ITINERIS

Deliverable D6.22 – Activity 6.17, CNR IRET LE

“RELEASE OF THE DATASETS ON SPECIES TROPHO-ENERGETICS AND SPACE USE BEHAVIOURAL TRAITS"

AUTHORS: ALBERTO BASSET, JESSICA TITOCCI, LORENZO LIBERATORE, FLAVIO MONTI AND TEODORO SEMERARO

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Deliverable number:	D6.22
Work package:	WP6
Intermediate Objective:	IO1.1
Deliverable type:	X Document, report
	☐ Websites, patent filings, videos, etc.
	☐ Other: please specify ....
Dissemination level:	X Public
	☐ Restricted
Estimated delivery (bimester):	B14
Actual delivery date:	15/03/2024
Author(s) (Partner-OU):	Lorenzo Liberatore, Flavio Monti, Teodoro Semerato, Andrea Tarallo, Jessica Titocci (CNR IRETLE), Milad Shokri, Parisa Taban (UNISAL), Vanessa Marrocco (LifeWatch ERIC)
Reviewed by:	ITINERIS Executive Board
Note:	

Table of contents

Index of tables

1.1. OVERVIEW OF THE ACTIVITY 6.17

The activity 6.17 entitled "*Data production experimental center on tropho-energetic and space use behavioral traits of aquatic organisms*" is part of the Working Package 6 "Terrestrial Biosphere" of the Italian Integrated Environmental Research Infrastructures System (ITINERIS).

ITINERIS is a project funded by EU - Next Generation EU PNRR- Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realization of an integrated system of research and innovation infrastructures".

The activity 6.17 is aimed at:

1) organize catalogue and make findable and accessible existing data resources on the tropho-energetics and space use behavior traits of aquatic and terrestrial organisms;

2) integrate existing data with new intensive laboratory and field data production by extending and updating the facilities already available at the URT IRET Lecce, in collaboration with the University of Salento realizing an experimental data production centre on tropho-energetics and space use behavior of aquatic organisms;

3) open the access of the facilities to the staff on the ITINERIS RIs and to external researcher groups. The structured data resources include individual body size, standard metabolic rates, trophic niche breadth, patch use, giving up time and giving up density, home range, connected with main abiotic metabolic rates, including water temperature, for their application to advanced case uses.

The experimental data production centre, which already equipped with 2 series of thermostatic rooms for experimental studies on functional biodiversity responses to climate change, makes available to the access the updated facilities for the data production on the following traits:

a) Individual level metabolic rates of aquatic organisms, including primary producers through high precision Strathkelvin systems for respiration rate assessment;

b) Individual space use behavior (i.e., patch selection, use, giving-up and density, home range size) both in the lab, extending and updating an innovative Noldus system and in the field, at the LTER site of the Aquatina lagoon, with individual tracking methodologies in connection with the European Fish Tracking Network;

c) Individual trophic niche within food web networks using stable isotope and NGS methodologies. The implementation of an intranet connectivity of all equipment with the LIFEWATCH-Italy National Hub of Biodiversity and Ecosystem Research Data ensure real time data storage.

1.2. STATE OF THE ART OF THE DELIVERABLE 6.22

The deliverable 6.22 refers to the “*Release of the datasets on species tropho-energetics and space use behavioural traits*” within the activity 6.17 of the WP6.

The lack of global cooperation in biodiversity research coupled with the poor use of common standards and guidelines for collecting, organizing and describing biodiversity data hampers the effective utilisation of data produced, ultimately jeopardising biodiversity management efforts. In the current scenario of global biodiversity loss, make existing heterogeneous biodiversity and trait-based data Open and FAIR (Findable, Accessible, Interoperable, Reusable) is needed.

To enhance our understanding of functional biodiversity responses to environmental changes, aiming at biodiversity conservation, we focused on different levels of investigation by working on:

A) An individual-level trait-based phytoplankton dataset from transitional waters;
B) Individual-based morphological trait dataset of phytoplankton from brackish and saline coastal lagoons in the Mediterranean and Black Sea;
C) A Mediterranean phytoplankton community dataset from Salento peninsula coastline;
D) Composition and Functional Traits of Freshwater Phytoplankton from Umbrian Lakes;
E) Composition and morpho-functional traits of zooplankton swimmers and sinkers collected by "Mooring Dirigibile Italia" sediment trap;
F) A global dataset of terrestrial and aquatic vertebrate behavioural and life-history traits.

Organisms modulate their spatial behaviour primarily to optimize their energy acquisition. Understanding the spatial and temporal contexts in which interactions among species occur, as well as the dynamics of resource-mediated interactions, is critical for investigating the mechanisms underlying interspecific coexistence. However, external factors can influence energy requirements and hence space use behavior, especially within the climate and environmental change era. This requires a huge effort in data harmonization by systematically aligning, integrating, and standardizing heterogeneous datasets from multiple sources to strengthen and expand spatial and temporal home range analysis.

In this context, LifeWatch ERIC, the European e-Science Infrastructure for Biodiversity and Ecosystem Services, technologically supports collaboration between researchers and stakeholders by providing several e-Science facilities such as a unified platform, web services and tools for data harmonization, exchange and analysis and long-term sustainability.

We are currently working on the application of data harmonization using LifeWatch data services for the production of Findable, Accessible, Interoperable and Reusable (FAIR) datasets, at all levels.

Enhanced availability and interoperability of these data will turn existing biodiversity information into actionable knowledge, to be used to model future trends on responses to anthropogenic threats and global change.

1.3. GENERAL PROCEDURE FOR DATASETS BUILDING UP

The initial phase of this activity focused on establishing the analytical framework and procedure for the development of the datasets containing information on both plankton individual-traits and vertebrate behavioural and life-history traits. We decided to work at individual level for the former and at global scale for the latter. The procedure adopted for building up the dataset followed the application of data harmonization steps, using LifeWatch data services for the production of Findable, Accessible, Interoperable and Reusable (FAIR) data.

The various steps of the procedure are graphically represented in Figure 1:

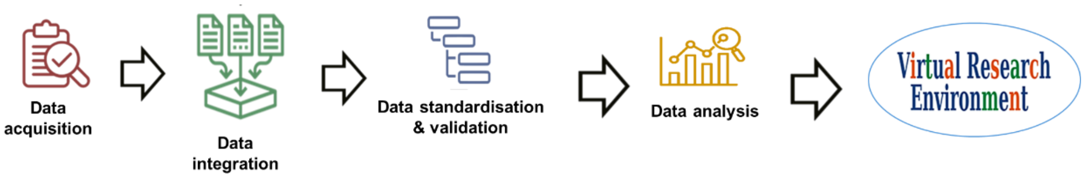


Figure 1: Procedure adopted for building up the datasets.

1- Data acquisition: Identifying and accessing existing data resources from internal experimental and monitoring activities conducted by the OU team, as well as gathering accessible data from literature and online repositories;

2- Data integration: combining data residing in different sources and providing a unified version of them. The data to be integrated were collected from heterogeneous datasets and transformed into a single coherent dataset.
3- Data standardization and validation: organize, curate and validate the data. This activity includes a process of data harmonization, metadata description and semantic and syntactic standardization of the data and metadata.
4- Data analysis: this step will focus on investigating how extrinsic and intrinsic factors affect taxonomic and functional composition and dynamics at different spatial and temporal scales. This step of analysis can include also the possibility of integrating remote sensing data in order to simulate provisional models for future projections and impact of climate change scenarios.
5- VRE: generated datasets will be used in specific case studies to test and validate the functionalities of the Phytoplankton and Biomass Virtual Research Environments and related workflows.

2.1. AN INTEGRATED INDIVIDUAL-LEVEL TRAIT-BASED PHYTOPLANKTON DATASET FROM TRANSITIONAL WATERS

Here below details of the different phases of the analytical framework and procedures represented above are reported:

Data acquisition:

Phytoplankton samples were collected in a single sampling event that took place between July 2010 and November 2012 in 24 transitional water ecosystems distributed across five biogeographical regions: Northern Atlantic Ocean (NAO-Scotland), South-Western Atlantic Ocean (SWAO-Brazil), South-Western Pacific Ocean (SWPO-Australia), Indo Pacific Ocean (IPO-Maldives) and Mediterranean Sea (MED- Greece and Turkey) (Fig. 2). Sampling was carried out according to a hierarchical sampling design: for each ecoregion, three ecosystems were selected and within each of these, a maximum of three habitat categories were chosen and three experimental stations per habitat type were sampled with three replicates each, for a total count of 116 sites and approximately 350 water samples.

Habitat types were classified based on sediment granulometry and type of bottom vegetation according to the EUNIS habitat type hierarchical classification, version 2012.

