|  |  |
| --- | --- |
|   | Approved by the order of the Chairman of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstandated May 23, 2018 63 |

**Methodology for assessing the quality of official statistical information**

**Chapter 1. General provisions**

1. This Methodology for assessing the quality of official statistical information (hereinafter - Methodology) refers to a statistical methodology formed in accordance with international standards and approved in accordance with the Law of the Republic of Kazakhstan dated March 19, 2010 "On State Statistics" (hereinafter - Law).

2. The methodology defines the main aspects and criteria for assessing the quality of official statistical information.

3. The methodology is applied by the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan (hereinafter - the Committee) in assessing the quality of official statistical information.

4. The following definitions are used in the Methodology:

1) asymmetry in mirror statistics of flows - the difference or absolute difference between incoming and outgoing flows between countries;

2) accuracy - a criterion determined by the degree of approximation of calculations to actual values;

3) imputation - the process of replacing missing, incorrect or inconsistent values with other values;

4) consistency - a criterion due to the ability to combine and share official statistical information obtained from various sources;

5) accessibility and clarity - criteria determined by the conditions and opportunities for users to receive, use and interpret official statistical information;

6) relevance - a criterion determined by the degree of compliance of statistical data with the current and possible needs of users;

7) comparability - a criterion determined by the degree of comparability of official statistical information over time, across regions, or across other areas of activity;

8) timeliness - a criterion determined by the time interval between the date of appearance (publication) of official statistical information and the event or phenomenon that it describes;

9) punctuality - a criterion determined by the delay time from the date of publication of official statistical information to the scheduled day.

Chapter 2. Quality assessment of official statistical information

5. The assessment of the quality of official statistical information is carried out on the basis of the information provided in the Quality Reports by the structural divisions of the Committee that form official statistical information.

6. The quality report contains information on the degree of compliance of the methodology for the formation of official statistical information with the quality assessment criteria for the purpose of presenting it to users. The structure for completing the Quality Report is given in Appendix 1 to this Methodology.

7. The quality report is generated in the Metadata component of the Integrated Information System "e-statistics" (hereinafter - the Metadata system) and uploaded in paper format for internal approval by the Committee.

8. Quality reports are approved by the Vice-Chairman of the Committee who supervises the issues of the quality management system. Approved quality reports are published on the Internet resource of the Committee.

9. If there are no changes in the Quality Report, the date of certification is updated (clause S.3.1 "Last confirmation of updated metadata"). When changes are made to the Quality Report, an updated and reapproved version of the Quality Report is published on the Internet resource of the Committee.

10. Information on items S.7 "Legal basis", S.8 "Confidentiality and data protection", S.9 "Publication policy", S.14 "Relevance", S.12.2 "Quality documentation" is submitted by the relevant structural units Committees whose competence includes these issues to the structural unit responsible for filling the Metadata system for the formation and updating of Quality Reports.

Chapter 3. Criteria for assessing the quality of official statistical information

11. The main criteria for assessing quality are:

1) accessibility and clarity;

2) relevance;

3) accuracy;

4) timeliness and punctuality;

5) consistency and comparability.

12. In assessing the quality of official statistical information in accordance with the above criteria, direct and indirect assessments are distinguished.

In a direct assessment, the results are numeric and are based on a limited number of criteria (timeliness, accuracy, and data comparability).

In an indirect assessment, information is provided that characterizes official statistical information in terms of its compliance with quality criteria.

Paragraph 1. Accessibility and clarity

13. To assess the accessibility and clarity of official statistical information, the following is submitted:

1) information about regular or special press releases related to the considered group of indicators;

2) a list of publications using the considered group of indicators with the presentation of the year of publication and, if available, links to documents in electronic form;

3) an available database in on-line mode for the considered group of indicators. The description includes the domain names of the website and links to the online database.

Paragraph 2. Relevance

14. To assess relevance, information is provided that includes:

1) description of users and their needs (providing a classification of users; information on unmet user needs and their causes; plans to meet future needs);

2) measures to determine user satisfaction with a description of the methods and regularity of collecting the views and opinions of users (survey of user satisfaction, other user consultations);

3) information on the completeness of data in comparison with requests from international organizations.

