Proceedings of MSP2009
Modernisation of Statistics Production 2009

Data Integration Activities on the Way to the Dutch Virtual Census of 2011[1]

Eric Schulte Nordholt[2]

Data from many different sources were combined to produce the Dutch Census tables of 2001. 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 found an alternative in the Virtual Census, using available registers and surveys. The table results are not only comparable with earlier Dutch Censuses but also with those of other countries.
The acquired experience in dealing with data of various administrative registers for statistical use enabled Statistics Netherlands to develop a Social Statistical Database (SSD), which contains coherent and detailed demographic and socio-economic statistical information on persons and households. The Population Register forms the backbone of the SSD. Sample surveys are still needed for information that is not available from registers.
For the 2011 Census more detailed information is required than for the 2001 Census. Therefore, many data integration activities take place between the Censuses of 2001 and 2011. This paper gives a sketch of all these activities that will help to make the 2011 Census a success.

Keywords: Census; data integration

1. Introduction

In a few years time a new census round will be held in all countries. For all European Union (EU) countries a 2011 Census is mandatory. The way this 2011 Census will be conducted is up to the countries. Since the traditional 1971 Census the Netherlands conduct virtual censuses. This means that census forms no longer exist; the relevant information is collected from already existing registers and surveys that Statistics Netherlands may use for statistical purposes. This way the Virtual Censuses of 1981, 1991 and 2001 were conducted. The Censuses of 1981 and 1991 were of a limited character. The data compiled on 1981 and 1991 were much less detailed than the set of tables of the 2001 Census. Moreover, they were largely based on a register count of the population in combination with the then existing surveys about the labour force and housing conditions. For the 1981 Census a higher than normal sampling fraction of 5 percent was drawn for the Labour Force Survey. The 1991 Dutch Census was largely based on a register count of the population in combination with the Labour Force Survey 1991 and the Housing Demand Survey 1989/1990. Contrary to 1981 and 1991, Statistics Netherlands has published census information for 2001 on the municipal level.
In section 2 the Virtual Census of 2001 is described in more detail. The method of compiling is explained in section 3. A short comparison of the Census results in the Netherlands can be found in section 4. A comparison of the Dutch 2001 Census to Censuses in other countries is given in section 5. Data integration activities between the Censuses of 2001 and 2011 are described in sections 6 and 7. Finally, some conclusions are drawn in section 8.

[1] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

[2] Statistics Netherlands, P.O. Box 24500, 2490 HA The Hague, The Netherlands;
Tel.: +31 70 337 4931; Fax: +31 70 387 7429; E-mail: esle@cbs.nl

2. The Dutch Virtual Census of 2001

In 2003, data were combined to produce the Dutch 2001 Census tables. In the Netherlands, this was done using data that Statistics Netherlands already had available rather than by interviewing inhabitants in a complete enumeration. This way, the Dutch taxpayer received a much lower census bill. The costs for a traditional census would be about three hundred million Euros, while the costs using this method are 'only' about three million. The estimate includes the costs for all preparatory work such as developing a new methodology and accompanying software. The costs of the registers are not included, but the analyses of the results are. Registers are not kept up-to-date for censuses but for other purposes. Saving money on census costs is only possible in countries that have sufficient register information. As an example, we can compare the costs of the Dutch Virtual Census of 2001 with the costs of the traditional Census that was held in Canada. In Canada, the census costs amounted to approximately 450 million Euros. Canada has about 31.6 million inhabitants, twice as many as the Netherlands. A virtual census would be impossible in Canada because of the lack of sufficient register data.

The 2001 Census related to forty extensive tables. Twenty-eight were about the Netherlands as a whole, nine were at the COROP level (NUTS 3) and three at municipal level (NUTS 5). The forty tables fell into a number of groups. Eight tables concerned housing, two tables concerned commuting and the other thirty tables were demographic tables, relating to occupation, level of education and economic activity. Additionally, demographic, housing and labour figures were compiled at sub-city district level for ten large cities that participated in Urban Audit II (Statistics Netherlands, 2003).

Except the financial aspect, other important differences exist between a traditional census and the virtual census conducted in the Netherlands. In spite of the mandatory character of a traditional census, a certain part of the population will not participate (unit non-response) and the part that does participate will not answer certain questions (item non-response). Correcting non-response by weighting and imputation techniques is worth trying. A well-known problem with traditional censuses is that participation is limited and selective. Traditional correction methods fall short of the need to be able to publish reliable results. The last traditional Census in the Netherlands (in 1971) met with much privacy objections against the collection of integral information about the population living in the Netherlands. This increased the non-response problem and the expectation was that non-response would be even higher if another traditional census were held in the Netherlands (Corbey, 1994). There are almost no objections to a virtual census and the non-response problem only plays a role in the surveys of which the data are used. If non-response can be corrected in a survey, it will certainly be possible to correct for the selectivity of that survey in the census where it is used.

