scientific data

Article in Press

Check for updates

OPEN

DATA DESCRIPTOR

A 40-year zooplankton dataset from the Mediterranean Sea supporting biodiversity and Digital Twin of the Ocean applications

Iole Di Capua[1]✉, Andrea Marcucci[1], Lorenzo Petrella[1], Maria Grazia Mazzocchi[2], Lynn Delgat[3], Cyril Radermecke[3], Ilaria Rosati[4] & Lorenzo Longobardi[5]

This paper presents the zooplankton dataset "40 Years of Zooplankton Biodiversity Assessment (ZOOGoN-40Y)", a long-term dataset of zooplankton abundance and taxonomic composition covering 1984–2024 from the MareChiara Long-Term Ecological Research site in the Gulf of Naples, western Mediterranean Sea. The dataset integrates and harmonizes historical records (1984–2015) with newly collected data (2016–2024), thereby incorporating 40 years of observation into a continuous, ongoing time series. Standardized quality assurance and quality control procedures were applied to ensure data consistency and temporal continuity, supporting sustained marine biodiversity data flows to the Digital Twin of the Ocean. The dataset contains 98,175 occurrence records from 1,295 sampling events and includes 259 taxa, of which 190 are identified at species level. It covers 181 copepod taxa, with 149 at species-level identified by juveniles and different sexes, as well as 83 additional holo- and meroplanktonic taxa recorded at varying taxonomic resolutions, including new records and non-indigenous species. According to FAIR (Findable, Accessible, Interoperable, Reusable) principles, the dataset is structured using the Darwin Core standard, the associated metadata are described using the Ecological Metadata Language (EML 2.2.0) and is published on the Global Biodiversity Information Facility (GBIF). The dataset provides a standardized and open-access resource for biodiversity, ecosystem monitoring, and modelling, as well as scenario-based applications for the Digital Twin of the Ocean in the Mediterranean Sea

Background & Summary

The oceans encompass nearly 90% of Earth's habitable space and host an estimated 2.2 million eukaryotic species out of the 8.7 million thought to exist globally, with approximately 91% still awaiting formal description. Within this vast diversity, zooplankton may represent ~38% of marine species, although these organisms are likely underestimated due to methodological limitations and taxonomic gaps[1].

The Mediterranean Sea, the world's largest and deepest enclosed basin, is recognized as a major biodiversity hotspot[2]. Its complex hydrological and climatic features support the coexistence of temperate and subtropical species, a high proportion of endemic species, and an increasing number of non-indigenous species (NIS) driven by anthropogenic pressures[3], and climate change[4]. Within this system, the metazoan plankton community plays a fundamental role in ecosystem functioning. Zooplankton act as a link between primary producers and higher trophic levels, sustaining pelagic food webs and contributing to biogeochemical cycles, including carbon transfer and export[5]. Changes in their abundance, distribution, and phenology can reveal ecological shifts and anthropogenic impacts across spatial and temporal scales[6–8]. Moreover, their short generation times and high sensitivity to environmental conditions make them effective early warning indicators of climate-driven

[1]Stazione Zoologica Anton Dohrn, Research Infrastructures for Marine Biological Resources Department, Marine Organism Taxonomy Core Facility, Naples, Italy. [2]Stazione Zoologica Anton Dohrn, Department of Integrative Marine Ecology, Naples, Italy. [3]VLIZ | Flanders Marine Institute, Ostend, Belgium. [4]National Research Council (CNR), Research Institute on Terrestrial Ecosystems (IRET), Lecce, Italy. [5]WorldFish, Penang, Malaysia. ✉e-mail: iole.dicapua@szn.it

change, often responding before higher trophic levels[8–10]. Despite their ecological importance, assessing zooplankton diversity and abundance remains a major challenge. The pelagic realm encompasses a wide range of marine phyla, many characterized by small body size, complex morphology, and multiple developmental stages. These features complicate species identification and require highly specialized taxonomic expertise. In this context, morphology-based α-taxonomy remains essential, but it is increasingly complemented by molecular approaches[11].

Long-term ecological time series are crucial for understanding ecosystem dynamics and detecting responses to environmental and climate change. The MareChiara Research Program, established in 1984 by the Stazione Zoologica Anton Dohrn (SZN), represents one of the few and longest planktonic research studies in the Mediterranean Sea. Conducted at a fixed station in the Gulf of Naples, this program has generated a unique dataset on zooplankton abundance and taxonomic composition.

The zooplankton dataset, covering the period 1984–2024, was produced within the DTO-BioFlow Financial Support to Third Parties (FSTP) program under the "Call for marine biodiversity (monitoring) data." DTO-BioFlow ("Integration of Biodiversity Monitoring Data into the Digital Twin Ocean"), a Horizon Europe project, supports the development of the biological component of the European Digital Twin of the Ocean by enhancing the accessibility, standardization, interoperability, and long-term flow of marine biodiversity data.

The project addresses a key challenge in marine biodiversity science: although extensive datasets already exist, many remain underutilized because they are stored in heterogeneous, project-specific formats and are not readily connected to international biodiversity infrastructures. DTO-BioFlow promotes the recovery and integration of these "sleeping" biodiversity data into European systems such as EMODnet and the EU Digital Twin Ocean, enabling their use in ecological research, environmental monitoring, marine management, and evidence-based policy.

