From Census To Integrated Population Data Or Socio-Demographic Accounts

by Mathieu Vliegenl
Netherlands Central Bureau of Statistics

Introduction

The census of population taken in the Netherlands in 1971 seems to be the last one in a series which started in 1830. In 1981 the government postponed the population census which according to the 1970 Census Law had to be taken in that year. Meanwhile, the government has announced to Parliament a proposal to revoke the 1970 Census Law. At the same time the government submitted an alternative statistical programme consisting of a set of register-based enumerations in combination with survey research during a period of circa ten years. First reactions indicate that Parliament will be in favour of such revocation.

In several publications attention has been paid to some underlying factors with regard to this development (Redfern 1986, Choldin 1987). Relevant for the topic under study here - post-censal surveys - are the public and parliamentary discussions on the privacy issue in relation to a census, in which pleas in favour of an absolute anonymity at taking a census are a central topic: data collection should take place without names and addresses. In this way - it was the argument - the privacy of the individual citizen would be guaranteed. Furthermore, it is worthwhile reminding the demands in these discussions for a voluntary participation in the census by the individual citizen. Every legal obligation in respect to this participation should be rejected, in particular a legally imposed penalty for not cooperating. Such obligations - the argument was - would infringe the fundamental rights of the individual citizen with respect to his willingness to give information about himself to others.

Ultimately those pleas and demands were not yet effective on the 1971 census as such. However, they affected the possibility laid down in the executive regulations of the 1971 census with regard to the keeping of a 10% sample from that census. By linking this sample to the 1981 census it was intended to obtain longitudinal data on among others occupational and educational mobility as well as on changes in household status. To carry out this longitudinal study it should be necessary to keep the names and addresses of the sampled persons during a period of more than ten years.

More and more, however, this procedure was considered in public opinion and in Parliament as an encroachment on the personal privacy of the citizen. Although the importance of such longitudinal research for statistical purposes was recognized by Parliament, utlimately it was stated that this kind of research connected with a census had to be subordinated to the interests of the individual citizen. Consequently, a few weeks before census date, the Minister politically in charge of the 1971 census was obliged to cancel a possible keeping a 10% sample from the 1971 executive regulations.

Since the 1971 census various sources and methods have been come into use for the construction of a system of population statistics as comprehensive as possible from a demographic as well as from a social and economic point of view (Vliegen and van de Stadt, 1988). The following developments should be mentioned specifically:

- starting enumerations from the municipal population registers and their respective enlargements in content;

- more extensive exploitation of the municipal reporting with regard to vital events and changes of residence;

- conducting regularly large-scale sample surveys on the labour and housing market;

- application of methods to obtain estimates at the level of the total population;

- the accomplishment of an automated (and yearly updated) register of all addresses in the Netherlands with geocodes, a so-called Geographic Base File (GBF), which - jointly with the municipal population registers - also is in use as sampling frame.

These developments have brought a statistical programme into practice which - in view of the results produced - is highly comparable to that of a conventional census in combination with post-censal surveys. Some parts of this programme show characteristics analogous to a conventional census. In particular this regards the system of demographic statistics which is based on periodical enumerations from the population registers in combination with the processing of monthly municipal reports on vital events and migrations.

To a certain extent this applies also to the large-scale sample surveys which generate benchmark data in the

social and economic field. These surveys have to fulfill this function, since the content of the municipal population registers is restricted to purely demographic data. Of course, in delivering the social and economic benchmarks they cannot completely compete with a census. For example, they cannot provide these data with the same regional detail as a census does. However, the extent of regional detail is sufficient for the level of which national policies regarding the labour and housing market are made.

At the same time, these large-scale sample surveys show characteristics inherent to so-called post-censal surveys. First of all, they provide in-depth information on some specific groups of the population. Secondly, the results of the system of demographic statistics (comparable to the results of a census in the classical sense) are used for making the relevant estimations on the level of the total population from the survey results. Finally, the sources for compiling the demographic statistics (the municipal population registers) themselves are sometimes used as frame for the selection of the sampling units.

These points will be discussed more deeply in the next sections. It should be pointed out, however, that in the system of population statistics built up until now, no use has been made of the technique of record linkage. In the near future the application of this technique will presumably not be used either. Reasons of public and political nature - rather than technical impossibilities - prevent such applications at present. Therefore, the above sketched statistical programme cannot be qualified as a register-based census (Redfern, 1986). Rather it is an approach in which both kind of statistical instruments are jointly applied in such a way that (a) societal changes relative to the field of inquiry in question can be discerned almost continuously by the statistical information provided, and (b) actual developments can be taken as a subject of inquiry in the research programme almost instantly.

The system of demographic statistics

2.1. The continuous population accounting: the basis of the system

The main source for compiling demographic statistics are the municipal population registers. These registers have been introduced in 1850 and set up using the data collected at the population census of 1849. Since then these registers have been continuously updated according to the regulations laid down in the system of population accounting (van den Brekel, 1977). Until World War II the registers consisted of family documents in which all members of the family were listed. Since then the personal card has been introduced. This personal card is made out at birth and follows the individual person during his whole life time.

An essential feature of the population registration system is its decentralization. This implies that each municipality keeps its own population register. Persons are registered in the population register of the municipality in which they normally reside.

The regulations for the systematical updating of the municipal population registers refer among others to the registration of all changes by birth, death, marriage and dissolution of marriage by the local Registrar of the civil registration to whom they have to be reported. Furthermore they consist of detailed rules with respect to taking permanent residence in and removals from a municipality as well as all changes of residence within a municipality (see figure 1). All these regulations are intended to guarantee the completeness and accuracy of the municipal population registers.

