

Distr.: General
18 April 2012

Original: English

Economic Commission for Europe

Conference of European Statisticians

UNECE-Eurostat Expert Group Meeting on Censuses Using Registers

Geneva, 22-23 May 2012

session 3: Availability, completeness and quality of data from registers and other sources

The population registers in Germany – the main data source in the 2011 Census

Note by Federal Statistical Office, Germany¹

I. The German Census Model – a short overview

1. Contrary to the last population census, the 2011 Census was conducted with a new method. To achieve the goal of minimising the burden on the population, administrative registers were used where possible. This allowed evaluation of all data contained in the register without having to contact the citizens directly, who form the universe of the census. Nevertheless, the information provided by registers on variables covers the whole of Germany and is entirely available for further use. This will allow small-area analyses of the census results without having to conduct a complete enumeration. This is why existing administrative registers are used and further survey components are added to be able to provide data on all census variables. Consequently, not all inhabitants were directly affected by the 2011 Census. Progress made in IT allows the methodological changeover from the conventional population census to the new census model. In addition to the benefit of placing a smaller burden on the population, registers provide area-wide variables allowing adequate analysis and combination of the results.

2. The register-based census consists of three major components providing information on the population as well as on buildings and dwellings in Germany. Those components are administrative registers, complete enumerations on buildings and dwellings and in specific areas and a household sample survey. Other variables, which cannot directly be covered by the register-based census, will be obtained by statistical methods applied in the household

¹ Prepared by Stephanie Hirner.

generating procedure. The generating process combines information of all components of the census and compiles new variables like the type of household or information about the family structure.

II. The use of population registers in the 2011 Census

3. In the 2011 Census the population registers mainly fulfil three functions.
4. On one hand the population registers were used for the build-up of a register of addresses and houses that is a steering file for the 2011 Census in Germany and the sampling frame for the household sample survey. For this purpose the addresses of the registered population are used to create a complete register by merging with other administrative registers such as the employment registers and registers from the surveying and mapping authorities. Detailed information about this register of addresses can be found in the supporting paper “The register of addresses and buildings – a combination of different registers” for this meeting.
5. On the other hand the population registers were used in support of the enumerators for the survey. For this purpose and for the planning and organisation of the survey the population registers were collected in 2010 thus six months before the reference date of the census. With this information the statistical territories could be planned and organizational papers could be prepared. Also the information from the registers could be used as an early indicator for the number of inhabitants the enumerators had to question.
6. The most important function of the population registers in the 2011 Census is that they are the main data source to cover the demographic and geographical census variables. The population registers are available area-wide in Germany and it is obligatory for inhabitants in Germany to register at their places of residence. Because of that the data from the residents’ registration offices are the basis for the identification of the official number of inhabitants. In the following the focus will be on the use of population registers as the main data source for determining the number of inhabitants.

III. Population registers: characteristics and data collection

A. Characteristics

7. Nowadays, technological progress enables us to use data that are already stored in administrative registers very efficiently. The most important administrative registers that are used in the census are the population registers. They are the main data source to be used to cover the demographic and geographical census variables. They contain information on every person registered with a place of residence in the Federal Republic of Germany. The population registers are managed by the residents’ registration offices that are legally obligated to transfer the data set to the official statistics for census purposes. As the residents’ registration offices are part of the public administration the availability and completeness of the registers is guaranteed at a very high level. The registered place of residence, according to population registration law, is stored there with the values of sole, main or secondary place of residence and, for the register-based census, corresponds to the compulsory EU variable of “place of usual residence”. Hence the data of the residents’ registration offices are the most important data source for determining the number of inhabitants in Germany. In addition, that administrative register contains the basic demographic data for every individual (sex, age, marital status, citizenship, place and state of birth).

8. Due to the requirement to register with the residents' registration offices, all municipalities in Germany maintain population registers that contain comparable information. The population registers are maintained in a decentralised way so every municipality uses its own register. Overall the raw structure of the registers is the same because it is based on the same parameters for the data set.

9. For an evaluation of the population register data, the decentralised data sets had to be combined to form one data set. According to the population registration law, several places of residence may be registered for persons who have one main place of residence and one or several secondary places of residence. But it may happen that persons are registered at the same time in several municipalities with their sole place of residence or main place of residence or only with a secondary place of residence. Such errors may occur where registration and deregistration are not done at the same time or they may be due to delays in administrative handling, to a person's failure to deregister, etc. After combining the population registers as one data source for the 2011 Census these errors had to be found and adjusted firstly.

B. Data collection

10. Based on the 2011 Census Act the data collection of the population registers was organised in a decentralized multilevel way. In the first step the residents' registration offices had to transfer their data set to the statistical offices of the Länder. For this purpose a standardised data record description was defined based on the institutional standard of the data sets in the population registers. In the next step the Länder imported the registers into a centralized database that is administrated by the Federal Statistical Office. The Länder were responsible for the data processing using computerised support. For the data integration and the plausibility checks an application was initialised. Both checks about valid parameter, values and combined values of variables were made with this application that was provided by the Federal Statistical Office. After the data transmission and validation of the Länder the Federal Statistical Office had data access to the different decentralised data. These registers had to be consolidated by using the information on addresses and personal characteristics of each data set. First the address variables were adjusted with the register of addresses and buildings, second the personal variables set up a population register linked with the register of addresses and buildings via a uniform identification number (address ID). In this way a temporary centralised population register for Germany was build up.

