# Establishing Harmonised Business Statistics Databases in the Nordic NSIs – challenges and achievements

*Peter Bøegh Nielsen, Statistics Denmark (corresponding author),* pbn@dst.dk

*Andreas Poldahl, Statistics Sweden*, Andreas.Poldahl@scb.se

*Henri Luomaranta, Statistics Finland,* *henri.luomaranta@stat.fi*

*Jelle van der Kamp, Statistics Denmark* jlk@dst.dk

*Kalle Emil Holst Hansen, Statistics Denmark* khs@dst.dk

**Abstract**

*Business statistics is characterised by being produced in a stovepipe system; each statistical domain in isolation fulfilling specific requirements. The requirements are very often based on different EU regulations, e.g. on short term statistics or structural business statistics. But user needs often require the combination of information from different statistical registers to address specific issues often being of cross-cutting nature, especially when analysing factors driving economic growth and the performance of enterprises. These requirements have put an increased focus on the need for partly more disaggregated data than the statistics traditionally published by National Statistical Institutes (NSIs) and partly for existing data combined in new ways.*

*For these purposes, the business statistics departments in the Nordic NSIs (Denmark, Finland, Island Norway and Sweden) have established harmonised databases holding a selection of harmonised and identical variables from different statistical registers at enterprise level. The underlying idea is that by matching information from different registers at micro level (enterprise), new information can be obtained without having to launch new surveys adding extra burden on the responding enterprises.*

*The databases are used for carrying out distributed micro-data linking for analytical purposes by means of centrally scripted syntax, run locally by each NSI. This approach allows for the design of harmonised, tailor-made statistical output across the Nordic countries for comparative analysis, without micro-data leaving the NSIs. The trusted partnership and the harmonised databases have gradually been established over the past nine years through varies projects and face-to-face meetings.*

*The databases contain data from a number of statistical registers, for instance Structural Business Statistics, Business Demography or International Trade in Goods and Services statistics. By establishing harmonised national databases, it has been possible for the Nordic NSIs, in a cost-efficient way to produce a number of analyses for external users such as the Nordic Council of Ministers and OECD addressing issues on the political agenda.*

This paper presents the organisation and implementation of harmonised databases within each Nordic NSI; the design and production of harmonised syntax; examples of the output achieved; and future potential of Nordic co-operation.

**Keywords**: Business statistics,internationally harmonised databases, micro-data linking, enterprise heterogeneity

## Introduction

The National Statistical Institutes (NSIs)[[1]](#footnote-1) are increasingly confronted with the challenge of balancing (new) user needs for business statistics covering emerging issues, such as internationalisation of the production arrangements and structures of the enterprises with the requirements from policy makers to reduce the respondent burden on enterprises. In this paper, we argue and give examples of how establishing databases consisting of variables from different statistical registers at enterprise level is a very powerful tool in meeting this challenge. By linking information from different statistical registers at enterprise level (so-called micro-data linking), it is possible to gain new knowledge about enterprise performance and illustrate the importance and impact of enterprise heterogeneity such as size, group relation or export status.

Introducing micro-data linking (MDL) has a number of advantages. Firstly, the method addresses the issue of establishing new knowledge without increasing the respondent burden on enterprises. Secondly, the MDL approach increases the return on investment for the existing statistical production. Thirdly, attempts to analyse causalities requires multiple points of departures and linking of different types of information at enterprise level. This clearly demands breaking through the traditional stovepipes of statistical production. Finally, this approach constitutes an agile and flexible tool for the NSIs to produce data on emerging policy issues in a timely manner.

Databases holding information from different statistical sources at enterprise level can be established at national level but by co-ordinating and harmonising the contents of the databases internationally, an extra dimension of comparability across countries and value added can be achieved. The business statistics departments in the Nordic NSIs have established harmonised databases in each statistical office. The databases are used for carrying out distributed MDL for analytical purposes by means of centrally scripted syntax, to be run locally by each NSI. This approach allows for the design of harmonised, tailor-made statistical output across the Nordic countries for comparative analysis, without micro-data leaving the NSIs. The trusted partnership between the NSIs, and the harmonised databases has gradually been established over the past nine years through varies projects and face-to-face meetings. It is important to emphasise that the financing of the databases mainly have been covered by projects commissioned by external partners for instance by the Nordic Council of Ministers.

