European Institute for Gender Equality

Gender Statistics Database

Quality Considerations for EIGE's Gender Statistics Database

Acknowledgements

The present report is part of the work of the European Institute for Gender Equality (EIGE) on its Gender Statistics Database.

The authors of this report are Dr Maris Goldmanis (senior lecturer, Department of Economics, Royal Holloway, University of London) and Dr Irene Rioboo (professor at Rey Juan Carlos University).

Jakub Caisl provided interim revisions to the document as a consultant with consultancy company ICF.

It also draws on the various consultation processes organised under the framework of EIGE's experts' meetings on the Gender Statistics Database. EIGE would like to thank the participants in the consultation meetings that took place on 7 November 2017 in Vilnius for their insights and expertise, which greatly contributed to validating the research(*).

EIGE's Gender Statistics Database officer, Ligia Nobrega, coordinated the work and the various contributions to this publication.

(*) Anissa Amjahad, Aida Eskic Pihljak, Tamas Antal Heizer, Cristina Martelli, Erica Aloe, Clotilde Talleu, Marjut Pietilainen, Susana Neves, Maria Jose Carrilho, Dragana Papic, Martha Kalana, Albert Motivans, Giusepina Muratore, Irena Varagic, Manuel Albano, Isabel Romao, Rita Penedo, Bujar Hajrizi, Ivana Naskova, Dilek Guder, Aida Pihljak, Martha Kalana. Further information on the event can be found on EuroGender.

European Institute for Gender Equality

The European Institute for Gender Equality (EIGE) is an autonomous body of the European Union established to strengthen gender equality across the EU. Equality between women and men is a fundamental value of the EU and EIGE's task is to make this a reality in Europe and beyond. This includes becoming a European knowledge centre on gender equality issues, supporting gender mainstreaming in all EU and Member State policies and fighting discrimination based on sex.

European Institute for Gender Equality (EIGE)
Gedimino pr. 16
LT-01103 Vilnius
LITHUANIA

Tel. +370 52157444

Email: eige.sec@eige.europa.eu

http://www.eige.europa.eu

http://twitter.com/eurogender

http://www.facebook.com/eige.europa.eu

http://www.youtube.com/eurogender

http://eurogender.eige.europa.eu

https://www.linkedin.com/company/eige/

Luxembourg: Publications Office of the European Union, 2021

Print	ISBN 978-92-9482-620-6	doi:10.2839/787	MH-02-20-966-EN-C
PDF	ISBN 978-92-9482-619-0	doi:10.2839/879831	MH-02-20-966-EN-N

Quality Considerations for EIGE's Gender Statistics Database

Contents

List of tables

Abbreviations

Country Codes

BE	Belgium
BG	Bulgaria
CZ	Czechia
DK	Denmark
DE	Germany
EE	Estonia
IE	Ireland
EL	Greece
ES	Spain
FR	France
HR	Croatia
IT	Italy
CY	Cyprus
LV	Latvia
LT	Lithuania
LU	Luxembourg
HU	Hungary
MT	Malta
NL	Netherlands
AT	Austria
PL	Poland
PT	Portugal
RO	Romania
SI	Slovenia
SK	Slovakia
FI	Finland
SE	Sweden
UK	United Kingdom

Frequently used abbreviation

BPfA	Beijing Platform for Action for Equality, Development and Peace
BSON	Binary JSON
EIGE	European Institute for Gender Equality
ESMS	Euro-SDMX Metadata Structure
ESQRS	ESS Standard for Quality Reports Structure
ESS	European Statistical System
JSON	JavaScript Object Notation
NGO	non-governmental organisation
NoSQL	Non-Structured Query Language
SDMX	Statistical Data and Metadata eXchange
SIMS	Single Integrated Metadata Structure
UNECE	United Nations Economic Commission for Europe
WMID	women and men in decision-making

Introduction

EIGE's Gender Statistics Database ([1]) is a comprehensive knowledge centre for gender statistics. Gender statistics measure to what extent equality between women and men is achieved across all areas of life in Europe. Such statistics give EU Member States and candidate countries the means to gauge their success in achieving gender equality against the dual benchmarks of other countries' achievements and their own histories. This brings Europe closer to EIGE's vision of 'making equality between women and men a reality for all Europeans and beyond'.

For gender statistics to serve this goal, they must measure an aspect of gender equality consistently, accurately and in a timely manner. It is also necessary to ensure that the statistics are presented clearly and with sufficient accompanying information (metadata) to allow users to understand what has been measured and how the measurement has been carried out. These properties constitute the basis for the multidimensional concept known as quality of statistics. In other words, quality is the extent to which statistics are fit for the purpose of measuring the concept of interest and conveying the result to the users.

The sections below explore all aspects of quality as it relates to EIGE's Gender Statistics Database. They should be useful both to our users and to our contributors (data producers and providers).

For users of the database, we aim to provide:

1. evidence that quality considerations have been (and continue to be) at the core of the database; and
2. the tools needed to understand the applicability and the limitations of the data and statistics available in the database.

We hope that after reading this document, users will not only have more confidence in the statistics contained in the database, but will also be better equipped to understand, through reading the metadata of their data sets of interest and reflecting on the production processes of these data sets, the possible biases and limitations in the interpretation and application of the available data and statistics to comparisons across countries and over time. As a result, users will be able to apply gender statistics to their work in a more confident, accurate, and effective manner.

For data producers and providers, we aim to provide recommendations and guidelines on how to:

1. produce data and statistics in a gender-sensitive way, avoiding common pitfalls that lead to gender biases;
2. document the quality of data and statistics, particularly focusing on the gender-relevant aspects; and
3. prepare data and metadata for use in the Gender Statistics Database.

This document aims to encourage more data producers and providers to collect and disseminate data and statistics on gender-relevant topics, and all data producers and providers to take care to avoid gender biases in the production and presentation of all data and statistics (even if the collection is not specifically focused on gender).

Incidentally, this document can be viewed as an integral part of EIGE's quality assurance framework. It contributes to the quality of the final product (the statistics available in EIGE's database) in multiple ways. First, it helps data producers to produce more relevant, reliable, accurate and comparable data and metadata. Second, it assists data providers and EIGE in selecting appropriate data for the database and presenting it in a clear and fitting manner. Last, but by no means least, it helps users to access, understand and correctly interpret these data and metadata. The important quality dimension of accessibility and clarity is really a property of the

([1]) https://eige.europa.eu/gender-statistics/dgs

interface between the statistical information and the users, and not merely an attribute of the data and metadata alone. This interface works correctly only when users are aware of it and proficient at using it. Informing users about data quality and its documentation in our metadata therefore itself contributes to improving data quality.

For most sections, we provide two versions: a detailed presentation and a very brief executive summary. Depending on your focus and interests, you can choose to read all or just part of the following articles, although we do encourage you to read at least the executive summaries of all sections.

1. Getting our terms right

Several key terms will keep recurring throughout this guide on quality of statistics. The terms specific to quality will be defined in Section 2 'The general framework'. Here we define the more elementary terms: those that refer to the contents database.

1.1. Data and statistics

In the context of social science, data numbers describe some aspect of social reality. More broadly, the *Oxford English Dictionary* defines data as 'related items of (chiefly numerical) information' that are 'used for reference, analysis, or calculation.' (Incidentally, while technically the Latin plural of *datum*, the word **data** in modern English is used both as a plural and a singular (mass) noun. In this document, it will still be taken as plural, as a nod to the word's origins and in line with the requirements of more conservative style guides.)

There are various levels of data. At the beginning, there is **raw data**, i.e. the original items of information that are recorded as part of a survey (such as the labour force surveys, which ask respondents about their characteristics and employment status) or an administrative data collection (such as the recording of births and deaths in a government register). Typically, these raw data are **microdata**: they describe individual actors (persons, firms or organisations). In order to draw conclusions from the data, they must be analysed to obtain summaries at an aggregated level (such as the number of births in a country in a year, or the ratio of this number to the total population, known as the birth rate). **Statistics** are the results of such analysis. Statistics are also data in their own right and can be used in further analysis, for example by relating various statistics to each other (such as exploring the relationship between national income and gender equality). By construction, statistics are **macrodata**: instead of providing information about individuals, they describe larger, aggregated units, such as entire populations or subpopulations (for example, all women of a country/region or all persons in a certain age group and/or occupation).

In summary then, data can refer both to the individual items recorded by a data collection (known as raw data) and to various aggregates/summaries of these items, while statistics specifically refer to these aggregates/summaries. The Gender Statistics Database contains only statistics and not raw data.

Official statistics are statistics collected by public bodies. In EIGE's Gender Statistics Database these are statistics collected by the national statistical offices/institutes of EU Member States and by Eurostat, the statistical office of the European Union.

Finally, the word **statistics** also refers to the scientific discipline of data analysis. Based on probability theory, statistics provide the body of mathematical methods used to reveal patterns in, and test hypotheses about, data. In this document, we will not use the term **statistics** in this sense.

1.2. Data sets and indicators

In statistical databases, such as the Gender Statistics Database, data are organised in collections known as **data sets**. Precise definitions of data sets in statistics vary depending on the intended application. A data set in the Gender Statistics Database is best understood in the Statistical Data and Metadata eXchange (SDMX) sense as 'a collection of similar data, sharing a structure, which extends over a period of time' ([2]). In practical terms, a data set contains measurements of a single concept (or set of similar concepts) for a number of countries/regions and time periods, possibly broken down by some addi-

([2]) https://sdmx.org/wp-content/uploads/SDMX_Glossary_Version_2_0_October_2018.htm#_Toc529282129

tional variables. For example, one data set in the Gender Statistics Database is 'Employment rates by sex, age and degree of urbanisation', which gives the percentage of people who are employed each year within each subpopulation defined by the intersection of a country, sex, age group and type of settlement (for instance, one such subpopulation consists of all women aged 15–64 living in urban areas in Germany). Each number in the data set is a **data point**; the more years and subpopulations a data set has, the more data points. Each data set can have thousands of data points (for example, 30 countries, three sex groups (men, women, total), five age groups, four levels of urbanisation and 20 annual observations would result in 30 × 3 × 5 × 4 × 20 = 36 000 data points).

A very special type of data set is one that contains **indicators**, which are the type of statistics that can be used directly in policy evaluation. Most users are often looking precisely for this type of statistics. Not all statistics are indicators. According to the European Statistical System (ESS) definition ([3]), a **statistical indicator** is a '[d]ata element that represents statistical data for a set of characteristics, one of which allows for **meaningful comparisons** of the data'. The contextual information for the ESS definition explains: 'An aggregation such as the number of accidents, total income or female members of parliament, are not in themselves indicators for comparison across countries, as they are not comparable between populations. However, if a transformation is applied to make the data comparable, e.g. number of accidents per thousand of population, average income, or female members of parliament as a percentage of the total, the result meets the criteria for an indicator.' In fact, as other sources point out, this restriction alone may not be sufficient to meet the criterion of meaningful comparisons: an indicator should not only be expressed in units that are comparable across space and time, but for these comparisons to be meaningful, 'a **reference point** should also be defined against which value judgements can be made. Indicators have a normative nature, in the sense that a change from the reference point (a norm or a benchmark) in a particular direction can be interpreted as "good" or "bad".'

1.3. Metadata

Numerical data alone (be they raw data or statistics) are of little use without a proper description of how the numbers have been obtained, what they mean and how they can be interpreted. This crucial description is contained in **metadata**, or 'data about data'. In the Gender Statistics Database, each data set is accompanied by a structured set of metadata providing all the information necessary to understand, interpret and evaluate the data. A significant proportion of metadata deals specifically with the quality of the associated statistics. Enabling users to understand that information is an important objective of this guide.

([3]) https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=21228653&RdoSearch=CONTAIN&TxtSearch=statistical%20indicator&CboTheme=&IsTer=&IntCurrentPage=1&ter_valid=0

2. The general framework

2.1. Detailed presentation

Quality is a complex, multidimensional concept that measures the fitness of data for their purpose ([4]). The identification of data quality with fitness of purpose is rooted in the belief that quality should not be assessed on entirely absolute grounds, but should rather be evaluated with respect to the target audience (that is, the users of the data) and the use to which the data are expected to be put. This is particularly relevant in the context of EIGE's Gender Statistics Database, which focuses exclusively on gender statistics that are intended for the very specific and clearly defined purpose of measuring and advancing gender equality.

The general framework EIGE uses to evaluate data quality is borrowed from the ESS quality assessment and assurance frameworks ([5]). The ESS is the partnership between the EU statistical authority, which is the Commission (Eurostat), and the national statistical institutes and other national authorities responsible in each Member State for the development, production and dissemination of European statistics. The ESS functions as a network in which Eurostat's role is to lead the way in harmonising statistics in close cooperation with national authorities with the objective of providing comparable statistics at EU level.

With the adoption of the *European Statistics Code of Practice* and its revisions in 2011 and 2017, Eurostat and the statistical authorities of the EU Member States have committed themselves to an encompassing approach to high quality statistics. Based on that code of practice, the ESS quality assurance framework evaluates quality according to three dimensions: institutional environment (principles 1–6 of the Code of Practice), statistical processes (principles 7–10) and statistical outputs (principles 11–15) ([6]).

Since EIGE is not involved in and has no control over the production of most of the required original data, it must focus on evaluating the quality of existing statistical outputs. Nonetheless, all three of the dimensions above are important, as the institutional and process aspects of data quality are considered when selecting new data providers for the database. Consequently, except in unusual cases (such as administrative data on gender-based violence), only well-established institutions at the international and EU level, (such as Eurostat and Eurofound) are selected as statistical providers. There may be data produced by other data providers that partially satisfy required criteria on statistical outputs, but if the institutional and process aspects of the data quality are unclear, they are not eligible for EIGE's database. More information on selecting data providers is available in Section 5 'Evaluation and selection of sources for EIGE's Gender Statistics Database'. The rest of this section focuses on the quality of statistical outputs.

The five dimensions of quality of statistical outputs in the ESS framework are as follows:

- the relevance dimension, which measures whether the outputs meet the current and potential needs of users;
- the accuracy and reliability dimension, which shows whether estimates and computations are consistently close to their exact or true values;
- the timeliness and punctuality dimension, which assesses whether outputs are released in accordance with an agreed schedule and soon after the period to which they refer;
- the coherence and comparability dimension, which shows whether concepts, definitions,

([4]) Eurostat, 2015; Bank of England Statistics and Regulatory Data Division, 2014.
([5]) Eurostat, 2015; Eurostat, 2019.
([6]) Eurostat, 2018.

methodologies and actual data are consistent internally and across space and time;

- the accessibility and clarity dimension, which indicates if data are available and accompanied by adequate explanatory information (metadata).

All five dimensions are clearly related to the overarching principle of fitness for purpose. On the most general and abstract level, statistical data have two purposes: first, to accurately measure the concepts of interest, and second, to clearly convey the results to the users. The first four dimensions (relevance, accuracy and reliability, timeliness and punctuality, and coherence and comparability) are crucial for the purpose of accurate measurement, while the fifth dimension (accessibility and clarity) and to some degree also the third dimension (timeliness and punctuality) are needed to achieve the purpose of clearly conveying the results to the users.

Table 1 lists these dimensions, along with abbreviated ESS definitions, itemised sub-dimensions and an example taken from the Gender Statistics Database referring to EIGE's primary data on the number of women and men (president and members) in the European Parliament. Figure 1 shows the entire framework (dimensions and sub-dimensions) as a hierarchical tree. Full, formal definitions of all the terms as defined in Eurostat's Concepts and Definitions Database (arranged following the structure in Figure 1) can be found in Section 7 'Definitions of quality dimensions in Eurostat's Concepts and Definitions Database'.

Table 1. Statistical output quality dimensions in the ESS framework

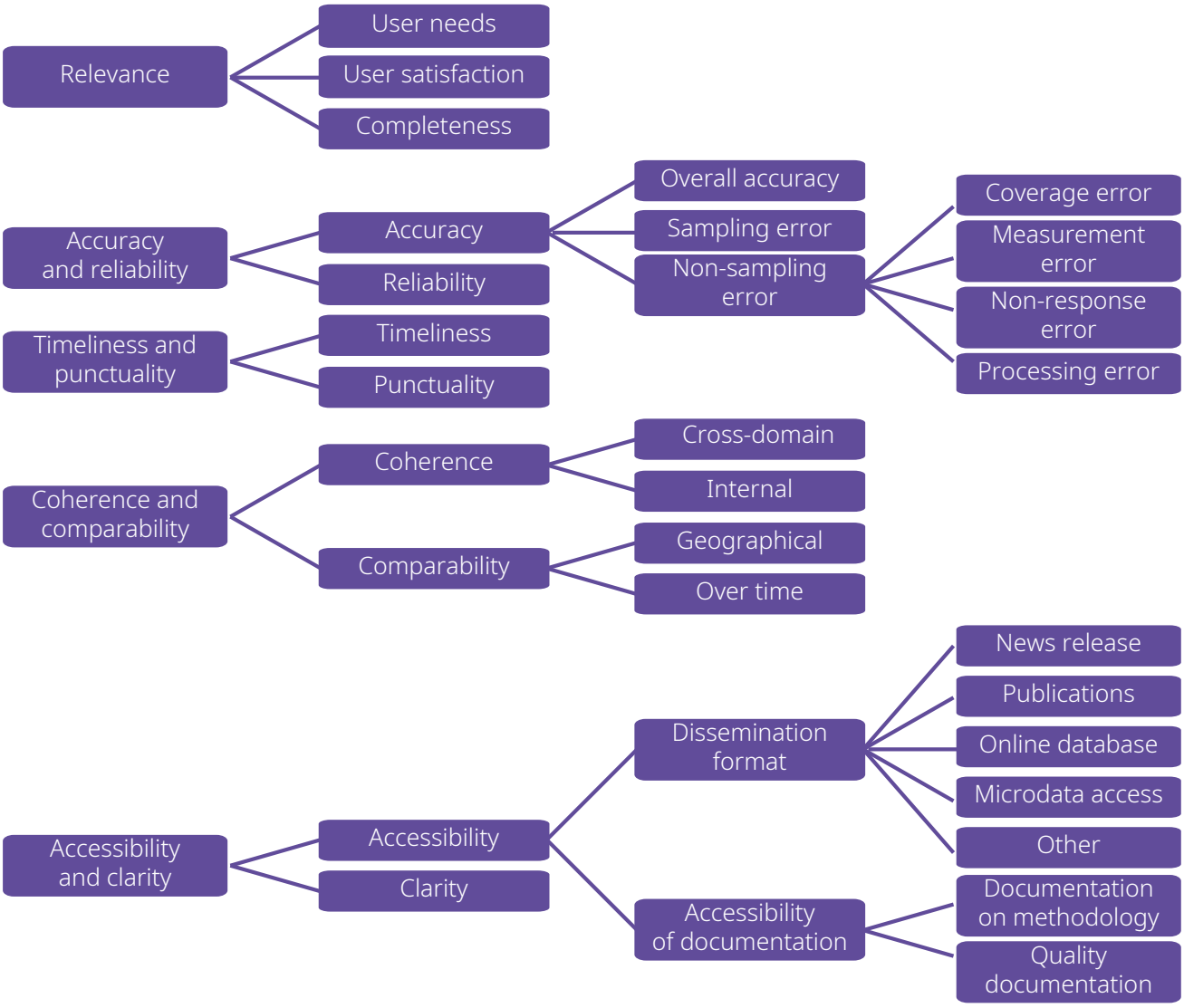
Dimension	Abbreviated ESS definition	Sub-dimensions	Example: number of women and men (president and members) in the European Parliament
Relevance	Relevance is the extent to which outputs meet the current and potential needs of users.	• User needs (12.1) • User satisfaction (12.2) • Completeness (12.3)	• 12.1. User needs Women and men in decision-making (WMID) data are the primary source of information for indicators to monitor the implementation of Area G (power and decision-making) of the Beijing Platform for Action for Equality, Development and Peace (BPfA). The data are therefore widely used by the European Commission (DG Justice and Consumers) and EIGE for analysis in this area and for reporting to the Council of the European Union. The data are also widely used by researchers in this area. • 12.2. User satisfaction No user satisfaction surveys are carried out. • 12.3. Completeness In the case of the European Parliament, data are complete.

