



EUROPEAN
STATISTICAL
SYSTEM

European Statistical System

Data collection

The link between national level and EU context

Data collection at national level

Many and different type of data sources....

- * Administrative data
- * Survey
- * Census
- * Other data sources...

Data collection at national level

Administrative data versus statistical data

- * Administrative data refers to information collected primarily for administrative (not statistical) purposes.
- * This type of data is collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.
- * Government departments are the main (although not exclusive) purveyors of large administrative databases, including welfare, tax, health and educational record systems.
- * These datasets have often been used to produce official statistics to inform policy-making.

Data collection at national level

Administrative data versus statistical data

- * Administrative datasets typically cover samples of individuals and time periods not normally financially or logistically achievable through survey methods. Alongside cost savings, the scope of administrative data is often cited as its main advantage for research purposes. Other advantages include relieving the burden on survey respondents and providing data on individuals who would not normally respond to surveys.

Data collection at national level

Administrative data versus statistical data

- * The criticisms levelled at these resources relate to the lack of control during the data entry stage and how this affects what can be done with the data.
- * More general concern has also been voiced about the lack of well established theory and methods to guide the use of administrative data

Data collection at national level

Advantages of administrative data	Disadvantages of administrative data
Already collected for operational purposes and therefore no additional costs of collection (though costs of extraction and cleaning).	Information collected is restricted to data required for administrative purposes – limited to users of services and administrative definitions.
Regularly (sometimes continuously) updated.	Possible lack of context information
Can provide historical information and allow consistent time-series to be built up.	Changes to administrative procedures could change definitions and make difficult comparison over time problematic.

Data collection at national level

Advantages of administrative data	Disadvantages of administrative data
Near 100% coverage of population interest.	Quality issues with variables less important to the administrator e.g. address details may well not be updated.
Regularly (sometimes continuously) updated.	Possible fragmentation of data sets
Captures individuals who may not respond to surveys.	Metadata issues (may be lacking or of poor quality)

Data collection at national level

Possible scenarios

- * 1. Direct use of administrative data to produce national economic and social statistics, for example crime rates, election statistics and employment statistics
- * 2. Linking different complementary administrative datasets. Data linkage is facilitated through concerted collaboration efforts between data holding authorities, and a well established unified system (including personal identity codes for accessing to different datasets).
- * 3. Combining survey and administrative data.

Data collection at national level

Data linkage

- * There are various ways in which extracts of administrative data can be linked with other data sources to create more comprehensive and effective datasets for analysis. The most obvious is the linkage of different years of data within a data source.
- * 1. Linking **individual level administrative data** with other **individual level administrative data**
- * 2. Linking **individual level administrative data** with **cross-sectional or longitudinal survey data**
- * 3. Linking **individual level administrative data** with **contextual information**

Data collection at national level

Survey

- * **OECD Glossary: survey** is an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their different profiles

Data collection at national level

Survey

makes international comparisons easier, compared to registration data

However


- * From one country to another, national survey may somewhat differ along a number aspects, such as the sample design, the coverage and timing of the survey, etc.

Data collection at national level

Census...

is the procedure of systematically acquiring and recording information about the members of a given population.

Data collection at national level

 STATISTICAL SYSTEM

Census dataMetadataData on qualityEnglish (e)

1.Select data2.Select layout3.Display data4.Download

Show data on **persons** ?

Geographic level

☒ Residence ?

☐ Place of work ?

☒ nations

☐ NUTS2 regions

☐ NUTS3 regions

☐ municipalities

Topic(s)

☐ Sex ?

☐ Age ?

☐ Marital status ?

☐ Family status ?

☐ Household status ?

☐ Current activity status ?

☐ Occupation ?

☐ Industry ?

Residence - nations
(0 of 32)**Residence - nations**

☐ all countries

☐ Belgium ?

☐ Bulgaria ?

☐ Czech Republic ?

☐ Denmark ?

☐ Germany ?

☐ Estonia ?

☐ Ireland ?

☐ Greece ?

☐ Spain ?

Select allDeselect allFind geo area

Cells selected0

Maximum selection100 000

BackNext

Data collection at EU level

Adopting the EU perspective

Business objectives:

- * Improve the overall efficiency of statistical production in Europe;
- * Improve the quality of statistical production;
- * Simplify/reduce requirements put on respondents;
- * Improve responsiveness to policy needs.