Within this framework, the conversion of datasets and associated metadata into standardized formats such as Darwin Core Archive represents a crucial step toward ensuring interoperability, quality control, data exchange, data integration, and long-term reusability across European and global marine biodiversity data networks. Building on this approach, the ZOOGoN-40Y dataset integrates, curates, and harmonizes four decades of zooplankton observations characterized by high taxonomic resolution. The dataset combines historical records (1984–2015) with newly collected data (2016–2024), ensuring temporal continuity and completeness through the application of rigorous and consistent QA/QC procedures. Structured according to the Darwin Core Archive standard and aligned to FAIR principles, the dataset significantly improves data accessibility, interoperability, and reuse. It includes 98,175 occurrence records from 1,295 sampling events and covers 259 taxa, including 190 identified species, of which 149 are planktonic copepods.

Beyond its value as a harmonized biodiversity resource, ZOOGoN-40Y provides an important basis for investigating long-term ecological changes in Mediterranean marine ecosystems, including variations in community composition, phenology, and the occurrence of newly recorded species, including non-indigenous species. These datasets are particularly relevant in the Mediterranean Sea, a region increasingly affected by climate change and multiple anthropogenic pressures driving rapid biodiversity transformations.

Overall, this dataset provides a high-quality, long-term resource for zooplankton biodiversity assessment at a high taxonomic resolution. Taxonomic information follows the World Register of Marine Species (WoRMS)[12] and is aligned with international frameworks such EMODnet Biology. By offering a long temporal perspective on zooplankton patterns in a climate-sensitive region, the dataset advances marine ecological research and supports evidence-based biomonitoring and management of coastal ecosystems. Furthermore, it represents a valuable resource for emerging digital infrastructures, including the Digital Twin of the Ocean, enabling the development, calibration, and validation of ecosystem models and data-driven simulations aimed at predicting future ocean scenarios.

Methods

Sampling. Sampling was carried out at the Long-Term Ecological Research site MareChiara (LTER-MC) https://deims.org/0b87459a-da3c-45af-a3e1-cb1508519411 with coordinates Latitude: 40.808 N Longitude: 14.25 E, located approximately 2 nautical miles off the coast of Naples (Italy), western Mediterranean Sea[13].

At LTER-MC site the zooplankton was collected from 26/01/1984 (eventID: mc1) to 02/02/2016 (eventID: mc1188) with an INDIAN OCEAN standard net, which has a mouth area of 1 m^2 (diameter 113 cm) and from 18/02/2016 (eventID: mc1189) to 12/12/2024 (eventide: mc1506) with a WP2 standard net, which has mouth area of 0.25 m^2 (diameter 57 cm) both fitted with a mesh aperture width of 200 µm (Fig. 1). In 2015, a year of combined sampling between the two net types was carried out to perform an intercalibration for zooplankton sample collection using the two different nets (Di Capua, unpublished data). Preliminary analyses of the intercalibration data did not indicate substantial differences in zooplankton abundance between the two nets, suggesting that the change did not substantially affect the overall patterns observed in the long-term time series; a comprehensive analysis of the intercalibration data, including differences in zooplankton abundance and community composition, will be presented in a dedicated future publication.

The nets were ballasted with a 2–5 kg weight and were towed vertically from –50 m depth to the surface at low speed (0.7–1.0 m s^{-1}). The filtered water volume ($V = AxL$, m^3) is calculated by considering the area of the net mouth (A, m^2) and the length of the released wire (L, m). The final thickness of the sampled layer (ΔD, m) and the depth limits of the layer ($\Delta L = Li\text{-}Lf$, m) are computed considering the wire angle α ($\Delta D = \Delta L \cos \alpha$). Attention is paid in maintaining the wire angle below 10°–15° during the sampling activities. When the wire is not perfectly perpendicular to the sea surface, the wire angle correction must be performed to calculate the correct wire length. The final wire length zf is the intended depth (e.g., 50 m) divided by the cosine of the wire angle α. The zooplankton samples were collected using a 1 L plexiglass filtering cod-end with a 200 µm mesh.

