

TURKISH STATISTICAL INSTITUTE

HANDBOOK FOR INSTITUTIONAL QUALITY REPORTS

SAMPLING AND ANALYSIS TECHNIQUES DEPARTMENT

DECEMBER, 2015

1. PURPOSE AND COVERAGE OF SURVEY

Turkish Statistical Institute is responsible for providing statistics to the decision makers, researchers and users according to the international standards in line with principles of reliability, consistency, impartiality, statistical confidentiality, timeliness and transparency. The basic principles are ensuring the accuracy of official statistics, presenting the data to all users on the same time and on impartial base, respecting the principles of confidentiality and protecting the rights of the public to access the information.

Turkstat aims to achieve the targeted level of quality at minimum cost and to produce accurate and precise statistics in line with European Statistics Code of Practice (COP)-Quality Components. Accordingly, study on institutional quality reports has been started in compliance with the European Union in order to improve Turkstat's strategic plan and quality policy by taking the views of all departments. European Statistics Code of Practice-Quality Components and all statistics disseminated by Turkstat were taken into consideration while structuring the context of institutional quality report; the principles and components of Turkstat's compliance with COP were determined.

A programmed period in Turkish Statistical System begun with the "2007-2011 Official Statistics Program (OSP)". TurkStat aims at determining the basic principles and standards for the production and dissemination of official statistics and producing reliable, timely, transparent and impartial data required at both the national and international level with OSP. Therefore, as the producer and coordinator of official statistics, TurkStat decided to prepare institutional quality reports compliance with European Union at both national and international level.

To evaluate the compliance of the TurkStat selected principles of European Statistics Code of Practice, to provide an external observation and to support the further improvement of TurkStat, the light peer review (LPR) was initiated by EUROSTAT at the request of TurkStat. According to activity report prepared after LPR, subject of "Preparation of Institutional Quality Report" has come up again. This subject has been stated within the context of Strategic Plan 2012-2016 in Article SH 1.1 and in order to produce statistics which are qualified in consistent with the international standards, the quality report has been decided to prepare by TurkStat. Meeting of "Preparation of Institutional Quality Report" held on 3 April 2012 and the framework of quality report has been formed according to related performance indicators. This was presented to President and after the meeting 5 statistics headlines were chosen as pilot for application.

The quality report contains all statistical processes and outputs. Statistical process refers to sampling survey, census, usage of administrative records, forming economic indexes and statistical data collection. The concept of quality report isn't finite. In other words; it has continuous. Therefore, the quality report provides a thorough explanation concerning the whole statistical process from the production of data till the presentation of data. Moreover, it contains the necessary proposals and concerning reasons for the development of quality.

2. THE EUROPEAN STATISTICS CODE OF PRACTISE AND QUALITY COMPONENTS

The European Statistics Code of Practice was elaborated and adopted by the Statistical Programme Committee on 24 February 2005. These principles were defined in coherence with the Treaty establishing the European Community, and in particular Article 285(2) thereof, with the Council regulation (EC) No 322/97 of 17 February 1997 on Community Statistics, and with the Fundamental Principles of Official Statistics adopted by the United Nations Statistical Commission on 14 April 1994.

The principles that need to be considered in order to improve “trust” and “the credibility standards of the statistics” on users of production and organization of the official statistics in official statistics production processes (collection, validation, analysis and dissemination) must be identified. The European Statistics Code of Practice is based on 15 Principles. These principles are:

Principle 1. Professional Independence: The professional independence of the institutions and organizations producing official statistics from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, ensures the credibility of official statistics.

Principle 2. Mandate for data collection: The institutions and organizations producing official statistics must have a clear legal mandate to collect information for Turkish Statistical System purposes.

Principle 3. Adequacy of resources: The resources allocated for the production and distribution of official statistics should be sufficient to meet Turkish Statistical System requirements.

Principle 4. Commitment to quality: The quality of statistical data should be regularly monitored and reported using statistical techniques.

Principle 5. Statistical confidentiality: The institutions and organizations producing official statistics have to take every precaution against unlawful access to or use of confidential data. The confidential data compiled, processed and preserved for the production of official statistics cannot be delivered to any administrative, judiciary or military authority or person, cannot be used for purposes other than statistics or as an instrument of proof.

Data confidentiality and security procedures and guidelines are regulated by the “Regulation of Procedure and Principles of Data Confidentiality and Confidential Data Security in Official Statistics” published in the Official Gazette No. 26 204 in 20/06/2006.

Principle 6. Impartiality and objectivity: The institutions and organizations producing official statistics produce and disseminate statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably. Data providers must notice which data to be collected, the reasons and the purpose of the data collection, the statistical methods and the essence of the data to be published.

Principle 7. Sound methodology: Standards and methods used for the collection, processing and dissemination of official statistics follows European and other international standards exactly.

Principle 8. Appropriate statistical procedures: All applied statistical processes from data compilation to validation, must form the basis of the production of qualified statistics.

Principle 9. Non-excessive Burden on Respondents: Resources must be used effectively while producing and disseminating official statistics. Response burden on data providers (households, enterprises, administrations and other respondents) must be monitored and set targets accordingly. Monitoring the user's needs and expectations, appropriate dissemination forms for different user groups must be developed.

Principle 10. Cost Effectiveness: Resources must be used effectively.

Principle 11. Relevance: Official statistics must explain their topic at the national level and have scope to meet users' need. Official statistics must be published with their definitions and classifications.

Principle 12. Accuracy and Reliability: Official statistics must reflect the current situation and the changes accurately and reliably according to their topic and scope.

Principle 13. Timeliness and Punctuality: Official statistics must be produced timely within the current database and disseminated in accordance with the National Data Release Calendar.

Principle 14. Coherence and Comparability: Official statistics must be consistent internally, over time and comparable between regions and countries; it should be possible to combine and make joint use of related data from different sources.

Principle 15. Accessibility and Clarity: Official statistics must be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance. Dissemination services must use easily accessible tools (print media, digital media, etc.). The main aim must be disseminating data with its metadata via web free of charge.

3. EUROPEAN STATISTICS PRODUCT QUALITY COMPONENTS AND QUALITY INDICATORS

The product quality components improved by EUROSTAT and also recommended for inclusion in National Statistics Institute's quality reports and quality indicators that should be produced for these quality components are given below.