The paper is divided into three main parts: the first one deals with establishing the foundation for MDL (sections 2-3) and in the second part a selection of results from utilising the MDL database is presented (section 4). Lastly, section five concludes the paper with a look on future challenges and goals for micro-data linking within business statistics.

## Organisation, data-sources and structure

The Nordic NSIs are particularly well placed to establish micro-level databases. This is founded in an already established and well proven production system for business statistics. In addition, the integrated use of unique identification numbers for enterprises across a range of different statistical and administrative registers makes information linkable. Finally, the statistical business register acts as the backbone of the production of any of the business statistics. All Nordic NSIs make intensive use of the underlying administrative sources and registers, covering a range of relevant variables at enterprise level, e.g. turnover (from VAT registers held by the tax authorities and based on the VAT declarations from the enterprises) and employment (from Tax registers based on salary information from the enterprises). By linking registers, a more holistic view of the enterprise can be achieved, based on a large number of observed variables at enterprise level, especially due to the use of administrative registers.

The business statistics departments have established an infrastructure that allows for micro-data analysis and work in close co-operation utilising the statistical registers in the Nordic NSIs. The national databases contain a range of sources and currently cover the reference period 2008-2016. The registers and their information are all linked by unique enterprise ID-numbers being available in all the statistical registers. The sources and their main contents are:

• *Business Register (BR)*: covering all enterprises and holding information about identification number of the enterprise, enterprise group identifier (where applicable), activity code, legal form and other basic structural enterprise variables. This forms the backbone of the database and is used to link the different statistical registers.

• *Structural Business Statistics (SBS)*: along with the BR considered the backbone of the database. The statistical register contains key variables on employment, turnover, and value added for enterprises within the scope of SBS. The SBS covers the non-financial business economy (NACE B to S, except K) and determines the activity coverage of the database, see box 1.

|  |
| --- |
| Box 1: Delineation of the non-financial business economyThe non-financial business economy consists of the NACE Rev.2 sections: Manufacturing, Electricity, gas, steam and air conditioning supply, Water supply, sewage, waste management and remediation activities, Construction, Wholesale and retail trade, Transportation and storage, Accommodation and food service activities, Information and communication, Real estate activities, Professional, scientific and technical activities.. |

• *Inward Foreign Affiliates Statistics (IFATS)*: contains foreign controlled enterprises within the country. The statistics enables comparisons and analyses of performance of domestically versus foreign controlled enterprises,

 • *Outward Foreign Affiliates Statistics (OFATS)*: contains information (employment and turnover) about the domestically owned affiliates abroad by geographical location. The statistics enables analyses of the global reach and activities of domestically owned enterprises.

• *International Trade in Goods Statistics (ITGS)*: contains imports and exports values for goods by detailed product categories and geographical destination/origin for each enterprise. The statistics enables analyses of trade patterns of different types of enterprises and their global reach.

 • *International Trade in Services Statistics (ITSS)*: contains imports and exports values by service product categories and geographical destination/origin for each company. This data source is sample-based, and the service product categories are not as detailed as is the case in the ITGS data. The statistics enables analyses of services trade patterns of different types of enterprises and also allowing for analyses of the possible mix of goods and services exports/imports by combining information from ITGS and ITSS.

• *International sourcing and organisation by business functions*: contains selected variables from ad hoc surveys on offshoring and related job effects by business functions and geographical areas in the periods 2009-2011 and 2014-2016. The statistics enables analyses of the impact of offshoring on domestic employment and the global organisation of the value chains.

• *R&D statistics and Community Innovation Survey (CIS)*: contains variables such as total, intra- and extra-mural R&D costs, indicators of different types of innovation activity. The statistics enables analyses focusing on the activity, scale and organisation of these types of activities by different types of companies. This enables the analyses of the performance of R&D intensive enterprises.

• *Business demography (BD)*: contains variables such birth, death and survival of enterprises. In the database, the BD is used as a source for defining the enterprise age, thus allowing carrying out analysis with an age variable that is, for example, clean from non-real entries resulting from administrative restructuring. This enables analyses of business dynamics for instance in terms of new enterprises contribution to growth.

