Western and Central Pacific Fisheries Commission

SCIENTIFIC COMMITTEE
TWENTY-FIRST REGULAR SESSION

Nuku'alofa, Tonga
13–21 August 2025

Data standardization to improve efficiency of WCPFC data provisions

WCPFC-SC21-2025/ST-WP-03
July 30, 2025

Paper prepared by

Tiffany Vidal[1] **and Eparama Loganimoce**[1]

[1] Oceanic Fisheries Programme of the Pacific Community

Executive summary

Ensuring timely and accurate provision of data for the work of the Commission is critical for the efficacy of this organization. Data needs continually evolve to meet growing demands for science, management, and compliance. The WCPFC aims to use the most up to date information available to support science and decision-making. As a result, members, the Scientific Services Provider (SSP), as well as the WCPFC Secretariat are increasingly operating on unreasonable short time frames to prepare information for the Commission's consideration. In recent years, members of the Commission have raised the issue of having insufficient time to thoroughly review papers and material for the meetings of the Commission and its subsidiary bodies.

One potential solution to improve the timing of data availability is to streamline and standardize the data submissions themselves. Without clear guidelines on how data should be submitted, data arrive in various formats. Processing these 'non-standard' data submissions requires time and at times interpretation, requiring correspondence with members. In addition, errors may be introduced due to the reliance on manual entry and extensive re-formatting (e.g., dates and times, species codes, conversion between weight units, etc.). In this paper, we detail several potential approaches to improve the standardization of annual reporting requirements to the Commission. Adoption of these approaches is intended to reduce the amount of time required to prepare, process, and disseminate these data for the work of the Commission, thereby freeing up time to provide more value-added services (e.g., enhancing data quality).

The SC is invited to:

1. Support the adoption of standardized data submission templates for SciData submissions; and
2. Task the SSP with further refining the proposed templates, companion data format documents, and comprehensive reference tables to facilitate successful implementation of these templates, for review at the TCC.

1 WCPFC reporting requirements and proposed standardization of data submissions

For catch and effort data, the required data submissions are hierarchical in nature, such that if 100% operational data is submitted, this information could be summarized to produced the aggregate catch and effort data, as well as the annual catch estimates (for the WCPFC area). One simple way to reduce catch reporting would be to submit full operational data[2]. Noting the challenges that may be associated with the provision of full operational data, aggregate and annual catch estimates may require separate submissions. The WCPFC scientific data reporting requirements are detailed in the Scientific data to be provided to the Commission, often referred to as 'SciData'.

For each of the key data source sections below we offer several options for standardized data submissions generally including Excel templates, e-reporting (ER), and JSON-formatted submissions via an API. These options should provide sufficient flexibility to work for all CCMs.

1.1 Annual Catch Estimates

Annual catch estimates (ACE) are catches reported for each calendar year by gear type. The reporting area differs by species, to capture stock-specific catch information, where applicable. A reporting template was developed in 2022 to streamline these submissions. A separate template must be submitted for each flag, fleet, and gear combination, which may be burdensome for members with diverse fleets and fisheries. There has been some uptake of the template, which has proven effective.

Alternatively, ACE data could be submitted in a simple, standardized Excel file. Use of this spreadsheet will require the data to be summarized appropriately into the desired format, or manually entered. We assume that for most, if not all members, the data used to generate the ACEs are stored in a database or digital file structure. If so, a simple script could be developed to format these data into the desired submission format[3]. If the template was used for manual entry, we have created data validations to facilitate reporting of key information in a standardized way (see Annex1).

	A	B	C	D	E	F	G	H
1	YEAR	GEAR_CODE	FLAG_CODE	FLEET_CODE	OCEAN_CODE	SP_CODE	RETAIN_MT	DISCARD_MT
2	2024	L	XX		WX	BET	100	5
3	2024	S	XX		WX	SKJ	200	
4	2024	L	XX		SP	ALB	4	
5								

Figure 1: Example of standardized ACE Excel template

[2] Annual catch reporting for other ocean areas, e.g., South and North Pacific, would still be required operational data outside the WCPFC area is not provided.

[3] For all proposed methodologies, the SSP is willing to assist with the initial formatting or script development to achieve the desired format.

1.2 Number of vessels

The number of active vessels in the Convention Area each year, by size class where applicable, are to be reported to the Commission annually. Although the vessel size categories vary by gear type, a single template is expected to suffice, for all gears excepting troll, which is region specific.