Relevance: Relevance is an attribute of statistics measuring the degree to which statistical information meets needs of the users. Therefore, it is necessary to take into account the basic needs of the users. 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 the varying needs of users.

According to EUROSTAT, the quality indicator that has to be in the quality report for relevance is “*Data Completeness Rate*”.

Accuracy: In the general statistical sense, accuracy denotes the closeness of estimations to the unknown true values. The difference between them is defined as an error. Two types of error are computed. These are:

1. Sampling error,
2. Non-sampling error.

Sampling Error

Sampling error is the error that arises only in research-based sampling. It does not occur in researches based on census. Calculating “Coefficient of Variation” and “Confidence Interval” as a traditional quality indicators is suggested by EUROSTAT.

Non-Sampling Error

Non-sampling error is defined as errors that may arise during survey, except sampling error. Contrary to sampling error, non-sampling error may occur in survey based on both sampling and census.

Quality indicators recommended for inclusion in quality reports by EUROSTAT for non-sampling errors are given below.

- *Over-coverage Rate,*
- *Unit Non-response Rate,*
- *Item Non-response Rate,*
- *Imputation Rate,*
- *Common Units Rate,*
- *Data Revision- Average Size.*

Timeliness and Punctuality: Timeliness of information reflects the length of time between its availability and the reference date. Punctuality refers to closeness between the release date of data and the target date.

Quality indicators recommended for inclusion in quality reports by EUROSTAT for timeliness and punctuality are given below.

- *Time lag - first results,*
- *Time lag - final results,*
- *Punctuality - delivery and publication.*

Accessibility and Clarity: Accessibility and clarity, one of the components of statistical quality, refer to the simplicity and ease, the conditions and modalities by which users can access, use and interpret statistics, with the appropriate supporting information and assistance, producing user-centered data in an acceptable time.

Accessibility refers to the physical conditions under which users can obtain data: where to go, how to order, delivery time, clear, availability of micro or macro data, various formats, etc.

Clarity refers to the data's information environment whether data are accompanied with appropriate documentation and metadata, illustrations, whether information on their quality is also available.

Quality indicators recommended for inclusion in quality reports by EUROSTAT for accessibility and clarity are given below.

- Number of metadata – consultations
- Number of data tables – consultations
- Metadata completeness – rate

Coherence and Comparability: Coherence refers to the consistency of statistics which are compiled on the basis of common standards with respect to scope, definitions, units and classifications in the different surveys and sources.

Comparability refers to the comparability of statistics between geographical areas, definitions or over time.

Quality indicators recommended for inclusion in quality reports by EUROSTAT for coherence and comparability are given below.

- Length of comparable time series
- Asymmetry for mirror flows statistics – coefficient

4. APPLICATION PROCESS AND INSTITUTIONAL QUALITY REPORT FORM

Institutional Quality Report was created by taking views of departments in the institution. This form contains information about statistics published by departments according to European Commission Quality Principles and indicators expected to be computed with respect to these principles.

4.1. PRINCIPLES OF FILLING THE INSTITUTIONAL QUALITY REPORT

Institutional Quality Report formed by Sampling and Analysis Techniques Department is filled for annual statistics. Explanations about how to fill the Institutional Quality Report are given below.

1- Subject of the Statistics

This field is filled by implementation unit. Subject of the statistics should be written clearly.

2- Contact Organization Department

This field is filled by implementation unit. Name of the implementation department is written to this field.

3- Contact Organization Unit

This field is filled by implementation unit. Name of the related organization unit under implementation department is written to this field.

4- Contact Name

This field is filled by implementation unit. Name-Surname, phone and fax number of person(s) who conduct this study or prepare this report is written exactly to this field.

5- Objective of Statistics

This field is filled by implementation unit. The aim and target of the survey is written clearly to this field. (Detection of level of poverty in household budget survey or minimum wage)

6- Confidentiality Policy

This field is filled by implementation unit. Census and survey legislation special to study is given in this field. It is filled by stating related articles of the Statistical Law of Turkey number 5429 of 10/11/2005. For administrative records, information about accessibility to data policy stated in institutions' own legislation is written.

Confidential Data: If the number of the statistical unit in any cell of the data table formed by aggregating the individual data is less than three or one or two of the statistical units are dominant even if the number of units is three or more, the data in the concerned cell is considered confidential. Confidential data can be published only as combined with other data so as not to allow any direct or indirect identification. The institutions and organizations producing official statistics have to take every/all kind of precaution against unlawful access to or use of confidential data. The confidential data compiled, processed and preserved for the production of official statistics cannot be delivered to any administrative, judiciary or military authority or person, cannot be used for purposes other than statistics or as an instrument of proof. Confidential data can be accessed only by the ones involved in the production of official statistics, to the extent that they need for performing their duties properly. Civil servants and other staff in charge of compiling and processing these data are obliged to comply with this rule. This obligation continues after the related personnel leave their duties and posts. Civil servants violating these bans and limitations shall be punished according to Article 258 of the Turkish Penal Code no. 5237. Data or information obtained from sources that are open to all people shall not be deemed confidential. Data confidentiality ceases when

a statistical unit gives written approval for the revelation of confidential data concerning itself.

The institutions and organizations producing official statistics have to take every/all kind of precaution against unlawful access to or use of confidential data. The confidential data compiled, processed and preserved for the production of official statistics cannot be delivered to any administrative, judiciary or military authority or person, cannot be used for purposes other than statistics or as an instrument of proof. Data confidentiality and security procedures and guidelines are regulated by the "Regulation of Procedure and Principles of Data Confidentiality and Confidential Data Security in Official Statistics" published in the Official Gazette No. 26 204 in 20/06/2006.

7- Data Source(s)

This field is filled by implementation unit. Data sources used during survey such as researches (all of the researches used), administrative sources, etc. is specified separately.

For example, Statistics of Plant Production are compiled through district and province directorates of Ministry of Food, Agriculture and Livestock. These data are based on available administrative records of district directorates and also agricultural data of their region.

