Food and Agriculture Organization of the United Nations
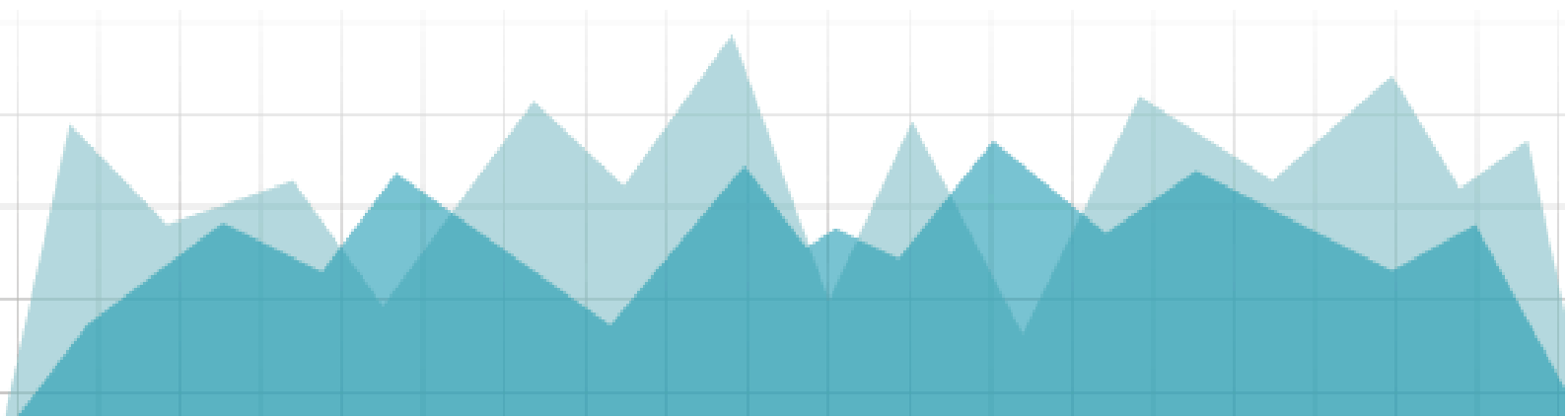
Statistical Standard Series

Metadata dissemination for FAO statistical databases
Version 2

This document provides guidance to support standardization and harmonization of web dissemination of reference metadata describing how various statistical outputs and processes were produced. It includes a set of concepts, recommendations and guidelines that can help users better understand the quality of the statistical data published by the Organization in its statistical databases.

The document was endorsed as FAO standard by the Technical Data Coordination Group (DCG-T) on 23 June 2023.

Contents

Acronyms

MSD   Metadata Structure Definition

QAPS   Quality Assessment and Planning Survey

SDMX   Statistical Data and Metadata eXchange

SDW   Statistical Data Warehouse

SDQAF   Statistics and Data Quality Assurance Framework

Background

The dissemination of statistical data and associated metadata is a core element of FAO's mandate and an important responsibility of the Organization. The publication of comprehensive metadata is one essential element of the overall requirements for improved comparability and consistency of statistics. Quality evaluation instruments, such as the Quality Assessment and Planning Survey (QAPS) and external user feedback collected through user consultations about FAO databases, underlined the need of making comprehensive metadata available in a consistent structure and format.

As stated in the Fundamental Principles of Official Statistics[1] and the FAO Statistics Quality Assurance Framework (SDQAF)[2], the dissemination of metadata is an integral part of accessibility and clarity, quality dimensions that require an adequate documentation along all stages of the statistical production cycle. The present document defines the standards to be followed in preparing and disseminating reference metadata for datasets in statistical databases and recommends a reporting framework on the statistical process across FAO's statistical units. This standard incorporates general guidelines and recommendations that cover all quality dimensions defined in the FAO SDQAF. The current version has been reviewed to include quality assurance provisions for the use of non-statistical data sources (including Big Data)[3] in the production of FAO statistics, and includes the requirement to communicate on this matter to users. The version 2 of this Standard therefore provides more guidance on what should be communicated to users in the reference metadata, when non-statistical data sources were used in the generation of the statistical outputs

As part of the continuing FAO's efforts to ensure the dissemination of high-quality and internationally comparable statistical data, the implementation of this standard will promote the compliance of statistical processes at corporate level with the quality principles and the harmonization of FAO's statistical outputs. The dissemination of metadata will maximize the added value of FAO's statistical work by enabling broader and deeper research and analysis. It will also add transparency and credibility to FAO's analytical work.

The standard proposes a common reference metadata structure for datasets included in FAO statistical databases, which consists of concepts and definitions compliant with the international standard for Statistical Data and Metadata eXchange (SDMX, 2021a). This will facilitate the migration of FAO databases in FAO Statistical Data Warehouse (SDW) while supporting the modernization of the statistical processes across all divisions as well as the integration and harmonization of the different statistical dissemination platforms.

This standard applies to datasets disseminated through FAO databases (e.g. FAOSTAT, AQUASTAT) and excludes reference metadata accompanying other types of data products, for instance microdata files or data contained geospatial platforms or dashboards. These will be covered by other Statistical Standards.

---

[1] https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
[2] http://www.fao.org/3/cc6683en/cc6683en.pdf
[3] **Non-statistical data sources (or non-traditional data sources) refer to** all data sets that are not created primarily for official statistical purposes but rather for administrative, commercial or other private purposes. In general, non-statistical data sources include administrative data sources and big data that have not yet being processed and transformed into a statistical outputs.

1. Technical terms and definitions

The technical terms and definitions are mainly taken from common statistical terminology and the SDMX glossary (SDMX, 2021b).

Metadata are the set of information that describe a particular dataset and the statistical processes used to generate the data contained therein. The metadata documentation ensures a proper use interpretation of the data by its owners and users. Two broad types of statistical metadata can be distinguished: structural metadata and reference metadata.