The Virtual Census of 2001 in the Netherlands was off to a later start than in other countries where a traditional Census was conducted. It did not make sense to begin the 2001 Census Project until all sources were available; some registers were available relatively late. Nevertheless, the Netherlands was quicker with the compilation of the forty census tables than most of the other countries that participated in the 2000 Census Round. In fact, the Netherlands was one of the first to send the complete set of forty tables to Eurostat, which coordinated the contributions of all European Union (EU) member states, accession countries and European Free Trade Association (EFTA) member states. The Netherlands had the advantage that the incoming census forms did not need to be checked and corrected. However, it must be noted that for some variables only sample information is available, which implies that it was impossible to meet the level of detail required in some Dutch tables.

Currently, the advantages of the virtual census in cost and non-response problems amply make up for the loss of some detail compared to a traditional census. Moreover, not all

required information will consistently be available for the users in traditional censuses. This is because traditional correction methods such as weighting and imputation sometimes do not correct for limited and selective participation. This means no reliable results can be published for some of the cells in the set of tables. One may wonder why simply applying mass imputation (filling in valid values for all missing scores) was not considered to overcome these problems. An important advantage of mass imputation is that once the records are imputed, any user will be able to reproduce results when using the same imputed file. However, mass imputation is not a viable strategy for raising survey outcomes to population totals. There are not enough degrees of freedom to sustain a sufficiently rich imputation model accounting for all significant data patterns between sample and register variables. Only if the interest is in totals of subsets of the population defined by the explanatory variables in the model does the imputation approach lead to (almost) design-unbiased and hence reliable estimates (at least if the variances are reasonably small) (Kroese and Renssen, 2000).

The Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have more variables available in registers than the Netherlands. So the problem of insufficient detail in the outcome does not play a major role there. Moreover, some Nordic countries conducted a (limited) enumeration for variables missing in the registers. Most of the other countries are in a similar position as the Netherlands where some variables relevant for the census can be found in registers, while other variables are available on a sample basis only. That's why much interest exists in the Dutch approach to combine registers and surveys and to use modern statistical techniques and software to compile the tables. It is, of course, crucial that statistical bureaus be able to make use of registers that are relevant for the census. For Statistics Netherlands, this possibility was strengthened in the new statistical law that came into force in the beginning of 2004. Nevertheless, Statistics Netherlands will have to maintain good contacts with register holders. Timely deliveries with relevant variables for Statistics Netherlands are essential for statistical production.

The reason why Statistics Netherlands has compiled the set of tables is a gentlemen's agreement. In 1991 the Census Act was rescinded, officially cancelling Statistics Netherlands' obligation to hold a census once every ten years (Corbey, 1994). There was no European obligation to supply census data of the 2000 Census Round, but it is inconceivable that the Netherlands would not compile census data for the international organisations just like all other European countries did in the 2001 Census Round. Eurostat has a coordinating role in collecting harmonised data on the EU and a duty to make international comparisons of the outcome.

It took several years before all countries participating in the 2000 Census Round had sent their final set of tables to Eurostat. Therefore, Statistics Netherlands took the initiative to compare the 2001 results of a limited number of European countries. The results of the Dutch 2001 Census were also compared to earlier Dutch Censuses. Such work had been carried out before as well.

3. Method of compiling

The last virtual census relates to 2001. The backbone of this census is the central Population Register (PR), which is the combination of all municipal population registers. PR data of 1 January 2001 were used as the basis for the set of tables. The set of tables focuses on frequency counts and not on quantitative information. Different variables, such as occupation and level of education, were obtained from the Labour Force Survey (LFS). The variable job size was obtained from the large Survey on Employment and Earnings (SEE). To obtain sufficient records, information on persons from the LFS 2000 and the LFS 2001 was