For statistics produced based on more than one data source, computations and explanations for indicators wanted to be in the report should be repeated according to each of these data sources and that information should be given separately. Computations and explanations for indicators should be repeated according to each of data sources, except for indicators listed below.

- Data revision - average size
- Timeliness - first results
- Timeliness - final results
- Punctuality - delivery and publication
- Number of metadata – consultations
- Number of data tables – consultations
- Dissemination format
- Length of comparable time series

For example, Red Meat Production Statistics have been collected from 2 sources.

1. Tanneries
2. Slaughterhouses

Therefore, only one quality report is filled for Red Meat Production Statistics but indicators and explanations of these two data sources is stated separately.

8- Frequency of Dissemination

This field is filled by implementation unit. Dissemination period of results is specified here. If press release is published periodically and monthly for the same survey, dissemination period belonging to all of the press releases is specified in this field.

When “periodically” or “other” options are marked, explanations of them should be written.

Example-1: In Household Labour Force Statistics, periodical estimations are published monthly and also annual estimations are produced. Therefore, both “Periodically” and “Annually” should be marked. Moreover, “Results based on 3-month moving average are published every month” is written next to “Periodically” field.

Example-2: Since Annual Industry and Service Statistics are published once in a year, “Annually” field is marked as a dissemination frequency.

For the surveys having dissemination period different than annually or periodically such as “twice in a year” or “once in an eighteen months”, other option should be marked and dissemination period should be written there as an explanation.

9- Date of preparation

This field is filled according to the completion date of the quality report by Sampling and Analysis Techniques Department.

10- Frequency of data collection

This field is filled by implementation unit. The frequency of data collection is specified. Data collection frequencies that belong to each of the data sources (researches, census, sampling, administrative sources) should be written clearly.

Example-1: For the Household Budget Surveys, dissemination period is annual whereas data collection frequency is monthly.

Example-2: For Tourism Statistics, data are collected once in three months from households living in the country with computer assisted personal interview.

Needed explanations are given under “Explanations” field.

11-Data Collection Method(s)

This field is filled by implementation unit. All data collection methods is written clearly such as face-to-face interview and mailing. Data collection method and data collection frequency should be complement of each other.

For example, Annual Structural Business Statistics are collected by interviewers on a face-to-face interview most of the time and also with e-mail. All related methods in the options should be marked in the report.

Needed explanations are given under “Explanations” field.

12-Reference Period

This field is filled by implementation unit. “Start date” and “End date” of reference period is written to this field. (Date of the field application is not written.) If reference period is one day, start date and end date is written as the same. Different reference period for different variables could be defined. In this case, reference period of the main variable of the survey is written to this field.

Needed explanations are given under “Explanations” field.

Example-1: In Income and Living Conditions Survey, reference period of the main variable “income” is written in the quality report. Variables contain information about ownership and housing are collected during field application but these variables are not the main variable of the survey; therefore, their reference period is not given. Since institutional quality report is filled for annual statistics, annual statistics is taken into account while writing reference period even if the survey is done periodically. Periodical reference period will not be stated here.

Example-2: Household Labour Force Survey is implemented periodically. In every application period, reference week is admitted as first week of the related month starting with Monday and ending with Sunday. However, since institutional quality report is filled for annual statistics of this survey, “Start date” and “End date” of reference period is the first day and last day of the related year, e.g. reference period for 2012 is written as 01.01.2012-31.12.2012.

13-Survey Coverage

This field is filled by implementation unit. Geographical and sectoral coverage of the survey, target population and non-coverage information should be given in this section.

Coverage: All sampling units and the area where these sampling units are chosen from are defined as coverage.

Information about target population and geographic and sectoral coverage of the survey should be specified clearly in this field.

Example-1: Coverage of Household Labour Force Survey is all private households who are living in the territory of Republic of Turkey are covered. Residents of schools, dormitories, kindergartens, rest homes for elderly persons, special hospitals, military barracks and recreation quarters for officers are not covered.

Example-2: Coverage of Life Satisfaction Survey is all private households of Republic of Turkey, which have 18 years of age and over persons who are Turkish citizens and foreign people are covered. Institutional population (dormitories, rest homes for elderly persons, special hospitals, military barracks and recreation quarters for officers etc.) are not covered.

Geographical Coverage: Geographical coverage describes the locations covered by the survey such as Turkey, rural area, urban area, NUTS1, NUTS2, NUTS3, etc.

For example, for the Household Budget Surveys, geographical coverage is the entire of the settlements within the borders of the Republic of Turkey.

Sectoral Coverage: It covers sectors related with aim of the survey.

For example, Sectoral coverage of “Structure of Earnings Survey 2010” includes the statistical classification of economic activities in the European Community (NACE) Rev.2

<i>NACE REV.2</i>	
<i>(B)</i>	<i>Mining and quarrying</i>
<i>(C)</i>	<i>Manufacturing</i>
<i>(D)</i>	<i>Electricity, gas, steam and air conditioning supply</i>
<i>(E)</i>	<i>Water supply; sewerage, waste management and remediation activities</i>
<i>(F)</i>	<i>Construction</i>
<i>(G)</i>	<i>Wholesale and retail trade; repair of motor vehicles and motorcycles</i>
<i>(I)</i>	<i>Accommodation and food service activities</i>
<i>(H)</i>	<i>Transportation and storage</i>
<i>(J)</i>	<i>Information and communication</i>
<i>(K)</i>	<i>Financial and insurance activities</i>
<i>(L)</i>	<i>Real estate activities</i>
<i>(M)</i>	<i>Professional, scientific and technical activities</i>
<i>(N)</i>	<i>Administrative and support service activities</i>
<i>(P)</i>	<i>Education</i>

<i>(Q)</i>	<i>Human health and social work activities</i>
<i>(R)</i>	<i>Arts, entertainment and recreation</i>
<i>(S)</i>	<i>Other service activities</i>

According to related survey, parts of the classification which are out of the scope are specified in this field.

14-Classifications Used

This field is filled by implementation unit. Names and sources of all classifications used in the survey are written to this field. If classifications were not used, this situation is specified in the “Explanations” field.

For example, International Standard Classification of Occupations (ISCO 88) for professional status, ISCE-93 for general occupational status and ISCED-2011 classification for educational status are used for Tourism Statistics. Therefore, three of these classifications are marked in the quality report.

