Seminar on Registers in Statistics - methodology and quality 21–23 May, 2007, Helsinki

Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case of Slovenia

Apolonija Oblak Flander
Statistical Office of the Republic of Slovenia
Apolonija.Oblak@gov.si

Chapter 1. Introduction

Similarly to Scandinavian countries, Slovenia has a long tradition of establishing various registers, linking them in a system and using register-based data for statistical purposes.

In line with the traditionally register-based orientation in Slovenia (Slovenian statistics was among the initiators of the establishment of a linked system of administrative registers on the basis of a unified identification code system) and on the basis of long-lasting efforts within the national statistics to use the administrative sources for statistical production, the Statistical Office of the Republic of Slovenia (SORS) has started with the preparations for the first (completely) register-based census of the population and housing in 2011.

Chapter 1.1 Lessons learned

Since the Slovenian National Statistical Institute (NSI) has always been among the initiators of the establishment of a system of administrative registers on the basis of a unified identification code system, the establishment of the first Central Population Register (CPR) was initiated by the NSI. The very beginnings go back to the early 1970s. The data from the 1971 Population and Housing Census were used for the establishment of the CPR and for the very first time personal identification numbers (PINs) were assigned to the people residing in Slovenia. This PIN was a precursor of the present PIN in use and the numbers in it only reflected the code of the way of the assignment, code of the municipality, serial number and control number. It was not the best solution and after some study visits to NSIs of Denmark and Sweden the Slovenian NSI suggested the introduction of a PIN as already used in Scandinavian countries.
There was a decision in the ex-Yugoslavia that all republics will go in the direction of establishing registers of persons and in the introduction of PINs. As a consequence of those decisions the federal Act on the PIN was adopted in 1976[1]. In 1980 new PINs were assigned to all persons in Slovenia and in the following year among other variables collected at the 1981 Census of the Population and Housing.

Until the 1991 Census of the Population and Housing, census data were used for the establishment of the CPR. In the 1991 Census there was a shift in the direction of the transfer of the data. For the first time the CPR data were used for the pre-printing of the census questionnaires (with PINs, names, surnames, the exact addresses, census district codes and municipality codes). In addition, before the census a pre-census database was established. In that database not only the CPR data were used. Also data on gender, educational attainment,

[1)] Irena Tršinar, Centralni register prebivalstva: oris razvoja, komentar zakona in prevod v angleški jezik. Ljubljana, Uradni list RS, 1999.

educational qualifications, occupation, industry, socio-economic status in employment, etc. were transferred from the Employment Database. Although at the 1991 Census some data were compiled in the pre-census database, all census topics were still on the census questionnaire. They were replaced in the phase of data processing if they already existed in the pre-census database.
The change in the way of conducting the census of population and housing was done in order to simplify the field enumeration, to rationalise the phases of census data capturing and to increase the quality of the main census data (the number, gender, age, geographical distribution of the census units).

In 2002 the so called combined Census of the Population and Housing in Slovenia was conducted (data partly obtained from administrative and statistical sources and in addition field enumeration was conducted). Although SORS followed the experiences gained at the 1991 Census, the 2002 Census was a huge step forward. In the light of the preparation for the 2011 register-based census there are four fields in which essential knowledge and experiences were gained:

- knowledge in the field of the content of the main administrative and statistical registers;
- knowledge on compiling the pre-census database and its use for the data processing and production of the census data;
- experiences on the use of the CPR data in the phase of logical controls;
- experiences and knowledge on defining the final census database.

Chapter 2. Opportunities of a register-based census

With a register-based census the 'census data are produced using the method of register estimation, in which several register sources are simultaneously used to define for each statistical unit the value of the relevant variable'[2].
Good experiences of countries which have already conducted or are conducting register-based censuses encouraged other countries to follow them as well.
What are the main opportunities which come out of a register-based census? In literature[3] we can find some (although there are also some unwritten). Among them:

- huge reduction of costs;
- shorter interval between censuses – some (or even all) census statistics can be compiled also annually and not only every 5 or 10 years;
- data collected once and processed only if and when they change (many census topics namely seldom change, such as marital status, number of rooms in the apartment, etc.);
- detailed geographical information on the census units;
- no special census legislation needed - the experiences of so-called 'register–based countries' show that the existing legal basis (for the national statistics in general) is sufficient when conducting register-based censuses. So, no additional burdening of the NSIs concerning special census legislation is needed.

