

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**UN/ECE Work Session on Statistical Data Editing**  
(Madrid, Spain, 20-22 October 2003)

Topic (iv): Data editing by respondents and data suppliers

**THE JUNCTION BETWEEN EXTERNAL DATA AND STATISTICS DATA**  
**Is it possible to optimise the roles of data suppliers and users?**

**Invited paper**

Submitted by Statistics Finland, Finland<sup>1</sup>

**Abstract**

Statistics Finland tries to use as much administrative data as possible. An external data supplier is not collecting these data for us, but just for themselves. There are, fortunately, good common keys, in most files, developed during recent four decades. This helps much our further work, especially linking and matching a number of files together, and the editing these further during the process. There have been occurred both drawbacks and progress during these years. Thus we have been able to improve our productivity and data quality but sometimes we have met difficulties. In the latter case, we have tried to tell to a supplier what are our needs and consequently, the needs of our users. But if our targets have not met the targets of a data supplier, we must have accepted this situation. And we must have adjusted for our data collection and editing systems to this new circumstance. Sometimes, we must have started even a new survey if the quality of information has been worsening too much without this operation. This paper discusses this kind of questions using some examples both from the point of view of suppliers and that of our statisticians and their users.

**1. Introduction**

A big amount of statistics data, especially data on facts but not such things as on attitudes and opinions, are in many statistical offices derived from external suppliers such as registers, other administrative sources or surveys provided by other bodies. A basic idea is that this kind of external supplier is capturing such data for its own purposes, primarily. If the same data seem to be appropriate for a statistical agency too, this should be exploited. The multi-exploitation would improve without doubt productivity, but for the quality it can be controversial. If agreed the data will be submitted to this agency, consequently, but the supplier is not necessarily responsible for data quality from its statistical point of view. Some external data thus will be received as such without any strict quality guarantee.

---

<sup>1</sup> Prepared by Seppo Laaksonen ([Seppo.Laaksonen@Helsinki.Fi](mailto:Seppo.Laaksonen@Helsinki.Fi)) and Ismo Teikari ([Ismo.Teikari@Stat.Fi](mailto:Ismo.Teikari@Stat.Fi)). Pekka Myrskylä (Statistics Finland), Mikko Pellinen (Finnish Centre for Pensions) and Kalevi Heiliö (Central Taxation Agency) have been interviewed.

However, the supplier's premises are known and these can be taken into account in further data editing and analysis. This is typically the starting procedure when inserting external data to statistics production. The second stage is to go forward so that as many statistics production requirements as possible may be met already at the junction between an external supplier and a statistical agency. This may mean among others that this external supplier provides more variables than minimally needed for itself. And moreover, the data may be edited more carefully than minimally needed for exploiting the data by themselves.

There are in Finland a lot of experiences on using external data for surveys and statistics, but no data are automatically reasonable for using without further editing. In this paper, we describe our circumstances from the external data capturing point of view. Administrative data or external data are used in almost all statistics, and understandably, we can here cover only some areas. Our focus is on multi-level and longitudinal information on businesses and people.

We use different examples which cover data from the taxation authority, from the central population register, from the employers' organisations, from the pensioners' organisation and directly from businesses. The last point is becoming very important especially concerning large and complex enterprises. In this context, the statisticians use the term 'profiling,' The purpose of profiling is to identify the enterprise, its legal structure, its operating structure, and the production units, that will be used to derive the statistical units of the enterprise. In order to succeed in this process, special profilers are required to recruit for statistical agencies. These will work in close connection with external data suppliers.

This paper is next organised so that Section 2 discusses biggest editing issues of our longitudinal census data. In Section 3, we give the three examples on how an external data supplier is taking care of data quality. Section 4 concentrates on profiling that could be a good response to improve business data quality if the requirements of editing could be included in this action. The two annexes give some impression on our huge register environment.

## 2. Experience on longitudinal census types of data

One of the most important and demanding data bases in Statistics Finland is the so-called *Linked Longitudinal Employment Data (LLED)* that has been conducted for each year from 1987 onwards. This provides a good longitudinal data base for each statistical year covering all residents in the end of the year. There are the three key questions in constructing and cleaning this data base:

- (i) Who is a resident and where this resident is living (to include the co-ordinates of the building is the target),
- (ii) What is the employment status of a resident,
- (iii) Where an employed resident is working (again the target is to get the co-ordinates of the building).