Phytoplankton samples were collected with horizontal tows from the subsurface (0.5 m) using a net mesh (6µm) and fixed with Lugol's solution (15mL/L). During the net sampling phase, the net was towed from the boat for a standard length of approximately 1.5–2m, repeated three times back and forth, with each haul consisting of a linear measure of approximately 10m. Phytoplankton taxonomic identification, cell abundances estimations and morphometric measurements were performed using an inverted microscope (Nikon T300E, Nikon Eclipse Ti) connected to a video-interactive image analysis system (L.U.C.I.A Version 4.8, Laboratory Imaging), following the Utermöhl method at 400x magnification. For each sample, a minimum of 400 cells were counted, measured and identified to the lowest taxonomic level possible, using specific manuals, monographs and phytoplankton Atlas. After taxonomic identification, cell volumes (expressed in µm3) were estimated according to the species/taxa specific shape association and using the geometric equations for simple and complex shapes recorded in the webservice "Atlas of Shapes" https://www.phytovre.lifewatchitaly.eu/vre/shapes-groups/. The geometric shape was attributed to the shape of the individual cell, even for coenobial, colonial and filamentous species where cells were not observable. Phytoplankton cellular carbon content (pg C) was obtained indirectly by converting cell biovolume to carbon using empirical or theoretically derived equations in accordance with Menden-Deuer and Lessard, 2000.

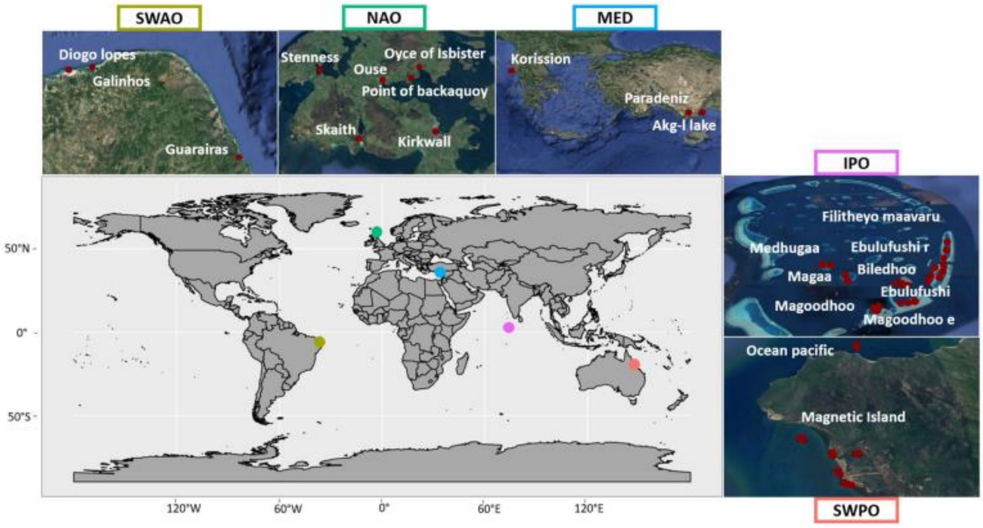


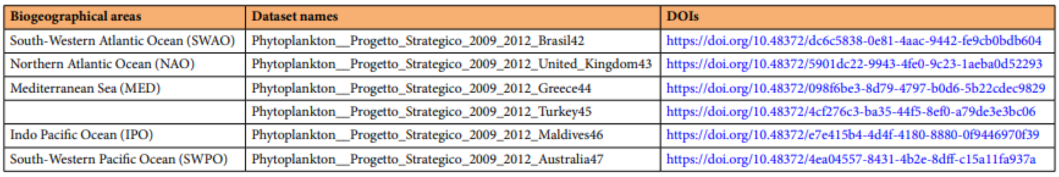
Figure 2. Distribution map of the five biogeographical areas included in the dataset: South-Western Atlantic Ocean (SWAO) in brown, Northern Atlantic Ocean (NAO) in green, Mediterranean Sea (MED) in light blue,

Indo Pacific Ocean (IPO) in purple and South-Western Pacific Ocean (SWPO) in pink. The red dots identify the phytoplankton sampling stations in each biogeographical area.

Data integration:

In total, 127,311 phytoplankton cells, representing 306 taxa, were counted, measured, and taxonomically classified. All data were collected in six individual datasets. As these datasets contain a different number of species each with available data, sometimes expressed even at the subspecies level, a long process of filtering, checking and cleaning the data was carried out. We proceeded by combining data from the six individual datasets generated from phytoplankton collected in each ecoregions in an integrated and unified individual-level trait-based phytoplankton dataset. Below the list of the datasets with their respective DOI (Table 1).

Table 1. List of datasets and respective DOIs



Biogeographical areas	Dataset names	DOIs
South-Western Atlantic Ocean (SWAO)	Phytoplankton__Progetto_Strategico_2009_2012_Brasil42	https://doi.org/10.48372/dc6c5838-0e81-4aac-9442-fe9cb0bdb604
Northern Atlantic Ocean (NAO)	Phytoplankton__Progetto_Strategico_2009_2012_United_Kingdom43	https://doi.org/10.48372/5901dc22-9943-4fe0-9c23-1aeba0d52293
Mediterranean Sea (MED)	Phytoplankton__Progetto_Strategico_2009_2012_Greece44	https://doi.org/10.48372/098f6be3-8d79-4797-b0d6-5b22cdec9829
	Phytoplankton__Progetto_Strategico_2009_2012_Turkey45	https://doi.org/10.48372/4cf276c3-ba35-44f5-8ef0-a79de3e3bc06
Indo Pacific Ocean (IPO)	Phytoplankton__Progetto_Strategico_2009_2012_Maldives46	https://doi.org/10.48372/e7e415b4-4d4f-4180-8880-0f9446970f39
South-Western Pacific Ocean (SWPO)	Phytoplankton__Progetto_Strategico_2009_2012_Australia47	https://doi.org/10.48372/4ea04557-8431-4b2e-8dff-c15a11fa937a

Data standardization and validation:

Data curation and technical validation were performed to ensure the accuracy of both the data and metadata (Figure 3). During data collection, a standardized sampling protocol, previously used in other studies, was consistently applied throughout the sampling campaign to avoid bias and ensure replicability. All samples were collected, identified, and measured by a team of qualified researchers and taxonomists, who ensured data quality by validating taxonomic and morphological classifications, as well as addressing formatting, nomenclature errors, or missing and inconsistent data. The data were harmonized using the Phytoplankton Data Template (https://www.phytovre.lifewatchitaly.eu/phyto-data-template/), which is based on the Darwin Core standards and the Phytoplankton Traits Thesaurus, part of the Phytoplankton Virtual Research Environment of LifeWatch Italy.

Additionally, the taxonomic and morphological information in all datasets were cross-checked and validated using the WoRMS and AlgaeBase repositories, as well as the web services "Atlas of

Shapes" and "Trait Computation" available through the Virtual Research Environment "Phyto VRE" of LifeWatch Italy.

The datasets were formatted as column-oriented tables in semicolon-separated values (.csv) format. Metadata were described using the Ecological Metadata Language (EML 2.2.0) standard in extensible markup language (.xml) format, ensuring long-term accessibility and proper data interpretation. Each phytoplankton record within the datasets is represented by a unique identifier (catalogNumber) and linked with ancillary data (e.g., sampling locations, temporal and spatial details), taxonomic classification, and morphological trait data (such as size and shape).

Upon completion of these curation and validation processes, the data were stored and preserved in the LifeWatch Italy Data Portal (https://dataportal.lifewatchitaly.eu/data), ensuring they are findable, accessible, interoperable, and reusable (FAIR) and have been published on Scientific Data Journal (Lairab et al., 2023. DOI: https://doi.org/10.1038/s41597-023-02785-w).

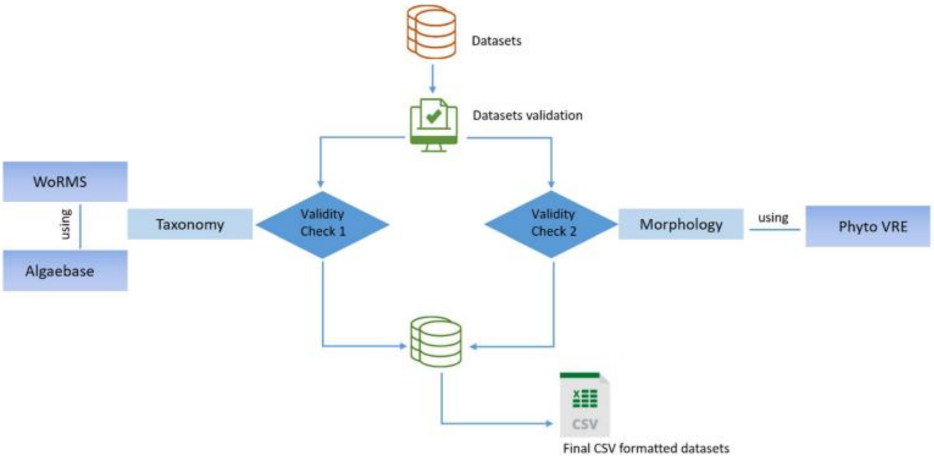


Figure 3. Schematic illustration of the data validation process. Taxonomical and morphological information were checked and technically validated through the use of WoRMS and Algaebase repository and the web services of the Virtual Research Environment “PhytoVRE”.

Data analysis:

The datasets have already been used for morpho-functional analyses and hypothesis testing on phytoplankton guilds at different levels of data aggregation and scale, from local to global (Lairab et al., 2024, DOI: https://doi.org/10.3390/d16020098).

