2015 European Establishment Statistics Workshop
7-9 September 2015, Poznan, Poland

Title of the paper*: Improved consistency over unit types in an integrated business statistics production system*
Author: Ms Hanna Heljala, Senior Statistician
Organisation: Statistics Finland, Business Register / Business Data Support

Table of Contents

1 Introduction

In this paper four different methods of improving consistency over unit types within a new integrated production system are introduced. The methods – presented under chapter 2 – are: Imputation of establishment level data, methodological compilation of local units, integration of enterprise group database and combining the annual publications of enterprise and establishment statistics.

In chapter 1, a brief introduction of Finnish Business Register and the integrated business statistics production system is given in order to provide the reader with the essential background knowledge.

1.1 Finnish Business Register

At Statistics Finland, the Business Register serves as an important data source and a sampling frame to all business and economic statistics and to a variety of population and social statistics. In addition, data – on aggregated and micro[1] level – is regularly delivered outside to various users and interest groups.

Finland has a long history of high quality administrative data and these sources are effectively utilized in the Business Register as well. As we are able to take full advantage of synchronized legal unit IDs with all the relevant authorities, we receive a good basis of information on the legal unit level from administrative sources. The most central source here for the basic Business Register data is the Finnish Tax Authorities database received monthly. The administrative data is further completed with our direct data collection inquiries focusing especially on the establishment (local kind of activity unit, LKAU) level.

1.2 The Integrated Business Statistics Production System

As a result of a large revision project, which has been running since 2010, Statistics Finland has in late 2013 implemented a new integrated business statistics production system. Within the revision project, the whole production process from data collection to data publishing was renewed. Revision included harmonization of methods and practices as well as elimination of overlapping work phases within and between different statistical domains.

The integrated system covers all the central business statistics; Business Register, Structural Business and Financial Statement Statistics, Regional Statistics, International Trade in Services, FATS statistics, Commodity (PRODCOM) Statistics and Short Term Statistics. It provides all the statistics with the same data content in order to produce consistent business statistics. Unit structures, as well as classifications (e.g. NACE) and concepts (e.g. turnover) used in several statistics were further harmonized during the revision project.

The system consists of a production database and a data warehouse. Production database is where all the checking, editing and imputation to the data sets of various statistics is done, whereas all the statistics are produced from the final stage data stored at the common data warehouse. Data warehouse also serves as an input for several other statistics as well as National Accounts.

Previously, business register as well as other business statistics domains operated on their own separate systems. Although business register was utilised as the main source of the units and as a sampling frame, in some cases parallel unit structures, classifications and characteristics data existed within the different domains. The main idea of the new system is that the existence of parallel data is no more enabled and all the corrections are done in one place and only once.

The main statistical units in the system are the enterprise (ENT) and the establishment (local kind of activity unit, LKAU). All the production is mainly based on these units and thus also the work input and especially manual checking is concentrated on these two units. In practice, enterprises are compiled from legal units (LeU)[2] although the unit model allows splitting of legal units (with the help of establishments) as well. All

---

[1] According to the Statistics Act of Finland (280/2004), certain Business Register variables are public on micro level as well.

[2] At the moment, 1 LeU = 1 Ent, but Finland has started profiling exercises in order to implement complex enterprise units in the near future.

legal units received from an administrative source are stored in the system but only active legal units are given an enterprise ID and an establishment with respective relationships and brought to as a part of the Business Register's and Business Statistics' production cycle. Other units in the system are kind of activity unit (KAU), local unit (LU) and enterprise group (EG).

2 Improving Consistency over Unit Types

2.1 Imputation of Establishment Level Data

To improve the quality and coherence of the establishment level data as well as the consistency between the two main statistical units – i.e. the enterprise and the establishment – a new imputation methodology was developed and implemented. The methodology covers Business Register, Structural Business, Financial Statement and Regional Statistics data, which serve as an input to the National Accounts.