The data sources for these data are derived from various administrative sources or registers (see also Ruotsalainen 2002). The Central Population Register (CPR) covers the 'full' information about residents, and homes and business buildings. A problem is that this register does not indicate residents to business buildings (premises) that would be ideal, or even in the neighbourhood of these buildings. Hence, we may get quite good information about people's living area, although there are naturally some errors due to especially some delay or missing data (for example reports of internal migration) in updating the register. Statistics Finland uses standard population information almost as such as received from CPR. Thus the data quality from the supplier is rather good for basic statistical purposes. But we are not satisfied with administrative information in order to identify residents to their working places. This thus leads to a quite complex editing procedure for certain sub-groups as illustrated in the next paragraphs.

What is especially needed, is the identification of employees to their local kind-of activity units (LKAU) on one hand, and secondly, we need to know the co-ordinates of this unit for small area statistics. Although we have a good Business Register (BR) file, the LKAU information of it is not reasonable for the LLED. It should be noted that the quality of the LLED would be excellent if the employees could be identified to buildings, but more realistically, we are still happy, if this could be done at some small area

level, at least at municipality level. Hence we try to create our LKAUS's using our editing techniques in such cases when we have no reasonably detailed information. We next give some examples about problematic cases:

- LKAU'S in municipality and government business entities are not well enough available in registers. As often as possible, the area of buildings have been tried to deduce from the address or other information. After publishing statistics with possible incorrect information, the local users have reacted that has led to corrections, naturally. Sometimes, it is better to organise an additional inquiry to these units using the information about the previous file and asking respondents to specify in which LKAU each person is working. This is to some extent made electronically but in some others manually. Since these entities are often using software firms for data service, and these have quite standard formats, Statistics Finland may receive correct information directly from this sub-contractor (naturally, there is a agreed permission from the data owner).
- LKAU information is not good for all large businesses which have a high number of local activities (e.g. banks, kiosks, franchise firms). The solution as in the previous point may be used, or good guesses based on employees' home residence.
- There are firms which rent workers to other firms that are locating around the big region or even around the whole country. It is not possible to get correct LKAU information. A standard solution is to use the home residence area as the LKAU residence.

Moreover, we want to notice that we have met a lot of difficulties in handling addresses. Earlier, the addresses in rural areas were very rough due to the lack of street/road names, but currently Finland's roads/streets are coded well for mailing purposes. Consequently, this helps in specifying the home and business buildings. On the contrary, there are used unclear mailing addresses (post box, post office), such that do not indicate well the location, even the municipality cannot be known in all cases definitely.

When the statistical units, people and businesses, have been regionalised reasonably well, we have still left much further editing work. Financial information such as pensions, unemployment and other social benefits, student supports, wages and salaries of employees, are substantial in order to identify the employment status of each person. Of course, these exact financial numbers are of importance, too. There have been used several sources for deducing the employment status. Some other statistics such as wage statistics were conducted from a certain period like from November, and naturally this information is very useful for the LLED, although the cross-section of the LLED is the end of the year. Currently, we can use complete taxation information on wages, for example, but this does not still give a perfect information for our needs.

First, there can be a high number of simultaneous work-contracts for one person, and hence in the editing, it is required to decide what is primary and what is secondary. This cannot be deduced completely from the administrative information. Secondly, and relating to the first, the period of earning is not always correct. Often this means the period (month) when the wage has been submitted, not the working period. There may be too long and short payment/working periods too in this source. A special question is concerned vacation or other leaves, when a person is paid but not working. In the future, we may get even worse information, since it is of importance for the initial data provider to know the working period at monthly or more detailed level, or at least at yearly level.

Some groups are problematic as students who may be working while studying. Our standard is to put such persons to the student category who are officially studying but not receiving wages at the same time. As above mentioned, there may be some errors due to a periodicity problem, among others.

The Finnish LLED is thus provided for each calendar year, the outcome being available about 1,5 years after the reference period (but preliminary data about 10 months after the reference period). This data base gives a good start for providing every five years a more complete census. This is also based on

administrative information. There are several additional variables in this census, but the most difficult task is include 'occupation' in the data set.