2.2. System of demographic statistics and register-based enumerations.

The system of population accounting also includes regulations concerning the municipal reporting of the various vital events and changes of address to the Netherlands Central Bureau of Statistics (see also figure 1). This reporting enables the CGS to compile continuously statistics on natality, mortality and nuptiality as well as statistics on internal and external migration.

The municipal information on vital events and migrations is also used at the CBS for updating a statistical file with aggregated data on the demographic composition of the population. This file - set up for the first time using the 1947 census data - enables the CBS to compile annually statistics on the size and composition of the population for every municipality.

The file with data on the demographic composition of the population has to be revised regularly. In the course of time deviations from the real situation are inevitably introduced due to the above mentioned method of updating this file. Consequently, the relevant statistics are becoming less reliable over time.

The revision of the 1947 file took place in 1960, still based on the results of the census taken in that year. Since 1971, however, complete enumerations from the municipal population registers are used for revising purposes. The underlying factors for this switch were among others satisfactory results from checks on the quality of the data in the municipal population registers obtained at the 1971 census as well as technological developments in data processing.

At every enumeration the amount of characteristics in the file has been augmented (see figure 2). At present by means of this file statistical information is supplied annually for each municipality on e.g. the total population by age, sex and marital status, and the alien population by age, sex and marital status.

2.3 The system of population statistics: its reliability

The basic demographic data in the municipal population registers have a high degree of accuracy. It is in the interest of the citizen that his data have been accurately recorded in the municipal population register in view of, among others, a request regarding a resident permit, a

driving licence, a passport, as well as several benefits or grants. Moreover, it is in the interest of the municipalities that the data of its citizen are accurately registered. The financial contribution of the central government to the municipalities depends on, among others, the number of its inhabitants. By law, these figures have to be determined annually by the CBS after mutual control with the municipalities. The procedure has been laid down in the above mentioned system of population accounting.

Therefore, the reliability of the demographic statistics is in general high. This can also be concluded from the results of the register-based enumerations carried out for revising purposes (for figures see table 1). The deviations found from the confrontation of the results from these enumerations with those from the yearly updated statistical file are usually very small for various age groups. They are higher for categories of marital status and for the alien population, preponderantly due to imcompletenesses in the reporting of the relevant changes to the CBS.

2.4. Register-based demographic statistics: some prospects

In addition to the data used in register-based enumerations for revising purposes, the municipal population registers contain still other data which from a statistical point of view are of great value. Up till now these data were not a topic for regular register-based enumerations due to the lack of proper automated processing systems with regard to these registers. Until recently, some municipalities had set up duplicates of their registers in an automated form, but the systems developed were for a great part different from each other. Other municipalities had duplicates of their population register in a mechanised form: either punch-cards or address-plates. Still other municipalities (mostly the smaller ones) had no duplicates at all.

At present an automated system of municipal population accounting - called the Municipal Administration of the Population (MAP) - is developed under the supervision of the Ministry of Home Affairs. Not only the municipal population registers itself is taken into regard, but also the reporting of the relevant changes between the municipalities mutually and between a municipality and its clients (including the CBS). This reporting will take place by means of an electronic network. The implementation of the whole system has been planned to take place within a few years from now.

It is obvious that the automation of the municipal population register in a similar way as described by the MAP will give rise to further developments on the statistical field (Verhoef and van de Kaa, 1987). For example, by means of a register-based enumeration on the population by status in the family (spouse/lone parent, child, not living in a nuclear family) it will also be possible to compile regularly benchmark statistics on families and (groups of) persons not living in a family nucleus. Already in 1987 such statistics have been compiled. From an organizational and financial point of view this enumeration had to be restricted to municipalities with an automated register[2]. Moreover, analyses from the results with respect to the number of 'family units' (i.e. nuclear families and persons not living in a nuclear family) living at one address have shown that by means of such a register-based enumeration it is also possible to compile statistics on households, provided that complementary statistical data from other courses (e.g. survey research) are available.

Decisions on the organization of such additional register-based enumerations have to be taken yet. These are dependent on the definite form the municipal reporting to the CBS on vital events and migration will take in the new system of population accounting. In consultation with the Ministry of Home Affairs several alternatives are discussed at the moment.

Finally, the implementation of the MAP will also offer possibilities for improving and enlarging the current demographic statistics. First, individual demographic events could be linked, so that statistics on life-cycles could be compiled. Secondly, demographic statistics could be presented on the territorial subdivision of municipalities formerly used in censuses.

Large-scale surveys

3.1 Introduction

Since the seventies two large-scale sample surveys have been conducted periodically by the CBS: from 1975 the Labour Force Survey (LFS) and from 1977 the Housing Demand Survey (HDS). Both surveys aim at describing regularly the situation at a specific field of inquiry (the labour market, the households and the housing market respectively) as completely as possible, as well as monitoring the developments which are taking place at those fields over time.

Consequently, the statistical information provided by these surveys includes both data which can be used to provide some benchmarks on the relative field in question and so-called in-depth data on those fields. These data are obtained by grossing up the survey results to the level of the total population, the annually compiled demographic statistics being the basis. The principal characteristics of both surveys are described below.

3.2. The Housing Demand Survey (HDS)

3.2.1. Topics:benchmark data and in-depth data

Since 1977 the Housing Demand Survey (HDS) is conducted every four years, partly at the request of the Ministry of Housing, Planning and the Environment. The principal aim of the HDS is to provide statistical information on the present housing situation of the population, the expenditures of the population for housing, realised residential moves in the two years following the survey date (Everaers, 1987a).

By means of this survey benchmark-data can be provided on (a) households, and (b) the dwelling stock and other housing units. Benchmark-data on households concern characteristics such as their size, type and demographic

composition. Benchmark-data supplied with regard to the dwelling stock are type of dwelling, period of construction, number of rooms and type of ownership amongst others.

