

The Register-based Statistical System

Preconditions and Processes

Johan-Kristian Tønder
Statistics Norway

International Association for Official Statistics Conference
Shanghai
October 14 – 18, 2008

This paper is based upon report UNECE (2007). The report was prepared by a Nordic working group. As head of this group I want to thank Finn Spieker (Statistics Demark), Pekka Myrskylä (Statistics Finland), Claus-Göran Hjelm (Statistics Sweden) and the secretary of the group, Harald Utne (Statistics Norway) for good cooperation.

1. Why register-based statistics?

All national statistical institutes (NSIs) have a duty to produce official statistics with the highest possible quality with reference to

- Relevance and completeness
- Timeliness and punctuality
- Accuracy
- Comparability and coherence
- Accessibility and clarity
- Cost efficiency
- Low response burden

When planning a census or a sample survey, the NSIs have some ideas of quality when defining units and variables. We want to use methods for compiling and editing data and for the presentation of statistics, based on the purposes of the statistics and on the ideals that the statisticians have learned at university or from their colleagues in the NSI. In practice, we have to adapt. We may not be able to interview all persons in a household, and have to accept the answers from one of them (proxy interview). The variable we really want to measure may be so complicated to explain that we have to simplify it, and in spite of all simplifications it may be difficult for some people to answer the questions. In addition, budget restrictions and restrictions on the response burden may reduce a big census to a small one, or a census to a sample survey (or a combination of the two methods). Thorough manual editing may be reduced to a rougher automatic editing. We may be able to estimate the effect of the "short cuts" and compensate for this by imputation at unit level or by corrections in the tables. We must at all times strike a balance between the quality wanted and the practical and economical realities.

Administrative data are produced on the basis of some administrative processes, and units and variables are defined out of administrative rules and demands. The definitions may differ from the needs of the official statistics, but the data are usually of good quality for their administrative purposes.

Statisticians always have to make compromises with their ideals in order to get practical results from the data collection. In a way, the use of administrative data has rearranged the situation: We have the product of a data collection process, and have to compare that product with our quality requirements for the statistics to see if the difference is acceptable. The administrative definitions of the target population may not correspond to our needs (employees, but not self-employed, in employment registers), the variables are not always defined in the way we want (de jure instead of de facto place of residence in the population register), and the time references are not always as precise as we want them to be (the time references of the register may not coincide with the "census day" of the statistics). However as these data are almost free for use for the NSIs, we can use our resources to supplement the information not covered by the administrative data, and make corrections just as we have to do when we are using traditional methods.

2. Historical development of register-based statistics

The history of the use of administrative register in statistics in the Nordic countries is briefly illustrated in the following table. It is limited to variables used in population and housing censuses.

The year of establishing registers/introducing registers in census statistics by type of register and country.

| Type of register | Denmark | | Finland | | Norway | | Sweden | |
|-------------------------------|-------------|----------------------|-------------|----------------------|-------------|----------------------|-------------|----------------------|
| | Established | First used in census |
| Central Population Register | 1968 | 1981 | 1969 | 1970 | 1964 | 1970 | 1967 | 1975 |
| Business Register | 1975 | 1981 | 1975 | 1980 | 1965 | 1980 | 1963 | 1975 |
| Dwellings | 1977 | 1981 | 1980 | 1985 | 2001 | 2011 | 2008? | 2011? |
| Housing conditions | 1977 | 1981 | 1980 | 1985 | 2001 | 2011 | 2008? | 2011? |
| Education | 1971 | 1981 | 1970 | 1975 | 1970 | 1980 | 1985 | 1990 |
| Employment | 1979 | 1981 | 1987 | 1990 | 1978 | 2001 | 1985 | 1985 |
| Family | 1968 | 1981 | 1978 | 1980 | 1964 | 1980 | 1960 | 1975 |
| Household ¹ | 1968 | 1981 | 1970 | 1975 | 2001 | 2011 | 2011? | 2011? |
| Income | 1970 | 1981 | 1969 | 1970 | 1967 | 1980 | 1968 | 1975 |
| Totally register-based census | | 1981 | | 1990 | | 2011 | | 2011? |

¹ Household-dwelling unit, i.e. all the persons living in one dwelling

In the period 1964 to 1969, Central Population Registers were established in all Nordic countries, introducing unique personal identification numbers. In the years that followed, several other administrative registers were established. Administrative data as a source for the production of statistics was introduced in the early 1970s. Registers were first used in several subject matter statistics, beginning with population statistics and income statistics. Subsequently, new register-based statistics were developed in all countries.