Dimension	Abbreviated ESS definition	Sub-dimensions	Example: number of women and men (president and members) in the European Parliament
Accuracy and reliability	Accuracy is the closeness of estimates and computations to the exact or true values. Reliability is the closeness of initial estimated values to subsequent estimated values.	• Accuracy ▪ Overall accuracy (13.1) ▪ Sampling error (13.2) ▪ Non-sampling error (13.3) ▪ Coverage error ▪ Measurement error ▪ Non-response error ▪ Processing error • Reliability	• 13.1. Overall accuracy In principle, WMID data accurately describe the situation for the area of decision-making concerned, though in some areas the coverage of organisations is restricted to limit the cost and burden of the data collection, and this could potentially have an impact on overall accuracy. In the case of the European Parliament, data can be considered fully accurate. • 13.2. Sampling error Not applicable. • 13.3. Non-sampling error Not applicable.
Timeliness and punctuality	Timeliness reflects the length of time between data availability and the event or phenomenon the data describe. Punctuality refers to the time lag between actual and scheduled data release dates.	• Timeliness (14.1) • Punctuality (14.2)	• 14.1. Timeliness Data are released within 1 month of collection. • 14.2. Punctuality • Punctuality is 100 %.
Coherence and comparability	Coherence refers both to the extent to which a data set is internally consistent and to the degree to which different data sets can be reconciled and combined. Comparability measures the impact of differences in applied statistical concepts, definitions and methodologies when statistics are compared between geographical areas or non-geographical dimensions or over time.	• Coherence ▪ Cross-domain (15.3, part 1) ▪ Internal (15.3, part 2) • Comparability ▪ Geographical (15.1) ▪ Over time (15.2)	• 15.3. Coherence – cross domain In general, there are few other sources of data on decision-making against which to assess the coherence of WMID data. In the case of the European Parliament, the official website of the institution publishes the distribution of women and men in the opening sessions of each parliamentary term (http://www.europarl.europa.eu/elections2014-results/en/gender-balance.html). However, data are not presented in absolute numbers (only the share by gender is shown) and do not reflect the variations between elections or the current composition. • 15.3. Coherence – internal Internal coherence of the data (e.g. through time or across countries) is ensured through careful application of the WMID methodology and routine validation of data. Additionally, the data collected goes through a verification process carried out by a senior researcher. • 15.1. Comparability – geographical Note that seats are allocated based on the population of each Member State and therefore the number of members by Member State will vary. Apart from this point, data are comparable between Member States. • 15.2. Comparability – over time • In general, WMID data are comparable through time in each area of decision-making. In the case of the European Parliament, there are changes in the number of members of the European Parliament due to changes in the number of EU Member States.

Dimension	Abbreviated ESS definition	Sub-dimensions	Example: number of women and men (president and members) in the European Parliament
Accessibility and clarity	Accessibility is the ease and conditions under which statistical information can be obtained. Clarity assesses whether data are accompanied by the metadata needed to fully understand the data	• Accessibility ▪ Dissemination format ○ News release (10.1) ○ Publications (10.2) ○ Online database (10.3) ○ Microdata access (10.4) ○ Other (10.5) ▪ Accessibility of documentation ○ Documentation on methodology (10.6) ○ Quality documentation (10.7) • Clarity	• 10.1. News release No regular news release. • 10.2. Publications Since 2017, EIGE has been publishing regular bulletins on gender statistics, which cover data on decision-making. The European Commission's annual report on equality usually includes a section on this topic. • 10.3. Online database EIGE's Gender Statistics Database • 10.4. Microdata access Microdata are not made available. • 10.5. Other WMID data are the primary source of information for indicators to monitor the implementation of Area G of the BPfA. • 10.6. Documentation on methodology WMID methodology published on EIGE's website • 10.7. Quality documentation Not applicable.

NB:
1. The definitions are adapted from Eurostat (2015) and Eurostat (ongoing).
2. The numbers in parentheses refer to sections in the Euro-SDMX metadata structure, version 2.0: http://ec.europa.eu/eurostat/data/metadata/metadata-structure

Figure 1. Statistical output quality dimensions and sub-dimensions in the ESS framework



Source: EIGE, based on the Quality Assurance Framework of the European Statistical System ([7]).

The dimensions of relevance, accuracy and reliability, and coherence and compatibility are closely related to the general measurement concepts of reliability and validity, which are widely used in the social sciences. **Reliability**, in this general sense, refers to the degree to which the measurement methods produce results that are stable and consistent. It assesses whether the same methods applied multiple times and/or by different people would yield estimates that are close to each other, provided that the underlying true values have not changed. **Validity**, on the other hand, refers to how well the test measures the underlying true concept or value. For example, a scale that on average shows the correct weight of an object but has a precision of plus or minus 10 kg provides a valid but unreliable measure of weight, whereas one that is on average 10 kg above the true weight, but is always within 1 g of this incorrect value would provide a reliable, but invalid measure. A yardstick (which measures height) would provide another invalid measure of weight. As demonstrated by this example, validity can be compromised either by measuring the wrong concept altogether (e.g. yardstick) or by measuring the right concept, but in a way that has an inherent bias (systematic error) (e.g. biased scale). Failures of reliability are generally a matter of random error.

For an example that is more closely related to social statistics, consider two statistical mod-

([7]) https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646

els, A and B, that have been developed to estimate the unknown age of an individual based on some observable characteristics (we assume here that for some reason age is not directly observable). For a certain person both models have been computed four times, with slightly varying inputs. The four estimates obtained from the four runs of model A (in years) are 30, 30, 30 and 30, while those from model B are 10, 18, 22 and 30. If the real age of this person is 20, we can say that model A is reliable (because it has consistently provided the same result) but invalid (because the mean age, obtained as the average of four results, is different from the actual age (systematic error)). On the contrary, model B is unreliable (because it has provided widely dispersed estimates) but valid because the average of those estimates is equal to the actual age of the person (random error).

Given these definitions, we see that the ESS quality dimensions of accuracy and reliability and coherence and comparability have complex relationships to the general concepts of validity and reliability.

- Accuracy.
    - Overall accuracy has both validity components (relating to systematic error) and reliability components (related to random error).
    - Random sampling errors are an aspect of (un)reliability (in the general sense, not the ESS sense). Because random sampling errors decrease with sample size, sample size evaluation is an important element of assessing reliability (in the general sense) and sampling error accuracy (in the ESS sense).
    - Non-sampling errors.
        - Coverage errors (inappropriate sampling frames) compromise validity, as they cause systematic biases.
        - Measurement errors in data collection introduce both validity problems (if they lead to measurements that are biased, e.g., due to leading questions) and reliability problems (for instance, if the questions are asked differently to different people (in an unpredictable way)).
        - Non-response errors also have both validity implications (if certain groups of people are less likely to respond than others) and reliability implications (because non-response is almost always random at least to some degree).
        - Processing errors (in data entry, editing, coding and imputation) can lead to validity or reliability problems (or both), depending on whether these errors create estimates that are systematically biased or randomly varying.
- Reliability in the ESS sense is a special case of reliability in the general sense. Note that reliability in the general sense is also influenced by accuracy, as defined above (via random error).
- Coherence and comparability have both validity and reliability components.
    - Reliability. Random errors in data collection and processing can lead to data that are either internally incoherent or not reconcilable across data sets and statistical domains.
    - Validity. In general data between different data sets, geographical areas or points in time fail to be reconcilable (coherent) or comparable due to systematic differences in definitions, classifications and methodologies, which lead them to measure different concepts.

While the reliability components of all the above quality dimensions (along with problems with timeliness and punctuality) are generally absolute, it should be made clear that the validity components are immensely dependent on the context and the purported use of the data, which links them to the relevance dimension of quality. An estimate or indicator that may be appropri-

ate for some general-purpose applications may be completely invalid for gender statistics. For example, while estimates of economic output that neglect off-market household production (especially care services) may be valid (although not ideal) measures for assessing a country's economic growth, they are completely inadequate for measuring the relative contributions of women and men: because women are highly over-represented in the (neglected) household sector, such estimates would grossly underestimate the contribution of women.

To identify such failures of validity, it is important to carry out a careful **qualitative analysis** of the definitions, classifications and procedures used in the production of the data, so that all sources of bias relevant for the intended use cases are taken into account. In the case of gender statistics, particular attention should be paid to **gender biases** (a thorough discussion of gender statistics and gender biases can be found in Section 3.1.1 'Gender statistics').

To enable the proper identification of limitations to reliability and validity, the accessibility and clarity dimension of data quality is paramount: a complete and thorough set of metadata must be available to make such an analysis possible. The backbone of any data set's metadata should be formed by a structured, dedicated metadata document (see below), but because all potential sources of bias may not be immediately obvious to the data provider, it is also important that original documentation is made available to end users and secondary distributors of data, such as EIGE. Such original documentation should include questionnaires, interview protocols and descriptions of data-processing algorithms.

According to current good practices, data providers use metadata structures to document data quality assessment. In particular, the current standard in the ESS is the Euro-SDMX metadata structure (ESMS), supplemented when necessary by the ESS Standard for Quality Reports Structure (ESQRS). A new standard combining the two structures was developed in 2012–2013. However, this standard, the Single Integrated Metadata Structure (SIMS), has not yet been widely adopted. In particular, the metadata for all data sets in Eurostat's online database are still organised according to the ESMS. EIGE has therefore adopted the ESMS as the standard for metadata (including quality reporting) in its database. The full definition of the ESMS structure can be found in Section 8 'EURO-SDMX Metadata Structure 2.0'.

2.2. Executive summary

Quality is a complex, multidimensional concept that measures the fitness of data for their purpose. It should not be viewed as absolute, but rather as dependent on the context and the purported use of the data. An estimate or indicator that may be appropriate for some general-purpose applications may be completely invalid for gender statistics.

The general framework EIGE uses to evaluate data quality is borrowed from the ESS quality assessment and assurance frameworks ([8]). The ESS quality assurance framework evaluates quality according to three dimensions: institutional environment (principles 1–6 of the Code of Practice), statistical processes (principles 7–10) and statistical outputs (principles 11–15) ([9]). EIGE uses the first two blocks to evaluate new data providers for the database. The actual data are evaluated using the five dimensions of quality of statistical outputs, which are as follows:

- the relevance dimension, which measures whether the outputs meet the current and potential needs of users;
- the accuracy and reliability dimension, which shows whether estimates and computations are consistently close to their exact or true values;
- the timeliness and punctuality dimension, which assesses whether outputs are released

([8]) Eurostat, 2015; Eurostat, 2019.
([9]) Eurostat, 2018.

in accordance with an agreed schedule and soon after the period to which they refer;

- the coherence and comparability dimension, which shows whether concepts, definitions, methodologies and actual data are consistent internally and across space and time;

- the accessibility and clarity dimension, which indicates if data are available and accompanied by adequate explanatory information (metadata).

All five dimensions are clearly related to the overarching principle of fitness for purpose. The first four dimensions (relevance, accuracy and reliability, timeliness and punctuality, and coherence and comparability) are crucial for the purpose of accurate measurement, while the fifth dimension (accessibility and clarity) and to some degree also the third dimension (timeliness and punctuality) are needed to achieve the purpose of clearly conveying the results to the users.

In EIGE's Gender Statistics Database, quality is systematically evaluated using structured metadata in the ESMS format, which is fully described in Section 8 'EURO-SDMX Metadata Structure 2.0'.

3. Gender statistics and gender-sensitive indicators

EIGE's Gender Statistics Database is focused on gender statistics. But what exactly are gender statistics? What distinguishes them from other types of social statistics? What specific quality considerations need to be taken into account? (Remember that quality of statistics is a fitness-for-purpose concept, and as such can never be considered in isolation from the specific purpose of the statistics in question.) We seek answers to these questions below. If you have not done so, we also suggest reading up on the general terms (such as data, metadata, statistics and indicators) in Section 1 'Getting our terms right' and Section 2 'The general framework'.

3.1. Definitions and scope

3.1.1. Gender statistics

According to the UN's *Gender Statistics Manual* ([10]), **gender statistics are 'statistics that adequately reflect differences and inequalities in the situation of women and men in all areas of life'**. According to the same source, in addition to being 'collected and presented **disaggregated by sex** as a primary and overall classification' and '**reflecting gender issues**', gender statistics data must also be 'based on **concepts and definitions** that adequately reflect the diversity of women and men and capture all aspects of their lives' and be collected using **methods** that 'take into account stereotypes and social and cultural factors that may induce gender biases in the data'.

This definition of gender statistics is in line with the BPfA adopted at the UN's Fourth World Conference on Women (Beijing, China, 1995), an agenda for women's empowerment, which asked nations to 'ensure that statistics related to individuals are collected, compiled, analysed and presented by sex and age and reflect problems, issues and questions related to women and men in society ([11]).

According to Hedman et al. (1996) ([12]), gender statistics not only cover the concept of statistics on women and men but also require that data are produced and presented to reflect women's and men's conditions, roles and contributions in society, and their needs and specific problems.

This concept is much wider than sex-disaggregated data, which are data collected and tabulated separately for women and men without guarantees of reflecting gender roles and social inequalities ([13]). The United Nations Economic Commission for Europe (UNECE) gender statistics guide ([14]) illustrates the need to go beyond disaggregation by sex in the context of gender-based violence, where data on victims of homicide disaggregated by sex have some value, but additional information is needed to understand whether a killing is gender related. For instance, information on the perpetrator and the perpetrator's relationship to the victim is also needed in order to know whether the homicide was committed in a family context or by someone unknown to the victim.

Moreover, it is worth highlighting that gender statistics do not need to consist solely of sex-disaggregated data. In fact, some relevant socioeconomic statistics reflect relevant gender aspects despite not being sex-disaggregated. For instance, national budgets or national

([10]) United Nations Statistical Division, 2013.
([11]) United Nations, 1996.
([12]) Hedman, B. et al., 1996.
([13]) United Nations Statistical Division, 2013.
([14]) United Nations Economic Commission for Europe and World Bank Institute, 2010.

accounts statistics produce different impacts and consequences for the lives of women and men and could be very useful when considering a component on gender budgeting or on planning infrastructure development.

Gender statistics play a key role in the improvement of national statistical systems. The inclusion of the gender perspective throughout statistical work implies a deep review of definitions and methods of data collection, analysis and dissemination taking into account gender challenges and gender biases in all statistical activities (read more on this in Section 4 'Guidelines for gender-sensitive data production and quality assessment').

3.1.2. Gender indicators

It is important to remember that statistical indicators are statistics that allow meaningful comparisons across countries and time periods, and in relation to a set of reference points ([15]). Gender indicators have several different reference points.

- First, any indicator measuring the situation of women in a given country at a given time has as its reference point the situation of men in the same country and at the same time.
- Second, insofar as the indicator is used to assess intertemporal changes, the reference point is the value of the indicator at a fixed earlier point in time.
- Third, when the indicator is used to assess international differences (as, say, the Gender Equality Index is), values of different countries serve as reference points for each other; in the context of the EU, the average value of all Member States is another natural reference point.
- Fourth, when an indicator has an obvious ideal value (such as complete freedom from gender-based violence), this ideal value serves as a natural reference point.

Finally, another natural comparison can be between different groups of women and men. It is possible to compare the situation of women and men across countries, for example by analysing women's income in terms of purchasing power parity, but also the gender pay gap expressed as a percentage of men's income. These comparisons may yield very different results – it is quite possible that the gender pay gap could be quite low in a specific country, although women's income in that country might be very low compared to women in other countries ([16]).

When evaluating the quality of gender indicators, in addition to the general quality considerations outlined above, it is necessary to assess whether the indicator adequately measures the underlying gender dimension and whether it is justifiable to regard a change in a particular direction as unambiguously good or bad. In indicators measuring gender differences (such as the Gender Equality Index), one must also decide when it is appropriate to look purely at the relative situation of women and men (percentage gaps) and when one should also take into account the absolute situation (levels). For example, is being equally poor preferable to being unequally rich? (The Gender Equality Index answers this question in the negative: the index scores are based not on pure gender gaps, but rather on gender gaps adjusted by the overall level of achievement ([17]).)

3.2. Uses and users

The UN manual ([18]) also lists a number of potential uses for gender statistics, which are:

([15]) See Section 1 'Getting our terms right'.
([16]) http://eige.europa.eu/gender-equality-index
([17]) European Institute for Gender Equality, 2017.
([18]) United Nations Statistical Division, 2013.