In-depth data relate mainly to characteristics of households which are of importance for the statistical description of the housing situation. In this respect special attention is paid to ownership or tenancy and the expenditures of the households on housing. Other topics on which in-depth data are provided, are residential moves and potential households, i.e., persons in private households of 18 years and over and personnel living in institutional households who want to move to another dwelling or another housing unit.

The benchmark data are only available for regions with circa 100 000 inhabitants or more. For the housing policies of the government this regional detail is sufficient, since mainly big cities and so-called housing market areas are the target areas in these policies.

As already has been pointed out, in the near future more regional detail in the benchmark data on families (and perhaps on households) will be obtained by enumerations from the municipal population registers. Furthermore, annual statistics are compiled on the size of the stock of dwellings at the level of the municipality. These statistics are based on a yearly updated statistical file set up at the 1971 census. In the coming years this file will be regauged by building up an automated register of dwellings by address.

3.2.2. Sampling procedures

At present two sampling frames are available: the decentralized municipal population registers and the Geographic Base File (GBF). The GBF - a joint project of the Postal Service, the Central Bureau of Statistics and the Government Physical Planning Service - is an automated (and yearly updated) register containing all addresses in the Netherlands including codes for the postal district, the grid square (500 by 500 meters) and the territorial subdivision of municipalities formerly used in censuses.

Theoretically the address together with its known occupants as sampling unit would be the best representation of the target populations of the HDS (living quarters, private households and potential households). In this respect, however, both sampling frames have disadvantages. At present it is not possible to draw such a sample from the decentralized municipal population registers, due to organizational and budgetary problems. Moreover, addresses with vacant dwellings are not included in the population registers. On the other hand, the GBF contains only an indication of the number of postal deliveries and type of building for each address.

Weighting the disadvantages of both sampling frames in connection with the target populations of the HDS, the municipal population registers have been chosen as sampling frame and the person as sampling unit. The disadvantage of having no information on vacant dwellings counterbalances strongly the disadvantages of having a very high underrepresentation of sub-tenant households (particularly one-person households). Such an underrepresentation was discovered from analyses of results of the 1971 census with those of surveys carried out around 1971 with the address as sampling unit. However, using this sampling procedure the probability of being selected into the sample is not necessarily the same for all households and dwellings. This probability is twice as high for a dwelling occupied by a household with a spouse as the corresponding probability for a dwelling occupied by a household without a spouse. This is a consequence of the research-design: questions regarding occupied dwellings are only to be answered by the main occupant or his (married or unmarried) spouse; questions regarding the composition of households only by the reference person or his spouse. Therefore, corrections are made afterwards for a great deal of the survey-data on more-person households and dwellings (see the next section).

Under the given budget, the sample fraction (1:150) is chosen in such a way that statistics can be compiled with sufficient precision for the big cities and the housing market areas.

The sample is drawn using a two-step method. In the first step a selection of municipalities is made by the CBS. The criteria used are the size of the intended sample fraction and requirements of the fieldwork. In the second step the selected municipalities draw a sample of persons aged 18 years or older from their population register according to written instructions by the CBS. Until now these instructions have to be made separately for municipalities with automated and mechanized data processing systems as well as for those municipalities which still administer their register by hand. Names and addresses of the selected persons (and some registered characteristics) are sent to GBS.

In spite of these instructions there are always some problems in getting samples of sufficient quality from the different municipalities. Some municipalities which have no automated system are unable to draw the sample. Sometimes preselection of persons has been taken place in drawing the sample. Many of these problems can only be solved adequately, when all municipal population registers will be automated.

3.2.3. Method of estimation

The grossed up HDS estimates are based on weighted observations. The respective weights are determined by using the method of post-stratification (i.e. stratification after selection of the sample). In this method the survey population is partitioned into a number of subpopulations, called strata, and all selected persons within a stratum are given the same weight. Per stratum the weight is calculated as the ratio of the size of the total population which is partitioned in the same way and the number of selected persons in the survey (Bethlehem, 1987).

By applying this method both the non-response bias can be reduced and the precision of the estimates at the level

of the total population can be improved. It is known that these effects are only obtained if a relationship exists between the target variables of the survey and the variables used to construct the strata. The data available for constructing the strata satisfy this condition to a great extent.

The respective weights are determined according to the following procedure (Everaers, 1978b). First, weights are calculated for correcting the over- and underrepresentation of population categories in various areas due to selectivity in non-response. The relevant strata for these areas are obtained using the information on a number of registered characteristics of all selected persons received in the sampling stage from the municipalities, such as sex, year of birth, marital status and family status. The selection of the areas is primarily based on the urban/rural distinction.

Secondly, weights are calculated indicating for the various areas the number of persons the selected person represents. The partitioning in strata is based on the municipal information on all selected persons and the results of the demographic statistics on age and marital status. The areas used in this reweighting are the areas for which data of the HDS are published.

Finally, the definitive weights are determined, first, by multiplying the above two weights and, then, by dividing the obtained results by two in those cases where the probability of being selected was twice as high (see section 3.2.2).

3.2.4. Main results and their reliability

The main results of the last HDS are presented in table 2. Their reliability can be checked by comparing these estimates with results from other statistics. Such comparisons can be made with respect to the estimated figures of occupied dwellings and households.

The estimate of occupied dwellings can be compared with the corresponding figure to be derived from two sources, namely: the already mentioned updated 1971 file on the stock of dwellings and the regularly published figures on vacant dwellings. It is found that the HDS estimate significantly deviates from the last figure. However, it has been already noted that the updated 1971 file will be regauged. There are indications that the information on some of the changes in this stock sent monthly by municipalities to the CBS, is unreliable.