Summarized information from each dataset is presented in Table 2.

Table 2. Summary of total abundance, taxa richness and shape occurrence from five biogeographical areas.

Biogeographical areas	Abundance (n. of cells)	Taxa Richness	Shape occurrence
South-Western Atlantic Ocean (SWAO)	32400	96	21
Northern Atlantic Ocean (NAO)	22396	120	20
Mediterranean Sea (MED)	24080	110	19
Indo Pacific Ocean (IPO)	28296	116	23
South-Western Pacific Ocean (SWPO)	20139	114	19

The integrated phytoplankton dataset will feed the workflows of the Phytoplankton VRE (Figure 4) and will be used to perform various analyses. The Phytoplankton VRE will use phytoplankton functional traits as essential biodiversity variables (EBVs) to advance the integrated collection, sharing and analysis of information on phytoplankton biodiversity. Below is an example of how the integrated phytoplankton dataset can be accessed and analysed in the Phytoplankton VRE workflow.

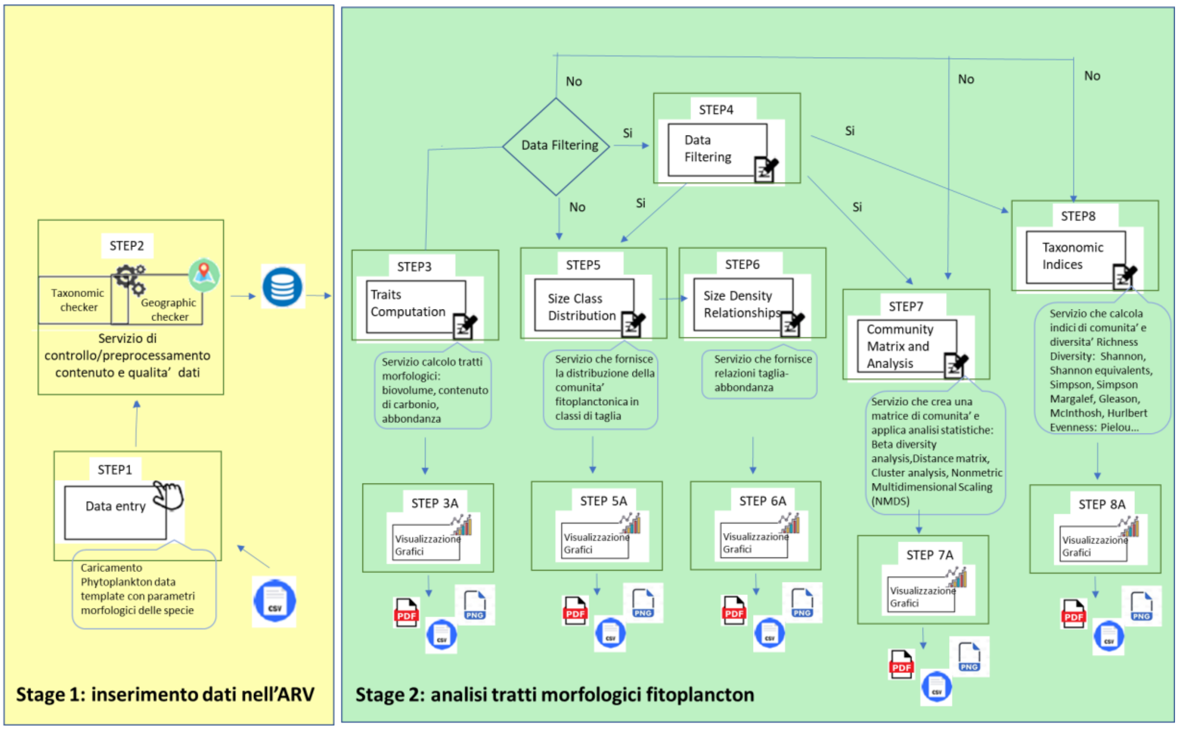


Figure 4. Analytical workflow of the Phytoplankton VRE (part extrapolated from the Term of Reference).

Through this workflow several analysis can be performed such as the computation of demographic traits (density, biovolume, carbon content) and then the size class distribution, size density relationships or community and diversity analysis.

A further application of this dataset was conducted to investigate the potential distribution of phytoplankton communities in both current and future scenarios, using an ecological niche-based approach. This has been done modelling the potential distribution of organisms around the world seas by considering their body shapes and relationships with some abiotic parameters in two scenarios: current one and 2100 future scenario. In this way the analysis, through application of *Machine Learning* model based on *Maximum Entropy* algorithm, can highlight which areas are more suitable for each body shape and how they evolve in warmer future scenario. Through niche model analysis, instead, it is possible investigate the variations of ecological niche parameters (breath and overlaps) in changing scenario. Thus, we can obtain distribution maps of the modelled suitable areas of phytoplankton (as body shapes) and parameters that influenced more the performances of the algorithms, showing a strictly relationships between rounded and globular shapes with temperature and nutrient concentration increases, as reported in literature too. Another conclusion is that all models reveal an increase in niche width in future scenario and a more overlapping distribution.

The methodological approach and preliminary results have been recently shown at the *CNR IRET Annual Conference* (Liberatore et al., 2025; see Tab. S1).

2.2. INDIVIDUAL-BASED MORPHOLOGICAL TRAIT DATASET OF PHYTOPLANKTON FROM BRACKISH AND SALINE COASTAL LAGOONS IN THE MEDITERRANEAN AND BLACK SEA.

Data Acquisition

This dataset comprises 121,764 individual-level phytoplankton records collected from 14 sites and 55 stations across the northern Mediterranean and western Black Sea lagoons (Figure 5). The data include taxonomic classification, ancillary environmental parameters, and morphometric trait data, standardized following FAIR principles. Phytoplankton and abiotic data were collected as part of the INTERREG IIIB TWReference Net Project in five countries: Albania, Bulgaria, Greece, Italy, and Romania. Data collection was conducted during two seasons: autumn 2004 and spring 2005, specifically in the months of October and April, within the habitat type of each lagoon. A hierarchical sampling design was followed to account for both seasonal and spatial variations in phytoplankton communities. Phytoplankton samples and abiotic parameter were collected from sampling stations distributed in 5 regions and 14 sites, with approximately 4 replicates per station. The water samples were collected from 0.5m depth using Ruttner bottle and then were transferred to PET bottles, fixed with 1% Lugol’s solution (15ml/l), kept cold at 4°C in darkness, and analysed for not more than two months after sampling. For each phytoplankton sample, 400 cells were counted measured and identified to determine the individual cell size and taxonomy. The taxonomic identification was done using specific monographs, manuals and atlases of phytoplankton (Al-Kandari et al., 2009; Haraguchi & Odebrecht, 2010; Pey et al., 2014; Rampi & Bernhard, 1978; Sournia, 1986; Tomas, 1997). All the cells were identified to the nearest possible taxonomic level when the species level identification was not possible different qualifiers i.e. sp., spp. and cf. were used to indicate the specimen to its relevant species/genus. The geometric shapes were attributed to the shapes of the cells, even for colonial, coenobial and filamentous species where cells were not observable. After the taxonomic identification the volume of each cell expressed in unit-µm3 was calculated according to the its associated shape, using the geometric equations and visual display of simple and complex shapes available on the web-service of Phytoplankton “Atlas of Shapes” https://www.phytovre.lifewatchitaly.eu/vre/shapes-groups/. Cell volume was then converted to individual cell weight (pgC) in accordance with Menden-Deuer and Lessard (Menden-Deuer & Lessard, 2000).

Abiotic water column parameters (Dissolved Oxygen, pH, water salinity and temperature) were monitored at each station during sampling activities using a multiparametric probe (WTW Multi 3500i). These study sites represent microtidal and nutrient-rich depositional ecosystems, selected based on anthropogenic pressure gradients affecting phytoplankton trait adaptations.

The dataset will be openly accessible on the LifeWatch Italy data portal, supplemented with metadata descriptions. The harmonized dataset can be used for the variety of analyses including trait-based analyses to for the study the intraspecific and interspecific trait variations of phytoplankton guilds.

The dataset consists of:

1. Phytoplankton Morphological Trait Dataset – Individual-based trait records from various sites.
2. Abiotic Dataset – Measurements of salinity, conductivity, temperature, pH, dissolved oxygen, total dissolved solids, and nutrient concentrations (NO_2, NO_3, NH_4, PO_4, SiO_2).

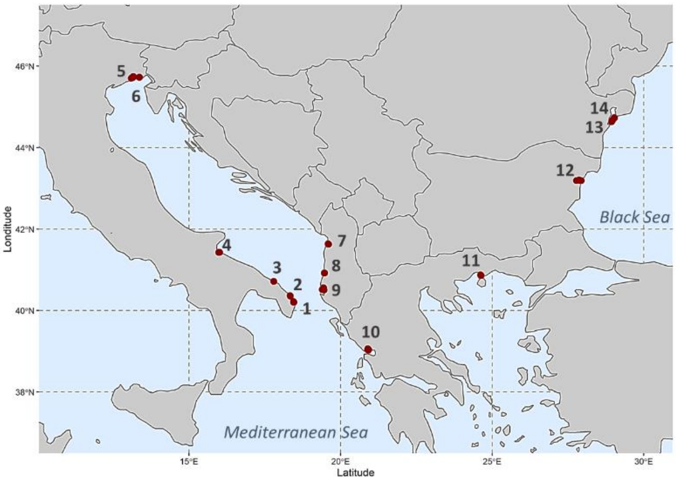


Figure 5. Map of Mediterranean and Black Sea lagoons included in the dataset. The 14 sites are 1-Alimini, 2-Cesine, 3-Torre Guaceto, 4-Margherita di Savoia, 5-Grado valli de Pesca, 6-Grado marano, 7-Patok, 8-Karavasta, 9-Narta, 10-Logarou, 11-Agiasma, 12-Varna, 13-Sinoe and 14-Lehaova.