Structural metadata are essential information needed to identify and process the data. They contain identifiers and descriptors of the data, including the name of variables, as well as their unit of measurement, code lists, data formats, frequency and time period, flag system and classifications used.

Reference metadata describe the content and quality of the statistical data. They should include: "conceptual" metadata describing the concepts used and their practical implementation; "methodological" metadata describing sources and methods used for the data generation (e.g. sampling, collection methods, editing processes); "quality" metadata, describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy). They can also be associated with different levels of data: entire collections, data sets from a given country, or a data item concerning one country and one year.

Reference metadata can be linked to diverse objects at different levels of generality (e.g. time series or observations). Nonetheless, they are often attached to high level statistical outputs (i.e. at data set or statistical domain level, or even at agency level).

Metadata producer is the officer who, on behalf of the technical unit, is responsible for providing the metadata of the statistical process, being its data owner. This officer serves as the focal point for receiving and replying the users' requests.

Data user is an individual or an institution, which accesses the published dataset for satisfying various needs. Data users can be classified in many different types with different needs (students/experts, academic institutions, government officials, media outlets, general public, etc.). The metadata documentation should cater to the needs of all these different types of data users to help them understand the content of the dataset and assess its suitability for the intended use.

2. Scope

The standard defines metadata to describe in a structured and systematic way the FAO statistical processes and outputs. It recommends the adoption of a reference metadata structure that incorporates core concepts and sub-concepts needed by the users to understand what the data outputs are and how they were compiled by FAO, in sufficient detail for the users to determine if they satisfy their needs.

The structure of the metadata should be filled in with information regarding the relevant FAO internal processes used to compile the datasets, starting from the moment when the data are submitted to or acquired by FAO. For instance, the metadata structure should include a description on the data imputation and editing processes implemented on the data submitted by the national statistical agencies, as recommended in the FAO statistical standards on "Imputation"[4] and "Data editing"[5]. However, it does not cover the statistical processes conducted at national level, such as the characteristics of the survey (e.g. data collection instrument, methodology, census, sampling error) used to compile the national estimates submitted to FAO. For non-statistical data sources used in the FAO statistical production process, the metadata structure should provide information about data access (e.g. full access to raw data, or access to pre-processed data by FAO or through third party), the rationale behind the use of non-statistical data sources (i.e. what are their value-added?), and the methods used to process the data, in accordance with the FAO statistical standard on the acquisition and use of non-statistical data sources (including big data) for statistical purpose[6].

The metadata structure should also contain a description of the quality of the FAO statistical process and outputs, using the quality dimensions of the FAO SDQAF: (i) relevance; (ii) accuracy and reliability; (iii) timeliness and punctuality; (iv) coherence and comparability; and (v) accessibility and clarity.

The metadata structure enables the metadata producer to report on the measurement of quality indicators, as recommended by the "Statistical Standard Series of Quality Indicators for External Users"[7]. This information supports FAO's efforts to provide external users with a transparent quality assessment of the published datasets; it also enables metadata producers to establish improvement plans to address the main quality issues identified.

---

[4] https://www.fao.org/3/cb9339en/cb9339en.pdf
[5] https://www.fao.org/3/cb9341en/cb9341en.pdf
[6] Under development as of June 2023
[7] https://www.fao.org/3/cb9309en/cb9309en.pdf

3. Description of the metadata structure

Many metadata schemes have been developed over time, depending on the data typology and the business requirements. For instance, FAO adopted the Data Documentation Initiative (DDI, 2021) as corporate reference metadata in the context of microdata (FAO Statistical Standard on Microdata dissemination[8]) and ISO and CKAN Standards for the dissemination of geospatial information and data disseminated through geospatial platforms respectively.

This standard on metadata dissemination has been developed to document FAO datasets in statistical databases which typically contain country-level time series, and it adopts a metadata structure consistent with the content-oriented Guidelines of the Statistical Data and Metadata Exchange (SDMX, 2021c), which were approved by ISO as an International Standard (ISO 17369:2013) in 2013. The compliance with SDMX enables the interoperability between internal and external data dissemination platforms and supports the exchange, reporting and dissemination of statistical data and related metadata.

The structure consists of 20 metadata concepts and 52 sub-concepts, SDMX-compliant, that cover core elements of reference metadata and data quality, such as accuracy and reliability, timeliness and punctuality, coherence and comparability, relevance, etc. (see Annexes 1 and 2).

Most of the concepts and sub-concepts in the metadata structure are SDMX cross-domain concepts (SDMX, 2021b and SDMX, 2021c), while some other concepts (e.g. COHER_COMPAR: Comparability and Coherence; ACCESS_CLARITY: Accessibility and clarity) are derived from the Eurostat Single Integrated Metadata Structure (Eurostat, 2021). This approach makes the concepts internationally comparable, the metadata machine-readable for automatic discovery of datasets and services, and thus the data interoperable and ready for re-use.

The use of standardized concepts pave the way for developing and managing a standard Metadata Structure Definition (MSD) compliant with the SDMX version 2.1, which will be an integral component of the FAO Statistical Data Warehouse dissemination portal. The MSD refers to the specification of the concepts required in terms of their attributes, how these relate to each other, their presentational structure and to which objects they should be connected.

This standard provides guidelines meant to assist the metadata producers in reporting on each concept (Annex 2). The guidelines include examples to facilitate the interpretation and applicability of the concept. The combined use of technical standards and content guidelines underpins the metadata standardization and ensures consistent documentation across FAO statistical data sets and dissemination facilities. This also supports the use of SDMX-based dissemination codes such as the observation status codes (FAO Statistical Standard on Observation Status Code List[9]).

---

[8] https://www.fao.org/3/cb9336en/cb9336en.pdf
[9] https://www.fao.org/3/cc6208en/cc6208en.pdf

4. Technical recommendations

- As recommended by the FAO SDQAF Principle 16 on accessibility and clarity, **FAO statistics and corresponding metadata should be made accessible and presented in a standard format that facilitates proper interpretation, meaningful comparisons and automatic machine-processing in line the open data format.** This will improve the findability, accessibility, interoperability, and reuse of FAO statistical outputs.