HDS estimates on hosueholds can be compared with similar information derived from the Labour Force Surveys which up to 1985 have been held every second year. Such a comparison shows a significant difference in the estimated number of one-person households between the two surveys. It is not clear yet which figure could be considered as more reliable.

An indirect check on the reliability of the household estimates can be performed by comparing the HDS population estimates calculated by means of the frequency distribution of the household-size with the corresponding figures from the population statistics. In table 3 the several figures are given for the total population and for age groups. Looking at this figures one may conclude that the HDS estimates on households on this point seem to be reliable.

3.3. The Labour Force Survey (LFS)

3.3.1. Topics: benchmark data and in-depth data

From 1975 until 1985 the Labour Force Survey (LFS) has been regularly conducted every two years; since January 1987 continuously. Data collection and data processing in the Continuous Labour Force Survey (CLFS) are completely automated (van Bastelaer, 1987). From the beginning these surveys have been designed to provide statistical information on the labour force, educational attainment and qualifications of the population and commuting of the currently active population.

The benchmark data which can be supplied by the LFS refer to among others the main categories of engagement (such as economically active, educational training, engagement in household duties) and not-engagement (such as retirement or disablement); the size and socio-demographic composition of the labour force (including educational attainment) as well as some economic characteristics of the employed persons such as occupation, branch or economic activity, status in employment and place of work. Due to the sample fraction (circa 2.5%) these data can only be published for the administrative areas of the Regional Labour Exchange (64 regions) as the lowest level of regional detail. However, this regional level suffices for national policy purposes with regard to the labour market.

In-depth data are compiled for the employed labour force (for example on several aspects of the time worked as well as of retirement of working; secondary occupation), and for the unemployed labour force (for example on job seeking, registration at a Regional Employment Exchange and social security benefits received).

Moreover, in due time flow data can be provided, since the Continuous Labour Force Sample collects data on the labour history in the preceding year for the population of 15 years or older. This regards among others the dates employment started or ended in the previous twelve months; the main characteristics of the jobs performed during this period such as occupation and branch of economic activity; the reason of terminating a job as well as job seeking activities for every period of unemployment in the previous twelve months. As yet, this kind of data are collected by means of retrospective questions. At present plans are worked out to use a panel for it.

Finally, in the near future further in-depth data will be collected on various additional topics according to a rotating system. Every year a specific topic will be chosen on which monthly information will be collected. At the moment plans are worked out for collecting data on not regular education and training next year.

In the long run the new design of the Labour Force Survey offers the possibility to provide a complementary

set of stock and flow data on behalf of which better insights in the dynamics of the labour market can be obtained. For the present annual figures are published, whilst the compilation of three month moving averages is worked on.

3.3.2. Sample procedures

The Geographic Base File is used as sampling frame and, therefore, the address as sampling unit. Organizational factors and budgetary reasons prevent the use of the municipal population registers as sampling frame, although the person as sampling unit fits the target population the best.

At present households living at ca. 12 000 addresses are visited monthly. This number is halved in the holiday season. In view of requirements of the fieldwork a stratified multistage sample is used. The first stage consists of a monthly revolving sample of municipalities stratified in ca. 80 geographical areas. This stratification is applied in order to obtain reliable annual figures at the levels of the relevant territorial sub-divisions (i.e. areas of the Regional Labour Exchange and areas covered by the regional subdivisions used by the European Communities). The revolving system is applied to municipalities with less than ca. 20 000 inhabitants only and is chosen in such a way that the territorial distribution of the sample over the whole year is as adequate as possible. The municipalities with more than this number of inhabitants are drawn every month.

In the second stage addresses in the selected municipalities are systematically selected. In drawing the sample a double selection probability is given to addresses with more than one "postal delivery". This procedure is applied in order to reduce evental cluster-effects, since households living at the same address are expected to resemble each other. Therefore, at addresses with a single delivery all households are interviewed; at addresses with more than one postal delivery only half of the households are interviewed. Addresses of institutional households are excluded.

It should be noted that only 4% of the addresses are addresses with more than one postal delivery. The greater part of these addresses regards addresses at which more than one household is living in a dwelling or another housing unit. For the lesser part it concerns addresses with two or more dwellings: not surprising, after all, since the municipalities are recommended to address each dwelling separately.

3.3.3. Estimation method

The grossed up LFS estimates are likewise obtained by assigning weights to the observations using the method of post-stratification. In the biennial surveys roughly the same procedure has been applied as the one mentioned in the section on the HDS. The calculation of weights for correcting non-response effects was based on information from the respondents and - for the non-response - on information from the municipalities sent to the CBS in connection with the fieldwork which was carried out by municipal civil servants.

For estimating annual figures from the CLFS the weighting procedure used in the biennial survey had to be adjusted. In the adjusted procedure the continuous character of the survey had to be taken into account, in particular the halving of the number of observations in the holiday season. Furthermore, the CBS does not have the relevant municipal information for correcting non-response effects at its disposal any more. Therefore, the number of steps in the revised weighting procedure has been extended.

The definitive weights used for grossing up the sample results are calculated as the product of five intermediary weights. First, a weight dependent on the monthly probability of being included in the sample is given. The second and third intermediary weights are calculated for correcting non-response effects. The second for correcting seasonal differences in the non-response; the third for differences in the non-response by various population categories. In calculating the correction weights for non-response in step two and three the same territorial subdivision is used; the population categories are determined by a combination of the characteristics sex, age and nationality. The partitioning in strata for these areas is based both on the characteristics of respondents and the corresponding demographic statistics.

The calculation of the weights in the fourth and fifth step is intended to get estimates which are representative for detailed populations categories (forth step) as well as for geographical areas on a detailed level (fifth step). In both calculations the same characteristics (sex, age and marital status) in determining the population categories are used, whilst one combination of those characteristics is reducible to the other. This principle of reducibility also applies to the geographical subdivisions used in both calculations. The calculation of both weights takes place simultaneously by iteratively proportional fitting. The strata for the various areas are obtained by using information both from the respondents and the system of demographic statistics.

