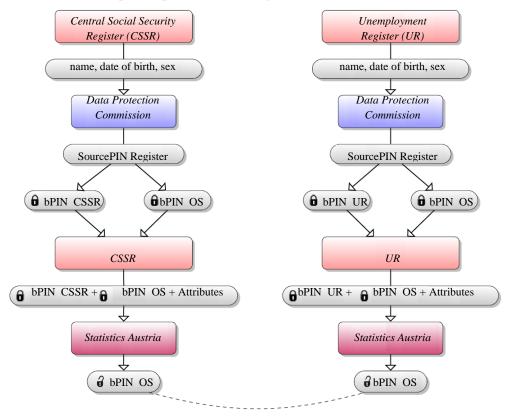
Figure 1:
Registers and topics in the Austrian register–based census

Figure 2

Achievement of data privacy protection in the register–based census

II. Implementation of preliminary censuses

A. The test census in 2006

11. In 2006, Statistics Austria conducted a test census with the scope of a full census of population, houses and dwellings as well as local units of employment to evaluate the transition from conventional to register–based census. For the first time, information from the various data holders was demanded for combined administrative statistics. Disregarding minor problems with the data sources, the test census was a general success. Particularly data on demographic issues, education and employment turned out to feature excellent quality and plausibility, while problems arised in identifying the job location for commuters and in associating individuals to dwellings (Statistics Austria, 2009, p. 7). Statistics Austria has validated the quality of the results by comparing them with a sample survey for the same reference date (Lenk, 2008, p. 14).

12. In 2008, the implementation of the fiscal equalization law marked a decisive turning point in the history of Austrian census. The law pegged the distribution of federal funds to municipalities to the test census–based population number, which converted the initial nonbinding test census into a politically relevant instrument.

B. Residence analysis

13. Ensuring the quality of the enhanced test census, it was important to detect missing values, multiple records or measurement errors and eliminate them for counting purposes. This procedure known as residence analysis numerates five reasons for excluding individuals from the census:

(a) The person has died before the reference date, but still shows up in the CPR;

(b) The individual shows multiple records at the reference day, for instance two or more main residences in the CPR;

(c) The person was registered less than 90 days with a main residence in Austria, which links the population number to a minimum stay;

(d) Detection of so-called "census tourism", i.e. persons who are registered in a municipality for not longer than 180 days but in another one before and after the reference period;

(e) Cases of possible discrepancies without a main residence, i.e. persons who are only recorded in the CPR but in no other administrative source

14. The latter case requires clarification by the NSO, since the signs of life are not satisfactory (see Table 1). Individuals who are only covered by the CPR implicitly need clarification and are questioned in a written form[1], which would be the case for individual ID3459 in this example. Following the test census, some 45,000 letters were sent to cases of clarification, whereof 9,000 individuals affirmed their main residence in Austria. Finally, approximately 0.5 percent of the initial population was not counted, including about 10,000 technical deletions (see points 1-3). Austrian municipalities have to be informed about non-counted individuals whereby registration authorities have the chance to prove the residence of these cases and possibly remove the individuals from the residence register. Due to the results of the test census in 2006, about 80 percent of the non-counted cases were removed from the residence registers by local municipalities.

15. For the register–based census of 2011, Statistics Austria applies the same procedure like in the test census of 2006. The residence analysis has been started in January 2012. In a first round, about 54,000 letters to clarify main residence were conveyed. The second round is planned in September 2012, because the data transmission from the external data holders will be completed.

---

[1] The letter, which is sent to all cases of clarification, contains only one single question: "Did you have your main residence in Austria at the reference date? - Yes or No."

Table 1
Signs of life in residence analysis

bPIN	OS	CPR	CSSR	TR	UR	SWR	CAR	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
ID3457		✓	✓	✓	⚡	⚡	⚡	
ID3458		✓	✓	⚡	✓	⚡	⚡	
ID3459		✓	⚡	⚡	⚡	⚡	⚡	
ID3460		✓	⚡	⚡	⚡	✓	⚡	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

CPR=Central Population Register, CSSR=Central Social Security Register, TR=Tax Register, UR=Unemployment Register, SWR=Register of Social Welfare Recipients, CAR=Child Allowance Register

C. Register–based labour market statistics

16.   Particularly comprehensive opportunities for real time statistics are offered by the annual implementation of register–based labour market statistics. From 2009 onwards, data on demography, education, commuters and labour markets are provided on an annual basis and permit detailed insights into Austrian economic and social issues even on a municipality level. For instance, the frequent request of administrative data allows in-depth analysis for the changes in youth unemployment between two particular years or the development of gender-specific differences in educational attainments. The data sources used as well as the processes applied for generating these attributes are identical to the methods for the register–based census. Some 35 different administrative and statistical data sources are linked to obtain the variables of interest. With the implementation of register–based labour market statistics, the NSO meets the standards of contemporary statistics and data analysis. The data are even publicly available in the statistical database on the homepage of the NSO (www.statistik.at).
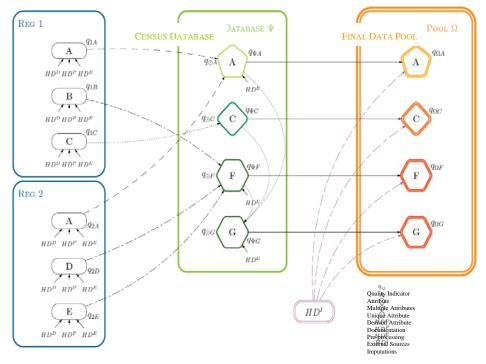
III.   Considerations of a framework for quality assessment

17.   The short transition time from a traditional to a register–based census has been a challenging task in Austria. The interim period for gradually substituting survey data with administrative data lasted about 20 years in some European countries (see Ruotsalainen (2008) for the example of Finland). This allowed an intensive discussion on quality assessment by the NSOs as well as the data owners in some countries, whereas the transition schedule was very ambitious in Austria. Since the NSO is not responsible for the maintenance of the external data, the necessity of quality assessment in the process of register–based censuses has to be singled out. The quality analysis of register data has to satisfy several requirements such as transparency, accuracy and feasibility. Our approach contributes a quality framework for the analysis of administrative data using three different hyperdimensions for the derivation of quality indices (see Berka et al., 2012). The framework is closely tied to the data flow yet independent from data processing, which ensures that the processing is not influenced but evaluated.

