

The usability of administrative data for register-based censuses

Eric Schulte Nordholt

Statistics Netherlands, P.O. Box 24500, 2490 HA The Hague, The Netherlands
Tel.: +31 70 3374931; Fax: +31 70 3877429; E-mail: e.schultenordholt@cbs.nl

Abstract. The Dutch 2011 Census tables were produced by combining existing register and sample survey data. Since the last census based on a complete enumeration was held in 1971, the willingness of the population to participate has fallen sharply. Statistics Netherlands no longer uses census questionnaires and has found an alternative in the register-based census, using only existing data. The register-based census is cheaper and more socially acceptable. The table results of the Netherlands are not only comparable with earlier Dutch censuses, but also with those of the other countries in the 2011 European Census Round, although most other countries do not yet have register-based censuses. The Dutch Census 2011 cannot be considered in isolation from its predecessors. This most recent census builds on a rich tradition of traditional population and housing censuses carried out in the period 1829–1971. For the European Union censuses of 1981 and 1991, Statistics Netherlands provided alternative census data consisting of a combination of register and survey data, but without complete numerical consistency between census tables and without detailed information about regions and specific categories. This combined data provision was continued for the censuses of 2001 and 2011, by which time it was possible to compile completely consistent data at a very detailed level. The 2011 Census data are more detailed than those of 2001, and include the hypercubes (high dimensional tables) that are mandatory for all EU countries under the 2008 European Census Act. The challenge for the future is to keep comparability, but to improve the timeliness of the European census results.

Keywords: Administrative data, traditional census, register-based census

1. Introduction

1.1. Data collection at Statistics Netherlands

At Statistics Netherlands, the data collection has been centralised in one input division. In this division, both register and survey data are collected. This change has been made to be able to work more efficiently and professionally. Now Statistics Netherlands uses the same infrastructure for social and economic statistics as much as possible. In its collection strategy the preferred order is to first look at register data. If variables or categories are missing and cannot be derived from register information, then sample surveys are conducted. In the Census 2011 Statistics Netherlands only used already existing sources. Therefore, questionnaires for the Census 2011 were not necessary.

1.2. History of censuses in the Kingdom of the Netherlands

Following the withdrawal of Napoleon, the Netherlands became a kingdom and a census was held once every ten years. The first census of the Kingdom of the Netherlands was taken in 1829. Before the official institution of a central bureau of statistics, the Ministry of Home Affairs organised another six censuses (in 1839, 1849, 1859, 1869, 1879 and 1889). In 1899 the Central Bureau of Statistics (nowadays called: Statistics Netherlands) was established, and was put directly in charge of the eighth census. Six more traditional censuses were carried out in the twentieth century: in 1909, 1920, 1930, 1947, 1960 and 1971. During the 1971 Census enormous privacy debates took place and thereafter the government decided to stop the traditional censuses in the Netherlands. In taking this decision also the growing unwillingness to participate

played a role, but the main aspect was the privacy issue [1]. The four most recent censuses (1981, 1991, 2001 and 2011) were not based on a complete enumeration but on registers and surveys available at Statistics Netherlands. Until now, 18 censuses have been held in the Kingdom of the Netherlands.

The Dutch data compiled for 1981 and 1991 were much less detailed than the sets of tables compiled for the 2001 and 2011 Censuses. Moreover, they were largely based on a register count of the population in combination with the then existing surveys on the labour force and housing conditions. The amount of information that has become available in the 2011 Census is again larger than that of the 2001 outputs. This is mainly the result of the introduction of so-called hypercubes (high dimensional tables): more variables have now been cross-classified than ever before.

In 1991 the Census Act was rescinded, officially terminating Statistics Netherlands' obligation to conduct a census every ten years [2]. For the 2001 Census there was no obligation to produce census data, but Statistics Netherlands compiled a set of tables based on a gentlemen's agreement between the European Union and its Member States at that time. The 2011 Census has its legal basis in the 2008 European Census Act [3] and three implementing regulations [4–6]. Eurostat has a coordinating role in collecting harmonised data on the EU and a duty to make international comparisons of the outcomes.

Originally, censuses had two aims. First, to correct errors in the municipal population registers. Second, to obtain extra information about socio-economic phenomena in the country. As the Netherlands conducts a register-based census, the first aim no longer exists. In addition, the quality of the central population register (PR), which unites all municipality population registers, has improved considerably: as central government funding is allocated based on population size, it is in the municipalities' own interest to keep their population registers up-to-date. Another reason for the improvement is that it is extremely difficult to function in Dutch society if you are not in the PR. Both municipalities and citizens, therefore, have enough incentive to maintain a high quality PR. The second aim is still valid and many census results are published in a historical or international context. At present, census data are also popular for comparisons between countries.

1.3. Preparations for the Dutch Population and Housing Census 2011

Census experts at Statistics Netherlands started preparations for the 2011 Population and Housing Census

in 2008. In 2009, they started work on the data collection procedures required to collect the census information about the 16,655,799 people living in the Netherlands on 1 January 2011.

Data from different sources were combined to produce the 2011 Census tables. These data were not obtained by interviewing inhabitants in a complete enumeration, as in traditional censuses in most other countries, but by using data from registers and sample surveys that were already available at Statistics Netherlands. The main sources were:

- Population Register (an automated population registration system used by Statistics Netherlands for all demographic statistics, see [7]);
- Jobs File (containing information about all employees in the Netherlands);
- Self-employed File (containing information about all self-employed in the Netherlands);
- Fiscal Administration;
- Social Security Administration;
- Pensions and Life Insurance Benefits;
- Housing Register;
- Labour Force Survey.