The estimates are calculated as averages for the whole year. The averages on the level of the total population for the various areas necessary to perform the calculations are obtained by linear extrapolation of demographic figures on the first of January of the relevant year. The extrapolation is based on the demographic developments during the preceding year. The method of extrapolation is applied, since the annual results of the CLFS ought to be published only a few weeks after the fieldwork in December has been finished.

3.3.4. Reliability of results

The reliability of the estimates from the Labour Force Survey - as far as they relate to persons in employment - can be checked by comparing these estimates with data on employed persons which are regularly obtained from (partly integral) surveys among private enterprises and public services. It should be noted that the last-men-

tioned data refer to jobs; the Labour Force Survey, however, to persons having a job. Moreover, in the LFS estimates data on the armed forces and persons employed in households are included; in the results of the establishment-based surveys they are not.

Taking these differences into account the main results of both kind of statistics did not significantly deviate from each other during the period 1975 to 1985. This situation changed at the introduction of the Coninuous Labour Force Survey. In comparison with the LFS 1985 the results of the CLFS 1987 show a higher increase in the number of persons employed than could be expected from the increase over this period derived from the establishment-based statistics. This extraordinary increase is but exclusively concentrated under part-time workers with less than 20 hours worked a week. Probably changes in the wording of the questions on employment and a better probing of the CBS-interviewers have led to these results.

Toward integrated population data or socio-demographic accounts

4.1. Separate collection of various benchmark data and coherency in statistical information on the population.

The preceding sections have shown that demographic, social and socio-economic characteristics of the population are collected in connection with the statistical description of a specific field of research and policy. This proceeding has the advantage that coherent statistics can be provided on certain benchmark data and in-depth data on a distinct field simultaneously. In applying this procedure it turns out that for the greater part the data are not tuned to each other. When data obtained in one field are also collected (usually as background information) in another field, very often the relevant figures differ from each other.

Incoherencies in statistical information on subpopulations also exist between results from the above-mentioned large-scale sample surveys and data regularly collected from surveys among private enterprises or institutions of public services. Some examples of the last kind of data are: the data on employed persons already mentioned in the last section, enrolment data obtained from educational establishments as well as data on persons in institutional households based on various surveys among e.g. health care institutions, homes for the aged and other social welfare institutions. Some examples of such incoherencies are given in tables 4 and 5.

The incoherencies are considered unsatisfactory by users of statistical information on the socio-demographic situation of the population. In order to meet the demand for more coherent information on this field the CBS recently started the compilation of Socio-Demographic Accounts (Koesoebjono, 1987).

The underlying aim in compiling these accounts is to provide a coherent statistical description of the socio-demographic composition of the total population in a twofold way. Firstly, on the level of stock data, reflecting size and structure of the population at a certain moment in time; secondly on the level of flow data, expressing changes in the size and structure of the population between two moments in time.

Achieving data coherency in these accounts necessarily implies a process of adjustments in existing data and of additional estimates for lacking data. A coherent system of stock and flow data requires one and the same reference period, uniformity in concepts and operationalizations as well as an identical target population, i.e. the total population of the country. In this respect it should be mentioned that the existing data (a) relate for the most part to different observations periods, (b) are often based on different operationalizations of concepts and sometimes even on conceptual differences, (c) show differences due to the application of sampling procedures (precision of sampling results, possible sampling errors) and of different estimation methods, and (d) refer to different population categories.

4.2. Methodology of the integration: mean features

As yet the stock data in the accounts refer to the situation at the first of January of each year; the flow data to the period between the first of January of two successive years. The accounts are presented in a matrix form: the stock data in the distributions of the marginal distributions relate, therefore, to the beginining, respectively the end of the period under review; the flow data to the transitions between the categories in these distributions. As a consequence, intermediate transitions are not taken into account.

In compiling the accounts the basic principle in the population accounting is followed: that is, the size of the population at the beginning of a period plus the number of persons entering the populations in the course of the period equals the size of the population at the end of the period plus the number of persons who left the population in the course of the period. This rule is consequently applied for each category which has been distinguished in the matrix.

The process of data integration occurs in various steps. First, the stock data (the marginal totals in the matrix) are compiled. For this purpose quantitative analyses are carried out with respect to the differences mentioned earlier in available data, and adjustments in data as well as minor additional estimates are made. In compiling the stock data the various figures are arranged in order of reliability. In all matrices the population figures are treated as the most reliable ones.

Second, the flow data (the cells in the matrix) are established analogously to the compilation of the stock data. However, in this step more estimates have to be made. Not all data are available, and if they are, they are not directly related to the categories used in the stock data.

In the next step stock and flow data are confronted with each other in the matrix. Explanations for differences and contradictions between stock and flow data are sought for. Thereafter, the relevant figures (on flow and even on stock data) are revised.

Finally, a procedure of iterative proportional fitting is applied in order to obtain a matrix which is internally consistent (that is: the basic accounting principle is valid for each category of the matrix) and which deviates as less as possible from the original matrix (established after the third step). During this process, the stock data are assumed to be fixed, only the flow data change.

4.3. Matrix construction and main results

At present two kinds of socio-demographic accounts have been compiled. One consists of coherent statistical data on the population with reference to type of engagement or non-engagement, the other one with reference to its status in household. The basic matrices - for men and women separately - are very detailed since they contain the data by engagement (status in the household respectively) and age group. The data on sex and age composition of the total population is the framework whereupon the data on type of engagement or not-engagement, and the data on household status are gauged.