Data Standardization & Validation

A uniform sampling protocol ensured data consistency across locations. Samples were collected, identified, and measured by qualified taxonomists, following a curated Phytoplankton Data Template (https://www.phytovre.lifewatchitaly.eu/phyto-data-template/), which includes metadata based on Darwin Core standards (Wieczorek et al., 2012) and the Traits Thesaurus (https://doi.org/10.48373/sa6p-ta25). Each phytoplankton record is assigned a unique identifier (catalogue Number) with associated morphometric and taxonomic data. Validation was performed using AlgaeBase (Guiry, 2013) and web services from LifeWatch Italy's Phyto VRE (https://www.phytovre.lifewatchitaly.eu/), including "Atlas of Shapes" and "Trait Computation" tools. After all the validation and curation steps, the data will be stored in the LifeWatch Italy data portal. Abiotic and biotic dataset were formatted as column-oriented table, with the data reported in comma-separated value (.csv) format. To ensure the understanding and long-term control of the data, the associated metadata were described using the Ecological Metadata Language (EML 2.2.0) standard in extensible mark-up language (.xml) format.

Data analysis:

Preliminary analysis and results have been made at the taxonomic and trait-based level. The dataset includes 253 accepted taxa across 10 phyla (Figure 6). Cryptista was the most abundant group, mainly represented by Cryptophyceae (genera: *Cryptomonas*, *Teleaulax*). Heterokontophyta was the most diverse phylum (96 taxa), followed by Dinoflagellata (71 taxa) and Chlorophyta (40 taxa). Chaorophyta, Chromista, and Ciliophora (e.g., *Mesodinium rubrum*) were minimally represented (<1%). 21% of taxa were classified as "others" (e.g., phytoflagellates and microflagellates with unidentified taxonomy). 24 different phytoplankton shapes were identified (Figure 7). The most abundant shape was Cone + Half Sphere, while others (e.g., Cylinder, Prism on Triangle Base 2) had <10,000 individuals each.

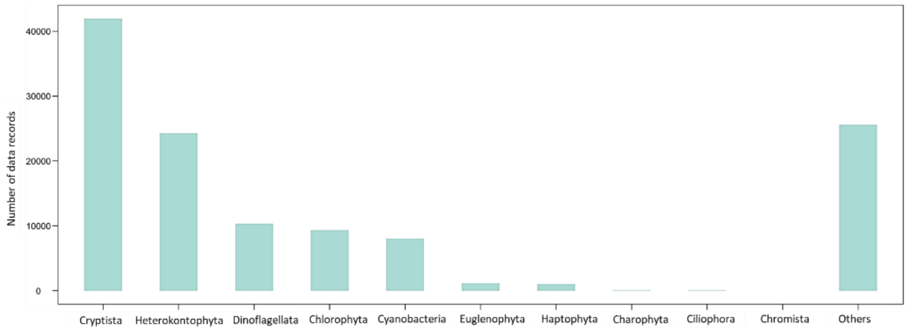


Figure 6. Taxonomic distribution of phytoplankton data records in each phyla. Others include phytoflagellates, microflagellates and phytoplankton undetermined cells.

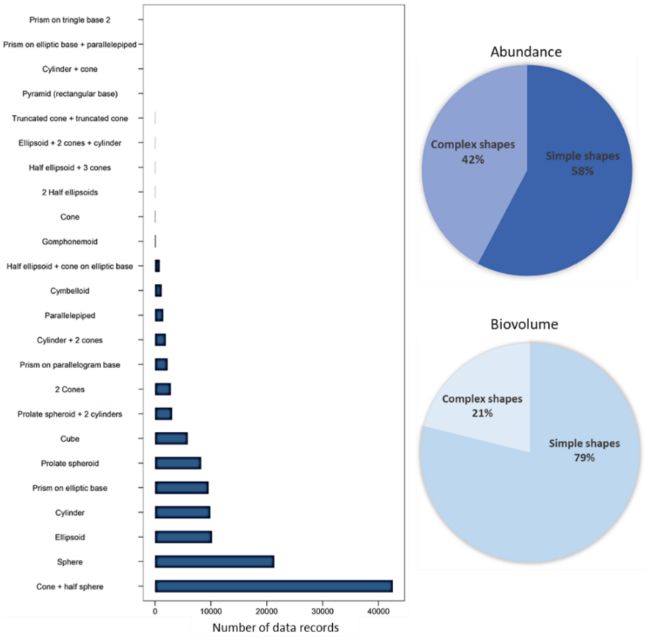


Figure 7. Shape distribution of phytoplankton cells in terms of data records and distribution of simple and complex shapes in terms of abundance and biovolume.

Further analyses will be made and the dataset will be incorporated into LifeWatch Italy's Phytoplankton VRE (workflow 1 and 2), enhancing trait-based research on intraspecific and interspecific variations in phytoplankton guilds. A data paper on the individual-based morphological trait data of phytoplankton from brackish and saline coastal lagoons in the Mediterranean and Black Sea is currently in preparation for Biodiversity Data Journal.

2.3. A MEDITERRANEAN PHYTOPLANKTON COMMUNITY DATASET FROM SALENTO PENINSULA COASTLINE.

Another dataset from a 2000 phytoplankton sampling campaign along the Salento Peninsula coastline is also part of this group's work. The harmonization and FAIRification process of this dataset is still in progress and is expected to be completed and published in the coming months.

Data acquisition:

Phytoplankton samples were collected during four seasonal cruises along the Southern Apulian coast (Adriatic and Ionian Seas) in March, June, September, and December, within the TWREFERENCENET, INTERREG IIIB CADSES project. Sampling was conducted along seven transects at five distances from the coastline and three depths. In the same time, abiotic data—including depth, temperature, DIP, DIN, silicate, and dissolved oxygen—were recorded at each sampling point.

The dataset includes phytoplankton taxonomic identification (genus/species or higher classification), biovolume data and abiotic data for each sampling point, defined by cruise, distance from shore, and depth.

In total, 40,688 phytoplankton cells, representing 364 taxa, were collected and taxonomically classified. For each taxonomic category, mean biovolume (V) value was calculated and, then, biomass values (W) were obtained, applying the following formula: $W=(V*1.03)/1000$, according to Hutchinson (1967). 320 taxa, the more representative (more than 10%) were categorized into size classes, calculated as natural logarithm of biomass values.

Data Standardization & Validation:

This process is still ongoing. Taxonomic data validation will be conducted using AlgaeBase and the LifeWatch Species Information Backbone (LW-SIBb). Morphological data will be annotated following the Traits Thesaurus and Darwin Core standards, while abiotic parameters will be standardized using NERC vocabularies. All datasets will be formatted as CSV tables, with metadata described according to the Ecological Metadata Language 2.0, before publication on the LifeWatch Italy data portal.

Data analysis :

All these data will be reported in specific tables and wille be useful to analyze the ecological responses to abiotic parameters, belonging an ecological niche approach, in particular through overlap analysis.

This dataset will be used as case study to test and validate the workflow 3 of the Phytoplankton Virtual Research Environment (Figure 8).

