

**Statistical**

**Standard**

**Series**

**Quality Indicators to Be Disseminated to External Users**

Endorsed by the IDWG on Statistics TTF

on xxxx

****

|  |
| --- |
| This standard provides the list of recommended quality indicators to be disseminated to FAO external data users. It includes key definitions, guidance on how to compile the indicators, general and technical recommendations on their compilation and dissemination as well as some governance considerations. |

Contents

[Background 5](#_Toc8046377)

[Definitions 5](#_Toc8046378)

[General recommendations 6](#_Toc8046379)

[Technical recommendations 6](#_Toc8046380)

[Governance procedures 7](#_Toc8046381)

[Annexes 8](#_Toc8046382)

Background

FAO is committed to produce and disseminate high quality statistics, where quality is measured in terms of (i) relevance; (ii) accuracy and reliability; (iii) timeliness and punctuality; (iv) coherence and comparability; and (v) accessibility and clarity. Measuring the quality of FAO statistical outputs is crucial to establish a roadmap for quality improvements. In addition, disseminating information on quality indicators facilitates a correct interpretation and use of the statistical outputs, contributing directly to improve the accessibility and clarity dimensions of quality.

Unfortunately, measuring quality is not straightforward. For some quality dimensions, like timeliness, a quantitative assessment can be easily obtained, while for other dimensions, this type of assessment may encounter several methodological difficulties and require non-negligible efforts. For instance, assessing the relevance of a statistical output would require carrying out a users’ satisfaction survey. Accuracy itself, which implies a quantitative assessment of how close an estimate is to the corresponding true (unknown) value, would require the estimation of the Mean Square Error (MSE), which is often unfeasible. For these reasons, the common approach to the measurement of the quality of statistics consists in the compilation of a series of quality indicators, directly or indirectly related to the quality dimensions. The indicators not directly measuring the quality usually focus on critical phases of the production process of the statistical outputs (performance indicators), based on the assumption that the fewer errors in the process the higher the quality of the final statistical outputs.

This document provides a list of recommended quality indicators intended to be disseminated to external users, jointly with the statistical outputs. The quality indicators are presented according to the dimensions they refer to.

Definitions

*Relevance*

Relevance is the degree to which the statistics produced meet current and potential users’ needs (see FAO Statistics Quality Assurance Framework). For a given statistical production process or, more generally, statistical domain, an indicator of relevance is **the percentage of totally satisfied users,** estimated from the corresponding users’ satisfaction survey. As reported in the Statistical Standard on users consultations[[1]](#footnote-1), the indicator is derived by adding all responses “Strongly Agree” or “Agree” in a five-items Likert scale (see Annex 1) for each of the SQAF Principles related to quality (P1-P5)[[2]](#footnote-2) and dividing this value by the total number of responses. Details on the calculation are reported in Annex 1.

*Accuracy and reliability*

**Accuracy** is the closeness of an estimate to the true value of what is measured. In this case, we are not interested in the accuracy of the incoming country data, rather in assessing solely FAO’s contribution to the overall accuracy of the final statistical outputs calculated and disseminated by FAO (usually referred to Regions or to the “World” level). This piece of the overall accuracy of the statistics calculated by FAO will depend on the errors that can happen in the FAO statistical production process; the higher the number of errors, the lower the accuracy. The best strategy to improve accuracy is to prevent errors from happening and correct the ones discovered (before the dissemination of the final statistical outputs).

The main errors that can affect FAO statistical production processes are:

* **Non-observation errors**: a respondent (country) does not report part or any of the required data.
* **Measurement errors** (in a broad sense): the final value for a given variable from a given respondent does not correspond to the one disseminated at the national level because of errors in the FAO process (data collection, data treatment, data processing). The most frequent reasons can be: wrong instructions or definitions provided in the questionnaire; classification/coding errors in the collection or treatment of country data; errors in converting the unit of measure during the data collection phase or in the treatment of country data; errors in the data validation phase (a non-erroneous value is identified as suspicious and modified when data are checked by FAO).
* **Estimation errors**: errors introduced by FAO during the calculation of the final aggregates. Typically, they correspond to errors in the software codes or in the model/assumptions underlying the aggregation of complex indicators (e.g. use of a simple average instead of a weighted average, etc.).

