

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

STATISTICAL OFFICE OF THE EUROPEAN UNION

**Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses
(The Hague, The Netherlands, 10-11 May 2010)**

Working paper 11
3 May 2010

Topic (iv) of the provisional agenda

METHODOLOGY

**Quality assessment in a register-based census – administrative versus statistical concepts in
the case of households**

Note by the Statistical Office of the Republic of Slovenia*

I. INTRODUCTION

1. Collecting data on households in Slovenia was obligatory topic of all traditional censuses after World War II. On the other hand, data on families were for the first time for the whole population derived not earlier than at the 1981 Census. The definition of household has not gone through fundamental changes. The basis at 1948-2002 censuses was self-declaration that persons are living together and sharing their income for covering the basic costs of living.
2. Our experience is that collecting data on household and family structure was one of the most demanding operations in traditional censuses. The household and family concept is relatively demanding to comprehend, so it is very difficult to instruct well enough a large number of enumerators (around 10,000). Besides that, there are many multi-generational households with more members and complicated relations among them in Slovenia, which leads to greater complexity.
3. At the 1981 and 1991 censuses the families were derived in the earlier phase of processing on the basis of enumerator's records of relation between household members, but data on relations were not captured in a database. At the 2002 Census for the first time all input data on households and families were coded at the fieldwork by enumerators and directly transferred into electronic form. In the later stage of processing the most inconsistency was found out exactly on generation of family nucleus.

* Prepared by Danilo Dolenc for the Q2010 - European Conference on Quality in Official Statistics, Helsinki, Finland

II. FIRST REGISTER-BASED CENSUS IN 2011

4. The Statistical Office of the Republic of Slovenia (hereinafter SORS) has already at the 1991 Census and to a greater extent at the 2002 Census used the data from different administrative and statistical sources, which enabled more efficient and simple data collection in the field, reduced response burden and gave opportunity for shorter, better and more cost effective data processing (Dolenc, 2009). But data on households and families still had to be entirely collected in the field as an adequate source had not yet been available.

5. For carrying out the first complete register-based census in 2011 the following important developments in the field of providing necessary data sources have been reached:

- Establishment of the Real Estate Register (hereinafter RER) on the basis of field enumeration of buildings and dwellings, which finished in 2007. From the register-based census point of view one of the most important elements was determination of dwelling numbers in multi-dwelling buildings;
- Supplementing of addresses in the Central Population Register (hereinafter CPR) with dwelling numbers;
- Informatization of paper household files and establishment of the Household Register (hereinafter HR) kept by the Ministry of the Interior.

III. HOUSEHOLD REGISTER – THE ONLY SOURCE FOR HOUSEHOLD AND FAMILY DATA

6. The HR is a part of the Register of Permanent Population. Data on household members are collected at the time of registration or de-registration of permanent residence by the statement of being a household member. Relevant changes of household members which are not the result of change of address (for example death, divorce, formation of a new household at the same address) are updated by authorized person by official duty.

7. The household identifier in the HR is the serial number of the household running from 1 to NNN at the same address. The most important advantage of the HR compared to other register countries is the availability of data on relation to the reference person of the household. The same classification was used at our previous 2002 Census, so it is in fact statistical classification taking into account the most common relations in households. Besides relations spouse, cohabiting partner, child, spouse of the child, parents, parents-in-law that are suggested in Recommendations for the 2010 Censuses of Population and Housing (hereinafter Recommendations, 2006) in paragraph 515, some other relevant relations are included (grand parents, grand child, siblings, nephew). By using data on relations to the reference person, for most of the population generation of households and families could be done automatically without using complicated statistical methods. For this purpose for every single combination of relations in household a unique matrix is being prepared which enables us to derive all obligatory topics from EU Regulation on population and housing Censuses (size of household / family, household / family status, type of household / family) directly from the HR.

