Quality indicators (draft)
FAO ESS review for questionnaire-based FAOSTAT domains

Javier Montero-Serrano, Irina Kovrova and Jean Marie V. Munyeshyaka
FAO ESS. June 2019

# Scope

The FAO Statistics Quality Assurance Framework (SQAF) (FAO, 2014) distinguishes three quality areas. This report focuses on one of them, the area of Statistical Outputs:

* **Statistical Outputs**
* Statistical Processes
* Institutional Environment.

Within the area ‘Statistical Outputs’, the FAO SQAF distinguishes five dimensions/principles. This review focuses on principle 2: Accuracy and Reliability.

* Relevance
* **Accuracy and Reliability**
* Timeliness and Punctuality
* Coherence and Comparability
* Accessibility and Clarity

The FAO Statistics Coordination Portal (http://intranet.fao.org/statistics\_coordination\_portal/) provides other relevant documentation for quality compliance, in addition to the SQAF. One of these documents is “Reporting on Data Quality Principles, Standards & Recommendations” (FAO, 2011), which in annex 1 proposes the indicators listed below for “Accuracy and Reliability”[[1]](#footnote-1). This review will focus on three of them: response rate, proportion of imputed data points, and ratio of the value of imputations.

* Expected variation of survey estimates resulting from random sampling (coefficient of variation)
* **Response rate**
* **Proportion of imputed data points**
* **Ratio of the value of imputations**
* Estimated Root Mean Squared Error of the imputed data
* Proportion of statistical discrepancy in the related totals
* Average relative size of revisions
* Proportion of rejected source data

# Reference documents

The three concepts selected for analysis in this review (response rate, proportion of imputed data points, and ratio of the value of imputations) cover the variables used in the **Programme Implementation Report 2016-17** (PIR) (FAO, 2018a; p.125) for outcome indicator 2.4.A: “*Number of countries with improved response rates and/or quality of contributions* […]”. This outcome indicator is based on average response rates and average *quality* ratings (with *completeness* being the quality dimension measured, as percentage of data points with the flag 'official data' in the corresponding FAOSTAT dataset).

The **FAO Statistics Coordination Portal** provides the document “Reporting on Data Quality Principles, Standards & Recommendations” (FAO, 2011). This report refers to an annex 2 titled “Proposed Formulas for Selected Quality Indicators”. Such annex could not be found online, but its most recent version may be the document “FAO Quality indicators for statistical processes: recommended indicators” (FAO, 2016a), which provides formulas for quality indicators. Looking at the dimension Accuracy and Reliability, it provides formulas for “data completeness”, “imputation rate” and “contribution of the non-imputed values to the final aggregate”. These three indicators approximately match the three concepts selected for this review.

The **Statistical Working System** (SWS) contains also preliminary developments of plugins for computing two of the indicators mentioned above: “imputation rate” and “contribution of the non-imputed values”. The names, however, seem to have been reinterpreted and partially swapped: in a 2016 draft document titled “Quality Indicators for the *aproduction* dataset in the agriculture domain” (received by email from the ESS Methodology Team, 30 May 2018), the indicators are named “imputation frequency” and “imputation rate”, respectively.[[2]](#footnote-2)

At the **Asia and Pacific Commission on Agricultural Statistics (APCAS)** was presented a document titled “Issues in the collection of FAO data” (FAO, 2016b; pp. 4-6). It included data on “*response rates, used to measure data availability*” and on “*data completeness, used to measure data quality,* […] *estimated by the proportion of total FAOSTAT records that are official records*”. This document also provided a description on how those indicators were calculated.

Finally, the concepts of response rate and imputation rate can also be found in a list of basic elements for measuring data quality provided by M. Trant in a 1993 article titled “Development of data quality indicators”, included in the **FAO Quarterly Bulletin of Statistics** (Trant, 1993).