The time elapsing from administrative registers being established to the point when the data are satisfactory for census use may vary from one subject matter to another, and also between countries. One example is the development of employment statistics. In Denmark, Finland and Sweden, the process lasted for a few years, but in Norway the situation was different.

The step-by-step development is, however, the same in all countries: First, subject matter statistics were tested and published in different areas. Register-based variables were introduced in the census as soon as the quality was considered sufficient. When statistics had been developed for all areas relevant for censuses, a totally register-based census could be conducted.

3. General preconditions

In the light of the Nordic experience, there are certain preconditions that facilitate the extensive use of administrative sources in statistics production.

3.1. Legal base

Legislation provides a key foundation for the use of administrative data sources for statistical purposes. National legislation must reflect the broadly held view that it makes good sense to take advantage of existing administrative data sources rather than re-collect data for statistical purposes. All Nordic countries have a *national statistics act* that gives the NSI the right to access administrative data on unit level with identification data and to link them with other administrative registers for statistical purposes. Furthermore, the statistics act provides a detailed definition of data protection.

All the Nordic countries have an *act on processing of personal data*, which contains provisions on the processing of personal data. According to this act, processing data for statistical purposes is allowed even if it was not the main aim of the data collection. Once data have been processed in a NSI, they must not be used for purposes other than statistics and research (the principle of "one-way traffic"). Data collected for statistical purposes are confidential irrespective of the source. The data collected from administrative sources are confidential in the possession of statistical authorities even if these data are public in the possession of administrative authorities. When handling personal data or business data, both direct and indirect identification shall normally be excluded.

3.2. Public approval

The existence of more and more administrative registers in society may of course trigger discussions on privacy issues. If the public attitude should become negative, politicians may become reluctant to establish new registers or upgrade existing ones. Statistical use of administrative data normally involves linking data from a number of different registers, which may give the impression that the NSI knows "everything" about every single citizen. ("Big Brother Syndrome")

On the other hand, people know very well that administrative authorities are collecting the same data that the NSI uses for statistical purposes. For example, the tax authorities hold information on everybody's employers and earned income, work pension institutes on all the working periods of employees, labour market authorities on all unemployed persons, and pension institutes on old age and other pensioners. In this situation it is very difficult to motivate people to report the same data for statistical purposes.

These advantages seem to be accepted by most citizens as good arguments for statistical use of existing administrative data. However it is important that the statistical agency always remains on its guard in this respect. It is very easy to lose the confidence of the general public, but a major effort to regain it.

3.3. Unified identification systems

One major factor that facilitates the statistical use of administrative data records is the use of unified identification systems across different sources. In the Nordic countries, unified personal identity codes (personal identification numbers) are currently present in nearly all registers used in the production of statistics. It may be possible to link different registers even without unified identification codes, but this is certainly more laborious and time consuming.

3.4. Comprehensive and reliable register systems developed for administrative needs

The compilation of administrative data registers has been initiated from the needs of the functioning of society and development of administration. It has been closely tied up with the development of social security, taxation systems and other administrative needs. These are mostly systems ruled by the state,

and therefore it has been necessary to establish registers on a state level. Very often the purposes of the administrative systems are connected and therefore register information is exchanged between the institutions. Such processes give useful corrections to administrative registers, and hence improve the quality of register-based statistics.

The official domicile of every individual resident in the country is determined on the basis of register information. Likewise, an extract from the population register serves as a basic document that is needed when applying for a passport, getting married or divorced, or when a funeral is held or an estate is distributed.