	A	B	C	D	E	F	G	H	I
1	year	flag_code	fleet_code	gear_code	nvess	cat1_nvess	cat2_nvess	cat3_nvess	cat4_nvess
2	2025	XX		L	46	0	22	24	0
3									

Figure 2: Example of standardized template for reporting the number of active vessels by gear and fleet on an annual basis for gears other than troll

The SciData specifies the vessel size class categories, by gross registered tonnage (GRT), to be reported annually. Those size classes are as follows:

- Longline: 0–50, 51–200, 201–500, 500+
- Pole-and-line: 0–50, 51–150, 150+
- Purse seine: 0–500, 501–1000, 1001–1500, 1500+

For troll vessels there is additional reporting by ocean area. Specifically,

- the WCPFC Statistical Area south of the Equator
- the WCPFC Statistical Area north of the Equator

and for troll vessels targeting albacore in the Pacific Ocean south of the Equator,

- WCPFC Statistical Area south of the Equator
- the Pacific Ocean south of the Equator.

For the South Pacific, the reporting of troll vessels is for the fishing period from July to June as opposed to the calendar year for other gears and areas.

	A	B	C	D	E	F	G	H	I
1	year	flag_code	fleet_code	gear_code	nvess	cat1_nvess	cat2_nvess	cat3_nvess	cat4_nvess
2	2025	XX		L	46	0	22	24	0
3									

Figure 3: Example of standardized template for reporting the number of active troll vessels by fleet and area, on an annual basis

1.3 Aggregate Data

Aggregate catch and effort data are provided when 100% operational logbook data are not. Aggregate data represents raised catch and effort data that capture all fishing activity by a fleet in a given year, spatially and temporally disaggregated. The details of the aggregate data vary by gear

type but the format and general information to be provided is largely consistent. Longline data are aggregated at a monthly and 5°x5° resolution whereas for purse seine, data are aggregated at a monthly and 1°x1° resolution. It should be noted that the SSP uses the **southwest corner** as the reference for spatial cells.

For aggregate data we are proposing a single flat file (wide format) as the initial template. An early version of the proposed templates for each gear are detailed in the attached documents.

1.3.1 Purse seine

For purse seine, the format is relatively straightforward with the flag, fleet, year, month, latitude and longitude (decimal degrees at 1°x1° resolution, with the coordinates representing the southwest corner), activity code (e.g., fishing set, searching, transiting; see activity reference table), school type (if a set is made whether it was a free school, drifting FAD, anchored FAD, etc. set type; see school type reference table), days associated for each activity and set type, and the number of sets associated for fishing set activities. The species column headers are the 3-letter FAO species codes for all of the key species that were caught in a given year. In the template, the column with header '**...**' indicates where additional species could be added as new columns, as necessary. All species catches are to be reported in metric tonnes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	flag_code	fleet_code	year	month	latitude	longitude	activity_code	school_type	days	sets	ALB	BET	SKJ	YFT	FAL
2	XX	XX	2025	1	6	-174	1	4	12	10	0	90	210000	3200	2
3															

Figure 4: Example of standardized purse seine aggregate catch and effort Excel template

1.3.2 Longline

Longline is very similar to the reporting for purse seine except that additional stratification by hooks between floats (HBF) is included and hooks are reported as the key effort metric. When HBF is not available, this field can be left blank. In the aggregate database tables there is one version without HBF (as it is not always available) and one with. Here the HBF tables would represent a subset of the main aggregate table, for that reason. For reporting purposes, if reported together, we will aggregate the catch and effort without HBF to produce the main table, thereby capturing the rows for which HBF is unknown.

Another difference between purse seine, is that aggregate catch by numbers is also reported for longline. Therefore, in this template we have two tables for catch and effort - one for reporting numbers and the other for reporting catch in metric tonnes. The tables are identical otherwise.

Similar to purse seine there is a tab dedicated to reporting of the number of active vessels by flag, fleet, year, month, latitude and longitude (5°x5°), to the extent possible.