combined. For the housing tables, we used PR data of 1 January 2001, the Housing Register 2001 and the Survey on Housing Conditions (SHC) 2000.
Some variables of the PR and Social Statistical Database (SSD) datasets are available on an integral basis. The SSD include integrated microdata on employees and self-employed. Examples of such variables are age, sex, marital and employment status. Survey variables are only available for a part of the population. Examples are the highest level of education attained (LFS) and whether someone rents or owns the property they live in (SHC). We guaranteed consistency among the tables by using the technique of repeated weighting. The method of repeated weighting has been described extensively in Houbiers (2004) and Houbiers et al. (2003). It generates a new set of weights for each estimated table and is based on the repeated application of the regression estimator. The results of five simulation studies testing various aspects of repeated weighting can be found in Van Duin and Snijders (2003). When using repeated weighting, the weights of the records in the microdata are adapted in such a way that a new table estimate is consistent with all earlier table estimates.
To apply the technique of repeated weighting, we used the software package VRD developed by Statistics Netherlands. The letters VRD stand for Vullen (Filling) Reference Database and the aim of the application is to fill and manage the reference database. The main functions of VRD are the estimating of tables via repeated weighting, the addition of these tables to the reference database, and the withdrawal of aggregates from the reference database. Under the condition of small, independent samples, the variances of the table values can also be estimated. The estimating of the tables does not occur in VRD itself, but takes place in Bascula 4.0 automatically without the VRD-user seeing this explicitly. Estimating the tables and the variances can be done in the batch or interactively.
To be able to estimate every table as accurately as possible, each estimate is based on the largest possible number of records. Tables that contain register variables only are counted from the registers. Tables that contain at least one variable from a survey are estimated from the largest possible combination of registers and surveys.
The figures of the 2001 Census relate to persons living in the Netherlands on 1 January 2001 (counting unit persons). The persons who were living in the Netherlands at the beginning of that day according to the PR were ‘counted’ in the Virtual Census. Most of the Dutch population lives in private households, the remainder being part of institutional households. The number of employees in the tables relates to the end of the year 2000 for which 22 December 2000 was used as reference date to fix the number of jobs of employees in the Netherlands. It was impossible to have a reference day in 2001 for the number of employees since the SSD datasets 2001 were not available in time to use in the 2001 Census. The SSD data used registers’ information on the jobs of employees. If an employee holds several jobs at the same time, he or she can appear several times in the employee register. The features of the main job are used in the set of tables. The main job of an employee has been defined as the job with the highest gross wage for the social insurances.
The 2001 Census was compiled partly on the basis of sample data. Therefore, margins of inaccuracy have to be taken into account for some results of the 2001 Census. Because of the reliability of the results, rules of thumb are being applied for cell values that are based on a sample from the census population. The exact margins of inaccuracy cannot be given because of the complex design of the sample surveys used for the Census. The rules of thumb for records of observations from the LFS run as follows:

- Table cells based on less than 10 persons are always suppressed.
- Table cells based on 25 or more persons are always published.
- Table cells based on 10–24 persons are only published if they form a part of a breakdown (by age or sex), in which no cells based on less than 10 persons occur, and at least 50 percent of the cells in the breakdown have more than 25 persons. The threshold of 25

persons corresponds to an estimated relative inaccuracy of at most 20 percent (i.e. the estimated margins amount to 40 percent at most).

The rules of thumb for records from the SHC are of the same form. However, somewhat higher threshold values are applied because the sample size of the SHC is somewhat more limited than the one of the LFS. For table cells with households or dwellings as counting unit, analogous rules of thumb are applied for the Dutch Census of 2001.

4. The 2001 Census compared to earlier Dutch censuses

The first Census in the Netherlands was held in 1795 for the purpose of establishing voting constituencies. At that time, the united provinces of the Netherlands were still a republic and the borders were different from the current borders. After Napoleon, the Netherlands became a kingdom and once every ten years a census was held. The first Census in the Kingdom of the Netherlands was held in 1829. Before Statistics Netherlands was established, another six Censuses were held in 1839, 1849, 1859, 1869, 1879 and 1889 under the responsibility of the Ministry of the Interior. In 1899, Statistics Netherlands was established and was put directly in charge of the eighth Census. In the 20th century six more traditional Censuses were carried out in 1909, 1920, 1930, 1947, 1960 and 1971. The three most recent Censuses (1981, 1991 and 2001) were not based on a complete enumeration but on registers and surveys available to Statistics Netherlands.

Originally, the censuses had two aims. First, they were meant to correct errors in the municipal population registers. Second, they were used to obtain extra information about the socio-economic phenomena in the country. Since the Netherlands conducts a register-based census, the first aim no longer exists. Also, the quality of the central Population Register (PR), which unites all municipality population registers, has improved considerably over time. This is because the incentive for municipalities to keep their population registers up-to-date is the allocation of central government funds among municipalities, which is generally based on the population size according to the local registers. Another reason is that it is extremely difficult to live in Dutch society without being included in the PR. So both municipalities and citizens have enough incentives to keep the PR of good quality. Recent actions in Rotterdam to improve the quality of the municipal population register for some old quarters prove this statement. The second aim is still valid and many census results are published in a historical or international context. Currently, census data are popular for comparisons among countries.