Earlier, we had better information on occupation than today, since the taxation authority collected textually occupation of each person. This textual information required, naturally, much automatic editing before deducing the code. After 1995 this kind of information has been worsened. The taxation authority uses now much more a so-called tax proposal to persons, not to entrepreneurs. At the same time, they have dropped out occupation information from their file. For the census of the end of 2001, Statistics Finland organised a hard effort to code the occupation of each employee. This was done in several ways using as much available information as possible. The best sources were found from the yearly wage statistics which cover almost completely municipalities and government entities, and the members of the two employers' federations. All these sources contain an occupation code for each employee in November or December of 2001. This code can be quite well converted to the ISCO standard.

There are still about 30 percent under-coverage in these available sources, and hence Statistics Finland organised the two techniques in imputing these missing values:

- (i) The occupation of employed for smallest enterprises (less than 5 persons based on the BR) was imputed logically so that the same occupation had given to all persons deducing from the main activity of the enterprise (e.g. if the activity of an enterprise was hairdresser's then the occupation for all was hairdresser, similarly to car mechanics). This, naturally, gives only a rough code.
- (ii) For other enterprises, a survey was conducted to indicate the occupation for each employee found from the employment register. This was the most expensive part of the 2001 census.

There were done several quality checks for this data comparing changes from the previous census (1995), and especially comparing these results with the parallel labour force survey. The coherence was looked to be rather good.

### **3. Points of suppliers**

In this section we present some editing practices used by the Central Taxation Agency, the Finnish Centre for Pensions (ETK) and the Employers' organisations. .

#### *Taxation Agency*

The Central Taxation Agency (CTA) is not naturally the only provider of their data. They are using different sources for collecting their data:

- some data are coming as data transfer from initial registers,
- some data are directly given by taxpayers (people, businesses, other societies),
- the third party has delivered data to the CTA,
- some data have deducted using the data from their own archives.

All data have been in principle checked keeping in mind the needs of the CTA. Alternatives for the data as follows:

- (i) The data arrives on time and are complete and correct, including the required format of the data.
- (ii) The data arrives late but are complete and correct.
- (iii) The data are not correct.
- (iv) The format of the data are not correct.
- (v) The form of the data are correct but do not correspond the truth.
- (vi) There are logical discrepancies in the data.
- (vii) The data are correct and arrived on time but do not correspond to the truth from the point of view of taxation rules (abuse).

Although there are rather good rules for data checking, it is observed that technical errors can be observed better than subject-matter-based errors. The routines are from year to year about similar. If the

taxation system or the technical infrastructure has been changed, there may be in the beginning occurred more problems and errors.

### *Finnish Centre for Pensions (ETK)*

The ETK maintains a number of pension data (registers) but these data are also coming from other sources, such as from other pension institutes and from the Central Population Register. The main purpose in this data transfer is related to the administrative needs of the ETK, they are not really thinking the needs of statistics. An exception is concerned statistics on pensions paid to foreign countries: this is motivated from international agreements in this area.

There are in the traffic of basic registers a lot of messages that have to be checked. The following steps have been used in this cleaning:

- (i) The form of message has been checked.
- (ii) The logic of the message within each message.
- (iii) Checked new information with the old one.
- (iv) Searching for problematic cases using a computer program made for this purpose.
- (v) Comparing the values with the values of pension institutes (coherence).

Every month has been taken an aggregate output (tables) and these have been compared with previous outputs. The outputs have been used as such but also these are useful for checking results at macro level.

The registers themselves do not include any proper statistical concepts, these have to be created. Following operations are used for this purpose. The starting point is that a new system has to be evaluated while introducing the system into use. After that, the volumes and respective aggregates will be always checked and assessed. Also, some random checks are done. The data mass is so big that dramatic program and other errors will be revealed quite easily. Naturally, some small errors may remain until the end.

A usual approach is to make all checks and tests using first quite small micro data, this is as a training data set. As soon as these results are satisfactory, they apply this experience on mass micro data. There are the certain expectations (based on small-scale test) for the results, if these are not met, a doubt on error will be arisen. Various types of distributions have been often compared in these checks. Naturally, the extreme values are also looked carefully.