Paragraph 3. Accuracy

15. To assess accuracy, the following steps are observed:

1) Describe the main sources of random and systematic errors in the output statistics to confirm overall accuracy and provide a summary estimate of all errors, with particular emphasis on their impact on the underlying calculations. An estimate of bias occurs in quantitative or qualitative terms, or both, including measures taken to reduce the number of biases;

2) when sampling, information is provided on the difference between the population value and its estimate obtained from a random sample, indicating information on adjustment for non-response, misclassification and other uncertain sources, such as the processing of outliers;

3) providing users with the level of non-responses and the risks of bias that are associated with them (coverage error; measurement errors; non-response error; processing error; model application error) and actions taken to reduce various types of errors;

4) provide information on the discrepancy between the general and sample populations in terms of excessive and incomplete coverage. Describes actions taken to reduce undercoverage with an assessment of the risks of data bias/bias;

5) provide information about the measurement error. An overall assessment of the main sources of measurement error and actions taken to correct measurement errors is provided. Describes the efforts made in the development of the statistical form and testing, information about the training of the interviewer and other types of work (based on comparison with external data, re-interview);

6) provide information about the non-response error. A qualitative estimate is provided for variables subject to non-response (sensitive questions) and non-response biases. Describes the efforts and measures being taken to reduce the number of non-response when collecting primary data and performing subsequent actions, including the technical processing of non-response.

Paragraph 4. Timeliness and punctuality

16. To assess the timeliness, information is provided on the receipt of official statistical information within the time frame corresponding to the objectives of the monitoring. The time required for the production and publication of official statistics is indicated, and efforts are being made to reduce it.

17. To assess punctuality, indicate:

1) the actual date of publication and the date planned in accordance with the annually approved Schedule for the dissemination of official statistical information;

2) the percentage of press releases published on time, based on the timing of publication;

3) reasons for press releases not published on time and measures taken to improve the situation.

Paragraph 5. Consistency and comparability

18. To assess consistency, information is provided on:

1) the degree of conformity of the methodology of the process of production of official statistical information with international standards (including the degree to which official statistical information is consistent with data obtained from other sources or statistical industries). Differences between the considered statistical materials and other information related to the output statistical data are described (including the main differences in concepts and definitions, statistical units or objects, classifications (nomenclature), geographical breakdown, reporting period, methods of adjustment);

2) the extent to which official statistical information with different production frequencies is compatible. Compares calculations for a period of less than a year and for a year and, if available, describes the reasons for the lack of consistency between the statistical results of different time periods;

3) the extent to which official statistical information is compatible with national accounts.

20. To assess the impact of differences in applied statistical methodologies, tools and procedures over time and space, describe:

1) the extent to which official statistical information is compared across countries. Comparability issues and their causes are described, and an estimate is provided on the impact of each reported difference on the output. Information is included on discrepancies in the concepts used in the field of state statistics and international concepts and definitions, as well as information on asymmetries in mirror flow statistics;

2) the extent to which official statistical information is comparable across time periods. Information is provided on possible limitations in the use of official statistical information for comparison over time periods.

When changes are made to the process of producing official statistical information from one period to another, information is provided on the possible consequences, including information on the length of comparable time series, reporting periods during which breaks in the time series occur, the causes of breaks and methods for reducing them.

21. For each quality criterion, the List of standard quality indicators given in Annex 2 to this Methodology is used.

|  |  |
| --- | --- |
|   | Appendix 1 to the Methodology for assessing the quality of official statistical information |