This approach of combining different sources has a number of advantages and disadvantages that will be explained in the last section of this chapter.

1.4. The Dutch Population and Housing Census 2011

A number of analyses based on the Dutch Population and Housing Census 2011 can be found in the book published about this census [8]. A number of census variables have been harmonised in Europe. According to [4] 35 variables are mandatory: 16 person variables, 7 household variables and 12 housing variables. An overview of these variables is given below.

Person variables

1. Place of usual residence
2. Location place of work
3. Locality size
4. Sex
5. Age
6. Legal marital status
7. Current activity status
8. Occupation
9. Industry
10. Status in employment
11. Educational attainment
12. Country/place of birth
13. Country of citizenship

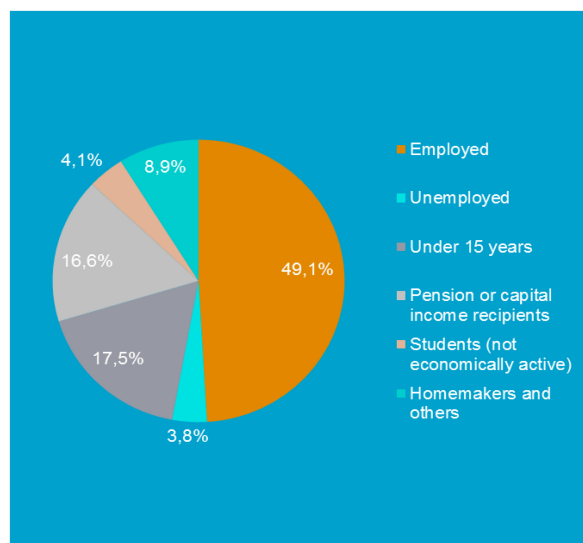


Fig. 1. Population in the Netherlands by current activity status. Source: Census 2011.

14. Year of arrival in the country
15. Place of usual residence one year before
16. Housing arrangements

Household variables

1. Household status
2. Family status
3. Type of family nucleus
4. Size of family nucleus
5. Type of private household
6. Size of private household
7. Tenure status of households

Housing variables

1. Type of living quarters
2. Occupancy status of conventional dwellings
3. Type of ownership
4. Number of occupants
5. Useful floor space and/or Number of rooms
6. Density standard
7. Water supply system
8. Toilet facilities
9. Bathing facilities
10. Type of heating
11. Dwellings by period of construction
12. Dwellings by type of building

The Netherlands could copy or derive most of these variables from the existing registers. Only for two variables (occupation and educational attainment) information from a survey (the Labour Force Survey)

was needed. Some derivations were rather complicated with a number of priority rules. Especially for the variable current activity status, many of such rules (see [4]) had to be applied. For example in ascribing a single activity status to each currently not economically active person, priority shall be given to the status of 'Persons below the national minimum age for economic activity' in preference to 'Pension or capital income recipients', to the status of 'Pension or capital income recipients' in preference to 'Students (not economically active)', and of 'Students (not economically active)' in preference to 'Homemakers and others'. In the Netherlands the national minimum age for economic activity is 15 years. In Fig. 1 the results for the census variable current activity status can be found.

Information about all European census tables for 2011 can be retrieved freely from the so-called Census Hub (see <https://ec.europa.eu/CensusHub2/>). This system contains information for 32 European countries: the 28 European Union Member States and the 4 EFTA countries.

1.5. Advantages and disadvantages of the Dutch register-based census

One of the advantages of this approach is a much lower census bill for Dutch taxpayers. A traditional census in the Netherlands would cost a few hundred million euros, while with this method it costs 'only' around 1.4 million euros. This bill includes the costs for all preparatory work, such as extending the methodology and updating and developing accompanying software, as well as the analyses of the results. It does not include the costs of the registers, as these are not kept for censuses, but primarily for other purposes. In addition, under Dutch law, Statistics Netherlands may access government registers free of charge. This low-cost census approach is only possible for countries with sufficient register information. A register-based census costing less than 1 percent of a traditional census is not exceptional [9]. Today, the huge costs of traditional censuses are often justified by pointing out the enormous implications of the census results for regional funding distribution. But a register-based census would be impossible in many countries, because of the lack of sufficient register data and access restrictions to the existing registers.

Apart from the financial aspect, there are also other important differences between a traditional census and the register-based census conducted in the Netherlands. A well-known problem with traditional cen-

suses is that participation is limited and selective. In spite of the mandatory character of a traditional census, part of the population will not participate at all (unit non-response) and those who do will not answer all questions (item non-response). Although correcting for non-response by weighting and imputation techniques is worth trying, traditional correction methods are inadequate to obtain reliable results. The last traditional census in the Netherlands, in 1971, met with many privacy objections against the collection of integral information about the population living in the Netherlands. This increased the non-response problem, and non-response was expected to be even higher if another traditional census were to be held in the Netherlands [2]. There are almost no objections to a register-based census in the Netherlands and the non-response problem only plays a role when survey microdata are reused.