In a normal case, the time spent for data cleaning is slight. And consequently, the data can be released to further users, including Statistics Finland, quite soon. The risk for essential errors is small. It may however happen that Statistics Finland is not happy with certain values. This leads to give feedback to the ETK and this part of the data will be checked again.

### *Employers' organisations for data of wages*

Most part of the private sector wage statistics have been based on the data derived from the two employers' organisations (we use later the Finnish acronyms 'TT' and 'PT' for these federations). This case differs essentially from the previous two cases in the sense that both the TT and the PT have their own statistical interest for these data, too. Earlier their needs for these data were even more important and detailed than currently but since these organisations had established the wage statistics systems and succeeded well in providing statistics for their users, for workers' organisations and for Statistics Finland, this tradition has been continued.

In the 1970's and 1980's Statistics Finland received only the aggregated (tabular) data from these organisations, but in the 1990's the agreement releasing micro data was made. This was motivated much from the requirements of the EU wage statistics regulation. Thus, currently Statistics Finland can combine the data from their own sources and from these external sources at micro level, and provide interesting outputs.

There are however some differences in the targets of both data providers. In general, Statistics Finland and Eurostat have more demanding targets for the data quality than these data suppliers have. Although on the other hand, the TT&PT has their own specific variables that have not used in Statistics Finland. Some examples about different targets:

- The business units of the TT&PT data are not similar in all cases as in the business register (BR). The TT&PT have the data only from their member enterprises, and these do not correspond always to BR units. It is possible that their units are closer to 'real-life', but it is not possible for Statistics Finland to use different units in different statistics.
- Some statistical classifications are not of high interest to the TT&PT. Hence the industry class, for instance, is problematic in some cases.
- The TT&PT have started to use some Statistics Finland classifications, especially our register of formal education. This has harmonised results for recent years, but in a long term, there are problems. Occupations are naturally in TT&PT statistics such that are used in real-life, the same is concerned Statistics Finland wage statistics, too. But the linked wage statistics from a number of micro data cannot be well harmonised although the initial codes have been converted to the standardised ISCO classification.
- Some variables are difficult to measure and to survey in all wage statistics, such as regular work time, daytime work time, weekend and other special work time, overtime work and over time wage, bonuses and tenure, but less motivated for the TT&PT than for Statistics Finland. This has sometimes led to impute some items to TT&PT data using auxiliary information from other statistics (Laksonen 2003).
- Part-time and temporary workers may be more often missing from the TT&PT data than from Statistics Finland data. In both data, this is a problem.
- The TT&PT data do not cover all businesses of the sectors desired to examine, especially small businesses are badly represented. Hence, some additional data collection has been made in Statistics Finland, the purpose being to fill these gaps. This kind of special data captures make, on the other hand, the statistics production system quite complicated, giving challenging jobs for statistical methodologists. So, we are happy and not so happy with these data. Hence, we have sometimes discussed to break up to use the TT&PT data, and to start our own production that can be designed better than using the current 'patch-quilt' approach.
- For econometric research the TT&PT data are much more used than any Statistics Finland data due to their long tradition with about similar quality (longitudinal analysis will be reliable). Hence the TT&PT data from 1980 onwards at micro level has been released to several researchers that seem to be fairly happy with the data (e.g. the ongoing International Wage Flexibility project with 13 countries, sponsored by the European Central Bank, The Brookings Institution and The US Reserve Bank, more information from S. Laaksonen, among others). The main problems are arising, as in general when using statistical data, from changes in terms (e.g. wage concepts) and classifications (e.g. occupation, education). Also, the business units have been changing quite much during the years.

#### **4. Profiling**

As far as the Business Register (BR) is concerned, there are three ways to update a BR. Firstly, it is important to take an advantage of survey feedback. However, most surveys do not collect information about the various levels of units in a business. It is thus needed other sources to update information of legal units, organizational units, physical locations etc. Another way to update a BR is to use surveys which are specially planned for this update. The third way is to use administrative sources. However it is probable that these records do not give just the right information about the structure of large units which is needed to define and maintain the sets of statistical units.

