



**Economic and Social  
Council**

Distr.  
GENERAL

CES/2001/21/Add.5  
7 May 2001

ENGLISH ONLY

---

**STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

Forty-ninth plenary session  
(Geneva, 11-13 June 2001)

**Annual System of Small Area Statistics Based on Administrative Records and Registers – the  
Possibilities and the Problems**

Supporting paper submitted by Statistics Finland<sup>1</sup>

**I. Annual System of Small Area Statistics Based on Administrative Records and Registers - the  
Possibilities and the Problems**

*The basic structure of statistics production based on registers*

1. Since 1987, all census data have been produced annually in Finland from data in administrative registers. Each year, Statistics Finland produces demographic and employment statistics, building, dwelling, household and family statistics and statistics on housing conditions. The most important of the exploited registers are the Central Population Register (CPR; total number of residents, demographics, families), the Register of Buildings and Dwellings (RBD; buildings, dwellings or premises), and the Register of Enterprises and Establishments (REE; all private sector enterprises and public sector establishments). Additional registers used include registers of employment pensions, taxation, the unemployed, pensioners, and students.

2. The RBD contains data of importance in defining area statistics. The information is linked via identification data to other statistical units. Units are identified through their personal identity code,

---

<sup>1</sup> Paper prepared by Kaija Ruotsalainen, Population Statistics, Statistics Finland

building number and enterprise number. Every building has its own building number comprised of the following parts:

1. Municipal code	3 digits
2. Village code	3 digits
3. Real estate code	4 digits
4. Checking number	1 digit
5. Building number	3 digits
6. Door code	1 digit
7. Dwelling number	3 digits

3. Code numbers 1 through 4 together form the real estate number, codes 1 through 5 the building number. Every flat is given a code consisting of the building number, plus numbers 6 and 7 that identify the flat. Numbers 1 through 7 together form the dwelling number, which is used as the domicile code for individuals in the CPR. The RBD contains data on buildings and dwellings, such as year of construction, year of completion, building material, volume, number of floors, number of rooms, floor area, heating system, equipment, etc., as well as map co-ordinates for each building. The co-ordinate data are needed for the production of small area statistics. Every individual with a permanent residence in Finland is registered in the CPR. Every person's record contains his/her personal identity code, as well as his/her domicile code, which is the same as the dwelling number. If the person moves to a new dwelling, the register authority corrects the domicile code accordingly. This provides the individual with a continuously updated data link to his/her dwelling. A household-dwelling unit comprises all individuals with the same domicile code (living in the same dwelling).

4. The REE contains all private sector enterprises and their establishments, as well as government establishments. Municipal establishments are recorded in a separate data register. These registers provide information on branch of industry, type of ownership, legal form and institutional sector.

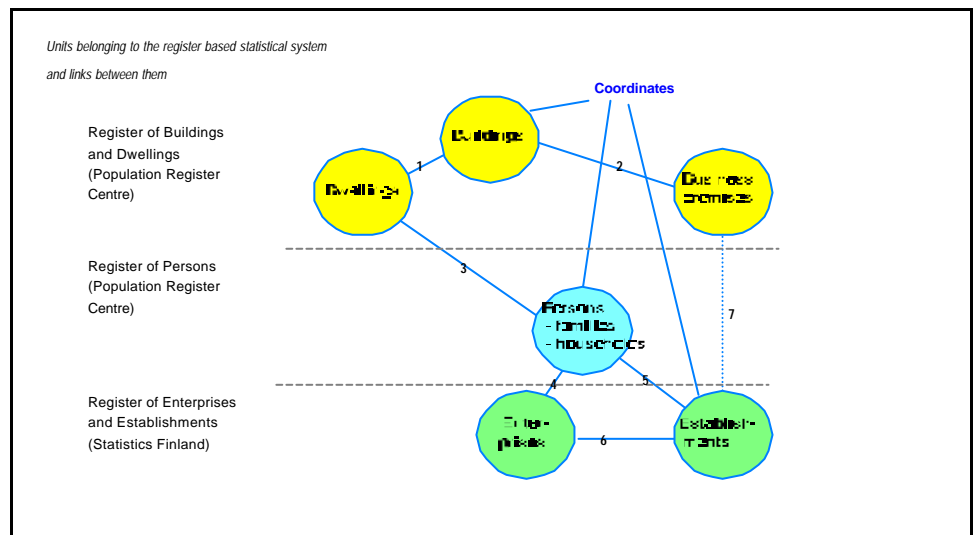


Figure 1. Units belonging to the register-based statistical system and links between them

5. Statistical units like persons, buildings, dwellings and establishments are linked together via different codes. All dwellings and premises (Links 1 and 2) are linked to a building via a building code

maintained by the CPR. The building code provides the co-ordinates for the respective unit. Persons (3) and dwellings (buildings and map co-ordinates) are linked via domicile codes.

6. Employed persons are annually linked (4,5) with employers and their establishments. Links for some business premises (7) and establishments are obtained via organisation codes. With the help of the address it is possible to link the establishment with the real estate number and co-ordinates in the RBD. In the case of some entrepreneurs, such as farmers, the industry is deduced from pension insurance data and from type of income. The location of the establishment is the home address of the entrepreneur.