Table 1 presents some key results of the Dutch Censuses in the period 1829-2001. The ageing of the Dutch population is worth noting, especially in the post-war period.

Table 1. Population by age group in the period 1829-2001

Census		*All ages*	*Age group*		
Number	*Year*		*0-19*	*20-64*	*65+*
		× 1,000	*in % of the total population*		
1	*1829*	2,613.3	44	50	5
2	*1839*	2,860.6	45	50	5
3	*1849*	3,056.9	43	53	5
4	*1859*	3,309.1	42	53	5
5	*1869*	3,579.5	43	52	6
6	*1879*	4,012.7	44	50	5
7	*1889*	4,511.4	45	49	6
8	*1899*	5,104.1	44	50	6
9	*1909*	5,858.2	44	50	6
10	*1920*	6,865.3	42	52	6
11	*1930*	7,935.6	40	54	6
12	*1947*	9,625.5	38	55	7
13	*1960*	11,462.0	39	53	9
14	*1971*	13,060.1	36	54	10
15	*1981*	14,216.9	31	57	12
16	*1991*	15,070.0	25	62	13
17	*2001*	15,985.5	24	62	14

5. The Dutch 2001 Census compared to other countries

More than fifty countries in the United Nations Economic Commission for Europe region participated in the 2000 Census Round. Many countries chose a day in 2001 as their reference day, although they chose many different days. As it takes a long time before all countries finish the tables required by the international organisations, the Netherlands took the initiative to make some simple comparisons among nine European countries that were relatively quick in compiling the set of tables for Eurostat and that were willing to join the comparison analyses.

The calculations in this paper are the author's own and are based on the set of standard tables produced from census data for Eurostat by nine different countries. It is expected that there are definitional differences among the countries which will affect comparisons. Also, the statistics produced by the author do not necessarily reflect the way countries usually choose to present their data. The nine countries are the Netherlands (NL), Norway (NO), Sweden (SE), Finland (FI), Estonia (EE), Switzerland (CH), Slovenia (SI), Greece (GR) and the United Kingdom (UK). The nine countries differ in size, but all except the United Kingdom have a fairly small number of inhabitants compared to France and Germany.

The nine countries are members of the European Union (EU) or the European Free Trade Association (EFTA). The Netherlands joined the European Community at the start in 1958, the United Kingdom joined in 1973 and Greece in 1981. The European Community became the European Union in 1995 when Sweden and Finland joined. Estonia and Slovenia joined the EU in 2004. Norway and Switzerland are EFTA members and work closely together with the EU countries. Norway is also a member of the European Economic Area (EEA). The EEA agreement came into force on 1 January 1994. EEA countries are the EU 15, Norway, Iceland and Liechtenstein. Switzerland did not join the EEA, but works together with the EU countries on a bilateral basis. Statistics is one of the issues on which the EEA countries work together. The aim of the statistical co-operation in the EEA is to build a European Statistical

System that gives a coherent and comparable description of the economic, social and environmental developments in the EEA countries.
The nine countries that are compared have different census reference dates: 31 March 2000 (Estonia), 5 December 2000 (Switzerland), 1 January 2001 (The Netherlands, Sweden and Finland), 18 March 2001 (Greece), 29 April 2001 (United Kingdom), 3 November 2001 (Norway) and 31 March 2002 (Slovenia).
Table 2 presents the estimated costs of the Censuses in the 2000 Round, and the population and area of the nine countries. Estonia, Slovenia, Greece and the United Kingdom held traditional censuses; Switzerland used a combination of a traditional census and register information to produce the census tables. Norway relied largely on registers, but conducted a census for some missing housing variables. Sweden and Finland held entirely register-based censuses and the Netherlands performed a virtual census based on existing registers and surveys. The Census costs for Norway, Estonia, Switzerland, Slovenia, Greece and the United Kingdom include enumeration costs. In the Netherlands, Sweden and Finland such enumeration costs do not exist for their 2001 Censuses, so the costs presented in Table 2 for these three countries are rough indicators of the extra costs of producing census tables for the international organisations and of analysing and publishing the results. Table 2 shows that the costs per inhabitant in those countries that required completion of a census form for the census were much higher than the countries that did not have enumeration costs. In Table 2, the population densities among the nine countries can be compared. The Netherlands has the highest population density, followed by the United Kingdom and Switzerland. The population density in the Nordic countries (Norway, Sweden and Finland) and in Estonia is relatively low. Slovenia and Greece occupy a middle position.