 **Structure for completing the Quality Report**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Section Attributes | Sections of the Report | Concept code | Description of sections | Explanations |
| S.1 | Contact details | CONTACT | Natural or legal person to contact for data or metadata, including contact details |  |
| S.1.1 | Organization | CONTACT\_ORGANI-SATION | Name of company | Full name of the organization |
| S.1.2 | Structural subdivision | ORGANIZATION\_UNIT | Name of the structural unit | Name of the structural unit in relation to the metadata file |
| S.1.3 | Name of contact person | CONTACT\_NAME | Surname, name, patronymic (if any) of the responsible person | Last name, first name, patronymic (if any) of the person responsible for monitoring (indicator) |
| S.1.3.1 | Name of the head of the responsible structural unit |  | Surname, name, patronymic (if any) of the head of the Department | Surname, name, patronymic (if any) of the head of the Department |
| S.1.5 | Postal address of the contact person | CONTACT\_MAIL | Organization postal address | Organization postal address |
| S.1.6 | Email address of the contact person | CONTACT\_EMAIL | Email address | E-mail address of the person responsible for monitoring (indicator) |
| S.1.7 | Phone number of the contact person | CONTACT\_PHONE | Phone number | Telephone number of the person responsible for monitoring (indicator) |
| S.2 | Introduction-Relevance | INTRODUCTION | General description of statistical observation: 1) definition of the branch of statistics to which statistical observation belongs, the purpose of the observation (definition of the indicator) and the scope of statistical data. It is also recommended to write a short history, which is accompanied by a general description, results and their evolution over time; 2) identification of the main users and ways of their participation; 3) determining the need for information through the means of obtaining user opinion (for example: feedback log, requests) | A brief description of the sector of statistics to which the statistical observation belongs, the purpose of the statistical observation (definition of the indicator, a brief history), the scope of the statistical data, the main users (classification of users, ways of participation of users (links where they can express their opinion)). Identification of information needs through means of obtaining user input (eg feedback log, inquiries). Show links to other surveys and indicate the contribution to the production of other statistics. |
| S.3 | Metadata update | META\_UPDATE | The date the metadata element was entered into the database or changed in the database |  |
| S.3.1 | Last confirmation of updated metadata | META\_CERTIFIED | Date of last confirmation by the statistical area administrator that the metadata entered is up to date, even if the information has not changed | Date of last confirmation that such a metadata file is up to date. This confirmation is also performed if the content of the metadata file has not changed. |
| S.3.2 | Last placed metadata | META\_POSTED | Date of last dissemination of metadata | The date of the last distribution of the respective metadata file will be automatically entered by the baseline metadata generation system |
| S.3.3 | Latest metadata update | META\_LAST\_UPDATE | the metadata content was last updated | The date of the last metadata update will be automatically entered by the baseline metadata creation system |
| S.4 | Presentation of statistical information | STAT\_PRES | Description of disseminated data that is provided to users in the form of tables, graphs or maps |  |
| S.4.1 | Data Description | DATA\_DESCR | Key features of the dataset related to disseminated data and indicators | Brief description of the main characteristics of the data set with reference to the main disseminated data and indicators |
| S.4.2 | Classification system | CLASS\_SYSTEM | Grouping objects or dividing them into groups based on their common characteristics | Providing information about the classifications used and links to their electronic versions |
| S.4.3 | Sector coverage | COVERAGE\_SECTOR | Main economic or other sectors covered by statistics | List of economic or other sectors covered by generated datasets and class/group parameters used (e.g. number of employees) |
| S.4.4 | Statistical concepts and definitions | STAT\_CONC\_DEF | Statistical characteristics of statistical observations, variables | Description of general information about statistical observation, including: range of respondents (coverage of respondents, for example, according to OKED, as on the form), survey date (submission deadline, for example, January 3 after the reporting year), data collection tools (index, name of the form, last revision where available), type of observation (sample, continuous, combined) |
| S.4.5 | Statistical object | STAT\_UNIT | The population for which information is sought and for which statistics are ultimately collected. | List of basic objects of statistical observations for which data are supplied. These objects of observation (e.g. enterprise, local unit, private households) may differ from the objects of reporting used in the main statistical surveys. |
| S.4.6 | Population (principle of selection of survey units) | STAT\_POP | Formation of the general and sample population of the survey | Formation of the general and sample population of the survey |
