



**Economic and Social
Council**

Distr.
General

ECE/CES/2006/SP/10
11 May 2006

ENGLISH ONLY

ECONOMIC COMMISSION FOR EUROPE

STATISTICAL COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fourth plenary session
Paris, 13-15 June 2006
Item 6 of the provisional agenda

**SEMINAR ON POPULATION AND HOUSING CENSUSES
SESSION II**

Planning the register-based Census 2011 in Germany¹

Submitted by the Federal Statistical Office of Germany

I. RESULTS OF THE CENSUS TEST

1. “Germany will take part in the new census round planned for 2010/2011 at the EU level, which should involve minimum burdens on the citizens and minimum costs.” Due to that provision in the government coalition agreement, German official statistics can continue its plans for a register-based census and implement the results of the census test performed in 2001/2002.²

2. On the basis of the results of the census test surveys, the statistical offices of the Federation and the Länder have recommended to the political decision-makers a census model for the next census which should consist of the following components:

- (a) demanding the population register data from all municipalities (about 12 600) for two reference dates; each of the two data stocks will contain about 88 million individual data records (about 6 million people are registered with a secondary place of residence);

¹ This paper has been contributed by Hans-Gerd Siedt, Federal Statistical Office of Germany.

- (b) demanding register data from the Federal Employment Agency (for about 34.5 million employees subject to compulsory social insurance contributions) and from the public administration (for about 2 million civil servants, army personnel and judges);
- (c) conducting a postal housing census by interviewing the about 17 million owners of the buildings;
- (d) checking the central population register for double (multiple) entries; it can be expected that about 90% of the discovered double entries can be cleared up without further inquiries with the persons concerned;
- (e) combining individual persons to form private households by matching the records of the population registers with the records of the housing census; by this means, about 38.5 million households will be created; this generation of households will also be used as a tool to discover erroneous entries in the population registers; for this purpose about 1.7 million persons (living in one-family or two-family houses) residing in municipalities with less than 10 000 inhabitants will have to be interviewed (by phone, mail or direct interview).

3. An **analysis of the test results** has quickly shown that there are very high error rates in population registers in the case of special buildings (e.g. institutions, students' halls of residence), which can be identified only with the help of primary surveys. Accordingly, the recommended surveys of institutions' administrators and students living in halls of residence are regarded as basic components of a future census just like the statistical methods for population register checking provided for in the census test law. By means of these components of the register-based census, the rate of outdated entries is reduced from 4.1 to 1.8%. The rate of missing entries, in contrast, amounts to 1.7%. For municipalities of differing sizes the error rates (rates of outdated and missing entries) vary strongly and thus also the accuracy to be expected of the numbers of inhabitants. Further corrective measures in the form of supplementary sample surveys are needed.

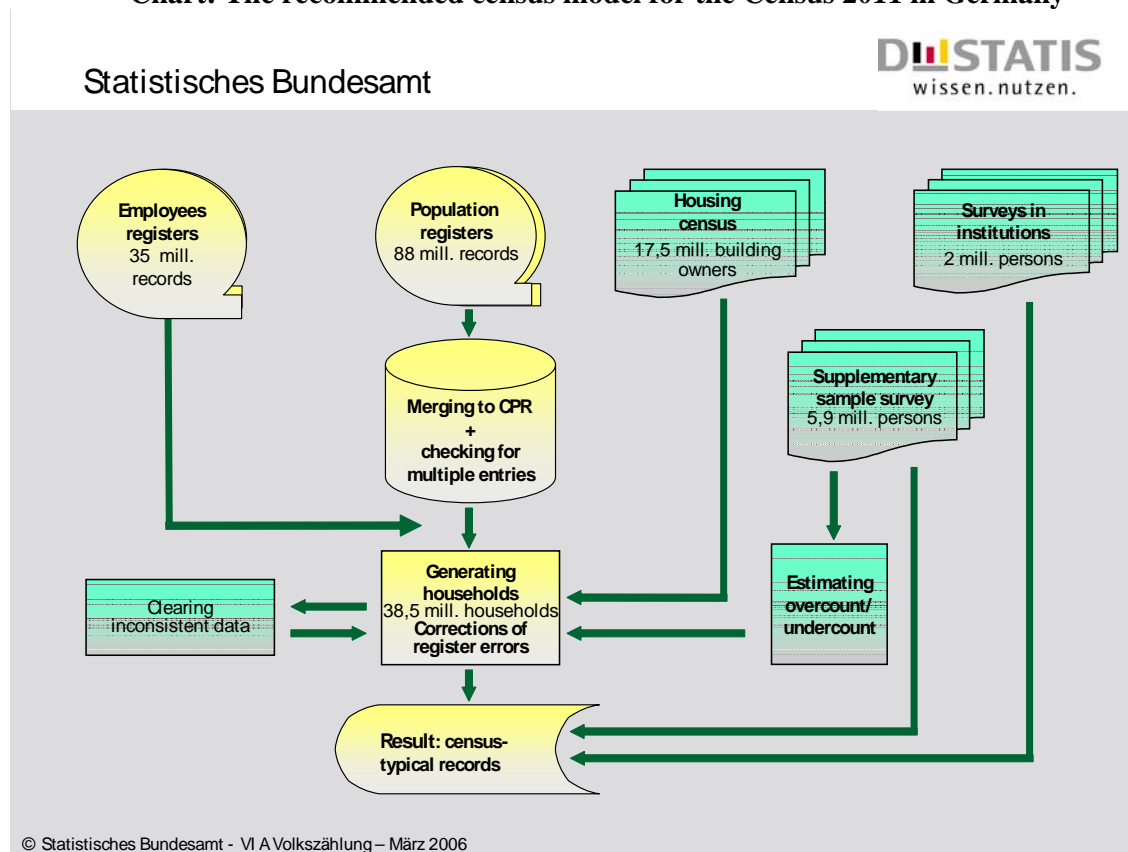
4. The **overcounts in the PR can be significantly reduced** by applying the tested adjustment procedures (comparing data stocks for two reference dates in order to discover and to clear up "temporary" overcounts, double-entry check, interviewing households living in one or two-family houses, for which inconsistencies between data from the housing census and PR data were stated in the household generation process). However, the remaining errors in the PR, that is in particular the remaining relatively big variance of overcount rates between the municipalities, cannot be accepted as a census result. The statistical offices of the Federation and the Länder have therefore suggested to extend the tested model of a register-based census with a sample survey (random sample of addresses from each municipal PR), that will allow limiting the size of register errors, at least by probability.³