The supplier information in tax records normally applies to legal entities. However statisticians normally want information about statistical enterprises and establishments which are the units to which industrial classification codes are assigned. There is no problem regarding to single establishment enterprises. Although these are about 99 per cent of all enterprises roughly half of the business sector output comes from few multi-establishment enterprises. For these complex enterprises the administrative records have a limited value because they do not facilitate the segmentation by region and industry. Instead of it tax records offer information about the enterprises as a whole. Sometimes they give information of groups of

enterprises where the unit enterprise can be identified from consolidated accounts. However there are no consolidated records for establishments.

It is thus needed profiling which means to keep direct contacts to define the structure of large businesses. At a minimum it involves a personal contact. Within many statistical agencies there is a unit with many persons to profile the largest enterprises. Thus profiling is thus the best way to clear a complex structure of the big enterprise. It is on the other hand expensive, time consuming and needs a trained staff. The problem on profiling is to clear how to collect information for statistical units. The extent to which standard statistical units correspond to legal units varies depending on the nature of available updating sources, legal structures and data availability for various statistical units.

Because profiling is expensive and time consuming there must be some criteria to choose business to be profiled. Following criteria are applied:

1. Size.

Due to the very skew distribution of population according to size measure selecting the largest 1 percent we achieve 50% of total contribution of quantitative response variable

2. Complexity

3. Propensity to change the structure.

If an enterprise tends to change often its structure it should be profiled often.

4. Respondent relations.

5. If a business has been a poor respondent, profiling is likely to improve its relationships with statistical agency.

6. Reliability of alternative sources.

7. If there reliable information got from other sources profiling is not so important

In Australia and Canada, there have been found following profiling experiences. Most of the profiled business units were already recorded but were shown under previous ownership. Therefore frames taken were correct in terms of identifying units, but were inaccurate in identifying the names and addresses to whom the forms to be sent. Many industry codes changed. Incorrect statistical units were recorded resulting in surveys collecting data from the wrong units.

Profiling can be extended to include not only the collection of the forms but also editing and even compiling the data received from large businesses. Because the staff of profiling will have less detailed knowledge than a subject matter specialist will have, relative responsibilities have to be carefully defined.

Inconsistencies between data reported in different collections can result from misunderstanding of data item definitions, problems of consolidation of data items. When inconsistencies are found they should be resolved with aid of subject-matter specialists.

In Finland profiling has taken place by following the press. If it is found remarkable changes in some business structures the business is contacted and tried to clear what kind of inconsistencies it will cause in business register. In this autumn it is found a group of four subject matter specialists to follow up changes in the big enterprises and in the group of enterprises. The corrections in business register will happen now faster than earlier.