- **The metadata producer should complete, whenever relevant, the metadata structure to the maximum extent possible, in line with the guidelines presented in Annex 2.** When a concept is not relevant to a given statistical process and/or output, it should be noted as "Not applicable".

- **The metadata content should be kept as stable as possible.** It should be updated only when changes are introduced in the statistical process, mainly due in particular to modifications of national data submissions. For instance, some datasets require frequent updates of metadata, but only for some specific concepts (e.g. timeliness, data validation).

- **Any reference to international or corporate statistical standards should be properly referenced in the metadata description.** For instance, reference should be made to the country codes or unit of measure, so that metadata can be aligned across datasets. The compliance with these standards should be briefly described.

- **One important element of the metadata documentation (in particular "16.2 Quality assessment") is the measurement of quality indicators**, as documented in FAO Statistical Standard on Quality indicators to be disseminated to external users[10]. A weblink to the standard can also be provided.

- For some specific metadata concepts, the use of a standard text across all statistical datasets, such as the one related to the Institutional mandate and data sharing arrangements, is recommended. References to the use of a standard text are highlighted in the guidelines (Annex 2).

---

[10] https://www.fao.org/index.php?id=115332

5. Governance

Each statistical unit should ensure the implementation of this standard and is responsible for the production of metadata across all the datasets it disseminates. The establishment of a corporate SDW, i.e. a unified data dissemination portal, would streamline the implementation of this standard metadata structure.

This metadata structure is expected to be relatively stable over time. It may evolve, for instance, by changing or merging concepts, expanding the guidelines with reference to new Statistical Standard Series and adding new quality indicators.

Proposals for extensions and revisions of this standard and its guidelines should be submitted for approval to the **Office of Chief Statistician** and copied to the **Technical Data Coordination Group (DCG-T) on Data for Statistics and Statistics**, which will be responsible for its review and endorsement.

The **Office of Chief Statistician** will be responsible for updating this standard accordingly and will document the different versions of the document in an annex.

Annexes

Annex 1: List of concepts and corresponding ID in Statistical Data and Metadata eXchange

Code	Concepts	Concept ID
1	Title	TITLE
2	Data source	DATA_SOURCE
3	Contacts	CONTACT
4	Metadata update	META_UPDATE
5	Statistical presentation	STAT_PRES
6	Unit of measure	UNIT_MEASURE
7	Reference period	REF_PERIOD
8	Institutional mandate	INST_MANDATE
9	Confidentiality	CONF
10	Release policy	REL_POLICY
11	Frequency of dissemination	FREQ_DISS
12	Accessibility and clarity	ACCESS_CLARITY
13	Relevance	RELEVANCE
14	Accuracy and reliability	ACCURACY
15	Timeliness and punctuality	TIMELINESS_PUNCT
16	Coherence and comparability	COHER_COMPAR
17	Statistical process	STAT_PROCESS
18	Quality management	QUALITY_MGMNT
19	Recommended uses and limitations	REC_USE_LIM
20	Comment	COMMENT

Annex 2: Concepts and guidelines of the reference metadata structure

(*) denotes concepts identifiers that are derived from the Eurostat reference metadata standard: Single Integrated Metadata Structure. All other concepts are from the SDMX glossary (SDMX, 2021b).

Code	Concepts	Sub-concepts	Identifier	Guidelines
1	Title		TITLE	Textual label used to refer to the reference metadata structure. It can be used as a semantic name.
2	Data source		DATA_SOURCE	Location or service from where data and/or metadata can be obtained. The location includes a resolvable URL which may download a file or be a query that returns data. Data citation (citation text) or DOI (Digital Object Identifier) may also be included alongside the URL.
2.1	Data source	Compiling Organization	COMPILING_ORG	Organization collecting and/or elaborating the data being reported. For instance, if the FAO collects data from Member States or/and various data sources, process and disseminate them, the compiling organization should be FAO. If the data was collected and already disseminated by another Organization, but acquired, minimally processed (e.g. through simple data manipulation to fit our FAO data dissemination structures) and re-disseminated by FAO, the compiling organization should be this other Organization.
3	Contacts		CONTACT	Organizational contact points for the data or metadata, including information on how to reach the contact points.
3.1	Contacts	Contact organization	CONTACT_ORGANISATION	Food and Agriculture Organization of the United Nations (FAO)
3.2	Contacts	Contact organization unit	ORGANISATION_UNIT	Name of the unit, branch or department responsible for the metadata file.
3.3	Contacts	Contact mail address	CONTACT_MAIL	Postal address of the unit responsible of the statistical domain.
3.4	Contacts	Contact email address	CONTACT_EMAIL	Generic e-mail address of the unit responsible for the statistical domain.
4	Metadata update		META_UPDATE	Date on which the metadata element was created or modified.