	C	D	E	F	G	H	I	J	K	L	M	N	O
1	year	month	latitude	longitude	activity_code	hbf	hooks	days	sets	ALB	BET	SKJ	YFT
2	2025	1	-5	165	1	19	5240	12	10	642	720	0	15285
3													

	C	D	E	F	G	H	I	J	K	L	M	N	O
1	year	month	latitude	longitude	activity_code	hbf	hooks	days	sets	n_ALB	n_BET	n_SKJ	n_YFT
2	2025	1	-5	165	1	19	5240	12	10	46	12	0	261
3													

Figure 5: Example of standardized longline aggregate catch and effort Excel templates (top: reporting in mt; bottom: reporting in numbers)

1.4 Operational Data

Operational (logbook) data provide high-resolution, event-level information on fishing activities. Each record generally corresponds to a single fishing activity (e.g., set, haul, or day) and contains critical information required for scientific analysis and compliance monitoring including date, precise location (latitude and longitude), effort metrics (e.g., number of hooks), gear specifications, and species-specific catch (by weight and/or number).

These data support a range of data analyses, including catch-per-unit-effort (CPUE) standardization, habitat modeling, species interaction studies, and compliance checks. Due to the fine-scale temporal and spatial granularity of these data, they are key inputs to regional stock assessments and ecosystem analyses.

Data submitted in non-standard formats introduce significant barriers to efficient integration and analysis as well as the timeliness of data availability. Specific examples of these issues include, but are not limited to:

- inconsistent field naming conventions;
- inconsistent position formats (e.g., S00-S05, N00-N05, 10°15'S, 140W, 140 with separate column to indicate N/S, etc.);
- merged fields (e.g., YFT/42.5kg, 08:15 12-Mar-2024)
- species names versus FAO codes;
- wide versus long format;
- missing fields and anomalous values;
- inconsistent coding (e.g., UTF-8 and UTF-16 files; special characters may break parsers); etc.

1.4.1 Proposed reporting options for operational data

1. **E-reporting (preferred):** The simplest option for timely, standardized reporting of operational data would be to use one of the regional ER applications. The SSP has developed an ER app for longline logbook operations OnBoard that is free to use and seamlessly integrated

into the databases housed by the SSP, on behalf of the Commissions. FIMS has also developed ER apps for purse seine and longline and they have worked closely with SPC to transfer those data directly to our databases. It should be noted that the iFIMS and SPC systems are not fully integrated, and there is additional work to be done to ensure consistency between the systems, but both offer practical solutions for e-reporting.

The use of these tools does depend on hardware access and connectivity and may require initial training and planning for national-level implementation.

2. **JSON format aligned with WCPFC ER standards:** Another option, albeit more technical, is the use of the SSP-developed JSON format to submit data from any third-party system (e.g., a national ER application) via an API. The JSON format is well suited to submit individual trips, as the different levels – trip, sets, catches – are grouped in a hierarchical way. This is a highly structured, machine-readable format that enables direct integration into our data systems. The JSON standards are published on a public domain website. This approach is best suited for regular data submissions, for example, after each trip.

   These formats require an initial development phase to map existing operational data into the standard JSON schema, but once set up, they require minimal maintenance and can support automated data exchanges reliably. The initial setup may require more advanced coding and processing knowledge; however, the SSP is willing to assist members to implement this approach, as necessary.

3. **Table format (normalized) with separate trip, set, catch files:** This option aligns more closely with relational database structures where the data are hierarchical in nature, typically one for each level of the data hierarchy (e.g., trip, set, catch). Records are linked using unique identifiers (such as primary keys or common fields) to reduce redundancy associated with repeating information e.g., from the trip level for every catch record. To use this format, the data provider must ensure that each record at every level (trip, set, catch) has a unique identifier, and that these identifiers are used to link related records across files. This challenge can be easily managed with initial planning and training, if necessary. In the templates provided there is no universal standard for a trip identifier, but a standard could be developed (e.g., vessel name combined with the depart date, in a standardized manner). These details could be further evaluated and discussed during the TCC, as necessary.

4. **Wide flat file format** This approach may offer the simplest approach as each row represents an individual event (e.g., set, searching, or transiting activity) including catch, with the full trip information repeated in each row (e.g., vessel name, departure date). The key species to report would be individual columns to enable reporting all catches (in numbers and or weight, in separate columns) in one row. It would be possible to add rows as necessary, for example to accommodate reporting of species not explicitly listed in the SciData (e.g., turtles). This approach may be comfortable or familiar for some, but it does present a less

efficient alternative than the normalized approach described above.

