**Data extraction “Basket analysis” project**

**Extraction performed by Pierre Hélaouët**

**14/05/2025.**

**DOI: 10.17031/68244c49126fe**

**Web page: https://doi.mba.ac.uk/data/3461**

**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 9 files:**

1. “CPR\_Data\_BasketAnalysis\_14052025.docx”: This document
2. “CPR\_BasketAnalysis\_ControlMap\_14052025.png”:

Map representing the selected samples from January 1961 to December 2010 (202723 samples)

1. “CPR\_BasketAnalysis\_Data\_LargeZooplankton\_14052025.csv”: Abundance data for all selected large zooplankton (192 taxa, see CPR\_BasketAnalysis\_List\_LargeZooplankton\_14052025.csv) and all samples available.

Rows: All samples for the selected area (202723 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 6: Spatio-temporal coordinates for each sample. The column “midpoint” corresponds to the spatio coordinates of the middle of each CRP sample (as DD/MM/YYYY hh:mm).

Columns from 7 to 198: Abundance data for all selected large zooplankton (192 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 3: 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. DRI is defined below, see Column 5 in “CPR\_BasketAnalysis\_List\_LargeZooplankton\_14052025.csv”

1. “CPR\_BasketAnalysis\_List\_LargeZooplankton\_14052025.csv”: List of large zooplankton

Rows: All selected taxa (192 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\_BasketAnalysis\_Data\_SmallZooplankton\_14052025.csv”: Abundance data for all selected small zooplankton (see CPR\_BasketAnalysis\_List\_SmallZooplankton\_14052025.csv) and all selected samples in the selected area. (82 taxa, 202723 samples).

Note: Same architecture as “CPR\_BasketAnalysis\_Data\_LargeZooplankton\_14052025.csv”

1. “CPR\_BasketAnalysis\_List\_SmallZooplankton\_14052025.csv”: List of small zooplankton (82 taxa).

Note: Same architecture as “CPR\_BasketAnalysis\_List\_LargeZooplankton\_14052025.csv”.

1. “CPR\_BasketAnalysis\_Data\_Phytoplankton\_14052025.csv”: Abundance data for all selected phytoplankton (see CPR\_BasketAnalysis\_List\_Phytoplankton\_14052025.csv) and all selected samples in the selected area. (194 taxa, 202723 samples).

Note: Same architecture as “CPR\_BasketAnalysis\_Data\_LargeZooplankton\_14052025.csv”

1. “CPR\_BasketAnalysis\_List\_Phytoplankton\_14052025.csv”: List phytoplankton (194 taxa).

Note: Same architecture as “CPR\_BasketAnalysis\_List\_Phytoplankton\_14052025.csv”.

1. “CPR\_BasketAnalysis\_Data\_PCI\_14052025”. Values for the Phytoplankton Colour Index (PCI)

Note: Same architecture as “CPR\_BasketAnalysis\_Data\_LargeZooplankton\_14052025.csv”