15- Type of Data Source(s)

This field is filled by implementation unit. Application method of survey such as census, sampling, administrative sources is specified. If both census and sampling were used, both of the options are marked.

For example, in fishery statistics survey-2010, the census is used for large scale fishermen and the sampling method is used for small scale fishermen. Both of the options are marked in the quality report for this survey.

Needed explanations are given under “Explanations” field.

16- Statistical Unit

This field is filled by Sampling and Analysis Techniques Department. Determined statistical unit such as household or individual should be specified here.

“Statistical unit” is an entity for which information is sought and for which statistics are ultimately compiled. According to subject of the survey, statistical unit can be an actual entity or a legal entity.

For example, statistical unit of Fishery Products Survey is all vessels belonging to professional fishermen fishing at freshwater.

17- Sampling/Census Unit(s)

This field is filled for the surveys based on census by implementation unit and for the surveys including sampling by Sampling and Analysis Techniques Department, except for administrative records. Sampling unit of survey and its number is specified.

While writing sample size, final sampling unit should be taken into account. Sample size should be written according to secondary sampling unit for the surveys used the stratified two-stage cluster sampling method. In most of the household surveys, primary sampling units are clusters (blocks) whereas secondary sampling units are households. According to that, number of chosen households is written as sample size.

For example, sampling unit of household budget survey is households.

On the other hand, sample size can be number of individuals, enterprises or local units.

For example, sampling unit of labour cost survey is a local unit. Sampling method of labour cost surveys was designed to produce estimations on region of the local unit the size of the enterprise that the local unit is associated and the economic activity of the local unit.

18- Estimation Level

This field is filled except for census and administrative records by Sampling and Analysis Techniques Department.

Estimation level of the survey (Turkey, Turkey-rural/urban, NUTS2, NUTS2-rural/urban, etc.) is written to this field. According to frequency of data collection of the survey (annually, periodically, etc.), changes in estimation level is specified clearly.

Estimation on different estimation levels can be used in same survey. In this case, all of the estimation levels should be stated in this section.

Example-1: In household budget survey, estimations are produced using annual and 3-annual data for different estimation levels. Therefore, all estimation levels of survey are stated separately.

Example-2: Structure of Earnings Survey 2010 was designed on the basis of local unit and employees to produce estimations on region of the local unit, the size of the enterprise that the local unit is associated and the economic activity of the local unit. Stratification variables with respect to estimation level are:

- The region of the local unit: NUTS 1
- The size of the enterprise that the local unit is associated: 10-49, 50-249, 500-999, 1000+
- The economic activity of the local unit: NACE Rev. 2 B, C, D, E, F, G, I, H, J, K, L, M, N, P, Q, R, S two digit level.

All of these explanations are given under “Estimation Level” field in the quality report.

19- Frame(s)

This field is filled by Sampling and Analysis Techniques Department. Frame used in the survey, information about its coverage and non-coverage and changes made in frame is given clearly under this section. Reference period of frame is also stated here.

For example, all local units in the 2009 business registers within the context of survey sectoral coverage and employing 10 and more employees were used as sampling framework in the Structure of Earnings Survey 2010.

Frame: A list, map or other specification of the units which define a population to be completely enumerated or sampled.

20- Sampling Method

This field is filled except for census and administrative records by Sampling and Analysis Techniques Department.

Used sampling method with respect to sampling design in the survey is written to this field. (Stratified multi-stage cluster sampling, stratified two-stage cluster sampling, etc.)

Moreover, all of the stages passing until selection of the final sampling unit and used selection methods in these stages are explained.

For example, for the survey used stratified two-stage cluster sampling method, both selection methods used at two-stage are written clearly such as clusters at the first stage are selected with probability proportional to size or systematic selection and households at the second stage with systematic selection.

Rotation information of longitudinal surveys and number of subsamples is stated under this section.

If stratified sampling is used, explanations about design domains beside estimation level should be given.

For example, while selecting a sample in a household survey having NUTS2 estimation level, if blocks are classified in terms of their size or being rural-urban area, this information is written in “Primary Sampling Unit Selection Method” and “Final Sampling Unit Selection Method” fields.

PSU: If multi-stage sampling method is used, primary sampling unit (PSU) should be stated.

FSU: If multi-stage sampling method is used, final sampling unit (FSU) should be stated.

PSU Selection Method: If multi-stage sampling method is used, selection method used at first stage should be written such as probability proportional to size or systematic selection.

FSU Selection Method: If multi-stage sampling method is used, selection method used at final stage should be written such as probability proportional to size or systematic selection.

For example, ICT (Information and Communication Technology) Usage in Households and by Individuals Survey 2012 is applied to sample households selected by a 2 stage stratified clustered sample. In first stage; in all areas in the frame is divided into clusters (blocks) of approximately 100 households each. These were selected as probability proportional to size (PPS). In the second stage; the sample addresses are determined from selected clusters (blocks) by using systematic selection method. Total 944 blocks are chosen as 383 blocks from rural and 561 blocks from urban. In this case, total 12806 households are chosen as 3810 households from rural (10 households from each block) and 8976 households from urban (16 households from each block). All of this information is explained in “PSU selection method” and “FSU selection method” sections.

If multi-stage sampling method (more than two-stage) is used in the survey, all of the selections done after second stage should be explained in “FSU selection method” section.

Cluster/Rotation Information: Information about clusters such as cluster size, rotation information of longitudinal surveys and rank order of subsamples is stated under this section.

21- Sample Allocation Method

This field is filled except for census and administrative records by Sampling and Analysis Techniques Department.

Sample allocation method of survey used stratified or cluster sampling methods and information about used formulations and method is stated under this section.

22- Estimations and Notations

This field is filled except for census and administrative records by Sampling and Analysis Techniques Department.

Information about weighting procedure, such as design weights, non-response adjustment, calibration and trimming, is given here. All formulations used in weighting procedures are stated.

If indirect calculation is used for estimation in the survey, all information about the calculation should be written clearly to this field.

23- Data Completeness Rate

This field is filled by implementation unit. This indicator is not calculated for non-regulatory studies.

