**Data extraction Calanus Irish Sea**

**Extraction performed by Pierre Hélaouët**

**23/06/2025.**

**DOI: 10.17031/68590fc397826**

**Webpage: https://doi.mba.ac.uk/data/3473/**

**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 identified project “*influences on fish biology (length/maturity-at-age, growth etc.) in Irish waters”*.

**Quick description of the dataset**

**1 – The dataset contains 5 files:**

1. “CPR\_Data\_CalanusIrishSea\_23062025.docx”: This document
2. “CPR\_CalanusIrishSea\_ControlMap\_23062025.png”:

Map representing the selected samples from January 1990 to December 2021

* 4873 samples in the area C03 (Irish Sea)

1. “CPR\_CalanusIrishSea\_Data\_23062025.csv”: Abundance data for all selected taxa or groups of taxa. Rows: All samples for the selected area (4873 samples).

Column 1: Unique sample id. For instance: “227B--17” corresponds to the 17th sample for the 227th transect on the B route.

Columns from 2 to 6: Spatio-temporal coordinates for each sample.

Columns from 7 to 10: Abundance data for all selected taxa/groups.

Note 1: The agreed taxa/groups were: (see emails sent to Pierre Helaouet for a previous extract on Thu 15/08/2024 11:17).

1. *Calanus finmarchicus*: Corresponds to CPR ID 40 with Aphia ID 104464.
2. *Calanus helgolandicus*: Corresponds to CPR ID 41 with Aphia ID 104466.
3. Large Copepods: Sum of all large copepods (>2mm) present in list “CPR\_CalanusIrishSea\_List\_LargeCopepods\_23062025.csv”
4. Small Copepods: Sum of all small copepods (<=2mm) present in list “CPR\_CalanusIrishSea\_List\_SmallCopepods\_23062025.csv”
5. Euphausiacea\_total: Corresponds to CPR ID 88 with Aphia ID 1128.
6. Decapoda\_larvae: Corresponds to CPR ID 83 with Aphia ID 1130.

Note 3: 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 4: 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 5: 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\_CalanusIrishSea\_List\_LargeCopepods\_23062025.csv”: List of large copepods (>2mm, 179 taxa)

Rows: All selected taxa (179 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\_CalanusIrishSea\_List\_SmallCopepods\_23062025.csv”: List of small copepods (<+2mm, 59 taxa).

Note: Same architecture as “CPR\_CalanusIrishSea\_List\_LargeCopepods\_23062025.csv”.