# Variables involved in the calculation of the indicators

The main variables selected in this proposal, to be measured for the compilation of the response rate and data completeness indicators, are briefly described below. For variables that would need to be calculated manually, a proxy for automatic calculation is provided afterwards (when possible), to be calculated in SWS. Alternatively, the disseminated datasets in FAOSTAT could also be used, although differences in the flags and the data categories disseminated may prevent an exact reproduction of the SWS-based calculation.

## Number of questionnaires dispatched

Number of questionnaires sent in a particular year for a particular dataset. Currently it may need to be calculated manually. In future, it may be recorded through the SWS Questionnaire Management System.

## Number of questionnaires acknowledged by focal points

Number of replies received from questionnaire recipients, largely irrespective of the content of the reply. It includes:

* + filled-in questionnaire returned (partial or complete)
	+ communication of work in process to fill in the questionnaire
	+ link to an official site provided
	+ lack of data availability reported
	+ empty questionnaire returned
	+ instruction provided to obtain data from another source (e.g. from multilateral agency)

The only cases that will not be counted in this number are those in which: i) no reply is received, ii) the reply indicates that the questionnaire was addressed to the wrong person, or iii) an automatic reply is received, which does not imply acknowledgement of the questionnaire being addressed to the right person.

The ‘number of questionnaires acknowledged by focal points’ indicates in how many cases contact with a focal point was established. At present, this number needs to be calculated manually, as it involves an assessment of the content of the replies received.

## Number of questionnaires filled

Number of questionnaires received containing some data (contact details, metadata and feedback are excluded) in a particular year for a particular dataset. The questionnaire may not be complete but it must be at least partially filled-in (as ‘overall response rate’ defined by UK ONS, 2013; p.81).

**Automatic proxy:** For a particular dataset, number of territories with at least one data point from questionnaire for the last year covered in the relevant questionnaire mail-out.

## Number of territories in the geographical scope

Number of geographical territories eligible for a particular year in a particular dataset. It represents the number of territories that would cover the entire world (or the relevant geographical scope) in a dataset. It may vary from year to year depending, for instance, on integration/disintegration of countries.

This number needs to be calculated manually (expert judgement) unless it is certain that all relevant territories have records in the dataset.

*Note: the number of territories in the geographical scope may vary between datasets because some domains may provide data with a higher level of geographical disaggregation than others. In fact, the list of M49 area codes endorsed by FAO Interdepartmental Working Group (IDWG) on Statistics as “Standard Country or Area Codes for Statistical Use” (FAO, 2018b) includes not only self-governing countries but also non-self-governing territories. It may be tackled in future the identification of a subset of “main country codes” in this list, as well as the correspondences with GAUL and FAO Geopolitical Ontology (FAO, 2018c). As a step in this direction, a note for discussion about ‘number of territories in the geographical scope’ has been prepared at ESS together with this review on quality indicators. That note aims at generating a harmonized list of M49 codes representing the default geographical scope of the questionnaires.*

## Number of expected data points

Total number of data points considered applicable in the geographical scope of a dataset, for a particular year. It may need to be calculated manually.

If there are data points in item-element combinations that by definition are derived from other data points (e.g. data calculated as ratio, data transformed by conversion of units, data calculated by aggregation), those data points are not counted in this total. This restriction also applies to ‘number of data points’ and ‘number of official data points’.

**Automatic proxy[[3]](#footnote-3):** Select those lines (area-item-element combinations) with at least one value different from blank-blank and zero (where zero includes all zero values[[4]](#footnote-4), also those with flags O, M, Q and N), and count those lines. Then subtract from the line count the number of data points within those lines with 0‑M for the selected year (example provided in next page).

## Number of data points

Total number of data points with values in a dataset, for a particular year. It excludes missing, suppressed and not applicable values, as well as blanks.

The observation flags to be excluded are: **O**, **M**, **Q** and, **blank-blank** (note that blank-blank is different from value-blank; the former is the result of no input whereas the latter means ‘official data’).

