**The statistical units’ definition and implementation as core element to guarantee the ESS’ consistency**

Giuseppe Garofalo (*garofalo@istat.it*)

Italian National Statistical Institute (ISTAT)

**Abstract**

The role of the statistical units to achieve consistency is well known. Even though the objectives of a specific statistical domain are clear and correctly identified, an inaccurate definition – and identification - of the statistical unit can cause a deviation from the objectives themselves. Sometimes the public statistician does not realize it. The statistical unit determines the data on the variables as well as the classification and how the data collected can be analyzed”. The “ESSnet on Consistency” has been one core ESSnet of the MEETS program, with the task to deal with the consistency issue in business and trade-related statistics in three Work Packages. The first WP dealt with the issues of the statistical units as a core factor in achieving consistency. Starting from the results achieved by the WP, in terms of analysis and evaluation on the identified inconsistencies, the paper analyses the problems in using different statistical units, or the different implementation of the same statistical unit, in terms of horizontal, vertical, spatial and time consistency.

The “route”, both conceptual and practical, to guaranty coherence in statistical units definition and implementation is proposed.

Keywords: Consistency, statistical units definition, European Statistical System

1. **Introduction: the meaning of consistency**

The European Statistics Code and Practices[[1]](#footnote-1) states in the principle 14 Coherence and Comparability: “*European Statistics are consistent internally, over time and comparable between regions and countries; it is possible to combine and make joint use of related data from different sources*”. The principle underline both the need to have comparable statistical outputs and the need that these outputs have to be able to jointly used for the production of the European statistical information.

Very often when talking in logical arguments about statistical products, the word consistency is treated as a synonymous of *coherence* and *comparability*.

Coherence and comparability refer to the extent to which differences between statistics can be attributed only to differences between to the true values of the characteristics they estimate. The difference between the two concepts is that comparability refers to comparisons between statistics based on usually unrelated statistical populations and coherence refers to comparisons between statistics for the same or largely similar populations. Using the SMDX content-oriented guidelines,[[2]](#footnote-2) we can have the following definitions:

***A1. Spacial comparability/coherence***: refers to the “*degree of differences between statistics measuring the same phenomenon for different geographical areas*”.

***A2. Time comparability/coherence***: refers to the “degree of differences between two or more instances of data on the same phenomenon measured at different points in time”.

***A3. Domain comparability/coherence***: refers to the “*degree of differences between the survey results which target similar characteristics in different statistical domains*”.

In statistics, the concept of “internal consistency” refers to the general agreement between multiple items that have a composite score of a survey measurement of a given construct. Generalizing this concept is possible to define consistency in the following way: *correspondence and uniformity among the parts of a complex system*.

In synthesis, the concepts of coherence (and comparability) refers to the comparison and evaluation of the homogeneity of the results obtained by different informative systems, while consistency refers to the evaluation of the internal homogeneity of an informative system.

Referring to the European Statistical Systems, consistency means the ability in terms of both statistical and space domains to describe, in a coordinate way in time and space, the wide European social and economic picture. In this way is possible to define:

***B1. 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 on such MS data. Virtual consistency is observed in statistical domains where the statistical objects are of cross-border nature[[3]](#footnote-3).

***B2. Horizontal consistency***: is the ability of the different statistical domains to be used for other statistical domains (e.g. from the Business Statistics domains to the National Accounts).

***B3. Time consistency:*** is the ability of the different statistics produced with the reference to different time (e.g. monthly, quarterly, annually statistics) do not produce conflicting results.

1. **The coherence and the consistency in the statistical domains referring to the Business and Trade related statistics.**

Inside the ESS, the Business and trade area covers a wide range of statistics[[4]](#footnote-4): from the *Structural Business Statistics* (SBS) (describing an economy through the observation of the activity of units in terms of employment and, both input and output, monetary characteristics), to the Short-term business statistics (describing a range of economic indicators in some case very different each other: e.g. turnover, production, price), to the employment statistics (describing phenomena as earning, labour cost, job vacancy), to some sectorial statistics (Energy, tourism, environment,…), to statistics supporting the international relationship (Trade, FDI, FATS), to Statistics on Information and Communication Technology. Furthermore some of these statistics are strictly connected to the so called “Tertiary Statistics” like National Accounts and Balance of Payments.

These statistics can be combined in homogeneous as in the following:

1. *Domains mainly financial and monetary oriented.* This typology can be subdivided in:
	1. Domains mainly related to national issues: e.g. SBS and STS (turnover and price indexes), R&D.
	2. Domains related to both national and global issues: e.g. FATS, External Trade, FDI.
2. *Domains mainly commodities oriented*: e.g. Prodcom, STS-production index, Tourism, Energy statistics, Waste statistics.
3. *Domains mainly oriented to the Innovation.*
4. *Domains mainly employment oriented.*

While the first two typologies support, even if in different way, the needs of the economic analysis, the other two also have an interaction with the social analysis.