2.1.1 Description of the Methodology

Previously Business Register inquired within its surveys the turnover data on establishment level. Turnover is also a part of the Structural Business and Financial Statement data. In order to aim for a higher consistency and to eliminate the overlapping work as well as to lessen the response burden on the enterprises, turnover data is no more part of Business Register's direct data collection. Main source for the turnover as well as for other Structural Business and Financial Statement Statistics' variables is the Business taxation data received from the Tax Administrative on a legal unit level. This data is further completed with surveys.

The basic data for the establishment (location (address) data, start/cessation date, activity code (NACE)) as well as the employment are inquired in the Business Register surveys, which cover all multi establishment enterprises. This serves as an important precondition for the establishment level estimation of the turnover as well as a set of other financial statement based variables. In case of non-response in the Business Register survey, the establishment level employment is estimated utilizing an administrative data based estimate compiled on legal unit level and establishment level data from the previous year.

The idea of the method is to divide the enterprise level data to the establishments utilizing the establishment level employment as well as an industry specific co-efficient. Industry specific coefficients are calculated by a simple regression analysis utilizing information on enterprise level employment and turnover among the mono-active enterprises. Other information utilized in the estimation process are e.g. the type of the establishment and in case of ancillary activities, data on ancillary NACE code. Variables covered by this method are e.g. turnover, the breakdown of income and expenditure as well as variables referring to current and fixed assets.

In the first phase of the process, the data is produced to the establishments independently of the enterprise level data, i.e. based solely on the establishment level employment and activity data. In the second phase, the data produced on the establishment level is summed up to the enterprise level and the possible difference between the sum and the enterprise level data is scaled to the establishments in order to ensure the coherence of the data.

2.1.2 Further Challenges

Since the method relies on the employment data, establishments with zero employment are a challenge. There are some specific industries, e.g. NACE 47302 *Automated retail sale of automotive fuel*, in which no employment is bound to exist but other variables should be attributed to the establishment. When these establishments belong to multi-active enterprises, no meaningful estimation method has yet been found and the subject is to be further discussed still. Applying direct data collection for these units is not excluded as they could be integrated into a survey already done on a business unit level with a specific variable inquired.

Further improvements into the calculation of the industry specific coefficients are planned as well. Utilization of the ancillary NACE code as well as information on global activity in the calculation have been discussed and will be further investigated.

2.2 Compiling Local Units Methodologically

In the new system local units are delineated methodologically based on the data on establishments (local kind of activity units) ensuring a full coherence over these different unit types. Previously the local units were a result of manual updating, which was prone to situations of incoherency between the establishment and the local unit level data.

2.2.1 The Methodology in the New System and Comparison to the Previous Situation

The delineation of local units is simply based on the street addresses of establishments under the ownership of a certain enterprise. Additionally, fixed building registry numbers assigned to the establishments are utilised. Local units are only produced to the data warehouse and thus no manual intervention is allowed. Local units are not used as data collection units but they can be utilised for example in sampling or data deliveries.

Compared to the previous system, there is no significant impact on published statistics as the publication has been based on establishments, but the key improvement is in the simplification of unit structures and updating processes. There is prominent improvement in the coherence over the whole population and especially in certain situations between the establishments and local units as well.

Initially, as local units were introduced in to the Business Register, they were created from the establishments methodologically, but afterwards the updating was mainly carried out manually. Nevertheless, there were a few automatic rules implemented in the previous system as well. For example, as the address of an establishment was updated, the system would automatically update the new address as the address of the local unit as well. In case of local units with multiple establishments, this sometimes lead to a situation where addresses were not coherently recorded; local unit had an address which did not match with all of its establishments. An example is illustrated in table 1 below.