**Figure 1: Structure of the databases**

The database contains selected variables from the data sources mentioned above. The Business register (BR) is the backbone, and thus forms the core population of enterprises for analysis. From the BR, information about enterprise identities and group identities can be derived. These are used as the keys between the different statistical registers at micro level. The structure of the database is based on the existing statistical registers and by using the enterprise ID-number as matching variable, the required data for a given project can be extracted in a flexible and cost-efficient way, see section 4

**Figure 2: Organisation of the Nordic micro-data linking (MDL) projects**



A typical internationally co-ordinated micro-data linking project is typically carried out in three separate phases. The first phase involves the construction of the linked micro-dataset. The project coordinators produce standardised guidelines explaining in detail how the datasets in each participating country are to be structured and provide a common code to ensure that identical tables are produced by all participating NSIs. Each NSI records information from all the data sources used in the project into its own national database. These linked micro-datasets are stored locally at the national statistical institutes and are not shared with users outside the national statistical institute.

In the second phase of the project, the dataset is tested for consistency. Although each dataset being used in the project has already been carefully edited as part of the on-going production of business statistics, it is necessary to carry out further checks. This is, for instance, done to ensure that enterprises are represented by the same ID-number across different datasets and over time, as the reporting units used for specific enterprises with a complex organisational structure can, and often do, differ across the data sources in each project. Tests used in this phase of the projects are devised by the project coordinators and implemented locally by the national statistical institutes.

In the third phase of the project standardised statistical output is created in each country to be used as input for descriptive analysis, see section 4 for examples.

1. **Centrally scripted, centrally corrected syntax**

While harmonised databases in the Nordic NSIs are the foundation of each of the outputs presented in the following section of the paper, coherent and comparative results have only come to be, as a result of centrally scripted SAS syntaxes. By centrally scripted we mean code that have been produced at one of the NSIs. This is done to ensure that the already aligned MDL data in each NSI is handled in the same way. Thus, circumventing the possibilities of 1) different code traditions 2) different ideas of concepts and 3) coding errors hampering the effectiveness and comparability of outputs. It also underscores the fact that a micro-data linking project across countries does not have to include sharing of individual micro-data, as the centralised scripts are distributed and run locally, with predefined and harmonised tabular outcome as a result. For the sake of conceptual cleanness, we recommend using at least two programs in the centrally scripted SAS syntaxes approach. Program one is used for aligning and preparing of base data. Program two is used for the production of common output from base data.

The above does perhaps seem a bit simplified, and that is because it is. A centrally scripted syntax does not just come to be. What follows is a short walkthrough of the steps taken by the NSIs to ensure common goal are met and that all NSIs are included in the building process. The output is produced in three steps:

1. Making the (draft) syntax simple enough to do the job intended. While a complex and compact syntax can be a thing of beauty, they are seldom easy to understand, review and mend by others.
2. Do not assume that people can read your mind. Much frustration in syntax collaboration can be avoided if the syntax starts by carefully stating the different steps, as well as the conditions and definitions that needs to be met for it to work. In the same regard, each new snip of code is most transparent when it is accompanied by a small text that outlines its purpose.
3. When the centrally scripted syntax it ready for testing it is distributed to the other NSIs for review and performance. The review goes through three main questions:
	1. Can the syntax run locally without errors occurring?
	2. Is it logically and theoretically correct?
	3. Does it deliver the agreed output?

Depending on how the specific review process plays out, the central mending of the syntax can go through a number of corrections. Often one would do well to place a face-to-face meeting in the latter part of this process to iron out practical and more conceptual issues discovered in the review of the syntax. Afterwards the syntax can be considered mature and ready to produce publishable results locally, aimed at answering the common goal.