Therefore, the first step in the matrix construction consists of the compilation of the demographic data matrix by sex and age. Table 6 shows an aggregation of the demographic matrix for the year 1984. Following the compilation of that matrix, the definitive matrix can be compiled step by step for each demographic population category (by sex and age group). The final result is a matrix by age and type of engagement, respectively household status for men and women separately. An aggregation of the first mentioned data matrix is given in table 7.

Both matrices reflects the composition of the population at two successive moments in time, as well as changes which take place between these two moments. Consequently, the destination of persons belonging to a certain category can be traced at the end of the period. Furthermore, the origin of persons belonging to a certain category at the end of a period can be derived. In this way the respective flows - the outflow and inflow - of each category can easily be calculated. This also applies to a calculation of the turnover flow for each category, that is the numbers flowing into and out of a certain category.

From this point of view the data in the respective matrices can serve as a basis for projections, as among others the destination percentages can, with due reserve, be considered as probabilities of transitions. The availability of a series of such figures over time allows to formulate hypotheses about the future developments with respect to processes of change, and in connexion with this, a projection of the various categories.

4.4. Some prospects

At present integrated stock data with regard to type of engagement or non-engagement are already available for five successive years (1980-1985); integrated flow data for three years. Further developments are directed towards (a) the extension with other categories of type of engagement (e.g. engagement in household duties or in voluntary work) and relevant categories of non-engagement (e.g. retirement, disablement); (b) the construction of such matrices for specific population categories, e.g. the alien population; and (c) the compilation of quarterly accounts in connexion with the development in compiling stock and flow statistics on employment and non-employment based on the CLFS.

The matrices with regard to household status will soon be available provisionally for only one year (1985), due to the lack of relevant data at present. In particular annually compiled statistics on households analogue to the population statistics are missing. Therefore, work is underway to compile such statistics for the short term using different sources, especially the demographic statistics and the CLFS. On the long-term it is expected that household statistics can be compiled regularly by register-based enumerations on family status together with survey data.

Finally, studies are progressing with regard to the presentation of transition within the population in order to have a better insight on its mobility.

Concluding remarks

In the preceding sections a broad outline of the statistical system in the Netherlands has been given as far as this system contains elements in reference to the general topic of post-censal surveys. It has been pointed out that various research and statistical techniques are jointly applied to obtain the relevant demographic, social and economic data on the total population. The instruments for compiling the basic demographic statistics, i.e. the system of population accounting (including the municipal reporting to the CBS) and enumerations from the municipal population registers have been described. Next to this special attention has been given to some methodological aspects regarding the large-scale surveys as being the relevant sources for providing data on the social and economic situation of the population. Finally, a description has been presented on the first efforts to generate coherent statistical information on the level of the total population by means of the development of socio-demographic acounts. This broad outline is summarized in figure 3.

The system of population statistics, the mean features of which have been presented above, may be considered as an alternative statistical programme to a population census and post-censal surveys. However, it should be emphasized that this system is not a substitution thereof in the sense that it aims at obtaining exactly the same statistical information. On the contrary, it is to be considered as a procedure of bringing up-to-date the formerly used instruments within existing possibilities and limits posed by (the Dutch) society. Within these possibilities and limits, the systems aims at producing the statistical information users in general are looking for.

References

Bastelaer van, A.M.L., 1987, *The continuous Labour Force Survey*. Netherlands Official Statistics, vol. 2, no. 4, pp. 30-32.

Bethlehem, J.G., 1987, *Weighting sample survey data.* Netherlands Official Statistics, vol. 2, no. 1, pp. 17-18.

Brekel van den, J.C., 1977, *The use of the Netherlands system of continuous population accounting for the population statistics*, Netherlands Central Bureau of Statistics, Voorburg/Heerlen, The Netherlands.

Choldin, H.M., 1987, *Statisticians' responses to the privacy issue.* Paper presented at the meeting of the American Statistical Association, Chicago.

Everarers, P., 1987a, *The Housing Demand Survey 1985/1986.* Netherlands Official Statistics, vol. 2, no. 4, pp. 42-46.

Everarers, P., 1987b, *The pitfalls of secondary data analysis with special reference to the Dutch Housing Demand Survey.* Paper presented at the fifth European Colloquium of Theoretical and Quantitative Geography. Bardonnechia (Italy).

Kosoebjono, S., 1987, *Socio-demographic accounts: framework to measure the dynamics of population.* Netherlands Central Bureau of Statistics, Voorburg/ Heerlen, The Netherlands.

Redfern, P., 1986, *Which countries will follow the Scandinavian lead in taking a register-based census of population?* Journal of Official Statistics, vol. 2, no. 4, pp. 415-424.

Redfern, P., 1987, *A study of the future of the Census of population: alternative approaches.* Statistical Office of the European Communities, Luxembourg.

Verhoef, R., and van de Kaa, 1987, *Population registers and population statistics.* Population index, 53 (4), pp. 633-642.

Vliegen, M., and H. van de Stadt, 1988. *Is a Census still necessary? Experiences and alternatives.* Netherlands Official Statistics, vol. 3, no. 4, pp. 27-34.

[1] Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989. The author expresses his acknowledgements to Santo Koesoebjono for his valuable comments on an earlier draft of this paper.

[2] Notwithstanding this restriction (nearly 75% of the total population has been enumerated), results have been presented for every municipality by generalizing the results obtained from the automated municipalities to the non-automated ones using their respective composition of the population by age, sex and marital status as a base. Given their geographical position and degree of urbanization, the family composition within the automated municipality was supposed to be equal to the family composition within the similar non-automated municipality.