## Example of calculation of “Number of data points” and “Number of expected data points (by the automatic proxy method)”:

The first table shows the value (‘#’ representing any value different from zero) and the observation flags. The second table shows the corresponding counts.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
| Area-Item-Element 1 | # - <blank> | # - <blank> | # - <blank> | # - <blank> | # - <blank> |
| Area-Item-Element 2 | # - <blank> | # - <blank> | # - <blank> | # - I | # - I |
| Area-Item-Element 3 | # - <blank> | # - <blank> | 0 - M | 0 - M | 0 - M |
| Area-Item-Element 4 | 0 - O | # - I | 0 - O | 0 - O | 0 - O |
| Area-Item-Element 5 | 0 - <blank> | 0 - I | 0 - I | 0 - O | blank-blank |
| Area-Item-Element 6 | 0 - M | 0 - M | 0 - M | 0 - O | 0 - O |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 2014 | 2015 | 2016 | 2017 | 2018 |
| Number of data points | 3 | 4 | 2 | 2 | 2 |
| Number of lines selected | 4 | 4 | 4 | 4 | 4 |
| Number of expected data points | 4 | 4 | 3 | 3 | 3 |

## Number of official data points

Number of data points classified as official in a dataset, for a particular year. Official statistics are those disseminated by a national statistical system (OECD, 2007).

SWS seems to allow dataset-specific selection of the subset of flags representing official values, but this might affect the comparability of the resulting ‘number of official data points’ across domains. Therefore, a harmonized approach is defined in the table below, where data points are classified as official or non-official depending on their observation flag.

|  |
| --- |
| **Observation status flags** |
| **Description** | **Flag** | **Official data point** | **Comment** |
| Official figure | <blank> | Yes |  |
| Revised | R | Yes |  |
| Provisional value | P | Yes | In this review, it is recommended to use flag “P” at ESS for provisional official statistics only (same as ‘P’ in FAOSTAT 2 flags), pending further discussion and harmonization with FAO OCS and IDWG-Statistics. |
| Estimated value | E | No |  |
| Imputed value | I | No |  |
| Unofficial figure | T | No |  |
| Missing value | O | No |  |
| Missing value (data cannot exist, not applicable) | M | No |  |
| Missing value; suppressed | Q | No |  |
| Figure from international organizations | X | No | If it is possible to know that a data point was originally disseminated by the statistical system of a country, even if FAO received the data through an international organization, such data point is considered official and flagged <blank> (e.g. data from Comtrade in ‘tariff line’). In consequence, the data flagged X are not counted as official. |
| Forecast value | F | No | In this review, it is recommended to use flag “F” at ESS for non-official statistics only, pending further discussion and harmonization with FAO OCS and IDWG-Statistics. |
| Not significant (negligible) | N | No | In this review, it is recommended to use flag “F” at ESS for non-official statistics only, pending further discussion and harmonization with FAO OCS and IDWG-Statistics. |
| Time series break  | B | No | In this review, it is recommended to use flag “B” at ESS for non-official statistics only. If the time series break applies to an official data point, it is recommended to mention the break in the “notes” field instead of the “flag” field, pending further discussion and harmonization with FAO OCS and IDWG-Statistics. |
| Exceptional event | S | No | In this review, it is recommended to use flag “S” at ESS for non-official statistics only. If the exceptional event mention applies to an official data point, it is recommended to include it in the “notes” field instead of the “flag” field, pending further discussion and harmonization with FAO OCS and IDWG-Statistics. |

## Number of data points received through questionnaires

Number of data points classified as ‘received through questionnaire’ (according to ‘method’ flag) in a dataset, for a particular year.

Only data points with method flag ‘q’ are counted in the number of data points received through questionnaires. All other flags, for instance ‘-’ (unknown collection method), are excluded.

It must be noted that there is no specific *observation* flag identifying the data points taken from a questionnaire. It is the method flag in the SWS the one indicating if the data point was collected through questionnaire. This method flag may not be available in FAOSTAT.

# Proposed formulas for selected quality indicators

As mentioned in section 1 (Scope), this review focuses on three concepts to be measured: one is response rate and the other two relate to data completeness (proportion of imputed data points and ratio of the value of imputations).