Code	Concepts	Sub-concepts	Identifier	Guidelines
4.1	Metadata update	Metadata last posted	META_POSTED	Date of the latest dissemination (last posting) of the metadata.
4.2	Metadata update	Metadata last update	META_LAST_UPDATE	Date of last update of the content of the metadata. The update can concern one single concept, but also the metadata file as a whole.
5	**Statistical presentation**		STAT_PRES*	(Information relating to this concept is provided by reporting on its sub-concepts.)
5.1	Statistical presentation	Data description	DATA_DESCR	Main characteristics of the data set described in an easily understandable manner, referring to the data and indicators disseminated. This summary description should provide an immediate understanding of the data to users (also to those who do not have a broader technical knowledge of the dataset in question). Data description can include the main variables covered.
5.2	Statistical presentation	Coverage	COVERAGE	Definition of the scope of the data compiled. In particular, this sub-concept should be used to describe the population covered by the statistics (including eventually its dimensions or levels of disaggregation such as income, sex, age group, disability status, etc.) as well as the geographical coverage of the data series.
5.3	Statistical presentation	Sector coverage	COVERAGE_SECTOR	Main economic or other sectors covered by the statistics. It lists groups and subgroups of related activities or populations covered by the data set produced. These sectors can be institutional sectors, economic or other sectors (e.g. agriculture, forestry, or aquaculture).
5.4	Statistical presentation	Statistical concepts and definitions	STAT_CONC_DEF	List of definitions and descriptions of the main statistical variables provided. There should be reference to FAO and internationally accepted statistical standards, guidelines, or good practices on which the concepts and definitions that are used for compiling the statistics are based (e.g. area under cultivation, species and harvested production). When relevant, deviations from statistical standards, guidelines or good practices, should be documented. In particular, deviations from international concepts and definitions caused by the use of a particular data source (e.g. a Big data source) should be explained.
5.5	Statistical presentation	Reference area	REF_AREA	Country or geographical or political group of countries or regions to which the measured statistical data set relates. Reference to the codelist of the FAO Statistical Standard on Country and Area Codes for Statistical Use[11].
5.6	Statistical presentation	Time coverage	COVERAGE_TIME	Period of time for which data set is described. The time period covered can be indicated as a time interval, e.g. "1985 to 2006" for annual time series data, or as several intervals or values of time.

---

[11] http://intranet.fao.org/statistics_coordination_portal/standards_for_quality_compliance/ Accessible internally only

Code	Concepts	Sub-concepts	Identifier	Guidelines
5.7	Statistical presentation	Base period	BASE_PER	Period of time used as the base of an index number, or to which a constant time series refers. It may not be applicable for many statistical data sets, but for some others a base period is chosen as a benchmark, indices can be applied to current values (e.g. base year 2000 for certain annual data).
6	Unit of measure		UNIT_MEASURE	The unit in which the data values are measured. The unit of measure in connection with the unit multiplier, provides the level of detail for the value of the variable (e.g. the following units are use: USD, national currency, tonne). The magnitude (e.g. thousands, millions) of numerical units should be included. Include reference to the FAO Statistical Standard on units of measure[12]. When multiple units of measure are used within the data domain, the UoM should be specified by data items
7	Reference period		REF_PERIOD	Timespan or point in time to which the measured observation is intended to refer, which can be a specific day or a specific period (e.g. a month, a fiscal year, a calendar year or several calendar years). In many cases, the reference period and time period will be identical, but there are also cases where they are different. This can happen if data are not available for the target reference period, but are available for a time period which is judged to be sufficiently close. For example, the reference period may be a calendar year, whereas data may only be available for a fiscal year. In such cases, "reference period" should refer to the target reference period rather than the actual time period of the data. The difference between target and actual reference period can be highlighted in a free text note.
8	**Institutional mandate**		INST_MANDATE	Set of rules or other formal set of instructions assigning responsibility as well as the authority to FAO to an organization (i.e. FAO) for the collection, processing, and dissemination of statistics. It also includes arrangements or procedures to facilitate data sharing and coordination between data producing agencies. **Use this standard text.** Article I of the FAO Constitution requires the Organization to collect, analyze, interpret and disseminate information relating to nutrition, food and agriculture (FAO, 2017)[13].

---

[12] https://www.fao.org/3/cc6207en/cc6207en.pdf
[13] D. Constitution of the food and agriculture organization of the united nations (fao.org)

Code	Concepts	Sub-concepts	Identifier	Guidelines
8.1	Institutional mandate	Legal acts and other agreements	INST_MAN_LA_OA	Legal acts or non-legal measures such as formal or informal agreements administrative arrangements employed to specific organizations that assign responsibility as well as the authority to FAO or an international institutions for the collection, processing, and dissemination of statistics. If available, a weblink should be provided. **Use a standard text, if available.** For example, in line with SDQAF Principle 7 on Coordination with other international organizations producing statistics, FAO collects this data with Eurostat in accordance with the Memorandum of Understanding (MoU) on agriculture statistics that was signed in January 2020.
8.2	Institutional mandate	Data acquisition and data transmission	INST_MAN_ACQ_TRANS	Arrangements, procedures or agreements in place for the acquisition, access and transmission of the data used to generate these data outputs. If relevant, a reference to the data sharing agreement or the terms of use of the data sourced should be provided as well as a weblink to these if publicly available If data accessed and transmitted include personal data, a reassurance that the data were accessed, managed and transmitted in accordance with FAO Data protection policy should be provided. If non-statistical data sources were used, the following should be described: • Modes of data access (full access to raw data or access to pre-processed data by FAO or through third party) • In case of access to pre-processed data by third-party/data provider: information about the methods applied to the raw data resulting in the pre-processed data • Time and method of transmission • Time horizon of the cooperation - Is a long-term access to the data guaranteed?" **Use a standard text, if possible.** For example, EU member country production data for the primary commodities are obtained from Eurostat (with some minor exceptions) since 2018, in accordance with the Memorandum of Understanding (MoU) between Eurostat and FAO on agriculture statistics
9	**Confidentiality**		CONF	Property of data indicating whether they are subject to dissemination restrictions. Data are protected by confidentiality in cases where unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.