11. Although a standard data record description was used and the variables had homogeneous definitions and values, the quality of the population registers was sometimes different so the validation of the data was an important step during the data processing. One of our basic tasks was to resolve discrepancies in the registers. The quality of the registers depends further on the conduct to register for example after removing or emigrating. This is why it may happen that a person is registered at several places or still registered at a place but does not live there any more.

IV. Quality assurance

12. To increase the quality of the population registers the Census in Germany contains two main methods of correcting errors in the population registers especially to determine the number of inhabitants. On one hand the data stock of registers must be corrected. On the other hand the sample survey has the function to assure the data quality of the number of inhabitants by quantifying, estimating overcoverage and undercoverage in the population registers and correct them statistically.

A. Correction of the data stock

13. Correctly representing the variable “place of usual residence” requires adjustment of the register data stock by multiple or incorrect entries which are not allowed according to the population registration law but nevertheless happened. It is laid down there that persons are registered either with their sole place of residence or with a main place of residence and one or several secondary places of residence. It is not allowed to be registered with more than one sole or main place of residence or only with one or several secondary places of residence. For this reason the central data stock of the population registers was being analysed with the purpose to find incorrect registration. During checks for multiple registration, double counts and registrations that were incorrect in terms of registration law had to be identified and corrected in the data stock. For referencing the flow of register data to census-day (2011/05/09) the population registers had been requested for two reference dates, which was the census reference date and three months after the census reference date. From the second data capture only a part of the data stock was used for the following steps of the Census. Only data sets with date specifications before the census reference date were accepted for further processes, more precisely the date of birth and the move in date had to be prior.

14. In order to identify incorrect data sets constant variables had to be compared. In the process of finding incorrect data sets the variables “name”, “first name”, “gender”, “date of birth” and “place of birth” were used as mostly fixed demographic characteristics.

15. For checking and correcting the combined data stock the different data sets had to be compared in order to find doublets, triplets etc. and singular data sets. The doublets were defined in a first step as identical in the main demographic characteristics. Because of the registers being organised in a decentralised way without an overall unique ID it was necessary to search for doublets with similar values too, especially in the variables “name”, “first name” and “place of birth”. In particular register items for secondary places of residents or for inhabitants with migration background may be likewise different in population registers. Among other things this is due to the fact that the values from registered secondary places of residents are not used for official documents like the identification card or the polling card. That is why the data content of the registers sometimes is not checked as accurate as contents of sole or main places of residence. For inhabitants with migration background and more than one registered entry there are sometimes different values due to the fact that the transcription from foreign languages, for example Cyrillic or Asian, was made in different ways.

16. After searching for doublets and singular data sets the type of the place of residence had to be checked. In this process different combinations emerged as listed in the following.

- (1) *singular data set with main/ place of residence*
- (2) *singular data set with secondary place of residence*
- (3) *doublet both with main/sole place of residence*
- (4) *doublet with main/sole and secondary place of residence*
- (5) *doublet both with secondary place of residence*

17. Based on the population registration law the combinations (1) and (4) are correct. The other combinations had to be cleared.

18. For this purpose for persons only registered with one or more secondary places of residence – combination (2) and (5) – it had to be clarified at what address the relevant person actually had to be counted at the reference date. In these cases the data set with a

secondary place of residence could not be connected to another data set with a sole or main place of residence. The place of residence had to be cleared by a survey. That survey to clarify the residence was conducted to ensure that people for whom discrepancies in the population registers had been detected were correctly counted when determining the official number of inhabitants. These people were contacted by letter. They got a questionnaire and had to specify their type of place of residence. This survey was organised by the Länder, they used a central application to chart the results. The respondent answered and told either that this was his or her main place of residence or his or her secondary place of residence and gave information where his or her main place was. It was also possible that the questionnaire was undeliverable. All the different results were collected and analysed centrally. In conclusion the incorrect data sets either were switched to main places of residence or got a delete flag.

19. Combination (3) also contains an incorrect combination of data sets. These doublets can be divided in temporary multiple entries and permanently incorrect entries. Reasons for the former are lags of time between registrations and deregistrations. For this purpose the second data delivery to be transmitted from the population registers three months after the census reference date was used to trace exactly (for the reference date) the registrations and deregistrations that belong together but were not done at the same time. Here the population register data which referred to a person's moving into a dwelling on a day before the census reference date were combined with the first data delivery to census-day to form one data set. These temporary multiple entries were corrected in an automatic process where the current data set remained active and the other got a delete flag. In this process the current data set was defined as the data set with a move in date more relevant which means the date was closer to the census reference date.