| S.4.7 | Territorial coverage | REF\_AREA | Country or geographic area to which the measured statistical phenomenon belongs | Regional coverage (regional cut) |
| S.4.8 | Temporal scope | COVERAGE\_TIME | Length of time for which data is available | Description of the time periods covered by the data set (i.e. the length of time over which the data is disseminated, e.g. 1985-2006, or 2000-… for certain annual data) |
| S.4.9 | Base period | BASE\_PER | A period of time that is used as a base for a set of indicators, or to which a set of constants refer | A description of the time period that is used as a base for a set of indicators, or to which a time series refers (for example, base year 2000 for certain annual data) |
| S.5 | Unit | UNIT\_MEASURE | Unit in which data values are measured | A listing of the units of measurement used for the dissemination of official statistical information (eg %, number, persons). You should also add an exact definition of the dimension (for example, thousands, millions) |
| S.6 | Reporting period | REF\_PERIOD | The period or point in time to which the measured observation relates | Statistical variables refer to specific points in time, such as a specific day, or to a specific period of time (such as a month, a calendar year, or several calendar years). If the specified reference period does not match the actual reference period, for example when there is no data for the specified reference period, this discrepancy should also be explained. |
| S.7 | Legal basis | INST\_MANDATE | Law, code of practice or other formal set of instructions that gives an organization responsibility and authority to collect primary and administrative data, process and disseminate official statistical information |  |
| S.7.1 | Legal framework | INST\_MAN\_LA\_OA | Legal acts or other formal and informal agreements that impose responsibility on the organization and give it authority in relation to the collection of primary and administrative data, processing and dissemination of official statistical information | Laws, by-laws, orders of the Committee and other regulatory legal acts should be listed |
| S.8 | Privacy and data protection | CONF | A property of data that determines to what extent its unauthorized disclosure could harm the interests of its source or other relevant persons |  |
| S.8.1 | Privacy Policy | CONF\_POLICY | Legal measures or other formal procedures that directly or indirectly prevent unauthorized disclosure of data in relation to a certain person or entity | Specify the legal act regarding the confidentiality of primary statistical data |
| S.8.2 | Privacy - handling data | CONF\_DATA\_TR | Regulations applied to the handling of micro and macro data (including data in tabular form) to ensure the confidentiality of primary statistical data and prevent their unauthorized disclosure | The following should be described: - regulatory rules applicable to the handling of micro and macro data to ensure the confidentiality of primary statistical data, as well as procedures and conditions for ensuring confidentiality and data security; - data aggregation methods during processing (summation, weighting, basis for determining weights) |
| S.9 | Publication Policy | REL\_POLICY | Rules for the dissemination of official statistical information among all users |  |
| S.9.1 | Publication calendar | REL\_CAL\_POLICY | Schedule for the dissemination of official statistical information, annually approved (hereinafter - Schedule) | The principles for the publication of relevant official statistics should be described, indicating the Schedule and its availability. |
| S.9.2 | Access to the Graph | REL\_CAL\_ACCESS | Access to information contained in the Graph | You should provide a link where you can access the Chart |
| S.9.3 | User access | REL\_POL\_US\_AC | Publication principles for users, distribution limits, method of notifying users about the publication of official statistical information | Describe the conditions for access to official statistical information for various categories of users, information about the approval of the Schedule and its location, the number of publications (characteristic of the simplicity and ease with which a user can obtain official statistical information) |
| S.10 | Distribution frequency | FREQ\_DISS | Time interval in some given time period through which the dissemination of official statistical information is carried out | The frequency with which official statistical information is disseminated should be indicated (e.g. monthly, quarterly, yearly) |
| S.11 | Distribution format, accessibility and clarity | DISS\_FORMAT / ACCESS\_CLARITY | Media, various means and formats for the dissemination of official statistical information between users and their availability. Accessibility and clarity refer to the simplicity and ease of conditions and regimes under which users access official statistical information for their use and interpretation, with appropriate supporting information and support. |  |
| S.11.1 | News publication | NEWS\_REL | Regular or special data press releases | Describe regular or special data press releases |