### *The possibilities of register-based statistics production from the point of small area statistics*

7. One of the main advantages of a population census system that is based on administrative registers is that total data by area can be produced annually. In the past, population censuses were taken at five or ten year intervals in Finland, as is still done in most countries of the world even today. Although data at the whole country or provincial levels were obtainable from e.g. the sample-based labour force survey, statistics by small area could only be produced at few years' intervals and the data could quickly become outdated when regional structures changed.

8. As the 1980s arrived a clear demand had arisen for annual data. Changing over to annual production became viable thanks to the considerably lower cost of compiling statistics from register data instead of from data collected by questionnaires. In today's money, the questionnaire-based 1980 Finnish census cost approximately FIM 200 million, while estimates put the cost of the 2000 census at around FIM 5 million. The transition to annual production was also supported by the fact that it was easier to maintain the processes and routines of continuous production than to build up a new system every few years.

9. The existence of identifying codes for individuals, enterprises or any other statistical units alike is a prerequisite of a register-based statistical system. It also enables the compilation of flow statistics. Flow statistics refer to the temporal monitoring of a certain population, e.g. studying how unemployed jobseekers become employed or completers of education get placed on the labour market, etc. Flow statistics can also embrace the regional aspect: what happens to a population in a certain area over a monitoring period of, say, five years.

10. The fact that individuals and buildings can also be assigned exact area co-ordinates also adds to the possibilities of producing different small area statistics.

11. There are three ways to produce small area statistics:

- a) **Administrative areas:** municipalities, provinces (counties), the whole country, etc. The dwelling number/domicile code is used to link the statistical units to the correct municipality and, by combining several municipalities, to higher level areas.
- b) **Sub-areas defined by the municipalities:** area borders defined by municipalities are digitised into machine readable form. Persons, workplaces, buildings, dwellings, etc., pertaining to these areas are linked to them using the co-ordinate data.
- c) **Map square procedure:** maps can be analysed by choosing the preferred square form (1 km x 1 km, 0.5 km x 0.5 km, or 0.25 km x 0.25 km) and selecting the number of squares to be analysed. The units pertaining to the selected square are then linked to it: individuals, workplaces, buildings, dwellings, etc. After the user has chosen the preferred area, the borders are digitised and the selected statistics are produced. The buildings within the area are linked to the persons living in the buildings and to the enterprises and establishments having their premises within the buildings. As all employed

persons are linked to the enterprise and establishment where they are employed, the linking provides information on the number of persons employed, and their field of work, within the chosen area.

12. Co-ordinate-based data also make it possible to produce time series by small area even if regional divisions change over time. Many municipalities change their internal area divisions at a few years' intervals because of e.g. structural area changes. Population may increase in a certain area and fall in another, new jobs are created in yet another area, etc. When such things happen there is also a desire to change the sub-area division to better serve the required purpose. The development of these newly established sub-areas can then be followed as required because individual persons and enterprises not only have their own sub-area codes but also exact co-ordinate data from earlier time periods.

13. In the register-based statistical system the data on e.g. a person's branch of industry and workplace location are obtained from the REE. Thus, all persons working at the same establishment are assigned precisely the same branch of industry and workplace location. In a questionnaire-based census the data on a person's branch of industry and workplace location were generally based on a person's own reporting. This meant that the data could contain discrepancies even when they concerned persons who worked at the same establishment. Data that are based on registers are thus more reliable and coherent than those collected with questionnaires are.

#### *The problems of register-based statistics production from the point of small area statistics*

14. From the point of producing statistics by region, register-based production also poses a few problems. The following are a couple of examples of problem situations that may arise when data by region are produced on e.g. workplaces, in particular. As already stated, the key prerequisite for register-based statistics production is the existence of unique identifying codes. If they are unavailable, in some cases an actual name may be used to identify a statistical unit or to link it to another register. For this reason an error in some data usually affects a larger group of persons in register-based production than would be the case with statistics produced from data collected with a questionnaire. For example, if there is an error in the REE in the location data of an establishment of an enterprise, all the persons in that establishment are assigned the erroneous data. If the same data were collected with a questionnaire perhaps only a few persons would give erroneous data concerning the location of their workplace. Thus, the advantage of consistency of the data in using registers may in some instances also turn to a disadvantage.

15. One example where the register that is used as the source of statistics lacks these identifying codes is the register of private sector employment relationships. The Finnish employment pension system obliges employers to pay employment pension contributions for their employees. Data on employment relationships insured at different employment pension institutes are collated centrally by the Central Pension Security Institute, from which Statistics Finland receives them. These data on employment relationships comprise their starting and finishing dates, employer's name and pension regulation number. Thus, these data on employment relationships cannot be linked direct to the data in the REE because they do not contain enterprise identification codes. Furthermore, an employer may have several pension regulation numbers in use in the employment relationship data. Employment relationships are mostly linked to enterprises automatically using the employer name in the register of employment relationships, the enterprise name in the REE and the tax authorities' annual control data. In other words, the names of employers in the data on employment relationships are compared to the names of the enterprises which according to the tax authorities' annual control data have paid wages to the persons concerned. If the names in the different sources match, the employment relationship is linked to the pertinent enterprise.

