
TERMINOLOGY RELATING TO THE IMPLEMENTATION OF THE VISION ON THE PRODUCTION METHOD OF EU STATISTICS



Authors: W. Radermacher, A. Baigorri, D. Delcambre, W. Kloek, H. Linden

INTRODUCTION

Two important documents have been issued recently by EU institutions which will have a strong impact on future statistical work in the European Statistical System (ESS¹). These documents are :

- *Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics, and*
- *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade (Document COM(2009) 404 final).*

The Regulation on European statistics² proposes new measures (temporary direct statistical actions, European approach to statistics, collaborative networks, transmission of confidential data, etc.) to increase the flexibility of the European Statistical System and hence its ability to respond to new needs and challenges.

The Commission Communication (the Vision document) proposes a full re-engineering of the production methods of statistics within the EU, going from a production system often based on numerous parallel processes to a more integrated production model. The objectives of this

¹ The European Statistical System (ESS) is the partnership between the Community statistical authority, which is the Commission (Eurostat), and the national statistical institutes (NSIs) and other national authorities responsible in each Member State for the development, production and dissemination of European statistics.

² European statistics are relevant statistics necessary for the performance of the activities of the Community.

new system are to increase efficiency and cost-effectiveness as well as improve the coherence and comparability of data.

Closely linked to the Regulation on European Statistics and the Vision Communication is the process of combining, integrating and/or aggregating data taken from different sources. This process also contributes to the efficiency of the production statistics since existing data are re-used to produce new indicators providing new angles of analysis.

The present document aims at clarifying the new concepts introduced in these documents and defining existing terms in the new context. When necessary the definitions are supplemented with comments and/or examples. The document is divided into three main chapters and one annex, as follows:

- Chapter I : Terms relating to the Regulation on European statistics
- Chapter II : Terms relating to the Re-engineering of Statistical Production Methods
- Chapter III : Terms relating to the Aggregation of Statistical Data
- Annex : Types of Statistical Information

This document can therefore be considered as a terminology tool supporting the implementation of the statistical law and the new vision on statistical production.

TABLE OF CONTENTS

I. TERMS RELATING TO THE REGULATION ON EUROPEAN STATISTICS.....	5
<i>EUROPEAN APPROACH TO STATISTICS</i>	5
<i>TEMPORARY DIRECT STATISTICAL ACTIONS</i>	6
<i>COLLABORATIVE NETWORKS</i>	7
<i>TRANSMISSION OF CONFIDENTIAL DATA</i>	8
II. TERMS RELATING TO THE RE-ENGINEERING OF PRODUCTION METHODS.....	11
<i>STOVEPIPE MODEL (also known as PRODUCT STOVEPIPE MODEL)</i>	11
<i>AUGMENTED STOVEPIPE MODEL</i>	13
<i>INTEGRATED MODEL</i>	14
<i>CLUSTERS OF STATISTICS</i>	15
<i>MICRO-DATA LINKING</i>	15
<i>WAREHOUSE APPROACH</i>	16
III. TERMS RELATING TO THE AGGREGATION OF STATISTICAL DATA	17
<i>PRIMARY DATA</i>	17
<i>SECONDARY DATA</i>	18
<i>ACCOUNTING SYSTEMS</i>	19
<i>INDICATOR</i>	20
<i>INDICATOR SET</i>	20
<i>COMPOSITE INDICATOR</i>	21
ANNEX : TYPES OF STATISTICAL INFORMATION	23
<i>DATA</i>	23
<i>METADATA</i>	23
<i>STATISTICAL INFORMATION</i>	24
<i>PARADATA</i>	24
<i>MICRODATA</i>	25
<i>MESODATA</i>	25
<i>MACRODATA</i>	26
SOURCES	27

I. TERMS RELATING TO THE REGULATION ON EUROPEAN STATISTICS

The Regulation on European Statistics, popularly called the Statistical Law, lays down principles which will enable all actors in the European Statistical System (ESS) to develop, produce and disseminate statistics more efficiently. It is flexible enough to respond to future challenges, such as rapidly emerging policy needs, financial constraints, reduction of response burden and developments in information technology. At the same time it gives a stable and transparent legal basis, which guarantees the independence, integrity and accountability of Eurostat and the national statistical authorities.

The most innovative measures provided by the Regulation are the following:

- European approach to statistics
- Temporary direct statistical actions
- Collaborative networks
- Transmission of confidential data

These measures introduce terminology that is either new or is in need of clarification. This terminology is described below.

EUROPEAN APPROACH TO STATISTICS

DESCRIPTION/DEFINITION

Statistics which are relevant at EU level *only* should ideally be produced *only* at European level.

COMMENTS

This approach will be implemented only in specific and duly justified cases and within the framework of the European statistical programme³. It consists of a pragmatic strategy to facilitate the compilation of European statistical aggregates, representing the European Union as a whole or the euro area as a whole, which are of particular importance for Community policies.

³ The European statistical programme provides the framework for the development, production and dissemination of European statistics, the main fields and the objectives of the actions envisaged for a period not exceeding five years. It lays down priorities concerning the needs for information for the purpose of carrying out the activities of the Community. Those needs are weighed against the resources needed at Community and national levels to provide the required statistics, and also against the response burden and the respondent's associated costs.

The two main aspects taken into account to decide upon this approach are the following:

- The information is not necessary at national level;
- The aggregates are crucial to respond to an urgent need linked to Community policies.

Its objectives are:

- a) to maximise the availability of statistical aggregates at European level and improve the timeliness of European statistics,
- b) to reduce the burden on respondents, National Statistical Institutes (NSIs) and other national authorities based on a cost-effectiveness analysis.

Measures to implement the European approach to statistics shall be carried out with the full involvement of Member States. If necessary, a coordinated release and revision policy shall be established in cooperation with Member States.

The underlying idea behind this new approach is that having data that are reliable at the national level is a sufficient condition for having reliable data at the European aggregate level, but it is not a necessary condition. If the only purpose of the data is to provide information at EU level, there is no need for a full set of national data, and there is therefore a potential efficiency gain in the system. EU sampling is a possible way to make this gain real. In areas where there is no need to have national data, EU sampling may lead to a reduced burden on respondents, better timeliness, and improved quality. As already mentioned, the European approach to statistics could also include the production of European statistics on the basis of non-published national contributions or contributions from a subset of Member States, as well as on the basis of partial information by modelling techniques.

The European approach to statistics can be applied in different ways:

- a) in the production of European statistics on the basis of:
 - i. non-published national contributions or national contributions from a subset of Member States;
 - ii. specifically designed survey schemes;
 - iii. partial information by modelling techniques.
- b) in the dissemination of statistical aggregates at European level by applying specific statistical disclosure control techniques without national dissemination provisions being impaired.

TEMPORARY DIRECT STATISTICAL ACTIONS

DESCRIPTION/DEFINITION

Collection of data that can be implemented by Eurostat to quickly respond to unexpected needs, which are not (yet) covered in the five-year programme.

COMMENTS

So far collection of data in the ESS was strictly governed by the five-year European Statistical Programme. Only the actions which were explicitly foreseen in this programme could be implemented. The main drawback of such a procedure was that it was impossible for the ESS to respond quickly to new and unexpected needs.

The response to this problem was the introduction of temporary direct statistical actions. The fact that these actions are conducted outside the scope of the five-year programme does not mean that they do not have to comply to certain rules.

And indeed these actions must be authorised through Commission Decisions under the comitology procedure.

Furthermore to qualify for acceptance, a temporary action must meet two major conditions :

1. the action does not provide for data collection covering more than three reference years;
2. the data are already available or accessible within the NSIs and other national authorities responsible, or can be obtained directly via European samples with the adequate coordination with the NSIs and other national authorities.

The Regulation on European Statistics also stipulates that the Community shall make financial contributions to the NSIs and other national authorities to cover the incremental costs incurred on them as a result of the implementation of the actions.

COLLABORATIVE NETWORKS

DESCRIPTION/DEFINITION

Synergies to be developed within the European Statistical System (ESS) by the sharing of expertise, tools and results or by fostering specialisation on specific tasks. [The expertise gained in these specific activities will then be used by the whole ESS community.](#)

COMMENTS

The joint structures, tools and processes established or further developed through collaborative networks between partners of the ESS will facilitate specialisation by certain Member States in specific statistical activities for the benefit of the ESS as a whole. This will avoid duplication of work and thus increase efficiency and reduce the unnecessary burden on respondents.

The outcome of those actions, such as joint structures, tools, processes and methods, should be made available throughout the ESS. The initiatives for the creation of collaborative networks as well as the outcomes are examined by the ESS Committee⁴.

EXAMPLE(S)

ESSnets. These ESS collaboration networks consist of projects carried out by a team of institutions aiming at developing results which can be used by the whole ESS community. ESSnet projects are co-financed by the Commission and participating institutions.

Examples of EESnets are: "ESSnet on a common reference architecture", "Partnership Health", "MEETS⁵ - Improving consistency in the methodology and legal requirements for different areas of business and trade statistics", "MEETS - Use of administrative data", etc.

The ESSnets are very important in their function of strengthening partnership, enabling formal networking, developing expertise, creating explicit methodological knowledge for the ESS, providing training and raising awareness in specific domains.

TRANSMISSION OF CONFIDENTIAL DATA

DESCRIPTION/DEFINITION

The purpose of this article is to regulate the transmission of confidential data between ESS authorities or between an ESS authority and an ESCB (European System of Central Banks) member. Access to confidential data which only allow for indirect identification of the statistical units may also be granted to the research community.

COMMENTS

The philosophy behind this initiative is to maximise the benefits of the data with the aim of increasing the quality of European statistics and to ensure a flexible response to emerging statistical needs.

⁴ The European Statistical System Committee provides professional guidance to the ESS for developing, producing and disseminating European statistics in line with the statistical principles set out by the Commission (Professional independence, impartiality, objectivity, reliability, confidentiality and cost effectiveness). It is composed of representatives of the NSIs who are national specialists for statistics and chaired by the Commission. The ESS Committee can be consulted on many issues: development, production and dissemination of European statistics, their justification on a cost-effectiveness basis, the means and timetables for achieving them, the response burden on survey respondents; statistical confidentiality; development of the Code of Practice; methodology; etc.

⁵ MEETS = Modernisation of European Enterprise and Trade Statistics (A programme established to develop target sets of indicators and to achieve a streamlined framework of business related statistics. Successor of EDICOM)

However, the transmission of confidential data will be only authorised under certain conditions:

- It is **allowed** provided that this transmission is necessary for the efficient development, production and dissemination of European statistics or for increasing the quality of European statistics, and that this necessity has been justified. Each transmission of data requires explicit authorisation by the authority that collected the data.
- It is **obligatory** when an act of the European Parliament and of the Council provides for the transmission of such data (this means that no confidentiality argument may be invoked to refuse the transmission of such data).
- Any further transmission beyond the first transmission shall require the explicit authorisation of the authority that collected the data.
- Confidential data shall be used
 - o exclusively for statistical purposes
 - o and only accessible to staff working in statistical activities within their specific domain of work.

II. TERMS RELATING TO THE RE-ENGINEERING OF PRODUCTION METHODS

The European Statistical System needs to streamline and improve its production process of EU statistics considerably in order to achieve productivity gains. These productivity gains are responding to stable or even decreasing resources available within the ESS and to user demands for new and additional statistics.

To improve this situation the Commission, in its document "Vision for the next decade", proposes to replace the present pre-dominant model generally referred to as the "stovepipe model" with an integrated model with a view to improve efficiency and reduce costs and response burden.

These models introduce new or updated terminology which will be explained in the first part of this section.

The proposed integration process will occur by the use of standardised tools and techniques (either at micro- or macro-level), the most promising being the following:

- Clustering of statistics
- Micro-data linking
- Warehouse approach

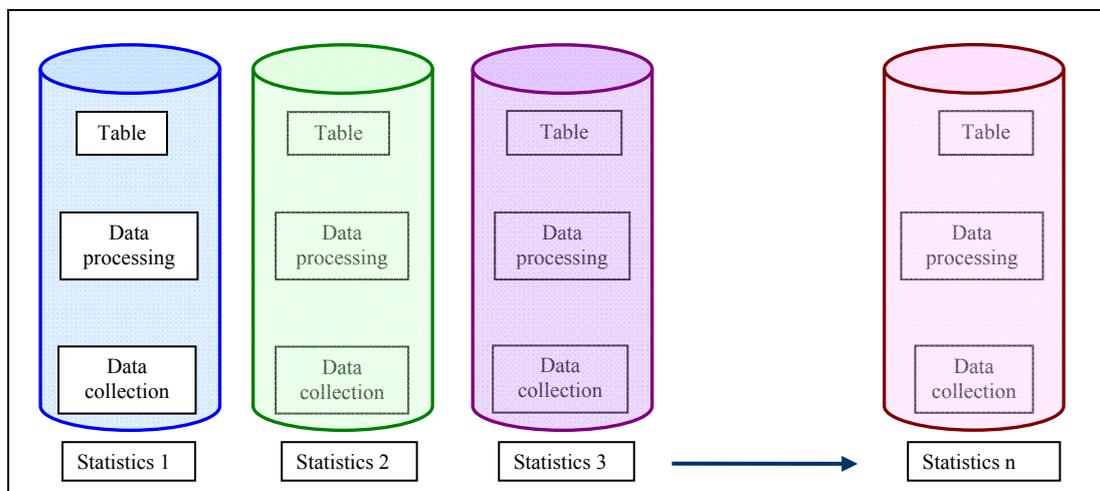
These aspects are further elaborated in the second part of this section.

STOVEPIPE MODEL (ALSO KNOWN AS PRODUCT STOVEPIPE MODEL)

DESCRIPTION/DEFINITION

Business model where the production processes are organised by distinct statistical products. This expression is used to illustrate how statistics are traditionally produced within the ESS, i.e. in numerous parallel processes, country by country (even sometimes region by region) and domain per domain. In such a model every single product stovepipe corresponds to a specific domain of statistics, together with the corresponding production system. For each domain, the whole production process from survey design over data collection and processing to dissemination takes place independently of other domains, and each domain has its own data suppliers and user groups. The stovepipe model is also reflected in the way statistical domains are regulated at the European level.

STOVEPIPE MODEL "DOMAIN BY DOMAIN"



COMMENTS

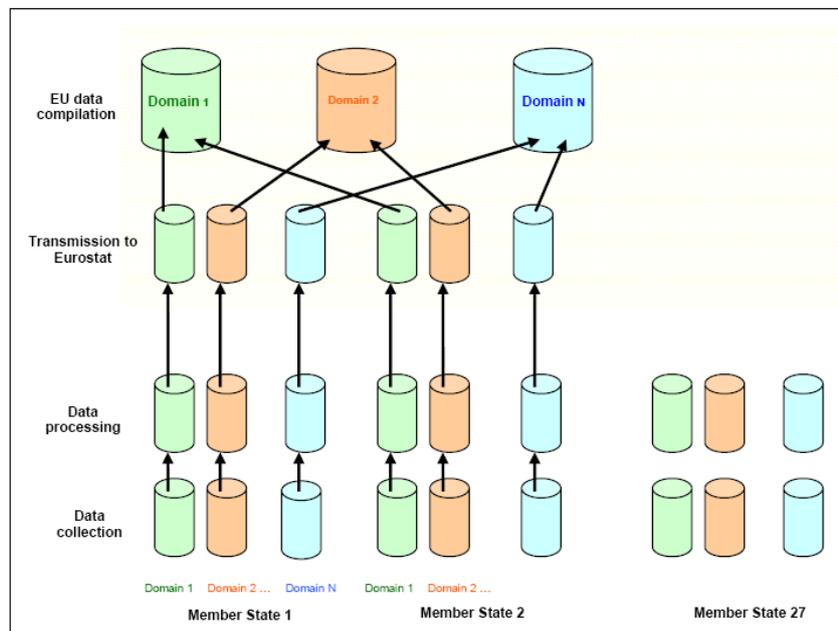
The stovepipe model is the outcome of a historic process in which statistics in individual domains have developed independently. It has a number of advantages: the production processes are best adapted to the corresponding products; it is flexible in that it can adapt quickly to relatively minor changes in the underlying phenomena that the data describe; it is under the control of the domain manager and it results in a low-risk business architecture, as a problem in one of the production processes should normally not affect the rest of the production. From a European perspective it has the advantage that it can be addressed by a relatively limited and specific Regulation.

However, the stovepipe model also has a number of disadvantages. First, it may impose an unnecessary burden on respondents when the collection of data is conducted in an *uncoordinated* manner and respondents are asked for the same information more than once. Second, the stovepipe model is not well adapted to collect data on phenomena that cover multiple dimensions, such as globalisation, sustainability or climate change. Last but not least, this way of production is inefficient and costly, as it does not make use of standardisation between areas and collaboration between Member States. Redundancies and duplication of work, be it in development, in production or in dissemination processes are unavoidable in the stovepipe model. These inefficiencies and costs for the production of national data are further amplified when it comes to collecting and integrating regional data, which are indispensable for the design, monitoring and evaluation of some EU policies.

AUGMENTED STOVEPIPE MODEL

DESCRIPTION/DEFINITION

As indicated in the previous paragraph, the stovepipe model describes the pre-dominant situation within the ESS where statistics are produced in numerous parallel processes. The adjective "augmented" indicates that the same model is reproduced and added at Eurostat level.



COMMENTS

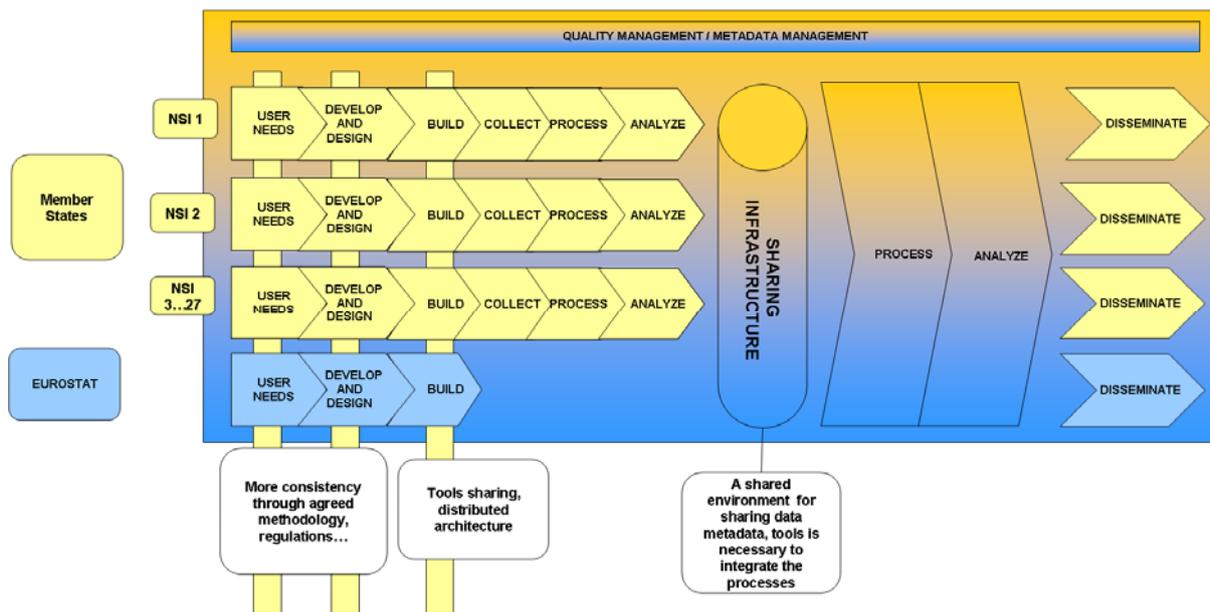
In order to produce European statistics, Eurostat compiles the data coming from individual NSIs also area by area. The same stovepipe model thus exists in Eurostat, where the harmonised data in a particular statistical domain are aggregated to produce European statistics in that domain. The traditional approach for the production of European statistics based on the stovepipe model can thus be labelled as an "augmented" stovepipe model, in that the European level is added to the national level.

INTEGRATED MODEL

DESCRIPTION/DEFINITION

Innovative way of producing statistics based on the combination of various data sources in order to streamline the production process. This integration is twofold:

- a) **horizontal integration** across statistical domains at the level of National Statistical Institutes and Eurostat. Horizontal integration means that European statistics are no longer produced domain by domain and source by source but in an integrated fashion, combining the individual characteristics of different domains/sources in the process of compiling statistics at an early stage, for example households or business surveys.
- b) **vertical integration** covering both the national and EU levels. Vertical integration should be understood as the smooth and synchronized operation of information flows at national and ESS levels, free of obstacles from the sources (respondents or administration) to the final product (data or metadata). Vertical integration consists of two elements: joint structures, tools and processes and the so-called European approach to statistics (see this entry).



COMMENTS

The present "augmented" stovepipe model, has a certain number of disadvantages (burden on respondents, not suitable for surveying multi-dimensional phenomena, inefficiencies and high costs). By integrating data sets and combining data from different sources (including administrative sources) the various disadvantages of the stovepipe model could be avoided. This new approach would improve efficiency by

elimination of unnecessary variation and duplication of work and create free capacities for upcoming information needs.

However, this will require an investigation into how information from different sources can be merged and exploited for different purposes, for instance by eliminating methodological differences or by making statistical classifications uniform.

CLUSTERS OF STATISTICS

DESCRIPTION/DEFINITION

Statistics that can be grouped together because they can be considered as covering various aspects of a more general domain and are based on common elements such as statistical units, classifications, etc.

COMMENTS

It is possible to reason in terms of clusters of statistics only if the basis for the collection and processing of these statistics (classifications, statistical units, etc) is sufficiently harmonised to qualify for potential integration. In business or household statistics significant efficiency gains could be generated via clustering.

EXAMPLE(S)

Examples are business statistics and households statistics where the statistical unit is the same (enterprise and household respectively), classifications used are identical (for instance, NACE and CPA for business statistics, ISCED and ISCO for household statistics).

MICRO-DATA LINKING

DESCRIPTION/DEFINITION

Establishing connections between data sets at micro-data level (individual statistical units such as enterprises or persons).

COMMENTS

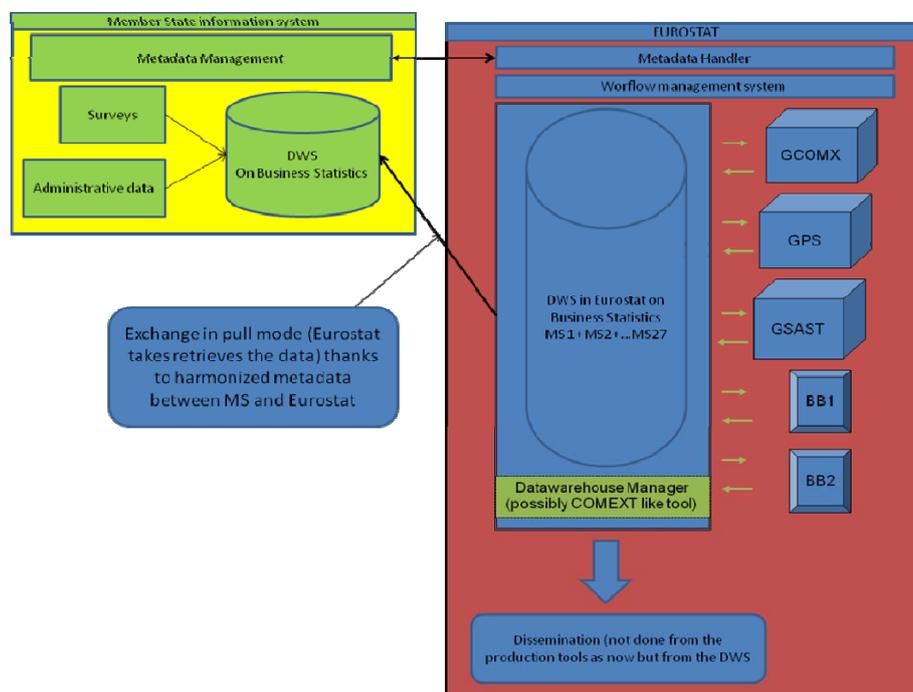
Linking micro-data helps improving the production of statistics in the sense that this will not only reduce the response burden (by not asking several times for the same piece of information) but also produce better comparable data sets and produce new kinds of output (since new links between variables will be established).

WAREHOUSE APPROACH

DESCRIPTION/DEFINITION

The warehouse approach provides the means to store data once, but use it for multiple purposes. A data warehouse treats information as a reusable asset. Its underlying data model is not specific to a particular reporting or analytic requirement. Instead of focusing on a process-oriented design, the underlying repository design is modelled based on data inter-relationships that are fundamental to the organisation across processes.

Conceptual model of data warehousing in the ESS (European Statistical System)



COMMENTS

Based on this approach statistics for specific domains should not be produced independently from each other, but as integrated parts of comprehensive production systems, called data warehouses. A data warehouse can be defined as a central repository (or "storehouse") for data collected via various channels.

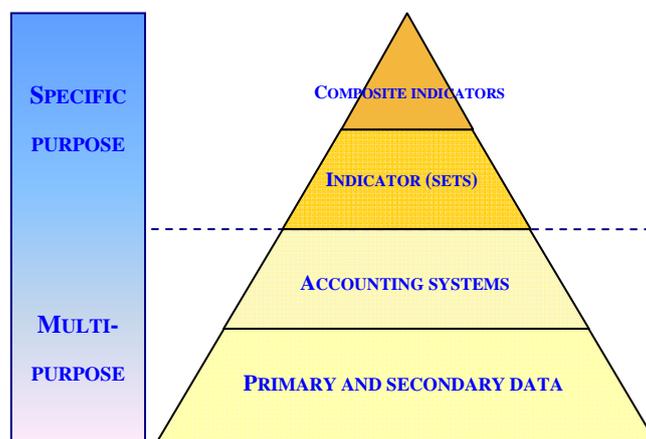
Data validation, integration, and provisioning into an dissemination database typically accounts for a large part of the statistical production costs, especially in the statistical fields where a lot of metadata would need to be harmonised or fine-tuned (for instance, statistical units, concepts and classifications). But by sourcing data into a reusable information asset in the form of a well-designed warehouse, these costs are incurred just once for the organisation, rather than multiple times in the case of multi-stovepipe deployments.

III. TERMS RELATING TO THE AGGREGATION OF STATISTICAL DATA

Although not explicitly mentioned in the Regulation on the statistical law or in the "Vision document" the maximum re-use of existing information or the development of new aggregation levels (based on existing data sources) aiming at creating intermediary levels of analysis are two important aspects in improving the efficiency and the sustainability of the ESS production system.

It was therefore considered useful to better define here the scope and coverage of these information levels as well as to describe the statistical terminology used.

PYRAMID OF STATISTICAL INFORMATION



PRIMARY DATA

DESCRIPTION/DEFINITION

Data collected by statistical authorities, via traditional statistical activities (sample surveys, censuses, etc.) for statistical purposes.

COMMENTS

"Use for statistical purposes" means that data collected will be used exclusively for the compilation of statistical tables or for statistical economic analyses; such data may not be used for administrative, legal or tax purposes or for verification against the units surveyed.

SECONDARY DATA

DESCRIPTION/DEFINITION

Data collected for administrative purposes but used by statistical authorities for statistical purposes (usually referred to as data from administrative sources).

COMMENTS

This wide definition complements the definition of primary data since it covers all data not primarily collected for statistical purposes, regardless of the origin. Governments for instance collect data for many non-statistical purposes, such as tax and labour market policies. Access to such data sets generates efficiency gains in the production of official statistics.

Administrative records can be used in many ways in the production of statistics. As a direct data source, statistics are compiled from:

- Single administrative sources (i.e. administrative registers), e.g. statistics on migrants, unemployment beneficiaries.
- Event-reporting systems, e.g. crime statistics, statistics of road accidents and statistics of the causes of death. In these cases the responsible administrative authority (police, hospitals, customs, etc.) report an event, including a number of variables characterising the event.
- Integration of several administrative sources in order to obtain new populations and variables (e.g. population censuses in some Scandinavian counties).

Administrative sources can also play auxiliary role in the statistical process: they can be used to create or supplement survey frames, to substitute survey data, as an input for validation, imputation, as aid in estimation, weight calculation or calibration and quality studies.

EXAMPLE(S)

An example for the use of administrative sources can be statistical tabulations that are produced from an administrative database maintained by the agency responsible for higher education.

ACCOUNTING SYSTEMS

DESCRIPTION/DEFINITION

Coherent, consistent and integrated set of accounts, balance sheets and tables based on a set of internationally agreed concepts, definitions, classifications and accounting rules. It provides a comprehensive accounting framework which ensures a high profile of consistency and comparability and within which primary and secondary statistical data can be compiled and presented in a format that is designed for the purposes of analysis, decision-taking and policy-making.

COMMENTS

The system of economic accounts is the oldest and most developed accounting system. The accounts present in a condensed way a great mass of detailed information, organized according to economic principles and perceptions, about the working of an economy. They provide a comprehensive and detailed record of the complex economic activities taking place within an economy and of the interaction between the different economic agents, and groups of agents, that takes place on markets or elsewhere. Other accounting systems can be either linked to the system of economic accounts (for instance environmental accounts) or can be separate systems (for instance material flow accounting).

In the future, i.e. for the future generation of statistical infrastructure, this layer of information will play a strategic role. If designed in an intelligent way it will ensure that a maximum of user needs (and many of them will not be known in detail beforehand) can be satisfied.

In the context of the pyramid of statistical information accounting systems is used in a very broad sense to indicate a common meso-level of information at which data from various domains can be presented. For instance the aggregation to broad branches of economic activity of both economic variables and environmental variables.

EXAMPLE(S)

European system of national and regional accounts, satellite environmental accounting system, economic accounts for agriculture and forestry, etc.

Based on the example of National Accounts (and particularly the Input-Output concept), one can distinguish the following building blocks:

- **Actors:** Households, natural persons, economic units (e.g. establishments)
- **Flows:** Activities of actors, transactions between them (production, income, consumption), migration, etc.
- **Stocks:** Produced and non-produced stocks, investments, depreciation
- **Products:** Goods and services, labour, energy, information
- **Measurement units:** Prices, joule, tonnes, meters, time, number of pieces

- **Aggregates:** Sectors, regions, periods, social strata, classes of economic activity, clusters of resources, commodity groups, and others (for instance, markets)
- **Systems:** National market economy in one period, National economy in one period (including household production), regional ecosystems, global ecosystem, etc.

INDICATOR

DESCRIPTION/DEFINITION

An indicator is a summary measure related to a key issue or phenomenon and derived from a series of observed facts. Indicators can be used to reveal relative positions and/or show positive or negative change.

COMMENTS

When evaluated at regular intervals, an indicator can point out the direction of change across different units and through time. In the context of policy analysis, indicators are useful in identifying trends and drawing attention to particular issues. They can also be helpful in setting policy priorities and in benchmarking or monitoring performance.

Normally the basic data are not directly applicable because they include besides the "signal" also different forms of "noise". Therefore, models of analytical statistics are used to separate the signal that is relevant for one specific purpose (from GDP to seasonal adjusted quarterly growth rates, from household income to poverty rates, from births data to fertility rates, from deaths to standardised mortality, from price information to price indices, etc.). If the analytical model is part of the production in official statistics it has to comply with specific requirements (for instance, published standards).

INDICATOR SET

DESCRIPTION/DEFINITION

Multivariate collection of indicators which should cover a broader field of application or a political area; the selection of the set is based on a model which aims at high quality of the entire set. The entire set of indicators has to be relevant; an optimal mix between theory-driven and policy-driven indicators is crucial in its composition.

COMPOSITE INDICATOR**DESCRIPTION/DEFINITION**

A composite indicator is formed when individual indicators are combined into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured.

COMMENTS

A composite indicator measures multi-dimensional concepts (e.g. competitiveness, industrialisation, sustainability, single market integration, knowledge-based society, e-trade or environmental quality) which cannot be captured by a single indicator. Ideally, a composite indicator should be based on a theoretical framework / definition, which allows individual indicators / variables to be selected, combined and weighted in a manner which reflects the dimensions or structure of the phenomena being measured.

ANNEX : TYPES OF STATISTICAL INFORMATION

In this annex the reader will find definitions of basic concepts (for instance, data and metadata) used in the everyday work within the ESS and which related to different data aspects. By explicitly defining them here it is hoped to avoid ambiguities about their scope and content. However, other well-established statistical terms found in the statistical literature are not included here and can easily be found in statistical glossaries.

DATA

DESCRIPTION/DEFINITION

Characteristics or information, usually numerical, that are collected through observation.

COMMENTS

Data are the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means. One important aspect of statistical data is that they are organized for analysis.

Statistical data are collected via censuses and surveys or elaborated from available primary (statistical) and secondary (non-statistical) data. In the ESS the data collection is based mainly on legal acts and based on strict methodology procedures which ensure consistency and comparability of the information gathered.

METADATA

DESCRIPTION/DEFINITION

Information that is needed to be able to use and interpret statistics. Metadata describe data by giving definitions of populations, objects, variables, the methodology and quality.

COMMENTS

A distinction is generally made between structural and reference metadata.

Structural metadata is used to identify, formally describe or retrieve statistical data, such as dimension names, variable names, dictionaries, dataset technical descriptions, dataset locations, keywords for finding data etc. For example, structural metadata includes the titles of the variables and dimensions of statistical datasets, as well as the units employed, code lists (e.g. for territorial coding), data formats, potential value ranges, time dimensions, value ranges of flags, classifications used, etc.

Reference metadata (sometimes called explanatory metadata) describes the contents and the quality of the statistical data from a semantic point of view. It includes explanatory texts on the context of the statistical data, methodologies for data collection and data aggregation as well as quality and dissemination characteristics. In Eurostat this information is presented in files based on a standardised format called ESMS (Euro SDMX Metadata Structure) This standard is based on previous standards (including the Standard Data Dissemination Standard developed by the International Monetary Fund) but has been substantially extended in order to include more information on data quality. These files are associated to the data series published by Eurostat on its web site.

STATISTICAL INFORMATION

DESCRIPTION/DEFINITION

Statistical data together with associated metadata.

PARADATA

DESCRIPTION/DEFINITION

Information related to a statistical data collection or production process.

COMMENTS

It is distinct from the information that is the objective of the statistical data collection or production process. Examples of paradata are: whether a unit has been selected into a sample; whether a selected unit has responded; number of attempts to reach a selected unit; mode of collection (telephone, personal, etc.).

Survey paradata, i.e. information about the process of survey data collection, has steadily increased in the breadth and depth of use. From modest beginnings in the 1980s as trace files of the fields entered, keys pressed, and timings during computer-assisted interviews, the value of the information for addressing significant challenges in the survey process has been explored and applied in increasingly diverse ways.

MICRODATA

DESCRIPTION/DEFINITION

Non-aggregated observations, or measurements of characteristics of individual units.

MESODATA

DESCRIPTION/DEFINITION

Mesodata is a concept that provides an intermediate conceptual layer between microdata and macrodata.

COMMENTS

The advantage of mesodata is that it provides a further level of analysis of statistical information independent of the levels of microdata and macrodata because statistical information is grouped in a way that opens up new angles of analysis of a phenomenon.

EXAMPLE(S)

Macrodata refer to tabulated aggregates for national or large geographic areas, mesodata to tabulated data for sufficiently small geographic areas so that the results can be used operationally to identify and target a vulnerable population subgroup, and microdata refer to identifiable records for individual units.

A concrete example of this are the hypercubes defined in the Census Hub Architecture⁶. These hypercubes make it possible to cross-tabulate variables, for instance age, sex, current activity status and geographical area.

⁶ The Census European Hub is the proposal for a conceptually new system to achieve the dissemination of the 2011 census data. It is based on the concept of data sharing, where a group of partners agree on providing access to their data according to standard processes, formats and technologies. For this purpose SDMX standards will be used.

MACRODATA**DESCRIPTION/DEFINITION**

The result of aggregation of data on individual units or collections of units to a higher level.

EXAMPLE(S)

Data on individual enterprises (= microdata) can be aggregated at the level of branches of industry (= mesodata) and at the level of the whole economy (= macrodata).

SOURCES

- BAIGORRI, Antonio, LAUX, Richard, RADERMACHER, Walter, "Building confidence in the use of administrative data for statistical purposes", paper presented at the 2009 ISI conference in Durban
- BROBST Stephen, "Compliance as a Trojan Horse: Funding Your Enterprise Data Warehouse", article published on "Business Intelligence Best Practices" web site.
- Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000
- European Statistics Code of Practice for the national and Community statistical authorities
- European Union, 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(2009) 404 final
- European Union, Council Regulation (EEC, Euratom) No 1588/90 of 11 June 1990 on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities
- European Union, Recommendation of the Commission on the independence, integrity and accountability of the national and Community statistical authorities, COM/2005/0217 final, Brussels, 2005
- European Union, Regulation (EC) No 223/2009 of the European Parliament and of the Council, of 11 March 2009, on European statistics, Official Journal of the European Union No L 87, 31.3.2009, p. 164 – 173
- Eurostat, "Business Register Recommendations Manual" (2003)
- Eurostat, CODED database (Concepts and Definitions database)
- [Eurostat, European Statistical System Committee \(ESSC\), 5th meeting, Item 6 of the agenda "Communication from the Commission to the European Parliament and the Council on the production method of EU statistics -joint strategy paper", Luxembourg, May 2010](#)
- Eurostat, "Guidelines for assessing statistical processes based on administrative source(s) with the ESPAC Checklist", September 2009.
- Eurostat, "Report of the Think Tank ESS Business Architecture", 16 June 2009, Internal document
- Eurostat, "Towards a New European Statistical System", document to be submitted to management bodies of the ESS (such as the ITDG, DIME or other similar groups)
- O'REILLY, Jim, "Paradata and Blaise. A Review of Recent Applications and Research", Westat, USA
- Organisation for Economic Co-operation and Development (OECD) and Joint Research Centre (JRC) of the European Commission, "Handbook on Constructing Composite Indicators: Methodology and User Guide", Paris, 2008.

-
- PATRUNO, Vincenzo, "The Census European Hub Project", Workshop on Data Transmission, Becici (Montenegro), June 2009
 - RADERMACHER, Walter, "Big potential in new Statistical Law", interview published on InSite, the website of the European Statistical System
 - RADERMACHER, Walter, Framework note "Terminology in Statistics: Some fundamental definitions", unpublished.
 - RADERMACHER, Walter, "The middle layer in a hierarchical multipurpose statistical infrastructure", November 2009", unpublished
 - RUŽIĆ, Genovefa, HREN Karmen, "New regulation on European Statistics. Increased flexibility of the European Statistical System", Statistical Office of the Republic of Slovenia, 2009
 - SDMX, Statistical Data and Metadata Exchange initiative, sponsored by BIS, ECB, Eurostat, IMF, OECD, UNSD and World Bank, 2009.
 - Statistical Programme Committee, 68th Meeting, "Proposal for the creation of a ESS Methodology Directors' Group", Luxembourg, 13 november 2008
 - Statistics Canada, on-line thesaurus
 - Statistics New Zealand, on-line "Indicator Guidelines"
 - The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003
 - United Nations, Economic and Social Council, Economic Commission for Europe, Conference of European Statisticians, Seminar on strategic issues in business statistics, Session 1: Reducing respondent burden, "The measurement and Reduction of Statistical Respondent Burden on Businesses: The Current State of Affairs", note by Eurostat, Geneva, June 2009
 - United Nations, "System of National Accounts (SNA) 1993", United Nations, New York, 1993
 - WALLGREN, Anders and Britt (Statistics Sweden), "Register-based Statistics. Administrative Data for Statistical Purposes", John WILEY & Sons, Ltd, England, 2007