- to promote understanding of the actual situation of women and men in society;
- to advance gender analysis and research;
- to monitor progress towards gender equality and the full and equal enjoyment of all human rights and fundamental rights by women and girls;
- to develop and monitor policies and programmes oriented towards increased investments in human capital and the labour force;
- to support gender mainstreaming in development and poverty reduction policies; and
- to develop and monitor policies on the reduction of violence against women.

Combining the lists from the UN manual and the UNECE gender statistics guide ([19]), the target groups of users for gender statistics are best defined as follows:

- the European Parliament;
- the European Commission;
- EU agencies;
- government bodies promoting equal opportunities;
- other government bodies (ministries of labour, social protection, education, etc.);
- women's organisations, feminist organisations as separate non-governmental organisations or within political parties, trade unions, parliaments, regional and municipal decision-making bodies;
- networks, faculties and libraries in universities and other parts of research environments focusing on questions of equality, equal opportunities, feminism and other gender-related issues;
- public libraries;
- women's magazines, publications and information centres;
- support centres fighting against harassment or violence against women, centres for young mothers and other gender-oriented social institutions;
- the media;
- international organisations.

3.3. Quality implications

The above definitions and use cases have important implications for several dimensions of quality assessment ([20]). In terms of **relevance**, only data that reflect topics relevant to gender (in) equality should be considered. In terms of **accuracy and reliability** and **coherence and comparability**, it is necessary to avoid gender bias arising from inappropriate definitions, flaws in sample coverage and sampling procedures, questionnaire design, interviewer training and interview protocols, coding and classification schemes and data processing. In terms of **accessibility and clarity**, data must be presented in a way that is accessible and understandable, including for target groups without specialised knowledge of statistics, and metadata must explain the gender-specific quality considerations. Detailed original documentation of data collection and processing procedures should also be available to enable further gender-sensitive analysis by expert users. Some relevant questions to be asked when evaluating gender statistics are as follows.

- [Relevance and accessibility and clarity] Are sex-disaggregated statistics available?

([19]) United Nations Economic Commission for Europe and World Bank Institute, 2010.
([20]) The following discussion assumes familiarity with the basics of the ESS framework for assessing the quality of statistics. Read more in Section 2 'The general framework'.

- [Relevance and accessibility and clarity] Are additional breakdown variables available (for example, breakdown by sex and age)?
- [Relevance] Are there questions on gender-relevant topics?
- [Accuracy and reliability] Do the working definitions used by the data producer fully cover the concept that the data attempt to measure? (For example, does the definition of the labour force adequately cover unpaid work? Is entrepreneurship properly distinguished from self-employment? Are all types of gender-related violence included in the operational definition of gender-based violence? Does non-consensual sex without the use of force count as rape?)
- [Accuracy and reliability] Does the sample adequately cover all groups of women and men? (For instance, are individuals living in collective households also covered? Is the migrant population also covered? Are individuals interviewed in their own language?)
- [Accuracy and reliability] Is the sample large enough to adequately describe small relevant groups (e.g. survivors of sexual violence) and to allow for simultaneous disaggregation by sex and other relevant variables?
- [Accuracy and reliability] Is the questionnaire designed in a way that takes into account gender perspectives and avoids gender biases? (For example, are questions worded in a way that is equally understood and interpreted by both sexes? Is gender-neutral language used when referring to concepts such as occupations? Are leading questions (that is, questions encouraging particular answers) avoided? Are sensitive questions asked in a way that minimises under-reporting?)
- [Accuracy and reliability] Are interviewers properly trained to avoid asking questions in a leading way, particularly one that is gender stereotyped?
- [Coherence and comparability] Do questions try to avoid bias arising from cultural differences in norms and perceptions? (For instance, is care taken to avoid under- or over-reporting of gender-based violence due to differences in the social norms of what is or is not acceptable in a given culture?)
- [Accessibility and clarity] Do metadata provide adequate information on how gender issues have been addressed? (That is, do the metadata contain answers to the above questions?)

Please note that the above list is provided as a sample only and is by no means exhaustive. Assessing data quality from a gender perspective requires close attention to all factors at all stages of data collection and processing that may introduce a gender bias into the data or limit comparability. We provide more detail on this process in Section 4 'Guidelines for gender-sensitive data production and quality assessment'.

4. Guidelines for gender-sensitive data production and quality assessment

This section aims to provide general guidelines for original data providers on producing gender statistics and data that are appropriate for computing gender statistics. These guidelines should also be helpful for users of gender statistics, who can employ them as a reference point when evaluating the production processes of existing statistics (as described in the metadata). The content of this section is mostly based on UNECE's gender statistics guide ([21]), but also draws on three other guides on the topic, the UN's *Gender Statistics Manual* ([22]) and *Human Rights Indicators* guide ([23]) and the Canadian International Development Agency's *Guide to Gender-Sensitive Indicators* ([24]).

4.1. Motivation

The need to produce and share EU-wide, comparable, reliable gender statistics and indicators has been regularly highlighted by the European Parliament, the Council and the European Commission. In the Council conclusions on the European Pact for Gender Equality 2011–2020, Member States and the Commission, in particular through Eurostat, are encouraged to further develop existing statistics and indicators disaggregated by sex and to fully utilise the capacities of EIGE ([25]).

Considerable progress has been achieved to date throughout the EU in the regular production and dissemination of data disaggregated by sex, even when gender is not the main focus of the statistical activity. Yet major gaps still exist in terms of availability and comparability. The main challenges are in producing relevant, accurate and timely gender statistics. A lot still needs to be done to further standardise indicators, to establish links between official statistics and policies and to expand the production and dissemination of statistics particularly in areas such as business statistics and transport and communication statistics. Furthermore, existing ad hoc gender-relevant data collection initiatives should be converted into regular statistical programmes.

The overarching general principle in the production of high-quality gender-sensitive data is that of **gender mainstreaming.** At the highest level of generality, the European Commission's strategic engagement for gender equality for 2016–2019 defines gender mainstreaming as 'the integration of gender equality perspective into every aspect of EU intervention (preparation, design, implementation, monitoring and evaluation of policies, legal measures and spending programmes)'. In the context of statistics, we can apply the same definition, with 'data production' in the place of 'EU intervention'. More specifically, as the UNECE guide ([26]) explains, 'mainstreaming a gender perspective in statistics means that gender issues and gender-based biases are systematically taken into account in the production of all official statistics and at all stages of data production.' Moreover, gender mainstreaming also requires that women and men be equally involved in the development and implementation of statistical strategies and work plans, so that indicators and data are gender-sensitive and the priorities and needs of both genders are taken into account within the statistical data collection process.

([21]) United Nations Economic Commission for Europe and World Bank Institute, 2010.
([22]) United Nations Statistical Division, 2013.
([23]) United Nations, 2012.
([24]) Canadian International Development Agency, 1997.
([25]) http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/lsa/119628.pdf
([26]) United Nations Economic Commission for Europe and World Bank Institute, 2010.

4.2. Data collection modes and stages

For the purposes of this section, it is useful to distinguish five different data collection modes:

- population censuses;
- population-based sample surveys;
- business surveys;
- extraction of data from administrative records;
- compilation of data from one or more sources produced by other data providers using one of the four modes above.

While most of the steps described below apply to all of these modes of data collection, others are only applicable to some. The relevant data collection modes are pointed out when each step is introduced.

The general steps involved in the production of gender-sensitive data are as follows.

1. [All data collection modes] Selection of research topics relevant from the gender perspective and identification of data required.
2. [All data collection modes] Analysis, modification and extension of definitions, concepts and research methods that allow the gender specificities to be captured.
3. [Only administrative data collection modes] Administrative register design considering the gender perspective.
4. [Only sample-based data collection modes] Survey design considering the gender perspective:
   a. sample design (definition of the unit of enumeration and the sampling frame; choice of sampling method and sample size);
   b. choice of interview mode (online completion, one-on-one interview, one-on-one interview with self-completion blocks, etc.);
   c. questionnaire development and testing;
   d. interview protocol development, interviewer training and actual interviewing.
5. [All data collection modes] Data processing (coding, validation and cleaning, weighting, imputation).
6. [All data collection modes] Presentation of data and metadata:
   a. macrodata (statistics);
   b. microdata;
   c. metadata;
   d. data analysis and preparation of tables and graphs.

The rest of this section provides detailed pointers for each of these steps. Specific technical guidelines for preparing data for inclusion in EIGE's Gender Statistics Database are given in Section 6 'Technical guidelines for contributing to EIGE's Gender Statistics Database'.

4.3. Selection of research topics and identification of data required

In addition to incorporating a gender perspective into the analysis and collection of data on all topics, a number of specific topics are of particular interest for gender statistics: these are topics that specifically deal with 'differences and inequalities in the situation of women and men in all areas of life' (as per the United Nations Statistical Division definition of gender statistics). Of particular interest are questions on equal decision-making power, equal economic opportunities and status, work–life balance, elimination of gender stereotypes and freedom from gender-based violence.

Several of these gender statistics topics are still underexplored. These include, but are not limited to, the following:

- violence against women;
- behaviours affecting health;
- reproductive health and reproductive rights;
- intra-household allocation of resources (including time and power);

- participation in informal and non-formal education.

In the area of violence against women, there is a shortage of all varieties of data on types of violence. Even for the types of violence that are the most common and in principle the easiest to record (such as physical and social violence), administrative data (from police, judicial and medical records) are often incomplete and not internationally comparable due to non-harmonised definitions and survey data are scarce (to date, the 2012 European Union Agency for Fundamental Rights survey on violence against women is the only EU-wide survey on the topic). The data availability situation is even worse for types of violence that often remain unreported to the authorities and support services (such as economic and psychological violence and harassment), where it is necessary to rely almost exclusively on survey data. Even less is known about practices such as female genital mutilation and child or forced marriage, which are restricted to particular groups within society and are therefore impossible to adequately detect using even large-sample general population surveys.

Over the last decade, EIGE has incessantly worked toward closing many of these gaps in data on gender-based violence, particularly through its efforts to improve administrative data collection on violence against women (which includes both collecting and disseminating existing administrative data from EU Member States and assisting them in improving their administrative records) ([27]) and through dedicated studies on female genital mutilation ([28]), cyber violence against women ([29]) and the economic costs of gender-based violence ([30]).

Health is an area where large differences between women and men exist, in large part due to differences in behaviours (such as nutritional habits and substance abuse). However, comparable data on health-related behaviours, health status and access to healthcare in all EU Member States did not exist until 2014, when Eurostat conducted the second wave of the European Health Interview Survey (the first wave from 2008 only covered a subset of EU Member States).

Intra-household allocation of resources is another key area where inequalities between women and men persist, particularly with respect to allocation of time for care activities and household chores. Despite this, data in this area are scarce. The most reliable source of data on time allocation, the Harmonised European Time Use Surveys ([31]), only covers 18 EU Member States and the most recent data (from the 2010 survey) from some Member States are over 10 years old (as the data collection in the various Member States took place between 2008 and 2015).

4.4. Analysis, modification and extension of definitions, concepts and research methods

In the early stages of any data collection project, appropriate choice of definitions, classification systems and research methods is of paramount importance. When analysing the existing definitions, classification systems and research methods from a gender perspective, one must consider whether they properly ensure **coherence** and **comparability**, whether they are free from **gender biases** and whether they are sufficiently fine-tuned to detect hidden inequalities. When inadequacies are detected, the existing definitions, classification systems and research methods should be modified accordingly.

To ensure **coherence and comparability** of data across countries, it is very important that

([27]) https://eige.europa.eu/gender-based-violence/data-collection
([28]) https://eige.europa.eu/gender-based-violence/female-genital-mutilation
([29]) https://eige.europa.eu/gender-based-violence/cyber-violence-against-women
([30]) https://eige.europa.eu/gender-based-violence/estimating-costs-in-european-union
([31]) https://ec.europa.eu/eurostat/web/time-use-surveys

the same definitions and classification systems are used across all countries. This is particularly problematic when statistical evidence and analysis are produced on the basis of administrative data, as the definitions used by national administrations are often far from being harmonised at the EU level. Crime statistics (necessary for administrative data on gender-based violence) are an extreme example of data where lack of international harmonisation of definitions severely limits comparability and coherence of data. EIGE has been encouraging and helping EU Member States to harmonise their definitions, particularly through its 2016 study on terminology and indicators for data collection on rape, femicide and intimate partner violence ([32]). In the meantime, given the current lack of harmonisation, EIGE's Gender Statistics Database only presents data on gender-based violence from national administrative sources separately for each country and never in a single table for all countries, so as not to invite misleading international comparisons.

The use of definitions that are inappropriate for gender-sensitive statistics may introduce a **gender bias** that compromises accuracy and reliability for gender-related questions. For example, the standard definition of unemployment, which requires someone to be both actively seeking work and available to work in 2 weeks, may lead to underestimating unemployment among women, who may have been forced to leave the labour market temporarily and would now like to resume working, but are unable to launch a full job search due to care duties and might not be able to start work immediately due to the need to hire care services before starting employment. Similarly, when unpaid work is not included in the definition of economic output, the contribution of women may be underestimated.

Classification systems are also important. For example, when considering horizontal segregation in education, the traditional grouping of 'Social sciences, business and law' obscures massive gender differences within this catch-all field, with law, business administration and mathematical economics being heavily dominated by men and areas such as sociology and psychology equally dominated by women.

Questions pertaining to labour market participation are particularly important from a gender perspective, as this is an area where large gender differences still prevail. When classifying the reasons for part-time employment and inactivity, it is important to make sufficiently fine distinctions, such as distinguishing care for children from care for adult family members and care for the home. It is even more important to distinguish disparities in choices due to differences in preferences from disparities in choices due to differences in opportunities.

4.5. Administrative register design

Administrative data are defined as information collected primarily for administrative (not statistical) purposes such as registration, transactions and record keeping. This type of data is typically collected by public sector organisations with a specific decision-taking purpose in mind, and therefore the identity of the unit corresponding to a given record is essential (in contrast with general statistical records, where the identity of individuals is of no interest).

The use of administrative records avoids further direct data collection costs and respondent burden because it is data that already exists.

An additional advantage is the possibility of linking several data sources that use common identification codes, such as personal identity codes, real estate identifiers and business identity codes. Denmark, where various population registers and administrative records can be linked via personal identification numbers, provides a prime example. This is particularly useful in the area of gender-based violence, where police records, information from shelters for women exposed to partner violence and population registers can be linked to each other via the identification numbers of the victims and

([32]) https://eige.europa.eu/gender-based-violence/data-collection#2016

the alleged perpetrators to obtain a comprehensive picture of the crime and the individuals involved. The completeness of these records is ensured by requiring shelter personnel to keep meticulous records on all contacts with victims and by maintaining a police administrative system (in operation since 2001) that regulates uniform data registration and updating.

In order to link several data sources, a high level of cooperation among the administrative authorities is needed. A good example of this is the French interdepartmental unit for protecting women against violence and for combating trafficking in human beings, created in 2013 under the Ministry of Women's Rights. The unit works in close collaboration with the statistical departments of the Ministry of the Interior, the Ministry of Justice, the national statistical institute and non-governmental organisations (NGOs), in order to gather available data and to set up relevant indicators.

Due to their ready availability and the possibility of linking multiple registers through common identifiers, administrative data sources are increasingly used for statistical purposes. Nonetheless, it is important to keep in mind that these data are not specifically designed for statistical purposes and as such may not meet the quality requirements discussed here. In particular, as already mentioned in the previous section (in the context of gender-based violence), the concepts and definitions used usually follow the requirements of national laws and regulations and as such may not be suitable for international comparison. Consequently, before being used for statistical purposes, administrative data should always be subjected to careful quality analysis, focusing in particular on relevance, coherence and comparability.

4.6. Survey design

4.6.1. Sample design (definition of the sampling unit and the sampling frame; choice of sampling method and sample size)

Appropriate sample design is important for ensuring the accuracy and reliability of data. Issues of particular importance from a gender perspective involve appropriate choice of sampling and analysis units (taking care not to use household-level aggregates to make sweeping conclusions at the individual level), selecting sampling frames in ways that do not cause gender-biased under- or over-coverage and selecting sufficient sample sizes to enable inference about subgroups created by the intersection of sex and other variables.

For many surveys, the sampling unit is the household. When this is the case, several gender-related concerns must be taken into account: first, we have to consider the reference person (the person answering questions about the household). How is this reference person selected? Could men and women have a tendency to answer the household-level questions differently? If questions are about the completion of specific tasks within the household (such as childcare), is the reference person the most knowledgeable household member? Even more importantly, when the survey deals with issues that could be an object of contention within the household (such as intra-household allocation of tasks and resources), it may be necessary to use two reference persons to avoid a single-sided interpretation of the actual situation. Second, for questions where the unit of analysis is an individual while the sampling unit is the household, it is important to be extremely careful with the use of proxies (individuals answering on behalf of other individuals). Are proxies used? If so, for what types of questions? It is extremely important that one spouse is not allowed to answer on behalf of the other when the question pertains to potential objects of conflict between the spouses (such as violence in the household) or to personal time use (such as time spent caring for the children or tending to the house).

Sampling frames are often limited to private households. This approach may be problematic from a gender perspective as it excludes a number of gender-imbalanced groups, such as people living in retirement homes (mostly women), women in temporary shelters for victims of domestic violence and the prison population (predominantly men).

The sample size must be large enough to have reliable estimates separately for women and men. Furthermore, due to interactions between sex and other characteristics (see Section 4.8.1 'Macrodata (statistics)'), it is often necessary to report data disaggregated simultaneously by sex and another characteristic (such as age or educational achievement), in which case the sample size must be further increased to allow for reliable estimates for the resulting smaller subsamples. A further concern related to sample size arises when the survey aims at gathering information about a subgroup that comprises only a small proportion of the overall population (such as victims of physical or sexual violence or drug addicts). In such cases, it may be necessary to increase the sample size well beyond what is necessary for estimates about the general population. It may even be necessary to use less conventional sampling methods (such as respondent-driven sampling) to obtain add-on samples for such groups.

The effective sample size can be further reduced by non-response, which must be taken into account when computing the required target sample size. In addition, when non-response rates are high, caution must be taken even if the effective sample size would otherwise be sufficient, because non-response may be non-random (for example, victims of violence may be more likely to refuse to answer questions about violence than women who have not been subjected to violence), introducing a bias (that is, causing the statistics to be under- or over-estimates).