## **5. Some concluding comments**

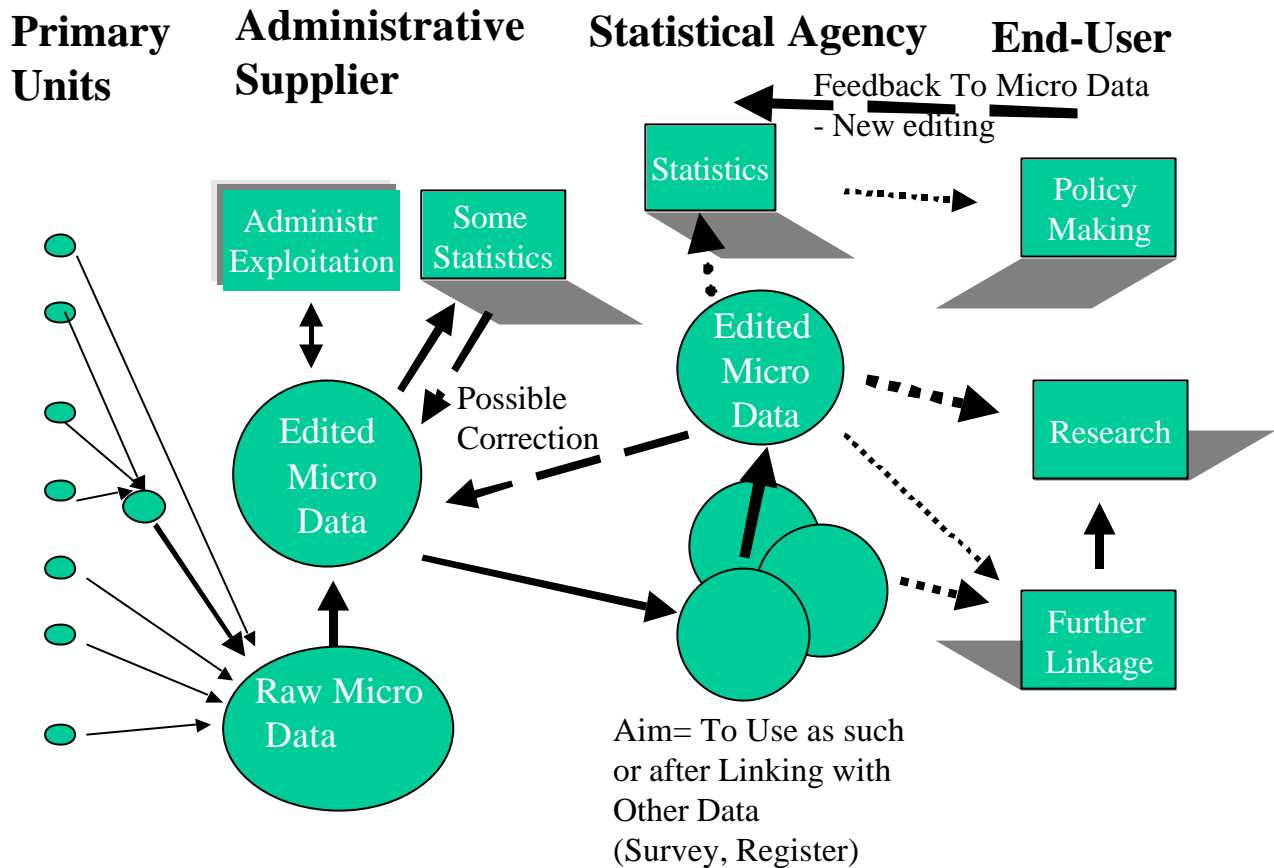
The data for statistics can be provided from a number of sources. It is naturally ideal first to exploit all existing data sources before going to start special data collection procedures such as surveys. This paper discusses the data quality received from outside a statistical agency, especially from administrative sources of public and private providers. If these data are already of high level, these are easily and with low cost usable for statistics, even so that just printing out and publishing by the agency. Also, the same data may be exploited as auxiliary data after linking and matching with internal data of the agency (see on auxiliary data sources, Laaksonen 2002).

The attached figure illustrates the cooperation between external data suppliers and statisticians both within the agency and outside it, and including end-user statisticians. This scheme is naturally simplified. The whole chain may be much longer than presented here. From the point of view of a statistical agency, the last preceding supplier is most important such as a taxation authority or an employers' federation. This supplier may be using several 'sub-contractors' such as local offices or adp companies for collecting individual data from primary units. In this stage also some editing rules have been used concentrating on such aspects that are crucial for administrative use of the data (e.g. we may trust well in wages paid for a worker since he/she usually will check how much money receives, but other information may be erroneous). This same target will continue in next steps, some information is usually almost complete.

It seems that in Finland, a typical external data supplier uses some basic checks in qualifying the data both at individual level and macro (tabular) level. They want to ensure that their data are enough good their own purposes. At the same time they hope that it satisfies the requirements of Statistics Finland, at least in some respects. They are willing to receive feedback from Statistics Finland if this observes some fatal errors (with influence at macro level), and if possible they will provide a new corrected file. But all aspects, especially many statistical classifications, are not of high interest for them.

There is need for further editing in a statistical agency, too, especially due to links of the external data to other data of the agency. We have presented in the paper some most difficult problems in this stage, especially arising from different business units used by the external supplier and the statistical agency. All problems cannot be solved; hence the agency concentrates on some crucial ones. Finally, there are further steps in this process, the role of end-users being very important in validating the data at least in a long term. The importance of researchers cannot be overrated in this context. In particular, the researchers who are exploiting micro data (often based on several data sources and from a number of time periods) have been very useful for us (E.g. Maliranta 2003). They are critical and ready to give feedback to statistics. It is not always possible to correct individual data a long time after the reference period, but these gaps may be marked, and the observed problems may be corrected in coming years.