Of course, there are not only advantages of a register-based census, but also some disadvantages[4]:

[2)] Recommendations for the 2010 Censuses of population and housing. United Nations, New York and Geneva, 2006, p. 158.
[3)] Recommendations for the 2010 Censuses of population and housing. United Nations, New York and Geneva, 2006, p. 158-159.
Myrskylä, Pekka, Use of Registers and Administrative Data Sources for Statistical purposes. Helsinki, Statistics Finland, 2004.
[4)] Ibid.

- dependence on the information available in the registers – some data items have to be dropped from the register-based census because there is no relevant information available in the registers;
- NSI dependent on the register authorities;
- the content in the registers dependent on the changes in legislation and administrative practices;
- when using several register sources simultaneously to define the value for each unit, decision rules have to be defined in such a way that the data come as close as possible to the data collected by means of questionnaires;
- consistency problems - when the same data are obtained from different data sources;
- reference period – a register-based system can create problems with the reference period. Namely, for reasons of statistical reliability it is important that changes of events are accurately recorded according to their true date. In the registration system this is rarely the case. Usually, data are good as regards recording the dates of births and deaths, because there are usually special certificates following those vital events. Accurate information is also very often obtained for the dates of employment, unemployment, etc., whereas for periods of studying the dates are less accurate. The same problem is also with the dates of change of a person's address etc.;
- linking of the data can be a very difficult or almost impossible task, if the unified identification code system does not cover all registers (databases).

Chapter 3. Challenges to meet the preconditions of a register-based census

It is well known that the development of a register-based population census system is a long process, which in many countries stretches over 2, 3 or even more decades. Consequently, the question 'WHY?' appears. The answers lay in the necessary preconditions which need to be met when NSIs are planning register-based censuses.

Chapter 3.1 General preconditions for a register-based census

Among the necessary preconditions for a register-based census are:

Legal basis

The appropriate legal basis provides a key foundation for the use of administrative sources for statistical purposes. National legislation must allow the use of existing administrative sources for statistical purposes rather than re-collecting data whenever it is possible. Furthermore, the appropriate legislation should provide a detailed definition of data protection.

Public approval

It is extremely important that the public in general appreciates and understands the benefits of using register sources for statistical purposes (in particular for the use in the census of population and housing). The precondition for that is an open discussion and debate, explaining the rationale and benefits of the use of registers. Very important contribution to the public acceptance of the use of administrative data for statistics production are open and transparent activities of authorities in charge of the maintenance of the registers and clear and up-to-date national register legislation.

Unified identification code system

For the linking of the data maintained in the administrative registers, it is vital that the countries have introduced a common identification number across different sources before they combine the data from different administrative sources. All (census) statistical units need to have such

identification that can be linked to another by means of an identification system. The data linkage must be possible at the individual level.

And more, the system needs to assure the location of census units by using local area codes or map coordinates or any other geographical identification.

Availability of reliable administrative registers

A register-based population census system is built around a set of basic registers that contains comprehensive data on units that are to be described in the population and housing census. These registers may include the data maintained in a population register and a register of buildings and dwellings, as well as data from a business register. But the precondition for a register-based census is not only the existence of the main registers, but also their reliability or, in other words, their stability.

Cooperation among administrative authorities

A harmonised system of administrative registers and consequently the register-based statistics requires a close cooperation of relevant responsible bodies and an explicit commitment from the highest possible level in the country.

Human resources

It is very important that the NSIs have human capital - people working on censuses not only with the knowledge of the census content ('philosophy' of counting the census units, core and non-core variables, etc,) but also with new knowledge related to the new techniques of conducting censuses and with the understanding of the functioning of the register-based environment and the system of administrative (and statistical) registers.

Chapter 3.2 Difficulties with meeting some general preconditions for a register-based census in Slovenia

Although the Slovenian NSI is in early preparations for the 2011 register-based Census of the Population and Housing, we are already aware of some major challenges which will have to be met and solved in order to produce not only relevant (and according to the proposed EU regulation of population and housing census obligatory topics) but also high quality census data.

The main obstacles which the Slovenian NSI is facing at the moment are the non-existence of a dwelling register, the non-existence of a stable and well maintained household register and bad data on the relationships among people in the CPR, which would give the necessary information on forming the families.

Register of Dwellings
The register will hopefully be established in the following years, although the Census of the Dwellings and Buildings, managed by the Surveying and Mapping Authority of the Republic of Slovenia, is at the moment underway. SORS is not only concerned about the establishment of the mentioned register in a proper way, but also on the quality of the data in it. The existing legislation will enable the transfer of some data on dwellings (building number, dwelling number, etc.) to the CPR, which will enable to form dwelling-household units.