Code	Concepts	Sub-concepts	Identifier	Guidelines
9.1	Confidentiality	Confidentiality - policy	CONF_POLICY	Legislative measures, rules or other formal procedures related to statistical confidentiality, which prevent unauthorized disclosure of data that identify a person or economic entity either directly or indirectly: It should provide the assurance that all necessary methods assuring confidentiality have been applied to the data. **Use a standard text, if possible.** FAO Data and statistics are produced in accordance with FAO Policy on Data Protection and the implementation modalities of FAO SDQAF Principle 4 on Data protection and Statistical Confidentiality..
9.2	Confidentiality	Confidentiality - data treatment	CONF_DATA_TR	Rules applied for treating the data set with regard to statistical confidentiality to ensure that private information from individual units cannot be accessed and to prevent unauthorized disclosure (e.g. cell suppression, aggregation of sensitive information, aggregation rules on aggregated confidential data). **Use a standard text, if possible.** If any, confidential data are suppressed and not disseminated by FAO. They are marked in the database with SDMX flag "Q - Missing value; suppress". To avoid residual disclosures, aggregates are also suppressed as necessary.
9.3	Confidentiality	Privacy	CONF_PRIVACY	Discuss how privacy sensitive is the information used to produce this or these statistical output data. State which treatments were applied to satisfy privacy concerns. In the case of personal information collected and acquired by FAO, describe how it complies with the organization's Data Protection Policy[14] throughout the data lifecycle, including obtaining/renewing the informed consent of data provider and deleting the data once the purpose for which it was collected is achieved. When data is coming from external data providers and contains personal information, state what steps have been taken to ensure that the third parties' security measures for processing and/or transferring the data are at least as comparable to those that are required for data of the same confidentiality classification in FAO.
10	**Release Policy**		REL_POLICY	Rules for disseminating statistical data to users and all interested parties.
10.1	Release policy	Release calendar	REL_CAL_POLICY	Schedule of statistical release dates. The policy regarding the release of statistics according to a preannounced schedule should be described. It should also be mentioned if a release calendar for the data set in question exists and if this calendar is publicly accessible.

---

[14] https://intranet.fao.org/fileadmin/user_upload/FAO_Communications/ac/1_AC2022-06-v2.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
10.2	Release policy	Release calendar access	REL_CAL_ACCESS	Description of how the release calendar can be accessed. The release date of the statistical output(s) should be included in FAO Data Release Calendar and a weblink to this calendar should be provided to access it (https://www.fao.org/fileadmin/user_upload/faoweb/statistics/Data_Release_Calendar/FAO_Data_release_calendar.xlsx
10.3	Release policy	User access	REL_POL_US_AC	Policy for release of the data to users, scope of dissemination (e.g. to the public, to selected users), how data users are informed that the data are being released, and whether the policy allows the dissemination of statistical data to all users. **Use a standard text, if possible.** The data are published in compliance with the SDQAF, Principle 14 on timeliness and punctuality, Principle 16 on accessibility and clarity, and Principle 1 on professional independence and impartiality. Moreover, data are disseminated according to the Open Data Licensing for Statistical Databases Policy (https://www.fao.org/3/ca7570en/ca7570en.pdf), under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO license (CC BY-NC- SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo). In addition to this license, some database specific terms of use are listed here: Terms of use of datasets (http://www.fao.org/contact-us/terms/db-terms-of-use/en).
10.4	Release policy	Open data license	REL_SPDX_CODE	This is a field introduced by FAO to describe the standard license in a machine readable format (based on standard coding systems, in particular SPDX ISO/IEC 5962:2021). For all databases listed under Open Data Licensing for Statistical Databases Policy, this field should be: https://creativecommons.org/licenses/by/3.0/igo/legalcode
11	**Frequency of dissemination**		FREQ_DISS	Time interval at which the statistics are disseminated over a given time period. The frequencies with which data are released (e.g. monthly, quarterly, yearly), which could be different from the frequency of data collection.
12	**Accessibility and clarity**		ACCESS_CLARITY*	Accessibility refers to the set of conditions and modalities by which users can obtain, use and interpret data. Clarity refers to the availability of various dissemination formats and adequate documentation: whether data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality are also available (including limitation in use), and the extent to which additional assistance is provided.

Code	Concepts	Sub-concepts	Identifier	Guidelines
12.1	Accessibility and clarity	News release	NEWS_REL	Regular or ad-hoc press releases or other kind of similar releases linked to data or metadata.
12.2	Accessibility and clarity	Publications	PUBLICATIONS	Regular or ad-hoc publications in which the data are made available to the public. This includes references to the most important data dissemination done through paper or online publications (e.g. yearbook), including a summary identification and information on availability of the publication means.
12.3	Accessibility and clarity	Online database	ONLINE_DB	Information and weblink to the on-line database in which the disseminated data can be accessed. As per FAO Statistical Standard on Quality indicators for external users[15], quantitative measures of accessibility of the data (in terms of number of visits and downloads) should be provided here.
12.4	Accessibility and clarity	Micro-data access	MICRO_DAT_ACC	Information on whether microdata are also disseminated which applies to only some statistical process and data sets. Description of if and how the data set is accessible as micro-data (e.g. weblink to the Food and Agriculture Microdata catalogue: http://www.fao.org/food-agriculture-microdata/en/).
12.5	Accessibility and clarity	Other formats	DISS_OTHER	References to the other main data dissemination outlets. This may include other publications, policy papers, and also as a sub-element, "Supplementary data", i.e. any customized tabulation that can be provided to meet specific requests (including information on procedures for obtaining access to these data).
12.6	Accessibility and clarity	Documentation on methodology	DOC_METHOD	Descriptive text and/or references to methodological documents available to better understand the methods used to produce the statistical output(s) and the resulting data. In particular, list reference metadata files, methodological papers, summary documents and handbooks relevant to the statistical process(es) applied and statistical output(s) produced.
13	Relevance		RELEVANCE	Degree to which the statistics produced, including through the use new data sources, meet the current and potential data user needs. The potential added value of a new data source (most probably a non-statistical data source) to an existing statistical product is thus an important element to highlight (e.g. detailed data on particular sub-groups, or information on grid-level instead of district level or potential replacement of questions of a survey through information of the new data source). Relevance also refers to the processes for monitoring the relevance and practical usefulness of existing statistics in meeting users' needs and how these processes impact the development of statistical programmes.

