MISSIONE 4
ISTRUZIONE
RICERCA

ITINERIS

Deliverable D6.8 – Activity 6.13, CNR IRET LE

Report and guidelines on the new developments of the LIFEWATCH-Italy Virtual Research Environments

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

Summary

Finanziato dall'Unione europea
NextGenerationEU

ITINERIS

Deliverable number:	D6.8
Work package:	WP6
Intermediate Objective:	
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:	31/12/2024
Author(s) (Partner-OU):	Lorenzo Liberatore, Flavio Monti, Teodoro Semeraro, Jessica Titocci (CNR IRETLE), Mariantonietta La Marra (UNISAL), Vanessa Marrocco (LifeWatch ERIC)
Reviewed by:	ITINERIS Executive Board
Note:	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)

Funded by EU - Next Generation EU

Mission 4 “Education and Research” - Component 2: “From research to business” -

Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

1. OVERVIEW OF THE ACTIVITY 6.13 – WP6

The activity 6.13 is part of WP 6 (Terrestrial Biosphere) of the Italian Integrated Environmental Research Infrastructures System (ITINERIS). ITINERIS is a project funded by the EU - Next Generation EU PNRR - Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures".

The action to strengthen the research infrastructure of LIFEWATCH Italy includes two main aspects:

1) Strengthening and updating the virtual research environments of the infrastructure and progressing the alignment with the LIFEWATCH ERIC VRE compositional structure;

2) Strengthening the commitment of the international scientific community to LIFEWATCH through pilot actions to open data production facilities to remote users; the proposed facilities include an advanced aquarium system for ecological experiments on freshwater and transitional water populations/communities, extendable to marine ones, and a top-level inverted confocal system.

Among the new VRE developments, the most important basic requirements are: a) a user-friendly interface accessible through web technologies; b) the possibility for the user to use existing and validated web services and modelling software; c) the interaction with the software interface of all the platforms, catalogues and other services already present in the LIFEWATCH Italy web portal; d) the interaction of the user with the VRE user experience based up to an advanced system where the skilled user will be able to generate, through an orchestrator, his own scientific workflows linking different digital objects, from data to advanced semantic services and modelling tools.

The present document illustrates the state of development of the final deliverable D6.08 "Report and guidelines on the new developments of the LIFEWATCH-Italy Virtual Research Environments" belonging to the Activity 6.13 and to be released on B14.

2. STATE OF THE ART OF PHYTO VRE

Virtual Research Environments are innovative, web-based, community-oriented, comprehensive, flexible, and secure digital working environments conceived to serve the needs of modern science. Virtual Research Environments provide user-centric support for discovering and selecting data and software services from different sources, and composing and executing application workflows. They play a direct role in the life cycle of research activities performed by scientists, such as experiments planning, search and discovery of resources from different sources (notably including RIs), integration of services into cohesive workflows and collaboration with other scientists.

LifeWatch Italy, the Italian node of LifeWatch ERIC, has designed the Phytoplankton Virtual Research Environment, a collaborative working environment supporting researchers to address basic and applied studies on phytoplankton ecology (https://www.phytovre.lifewatchitaly.eu/). The Phytoplankton VRE offers resources, tools and web services to create or discover, access and share harmonized datasets, with the associated metadata, on phytoplankton taxonomy and morphological traits. Moreover, it provides computational services that facilitate phytoplankton data analysis using automated procedures and algorithms.

Specifically, currently the Phytoplankton VRE comprises two atlases, a thesaurus on phytoplankton traits and six web services to perform calculations and analyses on phytoplankton data. They are currently available online and the web services are based on a client-server architecture. The scripts on the server side are written in the R programming language (https:/www.r-project.org.) and allow automating a set of calculations and analyses. Each web service runs independently from the others. The user is prompted to follow some operational steps and choose the input parameters and the services return an output as a .zip file.

Below are the currently existing services of Phytoplankton VRE.

