

# CPR product

## Product Specification

---

### Document Change Control

Version Number	Date of Issue	Brief Description of Change
Version 0.1	22-06-2021	Draft version for review and comment
Version 0.2	03-08-2022	Final

# Introduction

## Summary

This dataset includes Continuous Plankton Recorder (CPR) data, as part of the Australian Plankton Survey (AusCPR) and phytoplankton from the Southern Ocean Continuous Plankton Recorder Survey (SO-CPR). The dataset comprises the phytoplankton colour index (PCI), phytoplankton & zooplankton abundance, a biomass index and biovolume measurements. The biomass index is the dried mass of the plankton washed off the silk and dried in an oven at 60°C for 24 hours. This, therefore, represent the total plankton biomass. The PCI is an estimate of the amount of chlorophyll in the water. The colour of the silk is measured using Pantone colour charts against a numeric scale indicating increasing greenness or chlorophyll.

- 0 - No colour
- 1 - Very Pale Green
- 2 - Pale Green
- 3 - Green

To understand how CPR plankton data is analysed and how to use CPR data we recommend reading Richardson et al., 2006.

The data are made available in binned products for ease of use. As taxonomy changes over time and the ability of analysts improves with training more and more species can be identified over time. The products account for these changes and for real absences from the samples. A 0 abundance means that the species was looked for and not seen, -999 means that the species was not looked for in that sample. The products are therefore applicable for use in time series analysis.

There are a total of 15 products to choose from.

Phytoplankton data are available as either abundance or biovolume, and each of these at 4 levels of binning:

- Raw product (includes all data)
- Species product (includes all data where species have been identified)
- Genus product (includes all data binned to genus level)
- Higher taxonomic group (includes all data binned to functional groups).

Zooplankton abundance products are available at 5 levels:

- Raw product (includes all data)
- Copepods product (all data where species have been identified, Copepods only)
- Non-Copepods product (all data where species have been identified, non-Copepods only)
- Genus product (includes all data binned to genus level)
- Higher taxonomic group (includes all data binned to functional groups).

## Product Specifications

### General information

<b>Product name</b>	CPR Phytoplankton Abundance - Higher Taxonomic Groups CPR Phytoplankton Abundance - Genus Data CPR Phytoplankton Abundance - Species Data CPR Phytoplankton Abundance - Raw Data CPR Phytoplankton Biovolume - Higher Taxonomic Groups CPR Phytoplankton Biovolume - Genus Data CPR Phytoplankton Biovolume - Species Data CPR Phytoplankton Biovolume - Raw Data CPR Zooplankton Abundance - Higher Taxonomic Groups CPR Zooplankton Abundance - Genus Data CPR Zooplankton Abundance - Copepods Data CPR Zooplankton Abundance - Non-Copepods Data CPR Zooplankton Abundance - Raw Data
<b>Geographic coverage</b>	Australian coastal, open ocean region and Southern Ocean
<b>Temporal coverage</b>	2007 -- present
<b>Temporal resolution</b>	Variable
<b>Update frequency</b>	Weekly
<b>Delivery mechanism</b>	AODN Portal OGC WFS
<b>Format</b>	CSV

### Details of product contents

Name	Description	Units/format	Data type
TripCode	Code for the sampling trip	<Region_code>YYYYM MDD	string
Sample_ID	Unique code for the sampling trip		string
Region	Region where the vessel performed the sampling		string
Latitude	Nominal latitude (North) of the location where sample was collected	Degrees	double precision
Longitude	Nominal longitude (East) of the location where sample was collected	Degrees	double precision
SampleTime_UTC	UTC date and time of sample collection	YYYY-MM-DD	timestamp

		hh:mm:ss	mp
SampleTime_Local	Local date and time of sample collection	YYYY-MM-DD hh:mm:ss	string
Year_Local	Year of sample collection	YYYY	integer
Month_Local	Month of sample collection	MM	integer
Day_Local	Day of month of sample collection (local time)	DD	integer
Time_Local24hr	Local 24hr time of sample collection	HH:MM	string
SatSST_degC	Temperature of the water column near the surface, from matching satellite measurement	Degrees Celsius	floating point
SatChlaSurf_mgm3	Chlorophyll concentration from matching satellite measurement	mg m <sup>-3</sup>	floating point
PCI	Phytoplankton colour index		integer
BiomassIndex_mgm3	Plankton biomass	mg m <sup>-3</sup>	floating point
SampleVolume_m3	Volume of sampled water	m <sup>3</sup>	floating point
[taxonomic level name]  (e.g., "Noctiluca scintillans")	Taxonomic level abundance	Cells per litre	floating point

## Product delivery

### Format

The parameters of this product are available in comma-separated values (CSV) format.

### AODN Portal

Dataset collections:

- **IMOS - Phytoplankton Abundance and Biovolume (CPR)**
- **IMOS - Zooplankton Abundance and Biomass (CPR)**

The specific product is selected under the *Download as...* drop-down menu on Step 3.

