

**Economic and Social Council**Distr.: General
27 June 2019

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Twenty-first Meeting**

Geneva, 18–20 September 2019

Item 4 of the provisional agenda

Research on the use of administrative data for censuses**Selected Elements of the Integrated Census in Slovakia with
an Impact on the Use of Administrative Data Sources****Note by the Statistical Office of the Slovak Republic****Summary*

The 2021 Census is based on a new design in Slovakia. It is a combination of data from multiple administrative data sources with field data. In addition to ensuring access to the sources that were identified as relevant for the purposes of the 2021 Census, the preparatory phase of the 2021 Census also included work with the sources, their evaluation in terms of quality, the analysis of errors, proposals for correction and the system of their integration. The paper presents the main findings and recommendations from a procedural point of view.

* Prepared by Ms. Ľudmila Ivančíková, Ms. Zuzana Podmanická.



I. Introduction

1. In Slovakia, so far, all population censuses have been conducted using the traditional method. The planned 2021 Population and Housing Census in Slovakia is based on a new concept, and its design includes several new features compared to the previous censuses. The concept itself is based on the combined census method. It combines the use of administrative data sources (hereinafter referred to as 'ADSs') and data on the population collected in the field (exhaustive survey).
2. The use of an exhaustive census has resulted from the national requirement to have detailed data at the regional level (municipalities – Local Administrative Units (LAU) level 2) as well as the requirement to obtain data at the grid level for Eurostat purposes in accordance with the relevant implementing Regulation. Similarly, in the area of dissemination, the requirement of comparability of outputs with the previous censuses and the production of hypercubes for the European Union (EU) and Datacube of the Statistical Office of the Slovak Republic (SR) played a role.
3. One of the key elements of design is statistical data integration. This involves integration of data from multiple ADSs with each other, and the integration of data from ADSs and data from the population. This integration is multi-level, as will be shown throughout this paper.

II. The legislative framework for the use of ADSs

4. The census model presented at the national level by the legislative intent in 2018, and subsequently reflected in the legislative form of the draft Act on the 2021 Population and Housing Census, is the first integrated census in the history of Slovakian censuses. This census relieves the population of the obligation to fill in complex paper forms by self-enumeration. The Statistical Office of the SR will collect directly from the population only data that cannot be provided from ADSs. The population will not provide data on houses and dwellings at all.
5. Although the use of administrative data is based on the international recommendations for censuses, the Statistical Office of the SR has not yet implemented such use of administrative data in statistical practice in the area of social statistics; in the area of business statistics, it has already implemented it but only to a limited extent. Therefore, the first step was to identify the existing relevant data sources, managers of registers and their content that could be used for the census purposes. Overall, the Statistical Office of the SR communicated with 17 managers of ADSs during the years 2016 – 2018 and 12 sources were initially identified.
6. Selected ADS managers were contacted with a request for cooperation in the form of bilateral agreements on the provision of data. In the first step, the data served to verify their suitability for the purpose of the census. Despite the fact that the conclusion of the agreements was time-consuming and resulted from a number of negotiations, the process was successful, particularly in the case of registers relevant to the collection of data on the population. With regard to then-forthcoming entry into force of the General Data Protection Regulation (GDPR), which was implemented in Slovakia by a special law on personal data protection, some ADS managers raised the issue of the legitimacy of the request. The reservations of the ADS managers highlighted the need for a clear definition of the responsibility for the census and of the relevance of the use of the data for testing and verification and reflected the absence of national legislation.¹
7. A solution was brought by the amendment of the Act No. 540/2001 Collection of laws on State Statistics, which reads, in Section 15, that “The census of the population, houses and dwellings is a special type of statistical survey carried out by the Statistical

¹ The Population and Housing Census is the only statistical survey in Slovakia conducted under a separate law and not under the Act on State Statistics No. 540/2001 Collection.

Office of the SR which is a part of the State statistics and whose content, scope and manner of implementation shall be regulated by specific legislation. For the purposes of the preparation and implementation of the census, the Statistical Office of the SR also uses data from administrative sources pursuant to Section 13". By this amendment, the Statistical Office of the SR secured its access to all relevant ADSs even before the census itself. From a methodological point of view, this permitted their detailed analysis, design and validation of the quality evaluation methods, work with metadata, the system of checks for data processing, and design and testing of the statistical integration process.