## Illustration of the Roles of Data Suppliers and Statisticians

Nevertheless, we are not satisfied. There could be many tools to improve the situation. A question is what incentives to offer to an external data supplier that this would edit their data better, also including the needs of statistics and research. Secondly, there is need for better interaction between different data providers, both between various agencies and also within a statistical agency. For example, it is not a very common house-style to inform a business register unit about errors observed during a business survey process.

### References

- Laaksonen, S. (2002). Need for a High-Level Auxiliary Data Service to Improve the Quality of Editing and Imputation. *UNECE Work Session on Statistical Data Editing, 27-29 May 2002, Helsinki, Finland*. <http://www.unece.org/stats/documents/2002.05.sde.htm>
- Laaksonen, S. (2003). Alternative Imputation Techniques For Complex Metric Variables. *Journal of Applied statistics*.
- Maliranta, M. (2003). Economic Research with Firm Data – Practices and Experiences in Finland. Invited Paper of the 54<sup>th</sup> Session of the International Statistical Institute. Berlin, 19 August.
- Pietsch, L. (1995). Profiling Large Business to Define Frame Units. In: *Business Survey Methods* (Eds. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott. Wiley: New York 1995.
- Ruotsalainen, K. (2002). Use of Administrative Data in Population Censuses – Definition of Main Type of Activity as An Example. *UNECE Work Session on Statistical Data Editing, 27-29 May 2002, Helsinki*.

Finland. <http://www.unece.org/stats/documents/2002.05.sde.htm>

Smith, P. (2000). Statistics Canada's Broad Strategy for Business Statistics. *The Proceedings of the Second International Conference on Establishment Surveys*. Invited Papers, pp. 19-28. Printed in Canada.

### **Annex 1: Registers related to the Finnish Population census system**

The Finnish population census system draws on the basic registers maintained by society giving data on the population census target units and the links between them. These units comprise individuals, families, households, buildings (and summer cottages), dwellings (and business premises), enterprises and establishments (annex 1).

**The Central Population Register** is a basic register in register-based census system, defining the population included. The population contained in the employment statistics comprises all persons whose legal domicile on December 31 is Finland. All the basic demographic characteristics of the population (identification number, sex, age, marital status, place of birth etc.) are derived from the Central Population Register.

**The Register of Buildings and Dwellings** is a basic register defining all buildings and dwellings. It also contains the characteristics of buildings and dwellings. The basic data of the register were collected in the 1980 population and housing census and they have been maintained since by Population Register Center in which the register is situated.

**The Statistics Finland's Register of Enterprises and Establishments**, which is not an administrative register, but which has properties of the same kind. For example branch of industry, ownership type, juridical form and location of the workplace are characteristics which are linked to employed from the Register of enterprises and Establishments and from the **Register of Public Associations and Local Government Functional Units**.

Besides these basic registers the census system also uses about 30 other register (annex 2) files, from which data are combined with the census units. These files include:

#### ***The Ministry of Labour's Register of Job Applicants***

This register covers all persons who have applied for work through the employment exchange. The most important data derived from this register are:

- date when unemployment started and ended (each unemployment period of the year)
- reason for end of unemployment
- date when job placement started and ended.

Information of conscripts and conscientious objectors are obtained from **the General Headquarters Register of Conscripts and the Ministry of Labour's Register of Conscientious Objectors**. The data obtained from these registers are personal identification number and date when military service/non-military service started and ended.

The data on employment relationships are received from several different sources. These relationships can be divided into three different main groups: **employment relationships within the private sector (including the self-employed), within the government sector and within the local government sector (including municipal federations)**.

**The Central Pension Security Institute** provides almost all the data on employment relationships within the private sector. It records all employment relationships covering a period of over one month during a year. In addition, it contains the dates when an employment relationship started and ended.

The Central Pension Security Institute provides data of which the following data are the most important:

- name of the enterprise
- date when the employment relationship started and ended
- type of pension scheme
- pension regulation number.

Other sources of employment relationship data within the private sector are **the National Church Board's Register of Employees**, **the Register of Seamen's Pension Scheme (from the Central Pension Security Institute)**.

The major part of the data on employment relationships within **the Central Government sector** comes from the State Treasury register of central government employees. These data comprise for example:

- office number
- number of establishment
- date when the employment relationship started and ended
- occupation.

