

Identical variables in different business and trade-related statistics - a challenge for European statistics

Bjoern Witting¹, Joachim Weisbrod and Mirjam Weber
Federal Statistical Office (Destatis), Wiesbaden, Germany

¹ Corresponding author: Bjoern Witting, e-mail: bjoern.witting@destatis.de

Abstract

The programme for Modernization of European Enterprise and Trade Statistics (MEETS) encourages the development of an integrated system of business and trade-related statistics in which all statistical results can be combined easily to get new results without additional burden for respondents and to draw a coherent and consistent overall statistical picture. Within this system of economic statistics, different methods of data collection should be used and combined with each other. In this context the changeover from a stove-pipe approach to a data warehouse approach is an urgent need and consistency becomes imperative. To this end, a core project of MEETS “on consistency of concepts and applied methods of business and trade-related statistics” was launched. The third work package (WP3) of this project aims to identify inconsistencies regarding characteristics and their definitions in the area of business and trade-related statistics within the European Union and countries of the European Free Trade Association (EFTA). To achieve this goal, all available sources of information have to be evaluated and their relevance has to be assessed. The results of this project are also very important for the further development of a new Framework Regulation for Integrated Business Statistics (FRIBS) for the European Statistical System (ESS). Additionally, they are preconditions for the implementation of the recommendations of other related European projects within the ESS (ESSnets) like “on micro data linking and data warehousing in statistical production” (ESSnet Data Warehouse).

Key Words: consistency, modernization of business statistics, quality improvement

1. Introduction

Improving consistency within a coherent system of business and trade-related statistics becomes an important issue on the way towards an efficient and flexible ESS. The target is formulated in the so called “vision paper” (COM (2009) 404) of Eurostat, where it is described as the “changeover from a stove-pipe to a data warehouse approach”. The European Commission launched the MEETS programme to create the preconditions for the implementation of the “Vision” of a more effective production system of statistics within the area of business and trade-related statistics.

The ESSnet “consistency of concepts and applied methods of business and trade-related statistics” (ESSnet Consistency) is an important project of the MEETS programme. It attributed to objective 2 of the programme, which “aims to achieve a streamlined framework for business related statistics”. Action 2.1 foresees the integration of concepts and methods across the Member States (MS). The results of this ESSnet are also important for the further development of a new Framework Regulation for Integrated Business Statistics (FRIBS) for the European Statistical System and they are preconditions for the implementations of the recommendations of other related ESSnets like e.g. the “ESSnet on Data Linking and Data Warehousing” and the “ESSnet Admin Data”.

In the past, statistical regulations were created when there was a need for data in a specific area and the units, target populations, periodicities and characteristics were defined according to the special needs of these data demand. This was the so called

stove-pipe-approach in Official Statistics. Even though National Accounts (NA) are a common users of most of the data, an overall systematic approach does not exist. The "Vision" is now to develop an integrated system of business and trade-related statistics in which the results of the single statistics could be combined easily to receive new results without new surveys. Within this system of economic statistics different methods of data collection should be used and combined with each other. In this context consistency becomes an important quality standard. Furthermore, the term "system" implies that the final output is more than the sum of the constituent parts. Especially because we do not have the possibility to design a complete new system from the beginning, we have to transfer the running statistics into the new system stepwise and with a well-organized change management.

At first glance consistency means that the results of different statistics should be comparable in the sense that they refer to comparable statistical units, target population and variables. Therefore, the ESSnet Consistency consists of three work packages (WPs) which analyse the issue of inconsistency with regard to these three aspects. The ESSnet will end in December 2013 when MEETS expires.

For the results of the ESSnet it is important, that the analysis of the three work packages refer to the same set of statistical domains. From the European perspective, "business and trade-related statistics" groups all statistical domains based on the statistical unit "enterprise". There are also a couple of manuals to be analysed, which give methodological advice to the MS. Additionally, the analysis of the link between Balance of Payments Statistics (BoP) and NA needs a specialized discussion.