2.1. Phytoplankton Data Template

The Phytoplankton VRE users can work both with their own phytoplankton datasets or with datasets provided from the LifeWatch Italy Data Portal (https://dataportal.lifewatchitaly.eu/data). For the correct use and interoperability of the data, it is important to have homogeneous formats and terminologies. The structure of the dataset and the variables names have to be consistent, independently from who collected the data. Therefore, the Phytoplankton VRE provides and uses a *Data Template*, which is structured as a column-oriented table in CSV UTF-8 format, an open and non-proprietary format. The header row contains the fields related to ancillary information (e.g. time and location of the sampling) and phytoplankton features (e.g. taxonomical classification, linear dimensions, etc.) based on the Darwin Core standard and the phytoplankton terminologies derived from the *Phytoplankton Traits Thesaurus*. In this way, the records are syntactically and semantically interoperable and can be interpreted both by researchers and machines. The data are described by accurate metadata to ensure long-term usability.
Using this link: https://www.phytovre.lifewatchitaly.eu/phyto-data-template/ it is possible to navigate through the template and read the detailed description for each field of the Phytoplankton Data Template. Currently, the template has been designed to harmonise primarily abundance/biomass and morphological trait data (size and shape) at the individual level, but it can easily be used for data aggregated to species or higher levels.

2.2. Atlas of Phytoplankton and Atlas of Shapes

Changes in the nomenclature and categorisation of species are not uncommon, and new phytoplankton organisms are being described all the time. Therefore, the use of standardised and internationally accepted rules for taxonomic classification is essential. The Atlas of Phytoplankton (https://www.phytovre.lifewatchitaly.eu/vre/phyto-list/) provides a reference guide for the identification of marine, transitional and freshwater species. It includes species details, images and schematic drawings; information on taxonomy, synonyms, morphological and ecological characteristics and geographical distribution of species. It also provides formulae for calculating phytoplankton biomass in terms of biovolume, carbon content and surface area, based on the linear dimensions of phytoplankton cells (Figure 1).

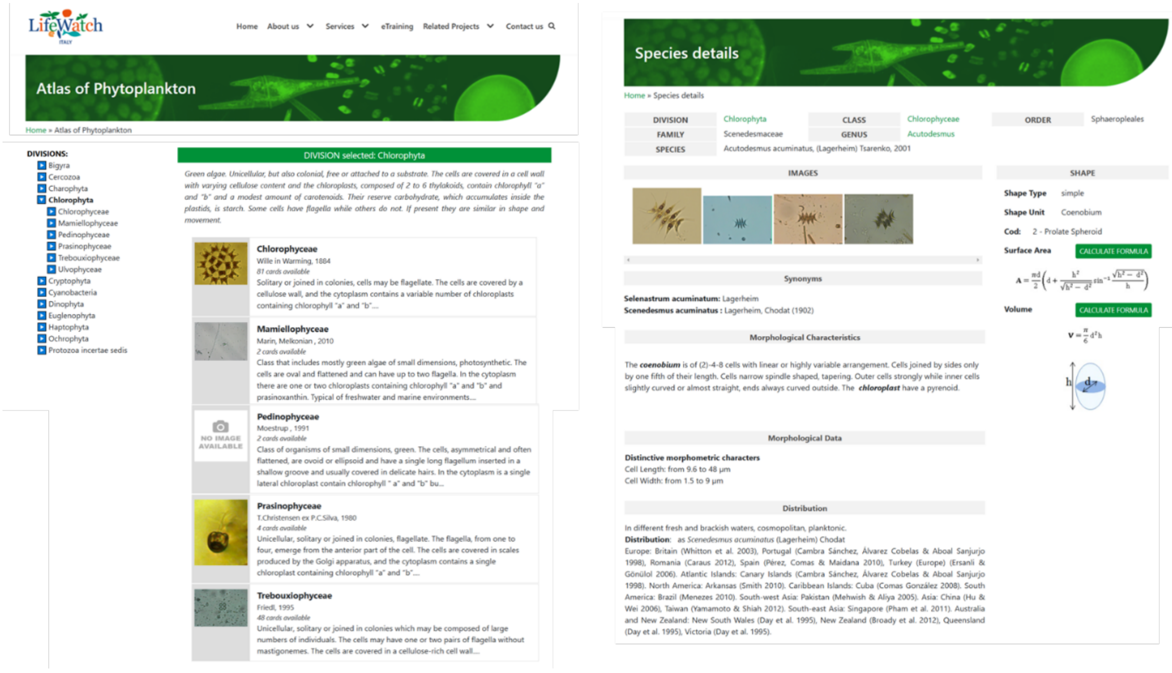


Figure 1: Atlas of Phytoplankton

An accurate calculation of phytoplankton biovolume is important for assessing the ecological status of water bodies. There are several methods for calculating biovolume. The most accurate method requires identification of the geometric shape that best approximates the actual shape of the cell, measurement of the linear dimensions and the correct formula for that particular geometric shape. It is important to have consistency between research groups in the geometric shape associated with each species and the associated formula. The Atlas of Shapes (https://www.phytovre.lifewatchitaly.eu/vre/shapes-groups/) contains schematic drawings of geometric models for the cell shapes, divided into 'simple shapes' and 'complex shapes', showing the relevant linear dimensions and cell shape views (Figure 2). The formulae for biovolume and surface area are given for each shape. The *Atlas of Shapes* and the *Atlas of Phytoplankton* are integrated and can be easily joint switching from taxonomic identification to morphological characterization of phytoplankton.