LeUID	LUID	LU_street_address	LKAUID	LKAU_street_address
0X0XXXX-X	4071242237	Hansakartano 4	708379788	Terveyskuja 2
0X0XXXX-X	4071242237	Hansakartano 4	643593865	Hansakartano 4

Table 1: Different addresses between LKAU and LU

Moreover, some situations lead to unnecessary local units being delineated. For example, when the address of an establishment (and respectively of the local unit) was updated, it could easily be overlooked that there already existed a local unit under the new address. An example is illustrated in table 2 below.

LeUID	LUID	LU_street _address	LKAUID	LKAU_street_address
0X0XXXX-0	4000007743	Oritkarintie 6	552092836	Oritkarintie 6
0X0XXXX-0	4000007826	Oritkarintie 6	500208062	Oritkarintie 6

Table 2: Unnecessary local units delineated (within the same street address)

As the concentration of the manual updating was on the main units - the enterprise and the establishment-local units did not get the attention needed to keep the unit structure correctly delineated and their data coherent. Moreover, systematic checking rules or warning notifications to help detect these situations were missing from the previous system.

2.2.2 Challenges of the Methodological Delineation

The fact that the delineation is based on the street addresses of establishments can pose a few challenges.

In case the street address doesn't truly reflect or define the actual location, the delineation does not go correctly. An example is a factory area, in which different street addresses exist but which constitutes a continuous area and should thus be considered as one local unit (also according to the EC Regulation 696/93). Here, the building registry number only recognizes the situations, which are situated in the same building; e.g. a building situated in a corner of a street having several street addresses can be recognized as one local unit with a help of a building registry number.

Standardized Euref-FIN (ETRS89) co-ordinates are stored in the Business Register and attached to the establishments as well. The co-ordinates could in principle be used as an additional help in the delineation as well, but as such they do not so easily fit into the definition of 'simple automatic rules'. In so the additional value obtained from their use in the methodological delineation is limited. Moreover, the cases in which the co-ordinates might be helpful in addition to the variables already used are assumed to be marginal and thus probably not worth the additional work.

Another challenge with the street addresses is that in practice, they are not always fully correct or complete. The address might be incomplete, incorrect or even missing altogether. However, the overall coverage and correctness of the street address data in the Finnish Business Register is very high and in problematic cases sometimes the building registry number can still be utilized.

2.3 Integration of Enterprise Group Data

2.3.1 From the Previous to the New System

Within the previous system, although the enterprise group data was closely linked to the Business Register, it was stored in a separate database located in a different server than the rest of the Business Register. Some tables needed to be copied between the servers and the updating procedures were carried out separately. The manual updating of the enterprise group structures and data was done with a separate software application. The unit structure depicted in the Business Register's software application did not include the enterprise group.

In the new system, the databases have been integrated and their updating procedures are more harmonised. Moreover, the enterprise group is brought to as part of the common software view in so that the full structure of the units in national territory can be viewed at once, which was not possible in the previous system. This enables a better utilisation of the data on enterprise groups in profiling is as well. A screen capture from the software can be seen in picture 1 below.

Tunnus	Tyyppi	Nimi	Toimiala	Kunta	Toiminnan AloitusPvm	Toiminnan LopetusPvm	Osoite	Postinumero	Postitoimipaikka
◢ 1005305 (1)	Kn	KESKO EG	NACE	Municipality			Street Address	Zip	Postal location
◢ 1231672355 (1)	Yr	Kespro Oy Ent	46390						
◢ 0911917-1 (23)	Oi	Kespro Oy LeU	46390	Helsinki				00016	KESKO
502361314	Tp	TURUN TUKKU LKAU	46390	Turku			PITKÄMÄENKATU 11	20250	TURKU
502361439	Tp	LAHDEN TUKKU	46390	Lahti			ALHONKATU 4	15610	LAHTI
502361587	Tp	KUOPION TUKKU	46390	Kuopio			TEHDASKATU 14	70620	KUOPIO

Picture 1. Screen view from the shared software application depicting the unit structure with basic information

Nevertheless, the manual updating of the enterprise groups is still done with a separate software application. The application is used to administrate enterprise group level data, relationship data between the group and the group head as well as relationship data between legal units within the group. The non-resident units of the enterprise group are only visible in the separate software. Within the renewal project, the application was amended to include functionalities for the FATS production as well.