4.6.2. Choice of interview mode (online completion, one-on-one interview, one-on-one interview with self-completion blocks, etc.)

Particular care must be taken when questions are sensitive and respondents may be reluctant to answer sincerely because they might be afraid of being judged (as for drug use and sexual behaviours), prosecuted (as for criminal behaviour or tax evasion) or confronted by their spouses (as for intra-household conflict or gender-based violence). Self-completion blocks might be better suited for such questions. However, fully self-administered surveys are subject to the potential risk that questions may not be fully and correctly understood.

When sensitive questions are being asked in face-to-face interviews, it is important that the interview protocol (see also below) ensures that no third persons can overhear the interview and that the interviewer does not exhibit any prejudices in the way she or he reacts to the respondents' answers.

4.6.3. Questionnaire development and testing

Careful questionnaire design is extremely important in ensuring reliable and valid answers. From a gender statistics perspective, it is critical that questions be stated simply, unambiguously and in a way that is not likely to be understood differently by men and women. Questions should use gender-neutral language and should not be leading (i.e. should not be asked in a way that steers the respondent toward a particular answer). In the case of internationally comparable studies, it is important to ensure that questions are understood in the same way in different countries and that answers are minimally influenced by international differences in societal perceptions, norms and expectations.

Clear and unambiguous questions are key to obtaining reliable and unbiased answers. From a gender perspective, it is particularly important that questions are not asked in ways that are likely to be interpreted differently by women and men and by people of different educational and cultural backgrounds. Subjective questions, especially ones involving self-assessment or evaluation of one's well-being or health, pose a particular challenge from a gender perspective, as women's and men's assessments may differ due to different attitudes to the same objective situation (for example, men tend to give a more positive self-assessment). Such questions should be supplemented by more specific questions that are less open to interpretation.

One of the most important concerns in gender-sensitive data collection is the avoidance of leading questions. All questions must be asked in a neutral way, so that the respondent is in no way steered toward a particular answer. In particular, wording that has a negative connotation and may suggest a value judgment against certain answers should be avoided, as should abstract concepts that can be interpreted more narrowly or more broadly than the researcher has intended. For instance, in the case of gender-based violence, it is preferable to refer to concrete behaviours (such as 'hitting', 'slapping', 'threatening' or 'insulting') instead of 'violence'. The use of gender-neutral language can also be seen as a way of avoiding leading questions (for example, labelling an occupational category 'housewife' clearly steers male respondents away from this category).

Another important potential source of gender-related bias arises from international differences in norms and expectations. To enable international comparisons, sufficiently specific questions must be asked to minimise the extent to which answers are influenced by different interpretations due to differences in norms. Studies on gender-based violence are particularly susceptible to problems arising from differences in norms: countries with higher awareness of and lower tolerance for such forms of gender-based violence as sexual harassment and psychological violence may exhibit higher measured prevalence of such violence, simply because women are more likely to recognise and report it.

4.6.4. Interview protocol development and interviewer training

Even if questions are worded in a non-leading manner in the questionnaire, interviewers may still present them to respondents in a leading way, say, by straying from the language of the questionnaire when actually asking the questions, or showing their prejudices and biases when reacting to respondents' questions via facial expressions and body language. To avoid this, interviewers must be trained to remain neutral and true to the original wording of the questions and the interview protocol. Interviewers must be made aware of potential sources of gender biases that may arise during an interview and trained to avoid such biases. Ideally, questionnaires and interviewers should be jointly tested during pilot/trial interviews.

It is also important that sensitive questions, especially questions about gender-based violence and ones on topics that may be contentious within the household, are not asked to one household member in the presence of others. In addition, proxy answers (where one person responds on behalf of another) should never be permitted for such questions. Interviewers should not exhibit any prejudices in the way they react to the respondents' answers. For sensitive questions particular to women (particularly on gender-based violence and women's reproductive health), the interviewers should also be women. Care must also be taken to ensure confidentiality.

4.7. Data processing (coding, validation and cleaning, weighting, imputation)

Coding schemes must be sufficiently fine-tuned to detect gender differences (see also discussion in Section 3.1 'Definitions and scope'). When choosing imputation models for missing data, the modeller must be wary of potentially underestimating gender differences by unconditional imputation or overestimating them by imputation that is conditional only or almost only on sex. Post-stratification weights should be computed separately for women and men, as it is often the case that non-response levels are considerably higher for men than for women, resulting in a biased sex distribution in the final realised sample.

4.8. Presentation of data and metadata

4.8.1. Macrodata (statistics)

Data should be presented disaggregated by sex and often by additional characteristics. Tables should be clearly labelled. Values that are problematic from a data quality perspective (for example, due to limited comparability or suspicions regarding reliability) should be flagged. The size of the sample used for computing each statistic should be carefully considered; estimates based on low sample sizes should be flagged and those based on very low sample sizes should be dropped. All data should be accompanied by appropriate metadata.

Presenting data disaggregated by sex alone may not be sufficient, as intersectional inequalities are often present: inequality between women and men is more severe within specific groups of the population (such as immigrants, the undereducated, women of childbearing age and the elderly). When such concerns are present, it is important to provide simultaneous breakdowns by sex and these other intervening variables, insofar as the sample size allows it. In particular, a simultaneous breakdown by sex and age is almost always desirable. The choice of age groups is also important: for example, for labour market analyses, where the disadvantaged status of women of childbearing age is of particular importance, particular care shall be taken to establish age groups that cut across the boundaries of the typical childbearing age. The nature and different implications of intersecting inequalities have to be taken into account in the production of gender statistics.

4.8.2. Microdata

For microdata, data in the data file should be anonymised and properly labelled. All data should be accompanied by appropriate metadata.

4.8.3. Metadata

All data should be accompanied by a complete and thorough set of metadata so users can correctly interpret the data. Such metadata should include a thorough analysis of the quality aspects of the data, using the general framework of the ESS (see Section 2 'The general framework').The backbone of any data set's metadata should be formed by a metadata document following the ESMS format (see Section 8 'EURO-SDMX Metadata Structure 2.0').

In addition, because all potential sources of bias may not be immediately obvious to the data provider, it is important that original documentation is made available to end users and secondary distributors of data, such as EIGE. Such original documentation should include questionnaires, interview protocols and descriptions of data-processing algorithms.

The quality analysis portion of the metadata should also include a thorough analysis of the gender-specific issues discussed in this document. Unfortunately, the ESS metadata structures (ESMS, ESQRS and SIMS) do not have dedicated positions for gender analysis. Therefore, quality reports must include gender analysis under the generic quality-related sections.

One example of a good quality report is the methodology of the 'Women and men in decision-making' data collection available on EIGE's Gender Statistics Database web page ([33]).

4.8.4. Data analysis and preparation of tables and graphs

The analysis of gender statistics and their visualisation in tables and graphs follow the general rules of data analysis. Nevertheless, it is important to gear the analysis toward the main aim of highlighting gender challenges and inequalities. In this regard, the analysis and presentation of data requires cross-tabulation

([33]) Accessible through any indicator under the 'Women and men in decision-making' entry point.

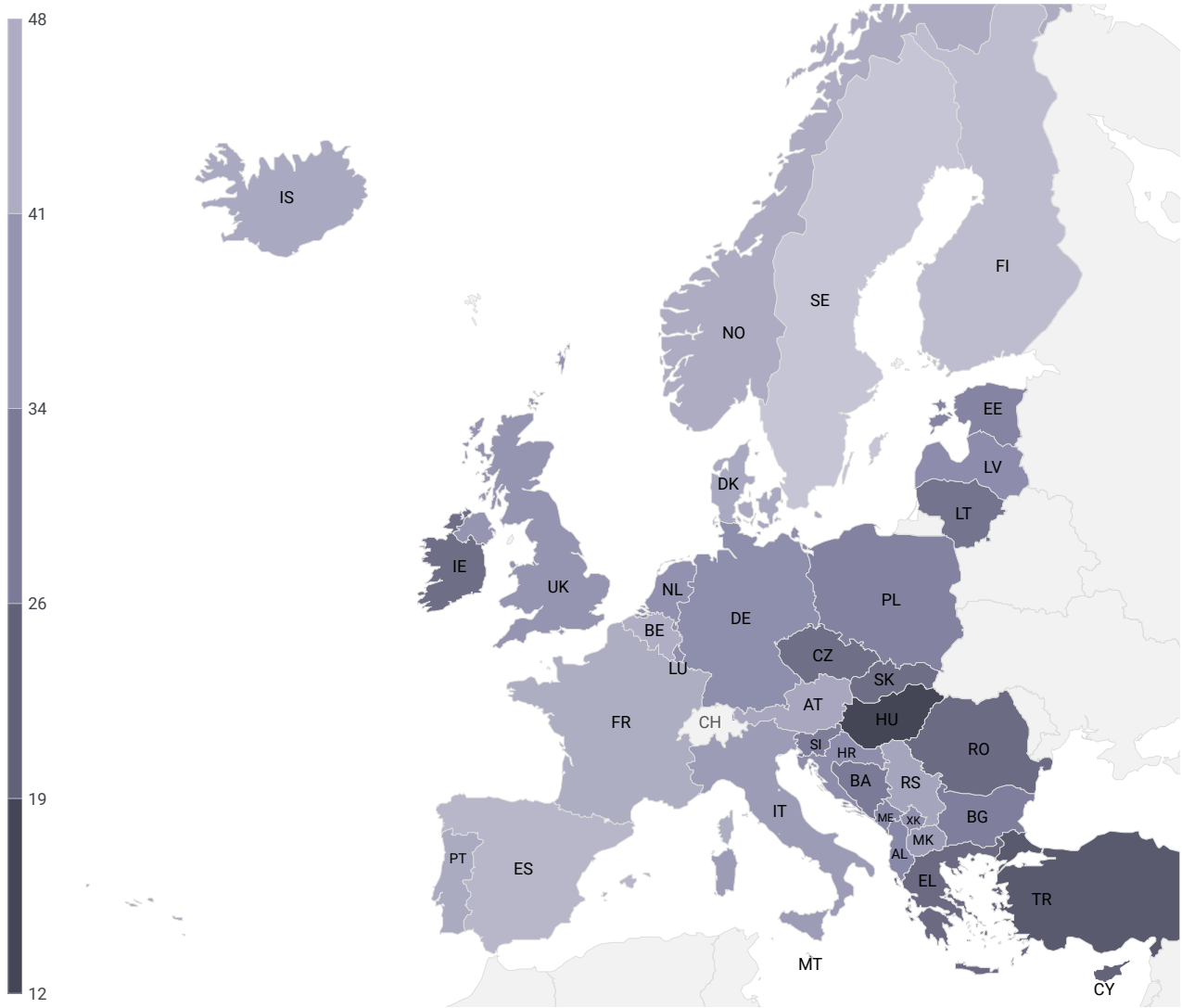
of sex with the characteristic or characteristics being studied, providing clear messages to users and stimulating their demand for more information.

Data analysis should start with descriptive statistics, which help to simplify large amounts of data in a sensible way, providing summaries of the basic features of the data. These summary measures may be, for example, proportions, rates, quartiles, averages or measures of the dispersion of the data. Additionally, it may be useful to study the degree of relationship between two variables (correlation analysis) as well as several measures of association to study the statistical strength of the relationship between the variables of interest.

Data and statistics in EIGE's Gender Statistics Database are presented as tables and visualisations (graphs and diagrams). While tables are an efficient tool to present detailed information and large amounts of data, visualisations are best at showing differences and trends at a glance. There are many potential visualisation options. The choice of the most appropriate visualisation depends on the nature of the data and the message to be visualised. Next, we present three examples of visualisations that are available in EIGE's Gender Statistics Database.

- Map: used to view all the available countries on a map, where each country is shaded on a gradient scale depending on the value of a certain indicator.

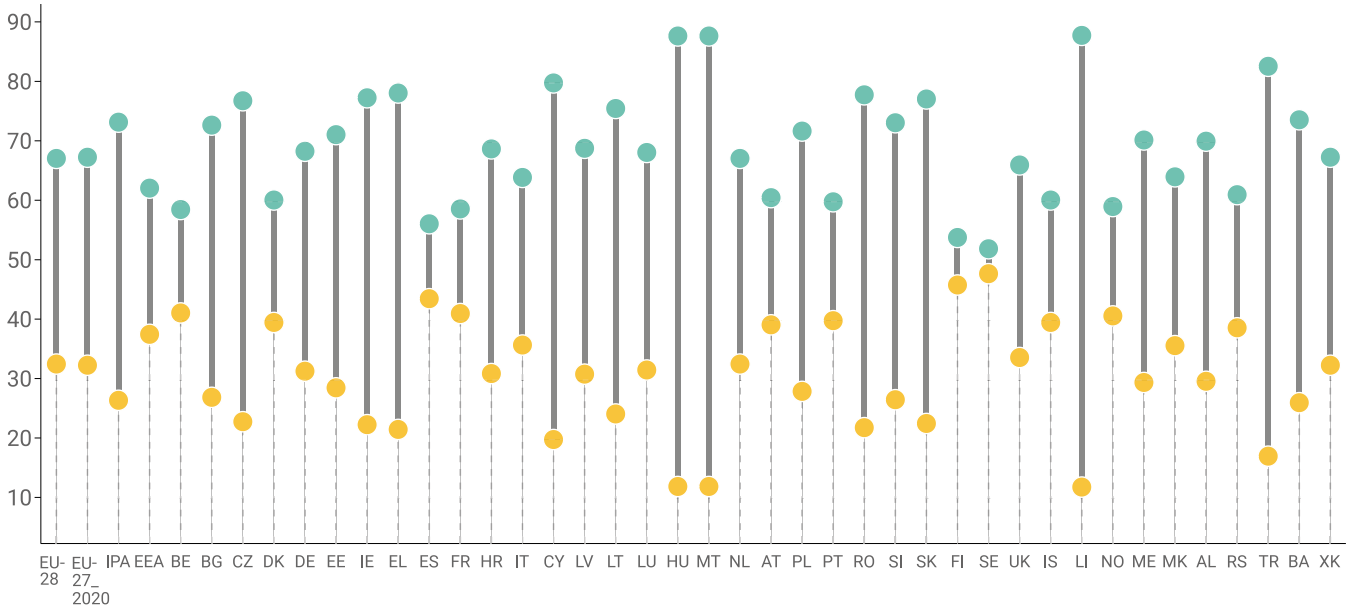
Figure 2. Number of women in national parliaments: presidents and members, 2020-Q2



Source: EIGE. For a list of country codes, see page 7.

- Bar chart: used to make comparisons among groups.

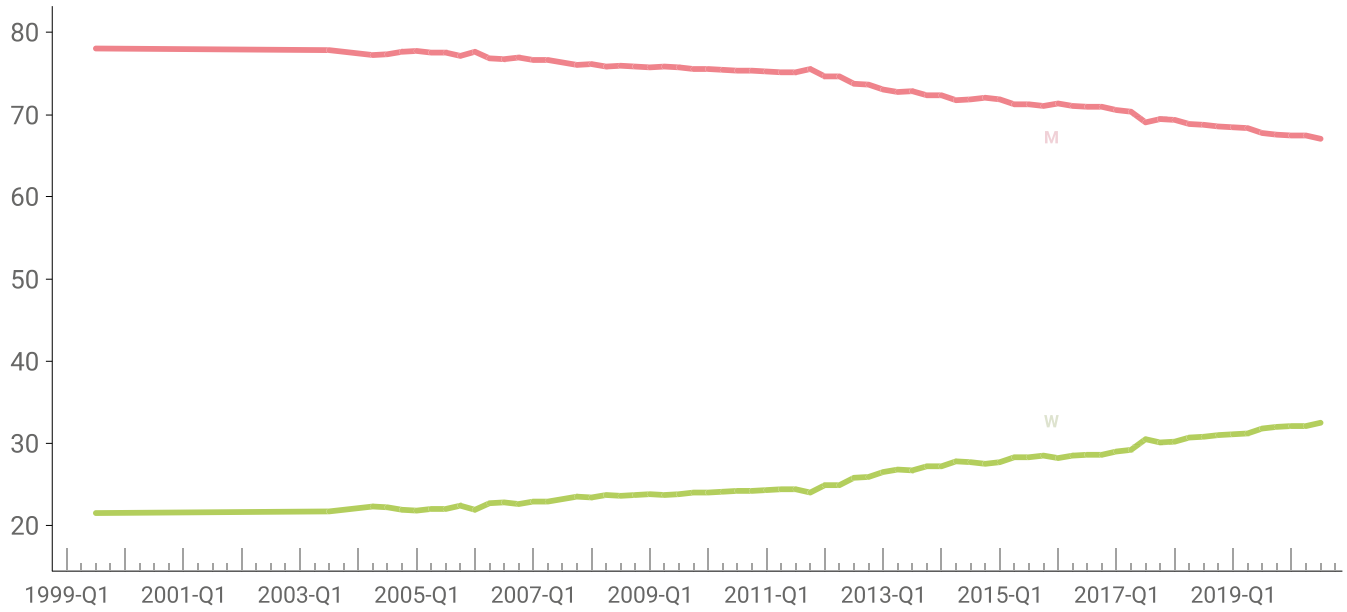
Figure 3. Number of women and men in national parliaments: presidents and members, 2020-Q2



Source: EIGE. For a list of country codes, see page 6.

- Line chart: used to track changes in indicators over time.

Figure 4. Proportion of women and men in national parliaments: presidents and members, EU-28, 1999-Q1–2020-Q2



Source: EIGE. For a list of country codes, see page 6.

4.9. Executive summary

The overarching general principle in the production of high-quality gender-sensitive data is that of gender mainstreaming. As the UNECE guide ([34]) explains, 'mainstreaming a gender perspective in statistics means that gender issues and gender-based biases are systematically taken into account in the production of all official statistics and at all stages of data production.'

First, one must ensure that the definitions, classification systems and research methods, when used properly, ensure coherence and comparability (particularly by being harmonised across countries and stable over time), are free from gender biases and are sufficiently fine-tuned to detect hidden inequalities.

