Methods and Challenges of Quality Measurement at Finnish Business Register

Hanna Heljala[1]

Abstract

Quality of a statistical Business Register is of high importance to various users and interest groups. As these often have different and sometimes conflicting interests, also their definitions of quality might not always be fully consistent with each other.

In the paper, methods and practices used at Statistics Finland to measure the different aspects of Business Register quality are presented. Both their advantages and restrictions as well as the challenges encountered in practice are studied.

Statistics Finland has implemented two formal methods and reporting tools to measure the Business Register quality. Firstly, in order to measure the accuracy and timeliness of location address and activity code data, an annual quality control inquiry is targeted at the units not covered by Business Register surveys. Secondly, a tool comparing Business Register data with different administrative data sources to check the completeness and relevance of the Business Register content is utilized. These tools have been found useful in providing valuable information on the state of the Business Register.

In addition, the recent development steps and possibilities to use additional measures to more comprehensively assess the quality of the Business Register are discussed. Here, the focus is also on utilizing the integrated business statistics production system in improving Business Register quality and its measurement.

Keywords: Business Register, Quality measurement

1. Introduction

The quality of Business Register data has different aspects. As for one user, the quality might equal how the content of the Business Register reflects reality while the other one might put more weight on the content being consistent with other statistics – for example SBS – when assessing the quality of the Business Register data.

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. The closest users inside our national statistical institute are the Short Term Statistics, Structural Business Statistics and National Accounts. In addition, data – on

[1]Hanna Heljala, Statistics Finland, P.O. Box 3 A, FI-00022, Statistics Finland, hanna.heljala@stat.fi

aggregated and micro[2] level – is regularly delivered outside to various users and interest groups.

In this paper, the most central methods for the quality measurement of Statistics Finland's Business Register are presented and their future challenges and possibilities are briefly discussed.

1.1 Finnish Business Register

Finland has a long history of effectively utilizing administrative data, which applies to 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 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 (LKAU) level data. For the units outside the scope of direct data collection, a valuable source of addresses on establishment level is the data of Finnish Postal Company updated to the Business Register every other month. Finally, we make use of data provided for us from different statistics, including also more informal methods like feedback as we work in a close co-operation with our most frequent in-house users.

To serve the different users and their needs we have implemented two versions of relevant variables; one that reflects the latest information and one that is frozen for a certain reference year. Variables that have two versions include classification variables like municipality of location, activity code (NACE) and institutional sector (SNA) code. This practice has been found necessary to answer both the needs of all the different statistics and the needs of a wide variety of other interest groups.

Due to certain conceptual and methodological differences, the comparability of our Business Register data with the information provided by other statistics – for example SBS – used to be restricted to some extent. As a result of a new integrated business statistics production system including more harmonized concepts and processes implemented in 2013, the comparing of data between different statistics is more meaningful. In the revision process, also the working processes of business statistics production were reorganized in so that the work from the data collection through the modifying and validating of the data is done in a closer co-operation between the Business Register and different business statistics. For example, the experts checking, modifying and validating the Business Register data now have also expertise in the same process steps with SBS data thus enabling a more thorough understanding of these comparable data sets.

2. Business Register Quality Control Inquiry

The quality control inquiry is used to examine the correctness of industry and location information for small single-establishment enterprises. As the method is direct data collection, correctness is here to be understood as timeliness and punctuality of our data content; that is, how well it reflects reality.

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

2.1 Purpose of Quality Control Inquiry

Quality control inquiry is carried out as a part of Statistic Finland's routine Business Register inquiries and it has been repeated annually since 2003. The inquiry is identical to the one used for the other inquired units and thus the responses are automatically part of Business Register's normal updating procedures. The population is restricted to units not in the scope of normal annual inquiries of Business Register, which means that the units are small in size and classified as single-establishment enterprises.

The report presenting the results of the inquiry is compiled mainly for in-house purposes. The report examines the results by industry giving the possibility to tackle specific issues and when needed, also to target additional inquiries to certain industry sections. In addition to Business Register staff, the results are of specific interest for the business statistics experts engaged in the micro data services, who regularly utilize Business Register industry and location information at their most accurate level and deliver these data subject to a charge. Thus, in addition to the overall Business Register quality, the benefit from this method of quality measurement is realized especially in the micro data services.