Figure 1 : From reporting of the population to statistics on the population

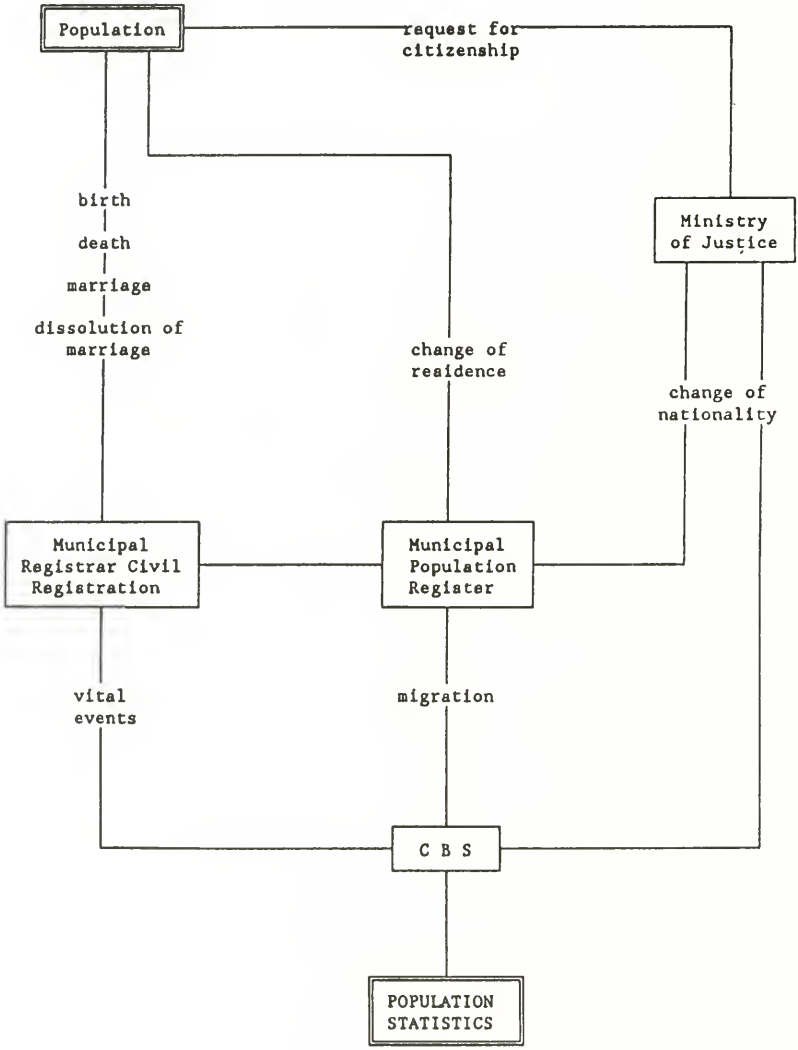


Figure 2 : Synopsis of (planned) enumerations from the municipal registers for revision purposes, 1971-1990

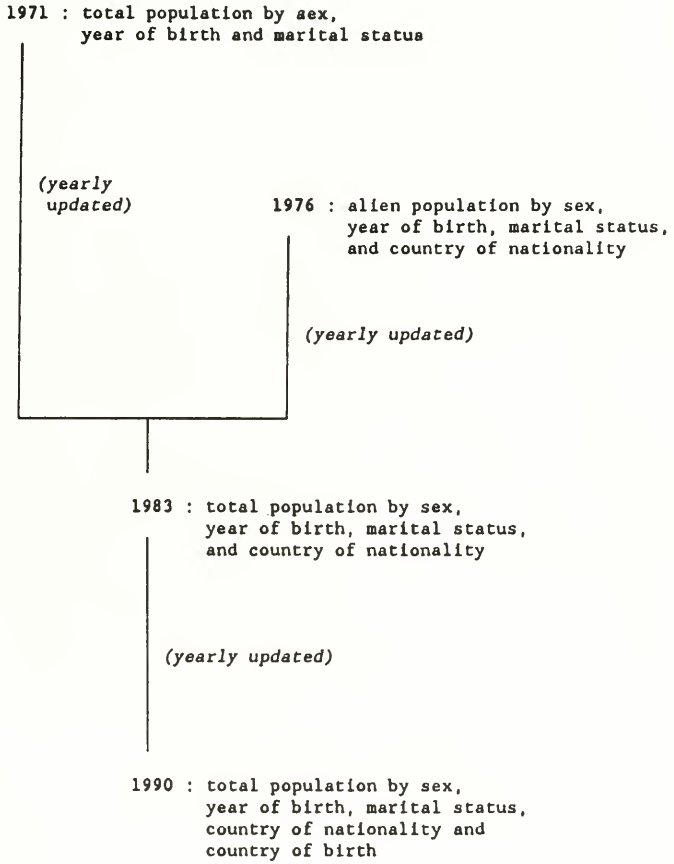
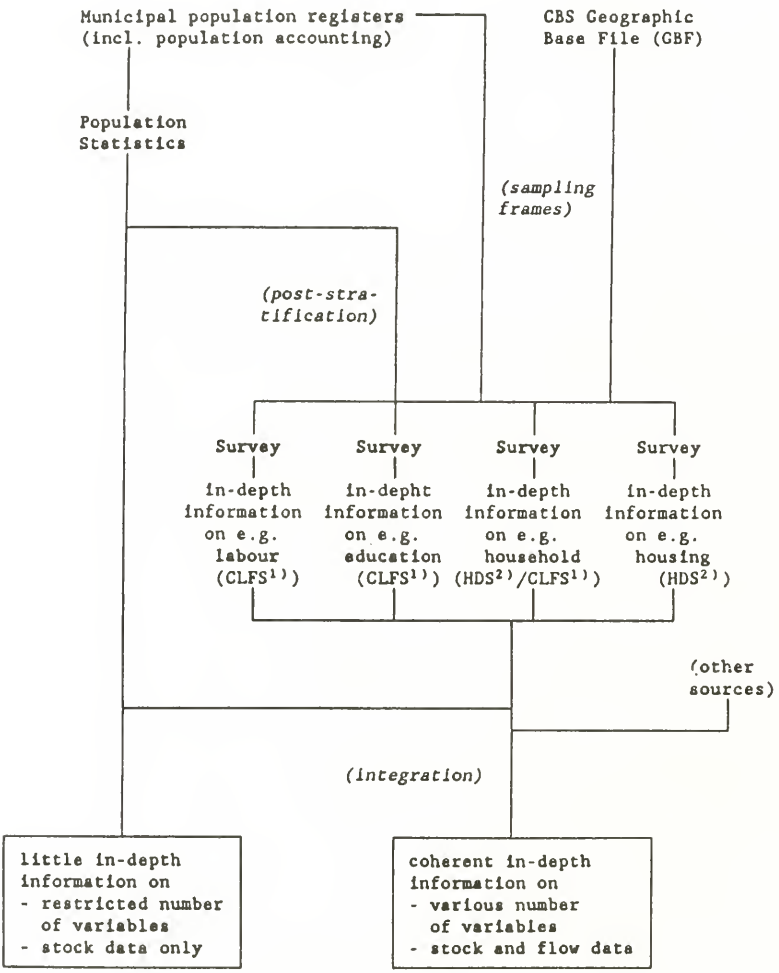