The collections can be subset by

- Geographic bounding box
- Temporal range

## AODN Geoserver

The products can be accessed via a Web Feature Service (WFS) query to the [AODN Geoserver](#). The relevant layers are:

- imos:cpr\_phytoplankton\_abundance\_raw\_data
- imos:cpr\_phytoplankton\_abundance\_species\_data
- imos:cpr\_phytoplankton\_abundance\_genus\_data
- imos:cpr\_phytoplankton\_abundance\_htg\_data
- imos:cpr\_phytoplankton\_biovolume\_raw\_data
- imos:cpr\_phytoplankton\_biovolume\_species\_data
- imos:cpr\_phytoplankton\_biovolume\_genus\_data
- imos:cpr\_phytoplankton\_biovolume\_htg\_data
- imos:cpr\_zooplankton\_abundance\_raw\_data
- imos:cpr\_zooplankton\_abundance\_copepods\_data
- imos:cpr\_zooplankton\_abundance\_noncopepods\_data
- imos:cpr\_zooplankton\_abundance\_genus\_data
- imos:cpr\_zooplankton\_abundance\_htg\_data

The Web Map Service (WMS) layers used for preview in Step 2 on the AODN Portal are

- imos:cpr\_phytoplankton\_map
- imos:cpr\_zooplankton\_map

## Data Lineage

### Provenance

The CPR is a robust instrument that can be easily deployed and retrieved in minutes while a vessel is under way. This means that CPRs can be deployed from ships of opportunity, comprising a combination of volunteer commercial and research vessels.

### Harvest to AODN

The phytoplankton products are based on raw data harvested from the following WFS layers in the CSIRO Geoserver:

- **cpr\_samp**- metadata for sampling trips and samples taken, plankton biomass index and PCI. Sea surface temperature and chlorophyll from satellite measurements.
- **cpr\_phyto\_raw** - sample, observed abundances, biovolumes and taxonomic details of individual taxa identified in each sample
- **cpr\_phyto\_changelog** - details when specific taxa were first looked for and first observed
- **cpr\_zoop\_raw** - sample, observed abundances and taxonomic details of individual taxa identified in each sample
- **cpr\_zoop\_changelog** - details when specific taxa were first looked for and first observed
- **phytoinfo** - metadata related to phytoplankton, it will be stored in S3 at IMOS/BGC\_DB/harvested\_from\_CSIRO to be publicly available.
- **zoopinfo** - metadata related to phytoplankton, it will be stored in S3 at IMOS/BGC\_DB/harvested\_from\_CSIRO to be publicly available.

Each of these layers is downloaded into a CSV file, then processed by the AODN data ingestion pipeline.

### Processing at AODN

Harvest all input data into the Harvest database

Create **Raw Data** product:

1. Join sample metadata onto raw data on Sample code
2. Pivot unique values of TaxonName to separate columns

Create **Higher Taxonomic Groups** product:

1. Group the raw data by Sample and TaxonGroup, summing up the phytoplankton abundance values
2. Remove rows with TaxonGroup in ('Other', 'Coccolithophore', 'Diatom', 'Protozoa')
3. Join metadata
4. Make a dummy line of data for samples that have nothing recorded against them and replace nulls in phytoplankton abundance with zero
5. Pivot unique values of TaxonGroup to columns, with values set from phytoplankton abundances

Create **Genus Data** product

1. From the taxon changelog find the TaxonNames where the genus was affected (i.e., genus in TaxonName  $\neq$  genus in ParentName)
2. From the raw data with genus not affected by changelog (TaxonName not in the list from step 1.):
  - a. Group by Sample and Genus, summing up the phytoplankton abundance values, and dropping rows where Genus is empty
  - b. Join sample metadata using Sample code
  - c. Replace missing values in phytoplankton abundance with zero
  - d. Group by all sample metadata columns (Station, Lat/Lon, date/time) and Genus, summing up the phytoplankton values
3. For raw data affected by the changelog
  - a. For each Genus observed,
    - i. replace missing StartDate with value from the changelog from parent taxon
    - ii. set default abundance values (NULL for 'not looked for', 0 for 'not found')
  - b. Group by all sample metadata columns (Station, Lat/Lon, date/time) and Genus, summing up phytoplankton abundance values
  - c. Append to the other data from step 2.

Create **Species Data** product

1. As per genus data product but based on taxon identified to species, i.e., group by genus and species and drop all taxa where species is not a defined species.

### AODN Harvest Database schema

**Schema name:** imos\_cpr\_db

Tables to hold data & metadata harvested from CSIRO:

- cpr\_samp
- cpr\_phyto\_raw
- cpr\_phyto\_changelog
- cpr\_zoop\_raw
- cpr\_zoop\_changelog
- satellite\_measurements

Materialised views for Geoserver layers:

- cpr\_phytoplankton\_abundance\_raw\_data
- cpr\_phytoplankton\_abundance\_species\_data
- cpr\_phytoplankton\_abundance\_genus\_data
- cpr\_phytoplankton\_abundance\_htg\_data
- cpr\_phytoplankton\_biovolume\_raw\_data
- cpr\_phytoplankton\_biovolume\_species\_data
- cpr\_phytoplankton\_biovolume\_genus\_data
- cpr\_phytoplankton\_biovolume\_htg\_data
- cpr\_zooplankton\_abundance\_raw\_data
- cpr\_zooplankton\_abundance\_copepods\_data
- cpr\_zooplankton\_abundance\_noncopepods\_data
- cpr\_zooplankton\_abundance\_genus\_data
- cpr\_zooplankton\_abundance\_htg\_data

Views for Geoserver layers:

- cpr\_phytoplankton\_map
- cpr\_zooplankton\_map

## Glossary and abbreviations

AODN	Australian Ocean Data Network
CSV	Comma-separated values
CTD	Conductivity Temperature Depth
IMOS	Integrated Marine Observing System
NRS	National Reference Station
QC	Quality Control
WFS	Web Feature Service

## References

Richardson AJ, Walne AW, John AWG, Jonas TD, Lindley JA, Sims DW, Stevens D, Witt M (2006) Using continuous plankton recorder data. *Prog Oceanogr* 68:27-74