Data completeness rate is calculated in accordance with “Relevance” principle of EUROSTAT. To calculate this rate, definitions and classifications of variables, sent by implementation department to EUROSTAT, need to be appropriate for EUROSTAT’s standards. Variables which are not appropriate for these standards are not included in the calculation.

The indicator shows to what extent statistics are available compared to what should be available.

In “Explanations” field, information about which variable are missing and which cannot be met also the reasons for incompleteness is stated clearly for each of them.

The rate of available data should be applicable to all statistical processes (including use of administrative sources).

The calculation formula of data completeness rate is given below.

$$R1_U = \frac{\#A_D^{rel}}{\#D^{rel}}$$

D^{rel} in the denominator is the set of data cells required (i.e. excl. derogations/confidentiality) and A_D^{rel} in the numerator is the corresponding subset of available/provided data cells. The notation # D means the number of elements in the set D (the cardinality).

If certain relevant variables are not reported, the statistics are incomplete. This can be due to data not being collected or data being of low quality or confidential. In this case, data completeness rate is low.

The target value for this indicator is “1” meaning that 100% of the required or relevant data cells are available. Sending micro data of the survey to EUROSTAT cannot be interpreted as data completeness rate is “1”.

Example-1: Data cells required by EUROSTAT are 113 for Household Labour Force Survey and provided data cells are 101 (12 variables cannot be provided for different reasons such as changes in codification, inconsistency in definitions of the variables, etc.). Therefore, data completeness rate is calculated like below.

Data Completeness rate= $101/113 = 0,89$

Example-2: Data completeness rate is not calculated for Household Budget Survey since there is no EUROSTAT-Regulation about it.

24-Name of the Main Variable

This field is filled by implementation unit on the basis of main variables of the survey.

Name of the main variable is written to this field also its definition and calculation is stated in “Explanations” field. Information about availability of international concepts and classifications of the main variable and sharing results at international level is stated here.

When “Not Applicable” or “Not Shared” options are marked, explanations are written to “Reason” fields.

Breaks such as NACE/NUTS and estimations calculated on the basis of these breaks are given in “Indicators” section. Methods used in calculating annual estimations of main variables are stated clearly.

Quality indicators of estimation (standard error, confidence interval, CV, DEFF) are filled by Sampling and Analysis Techniques Department.

Standard Error/ Deviation Values for Main Variables: Calculation of standard error/deviation for each estimate obtained by implementation unit is done by Sampling and Analysis Techniques Department. This field is not filled for surveys based on administrative records.

Main tables in annual press releases, breaks and variables requested by EUROSTAT in quality reports are taken into account in this field. It is calculated taking square root of main statistics’ variance ($\sqrt{\hat{V}(\hat{\theta})}$) computed depending on the sampling design. The smaller the standard error denotes the narrower confidence interval and the more accurate estimator. Therefore, small value of standard error is desirable.

Confidence Intervals for Main Statistics: This field is filled taking into account the standard error values by Sampling and Analysis Techniques Department.

This field is not filled for surveys based on census or administrative records. Only sampling part is taken into account while calculating confidence interval for surveys based on both sampling and census.

Calculation formula of Confidence Intervals for Main Statistics is given below.

$$\left[\hat{\theta} - d; \hat{\theta} + d \right] \text{ or } \hat{\theta} \pm d$$

$$d = t * \sqrt{\hat{V}(\hat{\theta})}, \text{ t depends on the distribution and confidence level.}$$

Computed confidence interval depends on sampling design, estimator and used method in variance estimation.

Narrower confidence interval is desirable.

CV Values for Main Statistics: This field is filled only for survey including sampling by Sampling and Analysis Techniques Department.

Only sampling part is taken into account while calculating CV (coefficient of variation) values for surveys used both sampling method and census.

CV (coefficient of variation) is calculated in accordance with “Accuracy” principle of EUROSTAT. CV values calculated regarding estimates of main variables is written on the basis of estimation level.

The CV is a relative (dimensionless) measure of the precision of a statistical estimator, often expressed as a percentage. In other words, CV is a measure of the dispersion for data set.

Calculation formula of CV is given below.

$$CV_e(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$$

The subscript "e" stands for estimate. The smaller CV value denotes the narrower confidence interval and the more accurate estimator. Therefore, small CV values are desirable.

Design Effect for Main Statistics: This field is filled except for census and administrative records by Sampling and Analysis Techniques Department.

Design Effect (DEFF) is calculated in accordance with “Accuracy” principle of EUROSTAT. The design effect is the ratio of the variance of a statistic with a complex sample design to the variance of that statistic with a simple random sample. Its calculation formula is given below.

$$DEFF = \frac{\hat{V}_C(\hat{\theta})}{\hat{V}_{SRS}(\hat{\theta})}$$

$\hat{\theta}$ is an estimation.

The ratio should be close to ‘1’ in order to say that the estimation is a reliable. Design effect is calculated for survey used cluster sampling method.

25- Data revision - average size: This field is filled except for census and administrative records by implementation unit.

Institutional quality report is prepared annually and so it is necessary to calculate indicators of annual statistics. The data revision average size is expected to calculate if periodically / monthly results are produced beside annual statistics in a study and there is a revision policy of this study. If the revision is made within year of institutional quality reports, the average size of the data revision is calculated under the definition given below by taking into account the date and frequency of revision of quality report (monthly or periodically). Household Labour Force is the only survey produced annual statistics as well as periodical statistics in accordance with revision policy, so the data revision average size is written in institutional quality report of Household Labour Force Survey.

If there is no revision for survey, this is stated in this field.

The “revision” is defined as the difference between a later and an earlier estimate of the key item.

The average size of revision is the average over a time period of the revisions of a key indicator.

In the following table this situation is illustrated for a revision analysis where the policy has K revisions and n reference periods are included in the analysis.

	Reference periods				
Releases	1	...	t	...	n
1 st release	X_{11}	...	X_{1t}	...	X_{1n}
...
kth release	X_{k1}	...	X_{kt}	...	X_{kn}
...
Kth and final release	X_{K1}	...	X_{Kt}	...	X_{Kn}

Revisions involve a time series: when publishing an estimate of the key indicator referring to time t, it is necessary to release the revised version of the indicator referring to a set of previous periods.

