

**Economic and Social Council**Distr.: General  
30 April 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****Lessons learnt from a mixed-mode census for the future of social statistics****Note by the Federal Statistical Office of Germany***Summary*

This paper presents the experience acquired so far in Germany with the 2011 Census, which is based on a new data collection method. The new census method is a combination of administrative register evaluation and field surveys. The biggest challenge of the new census model is that combining data from different data sources at the level of persons requires in part using plain text information, among other things, name and address, because neither a uniform person identification number nor a uniform building identification number is available. The experience acquired in combining data without numerical identifiers in the 2011 Census can be used to make concrete proposals for the further development of the new census model. The experience may also be to the benefit of other countries which have not yet used data from administrative registers for a census due to similarly difficult conditions regarding the combined use of administrative data and field surveys.

## **I. Legal and political requirements taken into account when developing the new German census model**

1. The last population and housing census in Germany before the 2011 Census was conducted in 1987, that is, as a complete German-wide survey with all inhabitants being questioned by interviewers. That census was met with massive opposition in relevant parts of the population, including a successful constitutional complaint which led to the prohibition to transmit the data back to the administration. According to that prohibition, which is still in force in Germany today, data that were provided to the statistical offices of the Federation and the Länder for the production of statistics must not be transmitted back to the administration. Also, in its judgment made at the time, the Federal Constitutional Court obliged the legislator to consider a specific question when taking decisions on the data collection methods of future population censuses. The question is whether, as a result of further developments of methods of official statistics and social research as well as of data processing, it is still necessary to conduct a traditional complete count or whether the information required for the census could be obtained through a different collection method which would place a smaller response burden on the population.

2. Therefore, after the 1987 population census, the legislator commissioned the statistical offices to develop a new census method. The following general requirements had to be taken into account: Where possible, existing administrative data should be used for the census, and the cost had to be markedly lower than that of a traditional population census, with the quality of the major census results, in particular the number of inhabitants, at least as good as in former population censuses.

3. As a result, a census model has been developed which combines register evaluation with field surveys. For the purpose, the register of addresses and buildings was set up as a first step. It contains all addresses in Germany at which residential space is available. It is used, first, as the statistical population underlying the census of buildings and housing and, second, as a sampling frame from which the sample for the household survey was drawn. In addition, the register of addresses and buildings is used as a central organisation basis for all survey components of the 2011 Census. The data of the 2011 Census are obtained from the following components (cf. chart 1):

(a) Register data of the residents' registration offices are the core data used to ascertain both the number of inhabitants and the data on the demographic structure of the population living in private households. They are combined into an overall German data set and are checked for any incorrect duplicate cases<sup>1</sup>;

(b) For residents of collective living quarters, institutions, residential establishments and similar facilities, population register data involve too many errors, so that data on these people are ascertained through a complete enumeration;

(c) At the level of individuals, register data of the Federal Employment Agency and register data on public service personnel complement the demographic variables obtained from the population registers and from the survey conducted at special facilities;

(d) A sample survey conducted among less than 10 per cent of the population assures the quality of the population register data by ascertaining, for every municipality,

---

<sup>1</sup> Incorrect duplicate cases may refer to persons registered with their sole residence in several municipalities.

the rates of outdated entries and missing entries and taking them into account when determining the number of inhabitants of the municipalities.<sup>2</sup> The survey also collects additional data on the population that cannot be obtained from registers<sup>3</sup>;

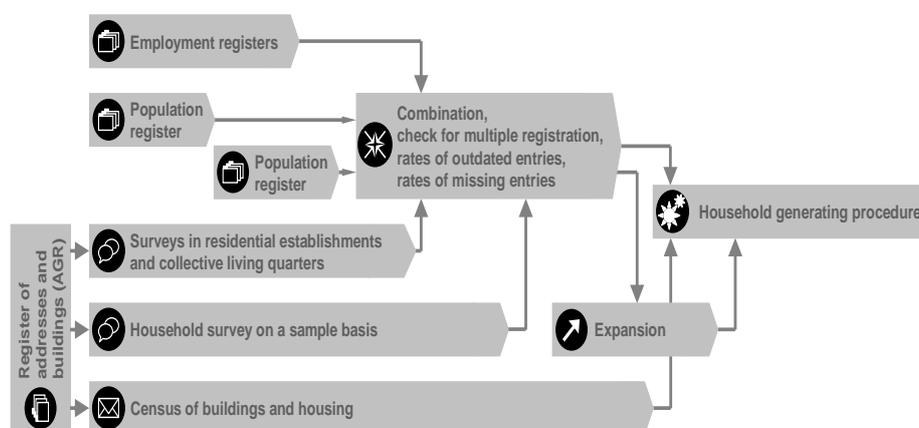
(e) Data on the number and structure of residential buildings and dwellings are collected by conducting a postal survey among all owners;

(f) Finally, the household generating procedure allows combining personal data into residential households by linking the population register data, which include information on family relationships, to dwellings data from the census of buildings and housing.

