
10TH MEETING OF THE FINANCE AND ADMINISTRATION COMMITTEE (FAC)

Manta, Ecuador 9, 10 & 14 February 2023

FAC 10 – Doc 06.2 Data Management Update and Development *Secretariat*

Executive Summary

An efficient and comprehensive database is critical for SPRFMO to carry out the objectives of the Convention. Science and management rely heavily on data-driven insights to monitor, assess, and develop policy frameworks for the fishery resources of SPRFMO.

This document outlines a proposal for the necessary transition to a new data services provider, as the current provider has notified SPRFMO they will no longer have capacity for continued support. In addition, a suite of changes to carry out during that transition have been proposed to strengthen the management of SPRFMO's data assets, increase transparency and efficiency, mitigate reporting burdens, and improve data reporting and dissemination mechanisms. An overview of the current data management framework, including identified limitations and challenges, is provided for context. The overview is followed by a proposed work plan spanning immediate needs to longer-term data-related proposals. Lastly, a brief proposal to establish a Data Working Group has been prepared with draft Terms of Reference.

The FAC10 is invited to consider:

- *noting the immediate need to source a new data services provider;*
- *endorsing the proposed work plan over the short and longer term;*
- *recommending that funds are secured to enable the transition the of the SPRFMO database and carry out the proposed short-term work plan; and*
- *supporting the establishment of a Data Working Group.*

1. Background

Over the past year, there were no substantive developments related to data management. As presented last year during FAC9, future database development was considered necessary but given that the Organisation had only recently recruited a new Data Manager, such development was planned for Financial Year 2023-24 to allow time for the Data Manager to assess the current state and future needs of the Organisation.

However, in November 2022, the current data services provider indicated that due to new management, the company is going in a new direction and will no longer have the capacity to support SPRFMO. This news was unexpected and will require SPRFMO to immediately secure a new data services provider and transition the current database to be hosted and supported elsewhere. This transition will require substantial effort and unanticipated costs over the coming year; however, the transition may also offer an opportunity to refresh a system that was developed more than a decade prior and has many limitations that inhibit efficiency with respect to the management of SPRFMO's data assets.

At the end of 2021, a new Data Manager was hired. In 2022, the FAC was notified regarding a one-year pause on data-related development projects until the Data Manager was acquainted with the systems (para 31 FAC9-



Report). The Secretariat has prepared this document to update the Commission on development needs, ranging from immediate needs to longer-term projects, with an aim of prioritizing funding to increase the efficiency and effective stewardship of the Organisation's data assets, and to propose a strategy to transition the database to a new provider.

Over the past few years there has been a substantial increase in the volume of data received and managed by the Secretariat. For example, with the adoption of CMM 18-2020, the data files processed annually by the Secretariat more than doubled between 2020 and 2021 (see Annex 1 of FAC8-Doc08_rev1 and FAC9-Doc09_rev2), with over 2,700 data files processed in 2021. This increasing trend has continued into 2022 (FAC10-Doc10). With respect to scientific needs, at the 2022 Jack Mackerel Benchmark Workshop, there was a recommendation to improve data management for the assessment process, and at the 10th Scientific Committee meeting there were discussions about data needs for new stock assessment objectives (i.e., jumbo flying squid) and potentially others.

To continue to support the Organisation's data needs, the data management framework should be adapted to ensure that it is appropriate to effectively manage the existing data streams; robust to future changes; reduces time burdens associated with data processing, management, and reporting; and increases accuracy and efficiency.

This document provides an assessment of the current data management framework and outlines a proposal over the short and longer-term to best meet the Organisation's needs.

2. Function and utility of the SPRFMO database

The SPRFMO database is a central repository for the Organisation's data assets ranging from the Record of Vessels to annual catches, fishing activity, observer, transshipment, and port inspection data. The database is essential to support data-driven science, management, and policy decisions to uphold the Convention.