8. The main emphasis of the paper are methodological solutions for using HR data for statistical purposes in a register-based census and how to improve the quality of input data in the administrative source. We will also present the preliminary results of deriving data on households

and families from the HR, which was for the first time used in the statistical process at all in Slovenia, so the rawest data (and of the worst quality) were processed.

IV. STATISTICAL HOUSEHOLDS IN THE REGISTER-BASED CENSUS

A. New concept of the household

9. The concept of the household (family) in the register-based census results from the changed statistical definition of population, which is harmonized with the definition of population and migrants in the Regulation on Community Statistics on Migration and International Protection (the intended stay of at least one year is the key criterion), follows the Recommendations (2006) and is closely connected to the availability of relevant data in three basic administrative sources (CPR, HR and RER) which form the framework of the Register-based Census 2011:

- Registered residence (permanent or temporary) from the CPR;
- Living in the same dwelling (with the same dwelling number in the case of a multi-dwelling building) (data from the CPR and the RER);
- Declared household (persons with the same household number at the address) from the HR;
- Relation to the reference person of the household, which should be stated for every person with permanent residence in the HR.

10. The new definition of the household in the Register-based Census 2011 is therefore a group of persons (or a person living alone) living in the same dwelling with the same household number.

11. Recommendations (2006) distinguish two concepts of private households (paragraphs 479-481):

- The housekeeping concept, which assumes that the number of private households is not equal to the number of housing units (two or more households can share the same dwelling);
- The household-dwelling concept, which considers all persons living in a housing unit to be members of the same household, such that there is only one household per occupied housing unit.

12. The analysis of the HR data and experiences from previous censuses lead to the implementation of the housekeeping concept in the register-based census in Slovenia. Cohabitation of more households in the same dwelling (mostly in detached houses) is very common. At the 2002 Census about 3.5% of households shared the same dwelling (mostly cohabitation of older and younger generations with separate self-declared households). Besides that, the organization and methodology of the HR enable simple implementation of the housekeeping concept.

B. Quality aspects of household data in the administrative source

13. The main part of the paper is devoted to the analysis of some quality components which must be taken into account while preparing the methodological solutions for generation of households and families at the 2011 Register-based Census.

1. Maintaining of the Household Register

14. The HR in fact is not a classical register because there are no transactions on entries or exits but only momentary and non-repeated cross-section of data for any point in time is available. Besides that, it is not completely simultaneous with the CPR data (although it should be) but in some cases there are delays in updating of the HR. As the time harmonization is crucial for linking population and household data, we have already informed the Ministry of the Interior about the problem and asked them to pay special attention in the last month before the census date.

2. Use of administrative data

15. As mentioned, the HR is a new electronic source and so far it has not been used in the statistical process. But on the other hand also the previous paper files were widely used for administrative purposes in the form of approval of common household which was intended for asserting different social rights together with tax data on income (for example for scholarship, social support). Permanent use of register data is a prerequisite for better quality.

3. Legislation

16. Register data are as a rule administrative records which do not necessarily correspond to statistical concepts but depend on legislation. From this point of view there are some main problems associated with the statistical concept of household:

- Every resident can have only one registered household which is at the address of permanent residence only;
- For the registration of temporary residence the legislation (Residence Registration Act) does not require data on households (more than 60,000 persons according to the new population definition are residents at the temporary address but data for their households are available only on permanent address);
- For foreigners with temporary residence in Slovenia there are no data on households;
- Institutional households are not defined in the legislation;
- Relations between household members refer only to the reference person of the household.

17. One can say that this is conceptual 'non-quality', which could be solved by methodological solutions and statistical methods.

4. Subjective aspect

18. Formally correct data in the register do not necessary correspond to the real field situation of population and households. The quality from this point of view depends on how strictly people respect the legislation on registration. According to the law, the registration must be done in eight days after settlement. But for Slovenia we well know from the previous census and from other surveys that mostly in larger towns around 10% of people live elsewhere but not at the registered address. There are three main reasons:

- Some personnel benefit (mostly of finance nature);
- Owners of private dwellings do not register their tenants or subtenants in order to avoid paying taxes;
- Many rights (in administrative sense) are connected to the permanent residence.

