**Data extraction “FONI” project, University of Dalhousie**

**Extraction performed by Pierre Hélaouët**

**03/02/2025.**

**DOI: 10.17031/67a0cc543b66b**

**Web page: https://doi.mba.ac.uk/data/3430/**

**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 8 files:**

1. “CPR\_Data\_FONIDalhousie\_03022025.docx”: This document
2. “CPR\_FONIDalhousie\_ControlMap\_03022025.png”:

Map representing the selected samples from January 1958 to December 2021 (58861 samples)

1. “CPR\_FONIDalhousie\_Data\_LargeZooplankton\_03022025.csv”: Abundance data for all selected large zooplankton (70 taxa, see CPR\_FONIDalhousie\_List\_LargeZooplankton\_03022025.csv) and all selected samples in the selected area (51°N to 61°N, -2.5°E to 10°E).

Rows: All samples for the selected area (58861 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 78: Abundance data for all selected large zooplankton (70 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\_FONIDalhousie\_List\_LargeZooplankton\_03022025.csv”: List of large zooplankton

Rows: All selected taxa (70 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\_FONIDalhousie\_Data\_SmallZooplankton\_03022025.csv”: Abundance data for all selected small zooplankton (see CPR\_FONIDalhousie\_List\_SmallZooplankton\_03022025.csv) and all selected samples in the selected area. (65 taxa, 58861 samples).

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

1. “CPR\_FONIDalhousie\_List\_SmallZooplankton\_03022025.csv”: List of small zooplankton (65 taxa).

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

1. “CPR\_FONIDalhousie\_Data\_Phytoplankton.csv”: Abundance data for all selected phytoplankton (see CPR\_FONIDalhousie\_List\_Phytoplankton\_03022025.csv) and all selected samples in the selected area. (139 taxa).

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

1. “CPR\_FONIDalhousie\_List\_Phytoplankton\_03022025.csv”: List of phytoplankton (139 taxa, 58861 samples).

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