A system of variables as an example for a systematic approach

Joachim Weisbrod joachim.weisbrod@destatis.de

Bjoern Witting bjoern.witting@destatis.de

Federal Statistical Office (Destatis), Wiesbaden, Germany

Abstract

The programme „Modernization of European Enterprise and Trade Statistics (MEETS)“ and the „Framework Regulation Integrating Business Statistics (FRIBS)“ aim at creating a more integrated and output-oriented system of Business Statistics. To reach this goal, consistency between the statistical domains is mandatory. The “ESSnet on Consistency” was one of the core ESSnets of the MEETS programme, with the task to deal with the consistency issue in business and trade-related statistics in three Work Packages. Work Package 3 (WP3) focused on consistency regarding characteristics and definitions. The WP3 proposed a systematic approach to reduce inconsistency, which is based on a metadata system that consists of an interrelated set of sub-systems. One of them is the system of variables, which contains standardized general definitions of all variables in Business Statistics. These definitions are equally structured, use the same terminology, unique identifiers and contain a precise description of content and objective of the variable. Furthermore, they include interrelations to accounting systems, National Accounts and other variables. To ensure consistency, it is important that for every statistical purpose there is only one variable. The proposals for a consistent system of variables have to be discussed and carried on in the relevant committees or in further projects.

1. Introduction

Modern Business Statistics faces new challenges requiring new concepts and methods of data collection. The programme „Modernization of European Enterprise and Trade Statistics (MEETS)“ and the „Framework Regulation Integrating Business Statistics (FRIBS)“ aim at creating an integrated and output-oriented system of Business Statistics. To reach this goal, coherence and consistency between the different parts of the European Statistical System (ESS) become more important within the overall quality concept of official European Statistics. The “ESSnet on Consistency of concepts and applied methods of business and trade-related statistics” (short: ESSnet Consistency) was one of the core ESSnets of the MEETS programme. Its task was to deal with the consistency issue in business and trade-related statistics in three Work Packages. It attributed to objective 2 of the MEETS 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.

The aim of the Work Package 3 (WP3) was 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.

2. Challenges for modern Official Statistics

In a society which is often referred to as “Information Society” new challenges for official statistics in general and especially for Business and Trade-related Statistics appear. These challenges on the demand side are e.g.:

- an increased demand for statistical information,
- more output flexibility to react to new data demands,
- improvement of timeliness of data availability and
- better assessment of the quality of statistical information.

For the supply side challenges are:

- resource restrictions in the NSIs,
- burden reduction for respondents and
- more efficient use of already existing data.

To meet these challenges several measures are envisaged. An improvement of the producer-user-dialogue allows the observation of the relevant data demands and a better priority setting. The commitment to the "Code of Practice"[1] helps the user assessing data quality. With the "Vision"[2] Eurostat initiated a fundamental change in data production of the ESS. This change is called the transition from a so called stove-pipe-approach to an integrated approach of data production. The MEETS programme has been designed to analyse the requirements of such a systematic approach and FRIBS will provide the legal framework for the new approach.

3. Integrated Approach for Business Statistics

The above mentioned challenges could be met by a system of Business Statistics with the following components:

- An integrated system of Business Registers,
- an uniform concept of statistical units in all Business Statistics,
- the implementation of a statistical data warehouse,
- the use of data linking (national and between the MS),
- the common use of standardised IT-tools and
- the secondary use of all available data sources for statistical purposes.

These components of a system of Business Statistics are part of a programme for a stronger integration of the European Statistical System (ESS) in general and especially the system of European Business Statistics. Comparability of data between related statistics, domains and Member States is a precondition for a changeover from individual stove-pipe structures to a more integrated approach. Within a set of different data sources comparability is a necessary condition to combine data or statistical results with each other 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 comparability could also help the National Statistical Institutes (NSIs) to save resources and to lower the burden for respondents.

---

[1] European Statistics Code of Practice, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF, 10.08.2012

[2] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL on the production method of EU statistics; a vision for the next decade, COM 404 (2009).

The most obvious requirement for comparability is that variables intended to measure the same phenomenon in different statistics are defined in a comparable manner. This is not a trivial task.

4. Comparability, Coherence and Consistency

Consistency is one of the quality criteria for official statistics in the ESS. 'The need for coherent treatment of data collected in different statistical domains is explicitly stressed in Article 12 of the European Statistics law[3] where 'comparability' and 'coherence' are key quality criteria which have to be met when developing, producing and disseminating European Statistics'.