4. Requirements to data from administrative sources

Contents

First and foremost, administrative registers must contain data covering the most important subject areas in a statistical system for elucidating patterns and trends in society. An important precondition for statistics based on registers giving comprehensive coverage, is that the data contained in administrative registers should be extensive and should cover many variables relating to the relevant units. Gaps either necessitate the supplementary collection of information using traditional methods or limit the content of the statistics.

Units and identifiers

Three central units are essential to the structuring of the statistics: persons, enterprises/establishments and dwellings.

Time references

The time dimension plays a very important role in statistics, revealing patterns and trends in society, and in all areas it is necessary to be able to make comparisons over time. It is therefore vital for statistical usability that reliable time references are contained in registers.

The most important is the *dates of changes or events*. Among the main events of interest are the "birth" and "death" of units. What we are concerned with here is the real point in time at which an event took place bringing about a change in a data item, for instance the date of a removal or the date of a change of industry for a business enterprise.

In addition to dates of events there is a need for *registration dates*, i.e. an indication of when the data value in question was entered in the register. The ideal situation therefore is that any item of information in the administrative register should be accompanied by two dates. Registers in reality often deviate significantly from this ideal.

Stability

An important characteristic of statistics is to describe a process over time, i.e. to show how a particular magnitude develops from month to month and from year to year. It is of great importance, therefore, that concepts in the administrative registers remain constant over the longer term. Otherwise major problems can arise in securing comparable figures from one period to the next. In some cases, but far from all, it is possible to adjust for changes with greater or lesser precision.

Quality

The quality requirements imposed by statistical use coincide to some extent with the requirements that must also be met in serving the primary purpose of the registers; the information must be reliable and recorded with sufficient precision. In addition, data must be relevant to the statistics.

Cooperation with register-keepers

In producing statistics based on administrative data it is not possible to exercise the same control over the content of basic data as in the production of questionnaire-based statistics. We cannot be sure that the registers cover the units of relevance with the same degree of precision or that data are defined in accordance with the needs of users.

It is of course desirable for statisticians to exert a certain influence on data content, but it must not be forgotten that registers are kept for quite specific administrative purposes and that the task of the register keepers is to serve those purposes in the best way possible. They cannot therefore pay too much attention to demands that "only" serve statistical purposes. For this reason, statisticians should not expect to make very extensive demands for additional data, different definitions and the like, and to have those demands met.

5. Developing a register-based statistical system

This means that all statistical registers, in principle, are regarded as being a part of the same system and not as single registers. Such a view has implications for all stages of statistical production: data collection, data processing, quality control and dissemination. The cornerstones are the existence of base registers keeping track of the target populations and the fact that all major registers can be interlinked. In this way, different subject matter statistics may be partly based on the same data sources. For instance, demographic variables are produced in the statistical population register and are used in all kinds of social statistics.

In practice, this does not mean that all variables are linked with all other variables available in the statistical system. When specialised registers are used for statistical purposes, they are always linked to the corresponding base register(s). Linking or not between specialised registers, is a matter of appropriateness. For instance, if variables will never be published or otherwise used in combination, there is no need for linking. In general, however, multiple use of register data will improve the quality of all data involved.

As pointed out earlier, developing a register-based data system has in all Nordic countries been a step-by-step process over a rather long period. Statistical registers have been established in several areas and by 2011 all Nordic countries, at least according to national plans, will have a totally register-based population and housing census system. And even if we call it a census system, the same data sources are used in the corresponding subject matter statistics.

A register-based statistical system should never be regarded as completed once and for all. As new user needs arise and new administrative registers are established, new information should be integrated in the system.

What is a register-based census?

At one end of the line we have traditional censuses collecting data by use of enumerators and questionnaires, using no register information at all. At the other end we have the totally register-based census. Some countries use mixed mode data collection with a combination of data from registers and questionnaires (as a total count or a sample survey). However, even countries conducting mainly

traditional censuses may use register information to some extent, for instance an address list for persons or households for mailing out census forms.