4. This census method ensures complete coverage of the target population<sup>4</sup>.

Chart 1

### The census model



## II. Acceptance of the new census method

### A. High acceptance in the population

5. Compared with the trauma caused among politicians and statisticians by the open rejection of the last population census of 1987 in western Germany by major societal groups, this year's census passed off virtually without any opposition. This is due not only to the changed legal and societal conditions (data protection provisions were considerably enhanced after the 1987 population census, the cases of misuse of personal data have so far not involved official statistics and many people today are rather generous in dealing with

<sup>2</sup> That method of assuring the quality of population register data is applied for municipalities with 10,000 inhabitants and over. In municipalities with fewer than 10,000 inhabitants, quality assurance is achieved by conducting a survey to clarify discrepancies and a survey to clarify the residence.

<sup>3</sup> The sample design and the expansion model for the household survey were developed in a specific project by Trier university and GESIS – Leibniz Institute for the Social Sciences in Mannheim.

<sup>4</sup> Covered are inhabitants at their main or second residence; however, the official number of inhabitants is calculated only on the basis of the population at the main or sole residence.

their own data) but also to the new census method. As data of the population registers are used, only a tenth of the population were directly concerned by the household survey.

6. The postal census of buildings and housing was limited to a population group with a lower potential for opposition (owners of residential space). The number of queries made by real property owners who were concerned because they have not received a questionnaire in time was much larger than the number of those who have not complied with their obligation to respond.

## **B. Considerable mistrust among local authorities**

7. There is an obvious lack of acceptance among regional and local authorities regarding the census results, which are not even available yet. One of the reasons is the mere fact that understanding the basic principles of the new census method requires knowledge of sampling theory, whereas people can understand the basic principles of a traditional population census or of a pure register count without having any special knowledge. Understanding the sampling and expansion methods applied in the 2011 Census would require thorough knowledge of sampling methodology.

8. As the official number of inhabitants is one of the underlying factors of the “equalisation of revenue” system, it is a major basis determining the financial resources of the regional and local authorities. Consequently, some municipalities prepare already now to take legal action against the official number of inhabitants that will have been ascertained by the 2011 Census. The main point of criticism is that the number of inhabitants, rather than being obtained through a complete count alone, includes an estimate, which involves a sampling error. What is estimated by means of a sample survey with regard to the municipalities is the share of outdated (= overcounting) or missing entries (= undercounting) in the population register data. That share is then taken into account in determining the official number of inhabitants. The level of the sampling error of the official number of inhabitants will not be available before the census has been completed. However, it is laid down in the Census Act that, through the sample design, the relative standard error of the Census results should arrive at a maximum of 0.5 per cent.

## **III. Combining data from various data sources without personal identification**

9. Even though collecting and processing data from various data sources of the 2011 Census is not yet finished, challenges are already emerging that will have to be overcome when combining the data at the address and personal levels. In Germany there is no personal identification (ID) and, considering the data protection legislation, it is inconceivable at present that a standard personal ID could be introduced even in the long term. The experience acquired so far in the 2011 Census has shown, however, that this situation does not have to preclude conducting a census where data from traditional primary-statistical surveys have to be combined with data from administrative registers. Data combination is achieved through a set of variables suited to identify persons (among other things, name at birth, family name, first name, date of birth, place of birth and residential address). However, this requires much more effort than data combination by means of numerical identifiers.

10. The centre of the 2011 Census in Germany is the register of addresses and buildings (*Anschriften- und Gebäuderegister – AGR*) which contains all addresses of buildings with residential space in Germany (chart 1). It is a temporary statistical register that has been set up specifically for purposes of the 2011 Census and may be used only to conduct the 2011

Census. Due to legal provisions, the register and all its data will have to be deleted six years after the census reference date, at the latest. To set up the AGR, the geo-referenced address data of the land surveying administration, data from the population registers and the registers of persons of the Federal Employment Agency were used. The population registers are maintained in a decentralised way by the approximately 12,500 municipalities in about 5,400 computing centres. Setting up the AGR required combining the data sources by matching text (street names).

11. The AGR contains the universe of buildings and housing. It provides the data required for dispatching the questionnaires (name and address of the owner of the building) and a sampling frame for the household survey. In addition, it is used in data collection and processing for combining data from the various sources at the level of persons. Generally, that combination is a two-stage process. At the first stage, the personal data records are linked to the relevant address (using the variables of post code, official municipality code, sub-municipality, street name, house number and house number supplement where applicable) and only in the second stage will the matching of persons occur within the address in question.

12. When combining the data sources to set up the AGR, three weak points emerged. First, renaming of street names (e.g. in the case territorial status changes) is not done simultaneously in the three data sources; second, street names have different spellings (between the three data sources and even within a data source); and third, house number supplements are used in different ways. In most cases, the different spellings of street names were clarified through computer-aided processing, e.g. by standardising abbreviations or removing special characters and numbers. However, many cases remained which had to be settled by manual checking. For the approximately 1.2 million streets in Germany, there were about 5.1 million different spellings in the three data sources used to set up the AGR. Using the knowledge obtained on the different street name spellings, a street thesaurus was compiled, which was used when updating the AGR with new data deliveries. That was necessary because new data deliveries contained the old, non-standardised street names as it was not allowed to transmit the standardised spellings back to the agencies maintaining the registers. So, when integrating the update deliveries into the AGR, in most cases, it was no longer necessary to perform either the automated standardisation steps or – in particular – the time-consuming manual checks.