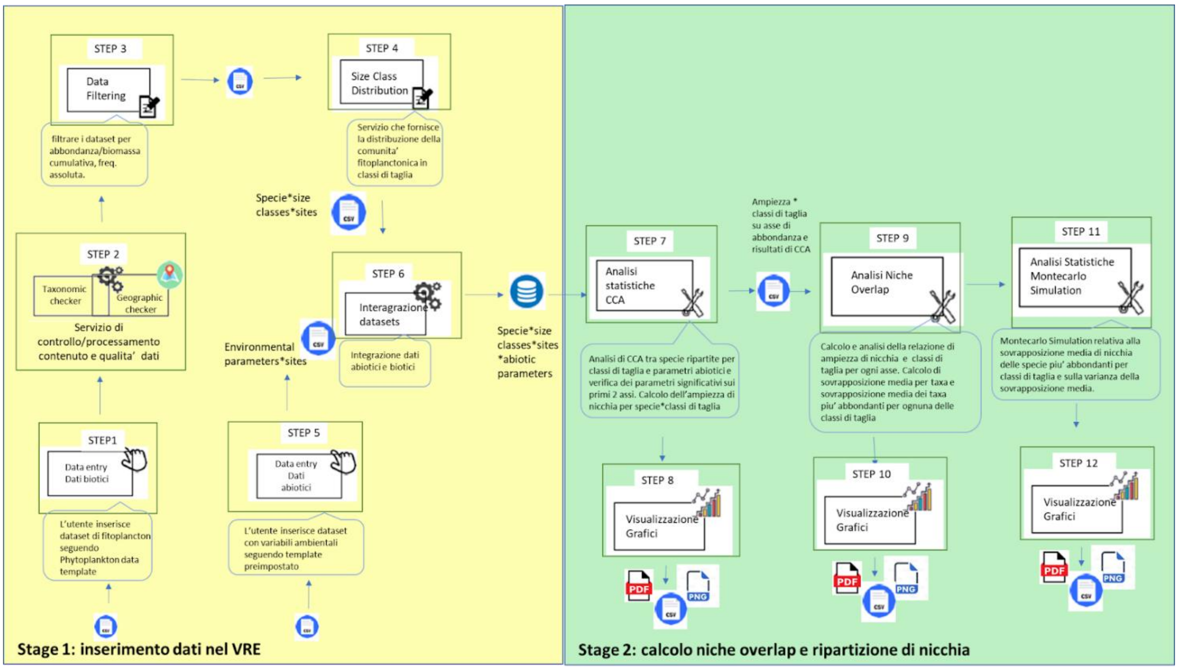


Figure 8. Example of Workflow 3 of the Phytoplankton Virtual Research Environment: Analysis of the mechanisms of interspecific coexistence in phytoplankton species and niche partitioning mediated by body size, nutrients, and temperature.

2.4. COMPOSITION AND FUNCTIONAL TRAITS OF FRESHWATER PHYTOPLANKTON FROM UMBRIAN LAKES

Data acquisition:

The dataset includes:

- Phytoplankton Data Table: This includes species density and biomass values collected between 2009 and 2022 with bi-monthly sampling. The data covers 692 species across 9 phyla, 22 classes, 64 orders, and 155 families.

- Functional Trait Table: A table currently under development, which includes functional groups (Reynolds 2002, Padisak 2009), morpho-functional classification (Kruk 2010), and traits categorical information (mostly binary values: 0-1) related to morphology (shape, presence of external structures), behavior (motility), physiology (trophic levels), and reproductive traits for all 692 species.

- Abiotic Data Table: This table contains physico-chemical environmental parameters, such as alkalinity, Total Nitrogen, Total Phosphorus, Dissolved Organic Carbon, Conductivity, , Dissolved Oxygen, pH, Depth, Silica, Temperature, Transparency.

Data have been collected from 9 water bodies across the Umbria region, during the monitoring phase carried out by ARPA UMBRIA which includes 8 lakes (2 natural – Trasimeno and Colfiorito) and 6 heavily modified water bodies (HMWBs) (Figure 9).

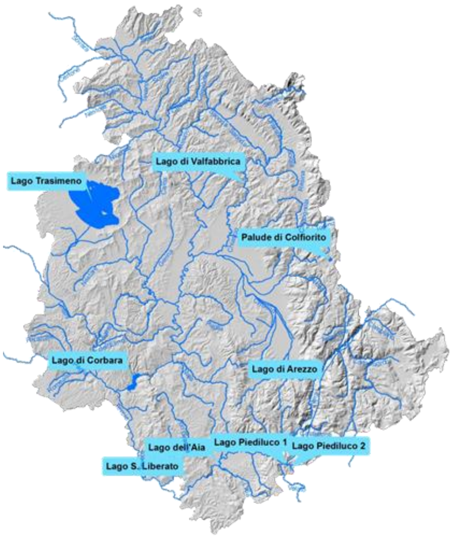


Figure 9. The nine water bodies across the Umbria region where sampling was carried out.

Data Standardization & Validation:

The data curation and standardization of all data tables is still ongoing. Taxonomic data validation has been conducted using AlgaeBase and abiotic parameters have been standardized and semantically annotated using NERC vocabularies. All datasets have been formatted as CSV tables. The metadata still need to be described according to the Ecological Metadata Language 2.0, before publication on the LifeWatch Italy data portal and the trait table still need to be completed with information mostly collected from literature and harmonized according to the Traits Thesaurus.

Data analysis :

This phase is still ongoing. A preliminary evaluation of the spatio-temporal variability in phytoplankton composition, density and biomass at the taxonomic level has been conducted (Figure 10), with diversity indices calculated using the “Diversity Indices” web service from the Phytoplankton Virtual Research Environment (Phyto VRE).

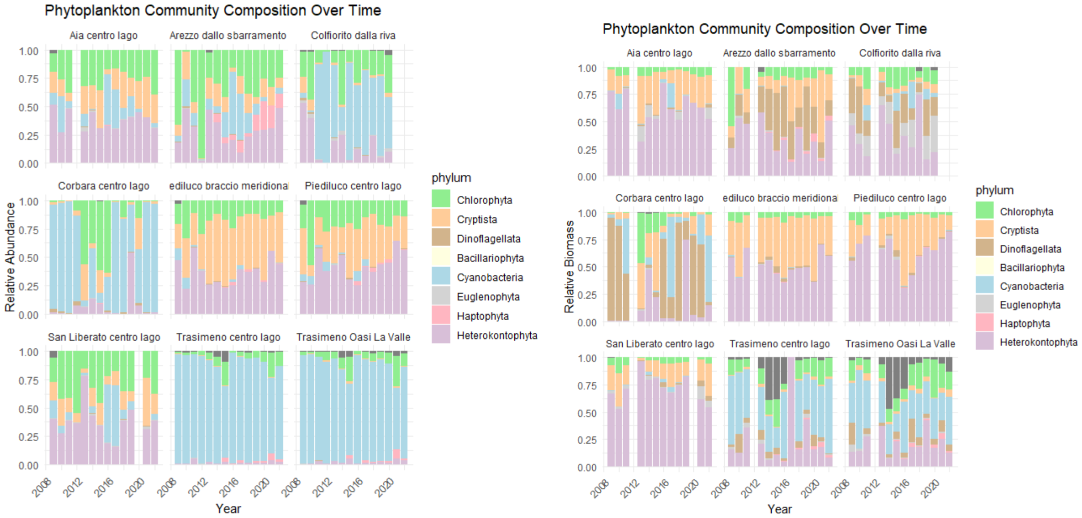


Figure 10. Phytoplankton density and biomass composition in the nine water bodies across the Umbria region, sampled between 2008-2022.

The trait-based analysis is yet to be applied but will be combined with the Phyto VRE for investigating changes in functional diversity and phytoplankton functional responses across broader spatial and temporal scales. This dataset could also be integrated with remote sensing data to enhance analyses and modeling, enabling the prediction of future scenarios and a deeper understanding of biodiversity organization in aquatic ecosystems under changing conditions. The aim for this dataset is to apply a trait-based framework to assess functional diversity patterns and assembly rules under a taxonomic and trait based perspective and utilize this dataset as case study for validating the workflow 2 and 3 of the Phyto VRE for large-scale data standardization, integration and analysis.

2.5. COMPOSITION AND MORPHO-FUNCTIONAL TRAITS OF ZOOPLANKTON SWIMMERS AND SINKERS COLLECTED BY "MOORING DIRIGIBILE ITALIA" SEDIMENT TRAP.

Data acquisition:

Since 2010, oceanographic moorings have been collecting zooplankton and sinking particle samples using sediment traps from the Mooring Dirigibile Italia (MDI) in Kongsfjorden, offering valuable insights into biodiversity dynamics in the Arctic Ocean. Initially, these data were not FAIR (Findable, Accessible, Interoperable, Reusable) or openly accessible, so the OU team, through the KAIROS project, has worked to make them compliant with FAIR standards. This process involved a comprehensive re-examination of zooplankton samples collected during the 2016-2017 and 2022-2023 periods. To do so, a comprehensive re-examination of zooplankton samples was conducted. This involved several key activities, starting with the stereomicroscopic re-evaluation of specimens to ensure precise identification of taxa. Over 3,000 specimens were photographed and subjected to morphometric analysis, while a detailed trait-based dataset was created, documenting size structure, life-stage, and sex where applicable. From these data an individual-level morphological trait dataset has been created, where morpho-functional traits were recorded for various zooplankton groups, enabling ecological function analysis. Here in detail the traits collected at individual level for each taxonomic group:

taxonomic group	***morphological trait***	***life-stage***	***sex***
Cnidaria	bell length, bell width		
Chaetognatha	body length, body width		
Annelida	body length, body width	larvae, adult	
Amphipoda	total length	juvenile, adult	
Cirripedia	body length, body width (naupliar stage), carapace length, carapace width (cypris stage)	naupliar stage, cypris stage	
Copepoda	prosome length, prosome width, urosome length, total length, individual biovolume	naupliar stage, copepodite stage, adult stage	female, male
Cumacea	body length		

Euphausiacea	body length	zoea, adult	
Ostracoda	body length, body width		
Bivalvia	shell length, shell width	veliger	
Pteropoda	shell width (line of aperture), shell diameter, n of whorls		
Appendicularia	trunk length, trunk width, tail length, tail width		

These morphometric traits records will provide critical insights into zooplankton size structure, ecological roles, and their responses to environmental changes. By analyzing these traits, researchers can better understand biodiversity shifts in response to climate change and gain a deeper understanding of Arctic zooplankton community dynamics.