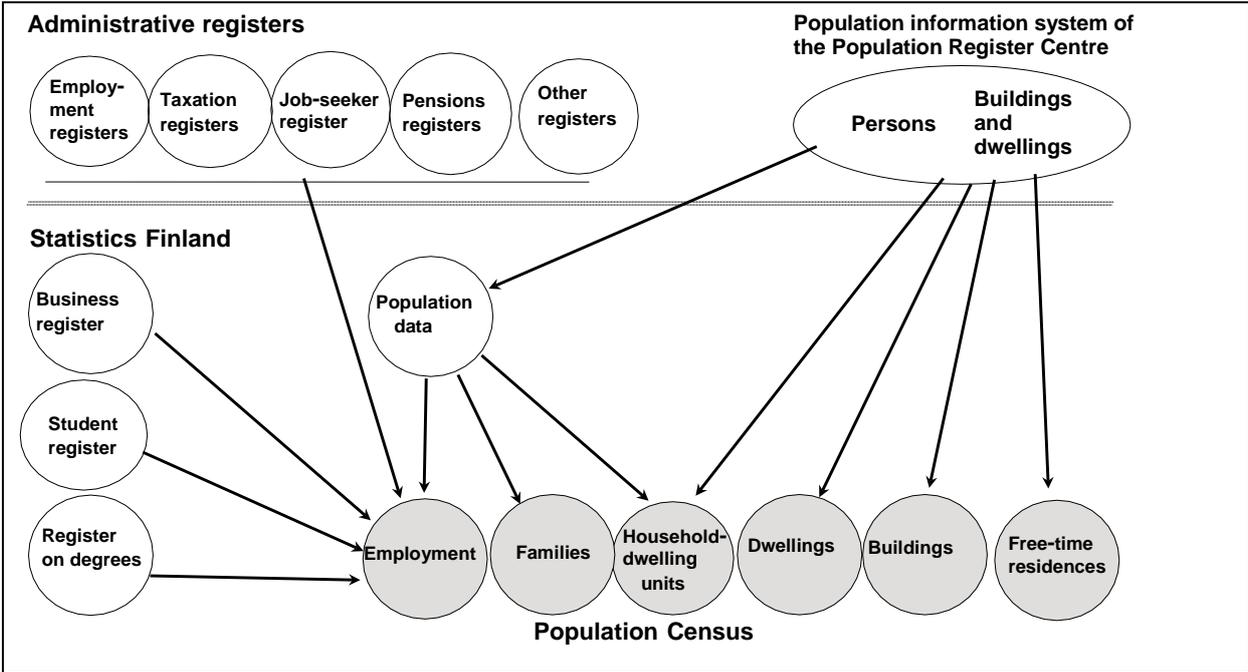
For some countries, such as in the Nordic region, the goal has been to develop fully register-based censuses. For other countries, the goal may be to use registers to a certain extent, even if it is not possible or desirable to conduct a fully register-based census. And for some countries, a register-based census may not be an option at all.

What is required in order to call a census "register-based"? The main requirement must be the existence of a population register and a dwelling register. After all, resident persons and housing units are the basic units in a population and housing census. The link between persons and their dwellings is equally important, giving the household unit. These requirements constitute the cornerstones for developing a register-based census.

Register-based population and housing censuses system

The population and housing census provides the best example of the use of administrative records in statistical production. Here we will use the Finnish system as a main example.

Example: Register-based population and housing census system in Finland



The base registers cover the statistical units relevant to a census: *persons* resident in the country, the *buildings and dwellings* in the country as well as all *enterprises and their establishments* (business register). All statistical units can be linked to one another by means of the identification systems: *persons* can be linked to *families* and *households*, to the *dwelling* and *building* in which they live, and to the *employer* for whom they are working. Similarly, all units can be located on maps using geographical coordinates, because all buildings have been provided by coordinates.