5. Inconsistency and incompleteness of administrative data

19. Missing obligatory data – the main problem associated with the HR is the fact that for almost half of population living in multi-dwelling buildings the dwelling number is still not updated in the CPR and consequently in the HR. This makes it impossible to link a household with its dwelling. Although the data on the dwelling number has already been collected by the Surveying and Mapping Authority of the Republic of Slovenia (hereinafter SMA), we assume that the only solution is to invite people in multi-dwelling buildings to provide the data once again to the adequate administrative bodies.

20. Accidentally wrong data are mostly the consequence of subjective mistakes in all stages of the administrative process. Many of these mistakes are easy to solve with more exact guidelines for work and simple rules when entering data into electronic form (for example a one year old child can't be a father by relation to the reference person). It is very important that mistakes of this kind are corrected already in the original administrative files and not later in the statistical process. SORS has already prepared a very consistent document with detailed description of different types of mistakes together with the proposals of solutions.

21. System inconsistency of data in the administrative source is the most demanding problem to solve in the statistical process. Incorrectness could be the consequence of the change of the methodology in the middle of data collection, incomplete coverage, different understanding of guidelines (especially in case of a great number of collaborators) and so on. We found out significant differences in the quality of input data among administrative units (58 of them in Slovenia), which are not coincidental. As expected, the most problematic are data from larger cities.

22. Updating of the HR data should be done simultaneously with every single change to any household member that also effected the changes of household generation. In practice the HR is often updated subsequently because of overloading and non-priority, so inconsistency between CPR and HR data appears.

6. Administrative versus statistical concepts

23. The influence of legislation was discussed in chapter 4.2.3. Besides that, some formally correct administrative data in the HR are not directly useful because of statistical concepts of household. Direct use could be misleading or methodologically improper. Typical cases are households in monasteries and other religious institutions with two administrative modes of registration:

- Every person is registered as a single household;
- There is one private household with a larger number of members with relation to the reference person as non-relative.

24. But according to the Recommendations (2006, paragraph 484) an institutional household had to be generated here.

V. FAMILIES IN THE REGISTER-BASED CENSUS

A. The family concept

25. The family concept is in principle the same as it was in previous censuses and it is harmonized with Recommendations (2006, paragraph 493). The administrative data were supplemented with relations between biological parents and children and between married couples. By combining CPR and HR data, we are able to apply also other non-core concepts of family such as reconstituted families, skip generation families or derive also generations in the household.

26. Basic objectives for defining families are:

- Identified private household (no families in institutional households);
- Completed relations to the reference person of the household from the HR;
- More than one family in a private household;
- Direct derivation of all core family topics in case of unique matrix of relations;
- Use of statistical methods for generating families;
- Application of methodological solutions of other register countries (for example for determine consensual unions).

27. Family variables are not available in the HR directly but are derived from relations to the reference member of the household. If the matrix of relations is unique and not too complicated, the generation of families is very simple. In most multi-person households in Slovenia all household members belong to the family nucleus.

B. Quality aspects of family data in the administrative source

28. The quality of family data is directly connected to the quality of household data described in chapter IVB. There are some methodological limitations for identifying families in the household:

- Some relations to the reference person are not identifying enough (other relative, non-relative);
- Other relative often means the cohabiting partner of one of the household members (except the reference person);
- In case of more than one family in the household, it is not possible to distinguish which members belong to certain family;
- Complexity of relations in households with many members.

VI. SOME METHODOLOGICAL SOLUTIONS

29. The principle rule in the statistical process is maximum automation and standardization and minimization of manual editing.