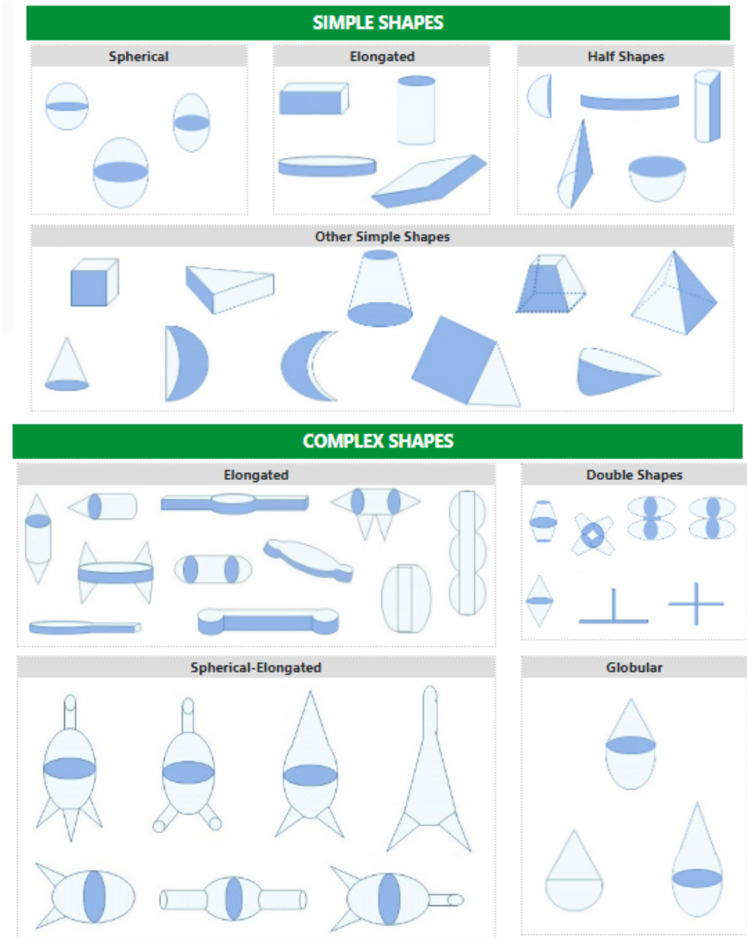


Figure 2: Atlas of Shapes.

2.3. Phytoplankton Traits Thesaurus

The Phytoplankton Traits Thesaurus (https://ecoportal.lifewatch.eu/ontologies/PHYTOTRAITS) provides unambiguous and standardised definitions and semantic properties of morpho-functional traits, resulting from the interdisciplinary collaboration of a scientific expert community within LifeWatch Italy. The thesaurus has been developed to address data harmonisation, integration and discovery, and to improve the interoperability and effectiveness of automated phytoplankton data exchange between different sources. The Phytoplankton Traits Thesaurus follows Semantic Web methodologies implemented in the Simple Knowledge Organisation System (SKOS), a common data model based on the Resource Description Framework (RDF) for encoding semantic relationships between data items and enabling computers to interpret these relationships. The thesaurus contains approximately 120 properties with defined data type properties (e.g. skos:prefLabel, skos:altLabel, skos:definition) for defining literal values as attributes (or qualities) of concepts, and object properties for defining hierarchical or associative relationships between concepts (e.g. skos:broader, skos:narrower, skos:related).

2.4. Traits Computation

The Traits Computation web service calculates demographic and morphological traits, in particular biovolume (µm³), surface area (µm²), surface/volume ratio, density (cells/L), cell carbon content (pg), total biovolume (µm³/L), total carbon content (pg/L). It automatically associates each taxa of the input dataset with the most similar geometric shape selected from the models proposed in the Atlas of Shapes. This service allows the calculation of biovolume, surface area, surface/volume ratio, density, cell carbon content, total biovolume, total carbon content, offering two types of calculation: simplified or advanced. The simplified calculation provides an approximate estimate of biovolume and surface area based on only two linear dimensions (length and width). The advanced calculation provides a more accurate estimate of biovolume and surface area by using more detailed measurements for the linear dimensions according to the shape and sedimentation view of the organism (Figure 3). The output of the service is a .csv file containing all the input data and the new calculated traits. The trait computation can be used following this link: https://www.phytovre.lifewatchitaly.eu/traits-computation/.