Criteria



EFFICIENCY

Efficiency means achieving maximum output from a given level of resources used to carry out an activity

EFFECTIVENESS

- Effectiveness means the extent to which the activity's stated objectives have been met

EFFICACY

- Efficacy means the power to produce a desired result or effect

Data collection at European level

Main constraints

- * 1. Diversity of national environments;
- * 2. Limited resources;
- * 3. Subsidiarity principle;
- * 4. Preservation of statistical confidentiality

Constraint: Subsidiary principle

The principle of subsidiarity is fundamental to the functioning of the European Union and more specifically to European decision-making. In particular, the principle determines when the EU is competent to legislate, and contributes to decisions being taken as closely as possible to the citizen.

The principle of subsidiarity is established in Article 5 of the Treaty on European Union. It appears alongside another principle that is also considered to be essential to European decision-making: the principle of proportionality.

Constraint: Subsidiary principle

The principle of subsidiarity aims at determining the level of intervention that is most relevant in the areas of competences shared between the EU and the Member States.

The Protocol on the application of the principles of subsidiarity and proportionality lays down three criteria aimed at establishing the desirability of intervention at European level:

- Does the action have transnational aspects that cannot be resolved by Member States?
- Would national action or an absence of action be contrary to the requirements of the Treaty?
- Does action at European level have clear advantages?

Statistical confidentiality

Unece Guidelines



Managing Statistical Confidentiality & Microdata

A common objective:

Quality data



The Quality assessment

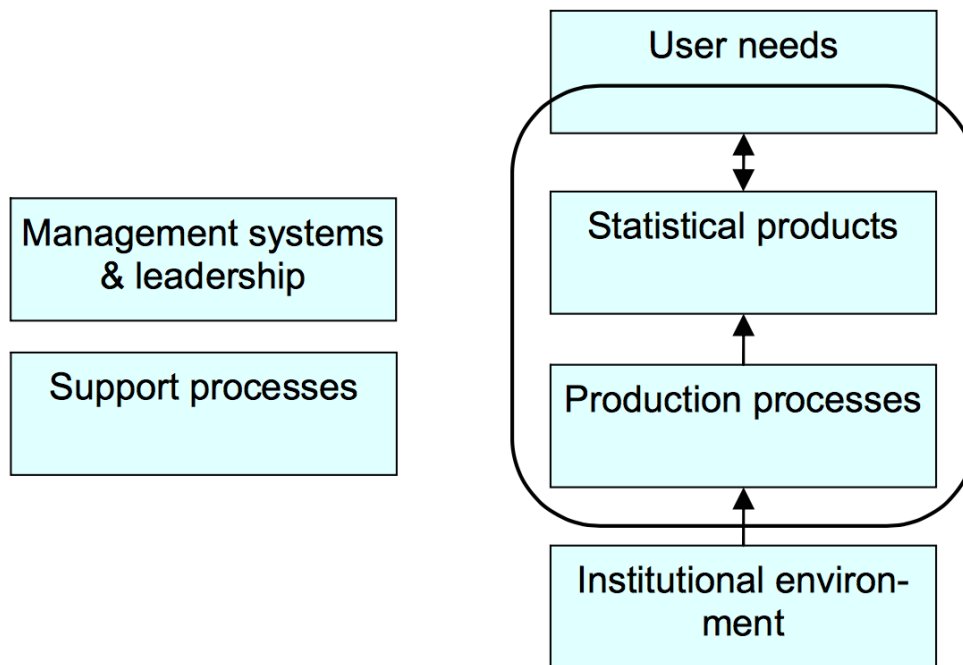
The European Statistics Code of Practice highlights the importance of data quality assessment in several instances. Its principles require an assessment of the various product quality components

The quality assessment

Production of high quality statistics depends on the assessment of data quality.

Without a systematic assessment of data quality, the statistical office will risk to lose control of the various statistical processes such as data collection, editing or weighting.