The database serves as a tool to not only store data but to implement quality control measures, facilitate data requests and summaries, update the website with current information, and to support stock assessments and research questions pertaining to monitoring, conserving, and sustainably harvesting SPRFMO resources. The database is an invaluable resource for Members/CNCPs as well as for the Secretariat to fulfil its obligations.

3. Overview and assessment of current data management framework

3.1. Data receipt and processing

As per CMM 02-2022, different SPRFMO data streams are submitted to the Secretariat based on specified timeframes. Data templates have been provided, but data are often received in various formats including Excel, Word, PDF, and text in the body of an email. The Secretariat uploads these data to either the central database or to an Excel spreadsheet on the internal server. All data submissions have minimum data requirements and standards to be loaded to the database. As a result, there is often an iterative process between the Secretariat and Member/CNCP to correct formatting issues and data errors identified during this data processing stage. It is not uncommon for initial data submissions to have hundreds and even hundreds of thousands of fatal errors preventing a file to load without further processing/corrections. This can be a time-consuming process for both the Secretariat and the Member/CNCP providing the data.

Receipt of data is tracked through various spreadsheets and email archives.

3.2. Data storage and management

SPRFMO currently contracts with a data services provider, FINNZ, to host and manage the central database. (i.e., a SQL Server relational database). Data are uploaded via Excel spreadsheets through a web portal. The upload routines have a suite of built-in data validation checks that must be met before



the data are accepted to the database. Access to the database is available to the Secretariat via Odata feeds. These data feeds are views of the data which join tables and perform data aggregations, in some cases, for ease of use. The Secretariat can update data (e.g., if a Member has submitted a correction to previously submitted data) by removing (i.e., “rolling back”) the Excel file that was originally loaded and reuploading a revised Excel file. There is currently no direct access to the database.

Data streams that do not have dedicated tables in the central database are managed individually on the internal server in Excel files.

3.3. Database structure

The database was designed with the main data streams in mind, during the early years of the Organisation (the database was linked to the SPRFMO website and went live in 2011). These include fishing activity, observer data, record of vessels, landings, and annual catch data. In more recent years, tables to support transshipment details and mitigation measures have been added. There remain additional data streams that do not yet have the support of database tables including: VMS¹, port inspections, port sampling, transshipment notifications, toothfish tagging, and abandoned and lost fishing gear. In addition, there are some data that have been loaded into existing tables that are not entirely appropriate, out of convenience and budgetary constraints. For example, port sampling by observers have been loaded into the at-sea observer data tables, even though there are fundamental differences between these data sources.

3.4. Standardisation/normalization of data

The development of the SPRFMO database was largely done with flexibility in mind. As a result, many of the data fields were formatted as free text, even when the input data are numeric or in situations where the information should be constrained to a limited selection of possibilities (e.g., port names). Although flexibility may be attractive, this approach to data management limits the utility of the data. As SPRFMO has matured, there have been increasing requests relating to data summaries/extracts for management decisions and research questions. The lack of standardisation creates challenges for meaningful analyses. For example, if a Member was interested in landings by port, this query would be affected by numerous representations or misspellings of port names, which would require substantial pre-processing of the data. As one illustrative example of this issue, Busan, Korea is reported in the database in the following ways PUSAN, KOREA; PUSAN IN KOREA; Pusan; Busan Korea; BUSAN; Busan, korea; and BUSAN/KOR. This example is one of many.

The issue of standardization extends beyond port names and affects most data tables. Data fields that could benefit from standardisation/normalisation include but are not limited to: port names, species codes, fishing method, species stock unit, length and measurement types, estimation methods, maturity stages², life history stage, sex, interaction type for species of concern, and product description.

3.5. Data integrity

Over the years, there has been great interest in collecting data to support the management objectives within SPRFMO, at times with a focus on flexibility over data integrity. For example, we have identified a suite of ‘orphan’ records in the database. Orphan records are those that are missing critical information, and as a result, are not as meaningful as they could be. As one example, there are biological samples from observer-collected data that do not have matching species catch records from the corresponding haul, and similarly there are catches associated with fishing events that do not exist (in the data). These data inconsistencies, although not prolific, create what we are referring to as

¹ VMS data are managed by CLS. The Secretariat has access to these data through THEMIS, but the platform is not designed to be used like a database in support of historic data queries and analyses. See CTC10-Doc06 for additional details on this matter.