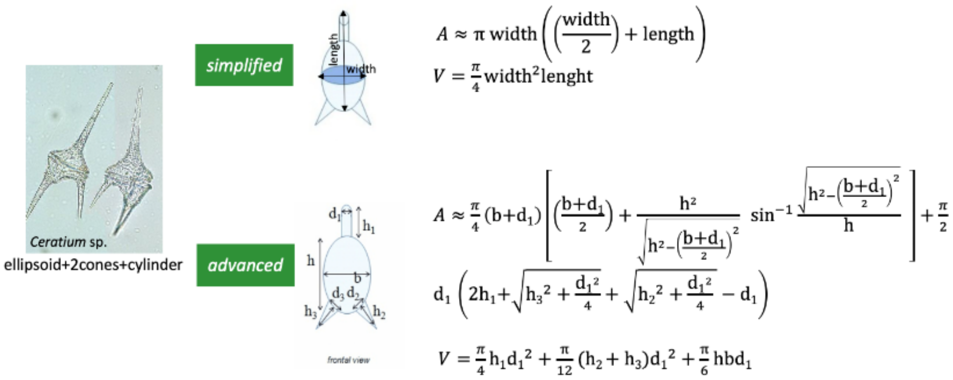


Figure 3: Different formulas to calculate the trait "biovolume".

2.5. Size Class Distribution

The Size Class Distribution web service (https://www.phytovre.lifewatchitaly.eu/size-class-distribution/) provides the distribution of phytoplankton cells into different size classes, based on logarithmic values of biovolume or cell carbon content, at different levels of data aggregation (spatial and temporal). Size classes can be calculated using a log base of 2, e or 10. As output, the service generates a .csv file with a summary table and a .pdf file with one or more bar charts, according to the selected spatial and temporal clusters.

2.6 Size Density Relationship

The Size Density Relationship web service (https://www.phytovre.lifewatchitaly.eu/size-density-relationships/) describes the relationships between phytoplankton cell size, based on calculated mean biovolume or carbon content, and phytoplankton density at different levels of data aggregation (spatial, temporal and taxonomic). It calculates and visualises a linear distribution model. As output, the service generates a .csv file containing a summary table with the values of density, average biovolume or average cell carbon content according to the selected taxonomic level and cluster, a summary table in .csv format with the result of the distribution linear model and a .pdf file showing the scatter plot with the trend power law ($y = k X\alpha$) and upper quantile bound regression (quantile 0.95).

2.7 Data Filtering

The Data Filtering web service (https://www.phytovre.lifewatchitaly.eu/data-filtering/) allows the removal of taxa from the dataset by selecting a threshold between 0 and 1 according to their cumulative contribution to total density, total biovolume or total carbon content. Filtering can be performed at any combination of spatial, temporal and taxonomic levels. The output of the service is a .csv file based on the original input dataset with the filtered taxa removed.

2.8 Community Indices

The Community Indices web service (https://www.phytovre.lifewatchitaly.eu/community-indices/) allows the calculation of phytoplankton community and diversity indices for a given combination of taxonomic, spatial and temporal levels. The calculation of the indices is based on feature density. The user can select one or more of the following indices: taxonomic richness, diversity (Shannon, Shannon equivalents, Simpson, Simpson equivalents, Menhinick, Margalef, Gleason, McInthosh, Hurlbert), evenness (Pielou, Sheldon, Ludwig-Reynolds) and dominance (Berger-Parker, McNaughton, Hulburt). A .csv file containing a summary table with the values of the calculated community indices and a .pdf file with one or more bar charts showing the graphical output of the calculation are provided.

2.9 Community Matrix

The Community Matrix web service (https://www.phytovre.lifewatchitaly.eu/community-matrix-and-analysis-lv/) generates a community matrix from the input data at the selected combination of taxonomic, spatial and temporal levels. The matrix can be calculated on the basis of either trait density, total biovolume or total carbon content. Only if the matrix is calculated on the basis of density, the user can choose to perform additional analyses on it: rarefaction curves (number of taxa vs. sample size), beta diversity calculation, distance matrix calculation (cross-site matrix of distances in community compositions), hierarchical cluster analysis, nonmetric multidimensional scaling (NMDS).