Register of Households
The informatization of households, which used to be kept on paper cards at the local registrars, was a huge step forward towards the Register of Households. At present the data were captured in 95-97%, but the register is not updated together with the changes in the e-Register of Births, Deaths and Marriages and not even with the Register of Permanent Residents. This in a very short period after the capture of the data in the Register of Households caused big discrepancies.

If the quality of the data in the Register of Households is not sufficient, SORS will follow the Scandinavian concept of forming households on the basis of dwelling-household units.

Families

The Slovenian NSI will have to reach for alternative ways of gathering and compiling of the core and non-core variables not available in administrative and statistical sources.

In the light of assuring high quality statistical data with a register-based census, for SORS data on families are among the most challenging ones. In the past traditional population and housing censuses represented the main source for the data on families, because the data in the administrative and statistical registers were not (and still aren't) available. Although the needs in the Slovenian society for the data on families due to the reorganisation of the family life, very low fertility, ageing and consequently the changed forms of families are huge. Not only the governmental sector is interested in the detailed data on the families (in order to prepare the analysis and consequently relevant policies and strategies for many fields of social life connected with the family life), many researchers are focusing on the family topics.

Due to the above mentioned facts, SORS is intensively preparing two strategies to assure the statistics on families.

The first strategy includes the use of existing data on relationships of persons in the dwelling (after the dwelling number is available in the CPR!). The biggest difference, if we compare the Slovenian situation and the situation in the Scandinavian countries is, that in the CPR some relationships between the family members exist, but only for the vital events such as births, marriages, etc., that took place after April 1988. The information on relationships in the family is not so complete as to enable simple forming of families. Families could be formed on the basis of data on age, marital status and gender of persons residing together in a dwelling unit, with the precondition that a dwelling number is available in the CPR. In addition, for some population groups also data on the relationships in the family from the 2002 Census could be taken (for example for people who have not moved since the last Census).

If the dwelling number is not available in the CPR, SORS will have to use alternative ways of forming the families. Alternative methods already applied in some countries were studied and further development of the statistical methods to compile 'statistical families' will be a priority task immediately when the situation concerning the dwelling number in the CPR will be clear.

Of course, there is still an option to introduce an 'ad hoc' module in one of the largest household surveys, which would enable to collect information on the families. However, since the data would be gathered on a sample (samples are usually made on the basis of selected persons and not households), it will be difficult to achieve a representative sample for the lower geographical levels, such as NUTS3 (in Slovenia this is the level of statistical regions) and LAU2 (in Slovenia the territorial level of municipalities).

Nationally interesting recommended census topics

In Slovenia there is a long tradition of gathering recommended topics such as national affiliation, mother tongue and language in the household. The collection of those data is connected with the origin of the Slovene identity and with the long tradition of cohabiting with other nations in larger states (firstly under the Austro-Hungarian Monarchy, later in Yugoslavia). Since there are also two autochthonous minorities (Italian and Hungarian) and one autochthonous ethnic group (Roma people) present on the territory of Slovenia and due to the long cohabitation with other ex-Yugoslav nations, the need for such data among researchers and governmental institutes still exists. Because the data such as national affiliation, mother tongue and language in the household do not exist for the whole population in the administrative records, the NSI is thinking about using alternative methods of gathering these census topics.

One of the options is to gather the recommended and nationally interesting census topics at the 'ad hoc' module in one of the largest household surveys together with the data on families. Here the representativeness of the collected data at the lowest geographical levels is not that problematic, because the data on national affiliation, mother tongue and language in the

household have already at the 2002 Census been treated more carefully concerning the confidentiality measures (data disseminated only down to the level of municipalities – LAU2).
One of the nationally interesting topics is religion. Data on religious affiliation were on the territory of Slovenia gathered at the 1910, 1921, 1931, 1991 and 2002 Censuses. In the light of the preparation for the 2011 register-based Census of the Population and Housing, SORS will slightly change the concept of gathering data on religion. In the past religious affiliation was a free declaration of a person about his/her religious belief, about the identification with a certain religion, etc. People also had the option not to declare their religious affiliation and the option not to answer the question on religion. In the future SORS will initiate the debate among various institutions on this topic and on using the statistics and available administrative sources of various religious institutions on formal membership and religious attendance.