² Recognizing that different Members have different staging codes; however, standardisation for SPRFMO purposes would be valuable.



orphan records.

Orphan records are generally identified through comprehensive data validation routines at the outset, or when using the data and oddities or unexpected results emerge. Addressing these records would require a dedicated inquiry but would help to maximize the utility of the data that have been and continue to be collected. One proposed approach to addressing these issues would be to: i) identify the orphan records; ii) work with Members/CNCPs to fill in the missing data or edit as appropriate; iii) when correcting/filling in the missing data is not possible, a decision (perhaps through the SC) may be required to decide how to deal with those data (and what utility they may have); and iv) strengthening of the data validation routines to ensure these orphans are identified at the point of submission and not a later date.

The limitations described above pose several challenges including: the potential for inadvertent errors to enter into the process (e.g., accidental keystrokes in Excel) due to the reliance on manual data processing; uncertainties when summarising data (due to lack of standardisation and compromised database structure); inefficiency when merging different data sources (largely due to the design of the entity-relationships and lack of primary/foreign key consistency); and lack of continuity and transparency (owing largely to bespoke procedures required for many data extractions/analyses). The lack of automation and standardised procedures also means that the current framework for managing data is extremely, and perhaps unnecessarily, time-consuming.

4. Proposal for data management upgrades

4.1. Immediate and necessary transition to a new data services provider

As indicated above, SPRFMO is in an unforeseen and extraordinary situation in which the current data services provider is no longer able to support the needs of the Organisation. As a result, we must transition to a new provider over the coming months. This is a matter of urgency as we have been given a 3-month termination notice from the current provider, albeit with an indication that they will be a sufficiently flexible with that timeframe to ensure continuity during the transition of our data services.

Although this change is abrupt and was unexpected, it does offer SPRFMO an opportunity to modernise an ageing data management framework, address many of the known limitations of the current database, and potentially build in new features during the transition that are expected improve the function and efficiency of the Secretariat and SPRFMO. The Secretariat will put out a call for proposals from prospective service providers and will aim to transition as soon as possible.

The Secretariat has outlined below the key elements to be addressed during the transition.

Preliminary transition proposal

- a. *Improved database schema / design (existing tables and additional tables that are currently in Excel tables)*
- b. *Redesign the infrastructure for database server to build the database and produce reports*
- c. *Improved data quality with additional validation³ to identify orphan and anomalous records and code-based reproducible approach for addressing data issues*
- d. *Guidance (e.g., training) about database infrastructure and process for cleaning data, adding tables and setting up reports, to build Secretariat capacity*
- e. *Scoping and estimates for development of web-based efficient data ingestion*
- f. *Scoping and estimates for development of web-based public reporting site*

³ The development of additional validation would be based on the existing checks and could be supported by a SPRFMO intern.



g. Ongoing hosting and support services.

This work (items *a-d* and *g*) could potentially be completed within 3-6 months, which would allow us to move from the existing provider with relative expediency. These items are discussed in additional detail in the following sections. Items *e* and *f* are items slated for the longer-term based on Commission needs and requests. Potential changes related to *e* and *f* are explored at a high-level further down in this document.

This proposal includes a larger role for the Data Manager in the management and development of the database, as appropriate (i.e., item *d*). Historically, the Secretariat has relied solely on the service providers to manage the database, which can be costly and carried out on a slower timeframe than desired. The staffing level is an important when considering the level of responsibility and involvement the Secretariat can take on, as the Secretariat only has one individual staff member to manage the Secretariat's data and science needs (refer also to FAC10-Doc12). Even so, developing a system that is more efficient is expected to free up time to allow the Data Manager to take a more active role in the routine management of the data systems, as appropriate. This is expected to have long-term benefits and cost savings.