In the European Statistics Code of Practice, to which the NSIs have committed themselves within the European Statistical System (ESS), principle 14 specifically 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'.

But consistency is only one out of five quality standards within a comprehensive quality concept. These are:

- Relevance
- Accuracy and Reliability
- Timeliness and Punctuality
- Coherence and Comparability
- Accessibility and Clarity

There might be conflicts between the different components of the quality concept. But focussing on statistical integration means that comparability, coherence and/or consistency have to get a higher weight in an overall quality concept of Business Statistics.

Unfortunately there is no unique terminology for this quality component. In MEETS "Consistency" is used as a generic term for:

> COHERENCE:"… of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes

[3] Regulation (EC) No. 223/2009 OF THE EUROPEAN PARLIAMENT AND THE COUNCIL on European statistics…, 11 March 2009.

coherence, as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency."[4].

NUMERICAL CONSISTENCY: Same numerical result for the same statistical question

COMPARABILITY of statistics over time and between regions and countries (in EU and EFTA)

In the context of the implementation of the "Vision" we have to be more precise and specify the relations between the variables of the system. That means we have to specify for each variable what we really mean by "consistency of definitions of variables".

As a starting point, concerning the issue of consistency two dimensions have to be differentiated:

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(s), the same coverage, the same classifications, the same definitions, the same frame and the same reference time and 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.

5. Causes and Reasons for Inconsistencies

In the course of analysing possible (horizontal and vertical) inconsistencies in the present state of business and trade related statistics WP3 distinguished between causes and reasons of inconsistencies because it is not only important to know "WHAT is different" between related variables in different domains and Member States but also to understand "WHY is it different (even if the inconsistency becomes obvious)".

The causes of inconsistencies are often difficult to identify because it requires a sound knowledge of the metadata in the different Member States. Therefore a standardized and comparable documentation of metadata is essential for a coherent and consistent supranational system of statistics like the ESS.

---

[4] Statistics Canada's Quality Assurance Framework 2002

A survey in the Member States showed that the use of different units, different coverage and deviating reference periods are obviously the predominant causes for inconsistencies between different statistical domains and between the Member States. The main reasons for existing inconsistencies could be summarized as follows:

- For inconsistencies between different domains the traditional stove-pipe approach on the national level and on the European level plays an important role. Also special analytical purposes of certain statistics and special conditions in certain branches could be responsible for slight deviations in definitions of related variables.

- Inconsistencies between Member States are often hidden and can only be identified by analysing the complete set of metadata. Definitions on the European level are not precise enough and could leave room for interpretations. Institutional frameworks and available data sources can also differ between Member States.

Therefore, a complete, comparable and transparent documentation of metadata and a cooperation based on mutual trust within all parts of the ESS are essential for coherence and consistency in European Statistics.

6. Definition of Variables

The WP3 proposed a systematic approach to reduce inconsistency, which is based on a metadata system that consists of an interrelated set of sub-systems. One of them is the system of variables, which contains standardized general definitions of all variables in Business Statistics. These definitions are equally structured, use the same terminology, unique identifiers and contain a precise description of content and objective of the variables. Furthermore, they include interrelations to accounting systems, National Accounts and other variables.

The implementation of the “Vision” requires a changeover from a domain oriented (stove-pipe) approach to an integrated or systematic approach in business statistics, which always has the system as a whole in mind. Therefore this changeover is more than just a “cosmetic” correction of a few definitions and storing all variables in a so called statistical data warehouse. It increases the interdependence within the whole production system of official business statistics. It also requires another quality of management and communication and therefore a better information

structure. The systematic approach also requires a systematic presentation of the information which is relevant for monitoring the system.

In general a system is more than a pure collection of its elements. It also consists of a description of relations between the elements. A system of business statistics means that the different elements of business statistics should fit together and supplement each other like the pieces of a jigsaw puzzle. On the national level this should be valid for the results of the different domains of business statistics and on the European level it means that the results for the same domains of the different countries should also have this quality. The analysis of this quality component requires a systematic description of the system of business statistics consisting of different parts or sub-systems.

Inconsistencies between different domains of business statistics within a national system of business statistics (horizontal view) and between countries within the same domain (vertical view) are generally caused by a lack of comparability (coherence) of the underlying metadata. Therefore harmonising the metadata systems within a country and between the countries play an important role in improving national and international consistency of statistical output. WP3 therefore proposes to use a standardised description for the output of business statistics consisting of fixed building blocks, which all together form the metadata system.

