**Data extraction Nicolas Djeghri**

**Extraction performed by Pierre Hélaouët**

**17/09/2024.**

**DOI: 10.17031/66f12be296d70**

**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 14 files:**

1. “CPR\_Data\_NicolasDjeghri\_17092024.docx”: This document
2. “CPR\_NicolasDjeghri\_ControlMap\_17092024.png”:

Map representing the selected samples from January 1958 to December 2021 (252890 samples)

1. “CPR\_NicolasDjeghri\_Data\_LargeZooplankton\_17092024.csv”: Abundance data for all selected large zooplankton (149 taxa, see CPR\_NicolasDjeghri\_List\_LargeZooplankton\_17092024.csv) and all selected samples in the selected area (30°N to 80°N, -80°E to 20°E).

Rows: All samples for the selected area (252890 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 157: Abundance data for all selected large zooplankton (149 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\_NicolasDjeghri\_List\_LargeZooplankton\_17092024.csv”: List of large zooplankton

Rows: All selected taxa (149 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\_NicolasDjeghri\_Taxonomy\_LargeZooplankton\_17092024.csv”: Taxonomy of all large zooplankton taxa selected.

Rows: All selected taxa (149 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 “counting\_method”: 1 for large zooplankton (eyecount), 2 for small zooplankton (traverse), 3 for phytoplankton.

Column 6 “is\_copepod”: flag taxa categorised as copepods

Column 7 “is\_diatoms”: flag taxa categorised as diatoms

Column 8 “is\_dinos”: flag taxa categorised as dinoflagellates

Column 9 “is\_total”: flag taxa considered as a group on different entities

Column 10 to 35: taxonomy.

1. “CPR\_NicolasDjeghri\_Data\_SmallZooplankton\_17092024.csv”: Abundance data for all selected small zooplankton (see CPR\_NicolasDjeghri\_List\_SmallZooplankton\_17092024.csv) and all selected samples in the selected area. (78 taxa, 252890 samples).

Note: Same architecture as “CPR\_NicolasDjeghri\_Data\_LargeZooplankton\_17092024.csv”

1. “CPR\_NicolasDjeghri\_List\_SmallZooplankton\_17092024.csv”: List of small zooplankton (78 taxa).

Note: Same architecture as “CPR\_NicolasDjeghri\_List\_LargeZooplankton\_17092024.csv”.

1. “CPR\_NicolasDjeghri\_Taxonomy\_SmallZooplankton\_17092024.csv”: Taxonomy of all small zooplankton taxa selected (78 taxa).

Note: Same architecture as “CPR\_NicolasDjeghri\_Taxonomy\_LargeZooplankton\_17092024.csv”

1. “CPR\_NicolasDjeghri\_Data\_Phytoplankton.csv”: Abundance data for all selected phytoplankton (see CPR\_NicolasDjeghri\_List\_Phytoplankton\_17092024.csv) and all selected samples in the selected area. (173 taxa).

Note: Same architecture as “CPR\_NicolasDjeghri\_Data\_LargeZooplankton\_17092024.csv”

1. “CPR\_NicolasDjeghri\_List\_Phytoplankton\_17092024.csv”: List of phytoplankton (186 taxa, 252890 samples).

Note: Same architecture as “CPR\_NicolasDjeghri\_List\_LargeZooplankton\_17092024.csv”.

1. “CPR\_NicolasDjeghri\_Taxonomy\_Phytoplankton\_17092024.csv”: Taxonomy of all phytoplankton taxa selected (173 taxa).

Note: Same architecture as “CPR\_NicolasDjeghri\_Taxonomy\_LargeZooplankton\_17092024.csv”

1. “CPR\_NicolasDjeghri\_Data\_PCI\_17092024”. Values for the Phytoplankton Colour Index (PCI)

Note: Same architecture as “CPR\_NicolasDjeghri\_Data\_LargeZooplankton\_17092024.csv”

1. “CPR\_NicolasDjeghri\_Traits\_Phyto\_17092024”. Information regarding functional traits and associated measurements for phytoplankton. Warning: this file is a prototype and does not have the same architecture of all the other files (e.g., eyecount and traverse are merged, rows are not in the same order, etc.)

1. “CPR\_NicolasDjeghri\_Traits\_Zoo\_17092024”. Information regarding functional traits and associated measurements for zooplankton. Warning: this file is a prototype and does not have the same architecture of all the other files (e.g., eyecount and traverse are merged, rows are not in the same order, etc.)