**Data extraction CPR data around PAP**

**Extraction performed by Pierre Hélaouët**

**11/08/2025.**

**DOI: 10.4124/68a5862f928f6**

**Web page: https://doi.mba.ac.uk/data/3538/**

**Note:** As always, this dataset has been carefully built and checked accordingly. However, it is the user’s responsibility to perform his own verifications.

**Quick description of the dataset**

**1 – The dataset contains 12 files:**

1. “CPR\_Data\_CPRaroundPAP\_11082025.docx”: This document
2. “CPR\_ CPRaroundPAP\_ControlMap\_11082025.png”:

Map representing the selected samples from January 1980 to December 2021 (9861 samples)

1. “CPR\_ CPRaroundPAP\_Data\_LargeZooplankton\_11082025.csv”: Abundance data for all selected large zooplankton (>= 2mm) (83 taxa, see CPR\_ CPRaroundPAP\_List\_LargeZooplankton\_11082025.csv) and all selected samples in the selected area (26°N to 54°N, -22°E to -11°E).

Rows: All samples for the selected area (9861 samples).

Column 1: Unique sample id. For instance: “240B--27” corresponds to the 27th sample for the 240th transect on the B route.

Columns from 2 to 8: Spatio-temporal coordinates for each sample.

Columns from 9 to 91: Abundance data for all selected large zooplankton (83 taxa).

Note 1: We may notice very small values (10^-10) or any other number with a very small fraction. Sometimes, our analysts can identify the presence of a specific taxa but are unable to quantify it. In that case, they report the taxa as “present”. This is hard-coded in our database as a very small value (10^-10) for statistical reasons.

Note 2: In a given sample, the abundance value of a specific taxon, is set to NaN (Not A Number) when the corresponding Data of Routine Identification (DRI) is posterior to the date of sample collection.

1. “CPR\_ CPRaroundPAP\_List\_LargeZooplankton\_11082025.csv”: List of large zooplankton

Rows: All selected taxa (83 taxa).

Column 1 “accepted\_id”: Unique identifier used by the CPR survey

Column 2 "Aphia\_id”: Identifier used by WoRMS

Column 3 "name-CPR”: Unique name used by the CPR survey.

Column 4 "Name\_worms”: Name used by WoRMS corresponding to the “aphia\_id”.

Column 5 "DRI”: Date of Routine Identification. Before that date, un taxon was not on our routine taxa list. For a given taxon, abundances associated with samples taken before the DRI are set to a NaN (Not A Number).

1. “CPR\_ CPRaroundPAP\_Taxonomy\_LargeZooplankton\_11082025.csv”: Taxonomy of all large zooplankton taxa selected.

Rows: All selected taxa (83 taxa).

Column 1 “accepted\_id”: Unique identifier used by the CPR survey

Column 2 "Aphia\_id”: Identifier used by WoRMS

Column 3 "name-CPR”: Unique name used by the CPR survey.

Column 4 "Name\_worms”: Name used by WoRMS corresponding to the “aphia\_id”.

Column 5 “counting\_method”: 1 for large zooplankton (eyecount), 2 for small zooplankton (traverse), 3 for phytoplankton.

Column 6 “is\_copepod”: flag taxa categorised as copepods

Column 7 “is\_diatoms”: flag taxa categorised as diatoms

Column 8 “is\_dinos”: flag taxa categorised as dinoflagellates

Column 9 “is\_total”: flag taxa considered as a group of different entities

Column 10 to 35: taxonomy.

1. “CPR\_ CPRaroundPAP\_Data\_SmallZooplankton\_11082025.csv”: Abundance data for all selected small zooplankton (<2mm) (see CPR\_ CPRaroundPAP\_List\_SmallZooplankton\_11082025.csv) and all selected samples in the selected area. (63 taxa, 9861 samples).

Note: Same architecture as “CPR\_CPRAROUNDPAP\_Data\_LargeZooplankton\_11082025.csv”

1. “CPR\_ CPRaroundPAP\_List\_SmallZooplankton\_11082025.csv”: List of small zooplankton (63 taxa).

Note: Same architecture as “CPR\_ CPRaroundPAP\_List\_LargeZooplankton\_11082025.csv”.

1. “CPR\_ CPRaroundPAP\_Taxonomy\_SmallZooplankton\_11082025.csv”: Taxonomy of all small zooplankton taxa selected (63 taxa).

Note: Same architecture as “CPR\_ CPRaroundPAP\_Taxonomy\_LargeZooplankton\_11082025.csv”

1. “CPR\_ CPRaroundPAP\_Phytoplankton\_11082025.csv”: Abundance data for all selected small zooplankton (see CPR\_ CPRaroundPAP\_List\_Phytoplankton\_11082025.csv) and all selected samples in the selected area. (128 taxa, 9861 samples).

Note: Same architecture as “CPR\_CPRAROUNDPAP\_Data\_LargeZooplankton\_11082025.csv”

1. “CPR\_ CPRaroundPAP\_List\_Phytoplankton\_11082025.csv”: List of small zooplankton (128 taxa).

Note: Same architecture as “CPR\_ CPRaroundPAP\_List\_LargeZooplankton\_11082025.csv”.

1. “CPR\_ CPRaroundPAP\_Taxonomy\_Phytoplankton\_11082025.csv”: Taxonomy of all small zooplankton taxa selected (128 taxa).

Note: Same architecture as “CPR\_ CPRaroundPAP\_Taxonomy\_LargeZooplankton\_11082025.csv”

1. “CPR\_CPRaroundPAP\_Data\_PCI\_06082025.csv”: Abundance data for Phytoplankton Colour Index (PCI).

Rows: All samples for the selected area (252890 samples).

Column 1: Unique sample id. For instance: “240B--27” corresponds to the 27th sample for the 240th transect on the B route.

Columns from 2 to 8: Spatio-temporal coordinates for each sample.

Columns from 9: Abundance data for the selected taxa.

Phytoplankton Colour Index (PCI) corresponds to the greenness of the silk and is associated to a semi-logarithmic scale as follow:

no colour = 0

very pale green = 1

pale green = 2

green = 6.5