4.2. Short-term work plan

4.2.1. Standardise/normalize existing data and create reference tables

This upgrade should be relatively straight forward and could be accomplished within the next 3-6 months. Many of the standardisation decisions can be made without consultation; however, input from Members/CNCPs on certain standardisation approaches may be required. This task should offer substantial benefit to the accuracy and usability of the existing data and will establish greater standardisation rules for data submissions moving forward. Updated data submission templates and guidelines will be developed to accompany these changes.

4.2.2. Creation of data tables for all SPRFMO data streams

The creation of tables for and upload of all data that are outside the central database. This should not require extensive development nor testing, as most of these data streams are peripheral and relatively low-volume. Having all SPRFMO-held data assets in one place will facilitate use of these data and have benefits for long-term management.

4.2.3. Address orphan records

Identify orphan records and work with Members/CNCPs to update/review the relevant information. Work towards strengthening data validation routines to mitigate the issue of data orphans in the future. Lastly, work with Members/CNCPs and the SC to evaluate how best to deal with data records that may not comply with data standards once they have been strengthened.

4.2.4. Data dictionary/documentation

As part of the transition process, the development of a single comprehensive document to describe the features of the database as well as the fields within each of the tables would be valuable. A data dictionary would contain important metadata relating to each table and field (e.g., format, units, code summaries), but would also detail the entity-relationship for reports and data outputs.



4.3. Longer-term work plan (2+ years)

4.3.1. *Improve data ingestion*

The aim of this item is to ease the reporting burdens of Members/CNCPs, increase the efficiency of the Secretariat, and minimize reporting errors. This goal is intended to be addressed partially over the short-term with increased standardisation and improved data templates, but there are longer-term goals to move towards electronic reporting (ER). Progress toward ER will require substantial Member/CNCP input to ensure that any reporting process is fit for purpose. In addition, development of ER should be embarked upon with a vision towards the future by ensuring that any framework developed has the ability to expand and adapt to new and different reporting requirements, some anticipated and others unforeseen.

When considering a move towards ER, transshipment notifications and details are potential candidates to explore. Transshipment notifications, in particular, are a relatively high-volume data stream (over 2,000 notification data files submitted in 2021, and approximately 4,000 transshipment related emails received in 2022). In addition to the volume of these data, usability can be difficult because notifications and details are received by both the fishing vessel and the carrier vessel (or respective flag-State). Advance notifications of transshipments can be modified or cancelled, for any number of reasons. Once a transshipment occurs, it is common for the date/time, location, and species amounts, to be slightly different between the fishing and carrier vessel reports (e.g., different day, difference in the latitude or longitude positions), which creates challenges for matching up transshipment details from both parties and to also match to the notifications to assess compliance.

To improve the monitoring of transshipment activities, working towards the implementation of a web-based transshipment application, such as the one used by the WCPFC, may prove valuable for SPRFMO. Such an app could allow for the transshipment notifications to be added, modified, cancelled as necessary. It could also allow for the creation of a unique transshipment identifier, that would facilitate that matching of data submitted by the carrier and fishing vessels. This identifier could then be used to report the transshipment details, thus creating a link between the vessels involved in a transshipment, the notification, and the final transshipment details.

Although this would represent a major change in reporting procedure, we expect it would offer many benefits. In addition, such platforms have been developed for other RFMOs (some Members/CNCPs may already be using them). SPRFMO could benefit from the development and experience of those organisations, which could potentially expedite our own development, if collaboration with existing platforms was not a possibility.

Once we have developed a framework for ER, we could work with Members to gradually move the reporting of data to a web-based system, if there were perceived benefits/interest in doing so.