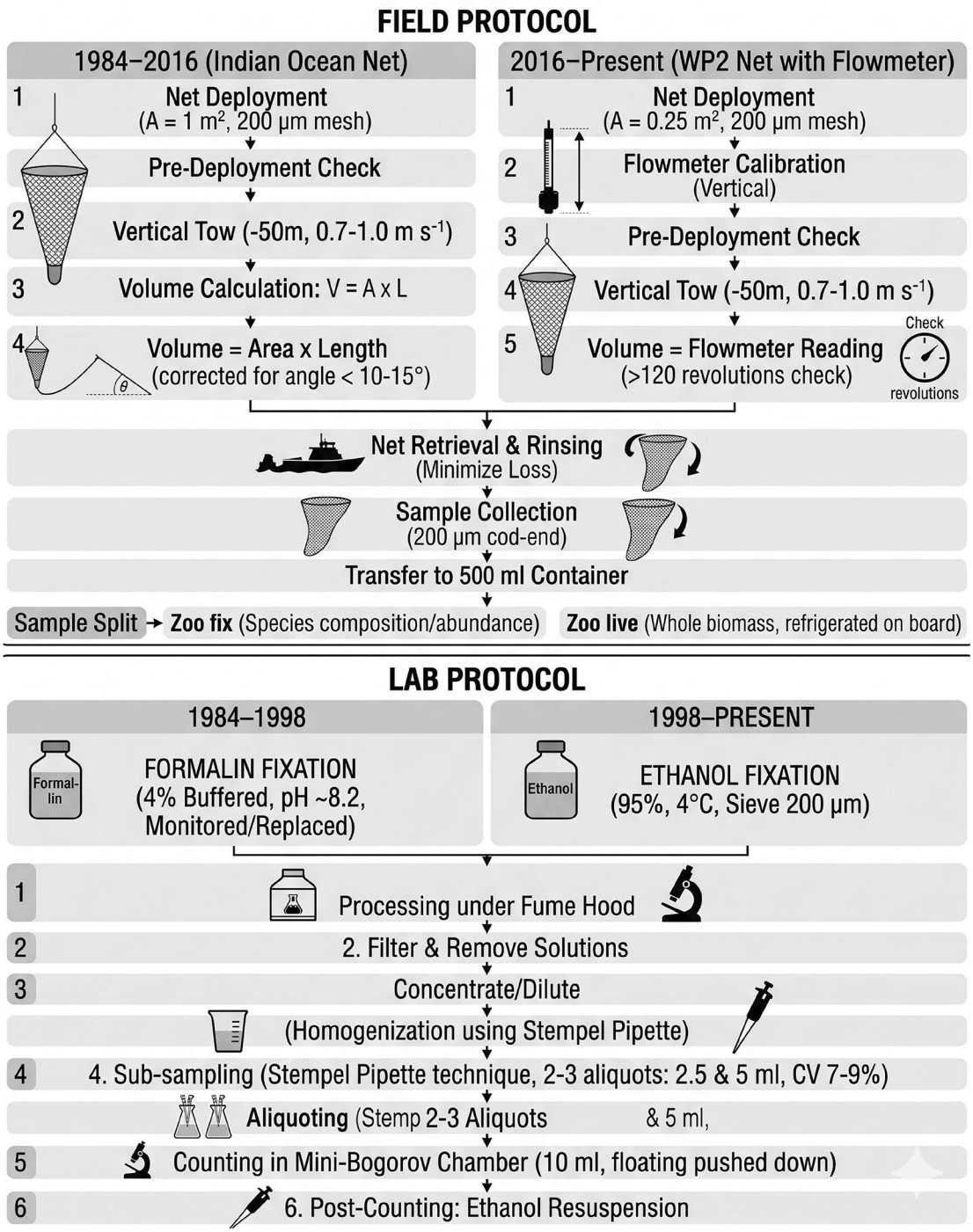


Fig. 1 Workflow for long-term zooplankton sampling and sample processing at the LTER-MC station (1984–2024). The field protocol (top panel) illustrates methodological shifts in sample collection. From 1984 to 2016, an Indian Ocean standard net (mouth area A = 1 m^2) was utilized, necessitating volumetric calculations based on Area (A). Post-2016, a standard WP2 net (A = 0.25 m^2) equipped with a flowmeter was adopted. Both net utilize vertical tows from -50 m depth to the surface at a standardized retrieval speed of 0.7–1.0 m s^{-1}. Shared post-tow procedures involve on-board net retrieval, standardized rinsing, and collection into 500 mL containers. The laboratory protocol (bottom panel) details processing conditional on historical fixation regimes. Formalin fixation (4% buffered) was standardized for samples collected between 1984 and 1998, requiring pH monitoring. From 1999 onward, Ethanol fixation (95% at 4 °C) became the standardized preservative. Sample analysis utilize a six-step sequential workflow: (1) handling saunder a chemical fume hood; (2) filtration; (3) homogenization; (4) precise sub-sampling using replicated Stempel Pipette aliquots (2–3 replicates of 2.5 or 5 mL); (5) counting in a 10 mL Mini-Bogorov Chamber, including manual submergence of floating organisms; and (6) final post-counting Ethanol resuspension. V denotes calculated volume; A denotes net mouth area; L denotes tow length; s^{-1} denotes inverse seconds; CV denotes coefficient of variation. LTER-MC refers to Long-Term Ecological Research – MareChiara site.

From 1984 to 2015, the sampled water volume (V) was estimated based on the net mouth area and the sampled depth layer (from the surface to −50 m) (Fig. 1). Since 2016, the sampled water volume (V) was calculated using a flowmeter (Hydrobios and MF315). The nets have been equipped with flowmeters, which are calibrated prior to each sampling event using vertical tows from 50 m to the surface to determine the calibration constant (Fig. 1). Before each deployment, both net integrity and flowmeter performance are checked. During sampling, the net is lowered vertically from −50 meters to surface, recording the initial flowmeter reading. Once at depth, it is held stationary for a few seconds and then retrieved at a constant speed (0.7–1.0 m s^{-1}). This speed ensures proper flowmeter operation and minimizes zooplankton loss due to pressure. The final flowmeter reading is then recorded; if fewer than 120 revolutions are counted, the sampling is repeated (Fig. 1).

After retrieval, the net, still attached to the cable, is rinsed from top to bottom and rotated to allow all organisms to collect in the cod end, minimizing losses. The zooplankton sample is then transferred into a 500 ml container and stored refrigerated on board.

For each sampling event at LTER-MC, two samples were collected: one for the analysis of species composition and abundance (zoo fix) and the other one for measurement of the whole zooplankton biomass (zoo live).

Sample processing. From 1984 to 1998, zooplankton samples were fixed immediately after collection in a seawater–formalin solution containing about 4% formaldehyde buffered with sodium borate, as recommended for animals collected for taxonomic purposes (Fig. 1). For the preservation of calcareous shelled zooplankton, the pH of the solution must be maintained at about 8.2. Frequent monitoring and buffering of the pH of the buffered formaldehyde with sodium borate and preserving fluid is done during the first days after the initial preservation of a sample and every few weeks thereafter for a period of 3 to 6 months. Since it is officially classified in the EU as a carcinogen, Category 1B, under the CLP Regulation (EC) No 1272/2008, official directives have required formalin to be replaced with 95% ethanol as a preservative.