Another advantage of the register-based census is the short production time. The register-based census in the Netherlands got off to a later start than traditional censuses in other countries. It would have been pointless to start the production phase of the 2011 census project before all sources were available, and some registers became available relatively late. In spite of this delay, Statistics Netherlands compiled its census tables faster than most other countries in the 2011 European Census Round. In fact, the Netherlands had one of the shortest production times for the complete set of tables required by Eurostat. Statistics Netherlands had the advantage that no incoming census forms had to be checked and corrected.

A disadvantage of the Dutch census is that for some variables only sample information is available, which meant it was impossible to meet the level of detail required in some census hypercubes. At the moment, however, the Netherlands perceives the advantages of the register-based census in terms of cost and non-response problems to amply outweigh the loss of some detail compared with a traditional census.

Statistics Netherlands is not the only country that uses registers to produce census information. Four Nordic countries (Denmark, Finland, Norway and Sweden), Austria and Slovenia have more variables available in registers than the Netherlands, and the problem of insufficient detail in the outcome does not play a major role there. Most of the other register-based countries are in a similar position to the Netherlands: not all variables relevant for the census can be found in registers. They are therefore very interested in the Dutch approach of combining registers and existing

sample surveys and using modern statistical techniques to compile the hypercubes. Obviously, it is essential that statistical bureaus are permitted to make use of registers that are relevant for the census. For Statistics Netherlands this is laid down in the statistical law that came into force in 2004 (see <https://www.cbs.nl/en-gb/about-us/organisation> for an English translation of the most recent version of the Statistics Netherlands Act). Nevertheless, Statistics Netherlands will have to maintain the good contact it has established with register holders over the last 25 years. For the census (and all other register-based statistics) Statistics Netherlands has become very dependent on the register holders. Timely deliveries with relevant variables for Statistics Netherlands are crucial for official statistics production.

All tables produced in the Dutch Census 2011 are numerically consistent. The results can be compared over time as Fig. 2 illustrates. In the period 1829–2011 the population has grown continually from 2.6 to 16.7 million people. The average private household size in the Netherlands has decreased since the beginning of the twentieth century. Together with the growth in prosperity in the post-war period, the average size diminished rapidly. Not only has the number of children per couple decreased since the 1960s, but children are leaving home earlier and live alone longer than half a century ago. From 4.8 persons in 1899 the average private household size has dropped to 2.2 in 2011. Specifically worth mentioning is the decrease since 1947, when the average size was still 4.0 persons.

In addition, international comparisons between census results are possible and the earlier mentioned Census Hub is facilitating such comparisons between European countries. Finally, Dutch census data are continuously available: there are no restrictions to official census years as census tables can easily be updated every year (see also Chapter 4).

2. Census methods in the UNECE region

2.1. *Census practices in the 2010 census round*

As mentioned in the previous chapter, a number of countries conduct register-based censuses. As it is interesting to know how countries conduct their censuses, in 2013 the United Nations Economic Commission for Europe (UNECE) conducted an online survey among its members to collect information about national census practices in the 2010 Census Round,

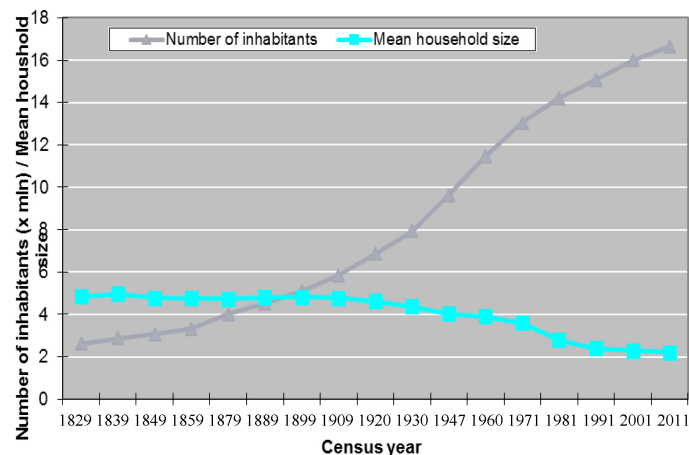


Fig. 2. Number of inhabitants and mean household size in the Netherlands, 1829–2011. Source: Statistics Netherlands.

and about plans for the 2020 round. All EU and EFTA countries are also members of the UNECE, which also includes Canada, the Russian Federation and the United States, among others. Response to the UNECE questionnaire was high and the results of fifty countries on important methodological issues were analysed [10].

As expected the countries used different methods, and some countries reported a different method for the population than for the housing census, often connected with the availability of registers for these domains. Using registers to produce official statistics reduces costs and bypasses the problem of declining survey response rates. Three main types of census method can be distinguished: the traditional census, the combined census, and the register-based census.

2.2. Traditional censuses

The traditional census approach collects basic characteristics from all individuals and housing units (full enumeration) for a specific point in time. More detailed characteristics can be collected either from the whole population or on a sample basis. Collection modes include personal interviews, self-completed paper questionnaires, and data collection by telephone and the internet. Across the world, this is still the most common approach to census taking. Most UNECE countries with a traditional census use personal face-to-face interviews with paper questionnaires as their main approach. However, in the Czech Republic, France, Ireland, Italy, Luxembourg, Slovakia, the United Kingdom and the United States, the main method is self-completion of paper questionnaires by respondents. In

Canada most respondents participate online (CAWI), while in Portugal self-completed paper questionnaires and online response were equally popular.