4.3.2. *Improve data reporting/dissemination mechanisms*

The Secretariat currently reports regularly on certain aspects of the data Members collect. Specifically, monthly catches, transshipments, jack mackerel vessel activity, Records of Vessels, SC and Commission papers, etc. These data summaries are either presented on the SPRFMO website, circulated via email correspondence, or detailed in meeting papers. There may be, however, added benefits of making data/summary reports more available to Members directly (acknowledging appropriate data sharing and confidentiality agreements outlined in CMM02) as well as to the broader SPRFMO community.

At this point, this is a general vision for greater interaction and use of the SPRFMO data assets,



and not a prescriptive proposal for what that might look like.

One approach could include moving towards more dynamic reports, available through the website. For example, instead of waiting for the monthly catch reports to be distributed, a Member could go to the SPRFMO website and query the current monthly catch data and produce a similar report, on command. Or perhaps, a Member may want to visualise a summary of their fishing activity data (e.g., plots, maps, tables) as a quick check to ensure the data are as expected (e.g., no outliers that may indicate a reporting error); features like this could be made available through the web platform.

A separate initiative could be to make the data more accessible to Members/CNCPs. Currently all data requests come through the Secretariat with custom queries filled, as appropriate and in line with the confidentiality standards outlined in CMM02-2022. There are other models of data dissemination that could be considered. For example, the [ICES data portal](#) allows individuals to query the data to produce standard data downloads, all in accordance with their data sharing agreements. SPRFMO could do something similar. This would enable Members to more easily ask questions of the data for monitoring, research, and management purposes.

5. Data Working Group

During the establishment of SPRFMO a Data and Information Working Group (DIWG) existed from 2007 until 2012. The purpose of the DIWG was to identify the types of data to be collected; develop standards for the collection, verification exchange and reporting of data; and develop standards for data security, and terms and conditions for making data available. That working group was dissolved in 2012 after developing the data standards, data submission templates, agreeing on appropriate levels for non-public domain data, setting data submission deadlines and providing input into the interim management measures. Recent discussions during the Scientific Committee suggest there may be valuable reasons to reinstate such a working group.

A Data Working Group is envisioned as a long-standing WG that may meet bi-annually, or more frequently as needs arise, to discuss issues ranging from data collection, management, and dissemination to activities more closely linked to the work of the SC to provide data analyses and investigations.

To facilitate discussions a set of draft Terms of Reference have been proposed, but with the expectation that Members/CNCPs may have additional/different ideas about the type of work this potential WG could take on to benefit SPRFMO. The WG is suggested to be placed under the purview of the SC but there may be good reasons to consider how it should fit into the wider SPRFMO structure as there are expected to be benefits and work items that may well be of interest to other subsidiary bodies (e.g., the Compliance and Technical Committee).

DRAFT Terms of Reference:

- The Data Working Group is to be established to primarily support the work of the Scientific Committee, and to support other subsidiary bodies as appropriate and specific requests from the Commission;
- The Working Group shall have a Chairperson appointed by the SC (as per SC10-Doc11);
- The Working Group shall be made up of SC Members and invited experts;
- All data made available to participants of this group, by the Secretariat, shall be treated as confidential information and used only for purposes described in these Terms of Reference;
- The Working Group should:
 - Advise on data collection, submission, and management needs to support the SPRFMO Convention
 - Take the lead on tasks to assess appropriateness of data collected to support SPRFMO management and to improve the quality and usability of SPRFMO data assets.



Specific examples may include:

- Develop and evaluate data standardization protocols and data submission templates
- Identify reporting needs/desires with a focus towards advancing electronic reporting capabilities
- Support the SC and other subsidiary bodies, by providing data summaries and exploring and preparing data analyses. Specific examples may include:
 - Develop and evaluate CPUE standardization approaches and prepared indices of relative abundance for stock assessments
 - Develop spatial mapping tools for SPRFMO fisheries
 - Update and evaluate observer coverage levels on an annual basis
 - Assess sampling effort in space and time
- The WG shall report annually to the SC (and other subsidiary bodies and/or the Commission as appropriate) on the group's activities over the previous year as well as current data-related challenges and opportunities.