From 1999 to the present, zooplankton samples were fixed with 95% ethanol (Fig. 1). The content of the 500 ml container is sieved out with the use of a 200 µm sieve. The sieved-out organisms are transferred to a 100 ml container, fixed with 95% ethanol and stored in the fridge at 4 °C.

The laboratory procedure starts with the filtration of the sample under a fume hood (Fig. 1). The sample is filtered and concentrated through a mesh smaller than the mesh size of the sampling net. The solutions (formaldehyde-sea and ethanol 95%) were removed, and the zooplankton sample was concentrated and diluted with tap water or ethanol 95%, in a graduated cup up to a volume of 100 or 200 ml according to the sample richness. The sample is carefully mixed until the organisms are distributed randomly allowing for randomization before collecting the aliquots to be analyzed. Repeated sub-sampling (two-three 2.5 and 5 ml sub-samples) through the Stempel pipette produces a coefficient of variation of 7–9%, lower than splitter technique (Folsom splitter 6-18%) and, furthermore, this treatment was reported to be 5 to 8 times faster than the splitter technique[13] (Fig. 1). Each sub-sample is transferred in a Mini-Bogorov Chamber (10 ml) for quali-quantitative analyses (Fig. 1). The organisms that tend to float at the surface, e.g., cladocerans, were pushed to the bottom of the chamber before starting the counting to avoid missing them. After identification and counting, the organisms were resuspended again in ethanol for long-term preservation (Fig. 1).

Each sub-sample was analyzed under a stereomicroscope (Leica M165C) at varying magnifications for species identification and abundance estimation. The number of individuals counted for each taxon was recorded in paper logs and subsequently used to calculate abundance, expressed as the number of individuals per unit volume of water (ind. m^{-3}). Abundance was calculated by considering the number of individuals counted in the sub-sample (n), the fraction of the original sample represented by the analyzed sub-sample (k), and the volume of water filtered by the net (V, m^3): $\mathbf{ind.\ m^3 = (n \times k)/\ V\ m^3}$.

Data processing. By harmonizing historical (1984–2015) and new datasets (2016–2024), we developed the Zooplankton dataset (ZOOGoN-40Y), covering the period 1984–2024. ZOOGoN-40Y is a standardized, FAIR-compliant dataset comprising 98,175 occurrence records and 1,295 sampling events. The temporal distribution and sampling density of the dataset, expressed as total zooplankton abundance per weekly bin, are visualized in Fig. 2, highlighting the longitudinal continuity of the series despite the historical gap in the early 1990s (1991–1995).

The data of abundance were organized in two temporal datasets, covering the periods 1984–2015, historical one, and a new dataset 2016–2024; each dataset includes taxonomic identification at different taxonomic level (Table 2), life-stage classification (female, male and juvaniles), and zooplankton abundance values expressed as individuals per cubic meter (ind. m^{-3}). The dataset from 1984 to 2020, previously maintained on local computers, has been migrated to centralized cloud storage on SharePoint to improve data accessibility and long-term preservation, and reproducibility. Since 2021, data collection has relied on digital survey forms implemented through KoBoToolbox, which replaced the previous locally managed workflow (Fig. 3). To facilitate subsequent analyses, the two datasets were merged into a single harmonized and consolidated unique dataset 1984–2024. To ensure consistency across the entire time series, particularly considering the changes introduced in the data collection workflow in 2021, taxonomic classification variables for the 1984–2020 period was harmonized using the *worrms* R package[12], which retrieves standardized taxonomic information from the World Register of Marine Species (WoRMS).

Data were processed through the ZooGoN R package, which implements a fully automated pipeline structured in two parallel workflows that generate a Darwin Core Archive and an interactive report, ensuring transparency, traceability, and reproducibility throughout the reporting process (Fig. 3). The workflow runs weekly at 00:00 UTC, ensuring that newly collected field data from KoBoToolbox is incorporated into the report and made available in the next weekly update. The rendered HTML outputs are automatically archived as downloadable artifacts via GitHub Actions, where they remain accessible for a 90-day retention period. Importantly,