Other sources of data on employment relationships within the Central Government sector are the registers of **the Social Insurance Institution, the Bank of Finland, Helsinki University and the Post and Telecommunications Office**.

The majority of the data on employment relationships within the local government sector comes from **the Municipal Pension Institution**, which takes care of the register of the employment relationships of the local government and municipal federations. The data received are for example:

- member association number
- date when the employment relationship started and ended
- occupation.

**The Employment Relationship Register of the Government of Åland is another source of information on employment relationships within the local government sector.**

In connection with the work of collecting all employment relationships from different sources, a separate **employment relationship register** is compiled. This register comprises all the employment relationships during the year.

Data on wages/salaries for all employment relationships are obtained from **the Income and Pension Register of the National Board of Taxes** (Employer's statement of earnings and tax deductions). These data are needed when choosing the employment relationship valid at the end of the year.

The data on students are also obtained from a number of sources. **Statistics Finland's Register of University Students** provides information on persons studying at a university during the autumn term or registered at a university, as well as the type of university studies.

**The Register of Study Aid to Students supplies data on** study aid granted during the year (during the autumn and/or spring term).

**The Joint Selection Register** provides information about the students who have started their studies at the senior level in an educational institution during the year in course or the two previous years.

**The Ministry of Education's Register of Students in Senior Secondary Schools and Vocational Schools** supplies data on codes and locations of educational institutions of the students in these schools.

**The Social Insurance Institution's Register supplies information on pensioners.** The data used in employment statistics are the initiation date and type of pension.

**The Register of the National Board of Taxation** provides many different kinds of data on income and benefits as well as taxes, assets and liabilities of the persons subject to taxes.

Statistics Finland's **Register of Completed Education and Degrees** contains all the degrees taken in Finland as well as the majority of the degrees taken by Finnish citizens abroad.

The Register of Completed Education and Degrees provides data on the highest degree for the employment statistics. Since this register contains all degrees that a person has taken, it is possible to use these too. The information describing the degrees taken includes the following:

- educational six-digit code
- date when the degree was taken
- location of the educational institution
- type of educational institution
- code of the educational institution.

## Annex 2. Producing occupational data from registers: background and status quo in 2000

In the 1990 and 1995 population censuses, occupational data were produced from registers. This will also be the case in the population Census of the year 2000. Occupational title data are gathered from diverse register sources and machine-coded using an occupational dictionary. Where machine coding has not been possible, manual coding has been done using computer terminals.

The sources for occupational data in the register-based population censuses have been:

1. Occupational data on persons employed by central government
2. Occupational data on persons employed by local government
3. Occupational data on persons employed by the state church
4. Occupational data on persons employed by the Social Insurance Institution
5. Employer organisations' data on the occupations of persons employed by organised enterprises
6. Occupational data collected with tax returns of entrepreneurs
7. Population Register System's data on the occupations of persons who have moved
8. Occupational data on unemployed persons in the Register of Job Seekers

Source	Amount of Occupation titles taken from the source	The date of occupation	Administrative use
1. Register of the Central Government Employees	150 000	Year 2000	Registration of employment relationships for pensions etc.
2. Register of the local government employees of the Municipal Pension Institution	420 000	Year 2000	Registration of employment relationships for pensions etc.
3. Register of employees by the Social Insurance Institution's	5 000	Year 2000	Registration of employment relationships for pensions etc.
4. National Church Board's Register of Employees	17 000	Year 2000	Registration of employment relationships for pensions etc.
5. Registers of the Employer Organisations	700 000	The last quarter of the year 2000	Information about the affiliated for wage negotiations
6. Local government of Åland Registers and The Bank of Finland	2 000 1 100	Year 2000 Year 2000	Information of employees Information of employees

---

7. Data inquiry Registers of the Employer Organisations	33 000	Year 2000	Information about occupational titles
8. Register of Job-seekers of the Ministry of Labour	30 000	Year 2000, registration date	Information to authorities for finding the job
9. Entrepreneurs, farmers Taxation and work pension systems	290 000	Year 2000	Occupational titles
10. Population Information System of the Population Register Centre (Notice of removal)	30 000	Year 2000, the last move in this year	Occupational titles
11. Register of Job-seekers of the Ministry of Labour	30 000	Year 2000, registration date	Information to authorities for finding the job

---