"National Accounts is a framework for the systematic and detailed description of the total economy, its components and its relation with other total economies" (Richter, Engelage and Thomas (2010)) whereas the domain statistics shed light on special aspects and/or sectors of the economy. The concepts relevant are defined in the European System of National and Regional Accounts (ESA), which in turn is harmonized with the concepts and classifications used in other European social and economic statistics such as statistics on employment, on manufacturing and on external trade. ESA is looked upon as the central framework of reference for social and economic statistics of the EU and its MS. Therefore, there should be a conceptual relation between the aggregates of NA and the variables of business-related statistics which are the main data sources of NA. Related variables are output of goods and services, intermediate consumption, gross fixed capital formation and compensation of employees. Here the definitions e.g. of the Structural Business Statistics (SBS) show relations to the definitions in ESA. A second type of macro-economic aggregates which represent balancing items in the accounts, are variables such as gross domestic product, operating surplus of the total economy, national income etc. For these variables various business-related statistics of the inventory such as SBS, Statistics on Production and External Trade Statistics have to be integrated within a sometimes difficult transformation and compilation process.

BoP summarizes systematically the economic transactions of an economy with "the rest of the world". This information is necessary for monitoring the national and international monetary policy. But there is a close conjunction between BoP and NA because the essential concepts - e.g. the concepts of economic territory, residence, centre of economic interest or time of recording etc. - are identical in BoP and NA.

ESA could therefore be used as a kind of coordination framework for business-related statistics. It could serve as a conceptual framework ensuring the coherency of the system making variables used in different but related fields of statistics comparable and it could be used as an accounting framework ensuring numerical consistency of data taken from different sources.

2. Aim of the project

WP3 is a cooperation of 8 National Statistical Institutes (NSIs) under the coordination of Statistics Sweden. The formal basis is a Multi-beneficiary Grant Agreement (MBGA) between the European Commission and the NSIs of Estonia (SE), Germany (DESTATIS), Greece (ELSTAT), Ireland (CSO), Italy (ISTAT), Slovenia (SORS), Sweden (SCB) and Switzerland (SFSO).

The aim of this work package is to identify inconsistencies regarding characteristics and their definitions within the area of business and trade-related statistics, to assess the importance of the identified inconsistencies and to elaborate proposals by which these inconsistencies could be avoided or at least reduced. Therefore, all available sources of information have to be evaluated and the relevance of the information has to be assessed.

The first step in the analysis of possible inconsistencies was stocktaking the present situation. Therefore, SFSO has been developing a database for stocktaking and analysing all relevant variables and definitions of the legal acts under discussion. Afterwards hypotheses about the reasons and the relevance of these inconsistencies have to be formulated and proposals for prevention or reduction of relevant inconsistencies will be developed. Therefore, the members of WP3 will present a report, which will be produced in a stepwise process. They will take into account the results of the working steps already completed and draw conclusions for the following steps.

3. The issue of Consistency

As elaborated in more detail in the External Study (ES) (Richter, Engelage and Thomas (2010)) the criterion of consistency within a set of different data sources is a necessary condition to make data or statistical results comparable with each other so that they could be combined in a meaningful manner. This increases the flexibility for the production of statistical results in the overall system of business and trade-related statistics. A higher level of consistency could help the NSIs to save resources and to lower the burden for respondents. It is a precondition for a changeover from a stove-pipe to a more integrated approach.

Consistency is also mentioned as a target in the European Statistics Code of Practice (CoP (2012)) to which the NSIs are obliged within the ESS. Principle 14 of the European Statistics Code of Practice states that “European Statistics are consistent internally, over time and comparable between regions and countries; it should be possible to combine and make joint use of related data from different sources”.

Although the ESSnet has the title “*Consistency of concepts and applied methods of business and trade-related statistics*” unfortunately “consistency” and its relation to the terms “coherence”, “compatibility” and “comparability” is not always clear. Regarding characteristics and definitions the types of coherence/comparability should be distinguished in “Comparability over time”, “Comparability over region”, “Comparability over other domain”, “Internal coherence”, “Coherence between sub-annual and annual statistics”, “Coherence with the National Accounts” and “Coherence with other statistics”.

The most obvious obstacle to consistency and coherence are different characteristics and/or differences in the definitions of the characteristics in related areas of business and trade-related statistics. Concerning the issue of consistency, the ESSnet differentiates two dimensions.

- **Horizontal consistency** refers to the comparability between the various statistical domains. Data between statistical domains can be compared if they are elaborated using the same statistical unit, definitions, frame and reference period.

