**Data extraction Zooplankton project**

**Extraction performed by Pierre Hélaouët**

**17/11/2025.**

**DOI:** 10.17031/691af87e77d0f

**Website:** https://doi.mba.ac.uk/data/3580/

**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 6 files:**

1. “CPR\_Data\_Zooplankton\_17112025.docx”: This document
2. “CPR\_Zooplankton\_ControlMap\_17112025.png”:

Map representing the selected samples from January 1958 to December 2022 (258015 samples)

1. “CPR\_Zooplankton\_Data\_LargeZooplankton\_17112025.csv”: Abundance data for all selected large zooplankton (13 taxa, see CPR\_Zooplankton\_List\_LargeZooplankton\_17112025.csv) and all selected samples in the selected area (30°N to 80°N, -80°E to 26°E).

Rows: All samples for the selected area (258015 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 21: Abundance data for all selected large zooplankton (13 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\_Zooplankton\_List\_LargeZooplankton\_17112025.csv”: List of large zooplankton

Rows: All selected taxa (13 taxa).

Column 1 “Accepted\_ID”: Unique identifier used by the CPR survey

Column 2 "Taxon\_Name”: Unique name used by the CPR survey.

Column 3 "WoRMS\_Name”: Name used by WoRMS corresponding to the “Aphia\_ID”.

Column 4 "Aphia\_ID”: Identifier used by WoRMS.

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\_Zooplankton\_Data\_SmallZooplankton\_17112025.csv”: Abundance data for all selected small zooplankton (see CPR\_Zooplankton\_List\_SmallZooplankton\_17112025.csv) and all selected samples in the selected area. (14 taxa, 258015 samples).

Note: Same architecture as “CPR\_Zooplankton\_Data\_LargeZooplankton\_17112025.csv”

1. “CPR\_Zooplankton\_List\_SmallZooplankton\_17112025.csv”: List of small zooplankton (14taxa).

Note: Same architecture as “CPR\_Zooplankton\_List\_LargeZooplankton\_17112025.csv”.