## **IV. Developing further the new census method**

### **A. Using standard address identifiers in the administrative registers**

13. Combining different data sources could be done in a much more efficient manner for future mixed-mode censuses if uniform standards of address variables were introduced in the administrative registers used for the census. An obvious code that could be used here is the geo-coordinate. As a starting point for developing such a list of addresses, the thesaurus of streets could be used. Such a thesaurus was developed for the 2011 Census. However, some issues would have to be settled first. For instance, there are licence issues and the question of how the process of updating the street thesaurus (mainly adding new street names) can be organised efficiently. An official list of addresses might be linked to the List of Municipalities that is maintained by the statistical offices of the Federation and the Länder. This means concretely that the statistical offices would offer the list of addresses free of charge, with liberal licensing terms that are easy to understand, and through state-of-the-art application programming interfaces (APIs). In turn, this would ensure that a standardised stock of addresses would be available to the statistical offices when they access administrative data.

## **B. Abandoning the updating of census results**

14. To provide current data on the demographic structure of the population in Germany, the results of the latest census are updated in a detailed regional (down to municipality level) and demographic breakdown (age, sex, marital status and citizenship). This is done by means of data from vital statistics (births and deaths) and from the statistics of arrivals and departures (only if occurring across municipality borders). Such updating of population figures is rather error-prone, especially if it has to be done over a long period.

15. It should, therefore, be attempted in the medium term to replace intercensal population updates by regular evaluations (at least once a year) of the population registers by the statistical offices of the Federation and the Länder. This would involve at least the following two requirements.

(a) The overcoverage and undercoverage of the population registers as ascertained by the 2011 Census would have to be within tolerable limits;

(b) Another considerable quality improvement of a networked register could be achieved by regularly (e.g. once a year) checking and adjusting the non-networked registers for double entries.

16. It would also be necessary to better account for the requirements of official statistics when the contents of population registers are defined. For example, items that should be stored in the population registers are, for all persons, their former places of residence within the last five years and, for naturalised persons, their former citizenships. Such data are necessary to delimit the target population of the census according to international standard definitions and to record migration variables.

## **C. Setting up a permanent register of buildings and dwellings**

17. The census of buildings and housing accounts for half of the costs of the German 2011 Census. In the register-based census conducted in Germany, it is necessary to have that largest primary-statistical survey component because there are no Germany-wide registers of buildings and dwellings. Also, the census of buildings and housing in Germany has to provide auxiliary variables that can be used to generate households and to combine them with the dwellings data. For the purpose, not only data on the building and the dwelling are collected, but also up to two names of main tenants/owner-occupiers per dwelling.

18. The 2011 census of buildings and housing was conducted as a postal survey among all owners of residential buildings (in Germany, there are almost 20 million buildings with residential space). To this end, a register of addresses had been set up containing the name and address of the owner for all addresses. The data on the owners of buildings had to be researched in administrative registers, e.g. from the real property tax data kept by the municipalities or from the administrative data of the supply and waste management enterprises. That research has turned out to be highly work-intensive. Furthermore, the quality of the data on owners can differ considerably between regions. The partly poor quality and, in particular, the inadequate up-to-dateness of the data on owners not only necessitated huge efforts of data collection, but also led to considerable problems of acceptance on the part of the respondents.

19. The German statistical offices would like to extend the success story of population register use and of the business register to include other areas. Therefore, efforts are being made to arrive at a situation where also the data on buildings, dwellings and households can

be combined from registers, which would mean rapid and low-cost data collection for the census. A register of buildings and dwellings – still to be set up – would enable that.

20. Such a register would have to contain complete and always up-to-date data on all buildings and dwellings and the relevant basic data needed to meet the requirements of the EU census regulation (year of construction, size of dwelling, ownership structure, type of heating).

21. The first-time set-up of such a register could be assigned to the statistical offices of the Federation and the Länder because they have wide experience in setting up the register of addresses and buildings for the 2011 Census. Moreover, without a new register of buildings and dwellings, statistical authorities would have to set up a register of addresses and buildings again for the 2021 Census in a few years' time. Not only official statistics, but also the public administration would benefit from such a register. Therefore, it should not be maintained as a statistical register, because in this case it would not be allowed to transmit data back to the administration; it should rather be maintained outside the protected sphere of official statistics.

22. The availability of such a register would notably reduce the amount of work for the statistical offices every ten years when a census is to be conducted. In addition, it would benefit many other official statistics. For example, updating the statistics on building activity would achieve a level of quality and up-to-dateness that has not been achieved so far. Also, the additional survey on the housing situation, which is conducted as part of the microcensus, could be abandoned, while losing only a few variables such as rent and energy source of heating; this would further contribute to reducing primary-statistical surveys to what is absolutely necessary.

---