Metadata are data about data and comprise all information which is necessary to understand and use statistical results correctly and to assess the quality of the results. They cover

- technical and methodological metadata,
- process and quality metadata and
- reference metadata (e.g. definition of variables)

A reference metadata system consists of different parts or sub-systems which broadly correspond to the above mentioned aspects of comparability. This means that a consistent description of a system of business statistics requires a set of sub-systems, e.g.:

- a sub-system of statistical units,
- a sub-system of periodicities,
- a sub-system of classifications and
- a sub-system of variables.

Each of these sub-systems defines the elements in the respective area and the relations between the elements.

For the definitions of the output only elements of these sub-systems are allowed to use. This is a precondition for consistent output of different statistical domains or different parts within one system.

7. System of variables

In this section the focus is on one special metadata sub-system, the sub-system of variables, which is of special relevance for the definition of the variables. While imposing the same conditions in all MS is not feasible (at least not in the medium term) it is necessary to develop a comparable manner of description of the output of business statistics which could be the basis (but only a starting point) of analysing the consistency issue. Therefore, it is very important to use a common and unambiguous terminology. It is the task of the sub-system of variables to describe in a comparable way what business statistics tries to measure. If we want to improve consistency it is necessary to develop a system of variables covering all domains of business statistics.

A system of variables should consist of different parts which all fit together and supplement each other:

- A structured list (or better matrix) of all variables used including their standardised definitions and the description of relations to other variables.
- A consistent and unique terminology for all statistical domains. If variables have the same name their content has to be identical. Already slight differences in content must lead to a different name.
- Only standardised definitions. The definitions should all follow the same pattern consisting of the following parts:
    - Unique code of the variable: There should be a unique identifier for each variable at least, better still would be a unique system of coding (classification) for all variables in business statistics. If the definition (the contents) between two variables varies even slightly (for example value added at factor costs and value added at basic prices) there should be two different IDs.

- Unique name of the variable: If the definition (the contents) between two variables varies even slightly (for example value added at factor costs and value added at basic prices) there should be two different names should exist.
- Objective of the variable: Which phenomenon is the variable ideally intending to mirror ("What should be measured?")?
- Definition of the variable itself (how is it measured):
  - verbal description of the content of the variable;
  - all inclusions and exclusions;
  - methodological details directly connected with the definition (e.g. value at the end of the reference period or average value).
- Relations of the variable:
  - Link to company accounts
  - Link to other variables
  - Link to National Accounts

The content of variables has to be described entirely and as simple as possible with a special focus on the constituents (inclusions and exclusions).

Since data sources vary among countries (e.g. different tax laws when using admin data) it is not sufficient to have a list of inclusions and exclusions. Each variable should have an "objective" containing a short description of the purpose of the variable. This could help the statisticians in different countries to decide on special cases. Relations between variables need to be made explicit. The proposal is to include a special chapter with the links to other variables (like the current SBS regulation on definitions does). In addition there should be links to company accounts (as already present in the current SBS regulation on definitions) and links to National Accounts, where applicable. Finally, special consistency issues regarding the variable should be mentioned explicitly (e.g. at what point in time 'number of persons employed' should be measured exactly – first day, last day or an average across the reference period).

Every inconsistency to the proposed definition should be assessed, explained and quantified. This should be part of the standardised quality reports produced by each MS.

Such a systematic approach could, however, only ensure consistency on a formal level and is focused mainly on the reduction of the horizontal inconsistencies. Nevertheless, it improves the

transparency of business statistics considerably. The implementation of such a system could lead to the identification of inconsistencies as MS face different constraints, use different methodologies for data collection and have different data sources at their disposal. It is therefore necessary to solve those problems by the help of guidelines, handbooks etc. that could give support to transform data from existing sources into the variables required by the European statistical laws. Handbooks and guidelines will have to be developed based on all results and proposals of the different ESSnets.

8. Practical Impacts of using a Systematic Approach

First of all it should be emphasized that it refers to output definitions which are used for the presentation of the results of business statistics. This has to be distinguished from the variables that are collected from the different sources (input definitions). Input definitions should be clear for the respondents and adapted to the data source which is used. When input definitions deviate from output definitions the NSIs will have to implement a transition process that transforms the input variables into the output variables.

Another impact of this approach is that the different domains of business statistics can no longer act independently from each other. Regarding the system of variables it is necessary to maintain it centrally. Whenever a new statistic is designed, the responsible unit has to use existing variables and their definitions as far as possible. This requires an intensive amount of communication between the different units of a NSI and the implementation of a monitoring procedure. New variables have to be justified and they have to be integrated into the system of variables.