

EVALUATION GUIDANCE FOR ANNUAL REPORTS UNDER REVISED EU-MAP

version EWG 25-16

Introduction

The evaluation process

The content of tables, text boxes and accompanying annexes 1.1 and 1.2 is evaluated according to the guidelines and consists of three steps.

- Step 1: pre-screening: this is done either automatically through the IT platform and/or by expert analysis based on the guidance. The result of the pre-screening is communicated to the MS and a reply with possible amendments is expected by a deadline set by DG MARE.
- Step 2: screening during the EWG meeting: this is done by expert judgement based on the guidance for the EWG experts. The result of the screening during the EWG is also communicated to the MS and a reply with possible amendments is expected by a deadline set by DG MARE and is always before the end of the EWG meeting.
- Step 3: feedback from end users (where applicable)

Basic principle - what is expected from STECF

On the evaluation of Annual Reports (AR), the DCF regulation (EU) 2017/1004 Article 11.2 says that, in accordance with Article 10, STECF shall evaluate:

- (a) the **execution** of the national work plans; and
- (b) the **quality** of the **data** collected by the Member States.

Evaluators assess the execution of the work plans and the quality of the data collected by MS.

The 'quality of data' is assessed in the data transmission monitoring tool (DTMT), supported by the information and indicators provided in work plans and annual reports. In the AR there is no reference to the statistical quality of raw data, which in certain cases are available only to end users after data calls. As such, this aspect of the quality of data is not evaluated in the AR assessment.

When evaluating the quality of reporting on data collected by MS, two aspects should be considered (ICES, 2014¹):

First – if the Workplan (WP) contains information to indicate if national sampling schemes follow good practice in terms of the three key components: statistical design, method of selecting sampling units, and estimators and method of analysis. These aspects are necessary to ensure data quality and are assessed whether the data are likely to be representative of the population, or if there may be a bias related to these design aspects.

Second – irrespective of the quality of the sampling design, its execution may deviate from the plan, for example due to high refusal rates in taking observers on board vessels or answering telephone surveys, inadequate coverage of strata etc.

Taking this into account, STECF should base the quality evaluation on:

1. General information on sampling design and implementation.
2. Specific quality metrics.

The general information is in Annexes 1.1 and 1.2, which are evaluated by STECF during the work plan evaluation.

The specific metrics are provided in Annual Reports detailing the data collection achievements. Complete quality metrics would include the contribution of stratum size to the overall size of the population (e.g. catches for the biological sampling of commercial fisheries), information related to bias (e.g. refused access to vessels or catches; evidence of non-random selection of sampling units; failures of sampling coverage), number of primary sampling units achieved in each stratum, precision of estimates etc. Some of these metrics are in the AR.

This information should be regarded carefully in relation to the design to determine if a failure or an issue exists in the final sampling data. Statistics such as numbers of PSUs sampled per stratum, precision estimates etc. have their highest value as a time-series covering as long as possible period of time. Some form of diagnostics is needed to show where and how these problems affect subsets of data (e.g. feedback from end-users in the DTMT) and the likely impact on the overall quality of the scientific advice derived from these data.

[Link between DTMT/DTi and AR](#)

STECF evaluates the quality of the data collected by MS indirectly, through issues reported in the DTMT. The metrics included in the AR may help to track the source of data transmission issues and substantiate the DTi assessment. On the other hand, the information reported by the data end users in the DTMT and the comments by the MS can give indications on the quality of the data collection activities by the MS.

Recurrent issues linked to coverage and quality from end users should be given considerations during the AR and DTMT/DTi evaluation process. To do this, the evaluator should check the history of the AR if these issues are recurrent and explained. Where relevant, the history of the AR of the specific MS is consulted to check the background of the recurrent issues.

The link of the AR tables to the DTMT is not obvious and needs to be tested.

Table 1.1 is linked with the timelines of the data submission, so if there is a timeliness issue marked in DTMT the expert would have to check if a valid justification is given in the 'AR'.