The four typologies have a different relationship with the concepts of coherence and consistency as previously described. The space and the time coherence (A1 and A2) have be achieved for all of the typologies and all statistics, the domains coherence (A3) can be achieved only for some statistics or some variable of the statistics, for instance between some variables of the employment oriented domains and the variables on employment and labour cost of the SBS.

In terms of consistency only the first typology (financial and monetary oriented) have to be faced with the vertical consistency (B1) because only this typology can produce statistics (e.g. turnover) for which their cross-border nature determine that the sum of the National figures (or Regional) not correspond to the European (National) figures. Time consistency (B3) can be achieved only for the first two typologies, e.g. because the existence of the relationship between STS and respectively PRODCOM and SBS statistics.

Horizontal consistency (B2) can be achieved for all typologies especially because they are input for the National Accounts and for the employment oriented domains with the social area in particular with the Labour Force Survey.

1. **The role of the statistical units to achieve coherence and consistency in ESS.**

*Statistical units are the entities for which information is sought and for which statistics are ultimately compiled[[5]](#footnote-5)*. Statistical units can be both physical entity identifiable in the real wold or may be a “theoretical” statistical construct. In case of the Business Statistics, a statistical unit can be a legal entity (recognizable through administrative or legal database) or unit like the “Enterprise” or the “Kind activity unit” or the “Institutional Units” defined by the need for statistical purposes.

The statistical unit can be of analysis or of observation. The unit of observation is at the level at which the data are collect. The unit of analysis is the object (*who*) of study about which (*what*) a statistician may generalize: *it is the analysis you do in your study that determines what the unit is.* The unit of observation and the unit of analysis can be the same but they need not be.

From a general point of view the chain of statistical production starts from scope/domains/objectives, then it identifies the variables to be collected for the statistical analysis and at the end it identifies the “statistical units of analysis” to which the statistical information have to be referred and the “observational units” able to collect the required data.

In the context of ESS’ coherence and consistency, statistical information are not comparable if different statistical units are used in different statistical domains or in different EU’ countries: to have a common definition of the statistical units is the necessary, even if not sufficient, condition to achieve coherence and consistency in the European Statistical System. With referring to this problem one of the first statistical regulation of the EU was in 1993 the *Council Regulation (EEC) No 696/93 “on the statistical units for the observation and analysis of the production system in the Community”*.

After 20 years, it is clear that the objective (in terms at least of coherence) is not achieved. The results of the “*ESSnet on consistency of concepts and methods of business-related statistics - 2010 project on statistical units*” well demonstrate both in qualitative and quantitative terms the present *not coherent* and *inconsistent* situation at EU level as regards the Business Statistics.

1. **The present situation.**

The ESSnet on consistency – project on statistical units - identified a list of problems[[6]](#footnote-6) both as respect to the to the application of the statistical units’ definition in different domains and as respect to the application of the statistical units’ definition in the MSs. In synthesis there is a not homogeneous identification of the statistical units in all domains and the NSIs do not implement in the same way (and in some case do not use) the definitions of the Regulation No 696/93. The general result of the study can be resumed as in the following.

*The ESS’ present situation is:* because we have a “common” and “official” and “mandatory” definition we “believe” we are able (and the users are able) to compare information produced by the EU NSIs and derive the necessary analysis. *The actual situation is:* comparing some EU indicators like “enterprise productivity” or “enterprise dimension” or the “GDP by branches” one of the determinants of the differences are the different criteria and practices in the identification of the “enterprise unit” used by the different NSIs and in different domains.

The cause of this situation depends on the evolution of the different part of the ESS takes place in the last 10-15 years, and in the different view and practices developed in different countries. Furthermore the need to reduce the NSI’ cost and the respondents’ burden has led the NSIs to identify and acquire the statistical data in different way and with different methods (e.g. massive use of administrative information) reducing the use of the statistical surveys: this is especially true in the field of the business statistics.

At same time, the cause must be sought even *within the current European system of definition of statistical units*. The following critical elements can be identified:

1. The Regulation 696/93 has been defined “before” the identification of the scope/domains/objectives (the Regulations on the various business domains) for which the statistical units have to be observed.
2. Some definitions are complex, difficult or expensive to be applied: e.g. the Kind of Activity Unit definition.
3. The enterprise definition general but at same time “*generic*”: it refers to corporations, self- employees, public administrations, outworkers, producers for own final use, non-profit units. Without any specifications (to better characterize and identify population – or subpopulation – of reference for statistics), such definition means all and nothing.
4. The enterprise definition is ambiguous because from one side it states: “*the enterprise is the smallest combination of legal units*”, in another side it states: “an enterprise may be a sole legal unit”. It uses some terms without definition: e.g., the meaning of “*current resources*” is not explained.
5. **A consistent definition of a statistical unit.**

This paper do not aim to discuss the contents of a new system of definitions of the statistical units **“for the observation and analysis of the production system in the EU”**. Some convergent proposals have done by both ESSnet Consistency[[7]](#footnote-7) and ESSnet Profiling of MNEs[[8]](#footnote-8). The proposals have been discussed and changed in a several Eurostat TFs and WGs.

It is well accepted the necessity to overcome the identified limitations and contradictions of the Regulation 696/93 and the need to increase the coherence of the European economic figures and to build consistent statistics. At the same time, must be understood as to avoid the risk of falling back into the same limitations and contradictions.

The "political will” to change must be a precondition. Because we are speaking in terms of (European) “system”, the changing means the reduction of the “power”, and in some cases of the needs, of the individual parts (each NSI, each domain) to have guarantee a “government of the system” adequately homogeneous. Without this “will” the evaluation of the competitiveness, the productivity, the profitability, the innovativeness of each regional (as respect other regions) or national (as respect other Nations) economy and of the whole European economy (as respect other economic and monetary areas) cannot be supported at all.

The second condition is to identify a conceptual “route” able to produce *consistent* *definition*s of the statistical units.

From the logic point of view, “*a deductive theory is called consistent or non-contradictory if no two asserted statements of this theory contradict each other”*. This statement means: the characteristic of a consistent theory is the lack of contradiction. The lack of contradiction can be defined in either semantic (the theory refers to a model) and syntactic (not permitting the deduction of a contradiction from the axioms) terms.

Using the previous logical approach is possible to identify the main *elements* that can characterize a definition of a statistical unit.

1. The *semantic* component of the definition describe the meaning of a statistical unit. This first element corresponds to the better identification of the context, objective and scope for which a statistical unit have to be defined. A statistical unit is defined referring to the need of a specific statistical analysis. The need analyze not only the national economy but also the European one and to have figures supporting the globalization of economy identifies in the different way the statistical units able both to collect data and to analyze the statistical information (Global/European enterprise instead only of the just National enterprise). In this case the achieving of the vertical consistency have to be take in consideration in the definition. The need to analyze the economic (e.g. productivity or competitiveness) or the production (physical quantification of the products) characteristics of a population can identity different “*meaning*” of the statistical units to be used. In the first case the meaning of the definition have to consider concepts like market/profit in the second it can consider other concepts. Prioritizing some objective instead others, is a necessary condition to avoid generic definitions and ambiguous (between counties and between domains) methods of statistical unit identification .
2. The *syntactic* of a statistical unit definition describe the “*how*” the definition have to be built. It is possible to define three main elements:
	1. Use of sentences not permitting the deduction of a contradiction from the different part of the definition (like the present definition of enterprise).
	2. To avoid the use of duplication of concepts.
	3. To avoid to use generic concepts not well explained. For instance in the last version of the proposed definition of enterprise what it means “sell in own will…”, what is the typology of concept that the sentence support? The juridical or economic type?
	4. Use of verbs that correctly specify the context of the definition, avoiding disambiguation or “free” interpretation. The use of the verb “*can*” (instead of “*are*”) can determine the following interpretation “*the definition may be what explicitly contains but can be other*”. This is the case of the sentence of “an enterprise can correspond…”.
	5. Need to describe explicitly all part of the definition. The presence of the following sentences “the statistical unit … have to respect the characteristics *a* and *b*“ need the exact description of both *a* and *b* parts.

A statistical unit definition is a theoretical (and synthetic) statistical construction. It have to be rigorous and strictly coherent between its specific parts. At same time, *how* in practice the units have to be identified is a fundamental element. To have a common definition of a specific statistical unit is a necessary element to guarantee coherence and consistency, but it is not sufficient.

1. The *semiotics* of a statistical unit definition describe the operative aspects of the definition. This element focuses on the structure of the definition more specifically. It is based on the identification of operative rules that can be both clarifications and specifications of the definition. The operative rules are fundamental elements for the implementation in practice and in homogeneous way (as regard different statistical domains and space) of the definition.The rules have not to be “recommendations”, they have to be considered in some way as part of the definition. Because “operative” they have to represent practical and exact solutions for specific aspects of the definitions that are considered relevant for a correct application of the statistical unit definition. Even in this case “generic” rules are not useful and dangerous.
1. Adopted by the European Statistical System Committee 28th September 2011. [↑](#footnote-ref-1)
2. <http://sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf> [↑](#footnote-ref-2)
3. ESSnet on consistency of concepts and methods of business-related statistics – 2010 project on statistical units. Deliverable 1.1: Direction Report. [↑](#footnote-ref-3)
4. The list can be found in: ESSnet on consistency, background of the project. [↑](#footnote-ref-4)
5. OECD Glossary of Statistical Terms. [↑](#footnote-ref-5)
6. For more precise analysis see: ESSnet on consistency of concepts and methods of business-related statistics - 2010 project on statistical units. Deliverables 3.1, 3.2, part 1 and 3.2 part 2. [↑](#footnote-ref-6)
7. ESSnet on consistency of concepts and methods of business-related statistics - 2010 project on statistical units. Deliverables 5.3.1 and 5.3.2. [↑](#footnote-ref-7)
8. ESSnet on profiling of large and complex multinational enterprises (MNEs). Deliverables WP\_B Statistical Units. [↑](#footnote-ref-8)