2.3.2 Further Challenges

To spread the knowledge and competence on the enterprise groups and their updating rules to a wider group of experts is a challenge to be tackled in the near future. Profiling exercises will cumulate the know-how and offer new possibilities to participate in the updating of the enterprise groups.

As the manual updating is still done with a separate software application, it will impose further challenges as the implementation of complex enterprise units takes place. It needs to be taken care of that as the enterprise group structure is updated, it will trigger a signal to the user to access the application which is used to update the structure of the complex enterprise as well. This will need to be done by the same user within the same session (e.g. block the progress with the enterprise group until the enterprise data is checked). Within the renewal process, it was also foreseen with a possible need to rethink the integration of the software applications as the enterprise unit is implemented to the Euro Groups Register.

2.4 Publishing Data on Enterprises and Establishments

2.4.1 Previously Several Separate Publications

Previously Business Register had its own production cycle as well as an annual publication covering the basic data on Finnish enterprises, establishments. Structural Business, Financial Statement and Regional Statistics based its own production on the Business Register's frame population and published within its annual publications partly parallel information. Moreover, in certain situations the information given by the different publications was not fully consistent. For example, case specific differences in industrial classification, turnover and personnel data as well as in the treatment of enterprise reorganisations (e.g. mergers) existed. The criteria of activity and the delineation rules of the population base, which is taken into account in the publication differed to some extent as well.

Moreover, prior to the renewal, the publication of Structural Business, Financial Statement and Regional Statistics was divided into a total of ten different publications according to the unit level and different industries. This was not an ideal solution for a user interested in gaining a comprehensive picture of the population as the information was spread between different publications.

2.4.2 New Combined Publications

In the new system, annual data on enterprises and establishments is produced and published only once. There are two publications: i) Structural Business and Financial Statement Statistics (incl. enterprise level data) and ii) Regional statistics on entrepreneurial activity (incl. establishment level data). Both publications contain information from both the Business Register and Structural Business and Financial Statement Statistics with the previous inconsistencies in data content and processes fully harmonized.

The new publication offers its users a more comprehensive and consistent picture of the enterprises and establishments population at once. An example of a table combining establishment level data from the two domains can be seen below in picture 2.

Establishments by region in 2013: Personnel, Turnover and Value added

Region	Establishments	Number of personnel	Turnover	Value added
01 Uusimaa	108 518	528 340	193 643	38 172
02 Varsinais-Suomi	36 819	123 380	25 976	7 229
04 Satakunta	16 910	60 260	14 540	3 887
05 Kanta-Häme	12 408	41 637	8 385	2 664
06 Pirkanmaa	33 878	131 212	29 254	7 937
07 Päijät-Häme	13 904	50 986	10 091	2 834
08 Kymenlaakso	11 455	40 423	9 295	2 602
09 South Karelia	8 992	31 414	7 811	2 393
10 Etelä-Savo	12 937	33 562	5 141	1 593
11 Pohjois-Savo	17 137	56 531	10 267	3 026
12 North Karelia	10 812	34 603	6 713	1 802
13 Central Finland	18 551	62 586	11 978	3 999
14 South Ostrobothnia	18 102	47 686	9 511	2 551
15 Ostrobothnia	14 207	50 033	12 612	3 493
16 Central Ostrobothnia	5 379	17 175	3 802	1 031
17 North Ostrobothnia	24 555	94 240	19 718	5 653
18 Kainuu	4 927	16 300	2 599	708
19 Lapland	12 312	39 677	9 986	2 713
21 Åland	2 798	9 635	2 137	616
ZZ Unknown	2 613	2 062	545	137
SS Total	387 214	1 471 743	394 003	95 041

Picture 2. Combining data from Business Register and Structural Business Statistics