The average size of revisions is applicable to statistical processes (national accounts, short term statistics etc.)

MAR (Mean Absolute Revision)

Calculation formula of Mean Absolute Revision is given below.

$$MAR = 1/n \sum_{t=1}^n |X_{Lt} - X_{Pt}| .$$

Where,

X_{Lt} : “later” estimate, Lth release of the item at time reference t;

X_{Pt} : “earlier” estimate, Pth release of the item at time reference t;

n: No. of estimates (reference periods) in the time series taken into account.

$n \geq 20$ is recommended for periodically estimates while $n \geq 30$ is recommended for monthly estimates.

For example, the Labour Cost Survey in Federal Statistical Office of Germany is implemented periodically. The value of MAR for revisions done between years 2009-2011 is given below.

In the following table, each cell denotes the ratio of the difference between the estimates calculated in two successive periods to the first period estimate. According to these values, MAR value is obtained as 0.4. The number of available reference periods is 11.

Releases	Reference Period										
	2009			2010				2011			
	2 nd quarter	3 rd quarter	4 th quarter	1 st quarter	2 nd quarter	3 rd quarter	4 th quarter	1 st quarter	2 nd quarter	3 rd quarter	4 th quarter
1 st release	-0.5	5.2	4.7	2.8	1.2	4.9	3.1	2.4	3.3	5.6	2.5

2 nd release	0.4	5.8	5.0	1.0	1.1	4.9	3.2	2.2	3.5	5.0	2.5
Differences	0.9	0.6	0.3	-1.8	-0.1	0.0	0.1	-0.2	0.2	-0.6	0.0
Absolute Differences	0.9	0.6	0.3	1.8	0.1	0.0	0.1	0.2	0.2	0.6	0.0
MAR	0.4										

26- Timeliness (first results) - T1

This field is filled except for administrative records by implementation unit for annual survey producing and publishing first result.

Timeliness- first results (T1) is calculated in accordance with “Punctuality - delivery and publication” principle of EUROSTAT. This indicator is applicable to statistical processes with preliminary data releases. The number of days from the last day of the reference period to the day of publication of first results is stated in “T1” field. The number of days admitted by EUROSTAT is stated in “Length of time” field.

Timeliness (first results): Length of time between the last day of the reference period and the day of publication of first results.

Calculation formula of T1 is given below:

$$T_1 = d_{first} - d_{refp}$$

d_{first} : Release date of first results.

d_{refp} : Last day (date) of the reference period of the statistics.

Smaller values denote higher timeliness. Measurement units are calendar days.

Length of time (day) is calculated as the number of days from the last day of the reference period to the delivery date of first results to EUROSTAT.

Length of time (day) = Delivery date of the first results to EUROSTAT - the last day of the reference period

For example; according to Regulation-58/97 and Regulation- 295/2008 determined by EUROSTAT, timeliness is 608 day for annual industry and services statistics. (Delivery date of data to EUROSTAT + 18 months). Calculations done related with first results using Table 1 are given below.

Table 1: Annual industry and services statistics

	Last day (date) of the reference period	31.01.2009
	Release date of the first results	26.02.2012
	Length of time (the number of days)	608
26- TIMELINESS (First Results) T1 :		787

T1: Release date of first results - Last day (date) of the reference period = 787,

Length of time (day): Delivery date of the first results to EUROSTAT - the last day of the reference period= 608

Timeliness of first results (T1) was delayed 179 days from length of time (608 days). Submission to EUROSTAT could be provided after 179 days from deadline. In this example, since T1 is greater than length of time, timeliness condition cannot be provided.

Needed explanations are given under “Explanations” field.

27- Timeliness (final results) - T2

This field is filled except for census and administrative records by implementation unit.

Timeliness- final results (T2) is calculated in accordance with “Punctuality - delivery and publication” principle of EUROSTAT. Fields are filled in terms of days. In calculation, final results are used instead of first result. That is the difference between timeliness (first results) and timeliness (final results).

Timeliness-final results (T2): The number of days from the last day of the reference period to the day of publication of complete and final results.

Calculation formula of T2 is given below:

$$T_2 = d_{finl} - d_{refp}$$

d_{finl} : Release date of final results.

d_{refp} : Last day (date) of the reference period of the statistics.

Smaller values denote higher timeliness. Measurement units are calendar days.

Length of time (day): The number of days from the last day of the reference period to the delivery date of first results to EUROSTAT.

Length of time (day) = Delivery date of the first results to EUROSTAT - the last day of the reference period

For example; according to Regulation - (EC) 808/2004 determined by EUROSTAT, timeliness is 5th October of the related year for ICT (Information and Communication Technology) Usage Survey in Enterprises. Calculations related with first results using Table 2 are given below.

Table 2: ICT Usage Survey in Enterprises

	Last day (date) of the reference period	31.01.2012
	Release date of the first results	16.11.2012
	Length of time (the number of days)	248
27- TIMELINESS (Final Results)- T2 :		321

T2: Release date of final results - Last day (date) of the reference period = 321

Length of time (day): Delivery date of the final results to EUROSTAT - the last day of the reference period= 248

While calculating length of time, the timeline between 31.01.2012 (the last day of the reference period) and 05.10.2012 (Determined submission date of the final results to EUROSTAT) was taken into account.

Timeliness of final results (T2) was delayed 73 days from length of time (248 days). Submission to EUROSTAT could be provided after 73 days from deadline. In this example, since T2 is greater than length of time, timeliness condition cannot be provided.

Needed explanations are given under “Explanations” field.

28- Punctuality - delivery and publication: This field is filled by implementation unit. Defined in accordance with “timeliness and punctuality” principle of EUROSTAT, this indicator is computed by taking into account pre-announced release dates calendar.

Punctuality is the time lag between the release date of data and the target release date. Target release date refers to scheduled date or announced date in an official release calendar.

Calculation formula of punctuality is given below.

$$P_3 = d_{act} - d_{sch}$$

d_{act} : Actual date of the statistics

d_{sch} : Scheduled date of the statistics (announced date by Turkstat for release)

Ideal P_3 value is zero. If P_3 value does not equal to zero (including statistics published before scheduled date), punctuality cannot be provided. Measurement units are calendar days.