The indicators defined below measure response rates for questionnaires and data completeness for datasets. This selection of indicators and objects of analysis intends to optimize the number of indicators to be monitored and to focus the analysis.

The proposal is to make available all the data indicated in the tables below, including the calculation of the indicator in the last cell, according to the formula provided.

## Response rate (questionnaires):

### Overall response rate by questionnaire

**Meaning**: Share of questionnaires dispatched that have been returned filled with some data.

**Objective**: Assist in ESS internal assessment of the relevance of the questionnaire dispatch exercise. To infer information about the completeness of the questionnaires received, this indicator needs to be analyzed in combination with the indicator “dataset completeness from questionnaire”.

**Calculation**: This indicator is calculated for a questionnaire.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Questionnaire | # territories in scope | # quest. dispatched | # quest. acknowledged | # quest. filled | Response rate |
| Name | A | B | C | D | D/B |

Comparability of response rates may be affected by the geographical scope, and by the degree of coverage of such scope with the questionnaires dispatched (possibility of selection bias). A narrower geographical scope or a lower coverage with the questionnaire dispatch may affect the response rate upwards or downwards depending on the likelihood of the selected recipients to respond.

$$\frac{Number of questionnaires filled (by year)}{Number of questionnnaires dispatched (by year)}$$

### Overall response rate by territory

**Meaning**: Share of all data requests via questionnaire for which a certain country provided some data.

**Objective**: Support discussion with countries to identify issues with contact details or data availability.

**Calculation**: Response rate calculated by territory. In this case, the denominator is the number of questionnaires (domains) that were dispatched to this territory.

$$\frac{Number of questionnaires filled (by year)}{Number of questionnaires dispatched to this territory (by year)}$$

## Data completeness (datasets)

**Data completeness** is understood in this review as related to the number of data items provided by a reliable source (FAO, 2016a; p.5), and more specifically to the number of official data items; i.e. data disseminated by a national statistical system (OECD, 2007).

The data completeness of a dataset may not be directly related to the response rate of the related questionnaire (if any). This may be because:

1. Questionnaires received may not be fully filled-in
2. Not all data points may be collected in the questionnaire
3. There may be additional sources of official data, other than the questionnaire.

To facilitate the analysis of data completeness and the comparison with the questionnaire response rate, two data completeness indicators are proposed: i) dataset completeness from all official data, and ii) dataset completeness from questionnaire.

### Dataset completeness from all official data

**Meaning**: Share of expected data points (scope) that are filled with official data.

**Objective**: Assist in ESS internal assessment of the quality, in terms of accuracy[[5]](#footnote-5) and reliability, of the data in a certain dataset. For a complete analysis of the “Accuracy and Reliability” dimension, additional elements would need to be considered (as indicated in section 1 “Scope”).

**Calculation**: This indicator is calculated for a dataset. It may be calculated for its entire geographical scope, by region or by territory. Those data points (item-element-year combinations) that by definition are derived from other data points are excluded, as mentioned in section 3.5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | # expecteddata points | # data points | # official data points | # quest. data points | Data completeness from official data |
| Name | E | F | G | H | G/E |

$$\frac{Number of official data points (by year)}{Number of expected data points (by year)}$$

### Dataset completeness from questionnaire

**Meaning**: Share of expected data points (scope) that are filled with data received via questionnaire.

**Objective**: Assist in ESS internal assessment of the relevance of the questionnaire dispatch exercise for gathering data for a certain dataset. Combined with the indicator “overall response rate by questionnaire”, information about completeness of the questionnaires received may be inferred.

**Calculation**: The numerator is to be restricted to the data points received through questionnaires, instead of counting all official data points. Therefore, information to identify the data points that were obtained through questionnaires is required (e.g. SWS method flag “q”). Data from questionnaires are assumed to be official and therefore a subset of all official data; if it is not the case, this needs to be highlighted.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | # expecteddata points | # data points | # official data points | # quest. data points | Data completeness from questionnaire |
| Name | E | F | G | H | H/E |

$$\frac{Number of data points received through questionnaires (by year)}{Number of expected data points (by year)}$$

As before, this indicator is calculated for a dataset, and it may be calculated for its entire geographical scope, by region or by territory.

