

**Meeting of the Group of Experts on Business Registers
organised jointly by UNECE, Eurostat and OECD**

21-23 September 2015

Centre Albert Borschette

36, rue Froissart, 1040 Brussels, Belgium

**SESSION II: Linking of statistical business registers and trade
statistics**

Statistics on trade by enterprise characteristics – brief review of past, current
and future projects linking business statistics in the European Statistical System

Pekka Alajääskö

Eurostat

1. Introduction

European business statistics compilers often face a dilemma: On the one hand, users and policy makers demand additional information on the structure and development of European enterprises. On the other hand, budget constraints and reluctance to increase the burden on survey respondents and national statistical institutes put tight restraints on the extension of data requirements.

Micro data linking (MDL) can provide an opportunity to discover new information and to develop new statistics and indicators both when using existing data sets but also when combining with new data collections. Very visible outcomes of the MDL business statistics projects within the European Statistical System (ESS)¹ are the regular compulsory production of international trade in goods by enterprise characteristics statistics (TEC) and inward Foreign Affiliate Statistics (IFATS). However, there are many other important MDL projects with interesting results; some of them are continuing and start to be mature enough to make proposal for further regular data production.

This paper briefly reviews the past, current and future projects linking structural business statistics (SBS) with other business statistics with the view of improving statistics on entrepreneurship and in particular its international aspects.

2. International trade in goods by enterprise characteristics (TEC) and Entrepreneurship Indicator Programme (EIP)

International trade in goods by enterprise characteristics is a statistical domain, which unlike traditional trade statistics, aims at describing the structure of trade by characteristics of the trading enterprises, for instance by their economic activities, their size or concentration of trade. It is based on linking international trade in goods (ITGS) micro data with business register (BR) information, allowing a deeper analysis of the impact of trade on employment, production and value added.

TEC has been part of the regular data collection since 2009 and contains data from reference year 2007 onwards² TEC provides very interesting and important statistics on the structure of trade by characteristics of the trading enterprises. From the Entrepreneurship Indicator Programme's (EIP) perspective particularly interesting TEC statistics is the enterprise size class information.

Information on young, small and high growth enterprises is the common and central theme for the EIP. In general one needs information on enterprise age, size class, employment and turnover growth rates and combinations of them. The TEC framework provides data for the EIP indicator on the export performance of small enterprises. An even better indicator, originally identified by the EIP would be the export performance of young enterprises; however, data for such an indicator is not yet widely available.

2.1. Current work and plans in 2016

The planned MDL project in 2016, extending the MDL of SBS with other business statistics will make first attempts to produce statistics on export performance of young and/or small enterprises by combining ITGS and Business Demography (BD) statistics:

¹ The ESS is the partnership between Eurostat and the national statistical institutes (NSIs) and other national authorities responsible in each EU Member State for the development, production and dissemination of European statistics. This Partnership also includes the EEA and EFTA countries.

² These data are available for some Member States from the reference year 2005 onwards.

- EIP indicator: Export performance, young or small enterprise
 - Currently one gets information on export performance of **small** enterprises from the TEC data collection;
 - In the 2016 MDL project ITGS and BD data will be linked and there will be an attempt to produce data on export performance of **young** enterprises.

In addition, current and planned (MDL) projects will take the development work done in the TEC framework forward to cover services enterprises i.e. international trade in services by enterprise characteristics (STEC). This will make it possible to produce the indicators above for enterprises engaged in international trade in services (ITSS).

Additional important development work carried out by the MDL SBS projects is the consistent linking of the enterprise population of ITGS with the SBS population. As a result one could compare and produce indicators on the performance and relative shares of traders and non-traders in the SBS population. An additional result could be the production of some indicators on a more regular basis (though still voluntary) in the SBS framework.

In 2016 it is also planned to extend the linking of ITGS and SBS enterprise level data with that of BD. It could become possible to produce the following EIP indicators:

- EIP indicator: Value-added of young or small enterprises
 - Currently one gets information on value-added of **small** enterprises from the SBS data collection;
 - In the 2016 MDL project SBS and BD data will be linked and there will be an attempt to produce data on value-added³ of **young** enterprises;
 - In the 2016 MDL project SBS and BD data will also be linked with ITGS trying to produce data on value-added of **young exporting** and **non-exporting** enterprises.

3. Further development of the EIP indicators - EIP Data needs and micro data linking

3.1. Medium to long term – no data currently available but some action on-going

Availability and relevance of the enterprise size-class information is essential for the indicators of EIP. Current SBS and TEC size-class information has a deficiency in the sense that size class data include all small and medium size enterprises (SME) and do not differentiate between independent SMEs and dependent SMEs, which are part of enterprise groups. Here development work to improve the situation within the SBS framework is already under way: Eurostat together with DG GROW has set up a task force. Denmark, Germany, Italy, the Netherlands, Poland and the United Kingdom also participate. Topics the task force works on are: linked SMEs, one person enterprises, the coverage of SBS as well as the consistency between SBS and BD data.