8. The draft Act on the 2021 Population and Housing Census states that at the written request of the Statistical Office of the SR, the public administration entity and the person who is the manager or operator of the public administration information system are obliged to provide the Statistical Office of the SR with data, information, metadata and population indicators, including the personal identification number used by the Statistical Office of the SR to integrate data from the public administration information systems and other ADSs that they manage or operate.

9. The Act also addresses the issue of two ADS managers who, by agreement, provided data for the validation within the census preparation, but only once, and they made the provision of further data conditional upon implementation of the Census Act. In addition, the Act defines the provision of ADSs within the census and post-census survey, but, crucially, it also defines the provision of data for the post-census.

10. The Act defines ADS managers in an enumerated manner. In total, there will be nine entities (The Ministry of Interior of the SR, The Ministry of Education of the SR, The Ministry of Environment of the SR, The Geodesy, Cartography and Cadastre Authority of the SR, The Social Insurance Agency, The Financial Directorate of the SR, The Corps of Prison and Judicial Guard, The Office of Labour, Social Affairs and Family, The Health Care Surveillance Authority) faced with the obligation that the relevant ADS manager will provide the Statistical Office of the SR with data from ADSs to the decisive moment of the census, by the established deadline. It also stipulates, for example, that the ADS manager shall provide the data by electronic means, free of charge and on the basis of a written request from the Statistical Office of the SR negotiated in advance with the ADS manager. The subject of the application and the technical details of the provision of data from ADS will be defined by agreement, and the requirements will be set by the Statistical Office of the SR in the form of a legislative decree.

III. The use of administrative data sources

11. For the purposes of the 2021 Census, the following administrative data sources, including registers, will be used:

- (a) Register of Natural Persons;
- (b) Register of Addresses;
- (c) Central Register of Public Health Insurance Policyholders;
- (d) Register of Accused and Convicted Persons, managed by the Corps of Prison and Judicial Guard;
- (e) Public administration information systems managed by the Ministry of Education of the SR;
- (f) Information System of the Social Insurance Agency;
- (g) Information Systems of the Geodesy, Cartography and Cadastre Authority of the SR;
- (h) Information System of the Financial Directorate of the SR;
- (i) Information System of the Central Office of Labour, Social Affairs and Family;
- (j) Information System of the Ministry of Environment of the SR.

12. When setting the individual processes in the framework of the Operational Plan for the 2021 Census Project, the Statistical Office of the SR made the correct assumption that the preparatory phase of the 2021 Census would take longer than in the traditional census and that it would also include, in addition to a number of studies, the analysis of ADSs.

13. Analyses of individual sources in the preparatory phase are based on a uniform approach. Given that some sources are analyzed repeatedly, the process of improving them is also verified. Nevertheless, it should be noted that despite the ADS managers' awareness and acceptance of the discovery of inconsistencies, in many cases the nature of registers and ADSs themselves (and the relevant legislation) does not allow the correction of data sets even for the purpose of the census.

14. The analyses cover technical characteristics of the provided data sets, the usability of the data content, the extent of completeness of selected items, the analysis of code lists and the possibility to convert them into binding statistical code lists, analyses of connections and logical links between variables, identification of errors, the possibilities for an optimal correction of emptiness or logical errors.