16. Problems in these cases arise if the name of an enterprise is spelled differently in the register of employment relationships and in the REE, or if an enterprise has changed its name and information about

the change has not become updated simultaneously in both registers concerned, meaning that the automatic linking of an employment relationship to an enterprise will fail. Cases like these are processed later manually by supplementing the above-mentioned data with enterprise history data (on e.g. name changes, mergers, etc.) available from the REE.

17. Problems in the manual linking process can also be caused by the fact that in some cases the wage payer may be an enterprise other than the actual employer of the person concerned. For example, the wages of enterprises belonging to a company group may be paid via the parent company. Efforts are made in these cases, too, to ensure that all the employment relationships are correctly linked to the right enterprise, but this does not always succeed. If a person's employment relationship becomes linked to a wrong enterprise it can distort data on both the total number of workplaces and on the industrial structure at the region level.

18. A second problem situation can arise if a multiple establishment enterprise fails to respond to a received inquiry about the establishments its employees work at. A similar kind of problem will also emerge if an enterprise mixes up establishments in its response or if enterprise codes become mixed up at the data input stage. The annual number of such cases is not high: under two per cent of enterprises fail to respond to inquiries and most of them are enterprises that have ceased to operate.

19. Such cases are corrected in the ways described below. If the employees of an enterprise remain at the enterprise level, i.e. they have not been assigned establishment codes, they are divided between establishments in a ratio according to the data in the REE. Information concerning a person's municipality of residence is used as assisting data in this so that persons are linked to the establishment that is closest to it as far as possible.

20. Corrections to errors where a person has been assigned a wrong establishment code are corrected by making a so-called distance matrix between the person's municipality of residence and the municipality in which an establishment is located. If most of the employees of an establishment come from a different municipality, it shows on the error list and the case is then investigated manually. This is done in an effort to eliminate the mixing up of establishment codes in either enterprises' responses or in the data input. Specific, individual errors can, of course, not be corrected but if e.g. the codes of two establishments have got mixed up or two establishments have been assigned the same code, this procedure corrects the error.

21. A third possible cause for errors is that the source data in the REE are incorrect. Data on the industry and location of establishments are drawn from the REE for employment statistics. Thus, if the original register contains an error, it also becomes transferred to other statistics using the register in question.

22. A new kind of problem is created by the recently sprung up enterprises that act as labour force intermediators and large national or international enterprises whose services cover the whole country, such as those in the cleaning branch. Quite often the enterprises that intermediate labour force only have one establishment in a major town, but can intermediate workers to any part of the country. In these cases, the worker has an employment relationship with the intermediating enterprise, not the one in which he or she actually works. This produces biased commuting from quite a distance to the Southern Finnish towns where most of these types of enterprise are located. Corrections to this bias have been made by changing the location of the workplace of a person who is employed by such an enterprise to his or her municipality of residence. However, exact co-ordinates cannot be given to these persons. The information will, of course, not get absolutely correct in all cases, but commuter journeys of hundreds of kilometres will, at least, get corrected.

23. Large national enterprises with regional organisations cause problems because they may only have a few establishments in major towns in different parts of the country, and yet their employees may work in different parts of a certain region. Such cases often relate to occupations in which persons do not necessarily have to work permanently in the same place but their work may be mobile by nature, as in the case of contract cleaners, for example.

24. Statistics on employed with no fixed place of work cannot be produced in the same way as statistics from data collected with questionnaires. Persons with no fixed place of work (e.g. transport or contract cleaning), are assigned as the location of their workplace the establishment from which their work is supervised. In some cases this can create concentrations of certain industries in certain localities. As an example of this in Finland, one could mention a large shipping line whose staff are assigned to its headquarters on the Åland Islands.

25. Tables by municipality are produced as a final check-up in which changes in the numbers of workplaces by industry (at 2-3-digit levels) are compared to the year before. If a deviation from the level for the whole country exceeds a certain limit, it will be examined more closely to see whether the apparent change is real or caused by an error, and any possible errors are corrected. In this examination, corrections are made to errors that have not become corrected through earlier measures and may have been caused by erroneous data in the REE, input error, erroneous reporting by an enterprise, etc.

26. All the above-mentioned error types are minor at the level of the whole country, but may in isolated cases cause major errors in data concerning a smaller region. However, it is justified to claim that the significance of all errors is negligible when compared against the advantage that detailed data are annually produced for each municipality.

-----

*Sources:*

Myrskylä, Pekka and Ruotsalainen, Kaija (1997). Annual System of Small Area Statistics Based on Administrative Records and Registers. *Contributed Papers for the 51<sup>st</sup> Session of the ISI, Istanbul*. Book 1, 511-512.

*Register-based Employment Statistics in Denmark, Finland, Norway and Sweden* (1996). Eurostat and Statistics Finland, Helsinki.