**Data extraction for Benthic Taxa in the Northeast Atlantic Ocean**

**Extraction performed by Pierre Hélaouët**

**08/08/2025.**

**DOI: 10.17031/6895c8d38edbf**

**Webpage: https://doi.mba.ac.uk/data/3537/**

**Note 1:** As always, this dataset has been carefully built and checked accordingly. However, it is the user’s responsibility to perform his own verifications.

**Note 2:** This dataset is associated with a Digital Object Identifier or DOI. Please, use the DOI for all citations. Please do not share this dataset outside the current project.

**Quick description of the dataset**

**1 – The dataset contains 4 files:**

1. “CPR\_Data\_BenthicTaxa\_08082025.docx”: This document
2. “CPR\_BenthicTaxa\_ControlMap\_08082025.png”:

Map representing the selected samples from January 1958 to December 2021

* 48447 samples in the area from 37°N to 70°N, and -11°E to 16°E.

1. “CPR\_BenthicTaxa\_Data\_08082025.csv”: Abundance data for taxa/groups (8) in all selected samples.

Rows: All samples for the selected area (48447 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.
* Columns from 7. Abundance for “Cyphonautes” noted “Cyphonautes”. CPR Id 35.
* Columns from 8. Abundance for “Echinoderm\_larvae” noted “Echinoderm”. CPR Id = 36.
* Columns from 9. Abundance for “Decapoda larvae (Total)” noted “Decapoda”. CPR Id = 83.
* Columns from 10. Abundance for “Cirripede larvae (Total)” noted “Cirripede\_Larv”. CPR Id = 350.
* Columns from 11. Abundance for “Echinospira larvae” noted “Echinospira”. CPR Id = 1543.
* Columns from 12. Abundance for “Cirripede nauplii” noted “Cirripede\_Naupli

”. CPR Id = 10535.

* Columns from 13. Abundance for “Bivalvia larvae” noted “Bivalvia”. CPR Id = 10590.
* Columns from 14. Abundance for “Gastropoda (unidentified)” noted “Gastropoda”. CPR Id = 10674.

Abundance per samples is expressed in number of individuals per 3m3.

Note 1: 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.

Note 2: 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.

1. “CPR\_BenthicTaxa\_ListTaxa\_08082025.csv”: List of all individual taxa (except large and small copepods\_

Rows: All selected 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).

**Note 3:** Samples collected in 2023 are still currently being analysed. 2023 sampling effort might therefore be substantially different from previous years.