The Community Indices, Community Matrix and Data Filtering web services still need to be tested and validated to ensure full functionality.

3. OPTIMISATION OF THE PHYTOPLANKTON VRE: GUIDELINES OF THE NEW DEVELOPMENTS

The ITINERIS project aims to optimise the Phytoplankton VRE by leveraging its existing resources and web services, and to create a stronger link with the other services offered by the LifeWatch ERIC e-Science research infrastructure. Specifically, the Phytoplankton VRE will interface with the LifeWatch Data Portal and Metadata Catalogue to facilitate the acquisition and harvesting of data resources, and will be able to integrate with the EcoPortal platform, a repository of biodiversity and ecological semantic resources, to ensure data harmonisation and semantic interoperability. In addition, the VRE will connect to Data Labs, LifeWatch's collaborative coding platform, to provide open access to scripts and projects, and enable the deployment of web services.

The Phytoplankton VRE will also be improved by changing its structure. Whereas the current Phyto VRE represents a framework of different and separate web services, the new development of the Phytoplankton VRE will include a modular structure where all data analysis and modelling services will be linked and automated as workflows in the cloud. To achieve this, the existing Tesseract and NaaVRE technologies developed within LifeWatch ERIC RI will be used and the Phytoplankton VRE will be linked to them. This new structure of the Phytoplankton VRE will be accessible not only through the Phyto VRE website, but also within the CNR reference portal for the ITINERIS project.

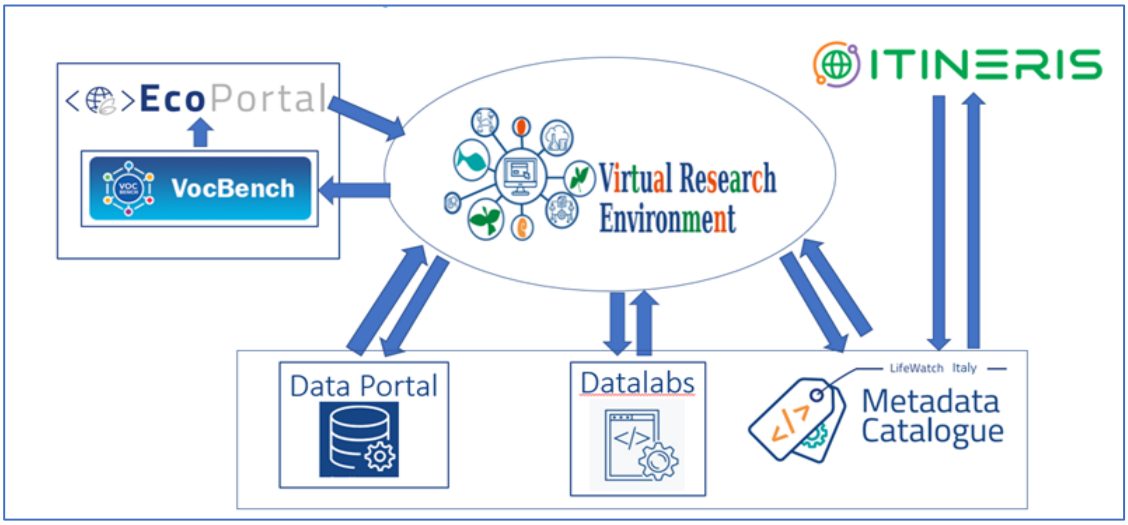
In addition, while the current version focuses primarily on phytoplankton organisms and their characteristics, the upcoming development of the Phytoplankton Virtual Research Environment will include all aspects of abiotic data associated with them. This will allow advanced functionalities such as modelling and analysis of the relationships between phytoplankton abundance, biomass and traits with key environmental parameters, including temperature, nutrient concentrations, pH and more, providing a holist framework for exploring the interactions between phytoplankton and their environment.

3.1. Functional and technical requirements

All proposed implementations were identified during the initial phase of activity 6.13 and focused on establishing the analytical framework for optimising and improving the phytoplankton VRE. This activity included the identification of essential data types and formats, analyses and models to be integrated into workflows and used for validation in case studies. A comprehensive review was carried out and subsequently a Terms of Reference document was produced to guide the development of the Phytoplankton VRE, as an external contractor will carry out its full implementation.

This document outlines the characteristics and structure of the workflows and the minimum functional and technical requirements necessary for the development of the VRE. It describes in detail all activities required for the full development of the VRE, including the creation of web interfaces for the VRE, the specification of workflows, a list of potential data sources, the types of analyses and models to be integrated, and the development of scripts and algorithms required for web services.