For Information and Communication Technology (ICT) Usage Survey in Enterprises -2012, scheduled date of the statistics is 16.10.2012. Also, actual date of statistics is 16.10.2012. According to that, the value for P_3 is 0.

	Scheduled date of the statistics:	16.10.2012
	Actual date of the statistics:	16.10.2012
28) Punctuality		0

Needed explanations are given under “Explanations” field.

29- Number of Metadata Consultations: This field should be filled by Metadata and Standards Department in coordination with Information and Technologies Department. This indicator is defined in accordance with “accessibility and clarity” principle of EUROSTAT. When number of metadata consultations is computed, internal page views are excluded. Calculation of the number of metadata consultations is done on the basis of first month after release date of data. Since this indicator is computed only by EUROSTAT, the metadata should be structured as ESMS.

Metadata – consultations: Number of metadata consultations (ESMS) within a statistical domain for a given time period. By "number of consultations" it is meant the number of times a metadata file is viewed. 1-month period after preparation of the metadata is taken into account as time period in institutional quality report.

Calculation formula of Metadata – consultations is given below.

$AC1 = \#ESMS$

#ESMS denotes the absolute number of elements in the set ESMS (Euro-SDMX Metadata Structure). In other words, ESMS contains headlines of metadata and explanations that provided summary information in the assessment of data quality and process of production. In this case, the set ESMS represents the ESMS files consulted for a specific subject-matter domain for a given time period.

Needed explanations are given under “Explanations” field.

30- Metadata Completeness Rate: This field should be filled by Metadata and Standards Department in coordination with Information and Technologies Department. This indicator is defined in accordance with “accessibility and clarity” principle of EUROSTAT. When number of metadata consultations is computed, internal page views are excluded. Calculation of metadata completeness-rate is based on monthly monitoring reports. The calculation is done at the level of ESMS files. All information is to be retrieved from ESMS files. In case the ESMS is empty for the different categories specified previously, no calculation is needed but a descriptive text should be replaced. Since this indicator is computed only by EUROSTAT, the metadata should be structured as ESMS.

Metadata completeness – rate is the ratio of the number of metadata elements provided to the total number of metadata elements applicable.

Calculation formula of Metadata – completeness-rate is given below.

$$AC3_c = \frac{\sum \#M_L}{\sum \#L}$$

#M_L : the number of available metadata elements in the set L

#L : the number of elements in the set L

L is defined as the set of metadata elements. Three groups of metadata follow as:

1. metadata about statistical outputs (3, 4, 5, 8.1, 9 and 10. Elements: Euro-SDMX Metadata Structure Mart-2009)
2. metadata about statistical processes (11, 20.1, 20.2, 20.3, 20.4, 20.5, 20.6: Euro-SDMX Metadata Structure Mart-2009)
3. metadata about quality (12-19: Euro-SDMX Metadata Structure Mart-2009)

Computations are made separately for each of the three groups and for each of the combinations.

The target value for this indicator is 1 meaning that 100% of metadata is available from what is required to the statistical process. This rate only reflects the existing amount of metadata for a certain statistical process but not the quality of that information.

Needed explanations are given under “Explanations” field.

31- Number of Data Tables- Consultations: This field should be filled by Information and Technologies Department. This indicator is defined in accordance with “accessibility and clarity” principle of EUROSTAT. When number of metadata consultations is computed, internal page views are excluded. Calculation of data tables’ consultations is based on monthly monitoring reports.

Data tables – consultations is the number of consultations of data tables within a statistical domain for a given time period. By "number of consultations" it is meant number of data tables views, where multiples views in a single session count only once.

Calculation formula of Data tables – consultations is given below.

$$AC2 = \#CONS$$

#CONS refers to the absolute number of elements in the set CONS.

While computing data tables – consultations, 1-month period after the release of the metadata is taken into account as time period in institutional quality report.

Needed explanations are given under “Explanations” field.

32- Dissemination format (CD, WEB, publications etc.): This field is filled by implementation unit.

Implementation unit is stated in which format results presented to the user. Dissemination format of survey (CD, website, publications, electronic publications, etc.) is written clearly.

33- Length of comparable time series: This field is filled by implementation unit. Computed in accordance with “Coherence and comparability” principle of EUROSTAT, the length of comparable time series is applicable to all statistical processes producing time-series.

This indicator refers to number of reference periods in time series from last break. Breaks in statistical time series may occur when there is a change in the definition of the parameter to be estimated or the methodology used for the estimation.

If there has not been any break, the indicator is equal to the number of the time points in the time series.

Revisions, breaks and reasons that occur in statistical time series are stated in this field.

Calculation formula of Length of comparable time series is given below.

$$CC_I = J_{last} - J_{first} + 1$$

J_{last} : the sequence number of the last reference period with disseminated statistics

J_{first} : the sequence number of the first reference period with comparable statistics

Example-1: For 2009 Annual Industry and Service Statistics, Statistical Classification of Economic Activities Sectors NACE Rev.1 was used between the years 2003-2009. Because of the first break in 2009, from onwards this date, NACE Rev.2 has been used. In this case, length of comparable time series in 2012 quality report is that: 2012-2009+1=4

Example-2: Breaks occurred due to population projections in Household Labour Force Statistics-2004. In this case, length of comparable time series in 2012 quality report is that: 2012-2004+1=9.

Needed explanations will be given under “Explanations” field.

34-Over Coverage rate: This field is filled for census by implementation unit and for sampling survey by Sampling and Analysis Department, except for administrative records. Computed in accordance with “Accuracy” principle of EUROSTAT, over-coverage rate’s calculation formula is given below.

$$OCr_w = \frac{\sum_O w_j + (1-\alpha)\sum_Q w_j}{\sum_O w_j + \sum_E w_j + \sum_Q w_j}$$

O: unit of non-coverage

E: unit of eligible

Q: units of unknown eligibility

w_j : weight of unit

α : The estimated proportion of cases of unknown eligibility that are actually eligible.

wj=1:Un-weighted rate

The calculation is done on un-weighted data.