1. **Statistical output**

The MDL approach has proven to be a successful analytical strategy in the development of statistical information on enterprises’ development, contribution to growth and their organisation of and integration into global value chains (GVCs). MDL serves as an appropriate method to analyse the current most addressed research questions on:

* enterprise heterogeneity asking ‘how are different types of enterprises contribute to the growth in the domestic economy’ ; and
* cross border activities asking ‘what kind of enterprises are trading’ instead of ‘what do countries trade’, and how the global value chains are organised in term of “what parts of the business organisation move up or down the value chain’

The Nordic NSIs have been the first to develop a bottom-up, collaborative response to the increased policy questions regarding globalisation, building on national data sources at the micro-level. By linking the data sources at enterprise level it is possible to reflect enterprise heterogeneity by identifying not only enterprises by employment size and trading activity (trader/non-trader). MDL have especially added value by integrating information on group status (independent/belonging to an enterprise group) or by nationality of ownership (domestic/foreign owned) or any combination of these enterprise characteristics. In the following, we present selected output from different Nordic projects, where distributed micro-data linking method has been the core approach.

**Figure 3: Enterprise heterogeneity: Key enterprise attributes for international orientation of enterprises**

****

* + 1. *When are SMEs small or medium sized?*

In the literature, Small and Medium sized Enterprises (SMEs) are described as a key drivers for economic growth and job creation (Birch, 1981). Influenced by this research, SMEs are also in focus for enterprise policy shaping combined with the assumption that SMEs have a huge and unexploited export potential[[2]](#footnote-2). The traditional way of identifying SMEs statistically is by the mere size of the enterprise. In a project for Nordic Council of Ministers, the NSIs (Statistics Denmark et al., 2014) have refined the traditional statistical definition of SMEs. The project moves on from only considering the employment size, i.e. enterprises with less than 250 persons employed, to also include information about group dependency and foreign ownership, see Box 2.

|  |
| --- |
| Enterprise group affiliation Enterprises are placed in two main categories, depending on whether the enterprises have group affiliations with one or more enterprises or not. Based on this, the analysis distinguishes between two groups:Independent enterprisesare not part of an enterprise group in the reporting country, e.g. in Denmark. The data sources do not contain information on ownership relations in other countries implying that an affiliate of a foreign enterprise group will appear as an independent enterprise in this analysis, just as domestic group head which only has affiliates abroad will also appear as an independent enterprise here. Dependent enterprises are part of an enterprise group in the reporting country, e.g. in Denmark, being either an affiliate or a resident group head. Affiliation is determined on the basis of controlling influence, i.e. the control of 50 per cent of stocks/votes. |

Based on these two main categories, it is possible to separate the SMEs into two groups; the independent ones and the ones belonging to an enterprise group and as such is a part of a larger organisation. Furthermore, it is useful to distinguish between those SMEs that are part of a large enterprise group, i.e. the total employment of the group accounting to 250 or more employees (measured in full-time equivalent number of employees (FTE)) – and those that are part of a small group, i.e. the total employment of the group being less than 250 FTEs. The SMEs belonging to a small group would still be considered within the SME range. Figure 4 below shows how these categorisations combine into new analytical categories:

1. Independent SMEs: SMEs that are not affiliated to a domestic group
2. Dependent SMEs: SMEs that are affiliated to other domestic enterprises, but where the employment in the domestic enterprise group is less than 250 FTEs
3. Large group SMEs: SMEs that are affiliated to other domestic enterprises, where the employment in the domestic enterprise group is at least 250 FTEs
4. Large enterprises: Large enterprises with employment of 250 FTEs or more are grouped together regardless of group affiliation.

**Figure 4: Analytical categories of enterprises**

This new type of breakdown of SMEs by type of dependency shows very interesting results for policy shaping. Between half and two thirds of the total exports from SMEs in the Nordic countries can be attributed to SMEs belonging to large groups with more than 250 employees or to foreign groups. These SMEs could more rightly be classified as large enterprises as they can be expected to benefit from different types of economies of scale. For instance, through easier access to finance, foreign markets or specific professional skills than independent SMEs, due to the affiliation with other enterprises within the group.

Building on these results, the importance of “true” SMEs for exports for the Nordic enterprises diminishes further as they only account for 26 per cent of total exports of goods in Denmark, 20 per cent in Norway, 12 per cent in Finland and Sweden.