2.2 Main Results of Quality Control Inquiry

The results of the quality control inquiry can be generalized to non-inquired units falling outside the scope of Business Registers' normal annual inquiries. The following results represent the latest quality control inquiry carried out in 2015.

On the basis of the inquiry's results it can be estimated that among units outside the normal Business Register annual inquiries, the section level industry code is correct for 98 per cent and the subclass level (national 5-digit level) is correct for 97 per cent of the enterprises. The industry code was most often reported to be incorrect in the industries M *Professional, scientific and technical activities* and N *Administrative and support service activities*. The 5-digit level industry code correctness stayed at the same level as reported in the previous quality control inquiry, but a little larger amount of the changes applied at the section level than in the previous inquiry.

Figure 1 depicts how the industry code correctness has developed during years 2003-2015. The improved quality is especially evident at the 5-digit level NACE codes. Firstly, the improved quality can be seen as a result of the efforts made to promote the correct and coherent application of NACE classification inside our national statistical institute. This has been done through an expert group consisting of representatives from different statistics and a close co-operation between them enabling Business Register to gain from their knowledge as well. Secondly, the positive development is due to a change made in the organizational structures of the Finnish administrative authorities, which has improved the timeliness and accuracy of the initial industry code data received to the Business Register.

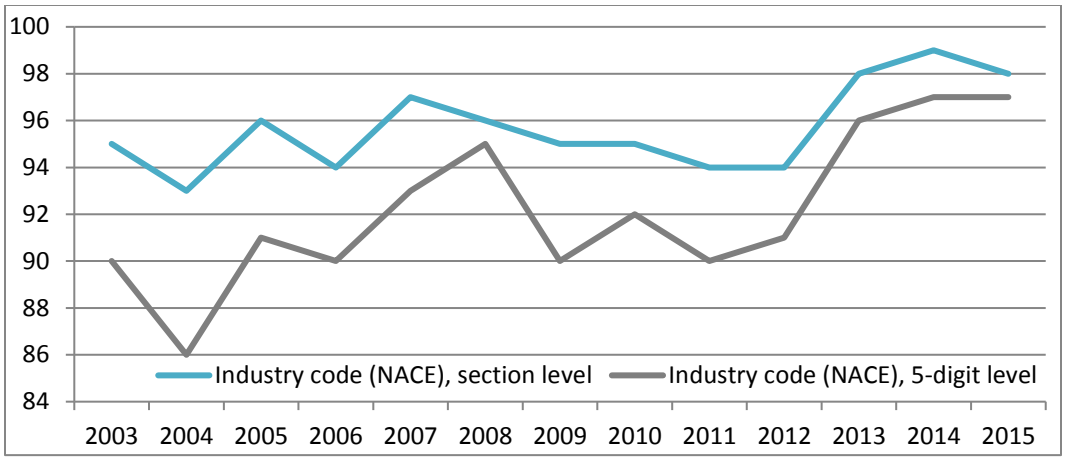

Figure 1. Industry code (NACE) correctness (%) among small, non-inquired enterprises.

The location address *(street address)* is estimated to be correct for 88 per cent of the small non-inquired enterprises. It should be pointed out, that the street address is not as such published in any official statistics, but the variable is nevertheless often used in the micro data services. The municipality of location is estimated to be correct for 100 per cent of these enterprises. Regarding the different location data variables, the industry section where changes were reported most frequently was *Q Human health and social work activities*.

Figure 2 describes the time series for the correctness of both the municipality of location and the street address. Correctness of the municipality of location has been on a very good level throughout the period under review. The quality of street address data has also significantly improved during the period. There are several reasons for the improvement. The first reason is the more efficient use of the public data sources available for street address, whereas the second one is the close co-operation with the micro data services providing us with valuable feedback. The third reason lies in the targeted checks completed for small units that have not been updated for a certain period of time. This has been made possible recent years by a more efficient way of working, providing us with a chance to re-target resources.