When administrative data are used, coverage deficiencies may be due to delays in reporting (typical for business statistics) and errors in unit identification, classification, coding etc. All of these are written in this field.

Used α value is stated in this field.

Over-coverage: there are units accessible via the frame, which do not belong to the target population. Units repeating more than one time are also defined as over coverage.

The rate of over-coverage is the proportion of units accessible via the frame that do not belong to the target population.

The target value of this indicator is as much as possible close to 0.

Needed explanations are given under “Explanations” field.

35-Common unit rate: This field is filled except for census and administrative records by implementation unit. This indicator is computed in accordance with “Accuracy” principle of EUROSTAT. If there are some variables come from the survey and some from the administrative data, this rate is calculated regarding coverage of both administrative records and survey. The proportion is applicable to mixed statistical processes where some variables come from survey data and others from administrative source(s).

This indicator does not apply if administrative data is used only to produce estimates without being combined with survey data.

Common units refer to those units that are included in the data stemming from an administrative source and survey data.

The proportion of common units is the proportion of units covered by both the survey and the administrative sources in relation to the total number of units in the survey.

Calculation formula of common units-proportion is given below.

$$A_d = \frac{\text{No. of common units across survey data and admin.sources}}{\text{No. of unique units in survey data}}$$

The “no. of unique units in survey data” in the denominator means that if a unit exists in more than one source, it should only be counted once. Linking errors related to common units should be detected and resolved before this indicator is calculated.

Units that are stemming from both survey data and administrative source are stated in this field.

In Fishery Statistics 2013, 5800 registered professional fishermen and 67 types of fish are covered. Data on blue fin tuna are taken directly from Ministry of Food, Agriculture and Livestock, rest of the data are compiled by the survey. Only 100 of the 5800 fishermen have the authority to hunt blue fin tuna. In this case, number of common units stemming from both survey data and administrative source is 100.

The proportion of common units is that: $(100/5800)*100=1.72\%$

Needed explanations are given under “Explanations” field.

36- Unit non-response rate: This field is filled for full census by implementation unit and for sampling survey by Sampling and Analysis Department, except for administrative records. This indicator is computed in accordance with “Accuracy” principle of EUROSTAT.

Unit non-response-rate is the ratio of the number of units with no information or not usable information (non-response, etc.) to the total number of eligible units.

Calculation formula of unit non-response-rate is given below.

$$NRr = 1 - \frac{\sum_R w_j}{\sum_R w_j + \sum_{NR} w_j + \alpha \sum_Q w_j}$$

R: the set of responding eligible units

NR: the set of non-responding eligible units

Q: the set of selected units with unknown eligibility

w_j : weight of unit j

α : The estimated proportion of cases of unknown eligibility that are actually eligible

The calculation is done on un-weighted data.

Used value of α is stated in this field.

The target value of this indicator is as much as possible close to 0.

Needed explanations will be given under “Explanations” field.

Distribution According to Reasons of Statistical Non-response Units

This field is filled for full census by implementation unit and for sampling survey by Sampling and Analysis Department, except for administrative records. The distribution information of non-response units according to reasons of non-response is stated in this field. If there is different reason of non-response for statistical unit (enterprise, local unit, household, individual), this should be added to list and the number of this should be stated.

Needed explanations are given under “Explanations” field.

37-Item non-response rate: This field is filled by implementation unit, except for administrative records. This indicator is computed in accordance with “Accuracy” principle of EUROSTAT.

Calculation formula of unit non-response-rate is given below.

$$NR_{Y}^{REQ} = 1 - \frac{\sum_{R_Y} w_j}{\sum_{R_Y} w_j + \sum_{NR_Y} w_j}$$

Y_R : the set of eligible units responding to item Y(as required)

NR_Y : the set of eligible units not responding to item Y although this item is required.

w_j :weight of unit j, described below

The calculation is done on weighted data.

Un-weighted rate: $w_j= 1$

For example, 2010-Industry/Service Institutions Research & Development Statistics is computed on the basis of NUTS1 and 3610 enterprises only for “other current expenditure” variable (materials, supplies and equipments taken without giving any money by the statistical unit in order to support R&D within the current year). Values for indicator are given below.

Name of the variable: other current expenditure

Number of non-response units (enterprises) for other current expenditure variable: 214

Total number of units (enterprises) for other current expenditure variable: 3610

In this case, item non-response rate is calculated as $214/3610 = \% 5.93$

Weighting values was taken as 1 in calculation.

The un-weighted item non-response rate should be calculated before the data editing and imputation in order to measure the impact of item non-response for the key variables.

The target value of this indicator is as much as possible close to 0.

The total number of unit and the number of non-response units is stated in this field.

Needed explanations are given under “Explanations” field.

38- Unit imputation rate and calculation method: This field is filled by implementation unit, except for administrative records. If imputation has been done for full of non-response unit, this indicator is computed and method of imputation is explained in this field.

Calculation formula of unit imputation-rate is given below.

$$\text{Unit Imputation Rate} = \frac{\text{the number of imputed value}}{\text{the number of imputed value} + \text{the number of response units}}$$

39-Item imputation rate and calculation method: This field is filled by implementation unit, except for administrative records.

If imputation has been done for non-knowledge variable, this indicator is computed and method of imputation is explained in this field.

Calculation formula of item imputation-rate is given below.

$$IR_{Yr} = \frac{\sum_{I_Y} w_j y_j}{\sum_{I_Y} w_j y_j + \sum_{K_Y} w_j y_j}$$

I_Y = the set of units for which variable Y is imputed

K_Y = the set of units for which the value of variable Y is kept / set of units for which variable Y is not imputed

y_j = value of variable Y (j described below)

The calculation is done on un-weighted data. ($w_j=1$)

The target value for this indicator equal or close to zero is desirable.

The total number of unit and the number of imputed value is stated in this field.

40- Quality control studies: If control studies of data quality are done by Coordination of Data Collection Department using CATI technology, this field is filled, except for administrative records.

In the context of quality control studies, required items are given below.

- Start date of the control study
- Period of the control study
- Frequencies of the control studies
- Number of responded units
- Stages of Control Study
- Calculation method and calculated indicators

If there is no quality control study, this is stated in “Quality Control Studies” field like “Quality control studies are not done.”