### Relevance and completeness of a category in a dataset

**Meaning**: No specific indicator is proposed, and therefore no *a priori* interpretation is provided in this case. The table below contains variables considered useful to analyze the data content of a category (item-element combination) in a dataset. This information can be analyzed in terms of frequencies (A-G) or values (I-J), depending on the specific needs.

**Objective**: Assist in ESS internal assessment of the relevance and completeness of a category in a dataset.

**Calculation**: For analyzing the relevance and completeness of a certain category (combination item-element) in a dataset, the following values may be provided:

* Number of territories in the geographical scope
* Number of expected data points
* Number of data points
* Number of official data points
* Sum of values of the data points
* Sum of values of the official data points

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | # territories in scope | # expecteddata points | # data points | # official data points | Σ valuesdata points | Σ valuesofficial data points |
| item-element-year | A | E | F | G | I | J |

These figures are calculated for a dataset, by item-element combination and by year. They allow subsequent calculations of proportion of imputed data points and ratio of the value of imputations (FAO 2011, annex 1; FAO 2016a, p.6).

Some values may be non-additive and therefore the sums of values cannot be calculated. In these cases, weights may be applied to the data points when counting them in order to obtain a weighed measure of completeness.

The use of weighted indicators not only for data completeness but also for response rates has been included in the next section, as a topic for future discussion.

# Topics for future discussion

* Could data points that are derived from others (e.g. data calculated as ratio, data transformed by conversion of units, data calculated by aggregation) be considered official? If so, in what cases?
* How should completeness be measured for domains that are not questionnaire-based (e.g. time series/domains that are entirely derived from other domains)?
* Should data completeness indicators be calculated in FAOSTAT database or in SWS?
A shortcoming of using SWS may be that its datasets could include data that will not be disseminated. A shortcoming of using FAOSTAT may be that the method flag required for calculating ‘completeness from questionnaire’ may not be available.
* How to deal with time series that should have data but have always been empty (i.e. no official data available and no imputations done)? Although those data should be counted in the number of expected data points, the automatic proxy described here does not include them.
One particular case would be that of countries included in the geographical scope but for which there are no data in the dataset. To identify those cases it could be informative to calculate the number of territories covered in the dataset divided by the total number of territories in the geographical scope.
* Should some form of weighted indicator be considered for inclusion together with the current indicator for monitoring response rates?

# References

EC (2007) *Handbook on Data Quality Assessment Methods and Tools. European Commission*. Eurostat. Available at: https://ec.europa.eu/eurostat/documents/64157/4373903/05-Handbook-on-data-quality-assessment-methods-and-tools.pdf/c8bbb146-4d59-4a69-b7c4-218c43952214

Trant, M. (1993) Development of data quality indicators. In: *FAO Quarterly Bulletin of Statistics*, v.6(4).

FAO (2011) *Reporting on Data Quality. Principles, Standards & Recommendations*. Statistical Standards Team, SCWG. Available at: http://intranet.fao.org/fileadmin/user\_upload/scp/Standards\_for\_quality\_
compliance/SST-03\_Data\_Quality\_Reporting.pdf;
*Annex 1* available at: http://intranet.fao.org/fileadmin/user\_upload/scp/Standards\_for\_quality\_
compliance/SST-03\_Annex\_1\_Examples\_for\_Quality\_Indicators.pdf

FAO (2014) *The FAO Statistics Quality Assurance Framework*. Available at: http://www.fao.org/docrep/019/i3664e/i3664e.pdf

FAO (2016a) *FAO Quality indicators for statistical processes: recommended indicators*. Available at: http://intranet.fao.org/fileadmin/user\_upload/scp/Standards\_for\_quality\_compliance/FAO\_Quality\_indicators\_-\_Short\_2016-11-24.pdf.