| S.11.2 | Publications | PUBLICATIONS | Regular or special publications through which official statistical information is made available to the public | Specify the type of publications, where to find, access to dynamic tables. Indicate the formats for the dissemination of official statistical information and their description (characteristic of simplicity and ease of understanding by the user of official statistical information) |
| S.11.3 | Databases in on-line mode | ONLINE\_DB | Information about databases in on-line mode, through which access to disseminated official statistical information is provided | The on-line database should be described for the corresponding official statistical information (domain names of sites and a link to access the database on-line) |
| S.11.3. 1 | AC1. Data tables - consultations | DATATABLE\_CONSULT | Number of requests for tables within the statistical industry for a certain period of time | Number of on-line views or spreadsheet downloads |
| S.11.4 | Access to microdata | MICRO\_DAT\_ACC | Information on the possibility of distributing microdata | Description of the conditions and possibilities for accessing the dataset in the form of microdata (for example, for researchers). Also briefly describe the rules for anonymizing microdata |
| S.11.5 | Other | DISS\_OTHER | Links to other key official statistical information disseminated. | Description of the means by which other essential official statistical information is disseminated (eg in other publications, policy documents); review various aspects of dissemination practices and their impact on the availability and legibility of official statistics |
| S.11.5. 1 | AC 2. Metadata - consultation | METADATA\_CONSULT | Number of metadata hits within the statistical area in a given time period | By "number of hits" is meant how many times the metadata file has been viewed |
| S.12 | Availability of documentation | ACCESS\_DOC | The conditions and modes under which users receive, use, and interpret data documentation, that is, descriptive text used to define or describe an object, design, specification, instruction, or procedure |  |
| S.12.1 | Methodology Documentation | DOC\_METHOD | Descriptive text and links to available methodological documents | Providing links to access methodological materials |
| S.12.2 | Quality documentation | QUALITY\_DOC | Documentation of procedures applicable to quality management and quality assessment | Describe the availability of all documents related to quality |
| S.13 | Quality control | QUALITY\_MGMNT | Systems and structures in place in an organization to manage the quality of statistical products and processes |  |
| S.13.1 | Quality assurance | QUALITY\_ASSURE | All organized, systematic activities that demonstrate and provide confidence that processes meet statistical output requirements | Brief description of the overall quality assurance system used in the organization |
| S.13.2 | Quality control | QUALITY\_ASSMNT | Overall assessment of data quality based on standard quality criteria | Assessing the overall quality of statistical outputs by identifying major strengths and possible weaknesses (standard quality criteria). Any pros and cons of quality aspects should be noted, as well as planned measures to improve quality. |
| S.14 | Relevance | RELEVANCE | Extent to which official statistical information meets current and potential user needs |  |
| S.14.1 | User needs | USER\_NEEDS | Description of users and their respective needs for official statistical information | Classify users by importance: - indicate why they need statistical results; - evaluate what key results/indicators are required by users of different categories and any missing results for important users; - provide information about the unmet needs of users, specifying the reason; - describe plans to meet future needs; - give details of definitions that differ from requirements |
| S.14.2 | User Satisfaction | USER\_SAT | Criteria for determining user satisfaction | Description of the frequency with which user needs and opinions are collected (e.g. user satisfaction surveys, other user consultations), as well as the main results of the user satisfaction analysis (in the form of an indicator) and data from the most recent user satisfaction survey |
| S.14.3 | Completeness / R1. Data completeness - share | COMPLETENESS/ COMPLETENESS\_RATE\_U | The ratio of the number of data cells available to the number of data cells required (this indicator is only applicable if there is a regulation or regulations/guidance at the level of the European Statistical System) | Providing data on the completeness of official statistical information compared to the relevant regulatory requirements/instructions |
| S.15 | Accuracy and reliability (to be filled in according to the type of observation) | ACCURACY | Accuracy: The closeness of estimates or estimates to exact or true values that are estimated using statistical data. Reliability: the closeness of the originally estimated value to the value subsequently estimated |  |
| S.15.1 | Overall Accuracy | ACCURACY\_OVERALL | Accuracy score consisting of multiple components associated with a particular data set or a particular area | It is necessary to analyze the causes of errors and indicate the next steps to prevent them further. |
| S.15.2 | Sampling errors - indicators/ A1. | SAMPLING\_ERR/ SAMPLING\_ERR\_IND\_U | Sampling error (standard error, confidence intervals, coefficients of variation) | Relative Standard (Coefficient of Variation) Marginal (Value of Confidence Interval) Sample Standard Error |
