**TITLE: AtlantECO [WP2] – Traditional microscopy dataset –** **Foraminifera abundance and biomass concentration data**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1.- INTRODUCTION**

This dataset contains **1 026 933** georeferenced abundance and biomass concentration records of **48** accepted scientific names of extant planktic Foraminifera of various taxonomic levels. This dataset is a compilation of the following five main global and regional datasets that reported abundances of planktic marine foraminifera:

* The Coastal & Oceanic Plankton Ecology, Production & Observation Database (NMFS-COPEPOD, O’Brien, 2014) from the National Oceanic and Atmospheric Administration - <https://www.st.nmfs.noaa.gov/copepod/atlas/html/taxatlas_2040000.html>
* The North Atlantic and North Pacific Continuous Plankton recorder (NANP-CPR) survey (Johns D & Broughton D, 2019) - <https://doi.org/10.17031/1629>
* The Southern Ocean CPR (SO-CPR) survey (Hosie, 2021) - doi:10.26179/ksds-s610
* The Australian CPR (AusCPR) survey (AusCPR) - <https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/c1344e70-480e-0993-e044-00144f7bc0f4>
* Various published cruise-level foraminifera counts from R. Schiebel and collaborators (Schiebel et al., 1995; Schiebel & Hemleben, 2000; Schiebel et al., 2001; Schiebel, 2002; Schiebel et al., 2002, 2004; Jentzen et al., 2018)

**2.- METHODOLOGY USED**

We compiled in situ planktic foraminifera abundance observations from large scale monitoring programs and existing data compilation programs to calculate total global foraminifer biomass and contribution to carbonate fluxes. The main data sources included were the Southern Ocean Continuous Plankton Recorder (CPR) (SO-CPR; Hosie, 2021), the Australian CPR (AusCPR) (IMOS, 2022), the North Atlantic and North Pacific CPR (NANP-CPR) (Johns D & Broughton D, 2019) and the Coastal and Oceanic Plankton Ecology, Production and Observation Database (COPEPOD) (O’Brien, 2014). We also gathered data from various individual oceanographic cruises (Schiebel et al., 1995; Schiebel & Hemleben, 2000; Schiebel et al., 2001; Schiebel, 2002; Schiebel et al., 2002, 2004; Jentzen et al., 2018). We matched all taxonomic information against the list of accepted taxon names of the World Register of Marine Species (WoRMS) to harmonize all classifications across datasets and correct for potential deprecated scientific species names (Horton et al., 2017). Observations lacking complete sampling metadata (date, depth, location, abundance value) were removed, together with observations of body parts and plankton observations that did not correspond to the taxa of interest (removal of 522 points). Biomass conversions were conducted using volume to biomass equations based on morphological shape groups. All the details of the abundance to biomass conversions are given in Knecht et al., (in prep., **manuscript available upon demand**).

The total foraminifer dataset consists of 1 026 933 points at 305 428 unique locations, with a mean (± sd) sampling depth of 106.69 ± 341.99m and collected between 1939 - 2021 (mean ± sd = 2 000.34 ± 13.27). Foraminifer abundances range between 0 and 152 170 ind.m-3, with a mean (± sd) abundance of 3.61 ± 162.63 ind.m-3. As for the pteropod dataset, there is a high prevalence of CPR data (73.94% of the total data) with 89.72% zero abundance observations, which causes a low median abundance value of 0.00 ind.m-3. 60.01% of the data are species resolved, followed by 32.89% of the observations on a phylum level. The dataset contains observations on 43 of the around 50 extant foraminifer species (A. Loeblich & Tappan, 1992), with most of the total abundance made up of *Globigerina bulloides* (25.6% of the total species-resolved abundance), *Neogloboquadrina incompta* (23.7%), *Turborotalita quinqueloba* (13.3%) and *Globigerinita glutinata* (11.3%).

The main R packages used to implement this dataset were: ‘tidyverse’ (Wickham et al., 2019), ‘reshape2’ version 1.4.4 (Wickham, 2007), ‘marmap’ version 1.0.6 (Pante & Simon-Bouhet, 2013), ‘lubridate’ version 1.8.0 (Grolemund, 2011), ‘raster’ version 3.5-15 (Hijmans, 2022) and ‘worms’ version 0.2.2 (Holstein, 2018).

**3.- DATASET DESCRIPTION**

**Data type:** Abundances converted to biomass concentrations.

**Latitude/Longitude format:** WGS 84 (-180°E/+180°E).

**Geographic area covered by the dataset:** Global Ocean.

**Depth range covered by the dataset:** From 0m to 4911m.

**Time period covered by the dataset:** From 14-10-1939 to 31-01-2021.

**Dataset format:** .csv file withsemicolon-delimited columns.

**Date of dataset creation:** 26/11/2022.

**Raw dataset repository:** AtlantECO’s GeoNode (<https://atlanteco-geonode.eu/>).

**4.- MAIN VARIABLE DESCRIPTION**

MeasurementTypeID: Has not been defined within AtlantECO

MeasurementValue: Organisms concentration (i.e., abundance) in ind.m-3

MeasurementID: Has not been defined within AtlantECO

occurrenceID: Combination of decimalLatitude, decimalLongitude, Day, Month, Year, Depth, MinDepth, MaxDepth, ScientificName, MeasurementValue..

**5.- DATA OVERVIEW**

**

**6.- CONTRIBUTORS**

- Nielja Knecht (nielja.knecht@gess.ethz.ch), ETH Zürich, Switzerland.

- Fabio Benedetti (fabio.benedetti@usys.ethz.ch), ETH Zürich, Switzerland.

- David Johns (djoh@MBA.ac.uk), MBA, United Kingdom.

- Sonia Chaabane (sonia.chaabane@fondationbiodiversite.fr), FRB-CESAB, France.

- Ralf Schiebel (ralf.schiebel@mpic.de), MPI, Germany.

- Meike Vogt (meike.vogt@usys.ethz.ch), ETH Zürich, Switzerland.