Non-observation errors are crucial, since just a small fraction of units not reporting their data may hinder the calculation of regional/global aggregates[[3]](#footnote-3). The favorite solution consists in imputing[[4]](#footnote-4) the missing values and then calculating the final aggregates using imputed values as if they were really observed (jointly with actually observed values). The calculation of the final statistical outputs discarding missing values (i.e. by processing solely the observed values) may need ad hoc assumption and/or methods; for instance, estimating the total amount of a variable by summing up just the observed values will be affected by a negligible bias (underestimation) if the missing values are assumed to be very small compared to the observed ones[[5]](#footnote-5). In some cases, mixed strategies can be adopted (e.g. imputation of only the most influential non-reported values).

A first indicator related to non-observation error is the **reporting rate** (or **response rate**), i.e. the number of returned filled in questionnaires (with complete or partial valid information) divided by the number of dispatched questionnaires.

When imputation is used to compensate for non-observation errors or for erroneous reported values (typically values identified as errors are deleted and replaced with valid plausible imputed values), it should be calculated the **imputation rate**, i.e. the ratio of the imputed data items over the total number of data items (imputed and observed).

When the final statistics (regional and global) are obtained as the sum of values (total amount), an important quality indicator to be calculated is the **contribution of the imputed values to the final sums** (for calculation details see Annex 1). This indicator may also prove useful when the disseminated outputs are obtained as a function of one or more sums (e.g. ratio of sums).

The values of the quality indicators related to imputation should be provided together with summary information (or links) on the methods applied to perform imputation.

**Reliability** indicates how close the initial estimates are to the subsequent or final estimates. Assessing reliability is preferred to measuring accuracy when dealing with complex statistical processes that involve multiple data sources being updated at different times, possibly, in some cases, with only provisional data. In this context, a common practice is to produce provisional estimates that are subsequently revised[[6]](#footnote-6) when the underlying data are updated. Revised estimates should be accompanied by a couple of revision indicators aimed at assessing the reliability of the estimates: the **mean revision** (MR) and the **mean absolute revision** (MAR). The MR provides an indication of the direction of the revision, while the MAR is used to assess the size of the revisions (the calculation details are reported in Annex 1).

*Timeliness and Punctuality*

**Timeliness** is the lapse of time between the end of a reference period (or a reference date) and the dissemination of the statistical outputs. The **overall time lapse** can be split in two parts: the timeliness of respondents in submitting the data requested by FAO and the time required for FAO to produce and disseminate the statistical output. Timeliness should be measured in months: the **“data provider” timeliness** is calculated as the number of months between the reference date of the statistical output and the last day of the FAO data collection window. The **timeliness of the FAO process** is the number of months between the first day after the data collection window and the date of dissemination of the FAO statistical outputs (key variables for regional and global aggregates). Details on the calculation of these indicators are reported in Annex 1.

**Punctuality** refers to the possible time lag (in months) between the actual delivery date of FAO statistical outputs and the target delivery date. In practice, punctuality can be measured only on the FAO side when a target date of dissemination exists (usually set in the official dissemination calendar).

*Coherence and comparability*

**Coherence** indicates how adequate the statistical output is to be meaningfully combined with other statistics in different ways and for various uses. Generally speaking, coherence refers to the extent to which statistics on the same phenomenon can be compared or combined. Coherence can be assessed at different levels: (i) in the same statistical domain, when comparing provisional with final estimates of the same aggregate; (ii) across statistical domains, when comparing similar statistics disseminated by different units (domains) within the same agency; and (iii) across agencies, when comparing statistics on the same topic produced by different agencies.

For FAO purposes, it may be worth calculating coherence indicators when provisional estimates are disseminated externally and are subsequently replaced by final estimates. For numerical aggregates, the **difference between the final and provisional estimate** (sometimes expressed in relative terms, as shown in the Annex 1) should be calculated.

**Comparability** refers to the extent to which different statistical estimates between geographical areas, non-geographical domains, or over time, can be attributed to real differences of the characteristic measured. Comparability is a stricter concept than coherence, as it assumes that the statistics being compared are produced by processes sharing the same concepts, definitions, classifications, methods, etc. A quality indicator related to comparability over time is the **number of comparable data points in the time series** since the last structural break. A break in a time series may occur when the definitions, the classifications or the procedures used in the statistical process are changed. When a break in the time series is introduced, then the indicator of comparability over time should be provided together with information on the main reasons of the break.