| S.15.3 | Non-sampling error | NONSAMPLING\_ERR/ UNIT\_NONRESPONSE\_ RATE\_U/ ITEM\_NONRESPONSE \_RATE\_U | Non-sampling error in estimates based on questionnaires | Providing a user-friendly summary estimate of non-sampling error, unanswered rate and associated risks of bias (coverage error: over/undercoverage and different listings; measurement error: impact of survey instrument, respondents and interviewers (if applied ) ; non-response error: rate of response (non-response) of the respondent, including reasons for non-response and corresponding scores, rate of unanswered questions for core variables; processing error: editing, coding, and substitution errors (if applicable); hypothesis error about the model: special models used in the evaluation and measures taken to reduce errors of various types |
| S.15.3. 1 | Coverage error | COVERAGE\_ERR | Error due to insufficient coverage of all components of the population under study, which leads to a difference between the target population and the selected group | Coverage errors include overcoverage, undercoverage, and classification errors. Incomplete sampled groups often lead to coverage errors. Unit non-response, non-response to individual questions, description of % non-response for the period (for example, non-response of units of observation about 5% per year) |
| S.15.3. 1.1 | A2. Overreach - proportion | OVERCOVERAGE\_RAT E | Characteristic of the simplicity and ease with which a user obtains statistical data, which is determined by the physical conditions under which users access statistical data | Overcoverage error |
| S.15.3. 1.2 | A3. Common units - ratio | COMMON\_UNIT\_SHA RE | Ratio of units covered by both the survey and administrative sources, depending on the total number of units in the survey (typical for censuses and structural statistics) | Description of units |
| S.15.3. 3 | Non-response errors | NONPESPONSE\_ERR | Non-response errors |  |
| S.15.3. 3.1 | A4. Unit of absence - share | UNIT\_NONRESPONSE\_RATE\_P | Not answered by the respondent, proportion of the number of units without information or information not used to the total number of covered (eligible) units | Description of non-response units |
| S.15.3. 3.2 | A5. Point of non-response - share | ITEM\_NONRESPONSE\_RATE\_P | Not an answer to any question | Description of unanswered questions |
| S.16 | Timeliness and punctuality | TIMELINESS\_PUNCT | Timeliness is a criterion determined by the time interval between the date of appearance (publication) of official statistical information and the event or phenomenon that it describes. |  |
| S.16.1. | Timeliness | TIMELINESS/TIMELAG\_FINAL\_U | Timeliness is a criterion determined by the time interval between the date of appearance (publication) of official statistical information and the event or phenomenon that it describes. |  |
| S.16.1.1 | TP1.Waiting period - first results | TIMELAG\_FIRST | Number of days (weeks or months) from the last day of the reporting period to the day the first results are published | Indicate the time spent on production, measured from the end of the reporting period to the date of publication of the first results. If the information loses its relevance due to the time spent on its processing, ways to reduce the time lag should be developed. |
| S.16. 1.2 | TP2.Waiting period - latest results | TIMELAG\_FINAL\_P | Number of days (weeks or months) from the last day of the reference period to the day of publication of full and final results | Indicate the time spent on production, measured from the end of the reporting period to the date of publication of the final and complete results. If the information loses its relevance due to the time spent on its processing, ways to reduce the time lag should be developed. |
| S.16.2 | Punctuality | PUNCTUALITY / PUNCTUALITY\_RELEAS E\_P | Punctuality is a criterion determined by the delay time from the date of publication of official statistical information to the scheduled date |  |
| S.16.2.1 | Punctuality/TP3 | PUNCTUALITY / PUNCTUALITY\_RELEAS E\_U | Criterion due to the delay time from the date of publication of official statistical information to the planned day, i.e. deviation from the planned results | Indicate the period of time between the actual date of publication of official statistical information and the planned date, in accordance with the Schedule. It is necessary to analyze the causes of nonconformities and indicate the next steps to eliminate them. |
| S.17 | Comparability | COMPARABILITY | A criterion based on the degree of comparability of official statistical information over time, across regions, or across other areas of activity |  |
| S.17.1 | Geographic Comparability | COMPAR\_GEO | Comparability of official statistical information between geographic areas | Description of comparability problems between countries, possible causes of problems and ways to solve them |
| S.17.1.1 | Asymmetry according to mirror flow statistics - coefficient/CC1 | ASYMMETRY\_COEFF | Difference or absolute difference between incoming and outgoing flows between a pair of countries, divided by the average of these two values (common in trade, migration and balance of payments statistics) | A description of the difference or absolute difference between incoming and outgoing flows between a pair of countries, divided by the average of the two values (common in trade, migration and balance of payments statistics) |