Census statistics and subject matter statistics from the same data source

The register-based population statistics system comprises several statistical sub-systems, including the following:

- Population statistics
 - Population structure
 - Population changes
- Family and household statistics
- Regional employment statistics
- Income statistics
- Education statistics
 - Completed education and degrees
 - Students
- Buildings and dwelling statistics
- Housing conditions statistics

This means that the users get annual (and in some areas even monthly or quarterly) updates of most variables included in the statistics.

6. Advantages and disadvantages of register-based production of statistics

Advantages

Reducing costs are particularly relevant in a statistical context. When conducting a register-based statistics there is no longer any need to design and test the forms, to send them out for printing, to pre-fill the forms, to mail them and send out reminders, to code and record the data. The NSIs no longer need to recruit large numbers of people to sift through the returned forms. In connection with the 1980 census in Finland, for instance, some 2500 employees were hired and trained.

The *reduced response burden* on the population is another significant advantage indeed. For example in Finland in the 1980 Census people had to complete a total of some 8.5 million forms and answer some 100 million questions. If reading the instructions, filling in the form and mailing the questionnaire took each person 10 minutes, the total labour input for the population (and this is still a rather conservative estimate) amounts to more than two million hours. If we consider the price of one hour to be, say, 10 euros, then the overall cost of completing the census form would amount to 15 million euros.

All the statistics are available *annually*.

Disadvantages

One drawback is the fact that register-based descriptions have to rely exclusively on the information contents that can be formed on the basis of the registers available. This imposes some restrictions with respect to the phenomena that are available for description, meaning that *some topics have to be dropped* from the register-based population system. Sample surveys can be used to collect data that are not available through registers. Imputation methods may be used to integrate these data into the file of the statistics.

Additionally, some restrictions are imposed on the definitions of units and variables. A private household is defined as all persons living in the same dwelling (*dwelling household*). *Housekeeping units* (persons living in the same dwelling with joint board) cannot be created using registers. The

population is normally counted according to *legal* place of residence (according to the population register) and not *de facto* place of residence.

7. Concluding remarks

The same statistics based on another type of data

To make statistics from administrative sources means that data collection, editing and other kinds of data processing are done by methods other than the traditional ones. The statistical product, however, is still the same. Instead of asking persons, households or enterprises for information, we ask the owners of the administrative registers. Instead of making quality controls of data received from the individuals, we have to modify administrative data based on our knowledge of differences in definitions and coverage between the administrative sources and the statistical needs. In both cases human resources are required, but the type of qualifications needed may differ. Once the registers are established and the information has passed a quality limit, the register-based solution generally gives lower costs than the traditional one.

Some base registers are necessary

Three base registers seem to be necessary for a statistical system based mainly on administrative sources: a population register, a business register (enterprises and establishments) and a register of addresses, buildings and dwellings. A unique identification for each unit in the base registers, and links between them are necessary. The unique identification should also be used in other administrative registers for these units (educational registers, taxation registers and so on).

Register-based statistical systems may be established without these three base registers, but in our experience the lack of such registers will give problems with the quality of the statistics produced.

Statisticians should give active support to base registers

Administrative registers are normally established and managed by the relevant authorities, but often the NSI provides active support in initiating and organising basic registers. In some cases it has even been necessary for the NSI to take responsibility for establishing and managing base registers as well as other administrative registers. In such cases, as soon as the register is established for administrative use, and a reasonable quality and well functioning routines for managing the registers are established, the responsibility for the register should be taken over by another official and qualified authority. However, even then the NSI should be active in the technical support (reporting quality problems) and in the political support (call for resources needed) of the register.

Many users of a register give better statistics

Establishing and managing administrative registers is very expensive. If many institutions (official as well as private) can argue that their goals can be fulfilled more reasonably by using an administrative register, the government may be willing to put money in it.

Even the administrative users can contribute to the quality of the register. If for instance one user reports that a unit that should be in the register is not present, the inclusion of that new unit will improve the quality for all other users, including the NSI. The NSI may inform the register keepers of quality problems, but in general not report on errors on individual units in the registers. In the active support of the register, the NSIs must assure themselves that they are not breaking the national statistics act or the UN Fundamental principles on official statistics.