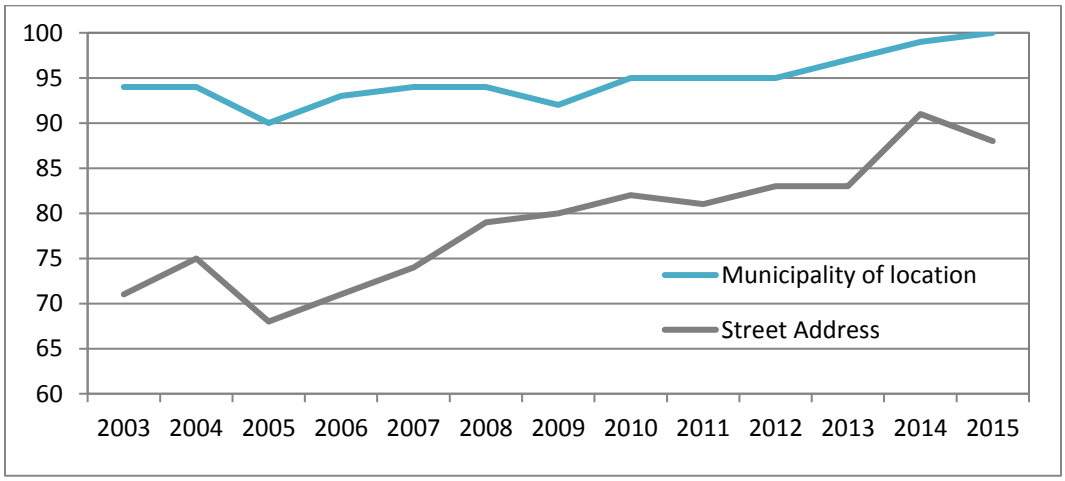

Figure 2. Correctness (%) of street address and municipality of location among small, non-inquired enterprises.

2.3 Further Challenges

Although the results of the quality control inquiry as such are encouraging, it is not completely without its challenges. One such challenge can be seen on the figure 3 depicting the response rates from year 2003 onward.

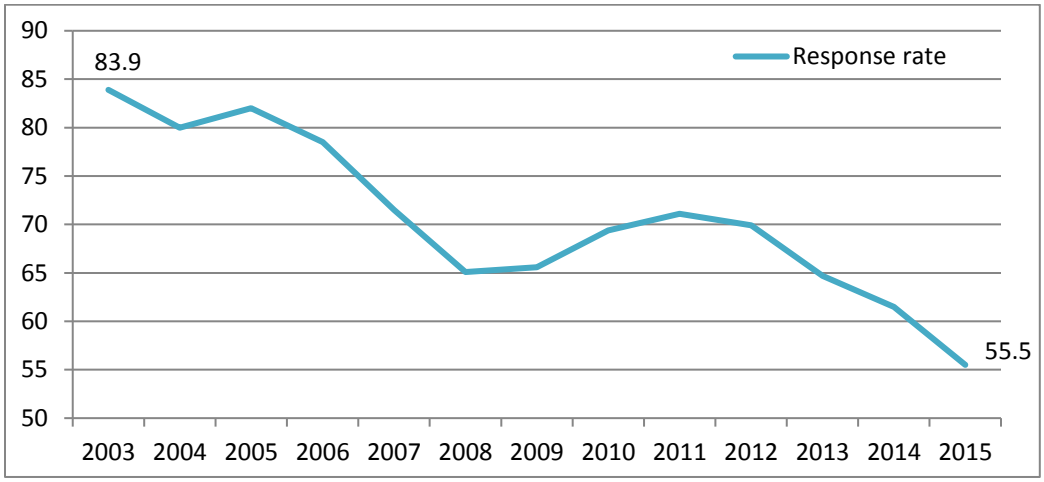

Figure 3. Response rates (%) of the quality control inquiry for years 2003-2015

In 2015, 1 666 enterprises were selected to the actual inquiry. 925 enterprises responded the inquiry leading to a response rate of 55.5%. In the previous quality control inquiry the response rate was 61.5%. The response rate was at its highest year 2003 (83.9%).

The descending response rate poses a risk for the possibility to generalize the results. To tackle this challenge the first methods would be either further motivation of the respondents or enlarging the sample size. However, here the real issue lies in the fact that these units are small and as such they are not just hard to motivate but also relatively insignificant when seen from the statistical point of view. Thus when examining the costs of such measures, they would most likely exceed the benefits gained. However, analysis of the non-response of Business Register surveys in general is planned to be carried out in the near future. As part of this work the non-response of the small units will be analysed to some extent as well.

Another challenge is more related to the nature of the new ways of doing business; that is the recent and ongoing trend of businesses' diminishing dependence on the physical location. This is especially evident on industries where more and more businesses are operating on a virtual platform. As such it poses a completely new challenge on defining, delineating and monitoring the actual location of the business that more often is not equal to its physical location anymore.

Additionally, when compiling the report for quality control inquiry's results, the data from direct data collection is assumed to be 100 % correct. This can be a little bit challenging assumption as respondents might not always understand the concepts or the reference time periods for the data fully correctly. However, automated checks as well as manual editing are extensively utilised to help the issue.

3. Coverage Analysis of the description area

3.1 Description Area and Coverage Analysis

According to the regulation on Business Registers *(Council Regulation (EEC) No. 177/2008)*, all enterprises carrying on economic activities contributing to the gross domestic product (GDP) should be gathered in a Business Register. It is pointed out that, non-market services contributing to the GDP, as well as direct and indirect holdings of active legal units shall be regarded as economic activities for the purposes of Business Registers. The description area of the Finnish Business Register has been built upon the following data received from administrative sources: i) Regular employers, ii) Units liable to value added tax and iii) Units entered in the prepayment register.

Units that fall outside the scope of the Finnish Business Register's description area have nevertheless been monitored regularly to ensure the completeness and relevance of our data content. The monitoring is carried out by comparing the Business Register data with the most central administrative sources containing data on legal units. As the administrative data sets compared with the Business Register data are the ones also used as sources in our regular updating process, coverage analysis can be seen as a tool for testing the proper functioning of our system updates as well.

The main target groups of the analysis are Business Register staff, business statistics micro data services and National Accounts. Whereas the micro data services' main interest is especially in the new units, the National Accounts makes use of the examination by legal form. The results of coverage analysis have served as an impulse for adding units of certain legal form – for example foundations and funds – to the description area of Business Register to better serve National Accounts' needs.

Coverage analysis of the description area was carried out annually since the statistical year 2002. After 2012 and the transfer to new integrated business statistics production system, the frequency of the analysis has been reduced as during recent years no more significant inadequacies have been found with the analysis.

3.2 Main Results of Coverage Analysis

The Tax Administration's customer database is the Finnish Business Register's main source for maintaining basic information on legal units. The coverage analysis tool searches for all the units in the Tax Administration's business taxation data, payment control data and annual wage and salary data that were missing from the Business Register's description area in the reference year.

As a result of cross checking the different data sets with an elimination of overlapping, a total of 8 376 units were found to be absent from the Finnish Business Register's description area in the statistical year 2012. When examined by legal form, the largest groups in number that were missing from the Business Register were *Housing and real estate corporations* (3 824) and *Foundations and funds* (609).

However, relative to the Business Register's frozen data for the reference year 2012, the turnover and sum of wages and salaries proportions were only 0.01 per cent and 0.02 per cent respectively. The results of the coverage analysis of the Business Register description area indicate that both the completeness and the relevance of the Finnish Business Register are very good.

Table 1 presents the aggregated results in number of units and as proportions of the turnover and the sum of wages and salaries for the reference years 2002-2012. The figures show the improvement of the Business Register coverage during the years under review. Especially, when comparing the results from reference years 2006 and 2005, the inclusion of new NACE sections in to our Business Register coverage as a result of the revised regulation can be seen in the improved figures. In addition, development of the conceptual model of birth and death of a unit in the Business Register in so that it better reflects the reality is seen in the improved coverage of the number of units.

Table 1. Aggregated results of coverage analysis of the description area for reference years 2002-2012

Reference year	*Units not in BR*	*Turnover proportion (%)*	*Sum of wages and salaries proportion (%)*
2002	44 417	0.6	0.3
2003	42 777	0.3	0.4
2004	46 355	0.4	0.4
2005	35 393	0.3	0.2
2006	25 236	0.04	0.2
2007	23 745	0.01	0.1
2008	23 791	0.01	0.2
2009	10 772	0.04	0.09
2010	6 943	0.02	0.03
2011	7 524	0.01	0.01
2012	8 376	0.01	0.02

3.3 Further Challenges

According to the regulation on Business Registers *(Council Regulation (EEC) No. 177/2008)*, direct and indirect holdings of active legal units belong to the description area of a Business Register. However, the regulation does not define "activity". According to the Business Registers Recommendations Manual (2010), activity of a unit in Business Register should not be seen just as the existence of turnover or employment but also to take into account whether the unit is part of an enterprise group.

Earlier some units in the financial sector have fallen outside the scope of the Finnish Business Register's coverage as they are pure holdings not having employment or turnover. When being part of an enterprise group, they are nevertheless important in completing the group's ownership structure. Thus, they have been stored in the enterprise group register database closely connected to the Business Register. This has, however, caused incoherence between the two registers' data contents.

As a result of the new integrated business statistics production system, the coverage of Business Register is further expanded. This is due to the fact that also all the units with a balance sheet total not equal to zero are regarded as active. Previously the variables used to indicate activity of a unit were just employment and turnover. In addition, all units belonging to an enterprise group are visible on the common software application to all users, no matter whether they are regarded as active or passive.

4. Recent Development Steps and Future Possibilities

4.1 Integrated Business Statistics Production System

Statistics Finland has since late 2013 implemented an integrated business statistics production system as a result of a large revision project since 2010. 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 consists of a production database and data warehouse, where all the central business statistics are connected. All the checking and editing to data sets of various statistics is done at the production database, whereas all publications 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.

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 main statistical units in the system are the enterprise (ENT) [3] 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. All legal units (LeUs) 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).

4.1.1 Quality Checks, Quality Coding and Metadata

Various quality checks are integrated throughout the process. For example, during the reception process for administrative data, the data goes through a technical validation based on the information stored in the Exist meta-database and the individual variables are checked for their correct data form and validity. In the electronic questionnaire application utilised in our Business Register surveys, the quality checks include, for example, that empty fields are not allowed and that the data needs to be in a certain format. In the updating processes of both administrative and survey data the quality checks performed include, for example, checks for the data form and logical checks. During the editing processes comparison of estimates and survey data as well as detection of big changes in the data are checked and marked.

Various quality codes are utilized as well. In the production database they guide the manual editing processes enabling us to target our resources and in the data warehouse they provide additional information to the users of the data. In the data warehouse, the quality coding

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

can be utilized for example in the sampling, which can be targeted to take into consideration the quality codes.

To manage the extensive production system, an XML-based statistical metadata system is utilized. The metadata system provides both data on the process management, as well as data on the data content itself; e.g. names and descriptions of variables and validation rules. A uniform process management application further improves transparency of the data process, reduces person dependency and offers new possibilities to divide work.

4.2 Quality Measurement as Part of Everyday Work

The new business statistics production model offers us a chance to not only improve the quality of our data content but also to bring the measurement of the quality closer to everyday work.

As all the most central business statistics are working on the same common platform and same experts are able to check and modify a combination of Business Register and business statistics data, the comparing of data between Business Register and business statistics becomes a more natural part of everyday work. In the integrated system experts are able to compare the Business Register data with all the other relevant, comparable statistics either already "on the spot" at the production database or when analysing the data on the common data warehouse.

Image 1 below presents a view of the software that can be used commonly by the experts. The view is from a page called 'a consistency report', which is a tool utilized especially for the large and complex units but also available for all units. The report depicts sum of wages and salaries data received from different administrative sources or different statistics. With this built-on report the user can easily compare the different figures and evaluate their correctness. The validation and corrections needed are done using the same software connection. Same kind of views are available also for other relevant variables as the common system enables the building of summary views and reports in a whole new way between different statistics and sources.

