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. 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.

In this paper, methods and practices used to measure the different aspects of quality of Finnish Business Register and the challenges of quality measurement encountered in practice are presented. In addition, the paper describes the possibilities for improving both the quality and its measurement on our recently implemented integrated business statistics production model.

Already from year 2003 onward, an annual quality control inquiry has been part of our regular processes. The inquiry is targeted at the units that are not covered by our direct data collection. Quality control inquiry is a good tool to measure the accuracy and timeliness of our location address and activity code data. Other methods we have implemented include a tool used to check the relevance and completeness of our register content. The tool compares our Business Register data with different administrative data sources and is especially useful to check the information concerning on one hand new units and on the other hand units that have ceased their activities. In addition to these reporting tools, more informal methods – like direct feedback both from other in-house statistics and from our outside-NSI users – play an important role in our quality measurement.

More recently, we have implemented an integrated business statistics production system providing us more intensive co-operation and more consistent data content with our most important in-house users. In this process, our business statistics production practices were renewed as well, enabling our statisticians to gain in more versatile expertise over different business statistics. Thus, the comparison between the Business Register and other business statistics data is not only more reasonable and more convenient but also a more natural part of the day-to-day work in practice.
# Table of Contents

Abstract ......................................................................................................................................................... 1  
Table of Contents .......................................................................................................................................... 2  
1. Introduction ............................................................................................................................................... 3  
   1.1 Building Blocks of Finnish Business Register’s Quality ................................................................. 3  
2. Business Register Quality Control Inquiry ............................................................................................... 4  
   2.1 Purpose of Quality Control Inquiry .................................................................................................... 4  
   2.2 Main Results of Quality Control Inquiry ............................................................................................ 4  
   2.3 Future Challenges ............................................................................................................................... 6  
3. Coverage Analysis of the description area ................................................................................................ 7  
   3.1 Description Area and Coverage Analysis ........................................................................................... 7  
   3.2 Main Results of Coverage Analysis .................................................................................................... 7  
   3.3 Future Challenges ............................................................................................................................... 8  
4. Recent Development Steps and Future Possibilities ................................................................................. 9  
   4.1 Integrated Business Statistics Production System ............................................................................... 9  
   4.2 Quality Measurement as Part of Everyday Work .............................................................................. 9  
5 Conclusions .............................................................................................................................................. 10
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 – both aggregated and on individual unit 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. Also their future challenges and possibilities are briefly discussed. The paper does not handle the issue of standardized quality reports that are produced as part of every official statistical publication as it concentrates on the quality reports specifically designed for and used in the Business Register.

1.1 Building Blocks of Finnish Business Register’s Quality

Finland has a 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 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 – has so far been restricted to some extent. As a result of our newly implemented integrated business production system including more harmonized concepts and processes, the comparing of data between statistics is more meaningful. In the revision process, also the working processes of business statistics production were reorganized. Now the work from the data collection through the modifying and validating of the data is done in a closer co-operation between 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.

2.1 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 Finnish 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 the 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 also 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 2013.

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 96 per cent of the enterprises. The industry code was most often reported to be incorrect in the industries P Education and G Wholesale and retail trade. The industry code correctness was overall higher than reported in the previous quality control inquiry.

Graph 1 depicts how the industry code correctness has developed during years 2003-2013. 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 administratives, which has improved the timeliness of the initial industry code data received to the Business Register.
Graph 1. Industry code (NACE) correctness (%) among small, non-inquired enterprises.

The location address (street address) is estimated to be correct for 83 per cent and the postal code of the location address for 91 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 correct for 94 per cent of these enterprises. Regarding the different location data variables, the industry section where changes were most frequent was Q Human health and social work activities.

Graph 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.

Graph 2. Correctness (%) of street address and municipality of location among small, non-inquired enterprises.
2.3 Future 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 graph 3 depicting the response rates from year 2003 onward.

Graph 3. Response rates (%) of the quality control inquiry for years 2003-2013

In 2013, 1 561 enterprises were selected to the actual inquiry. 1 010 enterprises responded the inquiry leading to a response rate of 64.7%. In the previous quality control inquiry the response rate was 69.9%. 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.

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.
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 are nevertheless 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.

Coverage analysis of the description area is carried out annually since the statistical year 2002. 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 a trigger 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.

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

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

3.3 Future 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 Register manual, 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 will be 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. So far, the variables used to indicate activity of a unit have just been employment and turnover. In addition, all units belonging to an enterprise group are visible on the common software to all users no matter active or passive.
4. Recent Development Steps and Future Possibilities

4.1 Integrated Business Statistics Production System

Statistics Finland is currently in the implementation phase of a large revision project of its business statistics production system that has been running since 2010. The project plan, contents and targets of the project were presented 2010 in the 22nd Wiesbaden Group meeting by Mr. Sami Saarikivi. The central principles of the data warehouse were also included in the presentation on Business Register and Metadata held by Ms. Tuula Viitaharju in the 23th Wiesbaden Group meeting 2012.

The integrated system consists of production database and data warehouse, where all the central business statistics are connected. All the checking, editing and imputing to data sets of various statistics is done at the production database, whereas all the statistics are produced from the final stage data stored at the common data warehouse. In the first phase Business Register, Structural Business Statistics, Regional Statistics, International Trade in Services, FATS statistics (inward and outward), Commodity (PRODCOM) Statistics and Short Term Statistics are integrated in the system.

The integrated system provides all the statistics with the same data content in order to produce consistent business statistics. The concepts used in Business Register and several other business and economic statistics – such as turnover and NACE classification – were further harmonized during the revision process.

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; depending on the working role of the expert in question. This provides us with completely new opportunities for both the quality and its measurement when compared to the previous situation with Business Register data on one and other business statistics each operating on their own system. Data warehouse also enables its users to make their own frozen versions linked to – for example – a certain statistical publication making it easy to identify exactly the frame used.

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 each their own combination of Business Register and one other business statistics data, the comparing of data between Business Register and other business statistics becomes a more natural part of everyday work. The concepts used are more harmonized making comparison of the information provided by different statistics more meaningful and convenient. The experts are more entrusted with knowledge of different statistics and they are able to compare and correct the data in the production database using a common software. 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.

Image 1 presents a view of the software that can be used commonly by the experts. The view is from a summary page depicting sum of wages and salaries data received from different administrative sources attached to the legal unit in question. In the summary page 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.

![Common software view depicting a summary page with data from different sources. Fictive example.](Image 1)

### 5 Conclusions

Quality of a Business Register has different aspects to be measured. Official reports compiled in a due form 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 has implemented both formal and 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 is monitored as well as the functioning of the system updates.

According to Statistics Finland’s experiences, the informal methods of quality control can be encouraged by ensuring a close co-operation with the most frequent users of the data. First experiences from the new integrated 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 content 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 can also be brought to as a more natural part of everyday work.