Requirements for register data from the statistician

Both units and variables need to be well defined, and the definition should be easy to understand for the person giving the information to the register (the reference person or the civil servant asking the questions and writing down the answers). For each subject it is preferable to have one single register covering all actual units with all information needed. However, in practice we have to accept that

information on one subject has to be collected from more than one source. To integrate data from two or more sources is one of the challenges for the statisticians. If it is difficult to identify the statistical unit from the administrative sources we must create some estimation procedures. Otherwise the data may be useless for statistical purposes. The time reference (point of time or period) must be clearly defined, and an event should ideally be marked with both the date it took place and the date the information was entered into the register.

Administrative registers shall serve administrative purposes, and the rules and procedures that are valid for the register, may change over time. The effect of changes on the quality of the statistics in general, and for the time series in particular, should be taken into consideration when the register is adjusted according to new administrative rules. The best way to safeguard the statistical needs is to have good relations between the persons responsible for the registers and persons responsible for the statistics. These contacts and experiences have to be documented and given to new persons in the register authority and in NSI when a person's position is changed.

Register data is a "common good" inside the NSI

Data from one administrative source may be direct input to more than one official statistics. The source may also be used in the quality control of other statistics, and the more administrative data are used the better the quality will be for the statistics.

All statistical data, register-based or not, are part of the statistical system, and all data may be used in subject matter statistics or in larger statistical data systems like population and housing censuses. Therefore, data are "common goods". The relations between the NSI and the administrative register may be organised through one unit in the NSI or (in practice) one person, but the responsible unit or person is acting on behalf of the whole NSI, and does indeed have to be conscious of this both in internal and external communication.

References

- United Nations Economic Commission for Europe (UNECE) (2007): "Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics."
http://www.unece.org/stats/publications/Register_based_statistics_in_Nordic_countries.pdf
- Houbiers, M., 2004. Towards a Social Statistical Database and unified estimates at Statistics Netherlands, in: *Journal of Official Statistics*, Volume 20, No. 1, 2004, 55-75.
- Longva S., Thomsen I. and Severeide P.I. (1998): Reducing Costs of Censuses in Norway through Use of Administrative Registers. Reprint from *International Statistical Review*, Vol. 66, 223-234
- Schulte Nordholt, E., M. Hartgers and R. Gircour (eds.), 2004. *The Dutch Virtual Census of 2001, Analysis and methodology.* (Voorburg/Heerlen, 2004).
(<http://www.cbs.nl/en-GB/menu/themas/dossiers/volkstellingen/publicaties/2001-b57-pub.htm>)
- Schulte Nordholt, E., 2005. The Dutch virtual Census 2001: A new approach by combining different sources, in: *Statistical Journal of the United Nations Economic Commission for Europe*, Volume 22, Number 1, 2005, 25-37.
- Statistics Denmark (1995): *Statistics on Persons in Denmark – a register-based statistical system*
- Statistics Denmark (2005): *New developments in the Danish system for access to micro data. Working paper for the joint UNECE/Eurostat Work Session on statistical data confidentiality in Geneva 9.11 November 2005.*
- Statistics Finland (2004): *Use of Registers and Administrative Data Sources for Statistical Purposes – Best practices in Statistics Finland.*
- Statistics Norway (2005): *Editing and imputation for the creation of a linked micro file from base registers and other administrative data. Working paper for the UNECE Work session on statistical data editing in Ottawa 16-18 May 2005.* (<http://www.unece.org/stats/documents/2005/05/sde/wp.8.e.pdf>)
- Statistics Norway (2006): *The role of censuses in a country with a register-based statistical system: Norwegian experiences and plans. Paper for the CES Plenary Session in Paris 13-15 June 2006.* (<http://www.unece.org/stats/documents/ece/ces/2006/32.e.pdf>)
- Wallgren A. and Wallgren B. (2006): *Register-Based Statistics - Administrative Data for Statistical Purposes.* John Wiley & Sons, Ltd