Just as in the census round of 2000, full field enumeration without register information (traditional census) is still the most popular method in the UNECE region in this census round. Almost two-thirds of countries collected data using ‘traditional’ methods. But although it is still the most common general approach in the region, it is less so than in the 2000 round, when four-fifths of countries used this approach. A substantial minority (33 percent) of the full field enumeration countries used information from registers only as a frame or control. The United States was alone in using traditional enumeration with yearly updates of characteristics on a sample basis. France used another alternative approach to the traditional model: the rolling census. This is a cumulative continuous survey covering the whole country over a period of time rather than on one particular day.

2.3. Combined censuses

Four countries (Estonia, Latvia, Liechtenstein and Lithuania) used a combination of register data with complete field data collection for selected population census variables, and six countries (Germany, Israel, Poland, Spain, Switzerland and Turkey) used a combination of register data with ad-hoc sample data collection for selected population census variables.

2.4. Register-based censuses

A growing number of EU and EFTA countries have switched to methods without field data collection, rely-

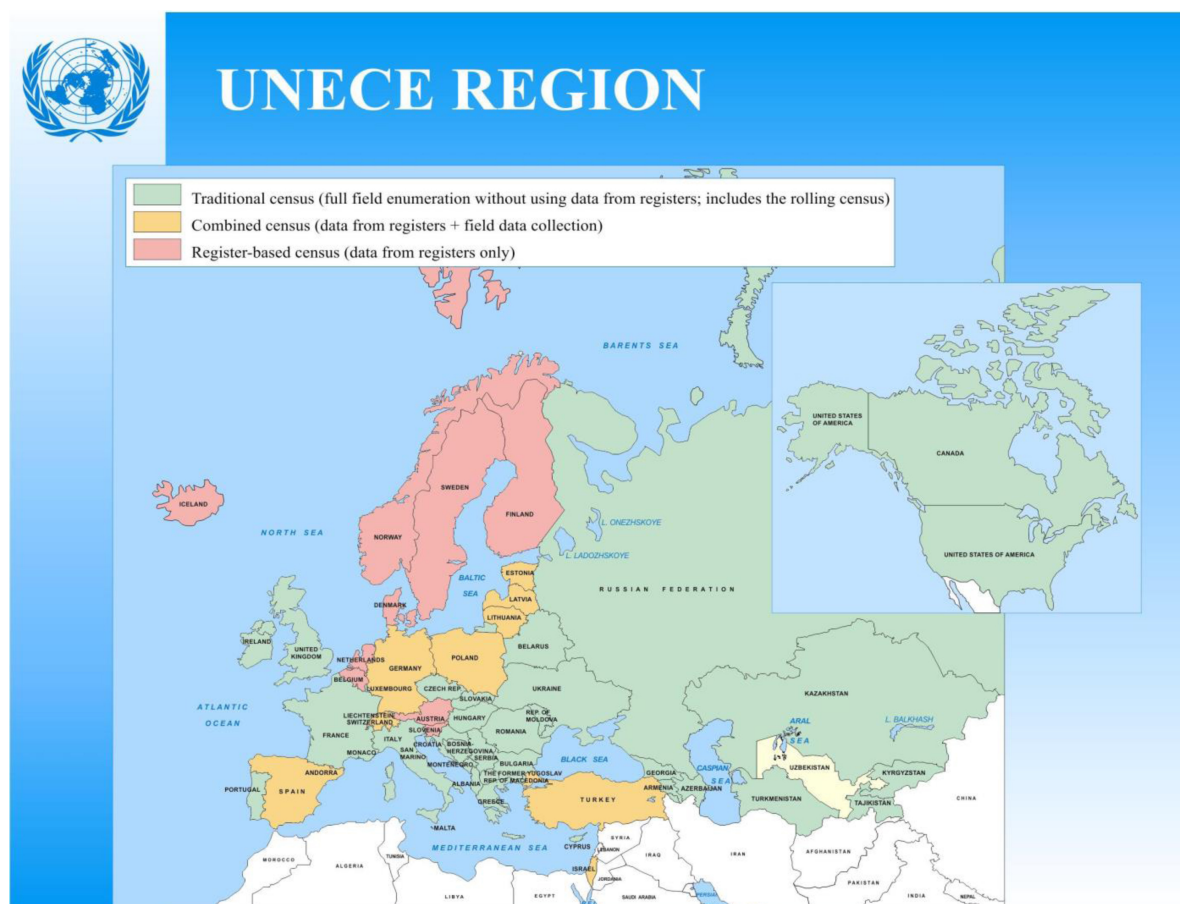


Fig. 3. Census methods in UNECE countries. Source: [10].

ing on registers for their 2011 Population and Housing Censuses, and skipping census questionnaires completely. Some of these countries ‘recycled’ information from their Labour Force Surveys, combining it with register data (Belgium, Iceland and the Netherlands). Lastly, some countries used registers only in this census round (Austria, Denmark, Finland, Norway, Slovenia and Sweden). All nine register-based countries collected census information relating to housing entirely from registers.

2.5. Overview

The map of the UNECE area reveals interesting east-west and north-south tendencies in census methods (see Fig. 3). Three main categories are distinguished on the map: traditional (31 countries), combined (10 countries) and register-based (9 countries). Register-based censuses are becoming increasingly popular in northern Europe, combined censuses are more often

found in central Europe. Traditional censuses continue to be more popular in English-speaking and Commonwealth of Independent States (CIS) countries. All UNECE countries outside Europe conduct traditional censuses. Only Uzbekistan did not conduct a census in this round, and had no plans to do so.