Data on independent SMEs are also one of the issues the on-going MDL project linking SBS and other business statistics is working on. The first results on this will be published in September. These results indicate that:

- Among SMEs, medium sized enterprises are very often part of an enterprise group, most visibly in manufacturing and to a lesser degree in the knowledge intensive business services sector
- Dependent SMEs, belonging to a group, are important in terms of employment and gross value added (GVA), especially in smaller countries such as Denmark, Norway

³ Number of employees, number of person employed, number of enterprises, turnover and personnel costs are other SBS variables included in the project.

and Finland but also in Germany where the dependent SMEs constitute 43 per cent of total GVA created by SMEs and employ 34 percent of total employment in SMEs

- Dependent SMEs are much more open to international trade than the independent ones; the dependent enterprises behave in this regard very much like the large enterprises. This is quite natural as dependent SMEs are part of an enterprise group, thus they are not SMEs but rather large enterprises.

3.2. Long term – no data currently available no action on-going nor planned

Since the reference year 2008 foreign controlled enterprises in the EU (IFATS) can be consistently compared with domestic enterprises. In the inward FATS domain, the SBS enterprise population is divided into foreign controlled and domestic enterprises, and SBS statistics for nine annual and two biennial variables and more than a hundred 3-digit NACE Rev2 groups or aggregation of them can be produced for both groups of enterprises. From the EIP perspective the deficiency is that there is no size-class or age information for foreign controlled enterprises in the EU that would allow a comparison of domestic- and foreign controlled small and/or young enterprises.

The TEC framework provides enterprise group information, so it should be possible to identify linked enterprises and improve the TEC size-class data. Additionally, information on the age of ITGS trader enterprises would be useful in order to identify young surviving trader enterprises among all surviving young enterprises and combine this information with information on size.

Additionally, the TEC framework includes information on ownership, therefore it is already possible to identify foreign controlled and domestic ITGS active enterprises and produce TEC statistics for both groups. From the EIP perspective it would be useful to combine size-class and ownership information and produce size-class data for foreign controlled and domestic ITGS active enterprises. Again, information on the age of enterprises would be useful to identify young surviving foreign owned trader enterprises among all surviving young (trader) enterprises.

4. Overview of the micro data linking possibilities in business statistics

Apart from the standard dimensions e.g. of location and economic activity, the population of enterprises in business statistics can be grouped along the following dimensions:

- Group status: to determine which enterprises are dependent (part of an enterprises group) and which enterprises are independent
- Size class: to identify (subgroups of) SMEs based on persons employed
- Control: to distinguish foreign and domestic controlled enterprises
- Age: to determine which are the young enterprises
- Involvement in international trade: to determine importers, exporters, two-way traders and non-traders

Furthermore these dimensions can be combined to form different groups of enterprises. Policy makers can then determine which of these groups contribute to employment and wealth creation in our economies and develop targeted policies.

4.1 Enhancement of the data on entrepreneurship

In this paper we have already listed some of the EIP data needs:

- Combining groups status with size class to identify dependent and independent SMEs

- Combining "control" with size class to compare foreign and domestic controlled enterprises in different size classes
- Combining "control" with age to identify young domestic and young foreign controlled enterprises and combine this information on growth to determine high growth enterprises (HGEs) and gazelles (young HGEs) which are young enterprises that have had 20 % employment growth or alternatively 10 % employment growth
- Using "control" in TEC data to distinguish foreign controlled and domestic controlled traders
- Use involvement in trade to divide the SBS population into exporters, importers, two-way traders and non-traders and further divide these by size class
- Identify high growth enterprises and gazelles in these trader populations by size class

Ideally we do not implement changes separately in each of our data collections but form a coherent plan on how to use micro data linking and other tools to create rich datasets that cater not only to the data needs identified and presented above but are capable of satisfying future data needs as well. Below the MDL approach used in SBS is briefly described.

5. Micro data linking in business statistics

As can be seen above most of the data needs if not all can be fulfilled by existing data or by combining data existing in National Statistical Institutes (NSI) by using MDL. MDL has been one of the priority areas in the business statistics in the recent past. It is considered a very useful investment and little by little with modest further investments, the still limited number of EU countries using MDL should be increased to eventually include all (see the below overview table), so that we will be in a position to produce very interesting data on a regular basis and also to full fill occasional ad hoc requests.

5.1 Co-ordinated micro data linking

When attempting to link two or more sets of data, two things are vital for a successful exercise:

1. There has to be a large enough intersection of responding units. Up to date SBRs play a key role here.
2. There has to be a unique identifier or at least a very reliable matching approach. Again up to date SBRs are vital.

Since confidentiality is an important issue in the field of European business statistics, it is important to note that the linked micro data files were stored locally at national statistical offices throughout the projects and were not shared with third parties. It was therefore important to decide on concrete research objectives before each exercise and to design the linked data sets accordingly (i.e. decide on the variables to retain and on the aggregates to be produced). To achieve a harmonised output, the participants were provided with precise data requirements and standardised guidelines explaining in detail how the linked datasets in each country were to be structured. The computer code to produce the tables was provided as well, to ensure that identical tables were constructed in all countries.

