NPFC-2024-TWG CMSA09-WP10

Scientific data management system: data inventory

**Abstract**: The intention of this paper is to develop an inventory of scientific data submitted by Members to the Secretariat and seek Members’ guidance on scientific data workflow to improve data management.

**Background**

The NPFC Performance Review recommended that the SC and the TCC each undertake a comprehensive assessment, updated annually, summarizing the NPFC data inventories and the status of data gaps and deficiencies in NPFC data and report the outcomes to the annual session of the Commission (Recommendation 4.2.5). SC08 tasked the Secretariat to draft data inventories and report progress to SC and its subsidiary bodies.

The primary objective of this paper is to develop an inventory of scientific data submitted by Members to the Secretariat. Additionally, the Secretariat seeks Members’ guidance on scientific data workflow to ensure effective data management.

**Scientific data inventory**

**Purpose**: to facilitate sharing, access and use of scientific data and information and improve data management.

**Scope**: scientific data and information provided by Members through annual reports and catch reporting systems or shared within a group of experts on the Collaboration site for stock assessment and ecosystem analyses.

**Data access and governance**

Scientific data sharing and security are regulated by two NPFC policies: (1) [NPFC Data Sharing and Data Security Protocol](https://www.npfc.int/system/files/2023-07/NPFC%20Data%20Sharing%20and%20Data%20Security%20Protocol%202023.pdf) and (2) [Regulations for Management of Scientific Data and Information](https://www.npfc.int/system/files/2023-04/Regulations%20for%20Management%20of%20Scientific%20Data%20and%20Info.pdf). They provide rules for access to the different types of data, for data collection, storage, and dissemination.

**Data inventory template**

Table 1 shows a draft scientific data inventory template. It includes the following metadata attributes:

* ID - The identifier of the data asset which is specific and unique to the NPFC.
* Title - The most common useful name by which the data asset is known by NPFC (e.g. CM CAS data).
* Description - A descriptive statement of the data asset.
* Purpose – A short description of the intention the data asset was developed for (e.g. Footprint summary, stock assessment, impact on fisheries resources, impact on ecosystems, adaptive management).
* SC group – relevant SC subsidiary body.
* Member – Who submitted the data.
* Point of Contact – Point of contact for data sharing.
* Data type – Type of data (e.g. dataset, text).
* Data format – Format of the data (e.g. csv, pdf …, DWH data warehouse)
* Date submitted/updated – The date when the data were submitted/updated.
* Temporal coverage from – The period which the data start from (year, month?).
* Temporal coverage to - The period which the data end with (year, month?).
* Spatial distribution – Spatial coverage of the data (e.g. CA, NW, CA+NW).
* Update frequency – How often the data are updated (e.g. annually, weekly).
* Location – Where the data are located (e.g. data warehouse, website, collaboration site).
* Access URL – Weblink to data storage.
* Access rights – Who has the right to access the data (e.g. Public, Members, CNCP, Expert groups, Observers).
* Data source – Where the data come from (Annual report, Catch report, Fishery, Survey, Fishery+survey, Observer).

**Scientific data workflow**

Data inventory is an essential element of data management systems. Figure 1 shows scientific data workflow in the NPFC. To ensure a sustainable data management system, it is important to (1) clarify the responsibility for collecting data, managing them and conducting analyses, and (2) develop standardized procedures for data collection, submission, management and processing.

Table 1. Scientific data inventory template



Figure 1. Scientific data workflow