1.5 Size composition data

Most size data reported to the Commission come by way of observer data and port sampling. This paper will not discuss observer data submissions, but will focus exclusively on aggregate size data submitted to the Commission as part of the annual data submission. The aggregate size data should not include observer data that have been submitted through standard observer reports to ensure data are not double counted.

Length and/or weight composition data that are representative of catches by the fisheries shall be provided to the Commission at the finest possible resolution of time period and geographic area and at least as fine as periods of quarter and areas of 20° longitude and 10° latitude. In addition to the data reporting, it is important to also notify the Commission of the following (from the SciData):

1. CCMs shall indicate whether lengths and/or weights are rounded up or rounded down to the unit.
2. The statistical and sampling methods that are used to derive the size composition data, including reference to whether sampling was at the level of fishing operation or during unloading, details of the protocol used, and the methods and reasons for any adjustments to the size data.
3. Information on operational changes in the fishery that are not an attribute in the data provided is to be listed and reported with the data provision.

There are two size data submission templates proposed, one for length data and one for weights. The common fields in both templates include year, month, fishing gear, flag, fleet, latitude, and longitude. For aggregate data, the preferred temporal resolution for all gears is monthly, but the spatial resolution varies by gear, as in the aggregate catch and effort data. For longline and troll, the preferred spatial resolution of these data is at 5°x5°; however, submissions at finer resolutions would be acceptable. For purse seine, pole-and-line, and other gears (e.g., ringnet, handline, gillnet, etc.) the preferred spatial resolution is 1°x1°. Latitude and longitude should represent the southwest corner of the spatial cell. All length measurements should be reported in cm whereas weights should be reported in kilograms.

In both the length and weight templates there is a field to report the measurement type. This is an important field as it allows the SSP to convert lengths and weights to comparable units for use in scientific analyses. A reference table detailing the appropriate measurement types to report are included in the reference tables in Annex A.2. Brogan (2002) developed a valuable port sampling manual which may serve as a useful guide for CCMs.

	A	B	C	D	E	F	G	H	I	J	K
1	year	month	gear_code	flag_code	fleet_code	lat	lon	sp_code	length_cm	freq	meas_type
2											
3											

	A	B	C	D	E	F	G	H	I	J	K
1	year	month	gear_code	flag_code	fleet_code	lat	lon	sp_code	weight_kg	freq	meas_type
2											
3											

Figure 6: Example of standardized size data template for reporting length frequencies (top) and weight frequencies (bottom)

Operational and aggregate data can be submitted together in the same template, as these data are combined into a single consolidated size data for scientific analyses.

2 Future developments

The proposals for standardized data submission templates detailed in this document offer simple options that could be implemented quickly to offer significant benefits to the Commission. These options may be enough, but the SSP is also interested in developing more sophisticated systems to further enhance the submission process, data review, and reporting needs. Although, the details have not yet been developed, some considerations may include: web-based data submissions (which would enable efficient tracking of submissions), built-in data quality checks (for real-time feedback to members on potential data issues); and greater access to data and data summaries for the work of the Commission (e.g., for compliance monitoring). We anticipate such a system would have significant value in terms of ease of meeting and tracking data reporting requirements, enhanced data quality, improved timeliness of data availability, and increased transparency around data. These components are expected to be part of a broader data management platform the SSP is envisioning. We would value feedback from CCMs on the design of such a system to ensure your needs and desires are addressed as part of this development.

3 Proposal and next steps

This document serves an initial guide, offering simple data submission templates to enhance the efficiency of processing and preparing Commission data to ensure more timely availability of data for the work of the Commission. We are seeking the SC support for this initiative and invite the SC to request the SSP to further develop these templates by preparing separate data format documents with comprehensive reference tables, for ease. These templates can be further refined, based on input from CCMs and the Commission, and could be ready for implementation by the end of the year, if approved by the Commission.

4 References

Brogan, D. (2002). *Port sampling manual*. Pacific Community (SPC).

