**Data extraction North Sea**

**Extraction performed by Pierre Hélaouët**

**29/01/2026.**

**DOI: 10.17031/697b3e2b3be3e**

**Webpage: https://doi.mba.ac.uk/data/3630/**

**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 11 files:**

1. “CPR\_Data\_NorthSea\_29012026.docx”: This document
2. “CPR\_NorthSea\_ControlMap\_29012026.png”:

Map representing the selected samples from January 1958 to December 2022

* 58048 samples in the area from 52°N to 62°N, and -3.5°E to 7.5°E.

1. “CPR\_NorthSea\_Data\_Diatoms\_29012026.csv”: Abundance data for all diatoms (75 taxa) in all selected samples.

Rows: All samples for the selected area (58048 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 to 81: Abundance data for all diatoms (75 taxa). Abundance per samples is expressed in number of individuals per 3m3.

Note 1: Not all taxa are present in each selected areas. The CPR survey records “true” zeroes, therefore a “0” can be interpreted as “we looked for it and did not find any”. A missing value “Not a Number or NaN” can be interpreted as “we did not look for it”.

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.

Note 3: 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\_NorthSea\_List\_Diatoms\_29012026.csv”: List of all diatoms (75 taxa)

Rows: All selected 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\_NorthSea\_Data\_Dinoflagellates\_29012026.csv”: Abundance data for all dinoflagellates (49 taxa).

Note: Same architecture as “CPR\_NorthSea\_Data\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_List\_Dinoflagellates\_29012026.csv”: List of small dinoflagellates (49 taxa).

Note: Same architecture as “CPR\_NorthSea\_List\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_Data\_LargeCopepods\_29012026.csv”: Abundance data for all large copepods (>= 2mm, 42 taxa).

Note: Same architecture as “CPR\_NorthSea\_Data\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_List\_LargeCopepods\_29012026.csv”: List of large copepods (>=2mm, 42 taxa).

Note: Same architecture as “CPR\_NorthSea\_List\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_Data\_SmallCopepods\_29012026.csv”: Abundance data for all small copepods (<2mm, 37 taxa).

Note: Same architecture as “CPR\_NorthSea\_Data\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_List\_SmallCopepods\_29012026.csv”: List of small copepods (<2mm, 37 taxa).

Note: Same architecture as “CPR\_NorthSea\_List\_Diatoms\_29012026.csv”.

1. “CPR\_NorthSea\_Data\_PCI\_29012026”: Data for Phytoplankton Colour Index (PCI).

Note: Phytoplankton Colour Index (PCI) corresponds to the greenness of the silk and is associated to a semi-logarithmic scale as follow:

no colour = 0

very pale green = 1

pale green = 2

green = 6.5