Second, when survey-based data collection methods are used, it is important to choose sampling units and sampling frames that do not obscure intra-household inequalities and do not overlook populations outside private households. Sample sizes must be adequate to draw conclusions about subpopulations defined by the intersection of sex and characteristics such as age and occupation.

Third, questions should be stated simply, unambiguously and in a way that is not likely to be understood differently by men and women or by people in different countries. Gender-neutral language should be used and questions should not be leading (i.e. should not be asked in a way that steers the respondent toward a particular answer). Interviewers must be trained not to exhibit any prejudice when asking sensitive questions (and self-completion blocks should be considered as an alternative to face-to-face interviews in the case of the most sensitive questions).

Finally, all statistics should be disaggregated by sex and often also by additional characteristics (such as age and educational attainment), to highlight both overall gender inequalities and intersectional inequalities. All data should be accompanied by adequate metadata. Data analysis methods and visualisations should be chosen taking into account the nature of the data and the overall aim of investigating and highlighting gender gaps.

([34]) United Nations Economic Commission for Europe and World Bank Institute, 2010.

5. Evaluation and selection of sources for EIGE's Gender Statistics Database

Data sources and particular data products considered for inclusion in EIGE's Gender Statistics Database are evaluated using the general framework of the ESS discussed in Section 2 'The general framework', paying particular attention to the specific requirements for the production of gender statistics, as outlined in Section 4 'Guidelines for gender-sensitive data production and quality assessment'. In addition, EIGE also produces original data (most notably, the collection on 'Women and men in decision-making' and the compilation of national administrative data on gender-based violence), observing the good practices outlined in this document to the fullest extent possible.

5.1. Evaluation and selection of data providers

The overall quality of data providers is assessed following the institutional and statistical processes principles outlined in the *European Statistics Code of Practice* ([35]). The institutional principles entail: professional independence (principle 1), coordination and cooperation (1a), mandate for data collection and access to data (2), adequacy of resources (3), commitment to quality (4), statistical confidentiality and data protection (5) and impartiality and objectivity (6). The statistical processes principle**s** consist of sound methodology (7), appropriate statistical procedures (8), non-excessive burden on respondents (9) and cost-effectiveness (10). For the purposes of selecting providers for the database, principles 1, 1a and 3–8 are the most important. The overarching goal is to ensure that the provider has the resources and the good practices to ensure high quality of statistical outputs. Only once the provider is deemed suitable is the quality of individual statistical outputs evaluated on the basis of metadata analysis.

When official sources ([36]) are available, they are prioritised over other sources. Within official sources, Eurostat data are prioritised over data from other official sources, in order to ensure international comparability of data. For the same reason, data collected by international organisations are generally prioritised over data provided by national organisations.

Finally, where external sources are lacking, a number of data sets are produced by EIGE itself. This includes a major data collection where EIGE is the primary provider ('Women and men in decision-making'), a compilation of data from national administrative sources on gender-based violence and several sets of indicators computed at EIGE from primary data provided by external sources (most notably, these include multiple indicators in the 'Beijing Platform for Action' entry point and the entire 'Gender Equality Index' entry point). For all of these collections, EIGE follows the quality principles and guidelines outlined in this document.

5.2. Evaluation and selection of statistical outputs

5.2.1. General principles

The evaluation of statistical outputs follows the five principles in the statistical output block of the *European Statistics Code of Practice*: relevance (principle 11), accuracy and reliability (12), timeliness and punctuality (13), coherence and comparability (14) and accessibility and clarity (15), discussed in detail here in Section 2 'The general framework'.

Relevance is an important first consideration when deciding whether to include a data set in

([35]) Eurostat, 2018.
([36]) That is, Eurostat and national statistical institutes, which are formally bound by the *European Statistics Code of Practice* (Eurostat, 2018).

the database. Given the focus of the database, this primarily entails deciding whether the data set meets the definition of gender statistics as outlined in Section 3.1.1 'Gender statistics'. This analysis is based on the available metadata and judgement of gender experts. If a statistic has no or very limited connection to gender equality, it is not selected, even if it can be sex disaggregated. As an example, the European Working Conditions Survey includes some questions related primarily to the characteristics of different occupations rather than gender differences. These are not included in the database, although it is possible to calculate sex-disaggregated statistics. The exclusion of such questions is necessary to avoid overcrowding the database with statistics that have very limited connection to gender equality.

Once a data set or a survey question (which can be converted into a data set in the database by computing summary statistics for the relevant subpopulations) is deemed relevant, the remaining quality dimensions must be evaluated. In principle, this can be done both qualitatively and quantitatively. However, given the resources available to EIGE and the large body of data to be evaluated, we are currently relying solely on a qualitative assessment of the documentation provided by the original data provider.

This evaluation process is straightforward when the source has provided adequate ESMS metadata, in which case careful examination of the metadata is generally sufficient for an overall quality assessment. Otherwise, all available documentation is scrutinised by EIGE to create a quality assessment following the ESS framework. In the case of microdata in particular, questionnaires and sampling schemes are examined to identify gender bias and limitations to accuracy due to sample under-coverage or insufficient sample size.

With respect to sample size, it must be noted that what matters is not the total sample size of the survey, but rather the effective sample size for the statistics to be displayed. This means that one must take into account response rates and the size of the subsample for which the statistic is reported. For instance, even if the overall sample size is large enough, there may not be enough women respondents in a particular occupational group to compute reliable statistics for that group. It is also necessary to examine response rates and to understand to what extent non-response may be non-random and a source of possible bias. For all of these reasons, it is necessary to examine the original data set and not just the metadata before a decision on reporting a statistic can be made.

5.2.2. Minimum standard

When a given statistic is available from multiple sources, the highest quality source is selected. When a statistic is available from only a single source (such as most attitude measures from the Eurobarometer ([37]) and survey-based estimates of the prevalence of various forms of gender-based violence from the EU-wide European Union Agency for Fundamental Rights survey on violence against women), that source is almost always included, as long as it meets a minimum quality standard: namely, data meeting the definition of gender statistics are available **(**[38]**)** and sufficient metadata are provided to identify the source and to understand how and what is being measured.

5.3. Presentation of quality-compromised data

When the minimum standards outlined above are met but quality assessment shows serious inadequacies, the data are included in the database,

([37]) https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/General/index

([38]) In most cases, a necessary (although not sufficient) condition for this is that sex-disaggregated data are provided (or computable from microdata).

but its presentation is changed to reflect these inadequacies. This is done on several levels. First, any such inadequacies are made explicit in the appropriate section of the structured metadata of the database ([39]). Second, when there are particularly severe problems with comparability across countries (as is the case with administrative data on gender-based violence), the data tables for different countries are displayed separately to discourage users from direct comparisons when such comparisons are not warranted by the data. Third, problems affecting only particular data points within the data set are highlighted by flagging (the flags can be seen in the data table view of each data set, as well as in exported data).

Flags can be thought of as footnotes to data points. In the data table, they are indicated next to the data point by a superscripted single letter in parentheses. Hovering the mouse cursor over a flag brings up a pop-up explaining the flag (see Figure 5). The following flags currently exist in the database.

- b break in time series
- c confidential
- d definition differs, see metadata
- e estimated
- f forecast
- i see metadata
- n not significant
- m imputed
- p provisional
- r revised
- s Eurostat estimate
- u low reliability
- x dropped due to insufficient sample size (n < 20)
- y unreliable due to small sample size (n < 50)
- z not applicable

For the most part, the flags are provided by the original sources (most notably Eurostat). The flags 'x' and 'y', however, are created at EIGE while computing statistics from microdata. When only a small number of observations are available for a statistic, the statistic is deemed unreliable due to the sampling error being high. Consequently, mitigating action is taken: when there are fewer than 20 observations, the statistic is not computed and the corresponding missing data point is flagged with 'x'; when there are at least 20 but no fewer than 50 observations, the statistic is computed and reported in the data point but is flagged with 'y' to indicate its unreliability.

Finally, we should note that when ESMS metadata are provided by the source (such as Eurostat), EIGE currently publishes the metadata in unaltered form. In the future, however, EIGE plans to carry out a complete re-evaluation of all data and metadata from a gender statistics perspective, augmenting the original metadata document with specific gender-related quality analysis sections.

([39]) See the structure of the metadata in Section 8 'EURO-SDMX Metadata Structure 2.0' and a detailed discussion of the framework in Section 2 'The general framework'.

Figure 5. Flags in a data table

✖ANSWER ▸	PCT_MEN	PCT_WOMEN	PCT_EQUAL	PCT_NONE	PCT_DK	PCT_REFUSE
isco08 ▸ geo ▾	ELEM	ELEM	ELEM	ELEM	ELEM	ELEM
EU28	9.1	45.4	23.4	18.9	3	0.3
BE	5	69.2	12.9	10.1	2.8	0
BG	4.2	60.1	25.3	10.2	0.2	0
CZ	[illegible]	64	11.2	8	0	0
DK	11.2(y)	52(y)	15(y)	19.8(y)	2(y)	0(y)
DE	6.2	51.9	24.9	11.6	5.4	0
EE	5.3	64	14.7	12.1	3.8	0
IE	6.2(y)	53.6(y)	33.3(y)	7(y)	0(y)	0(y)
EL	13.2	35.6	30.8	17	1.7	1.7
ES	10.1	47.6	17	24.3	0.6	0.4
FR	9.8	44.7	21.9	21.7	1.8	0
HR	5.2	43.8	20.7	29.1	1.3	0
IT	8	30	14.2	44.2	3.6	0
CY	16.9	38.7	24.3	20.1	0	0
LV	6.1	47.7	26	16.5	1.9	1.9
LT	12.5	61.2	16.1	9.2	0.9	0
LU	6	52	15.5	24.6	1.9	0
HU	10	45.5	29.1	11.1	4	0.3
MT	10.7	52	23.8	13.5	0	0
NL	8.9	54.5	24.6	4.2	7.7	0
AT	11.6	45.1	24.8	13.8	3.4	1.2
PL	11.2	38.6	37.4	6.6	4.8	1.5
PT	4.6	42.7	13.7	36.2	2.6	0.2
RO	5.3	47.9	26.2	18.5	2	0
SI	13.5	47.9	26.6	11.2	0.8	0
SK	5.9	58.6	29.5	4.3	0.4	1.3
FI	11.8	51.3	12.7	21.5	2.7	0
SE	26.6(y)	35.2(y)	27.6(y)	10.6(y)	0(y)	0(y)
UK	13.9	39	32.4	12	2.7	0

y - unreliable due to small sample size (n<50)

Source: EIGE's Gender Statistics Database (European Working Conditions Surveys (EWCS), original source: Eurofound). Retrieved on 29/03/2018.

Table 2. External data providers and corresponding data collections

Data provider	Data collections
Eurostat	Administrative data collections
	EU Labour Force Survey (EU-LFS)
	European statistics on Income and Living Conditions (EU-SILC)
	European Health Interview Survey
	European Information and Communication Technologies Survey
	Joint Eurostat–UNODC data collection on crime and criminal justice
Eurofound	European Working Conditions Survey
	European Quality of Life Survey (EQLS)
Eurobarometer	Eurobarometer special surveys
European Union Agency for Fundamental Rights	Survey on violence against women
European Commission, Directorate-General for Research and Innovation	'She figures' reports on gender equality in research and innovation
Organisation for Economic Co-operation and Development (OECD)	OECD programme for international student assessment
Women Against Violence in Europe	Reports on women's support services

5.4. Current data sources

5.4.1. External

Table 2 lists the external data providers and the corresponding data collections currently used in EIGE's Gender Statistics Database (as of the summer of 2020). The number of data sets varies widely by source, with Eurostat being the dominant provider.

5.4.2. Internal

The following data collections are carried out by EIGE.

- **Women and men in decision-making** ([40])
  This collection contains data on the numbers of women and men in key decision-making positions across a number of different life domains (politics; public administration; the judiciary; business and finance; social partners and NGOs; environment and climate change; media; science and research; and sports). The data are updated regularly (from once a quarter to once every 2 years, depending on the domain) and form a separate entry point in the database.

- **Administrative data collection on violence against women** ([41])
- Data on gender-based violence in Europe is still scarce. A major problem is the lack of harmonised definitions and data collection practices across EU Member States. EIGE is actively involved in helping Member States to improve their administrative data collection practices on gender-based violence. As part of this effort, EIGE regularly collects data on gender-based violence using a standardised tool. The results of this data collection are available throughout the 'Gender-based violence' entry point of the database, in the administrative data sections. Due to inconsistencies in the definitions and practices adopted in different Member States, the data from different Member States are kept in separate tables, which can be accessed through the metadata sections of the corresponding data sets.

- **Female genital mutilation** ([42])
- EIGE has carried out three studies on the subject and is in the process of carrying out a fourth. The data are available through the 'Gender-based violence' entry point of the database, in the female genital mutilation section.

([40]) https://eige.europa.eu/gender-statistics/dgs/browse/wmidm
([41]) https://eige.europa.eu/gender-based-violence/data-collection
([42]) https://eige.europa.eu/gender-based-violence/female-genital-mutilation

In addition to these data collections, EIGE has also computed several sets of indicators and estimates using external sources.

- **Gender Equality Index**
  First computed at EIGE in 2013 and updated regularly since then, the Gender Equality Index is a set of composite indicators that measures EU Member States' progress toward achieving gender equality, both overall and in several individual life domains. The results are available in the Gender Statistics Database via a separate entry point ([43]).

- **Beijing Platform for Action**
  In 1995, the Fourth World Conference on Women adopted the BPfA ([44]). The BPfA addresses 12 critical areas of concern requiring particular action towards the advancement of women. Since 1999 a number of quantitative and qualitative indicators have been developed by the presidencies of the European Council to monitor progress towards the achievement of the goals of the BPfA. The 'Beijing Platform for Action' entry point of the database presents these indicators, as computed by EIGE from the available external and internal sources ([45]).

- **Economic benefits of gender equality**
  In 2017, EIGE commissioned a study on the economic benefits of gender equality. Based on the E3ME econometric model developed by Cambridge Econometrics ([46]), the study models the potential benefits of gradually increasing gender equality in the EU over time (from 2020 to 2050). The results of this forecasting exercise are included in the Gender Statistics Database as a separate entry point. It should be noted, however, that this entry point contains forecasts, rather than statistics, and therefore is not fully subject to the same quality standards as statistics. In particular, it is important to note that the quality of the forecasts depends on the validity of the underlying econometric model ([47]).

5.5. Executive summary

Data sources and particular data products considered for inclusion in EIGE's Gender Statistics Database are evaluated using the general framework of the ESS discussed in Section 2 'The general framework', paying particular attention to the specific requirements for the production of gender statistics, as outlined in Section 4 'Guidelines for gender-sensitive data production and quality assessment'. In addition, EIGE also produces original data (most notably, the collection on 'Women and men in decision-making' and the compilation of national administrative data on gender-based violence), observing the good practices outlined in this document to the fullest extent possible.

The algorithm for the selection and presentation of data from external sources can be summarised as follows.

1. Identify all sources for the given data item (potential data set in the database).
2. Keep only the sources and data items that meet the following minimum requirements:
   a. the data provider is deemed acceptable in relation to the institutional and statistical processes principles of the *European Statistics Code of Practice*;
   b. the data item conforms to the definition of gender statistics;
   c. sufficient metadata are available to identify the source and to determine how and what is being measured.
3. Include in the database the highest-quality source among those meeting the minimum requirements.
4. When serious quality problems exist, change the presentation of the data to reflect these problems:
   a. reflect all quality issues in the metadata;
   b. do not display non-comparable data side by side;
   c. flag problematic data points.

([43]) https://eige.europa.eu/gender-equality-index
([44]) http://www.un.org/womenwatch/daw/beijing/pdf/BDPfA %20E.pdf
([45]) https://eige.europa.eu/gender-statistics/dgs/browse/bpfa
([46]) https://www.camecon.com/how/e3me-model/
([47]) https://eige.europa.eu/gender-mainstreaming/policy-areas/economic-and-financial-affairs/economic-benefits-gender-equality

6. Technical guidelines for contributing to EIGE's Gender Statistics Database

This section provides suggestions on how to best prepare data for inclusion in EIGE's Gender Statistics Database. It begins with a brief technical description of the database, namely its logical structure and technical implementation. However, this information should be considered optional: all information that is strictly necessary for the preparation of data for the database is included in the second part of this section, where step-by-step instructions are provided along with examples, with the main objective of improving the quality and availability of gender statistics.

6.1. Background: structure of the database

6.1.1. Principles and building blocks

EIGE's Gender Statistics Database is a collection of statistical data and associated metadata pertaining specifically to the area of gender statistics. The logical structure of the database is based on the SDMX standard, which is an international standard for the organisation, production and exchange of statistical information (data and metadata) among various data providers and users ([48]).

The basic (lowest-level) organising elements ('building blocks') of the database are data sets. While in general the term 'data set' tends to be used with a number of different meanings throughout statistics, EIGE follows SDMX's and Eurostat's convention in defining a data set as a set of observations that all meet the following two conditions: (1) they measure the same underlying concept (such as 'employment', 'employment rate', 'level of education', 'life expectancy', 'satisfaction with life'), and (2) they are defined in terms of the same criteria (to be defined formally below, but loosely meaning the same breakdown variables, such as 'sex', 'age', 'educational achievement'). This second condition implies that 'Employment by sex and level of education' and 'Employment by age' are two separate data sets, unless observations also exist for intersections of age and level of education (e.g. separate observations for young college graduates and old college graduates). Note that even with this fairly narrow definition, the division of data into data sets is still somewhat arbitrary. For example, this definition allows the storing of 'Employment in 1 000s of people' and 'Employment rate (%)' in either two separate data sets (each having 'unit of measurement' as a data set-level attribute) or a single data set, 'Employment and employment rates', which would then have an additional two-category criterion, which could be entitled 'unit of measurement' (with '1 000s' and ' %' as categories) or 'indicator' (with 'employment in 1 000s' and 'employment rate in %' as categories). In general, EIGE adheres to the following rule: whenever the immediate data source has provided data sets in SDMX format, EIGE respects the source's division of data into data sets (for example, Eurostat provides 'Employment in 1 000s' and 'Employment rate' as separate data sets when the criteria are 'time', 'country', 'sex', 'age' and 'nationality', but it provides the same two measures as categories within an 'indicator' criterion when the only other criteria are 'time', 'country' and 'sex'). In all other cases we group data at what EIGE deems to be the most natural level for that particular collection of data.