Elements of a quality management system

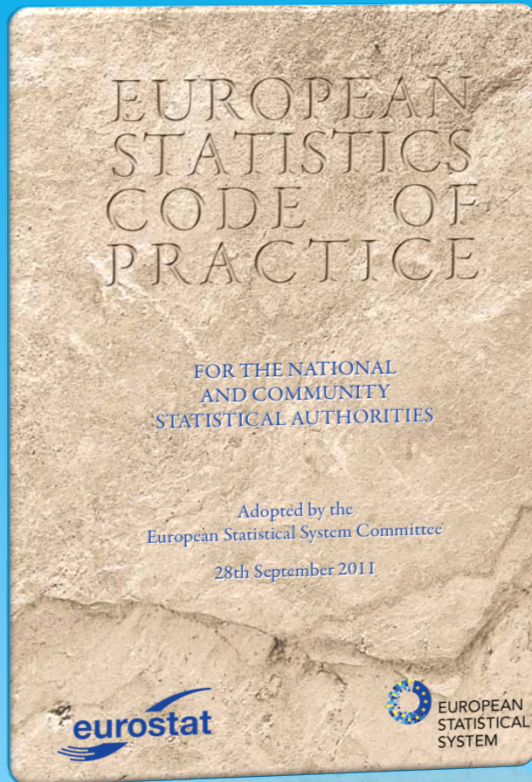


Corresponding principles from the European Statistics Code of Practice

Relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, accessibility and clarity

Sound methodology, appropriate statistical procedures, non-excessive burden on respondents, cost effectiveness

Professional independence, mandate for data collection, adequacy of resources, quality commitment, statistical confidentiality, impartiality and objectivity



Quality data criteria

- 1 relevance,
2. accuracy,
3. timeliness and punctuality,
4. accessibility and clarity
5. comparability and
6. coherence.

Quality criteria

* *Relevance*

Relevance is the degree to which statistics meet current and potential user needs. It refers to whether all statistics that are needed are produced and the extent to which concepts (definitions, classifications etc.) reflect user needs.

* *Accuracy*

Accuracy in the general statistical sense denotes the closeness of computations or estimates to the (unknown) exact or true values. Statistics are never identical with the true values because of variability (the statistics change from implementation to implementation of the survey due to random effects) and bias (the average of the estimates from each implementation is not equal to the true value due to systematic effects). A basic distinction is between sampling and non-sampling errors, which are both subject to variability as well as bias

Quality criteria

* *Timeliness and punctuality*

Timeliness of information reflects the length of time between its availability and the event or phenomenon it describes.

Punctuality refers to the time lag between the release date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar, laid down by regulations or previously agreed among partners.

Quality criteria

- * *Accessibility and clarity*
- * Accessibility refers to the physical conditions under which users can obtain data: where to go, how to order, delivery time, clear pricing policy, convenient marketing conditions (copy- right, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Inter- net etc.) etc.
- * Clarity refers to the data's information environment whether data are accompanied with appropriate documentation and metadata, illustrations such as graphs and maps, whether information on their quality is also available (including limitation in use etc.) and the extent to which additional assistance is provided by the NSI.

Quality criteria

* Comparability

Comparability aims at measuring the impact of differences in applied statistical concepts and measurement tools/procedures when statistics are compared between geographical areas, non-geographical domains, or over time. It is the extent to which differences between statistics are attributed to differences between the true values of the statistical characteristic

* Coherence

Coherence of statistics is their adequacy to be reliably combined in different ways and for various uses. When originating from different sources, and in particular from statistical surveys of different nature and/or frequencies, statistics may not be completely coherent in the sense that they may be based on different approaches, classifications and methodological standards.

Quality indicators

Quality indicators are specific and measurable elements of statistical practice that can be used to characterise the quality of statistics.

The use of indicators in describing and monitoring events/facts can be considered as a general method.

The indicators, as simplified and generally quantified measures – calculated according to clear rules – intend to characterise a complex phenomenon, in this case the many different quality aspects of data.

Examples of quality indicators

Quality policies

On 16th December 2014, Eurostat adopted a new policy on the management of errors in disseminated data.

This policy describes how and in which circumstances errors in the data are corrected and guarantees that users are informed about such non-scheduled revisions in an appropriate way.

The policy covers significance levels for errors' correction, commitment to correct significant errors as soon as possible, communication aspects and the approach for correction of past errors.

Quality policies

The quality policy includes:

- * The criteria used to decide the significance of the error and its treatment;
- * The guidelines on how to inform users about non-scheduled revisions (corrections) and their underlying causes.

The implementation of the policy is supported by a Eurostat internal procedure, internal communication mechanisms and reporting template supporting the implementation of the policy.