Table 2. Comparison of nine countries according to the Census results in the 2000 Round

	NL	NO	SE	FI	EE	CH	SI	GR	UK
Cost of the Census (in millions of Euros)	**3.0**	**14.6**	**1.0**	**0.8**	**10.2**	**99.1**	**8.0**	**49.7**	**367.4**
Population (× 1,000,000)	**16.0**	**4.5**	**8.9**	**5.2**	**1.4**	**7.3**	**2.0**	**10.9**	**58.8**
Area (× 1,000 km^2)	**41.5**	**323.9**	**450.0**	**338.1**	**45.1**	**41.3**	**20.3**	**132.0**	**244.1**
Cost of the Census per inhabitant (in Euros)	**0.2**	**3.2**	**0.1**	**0.2**	**7.3**	**13.6**	**4.0**	**4.6**	**6.2**
Population density (persons per km^2)	**386**	**14**	**20**	**15**	**31**	**177**	**99**	**83**	**241**

6. Data integration activities between the 2001 Census and 2008

At the end of 2003 the complete set of forty 2001 Census tables for the Netherlands was sent to Eurostat. The set of forty standard tables for the Netherlands (in Excel format) can be found at page http://www.cbs.nl/nl-NL/menu/themas/dossiers/volkstellingen/cijfers/incidenteel/maatwerk/2003-volkstelling-excel.htm and the table annotations at page http://www.cbs.nl/NR/rdonlyres/D8D55875-0630-492F-8125-BA71D7608009/0/tableannotationsNLcensus2001.pdf.
A book about the Dutch Virtual Census of 2001 was written afterwards (Schulte Nordholt, Hartgers and Gircour, 2004). This book provides a wide-ranging description of the socio-demographic and socio-economic state of the Netherlands based on the 2001 Census results. It discusses differences in size and composition among households, economic activity of households, individual activity status by region, age, education level and branch of economic

activity. There are separate chapters on the economic activities of young people and people of retirement age. The economic activities, levels of education and occupation of foreigners from various countries of origin are compared with each other and with the native Dutch population. Regional aspects are also examined, including commuting. The results of the 2001 Census are compared with the Census results of some other European countries and with earlier Dutch Censuses. Lastly, the Virtual Census methodology used is described in some detail.

The PDF version of the book can be found at the Statistics Netherlands website, at page http://www.cbs.nl/NR/rdonlyres/D1716A60-0D13-4281-BED6-3607514888AD/0/b572001.pdf. An extra Chapter (number 15) is available at page http://www.cbs.nl/NR/rdonlyres/7A45A707-D4F6-4F23-92E5-130C5BC1A144/0/b572001hoofdstuk15.pdf with an overview of the used data sources, methods and definitions. Hard copies of the book were sent to all authors of the book, to the management of Statistics Netherlands and to several libraries. The book was also offered to the Prime Minister, the Minister of Economic Affairs and the Minister of Education, Cultural Affairs and Science of the Netherlands and to Director-Generals of statistical offices in several countries. In August 2004, the book was publicly released at an official presentation in the Statistics Netherlands' office in Voorburg. The research process and the main findings were then presented to an audience of academics, press representatives, government officials, as well as Statistics Netherlands' employees. Several articles were written in national and regional newspapers about the Dutch Virtual Census of 2001 and its results. Announcements and interviews appeared in several mailing lists, newsletters and journals. Also articles in international journals and books were published as e.g. Schulte Nordholt (2005), Schulte Nordholt and Linder (2007) and Schulte Nordholt (2008).

Protected 1 percent samples of the microdata of the Dutch Censuses of 1960, 1971 and 2001 were produced in 2005 and disseminated via the IPUMS (Integrated Public Use Microdata Series) project, see http://www.ipums.org/international. These micro datasets contain a number of demographic and economic variables and can also be analysed via the institute DANS (Data Archiving and Networked Services), see http://www.dans.knaw.nl/en/. Bona fide researchers who want to make more detailed studies on these three censuses can work on-site at the premises of Statistics Netherlands. More information about this last option can be obtained via Statistics Netherlands' Centre for Policy Related Statistics.

In the last couple of years many lectures were given about the Dutch Virtual Census approach at conferences, universities, research institutes and statistical institutes. Examples are the ESTP (European Statistical Training Programme) courses given yearly at Statistics Norway in Oslo since 2006 and the 2007 International Statistical Seminar Eustat in Bilbao (Schulte Nordholt, 2007)

A lot of information on all censuses in the Netherlands can be found at the bilingual (Dutch and English) website http://www.volkstellingen.nl. Here also information about the digitalizing of old census results can be found. Many international census meetings have led to recommendations for the next census round (e.g. United Nations, 2006) and books about register-based statistics like United Nations (2007) and Wallgren and Wallgren (2007).