Figure 3. From restricted to in-depth information on the population



---

1) Continuous Labour Force Survey
2) Housing Demand Survey

Table 1. Differences between the results of the registered-based enumeration 1983 and the updated 1971 and 1976 file as a percentage of the relevant categories from the updated files

	Age 0 - 19 years	20 - 49 years	50 - 64 years	65 years or older	Total
	%				
Male	-0,0	0,0	0,1	-0,2	-
Female	0,0	0,0	0,1	-0,2	-
Total	0,0	0,0	0,1	-0,2	-
	Marital status never married	married	widowed	divorced	
	%				
Male	-0,6	1,0	-1,2	-2,6	-
Female	-0,3	0,3	-0,2	-0,5	-
Total	-0,6	0,6	-0,4	-1,4	-
	Nationality dutch	alien			
	%				
Male	-0,1	1,8			-
Female	-0,0	0,9			-
Total	-0,1	1,3			-

Table 2. Population, households and housing situation, HDS 1985/1986

	Total	Occupied dwellings	Other living quarters	Institutions
	x 1000			
Households				
one-person households	1 530,7	1 269,8	167,4	
multi-person households	4 034,5	3 993,6	26,2	
total	5 565,2	5 283,4	195,8	
Number of persons	14 402,1	13 988,8	254,2	251,3 1)

1) Persons of 18 years and older

Table 3. Population estimates by age, HDS 1985/1986 and demographic statistics by age, 1986

	To-tal	Age < 15 years	15-29 years	30-49 years	50-84 years	65 years or older
	x 1000					
Demographic statistics	14529,4	2788,2	3725,0	4107,1	2140,0	1769,2
HDS: population in						
-private households	14240,8	2821,1	3657,0	4052,4	2127,9	1582,5
-institut.households 1)	251,3	-	13,0	20,1	13,7	204,4
total	14492,1	2821,1	3670,0	4072,5	2141,8	1786,9
Difference with regard	%					
to the demogr. statist.	- 0,3	+ 1,2	- 1,5	- 0,8	+ 0,1	+ 1,0

1) Population 16 years and older

Table 4. Population in private households by status in household, LFS 1985 and HDS 1985/1986

	One-person household	Multi-person household			
		reference person	spouse 2)	child	other person
	x 1 000				
LFS 1985 1)	1 434	4 009	3 589	2 018	202
HDS 1985	1 531	4 035	3 622	4 655	199

1) Population of 15 years and older
2) Incl. living in consensual union

Table 5. Population of 15 years and older in full-time education by sex and age, ES 1984 (september) 1) and LFS 1985 (april) 2)

	To-tal	Age 15 - 24 years	25 years or older
	x 1 000		
ES 1984 (september)			
Male	672	621	51
Female	543	515	28
Total	1 215	1 136	79
LFS 1985 (april)			
Male	632	598	34
Female	571	524	47
Total	1 203	1 122	81

1) Educational Statistics
2) Labour Force Survey

Table 6. Total population by age, data matrix 1984/'85

	stock on 1-1-1984	of which on 1-1-1985				
		in population			not in population	
		to-tal	0-14 years	15-64 years	65 years or older	death	emi-gration
	x 1 000					
stock on 1-1-1985		14 454	2 650	9 673	1 730		
of which on 1-1-1984							
in population							
total	14 385	14 222	2 861	9 832	1 728	118	55
0-14 years	2 930	2 914	2 681	252		2	15
15-64 years	9 756	9 891		9 580	112	26	39
65 years or older	1 708	1 617			1 617	91	1
not in population							
birth		173	173				
immigration		80	16	42	1		

Table 7. Total population by type of engagement/non-engagement, data matrix 1984/'85 (provisional figures)

	stock on 1-1-1984	of which on 1-1-1985						
		in population					not in population	
		total	pre-school	full time education	full-time employment	otherwise	death	emigration
	x 1 000							
stock on 1-1-1985		14 454	700	3 378	4 645	5 731		
of which on 1-1-1984								
in population								
total	14 395	14 222	522	3 359	4 627	5 714	118	55
pre-school	709	704	522	180		2	1	4
f.t. education	3 421	3 405		3 162	138	104	1	15
f.t. employment	4 582	4 547		5	4 273	268	15	21
otherwise	5 682	5 565		11	215	5 339	102	14
not in population								
birth		173	173					
immigration		60	5	20	18	17		