Data sets are internally structured as follows. Each observation is a number linked to a set of qualifying criteria **([49])** (which both identify and

([48]) http://ec.europa.eu/eurostat/web/sdmx-infospace/welcome

([49]) The SDMX standard refers to these as 'dimensions'. In this report, we reserve the word 'dimensions' for the dimensions of data set views (to be defined below) and use the word 'criteria' when referring to the general data structure.

describe the observation) and attributes (which only describe the observation). Each criterion has a finite number of possible values (categories). Together, the criteria form a multidimensional coordinate system, also known as a cube, where each point corresponds to exactly one category of each criterion. Each observation (a real number) is associated with a point in the cube. In addition, the attributes of this observation provide supplementary information that help interpret the number. These ideas are best described by example. Suppose our data set represents a table containing the average annual employment rates for women and men in the EU-28 for each year between 2000 and 2019. In this case, the data have three criteria: 'sex' (with two categories, 'men' and 'women'), 'country' (with 28 categories, each Member State being one category) and 'year' (with 20 categories, each year being one category). The resulting three-dimensional cube will therefore have 2 x 28 x 20 = 1 120 data points (observations), one for each possible combination of categories (such as ('women', 'Germany', 2000)). Each observation will be the employment rate of the group defined by the corresponding categories (in our example, German women in 2000). In addition to the criteria, each observation will have a series of attributes, some of which will be at the data set level (the subject/concept measured (employment rate in our example) and unit of measurement (% in our example)), while others will be at the observation level (such as flags indicating whether the observation has been estimated or whether it represents a break in a series). Data set-level attributes are part of the metadata for that data set, while observation-level attributes are part of the data.

The number of criteria in a data set can vary, but two criteria defining the basic unit of observation are always required: a criterion defining a geographical area ('country' in the example above) and a criterion defining a time period ('year' in the example above). While most observations in the database are at the country–year level, there are data sets with subnational observations (such as the data sets on the composition of regional assemblies under the 'Women and men in decision-making' entry point) and data sets with higher-than-annual frequencies (such as the data set on the composition of national parliaments, which has quarterly observations). In addition, all data sets, except those that provide direct measures of the relative situation of men and women (such as values of the Gender Equality Index, gender pay gap measures, measures of gender gaps in other variables (prepared for the index) and indices of horizontal segregation in occupations and education) also provide sex-disaggregated data, i.e. contain the criterion 'sex'.

As stated in the previous paragraph, the database contains solely macrodata ([50]). However, most of these macrodata have been produced from microdata, either by the original or immediate source or by EIGE and its contractors. Depending on the way the data have been processed, we can distinguish three types of data: (1) data provided as macrodata by the source and used in the database as is (such as employment rates), (2) data computed by EIGE or its contractors from other data provided as macrodata by the source (such as the gender pay gap in monthly wages, computed from levels of monthly wages that are obtained as country-level aggregates from Eurostat), (3) data computed by EIGE and its contractors from microdata (such as data on attitudes and opinions, computed from Eurobarometer microdata).

Data sets are displayed to users with the help of data set views (DSVs), which define how the data set will be displayed to the user. The DSVs define data visualisations (line charts, maps and bar charts) and customisable data tables. A data table is a two-dimensional layout presenting all or part of a data set to the user. To define a table, one must specify one or more criteria

([50]) **Macrodata** are statistical data observed at the level of countries or other geographical regions. This includes both data that are directly measured at the country level (such as gross domestic product) and aggregates (country-level statistics) of microdata (such as unemployment rate estimated by the EU LFS or public opinion as gauged by the Eurobarometer survey), while **microdata** are statistical data observed at the level of individuals, households or firms (such as data from population surveys).

as row dimensions and one or more criteria as column dimensions. The table is now formed by the intersection of: (1) one row for each possible combination consisting of one category for each row dimension and (2) one column for each possible combination consisting of one category for each column dimension. The value displayed in any given cell is the observation pertaining to the categories defining the corresponding row and column.

Data set views are arranged in a tree structure, where they are grouped into several levels of **themes** (branches of the tree). The highest level themes are known as entry points. The current entry points are listed in Table 3. In principle, any given data set may have multiple associated DSVs linking this data set to a number of different themes, although this is uncommon; the current organising principles of the database aim to avoid redundancies in the tree.

Table 3. The entry points of EIGE's Gender Statistics Database

1. EU policies and strategies This entry point includes indicators that help monitor the progress made by EU Member States towards the achievement of the gender-relevant targets set out by the following EU policies and strategies. • Strategic engagement for gender equality 2016–2019 • European Pillar of Social Rights • Europe 2020 strategy • Barcelona strategy.
2. Thematic areas This entry point contains data that help assess gender differences in several life domains, specifically: • work and the labour market • research, science and digital society • education, training and skills • demography and migration • fertility and reproductive rights • health and mortality • living conditions • time use • public opinion.
3. Beijing Platform for Action In 1995, the Fourth World Conference on Women adopted the BPfA. The BPfA addresses 12 critical areas of concern requiring particular action towards the advancement of women. The 'Beijing Platform for Action' entry point presents the indicators adopted by the European Council to monitor progress towards the achievement of the goals of the BPfA in each of the 12 areas of concern: • women and poverty (A) • education and training of women (B) • women and health (C) • violence against women (D) • women and armed conflict (E) • women and the economy (F) • women in power and decision-making (G) • institutional mechanisms for the advancement of women (H) • human rights of women (I) • women and the media (J) • women and the environment (K) • the girl child (L).
4. Gender Equality Index This entry point includes the scores of EIGE's Gender Equality Index and the data used to calculate these. The index is a comprehensive measure developed by EIGE for monitoring progress in gender equality across the EU over time. It measures gender equality through a selection of indicators divided into six core domains (work, money, knowledge, time, power, health) and two satellite domains (intersecting inequalities and violence against women).

5.	Women and men in decision-making This entry point contains data on the numbers of women and men in key decision-making positions across a number of different life domains: • politics • public administration • the judiciary • business and finance • social partners and NGOs • the environment and climate change • the media • education, science and research • sports.
6.	Gender-based violence This entry point contains data on gender-based violence. The elimination of violence against women involves challenging the unequal division of social, political and economic power among women and men, and the ways in which this inequality is perpetuated through institutions at all levels of society. The sub-themes of this entry point are: • physical violence • sexual violence • psychological violence • economic violence • intimate partner violence • female genital mutilation • attitudes and perceptions • support services • statistical capacity and availability of data.
7.	Gender Mainstreaming This entry point presents indicators related to different aspects of gender mainstreaming. In particular, resulting from EIGE's studies developed under this area.

Last but not least, **metadata** (or 'data about data') are an integral part of the database, as the information they contain makes it possible for users to understand, interpret, evaluate and analyse statistical data. Metadata are crucial for ensuring data quality along the accessibility and clarity dimension. There are two main types of metadata: (1) structural metadata, which provide a structured description of the way the statistical data and the reference metadata are organised, and (2) reference metadata, which provide additional descriptive information on the concepts used, the data collection and generation methods employed and the quality of the data. Structural metadata essentially amounts to a formal definition and description of the data structure outlined above. The end user does not necessarily have to be aware of most structural metadata. Reference metadata consists of an extensive, mostly free-form description of the data, which allows the user to understand and evaluate various facets of the data, including (but not necessarily limited to):

- what the data purport to measure;
- how these measurements have been made;
- how the measurements should be interpreted;
- who is responsible for collecting and disseminating the data;
- how often the data are updated and disseminated;
- where the updated data and additional information can be found;
- how high the quality of the data is (within the framework described earlier).

For an in-depth discussion of statistical metadata, see Dippo and Sundgren (2000) ([51]).

In EIGE's database, structural metadata (such as the names and codes of data sets and dimensions and the code lists of criteria and attributes) are embedded in the data, and reference metadata are presented in the database alongside the data. The reference metadata follow the ESMS structure, described in detail below, as part of the practical guidelines for preparing data and metadata for the database.

6.1.2. Technical implementation

EIGE's Gender Statistics Database is a NoSQL database. A NoSQL database consists of a set of collections, each of which holds a set of documents. A document is JSON-style data structure composed of field-and-value pairs. Documents have dynamic schema, which means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data. The database stores documents on disk in the BSON serialisation format. BSON is a binary representation of JSON documents, though it contains more data types than JSON. For the BSON specification, see the following link: http://bsonspec.org

The collections that comprise the database are shown in Table 4.

Table 4. Collections in EIGE's Gender Statistics Database

COLLECTION	DESCRIPTION
Themes	This collection represents the browsing tree (themes tree).
DataSets	This collection contains a list of all data sets in the database, along with reference metadata, keywords, warnings, computation configuration and references to data set views. Data are stored in a separate collection.
Data_<DatasetCode>	These collections contain the actual statistical data (one collection for each data set).
DataSetViews	This collection contains information about how each data set should be displayed within each theme (each branch of the browsing tree). A data set view has its own code, name and keywords and contains information about the default row and column dimensions and the default and hidden data set criteria and categories.
DataSetViewsOptions	This collection contains some additional configuration options for DataSetViews. It is currently used to determine which views have associated predefined tables.
Criteria	This collection contains the criteria and their categories from all data sets. The collection also contains information about the usage of criteria and their categories in data sets.
Flags	This collection lists the available flags.
GeographicRegions	This collection lists all the countries and regions of interest (if an imported data set contains observations out of the defined geographical scope, the corresponding records will be skipped and warnings will be logged). The collection also contains information about the usage of geographic regions in data sets.
TimePeriods	This collection lists all time periods used in the database. Each time period includes a list of the data sets in which it is used and their overall number, the time format applied (e.g. P3M for quarterly data such as 2018-Q1), and the year to which the time period applies (e.g. 2018-Q1 applies to 2018).
Years	This collection is a list of all years used in the database, each with a list and count of the data sets that include observations for that year.
Criteria	This collection contains the criteria and their categories from all data sets. The collection also contains information about the usage of criteria and their categories in data sets.

([51]) Dippo, C. S. and Sundgren, B., 2000.

Keywords	This collection lists all keywords (candidates plus those specified for individual data set views) and which data sets and data set views each keyword applies to, with the total number of each.
Keywords.Candidates	This collections lists all keyword candidates and their synonyms.
SystemValues	This collection contains technical information about the database – loaded object counts by entity type, initial load date and date when last updated. This collection contains only one object (record).
Audit	This collection can be thought of as a log file. It contains audit information about the commands executed, their results, warnings and objects changed.

6.2. Preparation of data for the database: instructions and examples

In this section, we provide concrete instructions for preparing data for inclusion in EIGE's database. These instructions are aimed at data providers who intend to provide macro-level statistics to EIGE. The provider's policies permitting, we would also be grateful to receive the microdata, so as to enable the computing of additional statistics at EIGE. No special preparation is needed for microdata, beyond the usual anonymisation, validation, cleaning, weighting and possibly imputation. **For microdata, only those sections of these instructions that deal with metadata are relevant.**

6.2.1. Technical formats

We prefer to receive the data in the SDMX format used by Eurostat, with one DSD file (containing data structure definitions) and one SDMX file (containing data) per data set ([52]). Likewise, our preferred format for metadata is ESMS. That said, we are happy to process data and metadata in any other format, as long as adequate documentation is provided and the logical structure of the data is amenable to being translated into the logical structure of our database, as described in the previous section.

It should be noted that SDMX is an international initiative that aims at standardising and modernising ('industrialising') the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

6.2.2 Steps to be followed

Preparation of data and metadata for EIGE's Gender Statistics Database consists of the following steps.

1. Define the division of data into data sets.
   Each data set consists of data that measure the same underlying concept (or a set of closely related underlying concepts) and can be represented as a multidimensional cube, as defined in Section 6.1.1 'Principles and building blocks' and also explained under step 2 below. Examples of data sets would be, for example, 'Employment rates by sex, age and level of educational achievement' and 'Lifetime prevalence of intimate partner violence (self-reported) by age of victim and type of violence'.

2. Define the structure of each data set.
   This amounts to defining the criteria (dimensions) and the categories (values) along each dimension. For example, a data set on national-level annual employment rates and levels by sex and 5-year age groups contains five criteria: 'country' (28 categories, one for each EU Member State), 'year' (each year being a category), 'unit of measurement' (two categories: 'levels (1 000s of people)' and 'rates (%)'); 'sex' (two categories: 'men' and 'women'); and 'age group' (each age group being a category). Each criterion and each category must be assigned both a name and a (short) code. For concrete examples of data set structures, please see Section 6.2.3 'Examples of data set structures'.

3. Compute the observation value for each point in the multidimensional cube defined by the criteria (dimensions) from step 2.

([52]) http://ec.europa.eu/eurostat/web/sdmx-infospace/welcome

In the example above, one such point would be the employment rate of women in the 30–35 age group in the United Kingdom in 2015.

4. Flag outlying, unreliable or otherwise 'special' values.
   For example, in the case of microdata, values based on subsamples of fewer than 50 respondents would be flagged.

5. Convert the data to SDMX format using the structure, values and flags defined/computed in the first four steps.

6. Perform a quality analysis according to the ESMS framework.
   For a detailed discussion of this step, please consult the sections on quality analysis earlier in this document. Please try to ensure that the quality analysis also addresses the gender-specific concerns outlined in this document.

7. Construct a metadata file conforming to the ESMS format.
   We discuss this briefly in Section 6.2.4 'Preparation of metadata in ESMS format'.

6.2.3. Examples of data set structures

From macrodata

Consider the Gender Equality Index scores, as displayed on the index website: http://eige.europa.eu/gender-equality-index. The top-level page on the website shows a polar chart of the overall scores for all EU Member States, while linked lower-level pages present the sub-scores of all Member States in each individual domain. We then transform these tables into a single data set for EIGE's database as follows.

First, we define the title of the data set, its code and a short description.

Data set name:	**Scores of the Gender Equality Index and its domains**
Data set code:	gei_core_scores
Description:	This data set shows the scores of EU Member States on the Gender Equality Index and its individual domains.

Next, we define the criteria and the categories.

CRITERIA:	
Name	**Code**
Geographical region	GEO
Time period	TIME
Domain	DOMAIN

CATEGORIES:		
Criterion code	**Category name**	**Category code**
GEO	<Names of countries>	<Two-letter codes of countries>
TIME	<Years>	<Years>
DOMAIN	OVERALL	Overall Gender Equality Index
	WORK	Work (domain score)
	MONEY	Money (domain score)
	KNOWLEDGE	Knowledge (domain score)
	TIME	Time (domain score)
	POWER	Power (domain score)
	HEALTH	Health (domain score)

This completes the data structure definition.

Note that this is not the only possible way that the data set could have been defined. It is possible to extend the data set to also include sub-domain scores and individual indicator scores (this is the choice that has actually been made in the database). It is also possible to take the opposite approach and to break the data set into multiple smaller data sets, one for each domain. The advantage of multiple smaller data sets is tractability and ease of handling, while the advantage of a single, bigger data set is the ability to present more data side by side in a single table or graph.

Directly from microdata

Consider the following question from Eurobarometer 73.2:

> QC3. Please tell me whether you consider each of the following forms of domestic violence against women to be very serious, fairly serious, not very serious or not at all serious.

	Very serious	Fairly serious	Not very serious	Not at all serious	Don't know (DK)
Psychological violence					
Physical violence					
Sexual violence					
Threats of violence					
Restricted freedom					

The first step in defining a macro-level data set from this question for inclusion in the database is deciding what statistics should be reported at the country level. Here we choose to report the frequency of each answer for each type of violence as percentage of the relevant population. We can now define the title of the data set and assign the data set a code and a short description.

Data set name:	**Do you consider each of the following forms of domestic violence against women to be serious? (%)**
Data set code:	EB73_2_howserious
Description:	The data show what percentage of all respondents have selected each answer (from 1 = 'very serious to 4 = 'not at all serious', or DK = 'don't know') when asked the question 'QC3. Please tell me whether you consider each of the following forms of domestic violence against women to be very serious, fairly serious, not very serious or not at all serious.'

Note how the description includes the full original question. This should generally be the case for all data sets constructed from survey data.

The next step is defining the criteria and categories corresponding to the various subquestions and answer categories.

CRITERIA:	
Name	**Code**
Answer	ANSWER
Form of violence	FORM

CATEGORIES:		
Criterion code	**Category name**	**Category code**
ANSWER	Very serious	PCT_VERY
	Fairly serious	PCT_FAIRLY
	Not very serious	PCT_NOT_VERY
	Not at all serious	PCT_NOT_AT_ALL
	Don't know	PCT_DK
FORM	Psychological violence	PSYCH
	Physical violence	PHYS
	Sexual violence	SEX
	Threats of violence	THREAT
	Restricted freedom	RESTR

Note how the category code for each category of the 'ANSWER' criterion is preceded by the prefix 'PCT_'. We use this as a convention to indicate that the data will be frequencies in percentages corresponding to each answer. We encourage providers to use similar conventions.

In the final step, we define criteria and categories corresponding to the breakdowns defining the subgroups of the overall population within which we will be computing the percentages giving each answer. In this example, let us use breakdowns only by country and by sex of respondent.

CRITERIA:	
Name	**Code**
Country	COUNTRY
Sex of respondent	SEX

CATEGORIES:		
Criterion code	**Category name**	**Category code**
COUNTRY	<Names of countries>	<Two-letter codes of countries>
SEX	Women	W
	Men	M
	Total	T

The final structure of the database is defined by the union of the criteria in the two tables above.

6.2.4. Preparation of metadata in ESMS format

All data should be accompanied by detailed metadata files adhering to the ESMS format or another format that follows a structure that can be converted to ESMS. ESMS 'aims at documenting methodologies, quality and the statistical production processes in general. It uses 21 high-level concepts, with a limited breakdown of sub-items, strictly derived from the list of cross domain concepts in the *SDMX Content Oriented Guidelines* (2009)' ([53]). For a detailed description of the structure, with specific instructions for each position, please consult the full ESMS definition in Section 8 'EURO-SDMX Metadata Structure 2.0'. In Table 5, we list the 21 top-level ESMS concepts.

