**Data extraction for South North Sea**

**Extraction performed by Pierre Hélaouët**

**10/07/2025.**

**DOI: 10.17031/686f7bf9f3019**

**Webpage: https://doi.mba.ac.uk/data/3496/**

**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 6 files:**

1. “CPR\_Data\_SouthNorthSea\_10072025.docx”: This document
2. “CPR\_SouthNorthSea\_ControlMap\_10072025.png”:

Map representing the selected samples from January 1958 to December 2021

* 14694 samples in the area from 51°N to 57°N, and 0°E to 8°E.

1. “CPR\_NorthSea\_Data\_10072025.csv”: Abundance data for taxa/groups (9) in all selected samples.

Rows: All samples for the selected area (14694 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 “Echinoderm post larvae” noted “EchiPostLarv”. CPR Id = 460.
* Columns from 8. Abundance for “Bivalvia larvae” noted “BivaLarvae”. CPR Id =10590
* Columns from 9. Abundance for “Chaetognatha\_eyecount” noted “L\_Chaetognatha”. CPR Id = 89.
* Columns from 10. Abundance for “Chaetognatha\_Traverse” noted “S\_Chaetognatha”. CPR Id = 34.
* Columns from 11. Abundance for “Large\_Copepods” noted “L\_Cop”. Composed by the sum of 24 taxa. See associated list “CPR\_SouthNorthSea\_List\_LargeCopepods\_10072025.csv”.
* Columns from 12. Abundance for “Small\_Copepods” noted “S\_Cop”. Composed by the sum of 31 taxa. See associated list “CPR\_SouthNorthSea\_List\_SmallCopepods\_10072025.csv”.
* Columns from 13. Abundance for “Decapoda\_larvae” noted “DecapodaLarvae”. CPR Id = 83.
* Columns from 14. Abundance for “Para-Pseudocalanus\_spp” noted “ParaPseudo”. CPR Id = 3.
* Columns from 15. Abundance for “Rotifer\_eggs” noted “Rotifer”. CPR Id = 946.

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\_SouthNorthSea\_ListIndTaxa\_10072025.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).

1. “CPR\_SouthNorthSea\_List\_LargeCopepods\_10072025.csv”: List of all individual taxa in Large Copepods (>= 2mm, 24 taxa).

Note: Same architecture as “CPR\_SouthNorthSea\_ListIndTaxa\_10072025.csv”.

1. “CPR\_SouthNorthSea\_List\_SmallCopepods\_10072025.csv”: List of all individual taxa in Small Copepods (>= 2mm, 31 taxa).

Note: Same architecture as “CPR\_SouthNorthSea\_ListIndTaxa\_10072025.csv”.