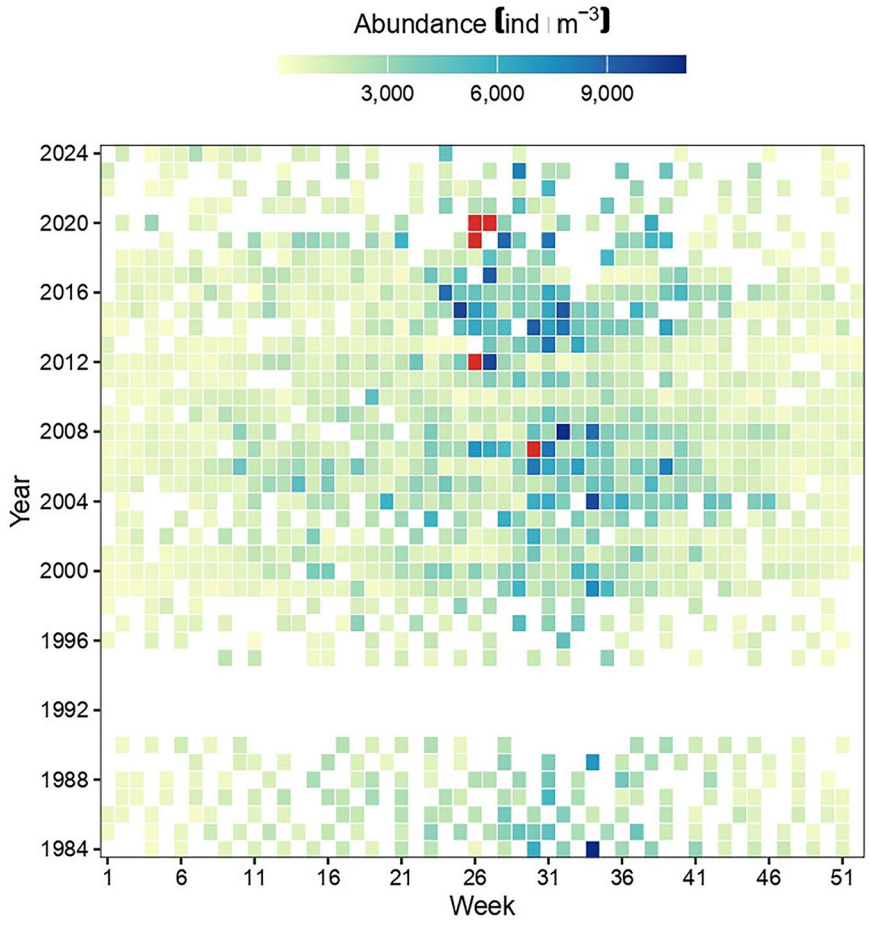


Fig. 2 Temporal coverage and abundance distribution of the ZOOGoN-40Y dataset (1984–2024). The heatmap represents the total zooplankton abundance (individuals m^{-3}) visualized by week (x-axis) and year (y-axis). Each tile represents a distinct sampling event; white spaces indicate periods where no sampling was conducted (notably 1991–1994) or data were unavailable. The color scale indicates total abundance on a gradient from light yellow (low) to dark blue (high), with the scale capped at 20,000 ind. m^{-3} for readability. Two observations exceeding this threshold are shown in red and represent exceptional high abundance events outside the range of the main color scale.

all new data collected via KoBoToolbox are not only included in the weekly report but are also converted into Darwin Core–compliant format, ensuring interoperability and facilitating integration with global biodiversity data repositories.

Darwin Core Archive and Global Biodiversity Information Facility. Darwin Core Archive (DwC-A) is a standardized format for packaging and publishing biodiversity data in a way that can be interpreted consistently by both people and information systems. Rather than being a single file, a DwC-A is a structured archive, usually distributed as a compressed file, that contains one or more delimited text files together with a descriptor file, meta.xml, which explains how the files are organized and how their columns correspond to Darwin Core terms, and metadata.xml file, which provides information about the dataset itself. Datasets require documentation such as a description (abstract) of the dataset, the agents responsible for authorship, publication and documentation, bibliographic and citation information, collection methods and much more. GBIF currently supports a metadata profile based on the Ecological Metadata Language 2.2.0. The concept of the archive is that its data files are logically arranged in a star-like manner: arranged around a central "core" data file, which can be linked to additional extension data files containing related information. Each extension record is linked to a record in the core file; allowing the fact that many extension records can exist for each single core record. This structure makes it possible to represent complex biodiversity datasets while maintaining a common vocabulary for key concepts such as taxonomy, geographic location, sampling events, dates, measurements, etc. By converting a dataset to DwC-A, the data are transformed from a project-specific structure into a standardized, exchangeable format that can be published through biodiversity data infrastructures such as GBIF, OBIS or EMODnet and more easily discovered, interpreted, integrated, and reused by the wider scientific community.

Rank Phylum	Rank Class	Rank Order	Phylum/Class	Order	Family	Genus	Species
Arthropoda	Copepoda	Calanoida		1	6	11	91
		Cyclopoida			2	8	49
		Harpacticoida		1		1	7
		Mormonilloida					1
		Siphonostomatoida				1	1
		Monstrilloida			1		
	Malacostraca	Amphipoda		1			
		Brachyura		1			
		Decapoda		1			2
		Euphausiacea		2			
		Isopoda		1			
		Mysida		1			
		Tanaidacea		1			
			1				
	Ostracoda		1				
	Cirripedia		1				
	Branchiopoda				1	2	7
Chordata	Appendicularia		1			2	2
	Ascidiacea		1				
	Thaliacea				2	1	4
	Teleostei			1			3
Chaetognatha	Sagittoidea		1				9
Mollusca	Bivalvia		1				
	Gastropoda	Pteropoda	1	1			
Cnidaria	Hydrozoa			1			12
	Hydroidolina		1				
	Scyphozoa		1				
Echinodermata	Echinoidea		1				
	Holothuroidea		2				
	Ophiuroidea		1				
	Asteroidea		1				
Ctenophora	Tentaculata					1	
Annelida	Polychaeta		1			1	
Platyhelminthes	Trematoda					1	
Bryozoa			1				

Table 1. Overview of the taxonomic composition of the ZOOGoN-40Y dataset, including detailed number of taxa identified and the level of the identification from order to species.

Taxonomic identification. Zooplankton dataset from 1984–2024 comprises 259 taxa belonging to 181 copepod and 78 others zooplankton taxa including olo- and meroplankton (Table 1). Among Copepoda six orders were identified: Calanoida, Cyclopoida, Harpacticoida, Mormonilloida, Siphonostomatoida, and Monstrilloida. A total of 149 copepods were identified at species level (Table 1). All adult individuals were distinguished by gender (females and males), while copepodite stages (CIII-CV) were identified and recorded separately. Moreover, 21 copepods were identified at genus level (e.g., *Calocalanus*, *Eucalanus*, *Diaxis* etc.), and 11 were identified at family and order level (Table 1). Other Zooplankton taxa were identified according to different taxonomic resolutions. Appendicularia, Branchiopoda, Siphonophora, Sagittoidea, and Thaliacea were identified also at species level (Table 1). Meroplanktonic larvae were classified at class or order level (Table 1). For Euphausiacea and Chordata, eggs and larvae were identified and recorded separately (Table 1). Supplementary taxonomic annotations, including new species records and non-indigenous species documented between 1984 and 2024, are provided in Table 2.