5. To **keep the costs of a register-based census low**, the supplementary sample survey to adjust population registers statistically may be limited to municipalities with more than 10 000 inhabitants. In that case, only the basic component surveys would be conducted in smaller municipalities. Additionally, direct enquiries would be made in these municipalities about the actual residents of one-family and two-family houses whose generation result is not plausible, thus reducing the rate of outdated entries considerably. Therefore, the components

- (a) collecting census data for persons living in institutional households and for students living in student hostels (by interviewing the managers), and

(b) conducting a supplementary sample survey in about 1 500 of 12 600 municipalities with 10 000 and more inhabitants (overall sample size: 5.9 million persons) are parts of the recommended census model, too. The supplementary sample survey affords the possibility of covering additional characteristics not included in registers. The flowchart of the census model with a supplementary sample survey would be as follows:

Chart: The recommended census model for the Census 2011 in Germany



6. The recommended **census model reduces the burden** placed on respondents and on taxpayers. In the census model, about 8.5 million persons are covered to collect the characteristics of population and employment statistics, that is only about 10% of the persons questioned in a conventional census. For the characteristics of building and housing statistics, the group of respondents is reduced by more than half because the persons questioned are not the owners or main tenants of dwellings but the owners/administrators of the buildings, and they are questioned by mail. The costs of the recommended census model are some EUR 350 mn (status: 2002), that is about one third of the costs of a conventional census in Germany.

II. METHODOLOGICAL FOCUSES OF PREPARATION

7. The **price of reduced burdens is higher complexity of the census** procedure, which has required redevelopment of many steps of the procedure and which will lead to a complete

rearrangement of population statistics in the long term. Some of the preparatory activities will be described in detail below.

8. To prepare the register-based census (census), the Federal Statistical Office and the statistical offices of the Länder will set up a **nation-wide register of all buildings** providing housing space and of inhabited accommodations (register of addresses/buildings). The **purpose** of the register of addresses/buildings is (a) to **check the completeness** of the buildings and dwellings to be covered in the census (determining the universe) and to control methodical evaluations; (b) to **coordinate** the required **matching** of the various data sources; (c) to control the process of the **postal census** of buildings and housing; (d) to **provide the universe** for the sample surveys required for the census, and (e) to **create the basis for small-area evaluations** of the census results.

9. To set up the register of addresses/buildings, files of the land surveying offices, the residents' registration offices and the Federal Employment Agency will be used. Those basic files have been created independently of each other and allow mutual checking of the survey unit addresses contained in them. Fundamental importance is attached to the files of the land surveying offices, which have set up a comprehensive and nation-wide register of buildings in Germany; that **data stock is referred to as "house co-ordinates"**. That data stock contains not only information on the residential address but also the official municipality code and a street code. What is of particular importance here is that the addresses of buildings have been extended to include the relevant building co-ordinates, so that spatial reference is included. By means of the house co-ordinates, **any address can be selected and shown in digital databases, maps or aerial photographs**. Using suitable software, they allow direct orientation and positioning. The "house co-ordinates" file is combined with the data from the residents' registration offices and the data of the Federal Employment Agency, based on the person-related addresses contained in them. This permits not only to detect and eliminate errors, but also – and especially – to mark the buildings as (not) providing housing space. Before starting to set up the register of addresses/buildings, German official statistics needs a legal basis, which is currently being prepared. Setting up the register of addresses/buildings will start in 2008. Particularly time-consuming and work-intensive is the determination of the addresses of the owners or administrators of buildings to be questioned.