In 2007 and 2008 the so-called CENEX (CENtre of EXcellence) on ISAD (Integration of Surveys and Administrative Data) (http://cenex-isad.istat.it) was running. In this project, that was sponsored by the European Union, the statistical offices of Austria, Czech Republic, Italy, the Netherlands and Spain were closely collaborating on probabilistic record linkage, statistical matching and micro integration processing. The state of the art in modern literature was described and practical examples were given. A very interesting practical example was the description of the results of the test that Statistics Austria executed for the 2006 Census. Herewith Austria proved to be able to move to a register-based census in the next census

round. A survey was conducted on the use of integration methodologies in the different ESS (European Statistical System) countries. A course on integration of surveys and administrative data was given in Hungary in November 2007. The final workshop on integration of surveys and administrative data was held in Austria in May 2008.
The Census of 2001 was based on a gentlemen's agreement, a European regulation for that Census did not exist. After long European negotiations this situation will be different for the 2011 Census. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses makes the 2011 Census an obligation for all EEA (European Economic Area) countries. This regulation defines the reference population, but leaves the countries free to decide on the data sources to use. In addition to the data delivery now also a quality report is mandatory. The census topics for the 2011 Census have been defined in the regulation, but final decisions have not yet been made about the hypercubes (detailed tables) to produce and level of detail of the variables in these hypercubes. However, it is already clear that the total number of cells in all 2011 Census tables will be much larger than that total number for the 2001 Census.

7. Data integration activities between 2009 and the 2011 Census

Statistics Netherlands foresees the main Census 2011 project in the years 2012-2014. To make this project a success four projects have to be conducted in the years 2009-2011. In the following four subsections these four projects are described in more detail.

7.1 Sources

It is clear that the Population Register (PR) will again be the backbone of the census. Information from other registers and surveys has to be added to be able to derive all 2011 Census variables. We have to realize that registers change over time. Also the quality of registers may change over time. Some registers like the new Housing Register were not yet available for statistics in the 2001 Census, but will be available in the 2011 Census project. Possibly the information from the new Housing Register can replace the census information from both the old Housing Register and the Survey on Housing Conditions (SHC). The fiscal and social security registers in the Netherlands have been merged since the last Census of 2001. This merging process led to reorganisations of government organisations and took much longer than planned. However, the hope is that this new combined register will be a good replacement for the former Survey on Employment and Earnings (SEE). In addition to register information for some variables information of the Labour Force Survey (LFS) will be essential for the success of the 2011 Census. In this project the current state of the different sources has to be checked on the consequences for the main Census 2011 project. European requirements have to be confronted with available sources.

7.2 Estimation method

In the ideal situation the software for repeated weighting is applied to all relevant microdata and that way all hypercubes are estimated consistently. However, we do not yet know if the much larger and more detailed set of 2011 tables can be estimated with this software in the same way as we estimated the 2001 tables. It is clear that first a number of tests have to be conducted. If there are practical problems it should be studied whether we can adapt the software or that we have to rely on another method than repeated weighting. A second best choice could be to simply weight the results to the PR only. However, this does not lead to a set of completely consistent hypercubes. This means that in that case we have to look carefully to the weighting model.

Finally, the question is if we have to impute some variables. Information about cohabiting couples is not fully derivable from the PR and thus has to be imputed. As more advanced derivation programmes are nowadays used, the percentage of imputed records will be smaller than during the 2001 Census.

7.3 Statistical Disclosure Control of the hypercubes

On 9 July 2008 the European Parliament and the Council adopted the Regulation (EC) No 763/2008 on population and housing censuses (Census regulation). The regulation is output oriented, i.e. it is open to the use of different data sources, but requires the respect of the essential features of population and housing censuses, the use of harmonized definitions, technical specifications, topics and breakdowns. The Census regulation foresees unified reporting years (the first being 2011), a common EU dissemination programme, technical standards for the data transmission and the establishment of quality reports for European purposes. Concerning the statistical confidentiality, the following aspects are of particular importance:

- Article 4 (2) foresees that the "Member States shall take all measures necessary to meet the requirements of data protection. The Member States' own data protection provisions shall not be affected by this regulation." That means that the protection of census data comes under the responsibility of the Member States, and has to be done at their level rather than by the Commission. Article 4 (2) provides further that the European Commission is not entitled to issue legislation on the disclosure protection of census data on the basis of the Census regulation. However, Article 6 (4) stipulates "The Commission (Eurostat), in cooperation with the competent authorities of the Member States, shall provide methodological recommendations designed to ensure the quality of the data and metadata produced, acknowledging, in particular, the Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing." (see also United Nations (2006)). Consideration 3 to the regulation explains that "in view of methodological and technological developments, best practices should be identified and the enhancement of the data sources and methodologies used for censuses in the Member States should be fostered."
- Article 5 (2) of the Census regulation foresees that the "Member States shall provide the Commission (Eurostat) with final, validated and aggregated data (…)". This excludes the transmission of microdata. Although aggregated data are not necessarily protected against disclosure of sensitive data, the spirit of Article 5 (2) implies that no confidential data shall be transmitted to Eurostat.
- Considerations 5 and 7 stipulate that the Statistical Law, respectively the European Statistics Code of Practice, constitute the framework for the Census regulation, both containing provisions on statistical confidentiality.
- Consideration 6 recalls the regulations on the transmission of data subject to statistical confidentiality. This means that, if Member States transmit data they feel is subject to statistical confidentiality, Eurostat has to ensure the physical and logical protection and that no unlawful disclosure or non-statistical use occurs when Community statistics are produced and disseminated. However, the census regulation does not foresee the transmission of confidential census data from the Member States to Eurostat. In a broad sense, Consideration 6 reminds indirectly that everything must be done to avoid inadvertent disclosure of any confidential data.

A Task Force on the EU Methodology for Census Data Disclosure Control was set up at the end of 2008 to identify and resolve areas of difficulty relating to the confidentiality data treatment of population and housing census data, adopting or developing a harmonised methodology which respects the national regulations. Eurostat and six countries participate in

this Task Force: Estonia, Germany, Italy, Portugal, the Netherlands and the United Kingdom. In principle, the Task Force follows up two major branches of thinking:

- A recommendation on the pre-tabulation noise protection at the microdata level. This seems to have advantages in the context of both a national and a European dissemination of 2011 Census results. However, this protection can only be done at the NSI (National Statistical Institute) level and Eurostat would have no means of even verifying that such a protection has been executed.
- A recommendation on post-tabulation protection (hypercube level). For the time being, the work is split into "cell suppression" and "post-tabulation noise protection". A simple solution would be to check which cells cannot be published (the so-called primary suppressions) and protect in addition a number of cells to prevent recalculations from the margins (the so-called secondary suppressions). However, the Task Force might also consider whether synergies between these two methodologies are achievable — given that the objective is limited to preventing the identification of individuals, i.e. to prevent certainty about cell values in frequency tables. This prevention action should ideally take place with minimum information loss.

The Task Force will report to the Working Group on Demography and Census in 2010.

7.4 Web service

All NSIs are now themselves responsible for the census data transmission to Eurostat. A set of Excel sheets that has to be filled in for all relevant regional levels no longer exists. It has become a national decision how to store all 2011 Census hypercubes. Irrespective of this decision a web service (programme) must be built to collect the relevant information from the national database and send it to Eurostat in the prescribed SDMX (Statistical Data and Metadata Exchange) format.

Statistics Netherlands makes all tabular output available via the Statline database. From this perspective it would be logical to add all 2011 Census hypercubes to Statline. Then the web service should find the 2011 hypercubes and bring the information to Eurostat. There is in addition some experience with data transmission in SDMX format.

However, the 2011 Census hypercubes are much larger than the Statline tables with which Statistics Netherlands has SDMX experience. It would simply be too costly to rebuild Statline and add all Census tables. The 2001 Census tables contained about 2 million cells and were too large for Statline. Therefore, the 2001 Census tables were made available in Excel format via the Statistics Netherlands website www.cbs.nl. As the Census 2011 hypercubes contain approximately 10 million cells and Excel is no longer an option for Eurostat, the best alternative is to build a separate database for the 2011 Census. An extra argument to do that is that the number of 2011 Census tables is higher than the number of tables currently available in Statline. This implies that the Census 2011 would dominate Statline and that is unwanted for a virtual census.

The web service to be built has to operate in a European infrastructure with guaranteed interfaces, response and availability. The web service has to be tested thoroughly and both the new database and the service have to be hosted. The costs involved are recurring yearly as long as the service is active.

8. Conclusions

The virtual census has proved to be a successful concept in the Netherlands. It has many advantages compared to traditional censuses. The costs are now considerably lower. Nevertheless, data on the Netherlands have become available that could be compared to

results of earlier Dutch censuses and to the results of other countries that took part in the 2000 Census Round. It was the third time that the Netherlands conducted a virtual census. However, the Dutch data that have been compiled for 1981 and 1991 were of a much more limited character than the set of tables of the 2001 Census. 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. Also for the Virtual Census of 2011 it is important that the final results are comparable both over time and with other countries

The technique of repeated weighting has been used successfully to produce a consistent set of tables for the 2001 Census. Before compiling tables with this new technique, micro-integration of the different sources in the SSD remains important. In the micro-integration process, the data are checked and incorrect data are adapted. It is strongly believed that micro-integrated data will provide more reliable results, because they are based on a maximum amount of information. Also the coverage of subpopulations will be better, because when data are missing in one source, another source can be used. Another advantage of micro-integration and repeated weighting is that there is no reason for confusion among users of statistical information: there will be one figure for each socio-economic phenomenon, instead of several figures depending on which sources have been used.