| S.17.2 | Duration of comparable time series/CC2 | COMPAR\_TIME / COMPAR\_LENGTH\_U | Number of reporting periods, in time series since the last break | Description of comparability over time, reference periods in which time series breaks |
| S.18 | Consistency | COHERENCE | Criterion based on the ability to combine and share official statistical information from different sources |  |
| S.18.1 | Consistency external, cross | COHER\_X\_DOM | Cross-consistency | Description of the compliance of the methodology of the statistical process with accepted international standards |
| S.18.2 | Consistency internal | COHER\_INTERNAL | Consistency internal | Description of consistency with the results of other departments of the Committee and other departments |
| S.19 | Load | COST\_BURDEN | Respondent burden | Description: - degrees of automation of operations - collection, coding, verification, calculation, registration, distribution, and so on (indicate the following degrees: partially, fully automated, not automated); - time burden on respondents (indicate the average time spent on filling out the form); - duplication with other surveys (yes (specify), no); - use of administrative and other sources (yes (specify), no) |
| S.20 | Data revision | DATA\_REV | Change of values in previously approved and published official statistical information, indicating the reasons |  |
| S.20.2 | Data revision /A6 | REV\_PRACTICE / DATA\_REV\_AVGSIZE\_ U | "Revision" - making any changes to previously approved and published official statistical information (difference between late and early calculation of a key indicator). | Re-reports after approval and publication of official statistical information. Description of the main reasons for data revisions and their nature (availability of a new data source, new methods) |
| S.21 | Processing of statistical data | STAT\_PROCESS |  |  |
| S.21.1 | Initial data | SOURCE\_TYPE | Characteristics and components of source statistics that are used to define statistical aggregates | Indicate whether the data set is based on questionnaires, on administrative data sources, on a combination of several data sources, or on data obtained from other statistical activities. If sampling is used, the sampling characteristics should also be given: - sampling procedure (design), describe sampling for sample surveys, eg one-stage cluster sampling (random sampling); - describe the sample size, sample proportion and sample units; - distribution to the general population (weighing, calculation of weights); - description of sampling rotation methods. Description of administrative sources (if applicable) |
| S.21.2 | Survey frequency | FREQ\_COLL | Frequency at which baseline data is collected | Specify the frequency of data collection (monthly, yearly). |
| S.21.3 | Method (way) of collecting primary statistical data | COLL\_METHOD | Systematic procedure for collecting primary statistical data for official statistics. | Description of the method used to obtain primary statistical data from respondents (mail, courier, registration, interview, on-line mode) |
| S.21.4 | Reliability of primary statistics | DATA\_VALIDATION | The process of monitoring the results of collecting primary statistical data and ensuring the quality of statistical results | Description of the procedure for checking the reliability of initial and output data and ways to control the results of such checks. Validation may include the following: checking that the coverage of the population and the proportion of participating respondents are consistent with what is required; comparison of statistical processes with previous cycles (if applicable); comparison of primary statistics with other relevant data (both own and third-party); analysis of inconsistency of statistical data; editing micro and macro data; comparing statistical data with expectations and information about a given subject area, identifying "outliers" |
| S.21.5 | Imputation - share / A7 | DATA\_COMP | The process of replacing missing, incorrect, or invalid values with other values, the number of replacements to the total number of values | Description of the procedure (method) of imputation. |
| S.21.6 | Adjustment | ADJUSTMENT | A set of procedures used to modify statistical data to bring them into line with national or international standards, or to eliminate differences in data quality when certain data sets are compiled | Description of the time series to be adjusted and the statistical procedures used to adjust the time series (eg seasonal adjustment methods, time series decomposition). Specify the type of adjustment (for example, seasonally adjusted, calendar, trend cycle) and, if applicable, the calendar used. If "outliers" are detected and replacement is performed, indicate which "outliers" are detected (impulse, temporal changes, level shifts). Description of the software and its version used for correction |
| S.21.6. 1 | Seasonal adjustment | SEASONAL\_ADJ | Smoothing price and seasonal fluctuations (if applicable) | Describe the procedure for seasonal smoothing, indicate for which indicator, the name of the software package, the number of values in the time series |
| S.22 | Remarks | COMMENT\_DSET | Final part | Conclusions, recommendations, conclusions, notes |
|   | Appendix 2 to the Methodology for assessing the quality of official statistical information |