10. With residents' registers maintained on a decentralised basis, it cannot be ruled out that persons are registered in several municipalities at the same time with their sole or main place of residence, or that they are registered only with a secondary place of residence. Such errors occurred in the past because persons did not register and deregister at the same time, administrative handling was delayed, or persons did not deregister at all. If the **residents' registration data** were used for census purposes without **checking the data stock for duplicates**, some persons would not be counted at all or would be counted several times, at the wrong place, or with a wrong residence status. This would lead to erroneous numbers of inhabitants. It is therefore necessary to check the data stock for duplicates by means of the key items sex, year of birth, month of birth, day of birth, place of birth, name at birth, and first names; such checks are automated. Checking for duplicates is designed as a similarity search because inaccuracies may occur in the handling of registration cases, different spelling of names may occur, for example when foreign-language names are transferred into German spelling, or abbreviations are used, for example for places of birth.

11. The implementation of duplicate checking in the census in terms of program technology must meet high requirements. Duplicates are searched for in a data stock of some 88 million data records. This must be finished soon after the reference day. To prepare duplicate checking, the data contents must be standardised in a particular manner and in some cases they must be extended to include their “phonetic” variant. The complex algorithm developed for the similarity search calculates the similarities of strings by trying to derive a target string from a given string (e.g. name at birth). What is used here is a generalised form of calculating the Levenshtein Distance⁴. For every manipulation (e.g. inserting, deleting, replacing, swapping, doubling, truncating of characters) that is required to obtain the target string, (different) scores are attributed. When added up, the scores indicate the manipulation costs. Those manipulation costs are put in relation to the string length and standardised to the interval between 0 and 1. The distance measure thus obtained will show the maximum of “1” for identical names at birth. In other cases, lower scores are attributed, depending on the manipulation. The similarity search will result in a file containing the records which have high similarity values for the primary variables.

12. As a result of the census, a data record is obtained for every person, containing the relevant variables in terms of demography, employment statistics, family statistics, household statistics and housing statistics. In the census model envisaged for Germany, that census-typical data record is obtained by **combining the different data sources** at the levels of municipalities, buildings and persons. As there is no personal identification number in Germany, such combining must be **based also on name, first name, date of birth, etc.** That work package, too, sets high standards to be met by the IT infrastructure.

13. The statistical offices of the Federation and the Länder have suggested that **supplementary sample surveys** be planned for selected addresses, that is especially in **municipalities with 10,000 or more inhabitants**. They should be used, first, to estimate additional person-related census variables that are not available from registers (e.g. in the areas of education and employment) and, second, to obtain information that can be used for statistical elimination of errors in residents’ registers in municipalities with 10,000 or more inhabitants (overcoverage and undercoverage). Two things – the understandable demand made by the scientific community and local statistical offices that census results be evaluated in a flexible manner as not every evaluation can be planned in detail, and the planned evaluations of census results for non-administrative territorial units by means of geo-information – require that overcoverage in the residents’ registers (outdated entries) be deleted from the data stock on the basis of individual data and that undercoverage (missing entries) be imputed on the basis of individual data. The deletion or imputation process is controlled by the estimates from the sample survey regarding overcoverage and undercoverage in the residents’ registers. What is basically done here is to adjust the corrections to the marginal distributions of the register errors estimated from the sample surveys (outdated and missing entries) regarding demographic core variables such as sex, age, marital status and citizenship as well as variables of employment statistics. What is also planned is an adjustment to the distribution of variable combinations (e.g. age groups by sex), where this can be estimated with sufficient accuracy from the sample survey. Optimising such estimations requires using modern and efficient estimation methods, which are developed with the help of external experts. Support by the scientific community is also required to optimise the estimation of variables that are covered only by the sample survey, especially for small-area

results. Even if small-area methods are applied, high random errors are to be expected for small-area results; such errors will be checked as part of an error calculation and will be subject to indication as uncertain results as is usual for publications of official statistics.