Quality policies

How significant errors can be identified:

- * Errors in disseminated highly market sensitive statistical data;
- * Errors in headline indicators or headline figures in news releases;
- * Errors in other parts of news releases following an agreement between the services concerned;
- * Errors in any other disseminated statistical data impacting the reputation of Eurostat or triggering substantiated major complaints from users.
- * Dissemination of confidential statistical data

Peer reviews

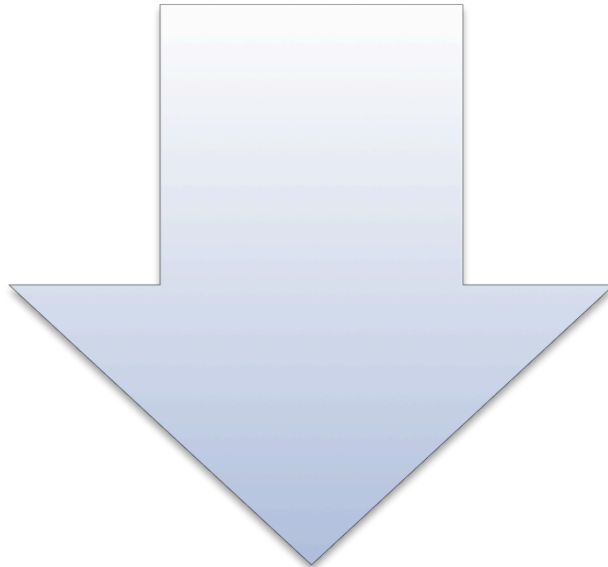
- * Peer reviews are part of the European Statistical System (EES) strategy to implement the Code of Practice (CoP).
- * Their objective is to enhance the integrity, independence and accountability of the statistical authorities which make up the ESS.

Peer reviews



The new round, which was launched in December 2013, seeks to:

- enhance the credibility of the ESS
- strengthen its capacity to produce high quality European statistics



- further reassure stakeholders about the quality of European statistics and the trustworthiness of the ESS

- assess progress made in compliance with the principles of the CoP
- assess progress made in the development of the ESS.

A common objective:

Data validation

Definition

The term validation refers to any process used to establish whether data conforms to specific criteria. Validation is a key task performed by all statistical domains.

In spite of its prominent role in the statistical business process, validation is not currently covered by common standards and procedures at ESS level.

To address this issue, Eurostat and the ESS launched the ESS.VIP project on **Common Data Validation Policy**

National versus EU level

The lack of harmonisation at ESS level has several negative consequences:

- * The validation process is not systematically considered when designing a new statistical domain or data collection.
- * Validation is very vulnerable with respect to staff mobility: Staff moving to new posts require specific training to adapt to local validation practices and procedures.
- * It is difficult to develop a set of common validation tools (building blocks) to support the different validation approaches.

National versus EU level

The lack of harmonisation at ESS level has several negative consequences:

- * It prevents common agreements with Member States. This results in a diversity of agreements on quality compliance depending on the statistical domain.
- * It may result in very inefficient validation processes, endless communications with data providers and revised data transmitted several times.
- * It may affect data quality as some validation tasks are performed in a redundant way while other tasks are not performed because there is no common agreement on the distribution of validation responsibilities between Member States and Eurostat.

Validation project objectives

- * Coherent validation policy in the different statistical domains, in cooperation with MS.
- * Standard validation language
- * Solutions for complex validation actions to ensure:
 - * Coherence between data files
 - * Coherence between Member States
 - * Integrity of the ESS data

A common objective:

Comparability

Comparability challenge

An attribute of statistics measuring the extent to which differences between statistics can be attributed to differences between the values of the statistical characteristics.....

or policy gaps... ???

Comparability challenge

Comparability aims at measuring the impact of differences in applied statistical concepts, definitions, measurement tools and procedures on the comparison of statistics between geographical areas, non-geographical dimensions, sectoral domains or over time.

