

14th MEETING OF THE SCIENTIFIC COMMITTEE

7 to 12 September 2026, Tórshavn, Faroe Islands

SC14-Doc14

Data working group information paper prepared by the chairperson

Chair of the Data Working Group

1. Summary

This paper is derived from the Data Working Group (DWG) proposed approach paper that was submitted at DWG01 meeting by the Chairperson and presents only those elements relevant to the Scientific Committee (SC). The Group was originally placed under the SC and now reports to the Commission.

Considering the change in the reporting line of the DWG, there is a need to define the SC's priorities and how the DWG can best support it. For that purpose the SC and its Working Groups (WG) should be consulted first, through a questionnaire prepared for the purpose (provided as an excel file)¹. It aims to record, for each WG, which data types are used and the importance of each to its analyses, the spatial and temporal resolution those analyses require, against the resolution currently received, its requirements for standardized templates and for a shared database. In addition, it requests information on the principal data gaps, ranked, with the effect of each on the advice the Group is able to provide.

The practical output could be a set of data submission templates, or revision and/or improvement of existing templates such as the ones already in use, for jack mackerel and for squid. In addition, the intention is to extend the approach to the further fisheries and data types identified by the WGs as priorities, and to determine whether those templates should be supported by a dedicated scientific database.

2. Questionnaire

The questionnaire records, for each Working Group, which data types it uses, how important each is to its analyses, and whether those data are used as reported under the CMMs or as derived estimates. It then sets the spatial and temporal resolution currently received against the minimum and preferred resolution those analyses require, together with the length of time series needed and the effect on the work where the current resolution falls short. A further section records whether standardized templates and a shared database are required, with what content, priority and format, and what access and query functions a database would have to provide. It then lists the principal data gaps, stating for each the advice that cannot be given and whether the gap is addressable through improved collection, through analytical methods, or only by decision of the Commission. A closing free-text section records the data types considered of low scientific value and the single improvement that would most improve the Group's advice.

3. Data templates

The purpose of this questionnaire is to have a starting point that would establish the requirements of the Scientific Committee's Working Groups. Once those requirements have been stated, they must be rendered into a template against which Members can submit: a defined set of fields, in a defined format, with defined units and codes.

Table 1: Example of functions of standardized data submission templates.

¹ The questionnaire is attached to this paper as Annex 1 (SPRFMO_DWG_Scientific_Data_Questionnaire.xlsx).



Function	Rationale
Comparability	In the absence of a common format, each Member submits according to its own structure, and effort that should be directed to assessment is instead expended in reconciling one submission with another. A template renders returns comparable across Members and across years.
Visibility of gaps	A template states what is expected. Where a field is returned empty, the deficiency is explicit and may be pursued, rather than being obscured by the difference between one Member's format and another's.
Validation	A defined structure permits checking on submission (of units, codes, ranges and completeness) so that errors are identified at the point of submission rather than in the course of an assessment.
Foundation for any database	A template defines fields, and fields define structure. Should the SC determine that a scientific database is required, agreed templates supply both its structure and its content from the outset, rather than the structure being designed first and the data fitted to it subsequently. Should the SC determine otherwise, the templates deliver the greater part of the practical benefit independently.

Two templates are already in use, one for jack mackerel and one for squid, which provide both a working precedent and an initial structure. The question before the SC and the WGs is therefore not whether templates are the appropriate instrument, but which further fisheries and data types require them, what those templates should contain, and in what order they should be developed. Part A of Section 4 of the questionnaire addresses precisely this, data type by data type: whether a template is required, at what priority, what it should standardize, and in what format.

The purpose of the work proposed in this paper is therefore to develop, harmonize and/or refine templates covering the data identified by the science bodies as priorities, and to determine whether those templates should be supported by a dedicated scientific database. The latter question is addressed below.

4. Scientific database

Whether the identified needs call for a dedicated scientific database is a decision for the Scientific Committee. The SC may conclude that the derived data and existing stores are sufficient, or that a purpose-built scientific database is warranted (for example to hold indices, composition or age-length key data that does not sit naturally in the current SPRFMO database).

It is necessary to be precise about what such a database would provide, since the term is frequently applied to what is in practice a shared folder of files. A database is not a file storage only: it holds data in a consistent structure, validates records as they are received, and governs which users may view and amend them. On that basis, the benefits to the science bodies would be as set out below.

Table 2: Benefits of a dedicated scientific database to the science bodies.

Benefit	Rationale
One authoritative version	All users work from a single source. Extracts presently circulate by correspondence and diverge over time, with the consequence that two analysts may reach different conclusions because they hold different copies of the same dataset. A single authoritative source eliminates that problem.
Validation on entry	Rules are enforced as data are stored, so that errors are rejected at the point of submission rather than identified by an assessor some months later. A file store accepts whatever it is given, whereas a database does not.



Benefit	Rationale
Reproducible assessments	With versioned holdings, an assessment may be re-run against precisely the data that produced the original result, and any subsequent change is recorded and attributable. This is material to peer review, to the revisiting of past advice, and to explaining why a result has altered between years.
Time returned to science	A substantial proportion of the effort preceding an assessment is expended in locating, assembling and reformatting data rather than in analysis. Query-ready holdings return that time to the work the Working Groups are constituted to perform.
Controlled access to confidential data	Role-based control over rights of access permits sensitive Member data to be held securely.
Data without an existing home	Standardized indices, age-length keys, age/length compositions and model inputs do not sit naturally within the SPRFMO database. These data are currently stored in teams.
Templates made cumulative	Templates improve each individual submission and a database behind them causes those submissions to accumulate as a consistent, queryable time series rather than as a growing set of discrete files.

5. Science related workplan

The aim is to have the questionnaire results ready in time for COMM14, so that the prioritized needs can be presented there. Depending on the decisions of the SC and the Commission, a detailed workplan will then be formulated.