Data Standardization & Validation:

The re-examination process involved classifying the zooplankton community into a number of taxonomic groups, with species and genus level identifications made wherever possible. This step was essential to validate the accuracy of previous taxonomic categorisations and to resolve any misidentifications or inconsistencies found in the original dataset. To ensure the data met international standards, a thorough process of data harmonization was carried out using R scripts. This included standardizing raw data formats to comply with the requirements of EMODnet Biology, validating the taxonomy with the WoRMS database, and integrating environmental parameters such as temperature and salinity extracted from the IADC ERDDAP repository. The data were then structured into three main tables—event, occurrence, and eMOF—and the R scripts were published on a restricted-access GitHub repository to ensure transparency and reproducibility. Taxonomic data validation continues using LifeWatch Species Information Backbone (LW-SIBb). For the trait data, they will be harmonized according to LifeWatch requirements and using the Traits Thesaurus.

Data analysis :

This phase has still not carried out, since the development of the dataset is still undergoing. This detailed work included in this dataset will improve the robustness of the trait-based information from

Arctic species and increase our knowledge of zooplankton size-structure composition in the Arctic Ocean and allow for more comprehensive comparisons. With the publication of the KAIROS dataset, the aim is to better understand how environmental conditions and climate change anomalies will impact the zooplankton biodiversity, community composition, and size structure over time in the Arctic. To support this, additional statistical modeling and multivariate analyses will be employed to assess the relationships between biodiversity composition (from sediment trap samples), abundance, and morphometric traits under future climate change projections.

2.6. GLOBAL DATASET ON VERTEBRATE SPACE USE AND LIFE-HISTORY TRAITS

Here below details of the different phases of the analytical framework and procedures applied to this dataset are reported:

Data acquisition:

We first assembled a dataset on behavioral and life-traits of a representative selection of aquatic and terrestrial vertebrate species, from the published literature. In particular, for each species, we compiled the dataset with the following key variables: i) home range; ii) body size and iii) maximum longevity. Data on home range estimates and body size were gathered from three different available resources: Tamburello et al. (2015); McCauley et al. (2015) and Udyawer et al. (2023). We ammended to this dataset, data on longevity/lifespan as the maximum recorded age at species level (hereafter “ML”), by relying on the AnAge dataset: a database of vertebrate longevity records and their relation to other life-history traits (De Magalhaes and Costa, 2009).

Data integration:

These datasets on home range and body size used multiple estimates available from different published studies from literature and provided averaged values per species. At the same time, as these datasets contained a different number of species, sometimes expressed even at the genus or subspecies level, a long process of filtering, checking and cleaning the datasets was carried out. We then proceeded by combining data from these different original datasets to provide a unified version in a single dataset. Data for individual species were averaged across datasets to obtain single home range (expressed in m^2) and body size (expressed in g) values for each species. If initially expressed in another unit, these variables were first converted for all datasets. More specifically, the dataset from Tamburello et al. (2015) already provided single species records, thus there was no need to average values. However, as in this dataset we found a case of a record "*Carphophis*" that had been ranked up to the level of Genus only, we decided to remove it. Similarly, a record that was reported at the subspecies level "*Coluber constrictor flaviventris*" was not considered. With regard to the dataset from McCauley et al. (2015), it included multiple records for several species (n=9). We averaged these multiple values for that species, in order to have just one record per species. For double species' entries present both in the Tamburello's and McCauley's datasets, we computed the mean of home range and body size between these two datasets, thus maintaining a single entry per species. Data were averaged in 308 cases (i.e. those species records in common between these two datsets). In the case of the dataset from Udyawer et al. (2023) that included multiple individual records for several species (n = 1596 records, for 79 species), values at individual-level were averaged in order to obtain just one record per species. From this latter dataset, we eliminated individual records from seven species (*Caranx ignobilis*, *Carcharhinus amblyrhynchos*, *Carcharhinus leucas*, *Carcharhinus melanopterus*, *Chelonia mydas*, *Plectropomus leopardus*, *Scarus rivulatus*), for which estimates were already available/present at species level in the two previous datasets: Tamburello et al. (2015) and/or McCauley et al. (2015).

From the AnaAge dataset (De Magalhaes and Costa, 2009), we initially gathered ML estimates for 803 species, for which home range and body sizes were available. We then relied on other external resources (i.e. ADW Animal Diversity Web: https://animaldiversity.org/; World Species: https://worldspecies.org/; FishBase: https://www.fishbase.se/), to integrate the dataset for this variable. That way, we added other 140 records, totalling n = 943 ML species records, representing 81% of the entire species list of the compiled dataset.

After this step, we proceeded by adding, for each species, the following four behavioural and life-history traits: realm (aquatic *vs* terrestrial), thermoregulation system (endotherm *vs* ecthoterm), locomotion type (crawling, flying, walking, swimming), primary diet (carnivore *vs* herbivore). For the latter category, omnivores have been included as carnivores or herbivores, based on their prevalent diet, following trophic classification databases or published literature (e.g. Tamburello et al., 2015; McCauley et al., 2015).

Data standardization and validation:

This step aimed at providing values estimates with the same unit of measurement and eventually correct/amend errors or double entries, as likely when working with multiple heterogeneous datasets.

For example, home range estimates expressed in km^2 were converted in m^2 for all datasets as well as body mass expressed in kg was converted in grams for all datasets.

Before being assembled, for each of these original datasets, we carried out a taxonomic identification via the GBIF (Global Biodiversity Information Facility - https://www.gbif.org/) "*species name matching*" validation tool, in order to avoid some misspelled names and unaccepted synonyms. Species names were thus corrected, as per the most recent internationally accepted nomenclature.

The resulting dataset was thus composed by 1164 records at species level, spanning fish (*n = 191*, including Elasmobranchs), reptiles (*n = 136*), mammals (*n = 642*) and birds (*n = 195*), from around the globe. More specifically, 191 fish species (including Elasmobranchs) ranging in size from -0.65 (log10 g; of the Rusty goby *Priolepis hipoliti*) to 6.08 (log10 g; of the Basking shark *Rhincodon typus*), with home range sizes ranging from -1.52 (log10 m2; of the Rusty goby) to 4.27 (log10 m2; of the Great white shark *Carcharodon carcharias*) and with maximum longevity spanning from 0.47 (log10 year; of the Bluehead *Thalassoma bifasciatum*) to 1.95 (log10 year; of the Quillback rockfish *Sebastes maliger*). Then 136 reptiles species ranging in size from 0.1 (log10 g; of the Clouded anole *Anolis nebulosus*) to 5.61 (log10 g; of the Leatherback sea turtle *Dermochelys coriacea*), with home range sizes ranging from 0.1 (log10 m2; of the Jamaican gray anole *Anolis lineatopus*) to 9.88 (log10 m2; of the Leatherback sea turtle) and with maximum longevity spanning from 0.1 (log10 year; of the the Jamaican gray anole and the Pygmy short-horned lizard *Phrynosoma douglasii*) to 2.10 (log10 year; of the Hermann's tortoise *Testudo hermanni* and the Greek tortoise *Testudo graeca*). The 195 bird species ranged in size from 0.71 (log10 g; of the Goldcrest *Regulus regulus*) to 4.94 (log10 g; of the Common ostrich *Struthio camelus*); home range were from 3.022 (log10 m2; of the Indigo Bunting *Passerina cyanea*) to 12.41 (log10 m2; of the Southern rockhopper penguin *Eudyptes filholi*); maximum longevity were from 0.60 (log10 year; of the Dartford Warbler *Sylvia undata*) to 1.83 60 (log10 year; of the Common Raven *Corvus corax*). The 642 species of mammals ranged in size from 0.60 (log10 g; of the Eurasian pygmy shrew *Sorex minutus*) to 7.9 (log10 g; of the Bowhead whale *Balaena mysticetus*); home range from 1.27 (log10 m2; of the Talas tuco-tuco *Ctenomys talarum*) to 11.06 (log10 m2; of the Northern fur seal *Callorhinus ursinus*) and maximum longevity from -0.09 (log10 year; of the Hairy-tailed mole *Parascalops breweri*) to 2.32 (log10 year; of the Bowhead whale).

Preliminary dataset explorations focussed on scaling relationships between home range with body mass as well as on the relationships of home range with behavioural and life-history traits, such as habitat type, thermoregulation, trophic behaviour and locomotion mode (Figure 11). We repeated the same exercise to investigate the relationship between maximum longevity and home-range as well as their interactions with the same behavioural and life-history traits (Figure 12).