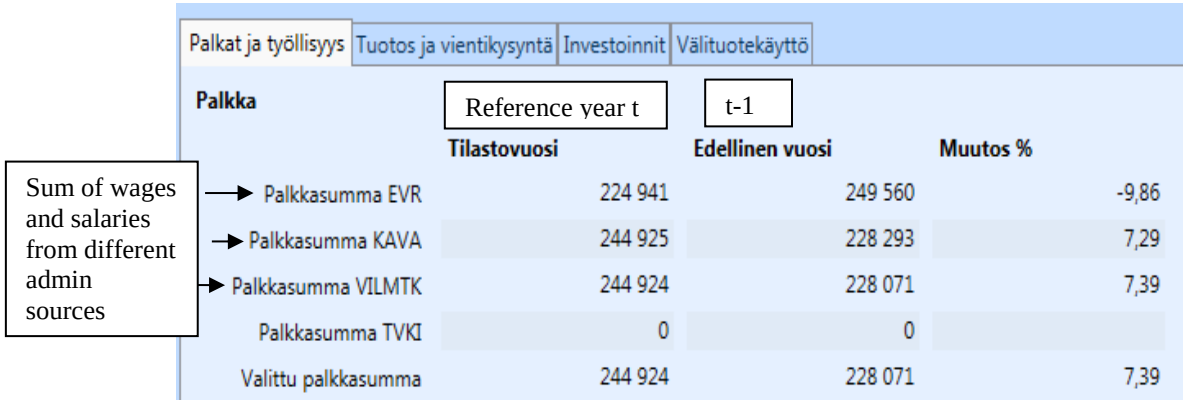

Image 1. A consistency report view depicting data from different sources. Fictive example.

As the concepts used are more harmonized, the comparison of the information provided by different statistics is more meaningful and convenient. Working on a common platform also encourages other more informal methods of quality control and measurement as the

quality of Business Register data is more genuinely a common interest between different statistics.

5 Conclusions

Quality of a Business Register has different aspects to be measured. Standardized reports make a sound and solid basis for quality measurement. These are suitable especially for situations when the register is new or changes have been made to the system. When conducted as a routine for several years, they give valuable information of the quality and its progress on time series level. Following this, also more informal methods should be utilized and brought to as part of everyday work.

Statistics Finland utilises standardised reports as well as more informal methods to measure the quality of its Business Register. Quality control inquiry gives a sound measurement for the punctuality and timeliness of Business Register's location and industry information as well as an opportunity to make full use of the data collected through the inquiry. Using a coverage analysis tool of the description area the completeness and relevance of Business Register data as well as the functioning of the system updates is monitored.

Both of the standardised reports are mostly related to measuring the quality of administrative data updated more or less automatically in to the business register. This is justified as the administrative data covers the majority of the information utilised in the business register. Nevertheless, one might argue that the quality should also be measured against the more substantial matters, which could also be related to e.g. the level of manual editing performed. A more thorough analysis of the non-response of the surveys would be one justified further improvement and is also planned to take place in the near future as well.

The implementation of an integrated business statistics production system has brought several improvements to Business Register quality and its control. The methods and concepts have been harmonised and overlapping work phases reduced between different statistical domains. Moreover, quality checks, quality coding and metadata are utilised more comprehensively than before.

According to Statistics Finland's experiences, the more informal methods of quality measurement can be encouraged by ensuring a close co-operation with the most frequent users of the data. Experiences from the integrated business statistics production system show that especially when operating on a common platform and utilizing a common database with shared concepts and methods, also the quality of the database, i.e. the quality of the Business Register, is ensured to be of a common interest. By entrusting the experts with knowledge of both Business Register and other relevant business statistics data, the quality control of the Business Register data can also be brought to as a more natural part of everyday work.

References

European Commission 2010: *Business Registers - Recommendations Manual*.

Heljala, H 2014: *Methods and Challenges of Quality Measurement - Finland's Experiences*. Paper presented at the 24th Meeting of the Wiesbaden Group on Business Registers - International Roundtable on Business Survey Frames, 15-18 September 2014, Vienna.

Heljala, H 2015*: Improved Consistency over Unit Types in an Integrated Business Statistics Production System*. Paper presented at the 2015 European Establishment Statistics Workshop, 7-9 September 2015, Poznan, Poland.