*Accessibility and clarity*

Accessibility is the set of conditions and modalities that determine how easy it is for the users to obtain data. A commonly indicator of accessibility is the number of visits of a given page or the number of data downloads. This indicator can be calculated using Google Analytics facilities.

General recommendations

* The data owner should evaluate which quality indicators presented in this document are applicable to the statistical process under his responsibility. All relevant quality indicators should be calculated and disseminated to external users jointly with the disseminated statistical outputs. A brief text should be provided to explain how to interpret and use the disseminated quality indicators.
* Quality indicators should be calculated at both global and regional levels (when relevant) in accordance to suggestions provided in the following Table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Quality dimension** | **Quality indicator** | **Geographical level** | **Disseminated outputs** |
| Relevance | Percentage of totally satisfied users (M) |  | Referred to the key variables/indicators |
| Accuracy | Response rate (RR) | Global and Regional |  |
|  | Weighted Response Rate (WRR) | Global and Regional |  |
|  | Imputation Rate (IR) | Global and Regional | Referred to the key variables/indicators |
|  | Contribution Imputed Values to Totals (CIVT) | Global and Regional | Referred to the key variables/indicators |
| Reliability | Mean Revision (MR) | Global and Regional | Referred to the key variables/indicators that are revised |
|  | Mean Absolute Revision (MAR) (or RMAR) | Global and Regional | Referred to the key variables/indicators that are revised |
| Timeliness | Overall Timeliness | Global and regional (if regional are disseminated later) |  |
|  | Timeliness incoming data | Global and regional (if regional are disseminated later) |  |
|  | Timeliness FAO statistical process | Global and regional (if regional are disseminated later) |  |
| Punctuality | Delay in publication  (if exists a dissemination calendar) | Global and regional (if regional are disseminated later) |  |
| Coherence | Difference or relative difference between provisional and final outputs | Global and regional | Referred to the key variables/indicators |
| Comparability | Number of comparable data items in a time series | Global and regional | Referred to the key variables/indicators |
| Accessibility | Number of web pages visits |  |  |
|  | Number of data downloads |  |  |

* Quality indicators different from those suggested in this document can be disseminated if they are considered relevant for the statistical outputs being disseminated.

Technical recommendations

* Quality indicators should be easily accessible from the web pages where the statistical outputs they refer to are displayed. Electronic publications disseminating statistical data should include an Annex with the pertinent quality indicators.
* Whenever possible, quality indicators should be disseminated via a quality report (a template is being prepared), i.e. a summary report providing the most important information about the quality of the process and the corresponding statistical output. A sub-optimal solution would consist in disseminating the quality indicators jointly with the reference metadata.
* A contact or a link to additional information about quality indicators (underlying definitions, calculation formula, main use, etc.) should be provided for users interested in getting more insights.

Governance procedures

* The list of quality indicators can be revised and/or updated (e.g. with the inclusion of additional indicators considered relevant for existing or new statistical production processes). In such case, the revised/updated list will be endorsed by the IDWG on Statistics and the Office of the Chief Statistician will be responsible for updating this standard.

Annex 1: Details on indicators

**Relevance – percentage of totally satisfied users**

A user satisfaction survey should allow to derive the following table where satisfaction of each of the quality dimensions is measured through a five-items Likert Scale.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Response | **FAO statistical outputs are:** | | | | |  |
| Categories  (*five-items Likert scale*) | Relevant | Accurate/ reliable | Timely/  punctual | Coherent/ comparable | Accessible/ clear | Total |
| Strongly agree |  |  |  |  |  |  |
| Agree |  |  |  |  |  |  |
| Neutral |  |  |  |  |  |  |
| Disagree |  |  |  |  |  |  |
| Strongly disagree |  |  |  |  |  |  |
| Total |  |  |  |  |  |  |

Calculation formula:

Note that in the absence of missing values or “Don’t know”, then , being the number of users participating to the survey (respondents to the survey).

**Accuracy – response rate (or reporting rate)**

Calculation formula:

is the number of returned questionnaires completely or partially filled with valid information; a questionnaire partially filled whose information is not used for calculation of the final statistical outputs (because not relevant or because being of poor quality) should be counted as a non-response.

is the number of dispatched questionnaires.