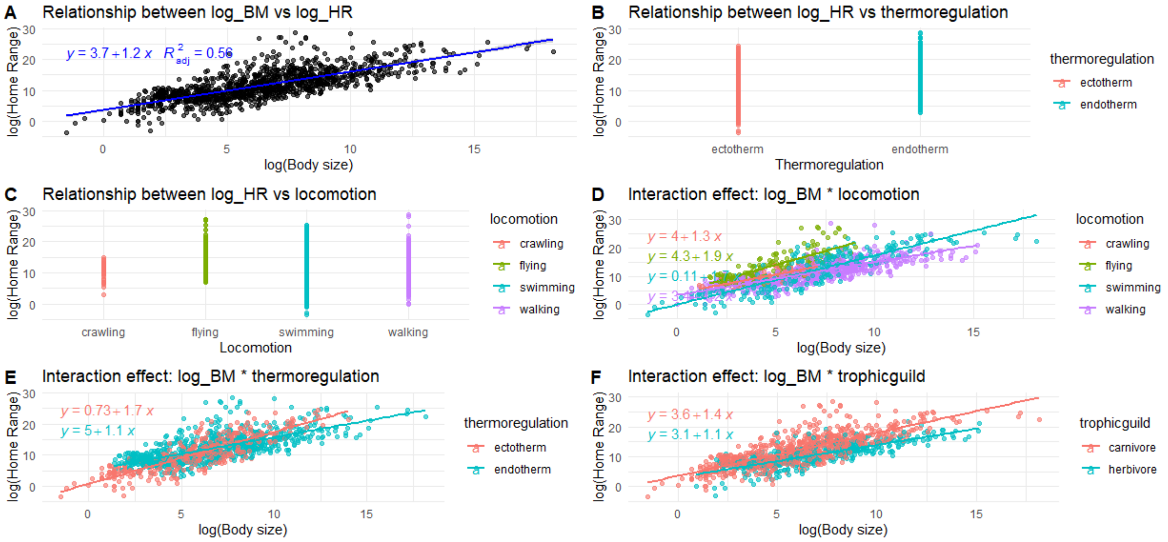


Figure 11: Combined plots showing home range variations in relation to body size, thermoregulation system and locomotion mode and factor interactions, across 1164 vertebrate taxa from around the globe.

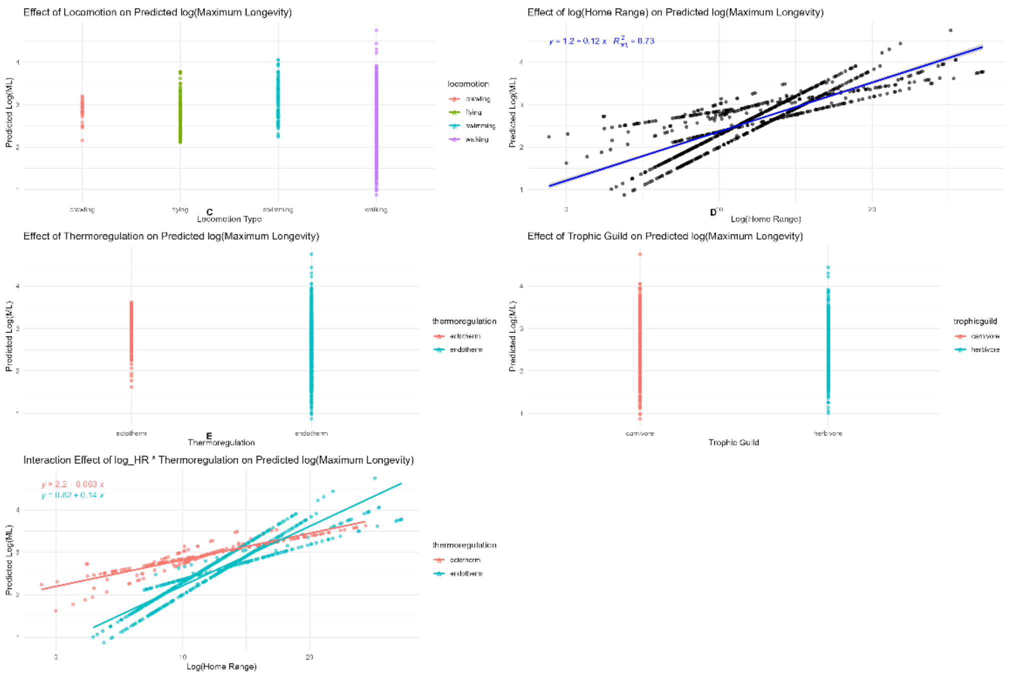


Figure 12: Combined plots showing maximum longevity variations in relation to locomotion mode, home range, thermoregulation system, trophic guild and factor interactions, across 1164 vertebrate taxa from around the globe.

The entire dataset, in a second step, will be further validated following the LifeWatch Eric web services and tools for data harmonization, exchange and analysis and long-term sustainability. All data will be harmonized and standardized according to a data schema and through the use of existing controlled vocabularies or created *ad hoc* semantic artefacts.

Data analysis:

Data analysis was divided in two different sections:

Section 1: For this analysis we run two model selections. First, to investigate the relationships between home range (response variable, transformed in log10) and body size (transformed in log10) while accounting for categorical ecological traits, we employed a linear modeling approach (Zuur et al. 2009; Zuur and Ieno, 2016). An initial full model was constructed, incorporating all potential combinations and interactions of predictors between body size and the categorical variables (realm, thermoregulation, locomotion, and trophic guild), also including the null model, because each combination could represent distinct a priori hypothesis. Secondly, we repeated the same approach to investigate the relationship between maximum longevity (response variable, transformed in log10) and home range (transformed in log10), while accounting for the same categorical ecological traits and interactions. In both cases, the model selection used Akaike’s Information Criterion corrected (AICc) for small sample size to identify the most parsimonious model, balancing model fit and complexity. Models were selected if they had ΔAICc ≤ 2, and if their AICc value was lower than that of any simpler alternative (Harrison et al. 2018). Therefore, we obtained either a set of top-ranked models or a single best model for each response variable. All statistical analyses were conducted in R (version 4.3.1). Inference about the effects of predictors was made on the best model, selected via a backward stepwise procedure through the *MuMIn* package for model selection.

Section 2: To investigate climate change impacts on species’ distribution range, we followed this approach. We made data request to the IUCN Red List to download species’ polygon shape files (if available). Then, part of the Aquatic Biomass VRE workflow was adapted and applied to analyse the variation in mean temperatures, as a proxy for global warming, across species' ranges and building future scenario of how these changes may evolve. To this aim, we used remotely sensed imagery derived from the Moderate Resolution Imaging Spectroradiometer MODIS (Christopher, 2007). For aquatic species, images were gathered from NASA's Ocean Color Sea Surface Temperature (SST), available on the EARTHDATA platform, and acquired from January 2003 to December 2024 with monthly resolution and 9 km spatial resolution (Ocean Color, 2024) (Figure 13). For terrestrial species, images of land surface temperature (LST) we will start the acquisition stage in the coming weeks/months from MODIS.

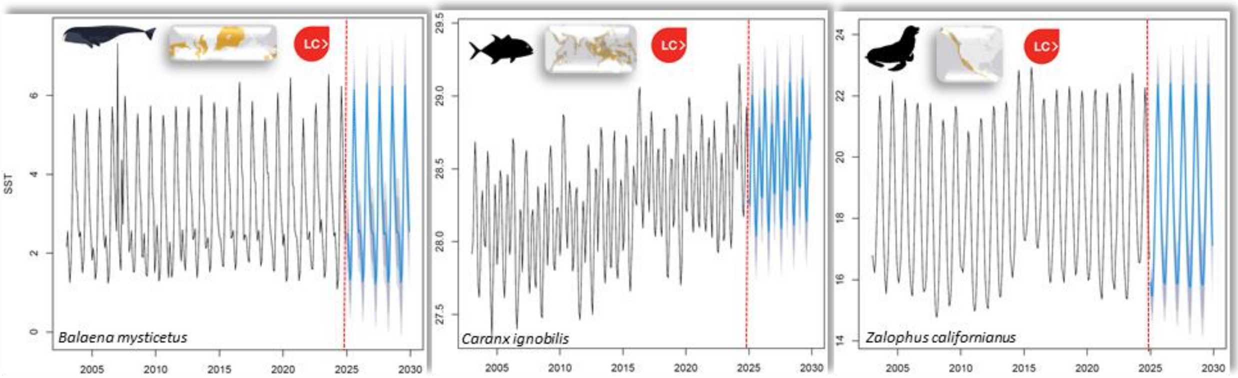


Figure 13: Output from the analytical framework applied to three marine species as an example. A temperature time series has been obtained for each species from 2003 to 2024 and projected into the future (up to five years, as a first step) to predict potential changes likely to occur in the species' range. In the boxes, for each species, is reported the map of the distributional range and its most updated IUCN conservation status.

This twice, not-mutually exclusive approach, facilitates comprehensive estimates of home range and maximum longevity relationships with both intrinsic and extrinsic factors and allows to understand the impact of climate change to biodiversity organization and conservation.

Once validated, both these procedures, currently run in the R software, will be converted into the software Python. A prototype representing part of the analysis processes will be developed on DataLabs by LifeWatch Italy, a virtual laboratory platform for collaborative coding and application management. This prototype included data acquisition activities to populate the models and perform the analysis processes into a dedicated VRE (see example below).

The resulting dataset from activity 6.17 will be available in the form of txt and/or .csv files. They will feed and be used for running the different workflows of the Biomass VRE.

For example, this dataset will be used for one of the workflows focusing on consumer biomass (Figure 14). This workflow focuses on the effects of thermal gradients on biomass at the consumer level, specifically examining metabolic rate and home-range parameters. The workflow allows the modeling of the responses of animal organisms to global warming in terms of metabolism and spatial use.