Data Records

The dataset is available in the Global Biodiversity Information Facility (GBIF) repository at https://doi.org/10.15468/3vrqmz[14].The dataset is organized as a Darwin Core Archive (DwC-A) comprising three core tables: Event, Occurrence, and Extended Measurement or Fact (eMoF). The dataset is published on Global Biodiversity Information Facility (GBIF), an international network and data infrastructure that provides open access to biodiversity data worldwide and hosted by GBIF Italy.

The Event table (event.txt) contains one record per sampling event, with 1,506 rows corresponding to the individual zooplankton tows collected between 1984 and 2024. Each record includes the sampling date (eventDate), geographic coordinates of the LTER-MareChiara station (decimalLatitude: 40.81, decimalLongitude:

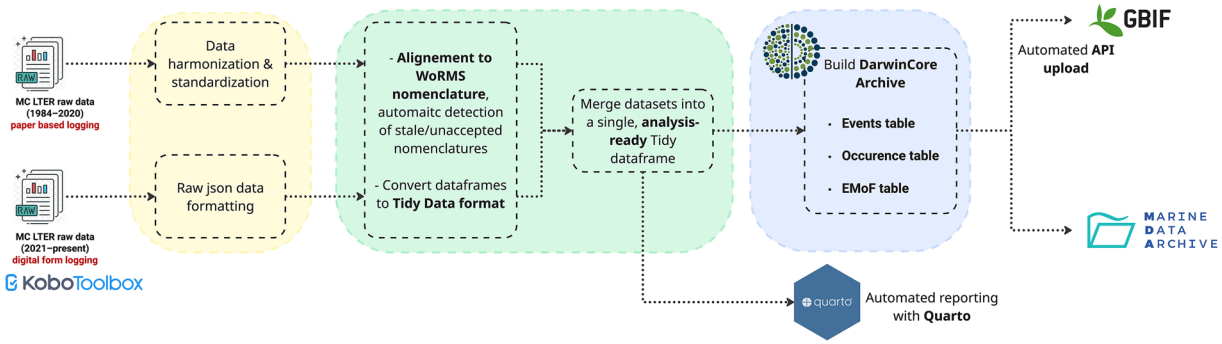


Fig. 3 Data management and automated processing pipeline for the ZOOGoN-40Y zooplankton dataset (1984–2024). The workflow illustrates the end-to-end provenance of the data, from raw acquisition to global repository dissemination. Data Input and Harmonization: Raw data are categorized into two streams: (left) records (1984–2020) originating from paper-based logging and contemporary records (2021–present) collected via digital survey forms using the KoBoToolbox platform. These streams undergo initial data harmonization, standardization, and tabular formatting from raw JSON files. Processing and Quality Control: The central pipeline (green shaded area) utilizes the worrms R package for automated alignment with the World Register of Marine Species (WoRMS) nomenclature, facilitating the detection of stale or unaccepted taxonomic names. Validated data are converted into a Tidy Data format and merged into a consolidated, analysis-ready dataframe. Standardization and Output: The processed data are structured into a Darwin Core Archive (DwC-A) (blue shaded area), comprising three relational tables: Events, Occurrence, and Extended Measurement or Fact (EMoF). In parallel, the pipeline utilizes Quarto and GitHub Actions to generate automated, interactive reports. Dissemination: The final FAIR-compliant outputs are disseminated via automated API upload to the Global Biodiversity Information Facility (GBIF) and preserved within the Marine Data Archive (MDA).

14.25), the sampling protocol (vertical tow, 0–50 m depth), and the sampling equipment used (net type and mesh size). Events are uniquely identified by a sequential identifier (eventID, e.g., mc_1 through mc_1506). The Occurrence table (occurrence.txt) links each species occurrence record (occurrenceID) to its corresponding sampling event via the eventID field. Each row represents a unique combination of sampling event and taxon, and life stage, and includes the scientific name (scientificName), the corresponding World Register of Marine Species Life Science Identifier (scientificNameID, in the format urn:lsid:marinespecies.org:taxname:XXXXXX), the specific nature of the data record (basisOfRecord) and an occurrence status flag (occurrenceStatus: present when individualCount exceeds zero, absent otherwise). The eMoF table (extendedmeasurementorfact.txt) stores quantitative and categorical measurements associated with each occurrence record. Abundance of mesozooplankton (size: >200 µm) per unit volume of the water body, determined by optical microscopy, is expressed as individuals per cubic metre (ind. m^{-3}) and linked to the BODC NERC Vocabulary Server URI for measurement type (http://vocab.nerc.ac.uk/collection/P01/current/ZU00M00D/) and measurement unit (http://vocab.nerc.ac.uk/collection/P06/current/UPMM/). For copepods, abundance is reported separately by gender (female, male) and life stage (juvenile), using controlled vocabulary identifiers from the NERC P01 collection for sex (http://vocab.nerc.ac.uk/collection/P01/current/ENTSEX01/) and life stage (http://vocab.nerc.ac.uk/collection/P01/current/LSTAGE01/). Measurement values for sex and life stage categories are encoded using the NERC S10 and S11 collections, respectively.