**Accuracy – Weighted response rate (or weighted reporting rate)**

The RR considers just if a country is reporting or not to FAO, but it does consider how important is the contribution of the country to the study of a given phenomenon; for this reason, it maybe necessary to calculate a *weighted response rate*:

Being a numerical value expressing the importance of country *j*. Typical variables used to represent the importance of a country are its GDP (when dealing with economic phenomena) or its total population (social phenomena).

Calculation of WRR is optional; when calculated it should be always come along with the RR indicator.

**Accuracy – imputation rate**

The calculation formula is based on the current statistical standard series related to flags[[7]](#footnote-7) associated to data:

: total number of units (contacted countries or number of questionnaires dispatched) of the considered key variable needed for calculation of the required aggregate (regional or “World”).

: number of units where the phenomena cannot exist (flagged with “M”)

: number of units where the phenomena is considered not significant (flag “N”).

: number of estimated values of the considered key variable needed for calculation of the required aggregate (regional or “World”).

: number of imputed values of the considered key variable needed for calculation of the required aggregate (regional or “World”).

In practice, it is considered a wide meaning of the term “imputation”, that jointly with imputed values (flag “I”) includes also all the values that come from estimation (flag “E” usually associated to back-casts, low quality values, experimental data, etc.).

If the standard on flagging the data is still not adopted, then the numerator should be replaced with the number of data gaps imputed using simple or advanced statistical methods (e.g. filling the data gap by “copying” the last observed value is considered as an imputation and should be counted in the numerator).

**Accuracy – contribution of the imputed values to the final aggregates**

This indicator is applicable to the cases where the final statistical output corresponds to the sum of the collected value. If Y is the variable for which the total is required then the following calculation formula should be applied

In practice, the denominator, , is the estimated total of the variable Y (at the “World” or regional level) that is disseminated externally and the numerator is the sum of the values that are achieved by estimation (flag “E”) and imputation (flag “I”). Whereas the standard on flagging the data is still not adopted, the numerator should be replaced with the sum of imputed values using simple or advanced statistical method (e.g. filling the data gap by “copying” the last observed value is considered as an imputation and should be considered in the numerator).

**Reliability – revision indicators**

Revision indicators are calculated when a disseminated statistical output is revised subsequently in time, as explicitly foreseen in the revision policy[[8]](#footnote-8).

The **Mean Revision** is an average between the differences between the latest and the previous disseminated aggregate values over the terms in the time series:

: latest available estimate of the aggregate (regional or “World” level) for the variable of interest at time t;

: previous available estimate of the aggregate (regional or “World” level) for the variable of interest at time t;

: starting year in the time series: the starting year should not be a year before a break in the time series;

: last year in the time series.

In this term, MR provides, on average, an idea about the direction of the latest revisions vs. previous ones. Since revision policies usually foresee that an aggregate is revised more than once, the “P” (previous) estimate considered here is preferably the first released estimate.

The stability of the revisions can be measured through the **Mean Absolute Revision**

It gives an idea of the average size of the revision. Sometimes it is expressed in relative terms, i.e. **Relative MAR**:

An alternative indicator can be calculated by changing the denominator, i.e. by measuring the relative size of the revision compared to the previous value (“P”), i.e.

The following example, refers to the case where an aggregate is revised twice (every year the data on the latest 3 years are collected).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date of** | **Reference year of the data** | | | | | | | | | | |
| **dissemination** |
| **of aggregates** | **…** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** |
| **2012** |  | L\_2008 | P2\_2009 | P1\_2010 |  |  |  |  |  |  |  |
| **2013** |  | L\_2008 | L\_2009 | P2\_2010 | P1\_2011 |  |  |  |  |  |  |
| **2014** |  |  | L\_2009 | L\_2010 | P2\_2011 | P1\_2012 |  |  |  |  |  |
| **2015** |  |  |  | L\_2010 | L\_2011 | P2\_2012 | P1\_2013 |  |  |  |  |
| **2016** |  |  |  |  | L\_2011 | L\_2012 | P2\_2013 | P1\_2014 |  |  |  |
| **2017** |  |  |  |  |  | L\_2012 | L\_2013 | P2\_2014 | P1\_2015 |  |  |
| **2018** |  |  |  |  |  |  | L\_2013 | L\_2014 | P2\_2015 |  |  |
| **2019** |  |  |  |  |  |  |  | L\_2014 | L\_2015 |  |  |