The workflow will be designed with a modular structure that integrates services for data discovery, harmonization, interoperability, analysis and modeling. These workflows will be incorporated into the ITINERIS Hub via LifeWatch Italy and will be accessible through APIs from the web portal.

The input data types for these analyses include text or tables in CSV or TXT format, raster data in HDF, Open NetCDF, or IMG formats, and vector data in SHP format. The outputs of the workflows are tables in CSV or TXT format, raster imagery of the newly derived products, and plot results, such as time series, in PDF, JPG, or PNG formats. The inputs and outputs can be adapted or modified during the development phase to accommodate specific needs, in collaboration with the scientific team at CNR-IRET-LE.

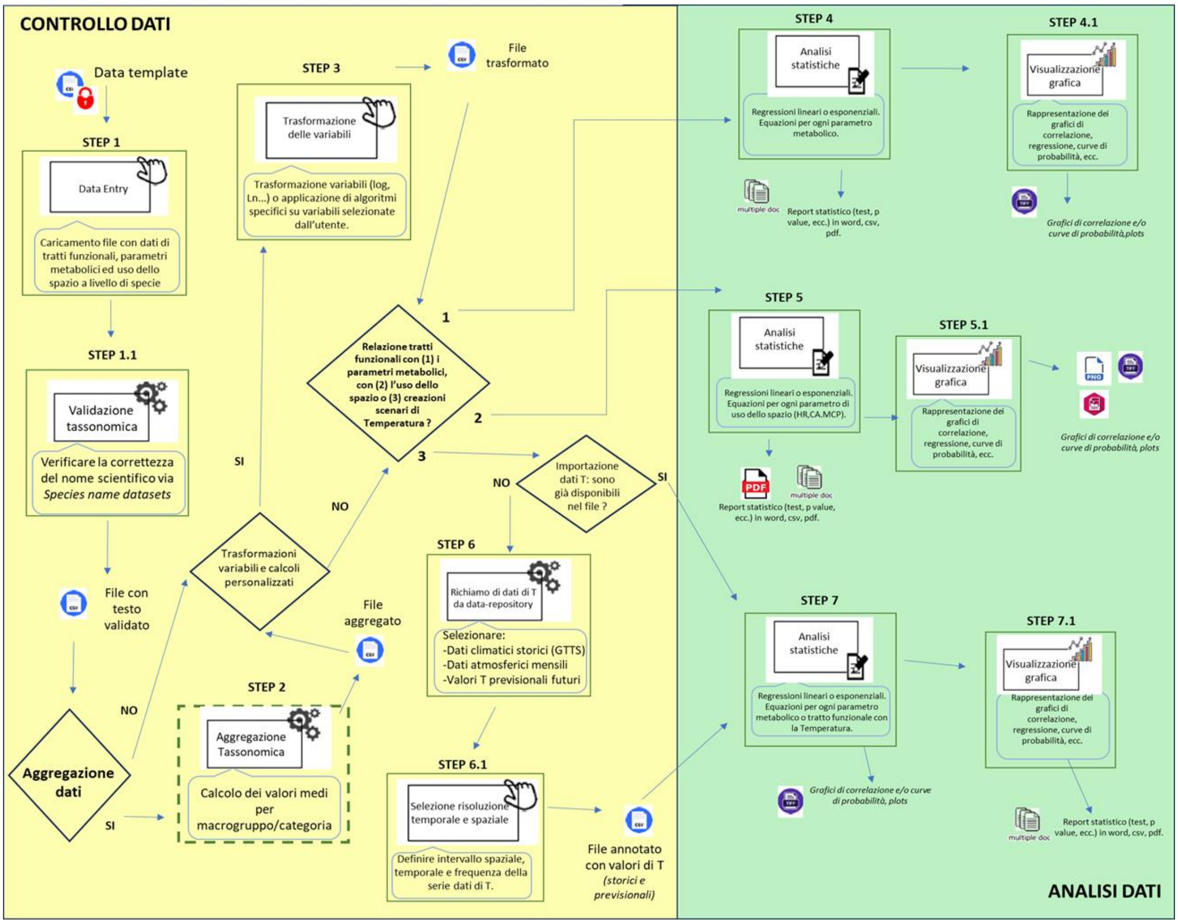


Figure 14. Example of representation of the workflow at the level of consumers of the Biomass VRE (part extrapolated from the Term of Reference).

3. LIST OF CONTRIBUTIONS

The rationale behind this work and steps carried out until now have been presented at various international and national conferences, as well as at internal project meetings or CNR conferences. Here below a summary of these contributions (Table S1):

Table S1: A list of the contributions presented both at international and national conferences and publications related to this deliverable.

✓ *Laraib M., Titocci J., Rosati I., Basset A. (2023). An integrated individual-level trait-based phytoplankton dataset from transitional waters. Sci Data 10, 897.* *Doi: https://doi.org/10.1038/s41597-023-02785-w*
✓ *Lairab M., Titocci J., Giannakouro A, Reizopoulou S, Basset A. (2024). Role of Rare Species on Phytoplankton Size–Abundance Relationships and Size Structure across Different Biogeographical Areas. Diversity 16, no. 2: 98.* *Doi: https://doi.org/10.3390/d16020098*
✓ *Monti F., Shokri M., Semeraro T., Titocci J., Liberatore L., Basset A. (2024). Enhancing biodiversity conservation analysis through data harmonization. Contributed session: Biodiversity XIX. ECCB – 7th European Congress of Conservation Biology. 17-21 June 2024, Bologna, Italy.*
✓ *Monti F., Titocci J., Liberatore L., Semeraro T., Basset A. (2024). Enhancing existing biodiversity knowledge through data FAIRness and e-Science tools. ITINERIS - 2nd General project Meeting - 09-10 July 2024, Rome, Italy.*
✓ *Scardia Scardia A., Titocci J., Liberatore L., Basset A. (2024). Decoding phytoplankton coexistence mechanisms through a meta-analysis across 24 transitional water ecosystems. Poster presentation, SItE, 23-26 Settembre, 2024, Roma*
✓ *Semeraro T., Titocci J., Liberatore L., Monti F., De Leo F., Ingrosso G., Shokri M., Basset A. (2024). Phytoplankton production response to global warming: estimating patterns of change across latitudes. Mediterranean Life Sciences Union Annual Meeting (MedLIFE-24). 08-10 December 2024, Istanbul, Turchia*

- ✓ *Semeraro T., Titocci J., Monti F., Liberatore L., Basset A. (2024). Risposta di biomassa e produzione fitoplanctonica al riscaldamento globale: stima delle variazioni nella biomassa stabile in ampie fasce latitudinali. ITINERIS - 2nd General project Meeting - 09-10 July 2024, Rome, Italy.*
- ✓ *Titocci J, LaMarra M., Cozzoli F., Semeraro T., Monti F., Liberatore L., Vaira L., Turrisi G., Fiore N., Rosati I., Basset A. (2024). Monitoring aquatic primary producers response to Climate Change: The Phytoplankton VRE, Oral presentation at the LifeWatch Thematic Service Workshop – Climate Change Impact on Biodiversity Patterns, Lecce, Italy, 21-22 February 2024*
- ✓ *Titocci J. et al. (2024). Sharing and reusing knowledge in trait-based research through semantic resources: The Functional Trait Thesauri. Oral presentation at the 7th ICES/PICES Zooplankton Production Symposium, 17-22 March 2024, Hobart, Tasmania, Australia*
- ✓ *Liberatore L., Titocci J., Semeraro T., Monti F., Basset A. (2025). Modelling transitional water phytoplankton communities by functional traits in climate change scenarios: an ecological niche-based approach. CNR IRET Annual Conference, 18-19/02/2025, Rome, Italy.*
- ✓ *Monti F., Semeraro T., Titocci J., Liberatore L., Basset A. (2025). Integrating species traits and environmental dynamics to predict spatial behaviour and climate change impacts on vertebrates. CNR IRET Annual Conference, 18-19/02/2025, Rome, Italy.*
- ✓ *Semeraro T., Titocci J., Liberatore L., Monti F., De Leo F., Ingrosso G., Shokri M., Basset A. (2025). Analytical Workflow to Study Ocean Production Response to Global Warming. CNR IRET Annual Conference, 18-19/02/2025, Rome, Italy*
- ✓ *Titocci J. et al. (2025). (in press) Pathways for converting zooplankton traits to ecological insight are paved with findable, accessible, interoperable, and reusable (FAIR) data and practices. ICES Journal of Marine Science. Doi: 10.1093/icesjms/fsaf017*
- ✓ *Titocci J. et al. (2025). e-Science tools for trait-based approaches in plankton ecology. Oral presentation, LifeWatch Italy Conference, Rome, January 29th-30th*
- ✓ *Titocci J., Semeraro T., Liberatore L., Monti F., Perrone L., Scardia Scardia A., Basset A. (2025). Investigating diversity distribution and assembly mechanisms of planktonic*

communities through a trait-based and e-Science approach. Abstract submitted for the CNR IRET Annual Conference, 18-19/02/2025, Rome, Italy.