The Phyto VRE will be implemented with a modular structure that includes data analysis and modelling services by automating computational operations in workflows in the cloud, without the need to install or manage software locally. In each block, service development can be supported by the R, Python and Matlab languages. All workflows will be functional to return outputs that can be used for research reports. The Phyto VRE will be developed using established technologies such as Tesseract and NaaVRE available from LifeWatch ERIC, with the aim of maximising integration between these two platforms (Figure 4). The Phyto VRE must ensure the integrity and security of all input data throughout the exploitation process. The Phyto VRE will be accessible via a dedicated website within LifeWatch Italy, accessible from the CNR reference portal for the ITINERIS project (ITINERIS hub). The VRE Interfaces website will comply with W3C accessibility standards. The portal will serve as a public forum for the dissemination of information about VRE, including general data, the composition of the design team, a catalogue of published works and an overview of related projects. There will also be a private area, accessible via single sign-on authentication, for content management and site administration.



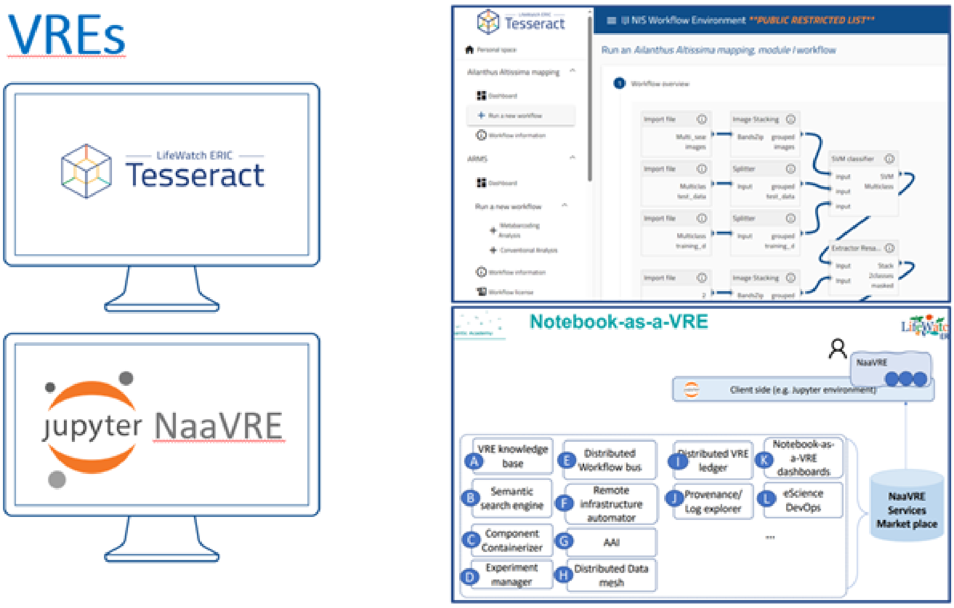


Figure 4. Current LifeWatch resources and their links in the optimized Phytoplankton VRE.

The proposed implementations will also provide:

- Improvements to the web services already present in the Phytoplankton VRE, in particular the Phytoplankton Atlas, the Traits Thesaurus and the three computational services Community Indices, Community Matrix and Data Filtering.
- A workflow that links all the existing services;
- Two new workflows including new computational and modelling services focusing on demographic, morphological and physiological-behavioural traits of phytoplankton species, community organisation and niche distribution, also in relation to abiotic parameters in the water column.

In February, the tendering process was launched using the Terms of Reference to identify the corporate contractor. During July and August the procurement process for software development services for the realisation of LifeWatch Italy's Virtual Research Environments Phytoplankton VRE was ongoing and the procurement approval was completed on 27th August 2024.

Below are the services and workflows to be implemented and included in the new release of the Phytoplankton VRE and to be developed in all parts by the contractor. The Term of Reference identify the minimum actions and results to be ensured in the VRE, but can be further implemented by the contractor with improved and additional proposals.