**List of standard quality indicators**

|  |  |
| --- | --- |
| Quality criterion | Index |
| Accessibility and clarity | AC1. Data tables - queries |
| AC2. Metadata - hits |
| Relevance | R1. Data completeness - share |
| Accuracy | A1. Sampling error - indicators |
| A2. Overreach - share |
| A3. Common units - ratio |
| A4. Unit of non-response - fraction |
| A5. Point of non-response - share |
| A6. Data revision - medium size |
| A7. Imputation - share |
| Timeliness and punctuality | TP1. Waiting period - first results |
| TP2. Waiting period - latest results |
| TP3. Punctuality - delivery and publication |
| Comparability and Consistency | CC1. Asymmetry according to mirror flow statistics - coefficient |
| CC2. Length of comparable time series |

1) "data completeness" - the ratio of the number of data cells available to the number of required data cells - R1 (this indicator is only applicable if there is a regulation or norms/guidelines at the level of the European Statistical System);

2) "sample error" - relative standard error (error), coefficient of variation or confidence interval (except for statistical registers) - A1;

3) "exceeding coverage" - the proportion of units obtained from the general population, but not belonging to the sample population - A2;

4) "common units" - the ratio of units covered by both the survey and administrative sources, depending on the total number of units in the survey (typical for censuses and structural statistics) - A3;

5) "unit of non-response" - the ratio of the number of units that do not have information or have inapplicable information to the number of covered (eligible) units - A4;

6) "point of non-response" - the ratio of covered units that did not respond to a specific item to the total number of covered units - A5;

7) "revision of data" - making any changes to previously approved and published official statistical information (difference between late and early calculation of a key indicator) - A6;

8) "imputation" - the number of replaced values to the total number of values - A7;

9) "waiting period - first results" - the number of days (weeks or months) from the last day of the reporting period to the day of publication of the first results - TP1;

10) "waiting period - latest results" - the number of days (weeks or months) from the last day of the reporting period to the day of publication of full and final results - TP2;

11) "punctuality - delivery and publication" - a criterion due to the delay time from the date of publication of official statistical information to the planned day, that is, deviation from the planned results - TP3;

12) "asymmetry in mirror flow statistics" - the difference or absolute difference between incoming and outgoing flows between a pair of countries, divided by the average of these two values (typical for trade, migration and balance of payments statistics) - CC1;

13) "duration of comparable time series" - the number of reporting periods, in the time series since the last break - CC2;

14) "data tables" - the number of requests for data tables within the statistical industry for a certain period of time (number of on-line views or table downloads) - AC1;

15) "metadata - hits" - the number of hits on metadata within the statistical industry for a given period of time. By "number of requests" is meant how many times the metadata file was viewed - AC2.

© 2012. RSE on REM Republican Center for Legal Information of the Ministry of Justice of the Republic of Kazakhstan