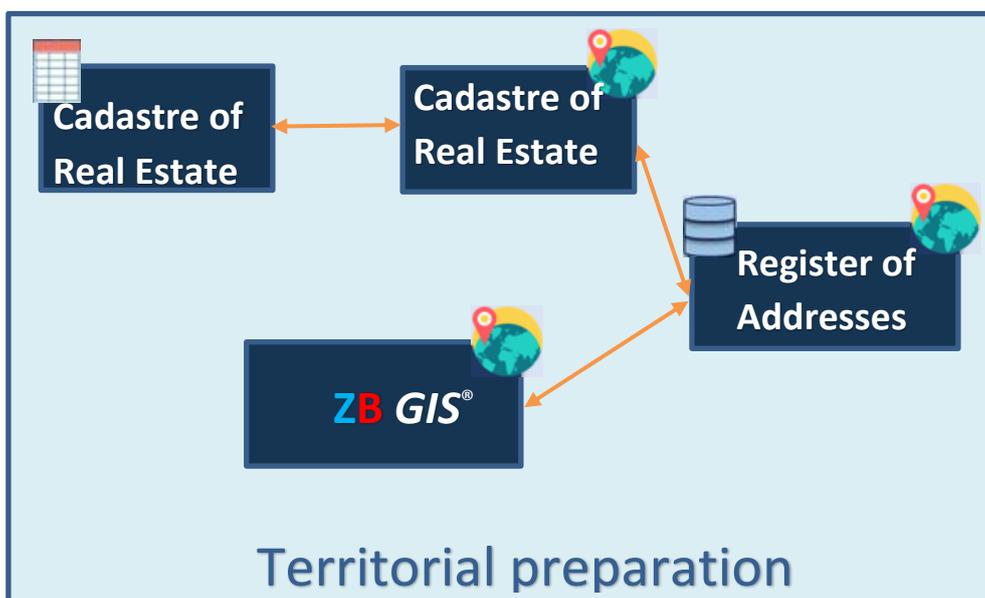
15. A prepared (ready-made) system of checks, conversions and logical edits will be verified, modified, and set up within the electronic information processing system.

16. ADSs will be used in the stage of the territorial preparation, in the census of houses and dwellings, in the electronic collection of data on the population and in the processing stage.

17. In order to ensure the territorial completeness of the census, the census process is based on the 'territorial preparation'. The territorial preparation for the purposes of the 2021 Census means the spatial and database identification of all objects in the statistical territorial unit, in which the SR population can live at the time of the census, regardless of whether or not they have a (any) registered residence there.

18. For the purpose of the territorial preparation, the following ADSs will be used: the Register of Addresses managed by the Ministry of the Interior of the SR, information systems of the Geodesy, Cartography and Cadastre Office (Cadastre of Real Estates and ZBGIS®) and the information system of the Ministry of Environment of the SR. The territorial preparation will result in a list of addresses that will be a maximum set of all possible inhabited and habitable objects (family houses, apartment houses, huts, garden cabins, schools, etc.) obtained from ADSs and a geographical layer of address points of these objects on base maps.

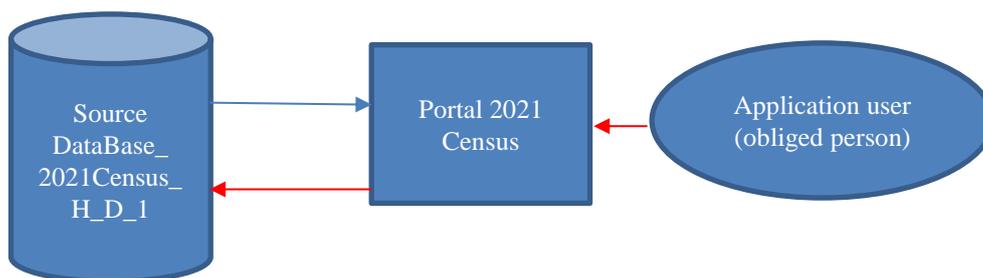
Figure I
Administrative data sources for the territorial preparation of the 2021 Census



19. The census of houses and dwellings will be based on the territorial preparation of the census. As already mentioned, there are efforts to reduce the reporting burden imposed on the population. Therefore for the purposes of the census of houses and dwellings, statistical data from ADSs available to the Statistical Office of the SR, (namely the Cadastre of Real Estate) will be used to pre-fill the form of each habitable object according to the territorial preparation and, subsequently, municipalities will supplement the missing data by means of registers (mainly municipal registers, registers of managers of apartment houses, and registers of management companies). This procedure removes the need for individual inhabitants of the SR to participate in the collection of data on the characteristics of dwellings and houses, and the census of houses and dwellings will be the exercise of state administration transferred to municipalities (LAU 2).