Technical Validation

Quality assurance of zooplankton counting relied on high-quality optical (instruments stereomicroscope), appropriate magnification and suitable background light conditions, as well as on sample-related factors such as sample richness and organism size and type (e.g., gelatinous forms). We ensured proper mixing before sub-sampling and counted enough individuals to guarantee representativeness and precision, while accuracy depends on taxonomic expertise. Our quality assurance actions include regular calibration of equipment, standardization of sub-sampling procedures, and ensuring a minimum number of counted individuals to achieve 95% confidence[15]. In addition, inter-calibration between operators was essential: throughout the year, approximately 5% of the analysed samples are re-counted by a second operator, often using the same sub-sample to verify consistency in organism counts. Continuous training and updated taxonomic literature further support consistency and accuracy. The most abundant species are counted with standard lab counters. As reported by Postel *et al.*[16], if all organisms are randomly distributed, following the Poisson distribution, the accuracy of a sample and the precision of a single count depends only on the number of specimens counted[17]. This can be inferred from Table which presents the precision in percentage terms that might be reached if the corresponding number of specimens is counted, together with the respective range within the organisms' actual number is expected to fall. Accordingly, doubling precision requires a fourfold effort in counting. For example, a precision of ± 20% requires the analysis of 100 specimens while those ± 10% needs 400 counts[18]. The 0.95 confidence limits (CL95) are also calculated from the number of counts (n): $CL95[\%] = \pm 2\,(100/\sqrt{n})$. In practice, one or more sub-samples (with the same concentration) should be analysed until 100 specimens of the most abundant taxonomic groups are reached in a sample. The estimation of the remaining, less common, groups is less precise. Finally, the remaining

Taxon/Species	Notes/Identification Status	MC ID/Range
Copepoda		
Ctenocalanus vanus (juveniles)	Pooled with other Calanoida juveniles	MC 1–6
Paracalanus (juveniles)	Pooled as Calanoida juveniles	MC 1–185
Clausocalanus juveniles	Reported as Calanoida juveniles until genus-level ID	Until MC 163; genus-level from MC 202
Corycaeus spp. (female and male)	Reported at genus level	MC 202–464
Corycaeus + *Farranula* juveniles	Pooled as Corycaeidae juveniles	MC 1–464
Oncaea spp. (females)	Includes *O. curta*, *O. scottodicarloi*, *O. media*	MC 1–552
Oncaea curta (females)	Identified from 2003	MC 566
Chiridius poppei	Identified from 2004	MC 609
Ditrichocorycaeus minimus	Identified from 2005	MC 687
Pontellina plumata (female)	Identified from 2005	MC 689
Triconia spp. (females)	Pooled until 2006; separated thereafter	–
Euchirella messinensis	Female (MC 787, 2007); male (MC 894, 2009)	–
Oncaea venusta (males)	Separated from MC 797; previously pooled	–
Triconia (males and juveniles)	Separated from MC 798 (2008)	Previously grouped under *Oncaea*
Parapontella brevicornis (juveniles)	Identified from 2008	MC 805
Pachos punctatum	Identified from 2008	MC 834
Juveniles of copepod species	Not identified at species level until 2008	MC 845
Oithona nana (juveniles)	Separated from females from MC 605	Earlier pooled
Mecynocera clausi (males)	Identified from 2001	MC 490
Lucicutia spp. (female)	Includes uncertain species IDs	MC 465 (2001)
Clausocalanus (males)	Includes specimens not identifiable to species	–
Euchirella juveniles	Includes *E. rostrata* and *E. messinensis* juveniles	Not species-level identifiable
Pseudodiaptomus marinus (female, male, juveniles)	New non-indigenous specie from 2014	MC 1110–1111
Rhincalanus nasutus	New record from 2017	MC 1249
Oithona davisae	New non-indigenous specie from 2018	MC 1302
Pontella lobiancoi	New record from 2019	MC 1323
Distioculus minor	New record from 2023	MC 1456
Paracalanus parvus	Treated as single species historically; now a species complex	–
Other Zooplankton		
Chordata	Eggs and larvae pooled	MC 1–613; separated from MC 614 onward
Evadne spp.	Identified at genus level	MC 210–268
Sagitta decipiens	Identified at species level	MC 801 (2008)
Pterosagitta draco	Identified at species level	MC 940 (2010)
Tomopteris	Separated from polychaete larvae	From MC 991 (2012)
Mnemiopsis	New record from 2016	MC 1223 (2016)
Evadne nordmanni	New record from 2021	MC 1381
Euphausiacea	Eggs and larvae pooled	MC 1385; separated from MC 1386 onward
Pleopis schmackeri	New record from 2024	MC 1496 (2024)
Liriope tetraphylla	Ephyrae and juveniles identified at species level	From MC 942 (2010)

Table 2. List of zooplankton taxa analyzed, indicating the taxonomic level of identification and temporal changes (MC) in the separation of juvenile stages, pooled groups, and identifications at genus or species level.

part of the total sample can be surveyed for the rare species which may be missed during sub sampling and therefore qualitatively noted as present with abundance of 0.02 ind. m^{-3}.

Usage Notes

Since 1984, sampling activities have been primarily based on data collected at fortnightly intervals (until 1991) and subsequently on a weekly basis (from 1995 to 2016). From 2016 onwards, due to logistical constraints, sampling frequency has been reduced to fortnightly and monthly intervals. An interruption in the time series occurred between 18 December 1990 and 1 March 1995. Supplementary taxonomic annotations, including new species records and non-indigenous species documented between 1984 and 2024, are provided in Table 2. Juveniles of copepods corresponding to developmental stages CIII-CV copepodites, depending on the species, were effectively retained using a 200µm mesh size.

Data availability

The data supporting the findings of this study are publicly available through the Global Biodiversity Information Facility (GBIF) at https://doi.org/10.15468/3vrqmz[14].