- **Vertical consistency** is the issue of comparability between the sum of MS data and the European aggregate. Concepts developed for the national implementation may not be suited to derive the consistent European aggregate.

The ESS, including business and trade-related statistics in its current state has been growing in its historical process with different stakeholders and different analytical objectives. Each statistical domain has developed and revised its legal basis. This is true for the definitions of the characteristics as well and little attention has been paid to the definitions of characteristics in related statistical domains. This is called the **“stove-pipe-approach”**. Within this system of economic statistics different methods of data collection should be used and combined with each other. In this context consistency becomes an important quality standard.

“When are two definitions of characteristics consistent or inconsistent?” is the central question of the project. Statistical characteristics are properties which are observed for and attributed to a statistical unit of a certain target population for a certain period or point of time. WP3 defines the determinants of characteristics (or variables) as follows:

- the name (standing for the contents of the characteristic),
- the description of the property itself (e.g. inclusions and exclusions, relation to other characteristics) and
- the description of a measurement concept.

These determinants are constituent parts of a complete definition. From a conceptual point of view cases for inconsistencies can be in all three parts of the definition of a characteristic. These cases are as follows:

- Although characteristics have the same name the description of the contents could be different (e.g. turnover in SBS and turnover in STS).
- Although the description of the contents is similar the measurement concept may be different (e.g. value added at factor costs and at basic prices).
- Although the description of the contents and the measurement concept is comparable the characteristic has different names in different domains or different data sources (e.g. personnel costs in SBS and labour costs in LCS).

WP3 therefore has checked the definitions of variables in business and trade-related statistics under the aspects mentioned above to find out whether identified differences are relevant and important for the consistency and the coherence of the overall of business and trade-related statistics. It has to be kept in mind that the removal of certain inconsistencies by changing the definition could be detrimental for the relevance of the characteristic for specific user groups.

4. Stocktaking

Looking at characteristics and definitions the variables required by the different legal acts are relevant. Although a large number of different statistical regulations exist and these regulations contain a large number of single characteristics the number of related characteristics is rather limited. Only a few variables appear in more than two statistical domains. Therefore, the stocktaking was concentrated on the 17 core variables of the SBS and all other domains where one or more of these variables appear.

A database has been developed, which allows the stocktaking and analysis of all relevant variables and definitions of the European legal acts under discussion. This database is based on information sheets for each MS and each variable considered in the context of this WP3 and contains input and output definition, national denomination, data sources and all information related to vertical and horizontal consistency. The information sheets were sent out to all MS in January 2013. Until the beginning of April 2013 we could reach a response rate of 100 percent. This shows the

big interest in consistency over the MS. All received data have been stored into the database and will enable us to understand the reasons for the identified inconsistencies, to assess their importance and to elaborate proposals by which these inconsistencies could be avoided or at least reduced.

There are originally observed characteristics and derived variables. The WP3 decided that inconsistencies within the definitions of 4 core economic variables and 4 employment variables will be analysed in detail. These are:

Core Economic Variables

- Turnover
- Intermediate consumption / inputs
- Value added
- Production

Employment Variables

- Employment
- Wages and salaries
- Personnel costs
- Hours worked

If the project discovers additional variables of major interest for consistency these will be considered in the project as far as it fits in the tight time table and resources are left.

5. First results

The analysis of all variables is basically divided in the vertical and horizontal view according to the two dimensions of consistency mentioned above. Table 1 shows the current situation from the vertical perspective in each domain equally weighted over all MS in relation to all valid answers; Table 2 accordingly from the horizontal perspective of each domain. Variables in domains with inconsistencies of 40% and more in the MS are marked in yellow and 50% and more in red.