- *Tables 2.1 and 2.2* are linked to the coverage or quality of the data submission for biological data.

The explanations on the deviations regarding these tables should be scored based on expert knowledge with emphasis on the effect of deviations in relation to end-user needs. Note that any deviation may affect the transmission of data to end users, an issue that can be raised in the DTMT.

Guidance for evaluation

The evaluation is purely done on the **grey part** of the Annual report.

Step 1: Pre-screening by IT platform

This takes place in the submission phase. The platform check the consistency between the submitted AR and the WP in the platform. Most validation are run on WP only during the WP submission phase. The pre-screening by the IT platform (validation) results in errors and warnings.

The errors have to be corrected before the AR is officially submitted through the IT platform to the Commission. The warnings will not block the AR upload. The pre-screening (validation) is automatic when MS uploads the files. The validation rules are listed in the [DCF IT Platform - Validation rules](#) and can be downloaded as an csv file. They are summarised in Table 1 below. Once the AR submitted without the errors in the platform, further pre-screening may be done by the pre-screeners before the EWG.

1. The annual report is checked against the workplan. The white part of the annual report may **NOT** be changed. e.g. numbers of planned sampling are NOT to be changed in the white cells
2. Short explanations should be included in the grey cells of the annual report Tables
3. Elaborate comments should be given in the Text Boxes

Table 1: Overview of current pre-screening done by the IT platform and what is done by expert

Issue	Expert	Platform validation (WP)
Editorial (e.g. respect of master code list)		X
Identification of blanks		X
Correct year		X
Merged cells		X
All type of data		X
Syntax errors		X
Cross checks between tables		X
White part = WP/grey part = AR and is left unchanged		X
Cross checks between lines and columns per table		X
Text box completed		X
Cross checks with external sources		X
Comments in tables	X	
Pilot studies	X	
Has MS followed the guidelines	X	
Check explanations (e.g. Deviations)	X	
Achieved sampling following NWP (e.g. oversampling)	X	
Multi-annual follow-up of the sampling		X
Cross check with RWP		X
Cross check tables & text	X	
Cross check if all MS relevant areas are referred to (incl. deleted/added)		X
Cross check if all MS relevant species are referred to		X

Cross check all tables with the regulation		X
Quality (by Table)	X	

Step 2: Screening by the expert

The screening by the expert is done by pre-screeners before the EWG and by experts during the EWG. The work done by pre-screeners supports the work of the experts during the EWG, while the responsibility of the evaluation of the AR, is with the EWG and validated by STECF plenary.

The evaluation is purely done on the **grey part** of the Annual report.

General section

In case of ping-pong exchange with MS resulting in updated AR versions, the IT platform gives a new version number each time an updated document is uploaded, but the version number does not change when the documents change status from national to EC prescreening etc.

The version number of the AR in pre-screening and EWG screening must be filled in in the assessment grid.

Evaluators should check that responses of MSs to the pre-screening exercise (if any) are inserted in the revised submission (same questions to be answered) and check for inconsistencies between the pre-screening comments and MS responses. The IT platform allows to compare the AR tables between versions and across the years.

Tables and texts

The AR assessment grid master template provides a description of the guidelines for filling in the different cells.

Row 2 of the AR assessment grid provides guidance on the meaning of the column headers.

Guidelines for traffic light evaluation

The aim of these guidelines is to ensure that each expert or subgroup of experts follows the same procedure to assess the AR and to provide a judgment for the traffic light evaluation for each MS.

To facilitate the evaluation and to anticipate the future use of the DT Platform, the evaluation of reporting and the execution of the AR should be separated. A third part on quality could be added for next year's evaluation.