Appendix - Reference Tables

A.1 Purse seine school association codes

Table A1: School Association Descriptions and Codes

school_association_desc	**schass_code**
No code	0
Unassociated	1
Feeding on baitfish	2
Drifting log, debris or dead animal	3
Drifting raft, FAD or payao	4
Anchored raft, FAD or payao	5
Live whale	6
Live whale shark	7
Other (please specify)	8
No tuna associated	9

A.2 Measurement types for size data

Table A2: Measurement codes, descriptions, and size data types (i.e., length or weight) for size data

meas_code	meas_desc	meas_type
AN	AN Anal fin length	L
BF	BF Bill to fork in tail	L
BL	BL Beak Length	L
CC	CC Curved carapace length	L
CK	CK Cleithrum to anterior base caudal keel	L
CL	CL Carapace length (turtles)	L
CW	CW Carapace width	L
CX	CX Cleithrum to caudal fork	L
EO	EO Posterior eye orbital to caudal fork	L
EV	EV Posterior eye orbital to vent	L
FF	FF 1st dorsal to fork in tail	L
FS	FS 1st dorsal to 2nd dorsal	L
GI	GI Girth	L
LF	LF Lower jaw to fork in tail	L
NM	NM Not measured	L
OW	OW Observer's estimate	L
PC	Nose to anterior tail portion (sharks)	L
PF	PF Pectoral to fork in tail	L
PS	PS Pectoral to 2nd dorsal	L
SC	SC Straight carapace length	L
SL	SL Tip of snout to end of caudal peduncle	L
T1	Finlet 1 to tail fork	L
T2	Finlet 2 to tail fork	L
TH	TH Body thickness (width)	L
TL	TL Tip of snout to end of tail	L
TW	TW Total width (tips of wings - rays)	L
UF	UF Upper jaw to fork in tail	L
US	US Upper jaw to 2nd dorsal	L
WL	WL Wing length tip of wing to wrist (birds)	L
FN	FN Weight of all fins (sharks)	W
FW	FW Fillets weight	W
GF	GF Gilled, gutted, headed, flaps removed	W
GG	GG Gilled and gutted	W
GH	GH Gutted, headed	W
GO	GO Gutted only, not gilled	W
GT	GT Gilled, gutted, tailed	W
GX	GX Gutted, headed, tailed	W
LW	LW Loin weight	W
OC	OC Weight of gills and guts only	W
TT	TT Trunk weight	W
WW	WW Whole weight	W

A.3 Fleet codes

Fleets are a construct to identify groups of vessels that may fish in similar ways as compared to other vessels. This designation has historically been used to identify coastal versus distant-water fleets or sub-fleets that fish under a common flag, such as the US fleets based in Hawaii, American Samoa, Guam, or the Northern Mariana Islands. These fleet definitions are used to group catch, effort, and at times selectivity or catchability within assessment models. For most flag-states, there are no sub-fleets identified. It is possible for new fleets to enter the fisheries, and when that happens, it would be helpful to notify the Secretariat of the characteristics of these fleets. Otherwise, using the fleet definitions as detailed below, is suggested.

Table A3: Fishing Fleet Codes and Descriptions by Gear Type

gear_code	**flag_code**	**subfleet_code**	**fleet_desc**
H	US	HW	Hawaii (US) Handline
L	AU	AU	Australia domestic and chartered longline
L	AU	JV	Australian joint-venture Longline
L	CN	DW	China distant-water longline
L	JP	CS	Japan coastal fleet longline
L	JP	JP	Japan offshore and distant water longline
L	TW	DW	Chinese Taipei distant-water longline
L	TW	OD	Chinese Taipei offshore domestic longline (REI)
L	TW	OS	Chinese Taipei small-scale Longline (STLL)
L	US	AS	American Samoa longline
L	US	GU	Guam (US) Longline
L	US	HW	United States of America longline (Hawaii)
L	US	MP	Northern Marianas Islands (US) Longline
L	US	US	United States of America longline
O	PF	PM	French Polynesia Potimarara
P	JP	CS	Japan coastal fleet pole-and-line
P	JP	DW	Japan offshore and distant water pole-and-line
P	US	HW	Hawaii pole-and-line
S	AU	AF	Australia purse seine, inside the Australian FIS
S	AU	EX	Australia distant-water purse seine
S	ID	DW	Indonesia distant-water purse seine
S	ID	ID	Indonesia domestic purse seine
S	JP	CS	Japan coastal fleet purse seine
S	JP	DW	Japan offshore and distant-water purse seine
S	PH	DW	Philippines distant-water purse seine
S	PH	PH	Philippines domestic purse seine
T	US	AS	American Samoa (US) Troll
T	US	GU	Guam (US) Troll
T	US	HW	Hawaii (US) Troll
T	US	HW	Hawaii (US) Troll and Handline
T	US	MP	Northern Mariana Islands (US) troll