[15] https://www.fao.org/3/cb9309en/cb9309en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
13.1	Relevance	User needs	USER_NEEDS	Description of requirements with respect to the statistical output. With respect to the statistical data to be provided, the main data users (e.g. official authorities, the general public or others) and user needs should be stated e.g. official authorities with the needs for policy indicators; e.g. the World Trade Organization uses the data for its trade policy review, etc.). The information is typically collected through user satisfaction surveys, user consultations and feedback from domain specific working groups. If available, a classification of users should be reported indicating their relative importance and an assessment of the key outputs desired by different categories of users.
13.2	Relevance	User satisfaction	USER_SAT	Description of how well the disseminated statistics meet the expressed user needs. If user satisfaction surveys are conducted, the main results should also be reported (in the form of a user satisfaction index if available). The date of the most recent user satisfaction survey should also be mentioned. As per FAO Statistical Standard on Quality indicators for external users[16], a quantitative measure of users' satisfaction should be provided here (average proportion of satisfied users). Any additional indication or measure to report on user satisfaction could also be reported here with reference to FAO Statistical Standard on User consultation[17].
13.3	Relevance	Completeness	COMPLETENESS	Extent to which all statistics that are needed are available. This normally compares the "ideal" data set required by the main users to the available one. As per FAO Statistical Standard on Quality indicators for external users[18], a quantitative measure of geographic completeness should be provided here
14	Accuracy and reliability		ACCURACY	Accuracy refers to the closeness of estimates to the true values that statistics were intended to measure. Reliability refers to the closeness of the initial estimates to the subsequent or final estimates.

[16] https://www.fao.org/3/cb9309en/cb9309en.pdf
[17] https://www.fao.org/3/cb9340en/cb9340en.pdf
[18] https://www.fao.org/3/cb9309en/cb9309en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
14.1	Accuracy and reliability	Overall accuracy	ACCURACY_OVERALL	Assessment of accuracy linked to a certain data set or domain. This metadata element is used to describe the main sources of random or systematic error (e.g. measurement error, non-response error) in the statistical outputs which may be generated in any phase of the FAO statistical process (FAO data collection, treatment, processing and dissemination), with special focus on the impact on key estimates at the FAO corporate level. It should provide a qualitative and quantitative assessment of the variability and potential bias (sign and order of magnitude) associated to each key indicator produced and disseminated by FAO. As per FAO Statistical Standard on Quality indicators for external users[19], applicable quantitative measures of accuracy should be provided here (for instance, questionnaire reporting rate, weighted questionnaire reporting rate, percentage of missing data and percentage of official data).
14.2	Accuracy and reliability	Model assumption error	MODEL_ASSUMP_ERR	Error due to adoption of statistical models needed to estimate a variable or an indicators that will contribute to the main statistical outputs of the process. This concept focuses on errors that occur are typically those related to not holding of the key assumptions underlying the model (including the risk of application of not fully adequate models e.g. a linear model when the relationship is not-linear). All FAO processes where production of final regional/global estimates is largely based on the application of statistical models should describe the potential model assumption errors and assess their impact on the accuracy of results.
14.3	Accuracy and reliability	Imputation indicators	IMPUTATION_RATE	Ratio of the number of replaced values to the total number of values for a given variable. Imputation should be documented with reference to FAO Statistical Standard on imputation[20]. If there is missing data, give detailed description of the methods used for imputation. For non-statistical data sources, indicate the reason why data were not collected (e.g. selection bias, technical issues, etc.). Imputation indicators (i.e. percentage of imputed values and contribution of imputed values to totals) should be reported in line with FAO Statistical Standard on Quality indicators for external users[21].
14.4	Accuracy and reliability	Data revision - policy	REV_POLICY	Policy aimed at ensuring the transparency of disseminated data, whereby preliminary data are compiled that are later revised. Description of the general guidelines for handling data revisions applied by FAO or the statistical unit.

---

[19] https://www.fao.org/3/cb9309en/cb9309en.pdf
[20] https://www.fao.org/3/cb9339en/cb9339en.pdf
[21] https://www.fao.org/3/cb9309en/cb9309en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
14.5	Accuracy and reliability	Data revision - practice	REV_PRACTICE	Information on the data revision practice. This provides documentation regarding the source data used and the way they are adjusted, in order to give compilers the possibility of incorporating new and more accurate information into estimates, thus improving their accuracy without introducing breaks in the time series. It also describes the revision status of available data. Data may also be subject to regular or ad-hoc revisions as a result of the introduction of new classifications, compilation frameworks and methodologies which result in the compilation of historical data that replace previously released data. Quantitative revision indicators should be disseminated here in line with recommended FAO Statistical Standard on Data revision[22] and FAO Statistical Standard on Quality indicators for external users[23].
15	**Timeliness and punctuality**		TIMELINESS_PUNCT*	
15.1	Timeliness and punctuality	Timeliness	TIMELINESS	The speed of dissemination of statistical outputs - i.e. the lapse of time between the end of a reference period (or a reference date) and the dissemination of the statistical outputs (e.g. quarter, month of the following year). Quantitative measures of timeliness (e.g Number of months between data availability and the event or phenomenon) should be provided here. Recommended timeliness indicators can be found in FAO Statistical Standard on Quality indicators for external users[24]. . Example of reporting on timelines indicators: Data are normally received 11 months after the reference year, and processed and disseminated by FAO within 2 months.
15.2	Timeliness and punctuality	Punctuality	PUNCTUALITY	Time lag existing between the actual delivery date of statistical outputs and the target date when they should have been delivered, for instance, with reference to dates announced in an official public release calendar or previously agreed among partners. When a target date of dissemination is set by FAO, a quantitative measure of punctuality should be provided here as per FAO Statistical Standard on Quality indicators for external users[25].
16	**Coherence and Comparability**		COHER_COMPAR*	Coherence is the adequacy of the statistical outputs to be meaningfully combined in different ways and for various uses. Comparability refers to the extent to which differences between different geographical areas, non-geographical domains, or over time, can be attributed to differences between the true values of the statistical characteristics.