3.2 Services

Optimisation of the Phytoplankton VRE includes the implementation of the Phytoplankton Traits Thesaurus, which will include not only morphological but also physiological and behavioural parameters and abiotic variables. The Phytoplankton Traits Thesaurus (https://ecoportal.lifewatch.eu/ontologies/PHYTOTRAITS) is the first initiative to address the semantics of functional phytoplankton traits, with a focus on morpho-functional traits. It was developed and published by LifeWatch Italy, the Italian node of the European e-Science Infrastructure for Biodiversity and Ecosystem Research (LifeWatch ERIC) on ECOPORTAL, the LW ERIC repository of semantic resources in ecology and biodiversity. The Phytoplankton Traits Thesaurus reflects the agreement of a community of scientific experts on the semantic properties (e.g. label, definition) of about 120 morphological traits. The implementation of the 'Phytoplankton Traits Thesaurus' must include terms with associated labels and definitions for the functional traits of phytoplankton species, in particular including information on physiological-behavioural traits. In addition, the thesaurus shall include information on the main abiotic variables used to describe the water column. The thesaurus shall be a SKOS thesaurus, conforming to Semantic Web standards, and shall be easily accessible through a set of APIs to meet different development needs. The thesaurus shall provide alignments and mappings with existing resources to increase the interoperability of the semantic resource.

In addition, a plugin or application for Microsoft Excel will be developed to allow the semantic resources contained in the Phytoplankton Trait Thesaurus to be accessed directly in the Excel spreadsheet to facilitate the harmonisation of phytoplankton datasets and abiotic variables. The service should provide a standard Excel spreadsheet with an extension or application/plugin to retrieve the semantic content of the Phytoplankton Trait Thesaurus and other semantic annotation resources for abiotic variables (e.g. NERC, BODC and others) to facilitate user harmonisation and integration of phytoplankton datasets. The spreadsheet model with its extension must be able to be shared, edited and reused like any other spreadsheet, allowing integration with SKOS ontologies or vocabularies. This service must be distributed and shared among scientists, and be able to ingest data and semantic annotations in RDF or CSV format.

3.3 Workflows

The optimisation of the Phytoplankton VRE will also include the development of three main workflows:

- The workflow 1 will include web services like "Traits Computation", "Size Class Distribution", "Size Density Relationships" and other services like "Data Filtering", "Community Matrix and Analysis" and "Taxonomic Indices" for phytoplankton trait analyses (Figure 5 a).
- The workflow 2 will be developed to study of the relationships between morphological traits and physiological characteristics of phytoplankton and to investigate trait variation in response to climate change (Figure 5 b).
- The workflow 3 will be developed to understand the mechanisms of interspecific coexistence of phytoplankton species and niche partitioning mediated by body size, nutrients and temperature (Figure 5 c).

Each workflow will be designed with a modular structure that integrates services for data discovery, harmonisation, interoperability, analysis and modelling. These workflows will be integrated into the ITINERIS Hub via LifeWatch Italy and will be accessible via 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 format and vector data in SHP format. The outputs of the workflows are tables in CSV or TXT format, raster images of the newly derived products and plot results, such as time series, in PDF, JPG or PNG format. The inputs and outputs can be adapted or modified to specific needs during the development phase, in collaboration with the scientific team of the CNR-IRET-LE.

