**Data extraction “Basking shark” project**

**Extraction performed by Pierre Hélaouët**

**03/02/2025.**

**DOI: 10.17031/67a08ad31812b**

**Web page: https://doi.mba.ac.uk/data/3427/**

**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\_BaskingShark\_03022025.docx”: This document
2. “CPR\_BaskingShark\_ControlMap\_03022025.png”:

Map representing the selected samples from January 1990 to December 2021 (23171 samples)

1. “CPR\_BaskingShark\_Data\_LargeZooplankton\_03022025.csv”: Abundance data for all selected large zooplankton (79 taxa, see CPR\_BaskingShark\_List\_LargeZooplankton\_03022025.csv) and all selected samples in the selected area (48°N to 56°N, -16°E to -6°E).

Rows: All samples for the selected area (23171 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 87: Abundance data for all selected large zooplankton (79 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\_BaskingShark\_List\_LargeZooplankton\_03022025.csv”: List of large zooplankton

Rows: All selected taxa (79 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\_BaskingShark\_Data\_SmallZooplankton\_03022025.csv”: Abundance data for all selected small zooplankton (see CPR\_BaskingShark\_List\_SmallZooplankton\_03022025.csv) and all selected samples in the selected area. (57 taxa, 23171 samples).

Note: Same architecture as “CPR\_BaskingShark\_Data\_LargeZooplankton\_03022025.csv”

1. “CPR\_BaskingShark\_List\_SmallZooplankton\_03022025.csv”: List of small zooplankton (57 taxa).

Note: Same architecture as “CPR\_BaskingShark\_List\_LargeZooplankton\_03022025.csv”.