18. The data flow of register–based census in Austria consists of three levels: raw data (i.e. the registers), combined dataset (Census Database, henceforth CDB) and imputed dataset (Final Data Pool, FDP). Figure 3 illustrates the data processing, beginning with the receipt of raw data from the various administrative data holders. The information is connected via the unique key (bPIN OS) and merged to data cubes in the CDB. Further, the CDB data are enriched with imputations of item non-response which complete the FDP containing real and estimated information.

Figure 3
Quality framework for register-based censuses

A. Quality assessment on the register level

19. Information on quality at the raw level is obtained via three hyperdimensions: Documentation (HD^D), Preprocessing (HD^P) and External Source (HD^E). Prior to seeing the data, HD^D describes quality-related processes in the register authority as well as the documentation of the data (i.e. metadata). The degrees of confidence and reliability of the data holders are monitored by the use of a questionnaire containing 16 open-ended and nine scored questions (e.g. administrative purpose or data treatment). The NSO is therefore able to check for data collection methods or legal enforcements of data recording which may significantly influence the quality of the data. The questionnaires are answered by experts from the respective data holders and should thus deliver convincing results.

20. The second hyperdimension HD^P is concerned with formal errors in the raw data. Range errors in the data, item non-response and missing primary keys are detected in this step of the quality framework. The final result of this hyperdimension is given by the ratio of *usable records* to the *total number of records*. Again, this procedure is carried out for each attribute in each register. If the proportion of usable records for an attribute in a certain register is smaller than that of the same attribute within another register, the quality measure will accordingly be lower.

21.   Finally, the third hyperdimension HD^E provides a comparison between the register–based data and an external source. In Austria, the microcensus is a common benchmark for representative surveys and is assumed to be the best comparative dataset available. Checking for consistency with the external source offers the third quality measure which is the ratio between the *number of congruent values* and the *total number of linked records*. If the attribute of interest is not covered by the benchmark, we rely on local expert opinions.

22.   Given these three quality measures, an overall quality indicator for each attribute and register can be derived as a weighted average. Hence, this indicator is able to capture quality-related effects ranging from the data generation to the final raw data in the registers.

B.   Quality assessment in the census database

23.   The entire information from the registers is combined in the CDB which covers all attributes of interest for the census. Since there may be more than one data source providing a certain attribute, a ruleset predefined by the NSO picks the most appropriate information from the underlying registers. Concerning the evaluation of quality for the CDB we distinguish three types of attributes by their origin:

(a)   Unique attributes exist in exactly one register, e.g. educational level (cf. attribute C in figure 3);

(b)   Multiple attributes show up in several registers, e.g. sex (cf. attribute A in figure 3). The information from multiple sources are combined by a ruleset to derive the most appropriate value in the CDB attribute;

(c)   Derived attributes are created based on different attributes, e.g. employment status (cf. attributes F and G in figure 3). The registers do not contain any information for these attributes in the required specification.

24.   A detailed description of the quality assessment for the three types of attributes in the CDB is given by Berka et al. (2010) and Berka et al. (2012).

25.   Current research is focussed on the Final Data Pool (FDP), which corresponds to the Census Database after the imputations were applied. The amount of item non–response is effectively reduced by imputations, however the imputation process itself has to be monitored. This is done by using information from the hyperdimension Imputation (HD^I), which is an ongoing task.

C.   Conclusion and outlook

26.   This paper presents the milestones in the transition from conventional to register–based census in Austria. Furthermore it describes a structural approach for the quality assessment of administrative data. A process based on three stages (raw data, combined data, imputed data) derives quality indicators for three hyperdimensions. These measures should cover all the available quality information for each attribute. To guarantee the applicability of the quality framework for the register–based census 2011, the procedure is tested by the use of data from the register–based labour market statistics 2009.

27.   A decisive advantage of the quality framework at hand is the independence of quality assessment and data processing. The separation from the processing procedure is required to evaluate the process without exerting influence on it. This offers the possibility to apply the methods on other register–based data sets. Moreover, the cooperation between the NSO and the register authorities is intensified because the data holder is integrated in the quality assessment process.

IV. References

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2010). A quality framework for statistics based on administrative data sources using the example of the Austrian census 2011. Austrian Journal of Statistics, Volume 39, Number 4, 299-308.

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2012). Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based census 2011. Statistica Neerlandica, Volume 66, Issue 1, 18-33.

Lenk, M. (2008). Methods of Register-based Census in Austria (Tech. Rep.). Statistik Austria, Wien.

Ruotsalainen, K. (2008). Finnish Register-based Census System (Tech. Rep.). Statistics Finland.

Statistics Austria. (2009). Bericht ¨uber die Probez¨ahlung 2006 - Ergebnisse und Evaluierung (Tech. Rep.). Statistics Austria.

Statistics Finland. (2004). Use of registers and administrative data sources for statistical purposes (Tech. Rep.). Statistics Finland.

United Nations. (2007). Register-based statistics in the Nordic countries. United Nations Economic Commission for Europe, New York and Geneva.