**Data extraction “PelagicFish” project, CEFAS**

**Extraction performed by Pierre Hélaouët**

**17/02/2025.**

**DOI: 10.17031/67b2f825d6e51**

**Web page: https://doi.mba.ac.uk/data/3433**

**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 3 files:**

1. “CPR\_Data\_PelagicFish\_17022025.docx”: This document
2. “CPR\_PelagicFish\_ControlMap\_17022025.png”:

Map representing the selected samples from January 1958 to December 2021 (119788 samples)

1. “CPR\_PelagicFish\_Data\_17022025.csv”: Abundance data for all selected taxa and all selected samples in the selected area (48°N to 62°N, -15.5°E to 10°E).

Selected taxa:

“Fish eggs”, CPR ID = 90, First recorded (DRI) in January 1958.

“Fish larvae”, CPR ID = 91, First recorded (DRI) in January 1958.

“Anchovy larvae”, CPR ID 10753, First recorded (DRI) in January 2020.

Rows: All samples for the selected area (119788 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 11: Abundance data for all selected 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.