[22] https://www.fao.org/3/cb9311en/cb9311en.pdf
[23] https://www.fao.org/3/cb9309en/cb9309en.pdf
[24] https://www.fao.org/3/cb9309en/cb9309en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
16.1	Coherence and Comparability	Comparability – geographical	COMPAR_GEO	Extent to which statistics are comparable between geographical areas, countries or regions. It includes description of the reasons of problems and as well the order of magnitude of the effects of the main sources of errors. It refers to the degree of comparability between similar results measuring the same phenomenon across geographical areas or regions (e.g. There is limited geographical comparability due to differences between countries in methods and coverage; e.g. Worldwide geographical comparison is possible; e.g. Data are mostly comparable over countries and regions).
16.2	Coherence and Comparability	Comparability - over time	COMPAR_TIME	Extent to which statistics are comparable or reconcilable over time. It provides information on the length of comparable time series, reference periods at which series breaks occur, the reasons for the breaks and treatments of them (e.g. For short time periods, reasonably good comparability over time can be expected, as there is stability in the product definition and classification. However, as the time series are very long (from 1955) full comparability over time is impossible). Additionally, provide information about the comparability over time of the methods used to produce the data; the stability of the data access, and the changes in the covered population over time Give also an assessment of how the over time comparability of the data will develop in the future. Finally, as per FAO Statistical Standard on Quality indicators for external users[26], this field should include a quantitative measure of comparability (number of comparable data items in time series).
16.3	Coherence and Comparability	Coherence	COHERENCE	Coherence is the adequacy of the statistical outputs to be meaningfully combined in different ways and for various uses. For this sub-concept, information should be provided on the extent to which statistics are reconcilable with those obtained through other data sources or statistical domains and extent to which the statistics are consistent within a given data set. As per FAO Statistical Standard on Quality indicators for external users[27], this field should include quantitative measures of differences or relative differences in the statistical results calculated on the basis of different statistical domains based on different methodologies.

[26] https://www.fao.org/3/cb9309en/cb9309en.pdf
[27] https://www.fao.org/3/cb9309en/cb9309en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
16.4	Coherence and Comparability	Classification system	CLASS_SYSTEM	List of classification(s) used, and description how these conform to internationally agreed standards. When relevant, deviations from international standard classifications should be documented. In particular, deviations from international classifications caused by the use of a particular data sources (e.g. a Big data source) should be explained. A weblink reference to the classification should be included if available.
17	**Statistical process**		STAT_PROCESS*	
17.1	Statistical process	Source data	SOURCE_TYPE	Characteristics and components of the statistical data used by FAO for compiling statistical aggregates from national primary statistics and also organizations. This is used to indicate whether the data set is based on a survey, on administrative data sources, on a mix of multiple data sources or on data from other statistical activities (e.g. The main national data sources are normally sample surveys, censuses and sometimes administrative sources. Other sources of the primary data are organizations). Only if available and relevant some sample characteristics sample surveys should also be given (e.g. gross and net sample size, type of sampling design, reporting domain etc.). If administrative registers are used, the description of registers should be given (source, year, primary purpose, potential deficiencies etc.). Description should be given at the level of FAO data collection not each country specific situation. A reference to FAO Statistical Standard on the use of external data sources[28] should be included.
17.2	Statistical process	Frequency of data collection and acquisition	FREQ_COLL	Time interval at which the source data are collected/acquired. This is used to indicate the frequency of data collection/acquisition and recording (e.g. monthly, quarterly, annually, continuous).

[28] https://www.fao.org/3/cb9343en/cb9343en.pdf

Code	Concepts	Sub-concepts	Identifier	Guidelines
17.3	Statistical process	Data collection method	COLL_METHOD	Method applied for gathering data for official statistics. Description of data collection methods used, including computer assisted interviewing (CAPI/CAWI), on-line surveys, mailed questionnaires administrative data sources, web-scraping and crowdsourcing sources (e.g. national data are collection through questionnaires that are automatically generated by the FAO Statistical Working System). A reference to FAO Statistical Standard on collecting data from national institutions[29] should be included. For non-statistical data sources (including Big Data): • Describe the methods used for data acquisition and recording, including methods used by data providers/third parties if data were acquired through data sharing agreements; • Add or specify where to find the hyperlink(s) of web data used, or add name of the API used to collect the data (when relevant)
17.4	Statistical process	Data validation	DATA_VALIDATION	Process of monitoring the results of data compilation and ensuring the quality of the statistical results (see also 16.3. Coherence). Data validation describes all methods and processes (e.g. such as data editing, imputation of missing values) for assessing statistical data, and how the results of the assessments are monitored and made available to improve statistical processes (e.g. The data undergoes comprehensive validation work that covers: detection of outliers, transmission errors and data consistency checks; e.g. Countries asked to examine the disseminated results for their country and either to confirm that they are correct or to provide remarks and / or revised data if they identify errors). A reference to FAO Statistical Standard on Data editing and validation of input data[30] should be included. For statistical outputs based on non-statistical data sources, discuss the linkage(s) of the data sources with other sources, and linkage methods used for the validation of the outputs (if relevant).
17.5	Statistical process	Data compilation	DATA_COMP	Operations performed on data to derive new information according to a given set of rules. Description of the data compilation process for deriving the main statistical outputs (aggregates, indicators, etc.) starting from the validated input data (e.g. FAO is responsible for compiling the data and generating regional aggregates).
17.6	Statistical process	Regional aggregation	REG_AGG	Methods used to produce global, regional and sub-regional aggregates. A reference FAO Statistical Standard on Data aggregation[31] should be included.