III. SUPPORTING MEASURES AROUND THE CENSUS

14. Based on the **2003 law on tax changes**, every natural person will be attributed an **identification number** – according to current planning, starting in 2007 – for purposes of identification in the taxation procedure. That identification number is unique all over Germany and is permanently allocated to every citizen registered in Germany. The Federal Central Office for Taxes (BZSt) is setting up a database for the purpose, containing the legally defined data for all taxpayers. Allocating the identification number is based on information provided by the local residents' registration offices. Also, the identification number will be stored in the data record of the registration office, so that a change of place of residence and other changes (e.g. change of name) will trigger a modification note to the BZSt. In that way it is to be ensured that the data stocks are, and remain consistent between the residents' registers and the BZSt database. It is assumed that the number of cases exceeds 10 million per year. To allow the BZSt to clearly identify a taxpayer, it is necessary that the data of the BZSt are consolidated, checked for double registration and, where necessary, corrected in co-operation with the residents' registration offices. Also, every taxpayer will be informed by post about his identification number. It is expected that this will draw the attention to outdated entries (e.g. persons who have moved abroad and have not deregistered) and that the residents' registers will be adjusted for those cases in co-ordination with the BZSt database. As it is **not possible to use the identification number for statistical purposes**, the census will benefit "only" from the consequences of its introduction, i.e. an improved quality of the residents' register data.

15. From **2007, feedback of the residents' registration offices will have to be done by data transmission**. Such feedback means that the registration office of "arrival" informs the registration office of "departure" about a registration performed there. Any inhabitant is obliged to register under his new address, so that he will be added to the residents' register with his new address, while he will be deleted with his old address. This is done through a task-related linkage between the registration offices via electronic networks using virtual post offices with mail boxes. This allows to reduce to a few hours or one day the interval in which outdated entries occur in the residents' registers during the administrative process because persons move from one place of residence to another. Thanks to the new procedure, some possible errors will no longer occur, such as forgetting to give feedback, the loss of the feedback document during postal transport, delays in registering and handling the feedback at the registration office of "departure" or the failure to do so. The quality of the residents' register data is further improved by the fact that, in case of an arrival, the registration office of "arrival" will request electronically from the registration office of "departure" the entire data record of the registering person; the official in charge will then show the data to the person concerned, will ask him to complete or correct the data and will then perform the registration. This reduces the burden on the citizen and allows the registration office not to enter the registration data, thus helping to avoid input errors.

16. The introduction of the **identification number in tax law and the networking of the residents' registers will improve the quality of the registers**. To what extent the quality will

be improved will depend on the concrete implementation of the work started. The census 2011 thus has the additional function to estimate the success of the consolidation and implementation measures. However, the implementation of the highly ambitious plans may also involve disadvantages for the census preparation. This may happen if the consolidation and implementation measures for the introduction of the tax number and the networking take place in the phase when the registration offices have to start the preparatory work for the census 2011. This will be the case from 2008.

IV. OUTLOOK FOR POPULATION AND HOUSING STATISTICS

17. The census will provide reliable information not only for the census year, but it will also put the system of population statistics on a solid basis for a long period. In current population statistics, the results of a population census are updated at municipality level, using the data of the residents' registration offices on arrivals and departures and using the reports of the registrar's offices on births and deaths. Due to the networking of the registration offices and the further progressing data technology, it may appear possible in the medium term that the above procedure of current population statistics be replaced by evaluations of the residents' registers.

18. If Germany wishes to satisfy in the future the data demand of the census of buildings and dwellings (GWZ) from register data, it would be an obvious choice to use the GWZ data obtained in the census 2011 for setting up a nation-wide buildings and dwellings register. Data of the buildings administration could be used for updating the buildings and dwellings register. The questions arising in this context (i.e., which variables needed for the updating procedure could be obtained from the buildings administration, or what are the estimated costs of the updating procedure) still have to be scrutinized. This should be done early enough as to allow official statistics to give a recommendation to politicians in connection with preparing the census law.

* * * * *

² The results of the census test are published in: Statistische Ämter des Bundes und der Länder, Ergebnisse des Zensusstests, Statistisches Bundesamt, Wirtschaft und Statistik 8/2004, pp. 813 et seqq. The publication is available in a German and an English version from Gruppe-VIA@destatis.de.

³ The underlying idea of supplementary sample surveys is to survey persons in the municipalities on a sample basis in addition to population register evaluations in order to determine the rate of outdated and missing entries for the individual municipality and to correct the number of inhabitants initially determined in the respective order of magnitude. For that purpose, addresses are selected in all municipalities with the help of a random sample stratified, among others, by municipality and address sizes, and the de facto inhabitants are identified by persons entrusted with the execution of the survey. By comparison with the population register data, deviations of the register are determined and extrapolated to the entire municipality. The sampling plan is designed in such a way that with a confidence probability of 95% the deviation of the extrapolated result from reality will amount to a maximum of $\pm 1\%$ on average for the municipalities. The error rates determined by sampling for quality checking are used to correct the official number of inhabitants.

⁴ For the basics cf. V. I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. Doklady Akademii Nauk SSSR 163(4) pp845-848, 1965; also Soviet Physics Doklady 10(8) pp707-710, Feb 1966.