Code availability

All code used to process the dataset and generate the Darwin Core Archive is available in the ZooGoN R package (version 3.0.0; R ≥2.10), hosted on GitHub at https://github.com/ioledc/ZOOGoN-40Y under the GPL (≥3) license.

Received: 18 May 2026; Accepted: 26 October 2026;
Published: xx xx xxxx

References

1. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How Many Species Are There on Earth and in the Ocean? *PLoS Biol* **9**, e1001127 (2011).
2. Coll, M. *et al*. The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLoS ONE* **5**, e11842 (2010).
3. Duarte, S., Vieira, P. E., Lavrador, A. S. & Costa, F. O. Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *Science of The Total Environment* **751**, 141729 (2021).
4. Haubrock, P. J. *et al*. Long-term trends in abundances of non-native species across biomes, realms, and taxonomic groups in Europe. *Science of The Total Environment* **884**, 163808 (2023).
5. Lomartire, S., Marques, J. C. & Gonçalves, A. M. M. The key role of zooplankton in ecosystem services: A perspective of interaction between zooplankton and fish recruitment. *Ecological Indicators* **129**, 107867 (2021).
6. Hobday, A. J. *Impacts of Climate Change on Australian Marine Life: CSIRO Marine and Atmospheric Research Report to the Australian Greenhouse Office, Department of the Environment and Heritage*. (Department of the Environment and Heritage, Canberra, A.C.T., 2006).
7. Carter, J. L., Schindler, D. E. & Francis, T. B. Effects of climate change on zooplankton community interactions in an Alaskan lake. *Clim Chang Responses* **4**, 3 (2017).
8. Beaugrand, G., Brander, K. M., Alistair Lindley, J., Souissi, S. & Reid, P. C. Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661–664 (2003).
9. Richardson, A. J. In hot water: zooplankton and climate change. *ICES Journal of Marine Science* **65**, 279–295 (2008).
10. Conversi, A. *et al*. A holistic view of marine regime shifts. *Phil. Trans. R. Soc. B* **370**, 20130279 (2015).
11. Kim, S., Cheon, S., Park, C. & Soh, H. Y. Integrating DNA metabarcoding and morphological analysis improves marine zooplankton biodiversity assessment. *Sci Rep* **15**, 7283 (2025).
12. Chamberlain, S. & Vanhoorne, B. *worrms: World Register of Marine Species (WoRMS) Client. R package v.0.4.3* https://doi.org/10.32614/CRAN.package.worrms (2023).
13. Di Capua, I. *et al*. From Phenotypes to Genotypes and Back: Toward an Integrated Evaluation of Biodiversity in Calanoid Copepods. *Frontiers in Marine Science* **9**, 833089 https://doi.org/10.3389/fmars.2022.833089 (2022).
14. Di Capua, I., Marcucci, A. & Longobardi, L. Zooplankton data from LTER MareChiara site in the Gulf of Naples, Mediterranean Sea, (Jan 1984-Dec 2024). *GBIF* https://doi.org/10.15468/3vrqmz (2026).
15. Hensen, V. & Hensen, V. *(1887) Ueber Die Bestimmung Des Plankton's, Oder Des Im Meere Treibenden Materials an Pflanzen Und Thieren. Fünf. Bericht Komm Wissensch. Unters. Deutschen Meere Kiel, 12, Bis* **16**, 1–108 (1887).
16. Postel, L., Fock, H. & Hagen, W. Biomass and abundance. in *ICES Zooplankton Methodology Manual* (eds Harris, R. P., Wiebe, P. H., Lenz, J., Skjoldal, H. R. & Huntley, M.) 83–192 (Academic Press, Boston, 2000).
17. Cassie, R. M. *A Manual on Methods for the Assessment of Secondary Productivity in Fresh Waters*. (Edmondson, W. T.; Winberg, G. G.).
18. Lund, J. W. G., Kipling, C. & Le Cren, E. D. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia* **11**, 143–170 (1958).

Acknowledgements

We thank all LTER-MC and NEREA teams, and the crew of the R/V "Vettoria" for assistance during the work at sea, and F. Margiotta for support during the initial phase of the DTO Bio-Flow project.

The LTER-MC project (https://deims.org/0b87459a-da3c-45af-a3e1-cb1508519411) is a member of the Long Term Ecological Research national and international networks: LTER-Italy (https://ror.org/05ma8mw15), LTER-Europe and ILTER

Author contributions

Iole Di Capua: Conceptualization; Quali-quantitative analyses; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Resources; Supervision; Original draft; Writing - review & editing. Andrea Marcucci: Data-set organization, Data curator, Formal analysis; Original draft; Writing - review & editing. Lorenzo Petrella: support for data organization; Formal analysis; Original draft; Writing - review & editing. Maria Grazia Mazzocchi: Formal analysis part of historical samples; Review & editing. Lynn Delgat: Support for data management; Data Validation; Original draft; Writing - review & editing. Cyril Radermecke: Support for data management; Data validation; Original draft; Writing - review & editing. Ilaria Rosati: Data management; Data Validation; Original draft; Writing - review & editing. Lorenzo Longobardi: Data management; Data curator; Investigation; Supervision; Original draft; Writing - review & editing.

Funding

Data transformation and publication was funded by the European Union under the Horizon Europe Programme with Grant Agreement No. 101112823 (DTO-BioFlow) through an FSTP grant. This work was supported by the Stazione Zoologica Anton Dohrn (SZN) within the framework of the Long-Term Ecological Research MareChiara (LTER-MC) research program.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.D.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.