[29] http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Collecting_Data_from_National_Institutions_Rev5_Jun2023_clean.pdf Accessible internally only
[30] Data editing and validation of input data (fao.org)
[31] Data aggregation (fao.org)

Code	Concepts	Sub-concepts	Identifier	Guidelines
17.7	Statistical process	Adjustment	ADJUSTMENT	Set of procedures employed to modify statistical data to enable it to conform to international standards or to address data quality differences when compiling specific data sets such as seasonal adjustment, exchange rates, prices, time series decomposition and other methods. Adjustments are in particular applied to compile consistent time series, but the concept is also used for describing adjustments related to other types of data (e.g. For quarterly data, seasonally and calendar adjusted data are compiled using an indirect approach, i.e. based on the submitted country figures).
17.8	Statistical process	Imputation	IMPUTATION	Procedures used to impute data items where response were missing, unusable or discarded due to expected low quality. A reference to FAO Statistical Standard on Imputation[32] should be included.
18	**Quality management**		QUALITY_MGMNT	Systems and frameworks in place within an organization to manage the quality of statistical products and processes.
18.1	Quality management	Quality assurance	QUALITY_ASSURE	All the planned and systematic activities implemented that can be demonstrated to provide confidence that the data production processes to fulfill the requirements for the statistical output, including measures for ensuring the efficient use of resources. The concept includes the design of programmes for quality management, the description of planning process, and other organizational arrangements to support and maintain planning function and brief description on how it is implemented for the domain-specific quality assurance activities (the use of best practices, quality reviews, self-assessments, compliance monitoring, internal audits, QAPS, checklist for assessment of databases.. etc.). **Use a standard text, if possible.** E.g. FAO is responsible for the quality of the internal statistical processes used to compile the published datasets. The FAO Statistics and Data Quality Assurance Framework (SDQAF), available at: http://www.fao.org/3/cc6683en/cc6683en.pdf, provides the necessary principles, guidelines and tools to carry out quality assessments. FAO is performing an internal bi-annual survey (FAO Quality Assessment and Planning Survey) designed to gather information on all of FAO's statistical activities, notably to assess the extent to which quality standards are being implemented with a view to increasing compliance with the quality dimensions of SDQAF, documenting best practices and prepare quality improvement plans, where necessary. Domain-specific quality assurance activities are carried out systematically (e.g. quality reviews, self-assessments, compliance monitoring).

---

[32] Imputation (fao.org)

Code	Concepts	Sub-concepts	Identifier	Guidelines
18.2	Quality management	Quality assessment	QUALITY_ASSMNT	Overall evaluation of data quality based on standard quality criteria. It includes the result of a scoring or grading process for quality (e.g. QAPS, overall assessment of the database). Scoring may be quantitative or qualitative. A qualitative assessment of the overall quality of the statistical outputs should be provided by summarizing the main strengths and possible quality deficiencies. Any trade-offs between quality aspects can be mentioned as well as planned quality improvements (e.g. if possible, the results of the FAO Quality Assessment and Planning Survey – QAPS should be reported).
19	Recommended uses and limitations		REC_USE_LIM	Metadata elements guiding users of the statistics presented and helping them to determine whether the product meets their requirements
20	Comment		COMMENT	Supplementary descriptive text which can be attached to data or metadata.

Annex 3: Document history

Revision Version	Revision Date	Author	Description of changes/status
0	12-11-2020	**Aymen Charef**	Draft 0
0.1	7-12-2020	**Aymen Charef**	Revisions according to comments received from members of the IDWG Statistics Adding concepts in the reference metadata structure and elaborating guidelines text
0.2	21-12-2020	**Aymen Charef**	Editorial changes to text according to comments received from members of the IDWG Statistics
1	7-1-2021		Endorsed by IDWG-TTF on Statistics
1.1	27-01-2023	**Ngarsaim Espoir Beram and Valerie Bizier**	Addition of provisions to inform users on the use of non-statistical data sources in the generation of statistical outputs and to ensure compliance with FAO Data Policies on Data Protection and Intellectual Property
1.2	23-06-2023	**Valerie Bizier**	Inclusion of comments received by DCG-T members, in particular the addition of narratives regarding the scope of the standard (datasets disseminated through DATA database) and the addition of a field REL_SPDX_CODE to provide machine-readable information on the license under which the datasets are made available
2	23-06-2023		Endorsed by DCG-T through written consultation

References

Data Documentation Initiative (DDI). 2021. Document, Discover and Interoperate. In: Data Documentation Initiative website. Cited 01 July 2021. https://ddialliance.org/

Eurostat. 2020. ESSNet Big Data II Project. In: Methodology and quality section. https://ec.europa.eu/eurostat/cros/sites/default/files/WPK_Deliverable_K6_Quality_report_template_2020_02_28.pdf

Eurostat. 2021. ESS Reference Metadata Reporting Standards. In: Metadata section. Cited 01 July 2021. https://ec.europa.eu/eurostat/data/metadata/metadata-%20structure

FAO. 2023. *The FAO Statistics and Data Quality Assurance Framework.* Rome. http://www.fao.org/3/cc6683en/cc6683en.pdf

FAO. 2017. Basic texts of the Food and Agriculture Organization of the United Nations. Rome. https://www.fao.org/3/K8024E/K8024E.pdf

International Organization for Standardization (ISO). 2021. ISO 17369:2013. In: ISO website. Cited 01 July 2021. https://www.iso.org/standard/52500.html#%3A~%3Atext%3DISO%2017369%3A2013%20provides%20an%2Cstatistical%20data%20and%20related%20metadata

Statistical Data and Metadata eXchange (SDMX). 2021a. SDMX Homepage. In: *SDMX website.* Cited 01 July 2021. https://sdmx.org/

SDMX. 2021b. SDMX Glossary Version 2.1. In: *SDMX website.* Cited 01 July 2021. https://sdmx.org/?sdmx_news=sdmx-glossary

SDMX. 2021c. SDMX Content-Oriented Guidelines (COG). In: *SDMX website.* Cited 01 July 2021. https://sdmx.org/?page_id=4345&Standardising

United Nations Statistics Division (UNSD). 2021. Fundamental Principles of National Official Statistics. In: *United Nations Statistics Division website.* New York. Cited 01 July 2021. https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx