

**Economic and Social Council**Distr.: General
1 May 2012

English only

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**The 2011 Census in the Czech Republic as a new experience****Note by the Czech Statistical Office***Summary*

The paper describes the Population and Housing Census operations in the Czech Republic for acquiring fundamental information on people, houses, dwellings and households. The paper explains the latest innovations and practices in different census operations, such as in organizing the field work, renewal of IT systems, monitoring of form life cycle and in data scanning and validation. Finally the paper presents the preliminary results of the census, carried out according to the national Act No. 296/2009 Sb. of 22 July 2009, on the 2011 Population and Housing Census and the EC Regulation No. 763/2008.

I. Preparation of data for field work of the Census

1. The census is a typical task of mass data processing– which meant punched cards and tapes in the 1970s and 1980s and deployment of scanning and other IT technology at the brink of this century, followed by the option of electronic census forms. The solution for electronic distribution and collection of census forms was one of the basic attributes of the 2011 Population and Housing Census in the Czech Republic. It was offered as an alternative to completing and returning paper forms.
2. The decision of the Czech Statistical Office (CZSO) that the distribution and collection of census forms will not be provided directly by the CZSO, but by an institution delegated to do so, was essential for the census round. Česká pošta, s.p. (Czech Post) was selected as the supplier of field work (SFW) in an open tender.
3. This meant that the SFW and the CZSO would have to cooperate closely via an interconnection of two different information systems. The CZSO had to prepare numerous pieces of background information, which had to be consolidated, electronically processed, and prepared in appropriate technical form to be submitted to the SFW. This information was also used by other ministries which participated in the Census and by all municipalities providing information to the population, and by some other entities, such as Státní tiskárna cenin, s.p. (State Printing Works of Securities).

A. Preparation of data from administrative sources

4. Concerning data processing, the classic relational database solution was chosen which was based on potentials that the CZSO had at disposal. Therefore, two main databases were developed – one dealing with data preparation prior the Census (census preparations) and the second dealing with processing of the collected data (data processing).
5. The database for data preparations prior the Census was named the Registration Database (RDB). It was designed to keep the central register of all entities and objects, subject to the census, up-to-date. Thus, the database primarily contained information on obliged persons¹, houses and dwellings of the census, census officers, paper and electronic forms created and information of the form life cycle, i.e. from the phase of their creation till their processing.

B. Territorial preparations for the Census

6. So called working meetings of CZSO field workers were held during the territorial preparations. The purpose was to identify changes to actual status compared to the status given in descriptions and maps of the Census districts, which were created on the basis of CZSO data. The changes found were recorded and transmitted back into the RDB environment so that conditions at the distribution and collection would meet the actual status. Preparations were needed to provide for the exchange of background data between the CZSO information system and the SFW information system. Then data on respective

¹ Persons that had the legal duty to participate in the Census.

census districts were generated in the SFW information system². Tasks of the territorial preparations were solved as batch tasks in the Registration Database.

C. Background documents for paper-based census forms

7. One of crucial tasks of the Census preparatory phase was the printing of approximately twenty million census forms of all types. The basic element of every census form was a unique identifier and a check code, which were used for the purposes of tracing the form life cycle and acquisition of the census form electronically. The reason for having unique census forms was the strict requirement for personal data protection throughout the Census.

8. Data of the forms was stored in the RDB in relation to data on the Census entities. Data to be handed over to the printing house (responsibility of the SFW) were, in the form of so-called data sentences, transmitted through an interface directly into the ISTEP system of Česká pošta s.p. This system then provided the data delivery and transformation for printing systems.

D. Background documents for printing of identification cards of census officers

9. In order to be able to prove the identity of every census officer, they were equipped with a service identification (ID) card (also called census ID card). Although the ID card was valid only if produced with the citizen ID of the census officer, it was a document of a sensitive nature³. Státní tiskárna cenin, s.p. was selected for the printing of the census officers' ID cards regardless if the census officer was a member of the SFW or CZSO.

II. Supporting systems

10. A highly comprehensive solution was developed for the Census information support. Teams provided for inevitable services by counting several hundreds of people from operators to specialists. Management of such a large team of people required organisational and project management work as well as information support to the management.

A. Internal and external information portal

11. The internal and external portals formed a doublet of portal solutions dedicated for the deployment or access to applications of the 2011 Population and Housing Census to the users at the CZSO (internal portal) or to third parties, namely to the SFW⁴ (external portal).

12. Both solutions were based on the technology of Oracle WebLogic Portal 11g and each of them was solved for a dedicated hardware infrastructure in order to separate security domains. Implemented applications were either deployed directly on the server, or, for instance for security reasons, were operated on other server and the portal was providing

² Česká pošta s.p. operated the information system called ISTEP.

³ Its template was depicted in the executive decree to the Act on Census.

⁴ The Supplier of Field Works (SFW) – a partner of the CZSO selected in a tender for the distribution and collection of paper-based forms. Česká pošta, s.p. was selected as the SFW for the 2011 Population and Housing Census.

them in the form of a “reverse proxy server”. In such mode the portal responds to a user request the way it hands the request over to the target server, takes the response from the server, and returns the response to the user.

B. Knowledge base

13. The knowledge base was a system with a primary objective to consolidate available pieces of information on the Census for the purpose of providing them to the public, and also internally within the CZSO. The knowledge base played a fundamental role in the distribution and collection of paper forms as well as electronic forms when it was used by workers of the Census Contact Centre.

14. The Census Contact Centre operation was provided by the SFW, including communication services (phone operators, e-mail, fax). The base of information provided was, however, the responsibility of the CZSO.

C. Administration of requirements

15. In the course of the 2011 Population and Housing Census it was necessary, namely in the phase of form distribution and data collection, to ensure communication among persons providing the Census. Although the standard service of electronic mail was useful, it was not sufficient, especially in situations as follows:

- (a) When communication of an issue addressed had to be monitored; and
- (b) When communication had to be provable.

16. For these purposes a system for administering requests was developed to register requests, monitor status and manage notifications. The system was used both for the communication within the CZSO and third parties contributing to the census tasks, such as the application of “help desk” incidents and other issues. The system operated on the hardware infrastructure of the external portal and was based on an open-source product of OTRS (Open Technology Real Services).

III. The census forms

17. One of the crucial characteristics of the 2011 Population and Housing Census, as in any census, was the inevitable work with personal data and confidential data in general. These include data about persons, such as birth number, and data on relations to other persons.

18. For this reason, maximum efforts had to be done so that the protected data could not be misused or abused, and so that any potential misuse and/or abuse could be identified. The set of measures to protect data was broken down into several areas, which had to be supported by IT. These areas were as follows:

(a) Security and safety – includes security and safety of buildings, persons, information systems and information in general. Such security and safety were provided by a team of security and safety specialists and resulted in a defined security and safety policy for the Census and its implementation;

(b) Statistical data protection – includes the protection of data on respondents so that in the moment when the work does not deal with direct personal data only derived data is processed. Even in such a case, there is still a possibility that data concerning individuals, or concrete groups of the population, may be identified from the aggregated data. For this

type of data protection, the CZSO applies numerous methods, defined under terms of primary and secondary data confidentiality;

(c) The form life cycle tracing – includes methods and procedures, which are necessary to monitor handling of forms and the primary data carriers. Although this was activity carried out as a result of a defined security policy, it turned out to be essential in the course of the Census taking as the following text will explain.

A. Registration database

19. Concerning logistics, the data were stored and processed in the so-called Registration Database (RDB). The Database content was subdivided into three areas as follows:

(a) Data integration – data and structures for data exchange, namely with the Supplier of Field Works⁵ and other systems supporting the 2011 Population and Housing Census;

(b) Input-output interface for administrative data sources – data and structures for data mining from administrative data sources and CZSO systems;

(c) Registration data – this includes the registration data of the Census (persons, houses, dwellings, forms, territorial breakdown, statistical breakdown, and characteristics of the given entities).

B. The form life cycle

20. The exchange of data on the status of paper and electronic forms was crucial for the census operations. For the purposes of the paper form life cycle tracing, the data exchange with the SFW was inevitable (through the system of ISTEP – Information System of Field Works). The SFW captured the primary data from the paper forms. From the technology point of view, the data exchange was performed at the level of the RDB, which implemented a mechanism of a batch data exchange for data integration in a similar environment to that on the ISTEP side. Thus, the communication used an encrypted connection, assembled as a tunnel in the Internet.

21. Once the data was exchanged, the logic of permitted states and their transitions was checked, on the basis of predefined processes of their distribution and collection. Information on the forms, which were not in assumed states and transitions, were “claimed” back against the ISTEP in order to identify the problem. Due to the number of printed forms, data transmissions and the logistics processing were time-consuming tasks, processed every day from midnight till morning hours.

22. Every paper form was bound to an electronic one, which the citizen could complete as an alternative. Thus, it was necessary to trace the life cycle of electronic forms as well. This task was, however, simpler in many aspects, namely due to a simpler model of states and transitions of the electronic forms and as there was no communication with a third party.

23. All paper forms, which finally entered the scanning line to be digitised, were in the first scanning phase identified by means of the form bar code and this information was

⁵ The Supplier of Field Works provided for the distribution and collection of paper-based forms. It was Česká pošta s.p., which came from an open tender.

transmitted into the system. If we had to wait for the delivery of the forms in envelopes, they would have arrived for scanning with a substantial time delay. The solution was that the forms were pulled from envelopes and put into archive boxes, before they were delivered to scanning, bar codes of the “stored” forms were read by readers. This information was then transmitted and downloaded into the ISTEP and further transmitted into the CZSO information system. This way the CZSO acquired information on returned forms, before they passed through the digitising line and could comply with the legally binding periods for investigation and producing notices on citizens not meeting their legal duty to respond.

C. Notices of the census officers

24. A register of persons who can (except for the respondents) get in contact with the forms was maintained. Majority of these persons were census officers providing the distribution and collection of paper forms. The census officers were employees of the SFW (ca. 11 thousand), CZSO (ca. 2 thousand) and a small number of other ministries as the Ministry of the Interior of the CR, Ministry of Defence of the CR, Ministry of Foreign Affairs of the CR, and Ministry of Justice of the CR. They provided for the census in areas under their responsibility according to law.

25. One of the legal conditions, which the law imposed on the CZSO, cities and municipalities, was the requirement to disclose notices with lists of names of the census officers and numbers of their ID cards to the public. The aim was to provide citizens with the possibility to verify the census officer’s identity.

IV. Scanning and validation of data

26. The process of scanning produced electronic images of paper forms as background documents for next phases of work. These scanning operations included:

- (a) Positioning of the form;
- (b) Suppressing of coloured background;
- (c) Reading of the form bar code.

27. The application tracing the form life cycle made use of the latter activity. This way information on that the form has passed through the scanning line was obtained earlier. In other case it would be necessary to wait till the validation of forms has been completed. The resulting images of forms were created in resolution of 300 dpi. The TIFF format enables to store multiple image pages in one file. Thus, respective (multiple-page) forms were stored in one file. Because of the significant file sizes and “price” of these digital images, images in the TIFF format were continuously stored as files on archival media.

28. In scanning, the forms had to be handled so that the form life cycle tracing was ensured and security incidents of lost forms were prevented. The forms were taken from the central warehouse of forms of the CZSO in transport boxes of Česká pošta s.p. (forms collected by the census officers), or in archive boxes (forms sent to the P.O. Box of the Census).

A. Recognising, verifying and validating

29. The recognising of information recorded in the forms was added with the verification of data obtained against registers, especially in the cases of addresses and

names. Verification of data was primarily carried out on the basis of data search in registers according to data similarity, using the application Netrics Matching Engine⁶. In the course of image processing, the verification was modified into alternative algorithms with more appropriated type of data coincidence for the verification purposes.

30. The validation of data obtained in the recognition process was the most massive manual activity carried out within the census results processing. As a standard, it was performed in a two-shift operation, and at peak times also round the clock. Sixty validators worked on each shift. Each validator had a dedicated PC with installed eFlow application in order to be able to carry out two types of activity as follows:

(a) Validation of individual characters – multiple-element matrix with identical characters identified by the automatic character recognition in numerous forms was displayed to the validator. In such a set of characters, however, the human brain shows the perception, in which by simple sight it can identify incorrect character– for example “8” instead “B”. Then the validator corrected such errors;

(b) Validation of word context – in the case the automatic character recognition was not able to identify characters of a word, the word was presented to the validators with appropriate context of the form cell, . The aim was to correct unrecognised characters – if the validators clearly identified the word “Prada” and its context clearly demonstrated the word should be “Praha”, they left the former. The data correction in terms of meaning was the subject of the first phase of processing – encoding.

31. The validated data were in the form of data sentences which corresponded to pieces of information recorded in respective forms. These data were transmitted by means of a defined interface to further processing of already statistical nature. Every data sentence was transmitted both in the status of data prior to their validation and after the validation.

V. Preliminary results of the 2011 Census in the Czech Republic

32. As at the decisive moment, the population of the Czech Republic was 10 562 214, 332 thousand people more than in the 2001 Census.

33. The population structure of the Czech Republic changed in favour of non-citizens, whose number attained almost 450 thousand, compared to the 2001 Census. Over the decade the number of non-citizens, that means persons having citizenship other than that of the Czech Republic, increased more than 3.5 times. Almost the whole increase in the population compared to the 2001 Census was due to the increase in the number of non-citizens. At the same time, compared to 2001, the concentration of non-citizens living in the Czech Republic also increased, thus over one half of the total number of non-citizens in the country is settled especially in the Capital City of Prague, where two fifths of them live, and in the Středočeský Region. In the Capital City non-citizens represent already even 14% of its population.