A. Identifying institutional households and/or collective living quarters

30. Criteria for potential institutional household and/or collective living quarters:

- At least 20 person without administrative household at the address;
- At least 6 citizens without administrative household in the same dwelling if their permanent residences are different;
- At least 7 foreigners without administrative household in the same dwelling;

- Selected addresses of religious institutions and monasteries in Slovenia.

31. More than 900 addresses with 60,000 persons have been determined and precisely checked in different public databases. An inventory of types of living quarters has been elaborated and classified into 8 basic categories. For only one category (worker dormitories) private households are foreseen but the other seven (secondary and tertiary student homes, old people's homes, welfare institutions, correctional and penal institutions, religious institutions) are typical living quarters with institutional households.

B. Households without a reference person

32. The most common is statistical relocation of the reference person to the temporary residence according to the new population definition while other members of the household still form a household. A matrix has been prepared to transform old relations to the new ones. The basic rule is the same hierarchical level of the new reference person (if exists in the household – for example a spouse becomes the reference person).

C. Family formation in households

1. Direct family formation from relations to the reference person

33. For every single combination of relations to the reference person in the household a unique matrix is prepared defining the type of household and the type of family (families). So far more than 50 basic unique household matrixes have been developed with almost 300 sub-matrixes. Sub-matrixes are needed because of possible combinations of relations in the household. For the most common and simple type of household (one-family household husband-wife couple with children without other persons) three sub-matrixes were produced.

Table 1 One-family household of married couple and children

Sub-matrix 1		Sub-matrix 2		Sub-matrix 3	
Code	Relation	Code	Relation	Code	Relation
0	Reference person	0	Reference person	0	Reference person
1	Spouse	5	Parents (both of them)	5	Parents (both of them)
3	Child*			10	Sibling*

*Number of children in sub-matrix 1 and number of siblings in sub-matrix 3 are not limited.

34. In the case of two lone parent families in household (household matrix) 20 sub-matrixes are foreseen and only for three of them there was no such combination in the test database.

2. Indirect family formation from relations to the reference person

35. If there are more complex relations in the household, the relation to the reference person is not enough for family formation, but also biological or legal ties between household members are

taken into account (if they exist). PIN of parents is almost completely available for younger generations.

Table 2 Household with two lone parent families

Code	Relation	PIN	PIN of parent	Family
0	Reference person	1	-	1
3	Child	2	1	2
3	Child	3	1	1
9	Grandchild	4	2	2

36. Sub-matrixes in this household matrix consider also the sex of the reference person and a combination of four sub-matrixes allows automated family formation in the later stage of the statistical process.

Table 3 Sub-matrixes for the household with two lone parent families

Sub-matrix	Sex of lone parent		Type of family
	Family 1	Family 2	
1	1	1	Two lone fathers
2	1	2	Lone father and lone mother
3	2	1	Lone mother and lone father
4	2	2	Two lone mothers

3. Use of statistical methods for family formation

37. Statistical methods will be predominantly used for the consensual union formation as we are ascertaining that their number is under-estimated in the HR because of the administrative methodology. The criteria for statistical formation of consensual unions are:

- Common children (both partners are biological parents of the child);
- Opposite sex, minimum age (18 years) and maximum age difference (15 years);
- Relations other relative and non-relative in the administrative source.

38. We are following the practice of some Nordic countries (Denmark, Norway and Finland) with very similar statistical concepts.

4. Complex households

39. A complex household is one with a larger number of members who are relatives and usually form more than one family in the household. There are many possible combinations between members and because of the complexity of relations it is almost unfeasible to produce a unique matrix for automated formation of families. The solution is an interface for manual formation of

families and coding family data. SWOT analyses have proven that much less time is spent on manual coding than on defining and programming very demanding and unique combinations.