Table 5. The top-level structure of the ESMS (version 2.0)

1. Contact	8. Release policy	15. Coherence and comparability
2. Metadata update	9. Frequency of dissemination	16. Cost and burden
3. Statistical presentation	10. Accessibility and clarity	17. Data revision
4. Unit of measure	11. Quality management	18. Statistical processing
5. Reference period	12. Relevance	19. Comment
6. Institutional mandate	13. Accuracy and reliability	
7. Confidentiality	14. Timeliness and punctuality	

([53]) See http://ec.europa.eu/eurostat/data/metadata/metadata-structure

7. Definitions of quality dimensions in Eurostat's Concepts and Definitions Database

The definitions in this document have been excerpted from Eurostat's Concepts and Definitions Database ([54]).

- **Relevance.**
  Definition. Degree to which statistical information meets the real or perceived needs of clients.
  Context. Relevance is concerned with whether the available information sheds light on the issues that are important to users. Assessing relevance is subjective and depends upon users' varying needs. EIGE's challenge is to weigh and balance the conflicting needs of current and potential users to produce statistics that satisfy the most important needs within given resource constraints. In assessing relevance, one approach is to gauge relevance directly, by polling users about the data. Indirect evidence of relevance may be found by ascertaining where there are processes in place to determine the uses of data and the views of their users or to use the data in-house for research and other analysis. Relevance refers to the processes for monitoring the relevance and practical usefulness of existing statistics in meeting users' needs and how these processes impact the development of statistical programmes.

  - **Relevance – completeness.**
    Definition. Extent to which all statistics that are needed are available.
    Context. The measurement of the availability of statistics normally refers to data sets and compares the required data set to the available one.

  - **Relevance – user needs.**
    Definition. Description of requirements with respect to the statistical output.
    Context. With respect to the statistical data to be provided, the main users (e.g. official authorities, the public or others) and user needs should be stated, e.g. official authorities with the needs for policy indicators, national users, etc.

  - **Relevance – user satisfaction.**
    Definition. Description of how well the disseminated statistics meet the expressed user needs.
    Context. In quality assurance frameworks this element indicates how the views and opinions of the users are collected. If user satisfaction surveys are conducted, the way users' views and opinions are collected should be described and the main results shown (in the form of a user satisfaction index if available); the date of the most recent user satisfaction survey should also be mentioned. Otherwise, any other indication or measure to determine user satisfaction might be used.

- **Accuracy.**
  Definition. Closeness of computations or estimates to the unknown exact or true values that the statistics were intended to measure.
  Context. The accuracy of statistical information is the degree to which the information correctly describes the phenomena. It is usually characterised in terms of error in statistical estimates and is often decomposed into bias (systematic error) and variance (random error) components. Accuracy can be expressed as either measures of accuracy (numerical results of the methods for assessing the accuracy of data) or qualitative assessment indicators. It may also be described in terms of the major sources of

([54]) http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL_GLOSSARY&StrNom=CODED2&StrLanguageCode=EN

error that potentially cause inaccuracy (e.g. coverage, sampling, non-response, response error). Accuracy is associated with the 'reliability' of the data, which is defined as the closeness of the initial estimated value to the subsequent estimated value.

- **Accuracy – overall.**
  Definition. Assessment of accuracy, linked to a certain data set or domain, which is summarising the various components into one single measure.
  Context. This metadata element is used to describe the main sources of random and systematic error in the statistical outputs and provide a summary assessment of all errors with special focus on the impact on key estimates. The bias assessment can be in quantitative or qualitative terms, or both. It should reflect the producer's best current understanding (sign and order of magnitude) including actions taken to reduce bias. Revision aspects should also be included here if considered relevant.

- **Sampling error.**
  Definition. Part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.
  Context. Sampling errors are distinct from errors due to imperfect selection, bias in response or estimation, errors of observation and recording, etc.
  For probability sampling, the random variation due to sampling can be calculated. For non-probability sampling, random errors cannot be calculated without reference to some kind of model. The totality of sampling errors in all possible samples of the same size generates the sampling distribution of the statistic which is being used to estimate the parent value.

- **Non-sampling error.**
  Definition. Error in sample estimates which cannot be attributed to sampling fluctuations.
  Context. Non-sampling errors may arise from many different sources such as defects in the sampling frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations, misunderstanding, bias, negligence or dishonesty on the part of the investigator or of the interviewer, mistakes at the stage of the processing of the data, etc.
  Non-sampling errors may be categorised as:
  — coverage errors (or frame errors) due to divergences between the target population and the frame population;
  — measurement errors occurring during data collection.
  — non-response errors caused by no data collected for a population unit or for some survey variables;
  — processing errors due to errors introduced during data entry, data editing and sometimes coding and imputation;
  — model assumption errors.

- **Reliability.**
  Definition. Closeness of the initial estimated value to the subsequent estimated value.

- **Timeliness.**
  Definition. Length of time between data availability and the event or phenomenon they describe.
  Context. Timeliness refers to the speed of data availability, whether for dissemination or for further processing, and it is measured with respect to the time lag between the end of the reference period and the release of data. Timeliness is a crucial element of data quality: adequate timeliness corresponds to a situation where policymakers can take informed decisions in time to achieve the targeted results. In quality assessment, timeliness is often associated with punctuality, which refers to the time lag between the release date of data and the target date announced in some official release calendar.

  Timeliness can be further broken down into 'Timeliness – output' and 'Timeliness – source data'.

'Timeliness – output' refers to the lapse of time between the end of a reference period (or a reference date) and the release of a version of the data: provisional, preliminary or final results. This reflects many factors, including some that are related to institutional arrangements, such as the preparation of accompanying commentary and printing. Usually, data are not released immediately at the end of the period they refer to, since data collection, data processing and data dissemination work needs to be performed.

'Timeliness – source data' refers to the time between the end of a reference period (or a reference date) and actual receipt of the data by the compiling agency. Compared to the parent concept – timeliness – this concept only covers the time period between the end of the reference period and the receipt of the data by the data compiling agency. This time period is determined by factors such as delays accommodating the institutional arrangements for data transmission.

- **Punctuality.**
  Definition. Time lag between the actual delivery of the data and the target date when it should have been delivered.
  Context. Punctuality may be calculated, for instance, with reference to target dates announced in an official release calendar, laid down by regulations or previously agreed among partners.

- **Coherence.**
  Definition. Adequacy of statistics to be reliably combined in different ways and for various uses.
  Context. When originating from different sources, and in particular from statistical surveys using different methodologies, statistics are often not completely identical, but show differences in results due to different collection methodology concepts, classifications and methodological standards. There are several areas where the assessment of coherence is regularly conducted: between provisional and final statistics, between annual and short-term statistics, between statistics from the same socioeconomic domain and between survey statistics and national accounts.
  The concept of coherence is closely related to the concept of comparability between statistical domains. Both coherence and comparability refer to a data set with respect to another. The difference between the two 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.
  In the Data Quality Assessment Framework of the International Monetary Fund, the term 'consistency' is used to indicate 'logical and numerical coherence'. In that framework, 'internal consistency' and 'intersectoral and cross-domain consistency' can be mapped to 'internal coherence' and 'cross-domain coherence' respectively.

  - **Coherence – cross-domain.**
    Definition. Extent to which statistics are reconcilable with those obtained through other data sources or statistical domains.
    Context. This metadata element is used to describe the differences in the statistical results calculated on the basis of different statistical domains, or surveys based on different methodologies (e.g. between annual and short-term statistics or between social statistics and national accounts).

    - **Coherence – sub-annual and annual statistics.**
      Definition. Extent to which statistics of different frequencies are reconcilable.Context. Coherence between sub-annual and annual statistical outputs is a natural expectation but the statistical processes producing them are often quite different. This metadata element is used to compare sub-annual and annual estimates and, eventually, describe reasons for lack of coherence between sub-annual and annual outputs.

- **Coherence – national accounts.**
  Definition. Extent to which statistics are reconcilable with national accounts.
  Context. This metadata element is used to report, where relevant, the results of comparisons with the national account framework and feedback from national accounts with respect to coherence and accuracy problems and should be a trigger for further investigation.

- **Coherence – internal.**
  Definition. Extent to which statistics are consistent within a given data set.
  Context. This metadata element is used to describe the differences in the statistical results calculated for the same statistical domain, based on stable or changing methodology (e.g. between provisional and final statistics or between different reference years showing a break in a series). Frequently, a group of statistics of a different type (in monetary value, in volume or constant price, price indicators, etc.) measure the same phenomenon using different methodologies. For instance, statistics on employment, depending on whether they result from employers' declarations or household surveys, do not lead to exactly the same results. However, there are often differences in the concepts used (*de jure* or de facto population, for instance), in the registration date, in the cif/fob ([55]) registration for external trade, etc. It is very important to check that these representations do not diverge too much in order to anticipate users' questions and to prepare corrective actions.

- **Comparability.**
  Definition. Extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.
  Context. Comparability aims at measuring the impact of differences in applied statistical concepts and definitions on the comparison of statistics between geographical areas, non-geographical dimensions or over time. Comparability of statistics, i.e. their usefulness in drawing comparisons and contrast among different populations, is a complex concept, difficult to assess in precise or absolute terms. In general terms, it means that statistics for different populations can be legitimately aggregated, compared and interpreted in relation to each other or against some common standard. Metadata must convey such information that will help any interested party in evaluating the comparability of the data, which is the result of a multitude of factors.
  In some quality assurance frameworks, e.g. the *European Statistics Code of Practice*, comparability is strictly associated with the coherence of statistics.

  - **Comparability – geographical.**
    Definition. Extent to which statistics are comparable between geographical areas.
    Context. Geographical comparability refers to the degree of comparability between similar survey results measuring the same phenomenon across geographical areas or regions. The surveys are in general conducted by different statistical agencies, referring to populations in different geographical areas, sometimes based on a harmonised methodology.

  - **Comparability – over time.**
    Definition. Extent to which statistics are comparable or reconcilable over time.
    Context. Comparability over time refers to the degree of comparability between the results of two or several surveys related to the same domain, carried out by the same statistical agency.

- **Accessibility.**
  Definition. Ease and conditions under which statistical information can be obtained.
  Context. Accessibility refers to the availability of statistical information to the user. It includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The

([55]) Cost, insurance and freight/free on board.

cost of the information may also be an aspect of accessibility for some users.

- **Dissemination format.**
  Definition. Media by which statistical data and metadata are disseminated.
  Context. This metadata element refers to the various means of dissemination used for making the data available to the public. It includes a description of the various formats available, including where and how to get the information (for instance paper, electronic publications, online databases).

  - **Dissemination format – news releases.**
    Definition. Regular or ad hoc press releases linked to the data.
    Context. This concept covers press releases or other kinds of similar releases linked to data or metadata.

  - **Dissemination format – publications.**
    Definition. Regular or ad hoc publications in which the data are made available to the public.
    Context. This metadata element provides references to the most important data dissemination carried out through paper or online publications, including a summary identification and information on the availability of the publication means.

  - **Dissemination format – online database.**
    Definition. Information about online databases in which the disseminated data can be accessed.
    Context. This metadata element provides a link to the online database where the data are available, with a summary identification of domain names as released on the website as well as the related access conditions.

  - **Dissemination format – microdata access.**
    Definition. Information on whether microdata are also disseminated.
    Context. This metadata element indicates whether microdata are also disseminated, e.g. to researchers. Access conditions should be described in short.

  - **Dissemination format – other.**
    Definition. References to the most important other data dissemination done.
    Context. Examples of other dissemination formats are analytical publications edited by policy users.
    This concept includes, as a subelement, 'supplementary data', i.e. any customised tabulation that can be provided to meet specific requests (including information on procedures for obtaining access to these data).

- **Documentation on methodology.**
  Definition. Descriptive text and references to methodological documents available.
  Context. 'Documentation on methodology' refers to the availability of documentation related to various aspects of the data, such as methodological documents, summary notes or papers covering concepts, scope, classifications and statistical techniques.

- **Quality management – quality documentation.**
  Definition. Documentation on procedures applied for quality management and quality assessment.
  Context. This metadata element is used to document the methods and standards for assessing data quality, based on standard quality criteria such as relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, comparability and coherence.

  - **Quality management – quality assurance.**
    Definition. Guidelines focusing on quality in general and dealing with the quality of statistical programmes, including measures for ensuring the efficient use of resources.
    Context. This metadata element refers to all the planned and systematic activities implemented that can be demonstrated to provide confidence that the data pro-

duction processes will fulfil the requirements for the statistical output. This includes the design of programmes for quality management, a description of the planning process, scheduling of work, frequency of planned updates and other organisational arrangements to support and maintain planning functions.

- **Quality management – quality assessment.**
  Definition. Overall evaluation of data quality, based on standard quality criteria.
  Context. The overall assessment of data quality may include the result of a scoring or grading process for quality. Scoring may be quantitative or qualitative.

- **Clarity.**
  Definition. Extent to which easily comprehensible metadata are available, where these metadata are necessary to give a full understanding of statistical data.
  Context. Clarity is sometimes referred to as 'interpretability'. It refers to the data information environment: whether data are accompanied by appropriate metadata, including information on their quality, and the extent to which additional assistance is provided to users by data providers.

- In the *European Statistics Code of Practice*, clarity is strictly associated with accessibility to form one single quality criteria: 'accessibility and clarity'.

8. EURO-SDMX Metadata Structure 2.0

The ESMS contains the description and representation of statistical metadata concepts to be used for documenting statistical data and for providing summary information useful for assessing data quality and the production process in general. The broad concepts used are compatible with the SDMX cross-domain concepts and with the common terminology as published within Eurostat's Concepts and Definitions Database.

The ESMS is addressed to the European Statistical System. It will be implemented at Eurostat and at national level: the application of the concepts and sub-concepts at European level and at national level is stated in the ESS guidelines.

The information to be entered is normally free text. However, in some cases, code lists will be used in the future: this is already indicated in the 'representation' column.

The ESMS allows the creation of different output files comprising information related to all the concepts listed or a subset of those concepts. These output files can be used for different purposes (data dissemination, quality reporting, etc.).

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
1.	**Contact**	CONTACT	Individual or organisational contact points for the data or metadata, including information on how to reach the contact points.		
1.1.	Contact organisation	CONTACT_ORGANISATION	Name of the organisation of the contact points for the data or metadata.	Text	The full name of your organisation.
1.2.	Contact organisation unit	ORGANISATION_UNIT	An addressable subdivision of an organisation.	Text	The name of the unit responsible for the metadata file (it can also include a unit number).
1.3.	Contact name	CONTACT_NAME	Name of the contact points for the data or metadata.	Text	The name of the person responsible for the statistical domain (first name and surname).
1.4.	Contact person function	CONTACT_FUNCT	Area of technical responsibility of the contact, such as 'methodology', 'database management' or 'dissemination'.	Text	The title of the person responsible for the statistical domain (this title can contain the precise area of responsibility, such as 'methodologist 'or 'database manager').
1.5.	Contact mail address	CONTACT_MAIL	Postal address of the contact points for the data or metadata.	Text	The postal address of the person responsible for the statistical domain.
1.6.	Contact email address	CONTACT_EMAIL	Email address of the contact points for the data or metadata.	Email	The email address of the person responsible for the statistical domain (this can be an individual email address or a functional mailbox).

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
1.7.	Contact phone number	CONTACT_PHONE	Telephone number of the contact points for the data or metadata.	Telephone	The phone number of the person responsible for the statistical domain.
1.8.	Contact fax number	CONTACT_FAX	Fax number of the contact points for the data or metadata.	Fax	The fax number of the person responsible for the statistical domain.
2.	**Metadata update**	META_UPDATE	The date on which the metadata element was inserted or modified in the database.		
2.1.	Metadata last certified	META_CERTIFIED	Date of the latest certification provided by the domain manager to confirm that the metadata posted are still up to date, even if the content has not been amended.	Date	The date of the latest certification of this metadata file in order to confirm that the metadata file produced is still up to date. Such a certification can also be done if the content of the metadata file has not been amended.
2.2.	Metadata last posted	META_POSTED	Date of the latest dissemination of the metadata.	Date	The date when this metadata file is disseminated will normally be inserted automatically by the reference metadata production system (for Eurostat: by the ESS Metadata Handler).
2.3.	Metadata last update	META_LAST_UPDATE	Date of last update of the content of the metadata.	Date	The date when this metadata file is updated will normally also be inserted by the reference metadata production system (for Eurostat: by the ESS Metadata Handler).
3.	**Statistical presentation**	STAT_PRES			
3.1.	Data description	DATA_DESCR	Main characteristics of the data set described in an easily understandable manner, referring to the data and indicators disseminated.	Text	Describe the main characteristics of the data set in an easily understandable manner, referring to the main data and indicators disseminated. This short description should be understood immediately and easily by the users.
3.2.	Classification system	CLASS_SYSTEM	Arrangement or division of objects into groups based on characteristics which the objects have in common.	Text	List all classifications which are used for the data set produced (with their detailed names).
3.3.	Sector coverage	COVERAGE_SECTOR	Main economic or other sectors covered by the statistics.	Text	List the main economic or other sectors covered by the data set produced, also adding the size classes used.
3.4.	Statistical concepts and definitions	STAT_CONC_DEF	Statistical characteristics of statistical observations.	Text	Describe in short the main statistical variables provided. The definition and types of variables provided should be listed, together with any information on discrepancies from the ESS/ international standards.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
3.5.	Statistical unit	STAT_UNIT	Entity for which information is sought and for which statistics are ultimately compiled.	Text	List the basic units of statistical observation for which data are provided. These observation units (e.g. the enterprise, the local unit, private households, etc.) can be different from the reporting units used in the underlying statistical surveys.
3.6.	Statistical population	STAT_POP	Total membership or population or 'universe' of a defined class of people, objects or events.	Text	Describe the target statistical population (one or more) which the data set refers to, i.e. the population about which information is to be sought.
3.7.	Reference area	REF_AREA	Country or geographic area to which the measured statistical phenomenon relates.	Text/Coded (code list: CL_REF_AREA)	**At European level.** The geographical area covered by the data set disseminated (e.g. EU Member States, EU regions, Japan, United States etc., as well as aggregates such as EU-28, European Economic Area). **At national level.** The country, regions and aggregates covered by the data set disseminated.
3.8.	Time coverage	COVERAGE_TIME	Length of time for which data are available.	Text	The time periods covered by the data set should be described (i.e. the length of time for which data set is disseminated, e.g. from 1985 to 2006 for certain annual data).
3.9.	Base period	BASE_PER	Period of time used as the base of an index number, or to which a constant series refers.	Text	The period of time used as the base of an index number or to which a time series refers should be described (e.g. base year 2000 for certain annual data).
4.	**Unit of measure**	UNIT_MEASURE	Unit in which the data values are measured.	Text/Coded (code list: CL_UNIT_ MEASURE)	The units of measures used for the data set disseminated should be listed (units of measures are for example Euro, % or number of persons). The exact use of magnitude (e.g. thousand, million) should also be added.
5.	**Reference period**	REF_PERIOD	Period of time or point in time to which the measured observation is intended to refer.	Text	Statistical variables refer to specific time periods, which can be a specific day or a specific period (e.g. a month, a fiscal year, a calendar year or several calendar years). When there is a mismatch between the target and the actual reference period, for instance when data are not available for the target reference period, the difference should also be highlighted.
6.	**Institutional mandate**	INST_MANDATE	Set of rules or other formal set of instructions assigning responsibility as well as authority to an organisation for the collection, processing and dissemination of statistics.		A law or other formal set of instructions that assign responsibility as well as authority to an agency for the collection, processing and dissemination of statistics (including arrangements or procedures to facilitate data sharing and exchange between data-producing agencies).