FAO (2016b) *Issues in the collection of FAO data*. Agenda Item 4. Asia and Pacific Commission on Agricultural Statistics. Twenty-Sixth Session. Thimphu, Bhutan, 15-19 Feb 2016. Available at: http://www.fao.org/fileadmin/templates/ess/documents/apcas26/docs/APCAS-16-4.2\_\_\_\_-\_FAO\_-\_Issues\_in\_the\_collection\_of\_FAO\_data.pdf

FAO (2018a) *Programme Implementation Report 2016-17*. Conference. Forty-first Session: Rome, 22-29 June 2019. Available at: http://www.fao.org/fileadmin/user\_upload/bodies/Progr\_Comm/PC\_124-documents/MV543e.pdf

FAO (2018b) *Standard Country or Area Codes for Statistical Use. Rev.2*. Available at: http://intranet.fao.org/fileadmin/user\_upload/scp/Standards\_for\_quality\_compliance/StandardsSeries\_M49\_July\_ENDORSED.docx

FAO (2018c) *IDWG on Statistics Technical Task Force Meeting. Minutes 23 May 2018*. Available at: http://intranet.fao.org/fileadmin/user\_upload/scp/IDWG/Minutes\_IDWG\_230518-final.docx

OECD (2007) *Glossary of Statistical Terms*. Available at: https://stats.oecd.org/glossary/

UK ONS (2007) *Guidelines for measuring statistical quality*. Version 3.1. Office for National Statistics. Available at: https://unstats.un.org/unsd/dnss/docs-nqaf/UK-Guidelines\_Subject.pdf

UK ONS (2013) *Guidelines for Measuring Statistical Output Quality*. Version 4.1. Office for National Statistics. Available at: https://www.statisticsauthority.gov.uk/wp-content/uploads/2017/01/Guidelines-for-Measuring-Statistical-Outputs-Quality.pdf

1. Annex 1 of the ‘Reporting on Data Quality Principles […]’ report (FAO, 2011) refers mainly to two publications: Eurostat “Handbook on Data Quality Assessment Methods and Tools” (EC, 2007) and UK “Guidelines for measuring statistical quality” (UK ONS, 2007; UK ONS, 2013). [↑](#footnote-ref-1)
2. The tree plugins related to quality indicators identified in SWS-QA are: “Quality Indicator Report”, “Quality indicators - aproduction” and “Write Quality Indicators – Production module”. [↑](#footnote-ref-2)
3. The automatic proxy intends to overcome some issues that may exist with the registration of zero (0), missing value (O) and not applicable (M) data points in SWS. For example, missing values may be recorded in SWS as 0-O, and not applicable values as 0-M. However, in some cases both missing values and not applicable values may have been recorded as blank-blank. It may also happen that zero, negligible and not applicable values all appear as missing data records (0-O) in SWS, if respondents do not appropriately flag ‘not applicable’ data points (0-M) and also leave blank the cells associated with zero values (0) and negligible values (0-N). Therefore, for the sake of comparability, some of the 0-O records (those that may represent not applicable data points) are excluded from the number of expected data points, while some of the blank–blank records (those that may represent missing values) are included. [↑](#footnote-ref-3)
4. These ‘zero’ values may include some ‘true’ values (even official values). However, they are excluded from the scope of this count (if no value different from zero appears in the entire time series) because zero values may not be registered in certain datasets while they may represent a large share in certain others. Therefore, counting them could hinder clarity and comparability in the resulting indicator. [↑](#footnote-ref-4)
5. Measuring accuracy through this indicator implies assuming that the official data disseminated are more accurate than the non-official or imputed data disseminated. FAO recommendations on reporting on data quality include the proportion of imputed data points as one of the indicators related to accuracy (FAO 2011, FAO 2016a). Notwithstanding, additional analyses of the accuracy of data could be used to complement this information. [↑](#footnote-ref-5)