**Figure 5: Employment, gross value added and exports of goods. By size class-group affiliation status. 2014**

|  |  |
| --- | --- |
| Source: | Statistics Denmark, Finland, Iceland, Norway and Sweden; national database on non-financial business sector linked with international trade in goods statistics |

* + 1. *Importance of enterprise groups*

Recently, the performance of large enterprise groups, and the possible effect on the economy has risen, with the introduction of the granular hypothesis[[3]](#footnote-3), where the size of the enterprises is a key transmission mechanism of shocks to the macro economy. This arises because of fat-tailed distribution of enterprises (in terms of size), so that the value added of few big enterprise groups accounts for a proportionally much larger share of economic activity and because idiosyncratic shocks to the most important enterprises are not eliminated due to diversification.

Nordic countries are a perfect example of how large the granular effects can be[[4]](#footnote-4). This is why proper assessment of the economic consequences of such concentration are vital in policy design, because one can understand a large part of economic activity of the host country, and its fluctuations, by focusing only on the activities of few large enterprise groups.

It needs to be underlined that the Nordic large enterprises are typically organised in *enterprise groups*, which may have global reach. This means that the global economic fluctuations will likely have an effect through these large multinationals – and this is why having a detailed view of not only the domestic activities, but also international activities is important.

In an ongoing MDL project, the Nordic NSIs have grouped together activities of the most important enterprise groups and created statistics that have the potential to provide more details on the aspects of concentration (and granular effects)[[5]](#footnote-5). In particular, the new statistics show how the activities of leading groups are underrated in the standard statistics if one only considers individual enterprises as the statistical unit. This is because the controlling rights to productive assets in the economy are even more concentrated than usually understood, because large groups can control multiple affiliates, and because such ownership links between enterprises are usually hidden in the standard statistical output.

Moreover, the new statistical output can provide more insights of large groups’ performance and links to global economy, when they are analysed as a whole, incorporating also the foreign parts to the analysis. This can change somewhat the perceived dynamics of large groups, since for example the employment generating behaviour can be dependent on which parts of the group are observed, possibly due to offshoring.

Figure 6 is an example of an output that demonstrates how only looking at domestic parts might leave important dynamics hidden. The figure shows that the top groups have significant workforce (and global reach) outside domestic borders. This is especially true in Sweden. Moreover, one can observe that the employment creation can be different domestically, or abroad. This is especially true in Denmark, where foreign employment of top groups has increased, while domestic employment has largely stayed the same between 2008 and 2016.

**Figure 6: Top 100 enterprise groups: Employment domestically and abroad**

****

****

****

* + 1. *MDL as input to macro statistics*

This project had the ambitious goal of aligning business data to national accounts concepts by integrating the output derived from the linked Nordic micro-data with OECDs Trade in Value Added database (Statistics Denmark et al., 2017).

The TiVA database follows the practices of the System of National Accounts (SNA), where imports by enterprises are included as direct imports even if they pass through resident wholesale and retail activities first. In other words imports of goods by wholesalers and retailers for subsequent sale without any further processing are not recorded as their imports in the SNA. The same holds for exports of goods that have not been the subject of any further processing by wholesalers and retailers. In the linked micro-data used in this study, trade is matched to those enterprises that are immediately responsible for imports and exports, including to wholesale and retail enterprises. To align with national accounts concepts, the export and import values for the wholesale and retail sector as reported in the linked micro-data were constrained to the levels reported in in Nordic Supply-Use Tables (SUTs), i.e. the ratio of exports or imports in total output. The additional trade (on average about half of what was reported) was subsequently distributed to other sectors in a two-stage procedure by first identifying the products involved (using official national Trade by Enterprise Characteristics data) and then proportionately allocating these products to using (importing) or exporting industries and enterprise types on the basis of information included in the micro-data and national SUTs.

The purpose of the project was to overcome some of the shortcomings of the current TiVA database by introducing enterprise based characteristics to better reflect the heterogeneous nature of GVC integration - including size, e.g. SMEs (dependent and independent); ownership, (i.e. foreign and domestically owned enterprises) and trading status, (i.e. trading and non-trading companies).