Table 1: Vertical inconsistencies per variable within the domains in all MS

Variable	Domain									
	SBS	STS	iFATS	oFATS	CIS	R&D	LFS	SES/LCS	LCI	ICT
Number of enterprises	37%	.	30%	41%	28%
Number of local units	43%	19%	.	.	.
Turnover	40%	59%	25%	28%	0%	13%
Production Value	30%	39%	21%
Value added at factor cost	27%	.	21%
Gross operating surplus	23%
Total purchases of goods and services	33%	.	18%	20%
Purchases of goods and services purchased for resale in the same condition as received	34%	.	21%
Payment of agency workers	32%
Personnel Costs	30%	33%	18%	50%	55%	.
Wages and salaries	23%	45%	32%	44%	.
Social security costs	33%	18%	34%	44%	.
Gross investment in tangible goods	33%	.	15%
Number of persons employed	20%	27%	21%	30%	.	17%	25%	.	.	23%
Number of employees	17%	28%	8%	25%	.	.	4%	31%	.	.
Number of employees in full-time equivalent units	23%	18%	17%	24%	.	.
Number of hours worked by employees	19%	26%	7%	29%	.	.

At first glance the results show that the challenge with horizontal inconsistencies is significantly higher than with the vertical ones. That corresponds to the expectations, as each domain could define more or less independently their variables. In detail, however, it is surprising that few domains have high vertical inconsistencies in a small number of variables. This could be explained by the reasons that a domain has used a definition from another source or that the domain has created a difficult implementable definition itself. The vertical inconsistencies caused by the first reason could be treated as horizontal ones; the second reason should be avoidable easily. The numbers to horizontal inconsistencies demonstrate the variety of variables and domains with the need of modification to reach an acceptable grade of consistency in business and trade-related statistics.

Table 2: Horizontal inconsistencies per variable in the domains in all MS

Variable	Domain									
	SBS	STS	iFATS	oFATS	CIS	R&D	LFS	SES/LCS	LCI	ICT
Number of enterprises	32%	.	32%	41%	35%
Number of local units	45%	74%	.	.	.
Turnover	37%	58%	36%	32%	33%	31%
Production Value	47%	63%	39%
Value added at factor cost	20%	.	14%
Gross operating surplus	0%
Total purchases of goods and services	27%	.	27%	29%
Purchases of goods and services purchased for resale in the same condition as received	20%	.	14%
Payment of agency workers	0%
Personnel Costs	47%	65%	43%	65%	60%	.
Wages and salaries	49%	48%	67%	56%	.
Social security costs	46%	38%	57%	46%	.
Gross investment in tangible goods	17%	.	14%
Number of persons employed	38%	52%	36%	40%	.	45%	72%	.	.	39%
Number of employees	43%	55%	38%	38%	.	.	69%	60%	.	.
Number of employees in full-time equivalent units	48%	49%	73%	59%	.	.
Number of hours worked by employees	53%	55%	75%	57%	.	.

Table 3 shows the percentage of causes and reasons for the inconsistencies, based on all valid answers for turnover in STS. They are important to find the origin of the inconsistencies and therefore inevitable to get useful solutions for the future. The results indicate that the main problems can mostly be found in a few causes and reasons in the vertical and horizontal perspective.

Table 3: Causes and reasons for inconsistencies regarding turnover in STS in all MS

Causes	vertical	horizontal	Turnover STS	vertical	horizontal	Reasons
						no inconsistency
reference time and period	3%	35%	0%	4%	EU legislation unclear	
coverage	7%	34%	7%	27%	timeliness of data collection	
observation unit	31%	22%	14%	4%	limited data availability	
denomination	0%	1%	10%	4%	multipurpose data collection	
use of proxy	3%	0%	28%	4%	reduction of external burden	
measurement concept	7%	21%	0%	2%	reduction of internal production cost	
elements included more	17%	0%	3%	2%	lack of/limited metadata	
elements included less	0%	2%	14%	4%	national information requirements	
elements excluded more	7%	3%	3%	29%	use of different methods/concepts	
elements excluded less	7%	3%	0%	0%	translation issue	

The follow up detailed analyses will be done for the core economic and employment variables and all variables with a high rate of inconsistency. The development of a weighting scheme for the inconsistencies is foreseen as well as the reconsolidation with NA and other ESSnet projects. Finally, the project will provide a set of recommendations for standardization and adjustments of definitions of variables for all business and trade-related statistics as input for FRIBS.

References

COM (2009) 404. *Communication of the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade*. Brussels, 2009.

CoP (2012). *European Statistics Code of Practice*. adopted by the European Statistical System Committee, 28 September 2012.

Richter, Engelage and Thomas (2010). *External study on detail evaluation of the legal acts in the areas of statistics which were identified by member states as areas to revision*. submitted to Eurostat, July 2010.