20. Apart from this, the combined stock of population register data contained permanently incorrect entries. These doublets had to be cleared. For clarifying the incorrect data sets they were parted in doublets that belong to addresses placed in municipalities with less than 10 000 inhabitants on one hand and in municipalities with 10 000 or more inhabitants on the other hand. Background for this differentiation was that the population registers were also being corrected by the sample survey. But because of the sampling method and the minor sample size in small municipalities it was not efficient to adjust the population registers in these municipalities with the results of the sample survey. For this purpose the incorrect entries of the population registers had to be cleared by questioning in contrast to the adjustment of the registers in big communities where incorrect register entries were cleared automatically. With this differentiation it was possible to ensure that both in small and in big municipalities primary data were used to adjust the population registers statistically.

21. Incorrect doublets in big communities were corrected in the way that the data set with the previous move in date was accounted to be older. Thus the data set with the move in date closer to the census reference date remained in the data stock; the other data set got a delete flag. In a different way the incorrect data sets in small municipalities had to be adjusted. In this case every part of the doublet got a questionnaire that was sent by the Länder. Just as the questioning of inhabitants only registered with secondary place of residence the respondent had to inform about the type of the place of residence at the address he or she received the questionnaire. As a result of the questioning the different answers were put together and were compared. The answers of the questionnaires of a doublet match if either the results of two data sets could be combined as main and secondary place of residence or one questionnaire was undeliverable and the other was replied as main place of residence. If both respondents of a doublet answered that the address was his or her main place of residence, the two data sets had to be disconnected. Results and combinations that did not match or questionnaires with non-response had to be corrected centrally in the same way as doublets in big communities were cleared up.

22. That stock of population register data adjusted for multiple entries or entries with only a secondary place of residence are used for the further process of implementing the census model. The results of the survey or the automatic clearing of the data stock will only be used to determine the official number of inhabitants. The data will not be fed back either to the registration offices or to other authorities.

B. Correction via the sample survey

23. After the adjustment of the central data stock of the population registers by using information given in the registers the household survey assured the data quality of the number of inhabitants. This sample survey covered up variables and population groups that were not containing in registers, too. For example, it was not possible through the employment registers to cover employment variables of self-employed or of persons with mini jobs, and there were no education variables at all in the administrative registers. To bridge that gap, a sample survey had been conducted to cover variables and population groups not contained in registers. Apart from this function the household survey covered up the number of inhabitants living in the sample unit. This information assured the data quality of the number of inhabitants by quantifying and correcting overcoverage and undercoverage in the population registers statistically. For that purpose the interviewer had to determine all persons living at an address. To support the interviews, the interviewer got a list of names of all persons registered at the relevant address. In that way, the persons living at an address and identified in the household sample survey could be matched with the persons registered at the address as shown in the population registers. This allowed identifying the extent of overcoverage and undercoverage in the population registers and making the necessary statistical corrections. In municipalities with 10 000 and more inhabitants the information of the sample survey was used to correct the number of inhabitants by extrapolation of the results of overcoverage and undercoverage. In municipalities with less than 10 000 inhabitants the coverage of the sample was not high enough to use the results for extrapolation. In those municipalities the results of the questionnaires from the adjustment of the central data stock were used to adjust the registers only on the sample unit.

V. Population registers – the reference data stock

24. Besides the using of population registers as a source for census variables the data sets were used for building up the census reference data stock. This data stock has a steering function for the interaction of the different parts of the census because it allows consolidating the information of the different data sources. For this purpose the consolidation is organised in two steps. First the addresses in the different sources were linked by using the register of addresses and buildings. Second the correct inhabitants registered at the addresses were identified. That means that all information based on one of the census components (register, household sample survey and complete enumerations) has to be merged with the data sets from the population registers with help of the register of addresses and buildings. Thus the data stock based on the population registers is called reference data stock. Every data set of the reference data stock is linked to an address in the register of addresses and buildings. During the process of consolidation the information and data from the different components based on individual variables were adjusted with the data sets from the reference data stock in order to complete or correct the variables that are necessary. That means that on one hand variables or whole data sets were added such as the information based on the sample survey, on the other hand variables were corrected, for example information about the type of the place of residence. The function of the reference data stock in the census model is to assure that all information is combined in one data

stock. For that purpose every other component of the census model had to integrate its information and had to use the information being contained in the reference data stock. As in Germany there are different statistical offices in the Länder it is important that all of these Länder can correct and use the reference data base. Therefore a web-application has been programmed.

VI. Conclusion

25. The new Census in Germany is a complex procedure requiring co-ordination and communication between the individual survey components and data sources. Although providing benefits, using registers also involves various problems which have to be taken into account and solved in planning and implementing the concepts. Especially the differences in data quality between the individual registers make it difficult to combine information. Also, the fact that there are no standard identifiers for persons or dwellings in Germany and the decentralised storage of data in many registers make data combination difficult. Nevertheless, the population registers provide area-wide information about all inhabitants which allows detailed interpretation of the data. In particular for population registers it is necessary first of all to set up a federal data stock, which then has to be co-ordinated with the other data sources. This central reference data stock is necessary to conduct the complex census model and to consolidate the different data sources.