The table shows that the latest revision of an estimate (final) is in the cell with yellow background color and is achieved after two years from the initial released estimate (P1, cell with red background color). Comparing values in the yellow cells with the corresponding ones in the red ones allows calculating the revision indicators of latest estimate vs. the first initial estimate, in practice in the table the MR, MAR and RMAR indicators are averages of 6 couples L-P1 (from 2010 to 2015).

**Timeliness**

The **overall timeliness** is calculated as:

It corresponds to the difference between two dates expressed in months:

is the date of dissemination of the statistical outputs

is the last day of the reference period the statistics refer to; e.g. if statistical outputs refer to 2017 then

The overall timeliness can be split in two components:

1. Timeliness of incoming data (data source), i.e. the number of months from the reference date up to the data FAO collection:

Where is the **last day** of the FAO data collection period of the data needed for producing the statistics of interest (day of arrival of the last dataset, in case of data provided by other international organizations). In case of FAO statistical processes based on both data collected from the countries and dataset provided by other organizations, the date to considerer for is the latest one.

1. Timeliness of FAO statistical process, i.e. the number of months between the collection of the data and the dissemination of the statistical outputs based on them

The sum of the two components returns the overall timeliness:

**Punctuality**

Is the delay in disseminating the results, i.e. number of months between the actual dissemination date and the one scheduled (typically indicated in an official dissemination calendar):

**Coherence**

When provisional statistical outputs (Regional or World) are disseminated and then they, after some time, are replaced with final values, it may be worth calculating the following coherence indicator:

It gives an idea on the closeness of the final estimate with respect to the provisional one. It can expressed in relative terms:

Sometimes terms in the formula can be exchanged by considering as the reference, i.e.

and

**Comparability**

**Comparability in time** is measured when statistical outputs contribute to a time-series. It is calculated as the number of comparable data items in the time series since last break.

Annex 2: Document history (for internal purposes)

|  |  |  |  |
| --- | --- | --- | --- |
| **Revision Version** | **Revision**  **Date** | **Author** | **Description of changes/status** |
| 0.1 | 2019-07-08 | Marcello D’Orazio | First draft |
| 0.2 | 2019-07-11 | Marcello D’Orazio | Revisions by Pietro  Small changes by MDO |
| 0.3 | 2019-09-25 | Marcello D’Orazio | Revisions by Valerie  Added indicator on the overall response rate and other changes |
| 0.4 | 2019-10-07 | Marcello D’Orazio | Editorial changes to text; added weighted response rate |
|  |  |  |  |

1. Statistical Standard Series User Consultations, endorsed by the IDWG-TTF on Statistics on 5 April 2019 <http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_User_Consultation__endorsed_5_April_2019_.pdf> [↑](#footnote-ref-1)
2. <http://www.fao.org/docrep/019/i3664e/i3664e.pdf> [↑](#footnote-ref-2)
3. See corresponding Statistical Standard Series (forthcoming) [↑](#footnote-ref-3)
4. Imputation is used here in its broader meaning, including also the replacement of missing values with values from nonofficial sources, historical data, etc. For further details, see the corresponding Statistical standard Series

   <http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Imputation__endorsed_5_April_2019_.pdf> [↑](#footnote-ref-4)
5. The sum of the missing values contributes to 1% or 2% of the total amount. [↑](#footnote-ref-5)
6. For more details on recommended practices for data revisions see the corresponding Statistical Standard Series <http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Data_revision__endorsed_30_January_2019_.pdf> [↑](#footnote-ref-6)
7. [http://intranet.fao.org/fileadmin/user\_upload/scp/Standards\_for\_quality\_compliance/SSS\_Observation\_‌Status\_Codes\_\_Flags\_\_endorsed\_\_December\_2016\_.pdf](http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Observation_Status_Codes__Flags__endorsed__December_2016_.pdf) [↑](#footnote-ref-7)
8. <http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Data_revision__endorsed_30_January_2019_.pdf> [↑](#footnote-ref-8)