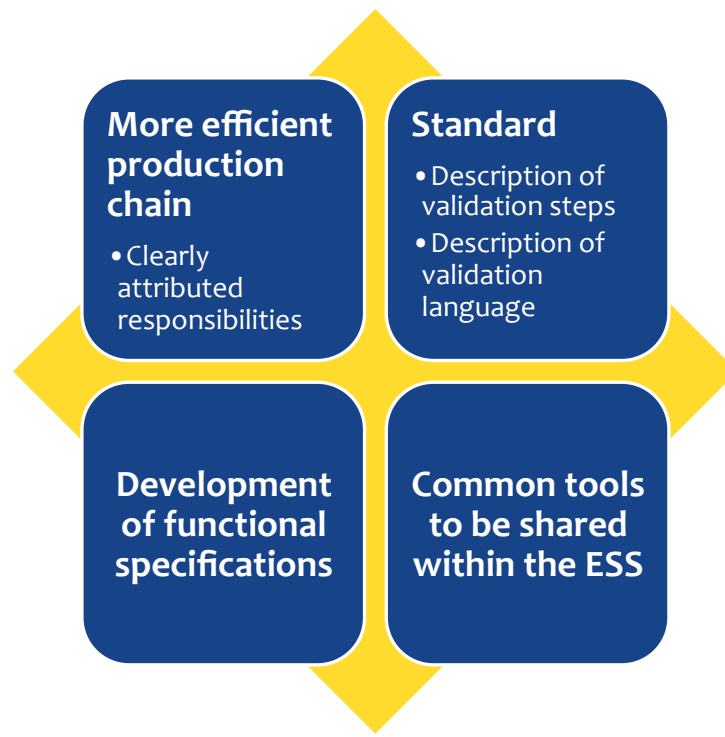
Comparability challenge

- * The concept can be further broken down into:
 - * (a) Comparability - geographical, referring to the degree of comparability between statistics measuring the same phenomenon for different geographical areas
 - * (b) Comparability over time, referring to the degree of comparability between two or more instances of data on the same phenomenon measured at different points in time.
 - * (c) Comparability between domains, referring to the comparability between different survey results which target similar characteristics in different statistical domains.

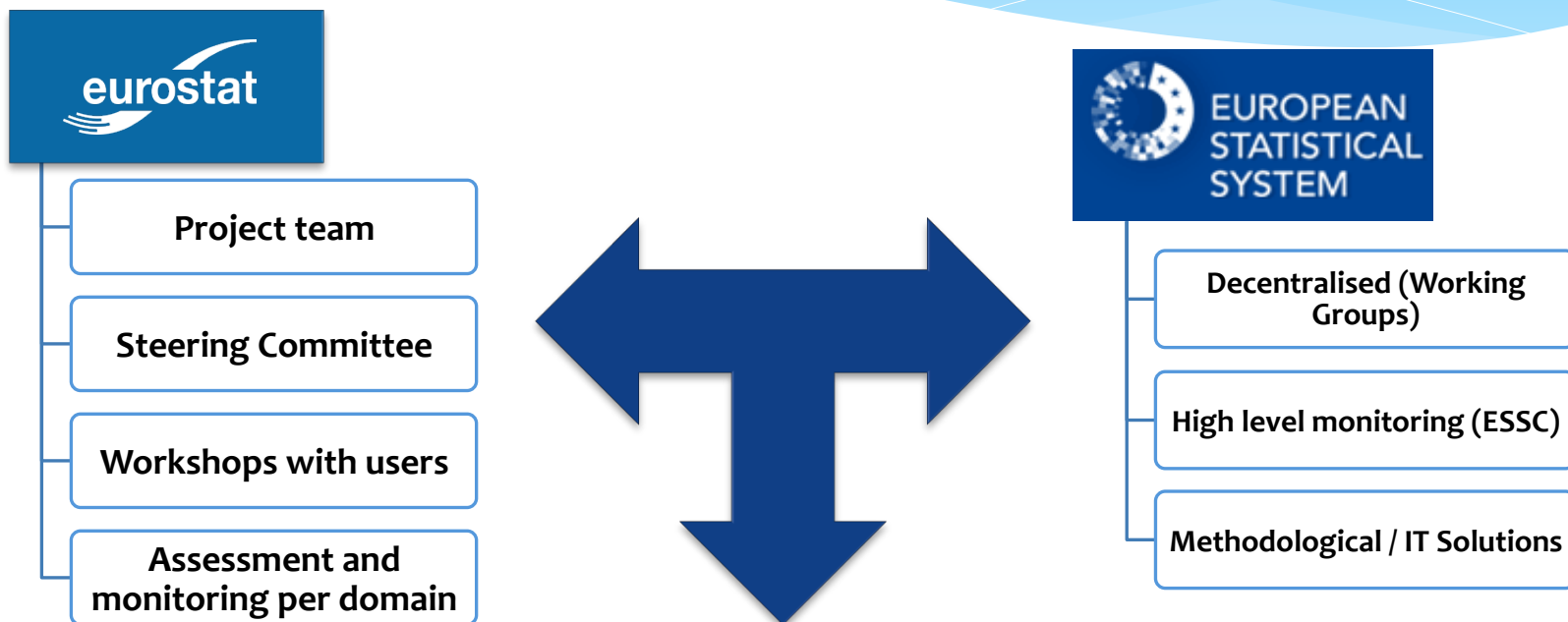
Comparability challenge

According to the European Statistics Code of Practice, European statistics should be 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.....

Comparability challenge



Proposed approach & milestones



Links with other ESS VIP

Validation

Warehouses

SIMSTAT

Shared services

...