NPFC-2024-SC09-WP06

SC database: concept paper

**Abstract**: The Scientific Committee (SC) has recently initiated several data-related developments including data provision templates, data inventory and suggestion to establish a Data Small Working Group. As the amount of scientific data has grown significantly, there is an emerging need to develop a database for efficient data management. This concept paper drafted by Secretariat attempts to outline some ideas and identify decision points for consideration by SC.

**Background**

Members of the SC and its subsidiary bodies share data for stock assessment and vulnerable marine ecosystems (VME) analyses (Annex). These data are usually posted on the Collaboration site in the Excel or CSV format as separate files submitted by each Member. Currently, there is no centralized database to store scientific data. The format of data submitted may differ, there is no version control of data files and no check of data completeness and quality. This results in significant time for expert groups to prepare data for analyses.

As the amount of scientific data has grown significantly over the past 8 years and will continue to grow, there is an emerging need to develop a database for efficient data management. This concept paper attempts to assist the SC in this process. It outlines some ideas for the first stage of developments and identifies some major decision points for consideration by SC and, if established, by its Small Working Group on Data.

**SC database**

**Purpose**: to efficiently and securely store, organize, and retrieve scientific data to facilitate data analyses and modeling.

**Scope**: data for stock assessment (catch, effort, size, age, maturity, etc.), VME identification, assessment of significant adverse impact (SAI) on marine ecosystems, annual catch and effort statistics and other data which may be shared in future to fulfill the SC’s functions.

Designing a database involves several steps starting from conceptual modeling to understand the problem domain, develop business requirements, identify entities, their attributes and their relationships. The output of this step is a high-level, abstract representation of the system. This working paper tries to outline elements of this process for further development by SC.

**Concept database considerations**

*Draft business requirements*

* Flexibility: adaptability to changing needs if new types of data need to be shared.
* Maintainability: easy updates and modifications, preferably internally by Secretariat or Members’ experts.
* Scalability: ability to handle increasing amounts of data, numbers of users and types of requests.
* Security: ensure that database is protected against accidental and intentional threats.
* Compatibility with the existing infrastructure: ability to connect to NPFC’s existing databases/data warehouse (NPFC account management system, vessel registry, annual report, etc.).
* Good performance and cost-effectiveness.

*Data flow*

In a very simplistic way, a database system consists of an input interface, storage and output interface. The data currently shared by Members are summarized in Data Inventory (NPFC-2024-SC09-WP03). The issues and options for consideration at this first step of development include:

* Architecture: client-server architecture versus three-tier architecture when choosing a database management system
* Input interface: bulk upload versus data fields
* Input interface: Initial data quality check and data revision procedure
* Output interface: downloading data sets for analyses, data visualization, data inventory
* Access and user roles: data submission, revision, retrieval procedures.

*Development and maintenance*

The development of a database may be conducted by a contractor directed by the SC/SWG Data and the Data Coordinator. Database maintenance can be done by the Data Coordinator who also can make modifications to the system if requested.

SC Members are invited to consider the development of a structured database for scientific data, provide feedback on this concept paper and identify next steps.

Annex