Figure II

Electronic collection of data on houses and dwellings in the 2021 Census



20. The person responsible for the census of houses and dwellings will be the person authorized by the municipality. The census of houses and dwellings will be realized only electronically, via the web portal of the Statistical Office of the SR.

21. The collection of data will begin on 1 June 2020, when the module named “COLLECTION_Census of Houses and Dwellings_2021” (“ZBER_SDB_2021”) containing electronic forms on houses and dwellings, pre-filled to the maximum extent possible, will be made available to all municipalities through the portal of the Statistical Office of the SR. Each municipality’s access will be secured by a password and name that will be made up of the statistical code of municipality according to the binding code list. The access name and password will be distributed to municipalities via the electronic mailbox according to the Act No. 305/2013 Collection §12. Persons authorized by the municipality will edit records on houses and dwellings on the basis of the recommended data sources. The data will be collected until 12 February 2021; the decisive moment of the census is midnight from Thursday 31 December 2020 to Friday 1 January 2021.

22. Within the processing of collected data, the database will be filled with calculated variable (e.g. the number of persons in dwelling) and derived variables (e.g. ownership of house by ownership of dwelling). This statistically processed set of data on houses and dwellings, including calculated and derived variables, will be provided to each municipality for its cadastral territory after the end of the entire census of the population, houses and dwellings. This means that the municipal data sets will be provided to each municipality; they will be provided in different formats to meet the specificities of data imports to the information systems of individual municipalities. At the same time, this data base will allow municipalities to continuously update the data and, thus, to ensure the sustainability of the solution and its use for the local needs of municipalities, as well as national requirements for the 2021 post-census.

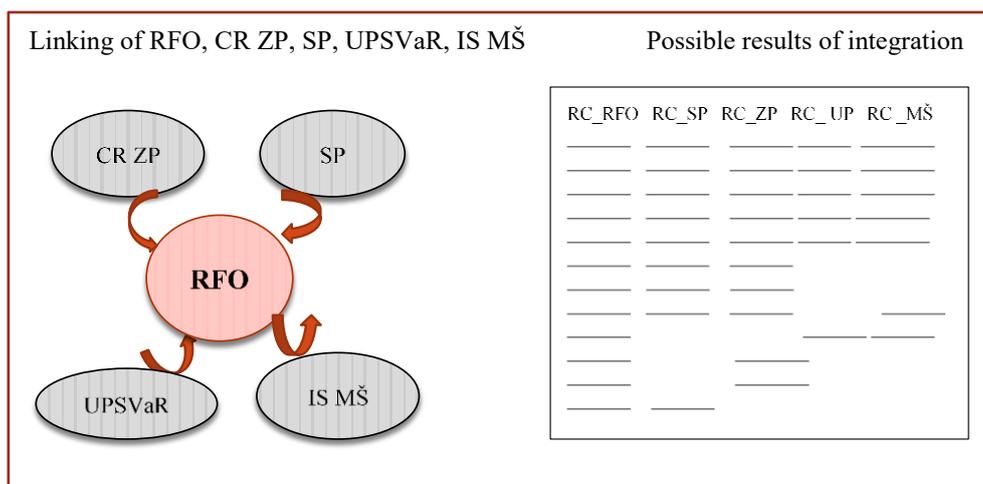
23. For the purposes of the population census and the collection of data on the population, data will be drawn from the Register of Natural Persons (RFO), Register of Addresses, Central Register of Public Health Insurance Policyholders (CR ZP), Register of Accused and Convicted Persons, public administration information systems managed by the Ministry of Education of the SR (IS MŠ), the information system of the Social Insurance Agency (SP), the information system of the Financial Directorate of the SR and the information system of the Central Office of Labour, Social Affairs and Family (UPSVaR). A minimized set of data that are not available in ADSs will be collected directly from the population and exclusively by electronic means.

24. The integration of data from ADSs as part of the preparation and data processing has three phases:

- (a) The first phase – the transfer and storage of data batches;
- (b) The second phase – checks and the initial processing of data in a particular source, e.g., in the case of the Register of Natural Persons (RFO) whose batch consists of several tables, it will be their linking through the personal identification number to the resulting database of RFO data. At the same time, potential sources of errors identified in initial analyses will be removed;
- (c) The third phase – the integration of individual ADSs with each other using a unique identifier and the integration code list.

Figure III

An example of the integration of ADSs on the population



25. The collection of data from the population will be fully electronic. The inhabitant is enumerated by means of the electronic form (e-form). The form exists in both online and off-line formats. The off-line form will be used only in exceptional circumstances. With regard to the nature of the data collected, self-enumeration is preferred when filling in the e-form. For a minor and for an inhabitant who is unable to provide the data himself/herself, the e-form will be filled in by his/her legal representative, a close person or a person living with the inhabitant in the same household.

26. The collection of data from the population will take place from 15 February 2021 to 31 March 2021. In the whole territory of the SR, the decisive moment of the census (the date to which the data refer) is midnight from Thursday 31 December 2020 to Friday 1 January 2021. The reference date was chosen with respect to the use of registers and other ADSs.

27. The data collection will be carried out in such a way that inhabitants can self-enumerate throughout the census period (from 15 February 2021 to 31 March 2021). Also, throughout the whole period, they can use contact points where they can self-enumerate as well, or they can be enumerated by an assistant working at a contact point.

28. The integration of ADS data and data from the population is based on the grouping of variables into four groups. The following four types of variables are identified for this purpose:

- (a) Variables of the first type – variables that enter the electronic collection: the electronic collection is filled with them, and there are, or there can be, checks (the personal identification number, first and last name, date of birth, gender, address) running in the background;
- (b) Variables of the second type – variables that are filled from ADSs but which will not be used for the electronic collection (the number of children, some types of economic activity);

(c) Variables of the third type – these are filled exclusively from ADSs (age, gender, economic activity, status in employment);

(d) Variables of the fourth type – these are collected exclusively in the field (religion, nationality, education).

29. The set of first-type variables created from all the ADSs used will be the starting point for the electronic collection.

30. The integration of ADSs with data from the electronic collection will take place in two phases. In the first phase, only first-type variables will be used: these will fulfil the authentication function in the electronic collection. In the second phase, after the end of the electronic collection, the data from the electronic collection will be linked to the processed data from ADSs.

31. Integrations to provide additional data (e. g. data on households or data on the population in houses and dwellings) will be statistically processed after the end of the electronic collection of data on the population.

32. Data obtained from ADSs which are evaluated as being of a high quality and relevance after the census can be the starting point for the 2021 post-census system.

IV. Conclusion

33. As the experience of countries with combined census and register-based census confirms, the implementation of censuses of this type is based on the assumptions of supporting legislation, public approval, co-operation with ADSs managers and other bodies and organizations involved in the census process, the existence of an appropriate system of registers (which did not exist in Slovakia and it has formed only after 2011) and the existence of a unique personal identification number². By meeting these assumptions, the possibility of the transition from the traditional census to the combined “integrated” census has been created in Slovakia.

34. The need to analyze new sources, to assess their quality and methods for their integration, as well as the use of these integration methods and tools, bring with them a change in working methods. At the same time, they create the possibility to set up processes to meet the requirements (though still formulated rather ideologically) for the post-census (and the associated more frequent production of part of the census data) and for a possible transition to a totally register-based census.

35. This concept is based on “wins and losses”. One loses the ability to collect data that are not in the registers, are not of sufficient quality, or are not methodically consistent with the census. Many countries abandon certain variables for which the census was a unique source. Registers are created for a non-statistical purpose. The nature of “registration” does not always capture the nuances of social changes, such as the composition of atypical households, as was the case with the traditional census. This aspect will necessitate more in-depth analyses and, perhaps, a search for alternative sources, since the strengthening of sample surveys seems to be difficult. At the same time, the decreasing participation of the population in the census calls for the use of the sources that are available, or can be made available, for statistical producers. Their use enriches statistics in terms of both content and development of new methods. At the same time, it creates some dependence of statistical offices on other entities, namely ADS managers.

² Schulte Nordholt, Eric, (2018) The usability of administrative data for register-based censuses, *Statistical Journal of the IAOS* 34:4, pp.487–498.