2.6. Essential features of a population and housing census

One may wonder whether register-based censuses still meet the well-known United Nations essential features of a population and housing census. According to the Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing [11] the set of essential features that makes a population and housing census unique is as follows.

2.6.1. Individual enumeration

Information on each enumerated person (and each

set of living quarters) is obtained so that their characteristics can be separately recorded. This allows cross-classifying the various characteristics and obtaining data by more than one characteristic.

2.6.2. Simultaneity

Information obtained on individuals and housing in a census should refer to a well-defined and unique reference period (or specific moment in time). Ideally, data on all individuals and living quarters should be collected simultaneously. However, if data are not collected simultaneously, adjustments should be made so that the reported data have the same reference period.

2.6.3. Universality

The population and housing census should provide data on the total number of persons, households and housing within a precisely defined territory of a country. The counting (or benchmarking) of the population should include every person residing and/or present in the defined territory of a country at a defined singular point in time (commonly referred to as the Census Day). The data on the basic level of enumeration provided by the census should be validated with an independent coverage check.

2.6.4. Small area data

The census should produce data on the number and characteristics of the population and housing related to the smallest geographic areas of the country, and to small population sub-groups, consistent with the overriding requirement to protect individual confidentiality.

2.6.5. Defined periodicity

The census should be taken at regular intervals so that comparable information is made available in a fixed sequence. It is recommended that census data be produced at least every ten years.

An interesting question is to what extent register-based censuses in general and the register-based census in the Netherlands in particular meet the essential features described above. Register-based censuses completely meet the essential features of individual enumeration, universality and defined periodicity. However, the universality in register-based censuses refers to the registered population, not to the usual resident population as in traditional censuses. In the case of the Netherlands, where microdata of the LFS are reused, one can debate whether the essential features of simultaneity and small area data are met. Not all people in the LFS are interviewed on the same day. Therefore,

formally the essential feature of simultaneity is not met for two variables in the Dutch Census 2011. For the other variables the simultaneity criterion is met in the Netherlands. In addition, hypercubes based on LFS information could not be estimated properly for the cells with very few observations. One could thus conclude that formally the essential feature of small area data is not met for all hypercubes in the Netherlands. Register-based censuses in countries where all variables come from registers meet the essential features of simultaneity and small area data.

Statistics Netherlands plans to conduct a 2021 Census conform the Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing [11]. However, for the variable occupation in the Dutch Census 2021 already existing but incomplete microdata will be reused to estimate tables including this variable. The Netherlands will then come even closer to meeting all essential features of a population and housing census. It can thus be concluded that for the Census 2011 Statistics Netherlands met most of the essential features and that the same will hold for the Dutch Census 2021.

2.7. Necessary conditions for a successful transition to a combined or register-based census

If countries want to move from a traditional census approach towards a combined or register-based census, a number of conditions should be fulfilled before information from registers can successfully be used in the census. A number of these necessary conditions can present some challenges, and these are discussed briefly in the following paragraphs.

- Legal base: Censuses need legal underpinning for the chosen methodological approach (traditionally Census Acts had that role, but in the Netherlands nowadays there is one Statistics Act for all official statistics).
- Public approval: This is also important but more difficult to measure (it is clear that people mentioning that in the Census 1971 Statistics Netherlands was acting like ‘Big Brother is watching you’ certainly did not approve, nowadays public approval is not such a problem for the Dutch census).
- Cooperation among authorities: As statistical offices become very dependent on register holders, cooperation (mainly with other government organisations) has become crucial.

- Comprehensive and reliable register system: This helps a great deal when setting up a register-based census (although sometimes the administrative quality is not 100%, the statistical quality is still good enough for the census).
- Unified identification system: Such a system preferably has unique identification numbers.

As a huge interest exists in possibilities for a successful transition to a combined or register-based census, many presentations and courses are given about this topic (see e.g. [12]). In addition, country visits to countries planning a change in census methodology are often helpful to have a successful transition.

3. Compilation methods in the Netherlands

3.1. Introduction

The current census results in the Netherlands refer to 2011. The backbone of the Dutch census is the central population register (PR), which combines all the municipal population registers. PR data for 1 January 2011 were used as the basis for the set of hypercubes. The hypercubes focus on frequency counts, not on quantitative information. Data not available or derivable from the PR were taken from other registers. All register variables are now available from Statistics Netherlands' system of social statistical datasets (SSD), and their quality has been improved by applying micro-integration techniques. More information about the SSD can be found in [13]. Micro-integration entails checking the data and adjusting those that are incorrect. It is widely assumed that micro-integrated data provide more reliable results, as they are based on a maximum amount of information. They also provide better coverage of subpopulations: if data are missing in one source, another source can be used.

3.2. System of base registers

In the past micro-integration played a very important role in the SSD and in particular in censuses. However, in the future the role of micro-integration will be more limited, because in the Netherlands the new so-called system of base registers is better integrated than the set of registers that Statistics Netherlands received before. The system of base registers will ultimately consist of 13 registers on population (residents and non-residents), addresses and buildings, enterprises, real estate (boundaries, ownership, value, etc.), topography

(maps: land, water, roads), motor cars (model, colour, ownership, etc.), taxable income, labour (wages, employers, social benefits, etc.) and subsoil (sewerage, cables, etc.). The system of base registers is based on legislation. For each base register one cabinet minister is responsible. The Minister of Home Affairs is responsible for the system of base registers.