VII. QUALITY ANALYSES OF THE HOUSEHOLD REGISTER

40. For the first time data from the CPR and the HR were linked in the second half of 2009 and this test database has been studied in greater detail. First results are encouraging:

- The share of formally incorrect data is less than 1%. Most of the errors were found in combination of relation to the reference person and age. The most common mistakes are inverted relations to the reference person (child instead of mother and vice versa for example);
- The share of non-useful data on relation to the reference person for direct family formation is about 3% (unknown relation, other relative or non-relative);
- For 86% of population data on households and families were directly produced by using simple matrixes;
- For 8% of population only data on households are produced but families were not generated;
- For 6% of population neither data on households neither data on families could be provided because of missing data on household number and relation to the reference person. Two main population groups are included:
 - o Slovenian citizens with permanent and temporary residence in Slovenia (most of them statistically belong to institutional households);
 - o Foreigners with temporary residence in Slovenia (statistical private households will be generated);
- Number of complex households is less than one thousand. The maximum number of household members is 21 with 5 statistical families inside.

41. An important aspect of quality assessment is also comparability of register-based household data with data from surveys or previous censuses from the user's point of view. The concept of household itself is different. Besides that, the new definition of population has a significant influence on household types and structures. The communication campaign foreseen for the second half of 2010 will therefore concentrate on explaining the differences between methodologies and proper interpretation of household and family data.

42. Some first estimates of the number and structure of households and families are already available at a very aggregate level:

- Increasing number of households, above all one-person and two-person households. The share of one-person households will exceed 30% (compared to 22% at the 2002 Census);
- Much more institutional households with at least doubled share of population (less than 1% at the 2002 Census);
- We expect also structural changes in types of families (underestimation of consensual unions, more lone parent families, fewer children in families).

VIII. CONCLUSION

43. Generation of household and family data is the most demanding and complex statistical process in the 2011 Register-based Census in Slovenia. The main challenge is adaptation of input administrative data to the output statistical concepts. Quality assessments and implementing a quality assurance will depend a lot on the quality of input data, so recognized inconsistencies have

to be suppressed already in the administrative sources. From this point of view the very close cooperation with data providers is of great importance. SORS has already produced a very comprehensive document about non-quality findings in the HR (and the CPR, too) with detailed description of data errors and proposals of solutions. We put great emphasis upon integrated managing of all joined records at the same time for the whole household whenever a household member arranges any administrative proceeding directly at the administrative body. The data provider (Ministry of the Interior) is aware of its responsibility for the success of the register-based census and has already started to improve the quality of the data. Regular meetings and exchange of information are an important step to better quality.

44. Statistical generation of households and families from an administrative source has certain advantage in terms of quality compared to the field collected data. The whole statistical process could be unified and controlled by the same rules in all stages and the subjective influence on data is almost absolutely prevented.

45. SORS will carry out the 2011 Register-based Census with its own human resources without any additional employment as a project with approximately 30 experts involved. The main advantages will be reduced response burden (which is now an important political project in Slovenia also in other fields of public administration) and considerable savings of the state budget (over EUR 12 million). The data produced in the 2011 Register-based Census will be, however, of the same quality (if not even better) and useful for analysis, planning, scientific and research work as data from previous censuses.

REFERENCES

Apolonija Oblak Flander: Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case of Slovenia. Seminar on Registers in Statistics – methodology and quality. Helsinki, 2007

Censuses in Slovenia 1948-1991 and Census 2002. Statistical Office of the Republic of Slovenia. Ljubljana, 2001

Danilo Dolenc: Register based Census 2011 - a new challenge for the Slovenian national statistics. Statistical days. Radenci, 2009

Measurement of different emerging forms of households and families. Statistical Commission and Economic Commission for Europe Conference of European Statisticians. First meeting of the 2009/2010 Bureau. Washington, 2009

Quality assessment of the register-based Slovenian census 2011. Joint UNECE/Eurostat Meeting on Population and Housing Censuses. Geneva, 2008

Recommendations for the 2010 Censuses of Population and Housing. United Nations, New York and Geneva, 2006