Reporting evaluation

Following the decision trees below for dedicated tables, a EWG judgment is obtained for each of the sections:

For Table 2.1, 2.2, 2.5:

Table is consistent with AR guidelines	Yes		EWG Judgment -> Yes	
	No	MS has fixed it during EWG, via ping-pong	EWG Judgment -> Yes	
		MS has not fixed it during the EWG	Almost the entire Table does not follow the guidelines	EWG Judgment -> No
			Main part of the Table does not follow the guidelines	EWG Judgment -> Partly
			More than 50% of the Table does follow the guidelines	EWG Judgment -> Mostly

For Textbox 2.1, 2.2, 2.5, 4.2, 4.3

Textbox is consistent with AR guidelines	Yes		EWG Judgment -> Yes	
	No	MS has fixed it during EWG, via ping-pong	EWG Judgment -> Yes	
		MS has not fixed it during the EWG	Almost the entire Textbox does not follow the guidelines	EWG Judgment -> No
			Main part of the Textbox does not follow the guidelines	EWG Judgment -> Partly
			More than 50% of the Textbox does follow the guidelines	EWG Judgment -> Mostly

Execution evaluation

For Table 2.1, first step, calculate for each species 'covered by a commercial sampling scheme for length' (column N) = 'Y':

Achievement = $\frac{\text{total number of species with positive values } (\geq 1 \text{ individuals measured})}{\text{total number of species selected for length sampling}}$

-Total number of species with positive values (at least one individual has been measured) over the total number of species selected for length sampling.

-Exclude the new rows that were added to the AR from this calculation.

Result for Table 2.1	EWG judgement
<10%	No
10-50%	Partly
50-90%	Mostly

>90%	Yes
not applicable	NA

Second step, judgement:

The following criteria to make the final EWG judgement for execution should be taken into consideration:

- Selected species for sampling with '0' achievement, when such achievement cannot be avoided due to external factors (global issues, fisheries ban and closures, no landings etc.)
- Selected species for sampling with '0' achievement covered by multilateral agreement exception for the MS which is responsible for the sampling.
- Explanation and action proposed by MS to avoid the deviations (add some example).

Table 2.2: All the species in Table 2.2 should have a positive value (Achieved number of individuals measured at national level', column P) in at least one of the biological variables that are planned.

- Exclude the new rows that were added to the AR from this calculation
- Exclude species with a triennial sampling plan.

To have a quick overview we suggest creating a Pivot table with the species (column E) in the rows, the biological variables (column G) in the columns and to calculate the sum of the achieved number of individuals measured at national level (column P). The ratio will be computed as the total number of species having more than one individual measured over total number of species. All calculations must be done per region separately.

Achievement = $\frac{\text{total number of species having more than one individual measured}}{\text{total number of species}}$

Result for Table 2.2	EWG judgement
<10%	No
10-50%	Partly
50-90%	Mostly
>90%	Yes
not applicable	NA

Second step, judgment:

The following criteria to make the final EWG judgment for execution should be taken into consideration:

- Selected species for sampling with '0' achievement, when such achievement cannot be avoided due to external factors (global issues, fisheries ban and closures, no landings etc.)
- Selected species for sampling with '0' achievement covered by multilateral agreement exception for the MS which is responsible for the sampling.
- Explanation and action proposed by MS to avoid the deviations (add some example).

Table 2.5: Planned number of PSU (column R, except 'out of frame') over the total number of PSU achieved (column X).

Achievement = $\frac{\text{planned number of PSU (column R, except 'out of frame')}}{\text{total number of PSU achieved (col X)}}$

All calculations have to be done per regions separately.

Result for Table 2.5	EWG judgement
<10%	No
10-50%	Partly
50-90%	Mostly
>90%	Yes
not applicable	NA

Quality evaluation

How to identify final evaluation criteria?

Considering the importance of the execution and in the future of the quality of execution of the WP, a higher rate should be assigned to the execution evaluation rather than in the reporting evaluation. The final EWG judgement should be based on following distribution:

$$Final\ EWG\ judgement = Reporting\ EWG\ judgement * 0.3 + Execution\ judgement * 0.7$$

In cases where the judgement is not clear (e.g. Table 2.5: One region is “Mostly” and 2 regions “No”) the judgement should be discussed in a subgroup of experts.