A. Population ageing

34. The comparison of the population age structure found in the 2011 and 2001 Censuses has proven the trend of gradual population ageing. The share of persons in the productive age (15–64 years) has not changed notably since the last census and represents almost 70% of the population. The number of persons in the post-productive age (65+

⁶ http://www.netrics.com/netrics_matching_platform.php

years), however, in 2011 already outnumbered the child component of the population (0–14 years). In the Czech Republic, on 26 March 2011, there were 1 674 thousand persons aged 65+ years, which represents almost 16% of the population. The number of children under 14 years of age was 1 528 thousand, i.e. 14.5%. While the number of the population of the productive age has grown in the decade by almost 200 thousand and the number of people older than 65 years by 260 thousand. The share as well as the absolute number of children under 14 years of age has been declining in monotonous way since the 1980 Census. In between 2001 and 2011 this decline was not as prominent as in previous decades, nevertheless, the trend has been continuing.

35. The population structure by marital status is different in males and females. While in 2001 persons living in a marriage formed the highest shares of population (48% and 45%, respectively), in 2011 a similar distribution remained in females only where the largest share went to married females (41%) followed by singles (35%). In males the largest group was that of singles (45%), the share of married males was somewhat lower (43%). Both in males and females the numbers and shares of married persons in the population decreased over the decade. On the contrary, the share of singles and divorced increased. This development is no surprise in light of the numbers of marriages and divorces in the last twenty years (at least). The number of widowed females in the population is five times higher than the number of widowed males because of their higher life expectancy.

B. Educational attainment has been growing

36. The trend of the gradual increase in the educational attainment of the population continued. In 2011, the group of the highest number (almost 3 million) remained that of persons with the secondary education without A-level examination (including voluntary training), although their number as well as share decreased over the decade. The share of persons with higher educational attainment has been growing over a long term. In 2011, the number of persons with secondary education with A-level examination or with higher professional education increased by 15% to almost 2.8 million compared to 2001. This represents almost one third of the population aged 15+ years. The increase in the number of persons with higher education rose in even more significant way, by more than 40%, totalling more than 1.1 million, and thus represents over 12% of the population.

C. Composition by nationality (ethnic structure)

37. Compared to the previous 2001 Census the share of persons who assigned themselves to the major nationality, the Czech, significantly dropped. In 2001, approximately 9.25 million of the population of the Czech Republic declared themselves to be Czechs, which was over 90%. In 2011, the Czech nationality was indicated in the forms by 6.73 million people that were almost two thirds of the population. This decrement was, however, solely in part at benefits of other nationalities. The main reason for this decrement is the fact that a significant portion of the population used the legally offered option not to answer the questions on nationality and religious attachment. Over 2.74 million persons, almost 26% of the population, did not state their nationality at all. In the 2001 Census the question on nationality was not answered by mere 1.7% people. The Czech nationality represented full 92% in 2001 and only 86% in 2011 of all stated answers. The decline is clear.

38. Traditional minorities by nationality as the Slovaks, Germans, and Poles also recorded decrements in their numbers the same way as the Czechs. In 2011, solely 5.2 thousand people declared themselves to be of the Roma nationality, 7.1 thousand others stated two nationalities, Roma and Czech. In total, there were over 13 thousand people who

stated the Roma nationality (including combined with other nationalities). Compared to 2001, higher number of persons showed certain nationalities as Ukraine, Vietnamese and Russian, for instance.

D. Ways of housing

39. On 26 March 2011, there were over 2.1 million of houses, i.e. buildings dedicated to dwelling (for instance, as family houses, multi-dwelling buildings, pensions for the elderly, etc.), or serving primarily other purposes yet there was one dwelling in them, as minimum (for instance, a school building with a dwelling), or do not belong to any of the previous groups, but there are persons notified as permanent residents in them (such as municipal authorities with the residence notification office, for example).

40. There were stated periods of construction or reconstruction within the years 2001 and 2011 for over 200 thousand inhabited houses. The new construction of houses was unambiguously concentrated into the group of family houses, mostly with one dwelling. In the view of territorial distribution, there was the above average share of inhabited houses, which were constructed or reconstructed after 2001, first of all, in the Středočeský Region, where it counted for more than 16% of the total number of inhabited houses. In other regions the shares oscillated mostly around 11% or 10%. The lowest share of the new construction was in the Ústecký Region.

E. Dwelling conditions

41. There were 3 894 permanently inhabited dwellings included in the Census at the decisive moment. The dwelling stock structure and changes to it virtually copied trends and changes of ownership forms of houses. The dominance of dwellings in the own proprietary house corresponded to the dominance of family houses in the housing stock. Following a temporary decline in 1991 the number of dwellings in the own proprietary house has been increasing and in 2011 it surpassed the limit of 1.4 million (last time this happened in 1980).