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
6.1.	Legal acts and other agreements	INST_MAN_LA_OA	Legal acts or other formal or informal agreements that assign responsibility as well as authority to an agency for the collection, processing and dissemination of statistics.	Text	**At European level.** The legal basis or other agreement creating the reporting requirement should be listed (e.g. the EU legal act or agreement or the 5-year-programme related to the European Statistical System). **At national level.** National legal acts or other reporting agreements should be mentioned (including the implementation of EU directives).
6.2.	Data sharing	INST_MAN_SHAR	Arrangements or procedures for data sharing and coordination between data-producing agencies.	Text	**At European level only.** Agreements related to data sharing and exchange between international data-producing agencies should be described (e.g. a Eurostat data collection or data production which is carried out in partnership with the Organisation for Economic Co-operation and Development, the United Nations, etc.).
7.	**Confidentiality**	CONF	A property of data indicating the extent to which their unauthorised disclosure could be prejudicial or harmful to the interests of the source or other relevant parties.		The legislation (or any other formal provision) related to statistical confidentiality applied to the data set in question, as well as the actual confidentiality data treatment done (also with regard to the aggregated data disseminated), should be described.
7.1.	Confidentiality – policy	CONF_POLICY	Legislative measures or other formal procedures which prevent unauthorised disclosure of data that identify a person or economic entity either directly or indirectly.	Text	The EU and national legislations related to statistical confidentiality should be described.
7.2.	Confidentiality – data treatment	CONF_DATA_TR	Rules applied for treating the data set to ensure statistical confidentiality and prevent unauthorised disclosure.	Text	The rules applied for treating the data set with regard to statistical confidentiality should be described (e.g. controlled rounding, cell suppression, aggregation of disclosive information, aggregation rules on aggregated confidential data, primary confidentiality with regard to single data values, etc.). Main reference: *Handbook on Statistical Disclosure Control* ([56]).
8.	**Release policy**	REL_POLICY	Rules for disseminating statistical data to interested parties.		Rules for disseminating statistical data to all interested parties.

([56]) Hundepool, A. et al., 2010.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
8.1.	Release calendar	REL_CAL_POLICY	Schedule of statistical release dates.	Text	The policy regarding the release of statistics according to a preannounced schedule should be described. It should also be mentioned if a release calendar for the data set in question exists and if this calendar is publicly accessible.
8.2.	Release calendar access	REL_CAL_ACCESS	Access to the release calendar information.	Text	The link or reference to the release calendar should be given.
8.3.	User access	REL_POL_US_AC	Policy for release of data to users, scope of dissemination (e.g. to the public, to selected users), how users are informed that the data are being released and whether the policy determines the dissemination of statistical data to all users.	Text	The policy for data release to users should be described. This includes the scope of dissemination (e.g. to the public, to selected users), how users are informed that the data is being released and whether the release policy determines the dissemination of statistical data to all users at the same time. **For Eurostat only.** Reference is also made to the impartiality protocol linked to the *European Statistics Code of Practice*, principle 6, where the person responsible for the statistical domain should state all kinds of pre-releases.
9.	**Frequency of dissemination**	FREQ_DISS	Time interval at which the statistics are disseminated over a given time period.	Text / Coded (code list: CL_FREQ)	The frequency with which the data is disseminated (e.g. monthly, quarterly, annually) should be stated. The frequency can also be expressed using the codes released in the harmonised code list available for the European Statistical System.
10.	**Accessibility and clarity**	ACCESSIBILITY_CLARITY	Conditions and modalities by which users can obtain, use and interpret data.		The various means of dissemination used for making the data set available to users should be described (including the various dissemination formats available as well as their accessibility).
10.1.	News release	NEWS_REL	Regular or ad hoc press releases linked to the data.	Text	Regular or ad hoc press releases linked to the data set in question should be described.
10.2.	Publications	PUBLICATIONS	Regular or ad hoc publications in which the data are made available to the public.	Text	The publications using the data set in question should be described in short. Quality indicators
10.3.	Online database	ONLINE_DB	Information about online databases in which the disseminated data can be accessed.	Text	The online database available for the data set in question should be described. This includes the domain names as released on the website. Quality indicators
10.4.	Microdata access	MICRO_DAT_ACC	Information on whether microdata are also disseminated.	Text	Describe if and how the data set is accessible as microdata (e.g. for researchers). Also the microdata anonymisation rules should be described briefly.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
10.5.	Other	DISS_OTHER	References to the most important other data dissemination carried out.	Text	The most important other data dissemination means should be described (e.g. within other publications, policy papers, etc.).
10.6.	Documentation on methodology	DOC_METHOD	Descriptive text and references to methodological documents available.	Text	Describe the availability of national reference metadata files, important methodological papers, summary documents or other important handbooks. Quality indicators
10.7.	Quality documentation	QUALITY_DOC	Documentation on procedures applied for quality management and quality assessment.	Text	Describe the availability of quality reports and studies. For Eurostat: The person responsible for the statistical domain should also describe the availability of national quality reports.
11.	**Quality management**	QUALITY_MGMNT	Systems and frameworks in place within an organisation to manage the quality of statistical products and processes.		Describe briefly the quality management system used in the organisation (EFQM, ISO series, etc.) and provide a hyperlink to the *European Statistics Code of Practice*.
11.1.	Quality assurance	QUALITY_ASSURE	All systematic activities implemented that can be demonstrated to provide confidence that the processes will fulfil the requirements for the statistical output.	Text	Provide a hyperlink to the general quality assurance framework (or similar) and a brief description of how it is implemented for domain-specific quality assurance activities (the use of best practices, quality reviews, self-assessments, compliance monitoring, etc.).
11.2.	Quality assessment	QUALITY_ASSMNT	Overall assessment of data quality, based on standard quality criteria.	Text	The standard quality criteria are provided in concepts 12–15. A qualitative assessment of the overall quality of the statistical outputs should be provided by summarising the main strengths and possible quality deficiencies. Any trade-offs between quality aspects can be mentioned, as well as any planned quality improvements. Main reference: *ESS Handbook for Quality Reports* (Eurostat, 2015).
12.	**Relevance**	RELEVANCE	Degree to which statistical information meets current and potential needs of users.		
12.1.	User needs	USER_NEEDS	Description of users and their respective needs with respect to the statistical data.	Text	Provide a classification of users with some indication of their importance, an indication of the uses they will make of the statistical outputs and any users and uses given special considerations. Unmet user needs and the reasons for not meeting them should also be included.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
12.2.	User satisfaction	USER_SAT	Measures to determine user satisfaction.	Text	Describe how the views and opinions of the users are collected. In addition, the main results regarding user satisfaction should be shown (in the form of a user satisfaction index if available) as well as the date of the most recent user satisfaction survey. Quality indicators
12.3.	Completeness	COMPLETENESS	The extent to which all statistics that are needed are available.	Text	Provide information on completeness compared with relevant regulations/guidelines. **Applicable for Eurostat.** Explain if any Member States are not producing the statistics in question. Quality indicators
13.	**Accuracy and reliability**	ACCURACY	Closeness of computations or estimates to the unknown exact or true values that the statistics were intended to measure. Reliability of the data, defined as the closeness of the initial estimated value to the subsequent estimated value.		The accuracy of statistical outputs in the general statistical sense is the degree of closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Reliability refers to the closeness of the initial estimated value to the subsequent estimated value.
13.1.	Overall accuracy	ACCURACY_OVERALL	Assessment of accuracy, linked to a certain data set or domain, which summarises the various components.	Text	Provide a summary of the main sources of error and an assessment of the potential for bias (sign and order of magnitude) for each key indicator in quantitative or qualitative terms.
13.2.	Sampling error	SAMPLING_ERR	That part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a subset of the population is enumerated.	Text	If probability sampling is used, estimates of the accuracy should be provided, normally in the form of cv's, standard errors or confidence intervals. It should be stated if adjustments for non-response, misclassifications and other uncertainty sources such as outlier treatment are included. If non-probability sampling is used, the person responsible for the statistical domain should provide estimates of the accuracy, a motivation for the invoked model for this estimation and a brief discussion of sampling bias. Quality indicators

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
13.3.	Non-sampling error	NONSAMPLING_ERR	Error in survey estimates which cannot be attributed to sampling fluctuations.	Text	Provide an assessment, preferably quantitative, on the non-sampling errors and the bias risks associated with: • over-coverage, under-coverage and multiple listings; • the survey instrument, respondent and interviewer where relevant; • unit (non-)response including causes for non-response and measures to reduce it; • item non-response for key variables; • data editing, coding and imputation where relevant; • specific models used in estimation. Actions undertaken to reduce the different types of errors could also be provided. Quality indicators
14.	**Timeliness and punctuality**	TIMELINESS_PUNCT			
14.1.	Timeliness	TIMELINESS	Length of time between data availability and the event or phenomenon they describe.	Text	Provide, for annual or more frequent releases, the average production time for each release of data. **Applicable for Eurostat.** For national data deliveries, the agreed time frame for deliveries should be included as well as the achieved dates for deliveries during a past period. Quality indicators
14.2.	Punctuality	PUNCTUALITY	Time lag between the actual delivery of the data and the target date when it should have been delivered.	Text	Provide, for annual or more frequent releases: • the percentage of releases delivered on time, based on scheduled release dates; • the reasons for non-punctual releases. Quality indicators
15.	**Coherence and comparability**	COHER_COMPAR	Adequacy of statistics to be reliably combined in different ways and for various uses and the extent to which differences between statistics can be attributed to differences between the true values of the statistical characteristics.		Coherence measures the adequacy of the statistics to be combined in different ways and for various uses. Comparability is a measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas or over time.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
15.1.	Comparability – geographical	COMPAR_GEO	Extent to which statistics are comparable between geographical areas.	Text	Describe any problems of comparability between countries or regions. The reasons for the problems should be described as well as the order of magnitude of the effects of the main sources of errors. Information on discrepancies from the ESS/ international concepts and definitions should be included. Asymmetries for statistical mirror flows should also be described. Quality indicators
15.2.	Comparability – over time	COMPAR_TIME	Extent to which statistics are comparable or reconcilable over time.	Text	Provide information on the length of comparable time series, reference periods at which series breaks occur, the reasons for the breaks and the ways they are treated. Quality indicators
15.3.	Coherence – cross domain	COHER_X_DOM	Extent to which statistics are reconcilable with those obtained through other data sources or statistical domains.	Text	Describe the differences between the statistical outputs in question and other related statistical outputs. The order of magnitude of the effects of the differences should also be assessed.
	Coherence – internal	COHER_INTERNAL	Extent to which statistics are consistent within a given data set.	Text	Describe statistical outputs within the data set in question that are not consistent and the reasons for publishing such results.
16.	**Cost and burden**	COST_BURDEN	Cost associated with the collection and production of a statistical product and burden on respondents.	Text	Provide a summary of costs for production of statistical data and of the burden on respondents (in general measured in time used). Objectives/actions concerning burden reduction could also be provided. Main references: *Handbook for Monitoring and Evaluating Business Survey*; *Response Burdens* (Eurostat, 2007); *International Standard Cost Model Manual* (SCM network). Quality indicators
17.	**Data revision**	DATA_REV	Any change in a value of a statistic released to the public.		
17.1.	Data revision – policy	REV_POLICY	Policy aimed at ensuring the transparency of disseminated data, whereby preliminary data are compiled that are later revised.	Text	Describe the general revision policy adopted for the organisation and the data disseminated.

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
17.2.	Data revision – practice	REV_PRACTICE	Information on the data revision practice.	Text	Describe major scheduled revisions as well as their expected average size. As far as necessary minor revisions should also be described. **At European level.** A reporting template is provided in this respect. Quality indicators
18.	**Statistical processing**	STAT_PROCESS			
18.1.	Source data	SOURCE_TYPE	Characteristics and components of the raw statistical data used for compiling statistical aggregates.	Text	Indicate if the data set is based on a survey or on administrative data sources. If sample surveys are used, some sample characteristics should also be given (gross and net sample size, type of sampling design, reporting domain, etc.). If administrative registers are used, the description of registers should be given (source, year, primary purpose, potential deficiencies, etc.).
18.2.	Frequency of data collection	FREQ_COLL	Frequency with which the source data are collected.	Text/Coded (code list: CL_FREQ)	Indicate the frequency of data collection (e.g. monthly, quarterly, annually, continuously). The frequency can also be expressed using the codes released in the harmonised code list available for the European Statistical System.
18.3.	Data collection	COLL_METHOD	Systematic process of gathering data for official statistics.	Text	Describe the method used to gather data from respondents (e.g. postal survey, CAPI, online survey, etc.). Some additional information on questionnaire design and testing, interviewer training, methods used to monitor non-response, etc. should be provided here. Reference: *The Handbook of Recommended Practices for Questionnaire Development and Testing Methods in the ESS* (Eurostat, 2006).
18.4.	Data validation	DATA_VALIDATION	Process of monitoring the results of data compilation and ensuring the quality of statistical results.	Text	Describe the procedures for checking and validating the source data and how the results of these validations are monitored and used.
18.5.	Data compilation	DATA_COMP	Operations performed on data to derive new information according to a given set of rules.	Text	Describe the data compilation process (e.g. data editing, imputation, weighting, adjustment for non-response, calibration, model used, etc.). Main reference: *Survey Sampling Reference Guidelines – Introduction to sample design and estimation techniques* (Eurostat, 2008).

	Concept Name	Concept code	Descriptions	Representation	ESS guidelines
18.6.	Adjustment	ADJUSTMENT	The set of procedures employed to modify statistical data to enable it to conform to national or international standards or to address data quality differences when compiling specific data sets.	Text	Describe the statistical procedures used for adjusting the data (such as seasonal adjustment methods, time series decomposition or other similar methods). Main reference: *ESS Guidelines on Seasonal Adjustment* (2008).
19.	**Comment**	COMMENT_DSET	Supplementary descriptive text which can be attached to data or metadata.	Text	

References

Bank of England Statistics and Regulatory Data Division (2014), *Data Quality Framework*, Bank of England, retrieved from https://www.bankofengland.co.uk/-/media/boe/files/statistics/data-quality-framework.pdf

Canadian International Development Agency (1997), *Guide to Gender-Sensitive Indicators*, CIDA, Canada, retrieved from http://eugender.itcilo.org/toolkit/online/story_content/external_files/TA_Edu_CIDA.pdf

Dippo, C. S. and Sundgren, B. (2000), *The Role of Metadata in Statistics,* paper presented at the International Conference on Establishment Surveys II, Buffalo, New York, available online from the U.S. Bureau of Labour Statistics at https://ww2.amstat.org/meetings/ices/2000/proceedings/S57.pdf

European Institute for Gender Equality (2017), *Gender Equality Index 2017 – Methodological report*, Publications Office of the European Union, Luxembourg, available at https://eige.europa.eu/publications/gender-equality-index-2017-methodological-report

Eurostat (2015), *ESS Handbook for Quality Reports*, Publications Office of the European Union, Luxembourg, retrieved from http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf

Eurostat (2018), *European Statistics Code of Practice*, Publications Office of the European Union, Luxembourg, retrieved from http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice

Eurostat (2019), *Quality Assurance Framework of the European Statistical System*, Version 2.0, retrieved from https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646

Hedman, B., Perucci, F. and Sundström P. (1996), *Engendering Statistics – A tool for change*, Statistics Sweden, available at https://www.scb.se/contentassets/886d78607f724c3aaf0d0a72188ff91c/engendering-statistics-a-tool-for-change.pdf

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Naylor, J., Schulte Nordholt, E., Seri, G. and De Wolf, P-P. (2010), Handbook on Statistical Disclosure Control, ESSnet, https://ec.europa.eu/eurostat/cros/system/files/SDC_Handbook.pdf

United Nations (1996), *Report of the Fourth World Conference on Women*, (The Platform for Action), Beijing 4–15 September 1995, United Nations, New York.

United Nations (2012), *Human Rights Indicators – A guide to measurement and implementation,* available at http://www.ohchr.org/Documents/Publications/Human_rights_indicators_en.pdf

United Nations Economic Commission for Europe and World Bank Institute (2010), *Developing Gender Statistics: A practical tool*, United Nations, Geneva, retrieved from http://www.unece.org/fileadmin/DAM/stats/publications/Developing_Gender_Statistics.pdf

United Nations Statistical Division (2013), *Gender Statistics Manual: Integrating a gender perspective into statistics*, United Nations, New York, retrieved from https://unstats.un.org/unsd/demographic/standmeth/handbooks/05323 Integrating a Gender Perspective into Statistics Web Final.pdf

GETTING IN TOUCH WITH THE EU

IN PERSON
All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

ON THE PHONE OR BY EMAIL
Europe Direct is a service that answers your questions about the European Union.

You can contact this service:
— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
— at the following standard number: +32 22999696, or
— by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

ONLINE
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU PUBLICATIONS
You can download or order free and priced EU publications from: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU LAW AND RELATED DOCUMENTS
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU
The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to data sets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

MH-02-20-966-EN-N

www.eige.europa.eu

Publications Office
of the European Union