It is possible to conduct a register-based census in more and more countries. However, first it should be possible to use registers for statistical purposes. In most countries, not all census variables can be derived from register information. Additional surveying then remains a necessity, but a consistent set of census tables can be produced using the technique of repeated weighting. If the larger 2011 set of Census hypercubes can be produced by using repeated weighting still has to be found out.

It is not always easy to get attention for a virtual census, especially if the last traditional census was held a long time ago. It is a challenge to keep the knowledge up-to-date and the software running if a census is only conducted every ten years. However, the enormous international interest is heart warming. In the coming years there is a lot of interesting work to do to prepare the Virtual Census of 2011.

References

Corbey, P. (1994). Exit the population Census. Netherlands Official Statistics, Volume 9, summer 1994, 41-44.

Duin, C. van and V. Snijders (2003). Simulation studies of repeated weighting. Discussion paper 03008, Statistics Netherlands, Voorburg / Heerlen. http://www.cbs.nl/NR/rdonlyres/203C85C6-7075-47A0-97BA-A3B748D393FE/0//Discussionpaper03008.pdf.

Houbiers, M. (2004). Towards a social statistical database and unified estimates at Statistics Netherlands. Journal of Official Statistics, Volume 20, No. 1, 55-75.

Houbiers, M., P. Knottnerus, A.H. Kroese, R.H. Renssen and V. Snijders (2003). Estimating consistent table sets: position paper on repeated weighting. Discussion paper 03005, Statistics Netherlands, Voorburg / Heerlen. http://www.cbs.nl/NR/rdonlyres/6C31D31C-831F-41E5-8A94-7F321297ADB8/0/discussionpaper03005.pdf.

Kroese, A.H. and R. H. Renssen (2000). New applications of old weighting techniques, constructing a consistent set of estimates based on data from different sources. ICES II, Proceedings of the second international conference on establishment surveys, survey methods

for businesses, farms, and institutions, invited papers, June 17-21, 2000, Buffalo, New York, American Statistical Association, Alexandria, Virginia, United States, 831-840.

Schulte Nordholt, E. (2005). The Dutch virtual Census 2001: A new approach by combining different sources. Statistical Journal of the United Nations Economic Commission for Europe, Volume 22, Number 1, 2005, 25-37.

Schulte Nordholt, E. (2007). The Dutch Virtual Census: Combining data from registers and sample surveys, Eustat, Vitoria-Gasteiz, 2007, 82 pages. (http://www.eustat.es/prodserv/vol47_i.html).

Schulte Nordholt, E. (2008). The Dutch Virtual Census of 2001: A register-based approach combined with survey information. In: International Seminar on the Use of Administrative Data for Economic Statistics and Register-based Population and Housing Census, 19~20 May 2008, Daejeon Convention Centre, Daejeon, Korea, Korean National Statistical Office, 197-214 (including presentation).

Schulte Nordholt, E., M. Hartgers and R. Gircour (Eds.) (2004). The Dutch Virtual Census of 2001, Analysis and Methodology, Statistics Netherlands, Voorburg / Heerlen, July, 2004. http://www.cbs.nl/NR/rdonlyres/D1716A60-0D13-4281-BED6-3607514888AD/0/b572001.pdf.

Schulte Nordholt, E. and F. Linder (2007). Record matching for Census purposes in the Netherlands. In: Statistical Journal of the IAOS, 24, 2007, 163-171.

Statistics Netherlands (2003). Urban Audit II, the implementation in the Netherlands. Report, BPA no. 2192-03-SAV/II, Statistics Netherlands, Voorburg. http://www.cbs.nl/NR/rdonlyres/8C6E4C9D-4338-4E32-848B-8D43B9B3242D/0/urbanauditIINetherlands.pdf.

United Nations (2006). Conference of European Statisticians recommendations for the 2010 Census of population and housing. United Nations, New York and Geneva.

United Nations (2007). Register-based statistics in the Nordic countries: review of best practices with focus on population and social statistics. United Nations, New York and Geneva.

Wallgren, A. and B. Wallgren (2007). Register-based statistics – administrative data for statistical purposes. Wiley, New York, United States.