**Figure 7: Exported domestic value added as per cent of total domestic value added. 2013**

****

Although SMEs have less direct engagement in exports, see Figure 7, they have important integration channels as upstream suppliers. Indeed, for some categories of enterprises, the value added due to indirect exports is larger than through direct exports. Overall, when accounting for indirect exports of value-added, foreign markets are at least twice as important for SMEs compared with traditional business statistics published by the NSIs. For example, while only 5 per cent of total value added generated by independent micro SMEs in Sweden is exported directly, an additional 24 per cent of their value added is indirectly embodied in exports. For large enterprises these differences are also significant, but they have more value added due to direct exports than to indirect exports: direct exports accounted for 40 per cent of the total value added of large enterprises in Sweden for example but an additional 15 per cent was indirectly embodied in exports. Nevertheless, despite the smaller indirect links of large enterprises, in general, SMEs have lower export orientation than larger enterprises, especially independent SMEs, see also 4.1.1.

1. **Conclusion**

The selection of results presented in section four, show how harmonised databases across the Nordic countries and the use og centrally scripted approach is an effective tool in answering complex and diverse research questions. As such, this approach has played a central role in advancing the knowledge of enterprise heterogeneity and the answering of the question ‘how are different types of enterprises contribute to the growth in the domestic economy’, as well as aiding in the proses of understanding the magnitude and importance of cross border activities by addressing ‘what kind of enterprises are trading’ instead of ‘what do countries trade’. Thus, harmonised databased across the Nordic countries and the use og centrally scripted syntax have been verified, as a practical tool in meeting the challenge of (new) knowledge creation without putting any further burden on enterprises in term of increased data collection.

In the future, combining the business statistics data to social statistics by including variables available at individual employee level such as educational attainment, gender, income/salaries, years of work experience and more is expected to enrich and deepen linked micro-data analysis. Linking information from the business register with information on jobs enables the analysis of the impact of the (dynamic) structure of the business economy on the labour market. The advent of globalisation has been paired with intense debates among policy makers and academics about its consequences for a range of social issues related to employment, labour conditions, income equality and overall human wellbeing. On the one hand, the growing internationalisation of production may lead to economic growth, increased employment and higher wages. On the other hand, fears are often expressed that economic growth may be decoupled from job creation, partly due to increased competition from low-wage countries, or through outsourcing and off-shoring activities of enterprises.

The above coupled with the strong embeddedness of the Nordic economies to the world marked exposes enterprises and the Nordic societies to economic uncertainties. Furthermore, both the sort and long term effects of major geopolitical events such as BREXIT and the deteriorating trade relationship between the US its trading partners are yet to be uncovered. In this, micro-data linking could play a pivotal role in scrutinising the different effects of potential shocks to the economy.

1. **References**

Birch, D.L. (1981), ‘Who Creates Jobs?, Interest, 65:3, pp. 3–14.

Eksportrådet (2014), *Det handler om arbejdspladser*, Eksportrådets politik for små og mellemstore virksomheder.

Fornaro, P. and Luomaranta, H (2018), Aggregate fluctuations and the effect of large corporations: Evidence from Finnish monthly data, Economic Modelling, 70, pp. 245–258.

Gabaix, X. (2011), The granular origins of aggregate fluctuations, Econometrica, 79(3): pp. 733–772.

Statistics Denmark, Statistics Finland, Statistics Norway, Statistics Sweden and OECD (2017), Nordic Countries in Global Value Chains.

Statistics Denmark, Statistics Finland, Statistics Norway and Statistics Sweden (2014), Nordic Exports of Goods and Exporting Enterprises.

1. Core countries are Denmark, Finland, Norway and Sweden, while Iceland has also participated in a number of collaborative MDL projects. [↑](#footnote-ref-1)
2. In all Nordic countries as in the European Commission, dedicated export agencies have been established with the purpose of supporting especially the export opportunities of SMEs, see for instance *Det handler om arbejdspladser*, Eksportrådets politik for små og mellemstore virksomheder, (Udenrigsministeriet, Eksportrådet,2014) [↑](#footnote-ref-2)
3. Gabaix (2011) [↑](#footnote-ref-3)
4. Fornaro & Luomaranta (2018) provides a Finnish example [↑](#footnote-ref-4)
5. The project “Nordic countries’ role in the value creation of global firms” is carried out for Vinnova, the Swedish Innovation Agency.

 [↑](#footnote-ref-5)