All variables in the system of base registers are kept in only one register. For example, addresses are kept in the base register for addresses and other registers that need address information copy this information exactly from the base register for addresses. The data linkage is through identification numbers. All objects (persons, enterprises, addresses, etc.) have a unique identification number. The use of data from base registers is compulsory for governmental agencies. The objective is that all users of the system contribute to the quality of the data. Therefore, users are obliged to notify the holders of base registers if they have alternative data that are considered to be of better quality. However, due to legal restrictions Statistics Netherlands does not notify the holders of base registers.

Users of base registers can rely on their validity. Statistics based on base registers demand only a limited amount of data editing. As base registers are linked to one another (through identification numbers), it follows that statistical data are also coherent. Contrary to sample survey data, no estimation methods are needed for small area statistics and non-response problems do not exist. However, some data are delayed, for example information about self-employed persons. One should not forget that registers are not developed for statistical purposes and do not always contain the statistical concepts required. In some situations different statistical concepts are acceptable. In other situations one tries to derive information about the required statistical concept. If such derivations are not satisfactorily, one could finally decide to fall back towards sample surveys with the required statistical concept, but with more limited publication possibilities compared to register data.

It is important to manage the base registers properly. Therefore, each base register has its own project board. All groups of stakeholders are represented in these project boards. Project boards operate within the legal framework and see to it that the register data fulfil the legal requirements (quality, completeness, etc.) and that the data are correctly applied. Project boards act as advisory boards to the responsible cabinet ministers and meet a few times per year.

3.3. Survey variables

In the Census 2011, only two variables were not taken from a register: ‘occupation’ and ‘educational attainment’. Records from the Labour Force Survey (LFS) in a three year period around the enumeration date (1 January 2011 was the so-called Census Day) were used to estimate values for these two variables. Table consistency was guaranteed by using repeated weighting for tables including these variables. Repeated weighting is a method for the consistent estimation of multiple frequency tables from registers and sample surveys. The method of repeated weighting, described extensively in [14], is based on the repeated application of the regression estimator, generating a new set of weights for each table estimated. The weights of the records in the microdata are adjusted in such a way that a new table estimate is consistent with all earlier table estimates.

3.4. Repeated weighting

Statistics Netherlands had already used repeated weighting for the compilation of the Census 2001 tables. However, the set of tables for the Census 2011 was larger and much more detailed and therefore the application of the method of repeated weighting to the Census 2011 was not without problems. If cells were covered by only a few (or even no) observations, estimation became problematic. New additional solutions were developed (see Chapter 7 of [8]) and introduced for 23 hypercubes that were to be estimated. In [15] existing solutions for methodological problems are reviewed and new solutions are proposed when necessary. The problems and solutions mentioned are illustrated with experiences of the Dutch Census 2011 compilation. The general message of [15] is that repeated weighting can be applied to very complex estimation problems, although it still has its limitations.

Statistics Netherlands developed and used its own software for applying repeated weighting. Under the condition of small, independent samples, variances of table values can also be estimated. Such estimated variances were used to set publication rules for cells and to calculate variation coefficients for the quality hypercubes, which serve as a quality assessment of the census hypercubes.

To maximise accuracy, all estimates are based on the largest possible number of records. Tables containing only register variables are counted from the registers. Tables with at least one variable from the LFS are esti-

mated from the largest possible combination of register and survey data. Initial weights from the LFS were available for these estimations.

As part of the Census 2011 was compiled based on sample data, margins of inaccuracy had to be taken into account for some results. A rule of thumb was applied for cell values based on a sample from the census population: only estimated table cells based on a minimum number of persons are published. In addition, rare categories have been made confidential to prevent disclosure of individual information.

4. Current and future work

4.1. Introduction

The work on the Census 2011 has been finished at Statistics Netherlands and the Census 2021 project has not yet started. The 2008 European Census Act [3] will also be the legal basis for the European Census 2021. To prepare the European Census 2021 Statistics Netherlands actively joined the discussions on the future of the European censuses based on UN-ECE recommendations [11]. So far, these discussions have already led to three new implementing regulations [16–18]. Moreover, a number of preparatory projects for the Census 2021 have been initiated at Statistics Netherlands that are described briefly in the following paragraphs.

- Methodology: For the variable occupation there will not be a Dutch register with full information on this variable and therefore in the Census 2021 occupation tables have to be estimated as well.
- Software: Linked to the decisions taken how to estimate occupation in the Census 2021, new software has to be developed as the old programme for repeated weighting (that was used in the 2001 and 2011 Censuses) is outdated and can no longer be used.
- Level of education attained: The variable educational attainment will be based on the Educational Attainment File that will be imputed (this work is described in the following sections).
- Housing variables: Changes in some housing variables in the Dutch Housing Register will lead to a few changes in how the Housing Census 2021 will be conducted in the Netherlands.
- Population grid squares: A new phenomenon in the European Population Census 2021 will be information by grid square for which a legal base

has been drafted (see [https://circabc.europa.eu/sd/a/65be2124-27d2-43c4-b36c-9ccfdad9128d/CENS\(2017\)06%20Annex%201%20-%20Census%201%20km%20grid%20data%20-%20draft%20regulation.pdf](https://circabc.europa.eu/sd/a/65be2124-27d2-43c4-b36c-9ccfdad9128d/CENS(2017)06%20Annex%201%20-%20Census%201%20km%20grid%20data%20-%20draft%20regulation.pdf)).