Chapter 3.3 Main activities within the phase of the preparation for the register-based census in Slovenia

At SORS many necessary activities are planned in the phase of the preparation for the register-based census. The first step will be to **identify the relevant administrative and statistical sources**. Partly this task has been already covered at the 2002 Census of the Population and Housing, but since then many data were informatizied and many new administrative sources were established - mostly within the Ministry of the Interior and the Ministry of Public Administration. At SORS at the moment a big **Social Statistics Database** (SSD) is in the phase of the establishment (see Chapter 3.4), in which data on persons will be linked on the basis of unified identification codes. In the future also linking with other units (such as dwellings) is planned. At present persons are 'placed' in the dwellings which were collected at the 2002 Census (dwelling identifiers from the 2002 Census will be assigned to persons where possible) but in the future the dwelling numbers will be replaced with the new ones from the Register of Dwellings.

After the identification of the relevant sources, **all relevant identifiers for linking census units** and their data will be analysed. Options for linking data in cases where the identifiers are missing will be studied and best **matching methods** will be proposed.

When the main sources and common identifiers are identified, a task of **identifying the census relevant variables** will take place. If variables are available in several sources, the priority list of sources for each variable will be set.

On the basis of the available data in the data sources and according to the UNECE/EUROSTAT Recommendations for the 2010 Censuses of Population and Housing the compiling of the **core topics, non-core topics and nationally interesting topics will be studied**.

The majority of the above mentioned tasks will be covered within the **Transition Facility 2006 Project** (Preparation for the 2010 Population Census) in the period from September 2007 to December 2008.

Chapter 3.4 Register-based census as a core database for the social statistics

As already mentioned, SORS is establishing a large Social Statistics Database, which will consist of numerous databases from various statistical fields. Data in it will be linked on the basis of the unified code system and standard classifications. The general scope of the SSD is as follows[5]:

- to set up an integrated system for linking both survey data and other statistical and administrative data for selected statistical domains in order to improve reporting

[5)] Erika, Žnidaršič, Jug, Matjaž, Tina, Žnidaršič, Terms of Reference: Setting up the Social Statistics Database for the Implementation of EU SILC. Ljubljana, 2006 (unpublished document), p. 3-5.

mechanisms and decrease the burden laid on reporting units which have to respond to various surveys;
- to enable effective production of social statistics and easier and faster use of data for analyses in the social domain.
- to enable sampling procedures for various socio-economic subgroups of population.

SORS has already been linking various administrative data on persons. Our experiences show that with our level of legal system development and SORS's competence this is possible. The 2002 Census – in which more than ten statistical and administrative sources were linked – showed just that. The SSD project is even more complex since even more administrative sources will be used to gather data on persons. The project will improve the infrastructure needed for producing the Slovene statistics in line with the EU statistical requirements. The main focus of this project is the EU SILC and the 2011 register-based Census of the Population and Housing.
The integrated information system will enable combined use of both survey data and administrative data and facilitate the decrease of the burden on reporting units. The burden of data providers will be reduced by cross querying across surveys or editing against administrative data. The system will also provide the environment for improved imputation of missing data items and whole records (through the use of alternative data already present in the system) along with the ability to analyse the impact of individual and mass imputation on statistics. The process metadata will be stored and analysed in order to monitor the statistical process and to increase the efficiency of the system.

The role of the census data in the establishment of the SSD is **twofold.** The data from the census are the input and also the output.

How can the census data be the input as well as the output? Beside census data (at first from the 2002 Census), the SSD will have also other inputs, such as data from the CPR, the Register of Employment, the Register of Unemployment, the Register of Foreigners, the Register of Dwellings, the Register of Territorial Units, the Tax Database, social statistics surveys, etc. The linking of those data will enable us to conduct a completely register-based census already in 2011 or, if some of the necessary preconditions are not met, in the near future.

Chapter 4. Quality report – the final challenge and a mirror of the work done

In the phase of the preparation of the *UNECE/EUROSTAT Recommendations for the 2010 Censuses of Population and Housing* there were many discussions on the quality assessment of the census as one of the statistical surveys. Until 2000 census round the census quality reports have been in the countries prepared without an unified frame. Since the 'demand for regular and systematic assessment of quality of the statistical products is becoming more and more important in recent years, many different activities has been started by different (international) organizations on this field. SORS has been following these activities. The more intensive work on quality assessment at SORS has begun in year 2002. The basis for these activities were documents of Eurostat's working group on quality assessment, especially the methodological document *Standard Quality Report* and handbook *How to make Quality Report*. The document *Standard Quality Report* was also translated and adopted for our own reality with intention to become a basic document for the area of quality assessment'[6].

[6] Seljak, Rudi, Quality assessment at the Statistical Office of the Republic of Slovenia in years 2002-2004. Ljubljana, 2004 (unpublished document), p. 1.