Micro data linking can also be used to improve the quality of existing statistics. In 2013, under the umbrella of the ESSnet project (A network of several ESS organisations aimed at providing results that will be beneficial to the whole ESS) on Measuring Global Value Chains, the NSIs of Denmark, Norway and Finland linked statistics on the activities of affiliates based abroad (foreign affiliates - OFATS) with statistics on foreign controlled enterprises resident in the compiling economy (IFATS). Since IFATS is mostly based on

administrative (subset of SBS) data while OFATS information is collected by a survey, IFATS quality is generally assumed to be superior. The approach taken was to mirror IFATS and OFATS data sets between the countries, where control was exerted from an enterprise resident in one of the three countries and the foreign affiliate was located in another. In theory, this approach should have resulted in an identical set of affiliates in IFATS and OFATS; however the exercise showed that there are some discrepancies between the two statistics and gave important leads for the improvement of FATS data quality.

Apart from the SBS MDL exercises, TEC, and the FATS linking projects described above, the ESS has engaged in several other business statistics related MDL exercises. There has been an ESSnet on Linking of Micro data on ICT Usage, where business registers were linked with SBS and ICT usage and e-commerce data. Finally there has been an ESSnet on data warehousing (DWH) and MDL which touched more on theoretical aspects. The following table provides an overview and shows which countries were involved in each round:

	ICT Usage			Business Statistics				Trade					DWH	FATS	
	2006-08	2010-12	2013	2010-11	2012-13	2014-15	2015-16	TEC regul	TEC Meets 2009	TEC Meets 2011	TEC Meets 2013	sTEC voluntary	sTEC 2015-16	Data warehouse	iFATS vs oFATS
BE								X							
BG							X	X							
CZ	X			X				X				X			
DK	X	X	X	X	X	X	X	X		X	X	X	X		X
DE	X	X	X			X	X	X	X						
EE							X	X			X	X	X	X	
IE	X	X	X		X			X				X			
GR								X	X						
ES								X							
FR	X	X	X		X			X							
HR															
IT	X	X	X	X			X	X		X				X	
CY								X							
LT					X			X						X	
LV						X	X	X	X						
LU		X	X				X	X				X			
HU								X	X						
MT								X							
NL	X	X	X	X	X	X	X	X			X	X	X	X	
AT	X	X	X			X	X	X			X	X			
PL		X	X					X			X	X			
PT				X		X		X						X	
RO		X		X	X			X							
SI	X	X	X	X				X							
SK								X		X					
FI	X	X	X	X	X	X	X	X							X
SE	X	X	X		X	X	X	X						X	
UK	X	X	X					X				X		X	
NO	X	X	X	X	X	X	X	X			X		X		X
IS							X						X		

5.2 Benefits and complications of micro-data linking

"Benefits":

- Micro-data linking has significant potential in gathering new statistical evidence without increasing the burden placed on respondents;
- It can be an additional way of ensuring data quality and consistency between related data sets;
- MDL can be a tool to reduce discrepancies that occur in mirroring international statistics e.g. in inward and outward foreign affiliate statistics.
- A coordinated approach (all ESS MDL projects in business statistics applied this approach) to micro-data linking is a cost-effective way for national statistical offices to undertake micro-data linking and ensures harmonised, comparable results across countries;
- MDL can help reduce the burden on enterprises when conducting new or existing surveys by eliminating all questions that can be answered from existing data sources through MDL. A successful example here is the International Sourcing survey carried out in 2007 and 2012 (to be carried out in 2017).

"Complications":

- It is very important to decide on concrete research objectives before the exercise and to design the linked data set accordingly (i.e. decide on the variables for the analysis and on the aggregates to be produced). This is very difficult before the actual data analysis and in the projects with tight deadlines it has proved to be almost impossible to rerun the code and produce additional data needed.
- Sample surveys versus administrative registers. Administrative data is generally found to be superior to information collected by a survey. Micro-data sets based on administrative data are also more straightforward to gross-up to total population; they are often exhaustive census type of data. Linking survey based data at enterprise level is additionally complicated by the negative coordination of samples.
- From the analytical perspective it is desirable not only to link the latest data but also earlier records going back as many years as possible. Since business demography events like the creation, termination or acquisition of enterprises may affect the BR populations over time, the matching rates are the highest for latest data (in previous exercises more than 95% in all participating countries) and decrease for each year back in time (often just above 50 % a decade earlier). It is useful to have a common strategy to deal with demographic events.
- Statistical units are not always the same; legal units versus enterprise versus enterprise group complicates the micro-data linking. ESS is aiming at one common statistical unit for business statistics: enterprise. It is also important to implement the statistical unit in the same way.
- There are different national institutes responsible of the production official business statistics; this often causes unnecessary problems in the micro-data linking.
- Important dimension of globalisation: information about belonging to an enterprise group or not, and whether the enterprise group is all national or foreign is already part of the business register regulation in the ESS, however, this regulation is not implemented or it is implemented in different ways in various member states.