In 2016 and 2017, six countries (Finland, France, Germany, Hungary, Slovenia and the Netherlands) worked together in a European project (see https://ec.europa.eu/eurostat/cros/content/harmonised-protection-census-data_en) in which new methods for the protection of census data were developed that should lead to better comparable safe outputs between countries in the European Census 2021. A choice was made in this project for a combination of the pre-tabular method of data swapping and the post-tabular cell key method in which some random noise is added to the results. In 2018 and 2019 a further European project is foreseen to make the software for this combination of methods available in both R and software packages for Statistical Disclosure Control (e.g. the data swapping method in μ -ARGUS and the cell key method in τ -ARGUS). Then, it will be much easier for statisticians involved in the European Census 2021 to produce safe (according to their national criteria) and well comparable census tables (including the new grid squares).

4.2. Educational Attainment File (EAF)

In this section, some more information is given about the Educational Attainment File (EAF) that will be an important new source for the Dutch Census 2021.

A complex integration process of microdata from the LFS and examination registers has led to a new source for Statistics Netherlands: the EAF. Early versions of the EAF covered a smaller part of the population and missed some kinds of education and therefore they were not considered to be of good enough quality for the Census 2011. However, the new version, that is containing also information on private education institutions, is available since 2016. In this new version about 60 percent of the records have information on highest level of education attained. This huge set of records can be weighted to known marginals of the population for statistics on level of education. In such a way better quality and more detailed education tables can be produced than before when only LFS information was used.

4.3. Imputing the EAF

Statistics Netherlands launched a special project

(subsidised with a European grant) to find out how to impute the EAF for the Census 2021 so that education hypercubes can be produced based on this source. In this section, the different parts of this special project are described.

First, a good imputation model had to be found for the EAF. The choice was made for a logistic regression model. Then a set of hypercubes of the Census 2011 was produced again, now by using the imputed EAF. By comparing the newly produced hypercubes with their official variants of the Census 2011, the differences were studied. These differences are currently (last quarter of 2017) being studied. As about 40 per cent of the records is imputed, it is clear that one cannot make too detailed publications based on the EAF. However, more detailed publications than the official publications of the Census 2011 should be possible. The differences found are used to decide on a criterion how detailed the future Census 2021 education hypercubes based on the imputed EAF will be. Finally, in the first quarter of 2018 a plan, that is possibly also of interest for other countries, will be made for the variable highest level of education attained in the Dutch Census 2021.

4.4. Yearly census results

As mentioned at the end of Chapter 1, in the Netherlands there are no restrictions to official census years as census tables can be updated every year. In this section both the current national yearly tables and the future European plans are described.

Statistics Netherlands publishes over 100 census tables on a yearly basis about persons and households (not about housing) based on the SSD. Some extra variables and specific categories are included in these yearly tables, but it concerns lower dimensional tables than the census hypercubes produced every ten years. Examples of extra variables are:

- Age of the partner (in classes);
- Personal and household income (both in classes);
- Highest level of education attended.

Examples of extra specific categories are:

- The regional categories ‘municipalities around biggest cities’ and ‘postal code’;
- The socio-economic categories ‘employee private sector’, ‘civil servant’ and ‘unfit for work’.

4.5. Future work

The current situation in Europe leads to outdated

census publications (the European Census 2011 results were published at the end of 2014). This has caused users to look for other sources than the census to meet their data requests. Now that these requests can often be met (based on existing register and survey information), census budgets are no longer justified and are shrinking because of government budget cuts.

It is clear that additional publication of yearly tables will make the census information much more useful. The level of detail in such yearly European census tables should be more limited than in the current tables published every ten years. It is important that all countries can participate, either by using registers or by yearly updates (as the United States Census Bureau currently does with its American Community Survey). To make such a European transformation a success, time is needed to come to a change backed by new European legislation. The strategy (see [https://circabc.europa.eu/sd/a/18e61999-517a-45ed-b348-aeb08e44c321/CENS\(2017\)07%20Post-2021%20census%20strategy.pdf](https://circabc.europa.eu/sd/a/18e61999-517a-45ed-b348-aeb08e44c321/CENS(2017)07%20Post-2021%20census%20strategy.pdf)) first includes the European Census 2021 (the results of which will not be published before 2024) and thereafter yearly updates (from 2025 onwards with reference years one year earlier). In the yearly updates current output from censuses, demographic statistics and migration statistics as well as information on small areas could be combined to prevent an extra burden for the Member States of the European Union.

It is clear that a more flexible European legal framework is needed to accommodate future requests by users. After some experiences under the new framework, it will be decided which extra census information will be necessary for the European Census 2031. But first the challenge for the European Union and its Member States is to work out a new and flexible European legal framework to be able to meet future census requests in a timelier manner.

5. Conclusions

The register-based census has proven to be a successful concept in the Netherlands. It has many advantages compared with traditional censuses: costs are considerably lower; problems with non-response only play a role when survey microdata are reused; and the production time is much shorter. These advantages more than make up for the loss of some detail in tables based on survey variables. The Census 2011 provides data on the Netherlands that can be compared to

results of earlier Dutch censuses and to results of other countries taking part in the 2011 Census Round.