The concept of the Standard Quality Report has been at SORS tested within several pilot quality reports. Because those reports were prepared to asses 'classical' surveys, SORS decided to proceed the testing in 2007 – this time on surveys which are using administrative and statistical sources. It is expected that those pilots will prepare the ground for the comprehensive quality assessment of the register-based 2011 Census. The main focus in the quality assessment of the census will be the relevance of the data for the end users and the coherence of data from various administrative and statistical sources.

The *Proposal for the Regulation of the European Parliament and of the Council on Population and Housing Censuses* in Article 4(4) defines that the EU 'Member States shall report to the Commission (Eurostat) on the data sources used, the reasons for the selection of these sources and the effects of the selected data sources on the quality of the statistics (quality report)'. It is not clear if the Standard Quality Report will be used also for the assessment of the census products or if Member States can expect that the Commision (Eurostat) will adopt a special census quality report in the near future. The SORS will use the experiences gained in the last five years and in the year 2007 and will adapt the standards which will be set at the EU level.

Chapter 5. Still a census?

The new *UNECE/EUROSTAT Recommendations for the 2010 Censuses of Population and Housing* and the *Proposal for the Regulation of the European Parliament and of the Council on Population and Housing Censuses* give countries the opportunity to use different data sources. Data can be compiled on the basis of:

a) a conventional census;

b) a register-based census;

c) combination of a conventional census and a sample survey;

d) combination of a register-based census and a sample survey;

e) combination of a register-based census and a conventional census;

f) an appropriate survey with rotating samples (rolling census).

The question which rises from so many various ways to compile census data is, if the census of population and housing is still a census. Namely, some of the essential features[7] (among them individual enumeration, simultaneity, universality, small-area data, defined periodicity) that 'distinguish population and housing censuses from other surveys'[8] are no longer met. In the combination of a conventional census and a sample survey or a register-based census and a sample survey it is difficult to meet the essential census feature called simultaneity *('information obtained on individuals and housing in a census should refer to well-defined and unique reference period'*) and the assurance of small-area data *('the census should produce data on the number of characteristics of the population and housing related to the smallest areas of the country, and to small population groups'*). In the case of combining a conventional or a register-based census with surveys to cover the core or non-core variables or even some of the census units (for example families), it is difficult to conduct the survey on the same reference day as the census, and furthermore the data coming from the survey (depending on the size of the sample) can assure the reliability of data only for higher territorial (geographical) levels as NUTS1, NUTS2 and NUTS3. In some countries where LAU2s are already very small (the case of Slovenia), it is almost impossible to have a representative sample that the data would be reliable for LAU2 territorial level.

7) Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva, 2006, p. 9-10.
8) Ibid.

And furthermore, the register-based census offers the opportunity to conduct the census not only each 5 or 10 years. The experiences of the countries which are already conducting register-based censuses show that the annual census no longer represents a census in the traditional sense. It became only one statistical survey among others.

Chapter 6. Conclusions

When a country, especially keeping in mind benefits of a register-based census (if we still talk about the census!), decides to have a register-based census of the population and housing, a commitment at the highest level and public approval in the country is not enough. Countries experienced that preparations for this kind of a census take 2 or even 3 decades.

And even if the relevant and stable administrative and statistical sources for the census topics exist, the statisticians are faced with many challenges to meet the international recommendations for the census topics, national demands for the data, and last but not least, to achieve statistical quality measures.

References:

1. Myrskylä, Pekka, Use of Registers and Administrative Data Sources for Statistical Purposes. Helsinki, Statistics Finland, 2004.
2. Proposal of a Regulation of the European Parliament and of the Council on Population and Housing Censuses. Brussels, COM(2007) 69 final. 23. 2. 2007.
3. Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva, 2006.
4. Seljak, Rudi, Quality assessment at the Statistical Office of the Republic of Slovenia in years 2002-2004, Ljubljana, 2004 (unpublished document).
5. Standard Quality Report, Luxembourg, 2003, Item 4.2b of the Working Group 'Assesment of Quality in Statistics' (unpublished document).
6. Tršinar, Irena, Centralni register prebivalstva: oris razvoja, komentar zakona in prevod v angleški jezik. Ljubljana, Uradni list RS, 1999.
7. Žnidaršič, Erika, Jug, Matjaž, Žnidaršič, Tina, Terms of Reference: Setting up the Social Statistics Database for the Implementation of EU SILC. Ljubljana, 2006 (unpublished document).