Although most countries in the world still conduct traditional censuses, the Netherlands is not the only country with a register-based census. A number of countries in Europe have switched to combined and register-based censuses. The 2011 Census was the fourth that the Netherlands conducted without census questionnaires.

Just as in the 2001 Census, the repeated weighting technique was used successfully to produce a consistent set of tables for the 2011 Census. However, new additional solutions had to be introduced for the hypercubes that were to be estimated. All tables that had to be estimated were based on the largest number of records possible and the resulting hypercubes are mutually consistent. It is important to apply micro-integration of the different sources in the SSD before compiling tables using estimation techniques. The use of micro-integration and the applied estimation techniques guarantee the consistency between table results from different hypercubes. There is thus no confusion for users of census information, as there is one figure on each socio-economic phenomenon, instead of several figures depending on which sources are used.

The challenge for the future is to keep the comparability over time and between countries, but to improve the timeliness of the results. To that aim, a new and flexible European legal framework will be set up to be able to meet future census requests in a timelier manner.

References

- [1] Holvast J. The population census of 1971. Uitgeverij Paris, Zutphen, the Netherlands 2013.
- [2] Corbey P. Exit the population census. Netherlands Official Statistics 1994; 9: 41-44.
- [3] European Commission. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses. Official Journal of the European Union 2008; L218: 14-20.
- [4] European Commission. Commission Regulation (EC) No 1201/2009 of 30 November 2009 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns. Official Journal of the European Union 2009; L329: 29-68.
- [5] European Commission. Commission Regulation (EU) No 519/2010 of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council. Official Journal of the European Union 2010; L151: 1-13.

- [6] European Commission. Commission Regulation (EU) No 1151/2010 of 8 December 2010 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission. Official Journal of the European Union 2010; L324: 1-12.
- [7] Prins K. Population register data, basis for the Netherlands' Population Statistics, Bevolkingstrends 2016|01. Statistics Netherlands, The Hague/Heerlen 2016. <https://www.cbs.nl/nl-nl/achtergrond/2016/01/population-register-data-basis-for-the-netherlands-population-statistics>
- [8] Statistics Netherlands. Dutch Census 2011, Analysis and Methodology. Statistics Netherlands, The Hague/Heerlen, November 2014. <http://www.cbs.nl/en-GB/menu/themas/dossiers/historische-reeksen/publicaties/publicaties/archief/2014/2014-dutch-census-2011-pub.htm?Languageswitch=on>
- [9] United Nations. Measuring Population and Housing. Practices of UNECE countries in the 2010 round of censuses. United Nations, New York and Geneva 2014. http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf
- [10] UNECE Task Force on Census Methodology. Census methodology: Key results of the UNECE Survey on National Census Practices, and first proposals about the CES Recommendations for the 2020 census round. Paper presented at the Fifteenth Meeting of the Group of Experts on Population and Housing Censuses (30 September–3 October 2013, Geneva) by Eric Schulte Nordholt. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2013/census_meeting/3_E_x_15_Aug_WEB_revised_map.pdf
- [11] United Nations. Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing. United Nations, New York and Geneva 2015. <http://www.unece.org/publications/2020recomm.html>
- [12] UNECE Task Force on Register-Based and Combined Censuses. Draft UNECE Guidelines on the use of the registers and administrative data for population and housing censuses. Paper presented at the Nineteenth Meeting of the Group of Experts on Population and Housing Censuses (4–6 October 2017, Geneva) by Eric Schulte Nordholt 2017. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/GE_41_2017_20_ENG.pdf
- [13] Bakker BFM, van Rooijen J, van Toor L. The system of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. Statistical Journal of the IAOS 2014; 30: 411-424.
- [14] Houbiers M, Knottnerus P, Kroese AH, Renssen RH, Snijders V. Estimating consistent table sets: Position paper on repeated weighting. Discussion paper 03005, Statistics Netherlands, Voorburg/Heerlen 2003. <http://www.cbs.nl/NR/rdonlyres/6C31D31C-831F-41E5-8A94-7F321297ADB8/0/discussionpaper03005.pdf>
- [15] Daalmans J. Estimating detailed frequency tables from registers and sample surveys, discussion paper 2015|003. Statistics Netherlands, The Hague/Heerlen 2015. <http://www.cbs.nl:81/NR/rdonlyres/D4769B63-0D25-405D-8134-E11817D642EE/0/201503DPEstimatingdetailedfrequencytablesfromregistersandsamplesurveys.pdf>
- [16] European Commission. Commission Implementing Regulation (EU) 2017/543 of 22 March 2017 laying down rules for the application of Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns. Official Journal of the European Union 2017; L78: 13-58.
- [17] European Commission. Commission Regulation (EU) 2017/712 of 20 April 2017 establishing the reference year and the programme of the statistical data and metadata for population and housing censuses provided for by Regulation (EC) No 763/2008 of the European Parliament and of the Council. Official Journal of the European Union 2017; L105: 1-11.
- [18] European Commission. Commission Implementing Regulation (EU) 2017/881 of 23 May 2017 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission, and amending regulation (EU) No 1151/2010. Official Journal of the European Union 2017; L135: 6-14.