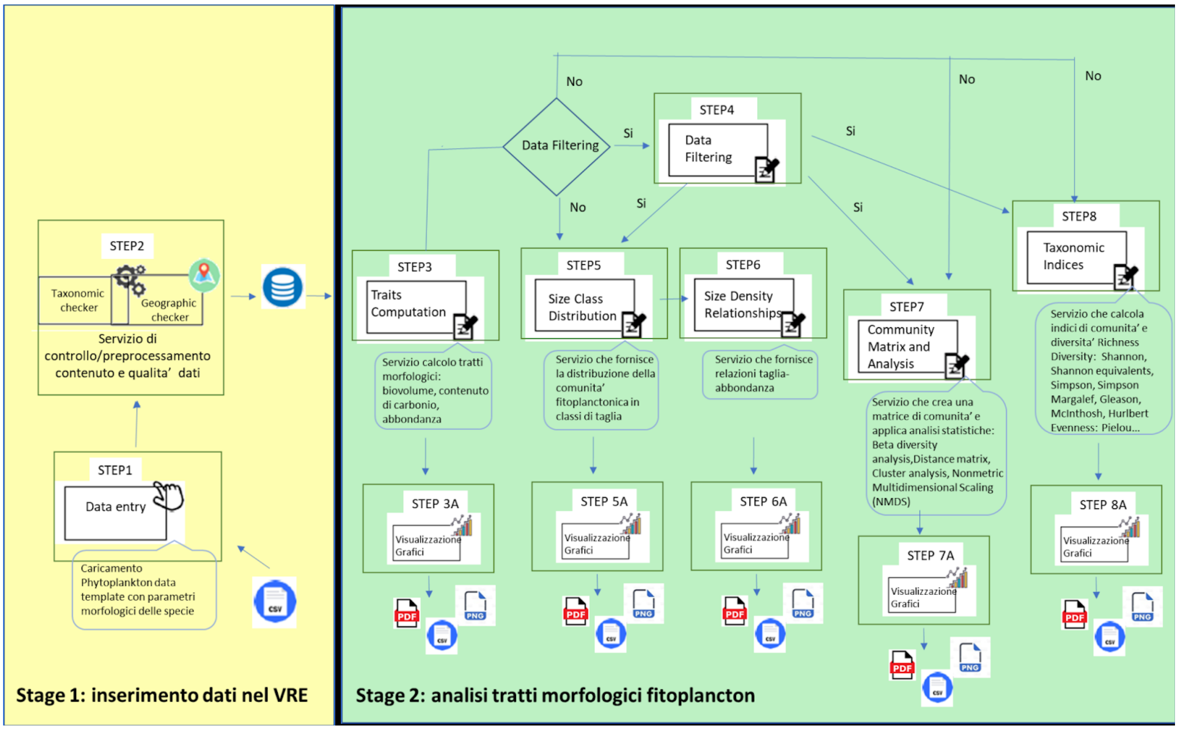


Figure 5a: Illustrative Diagram of Workflow 1, extrapolated from the *Term of Reference*.

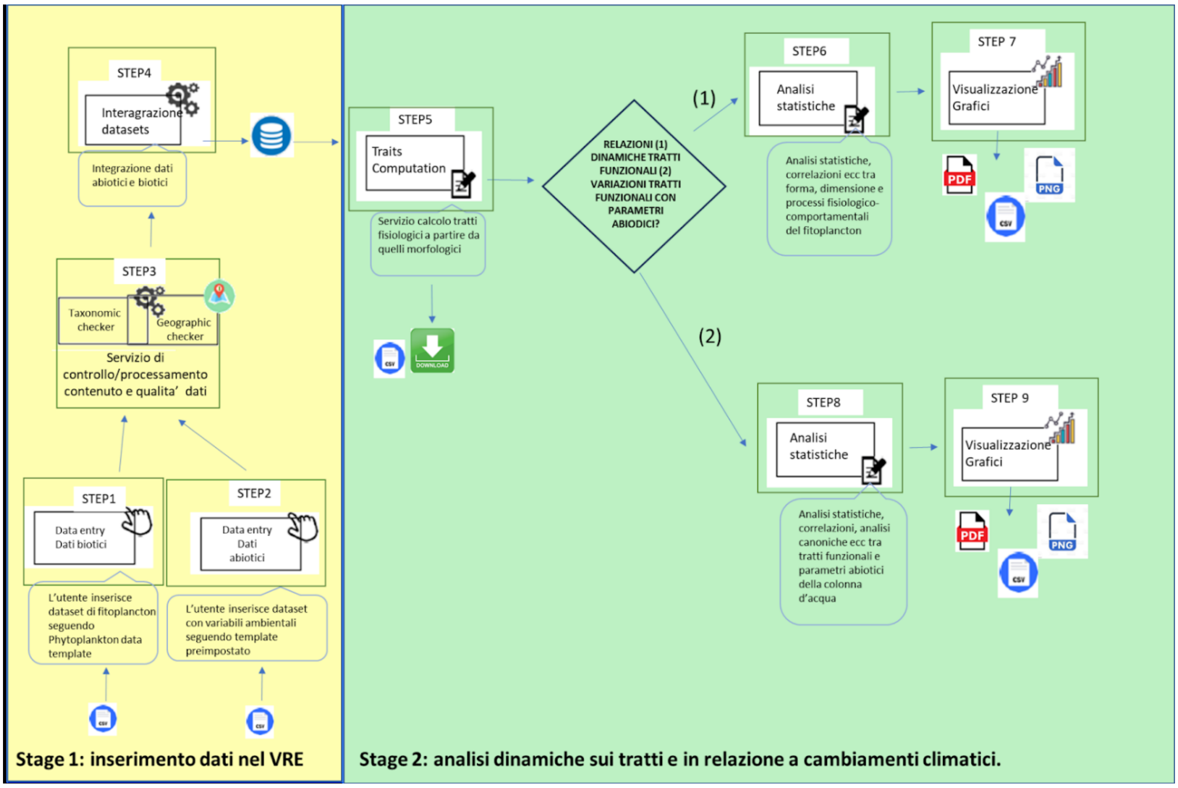


Figure 5 b: Illustrative Diagram of Workflow 2, extrapolated from the *Term of Reference*.

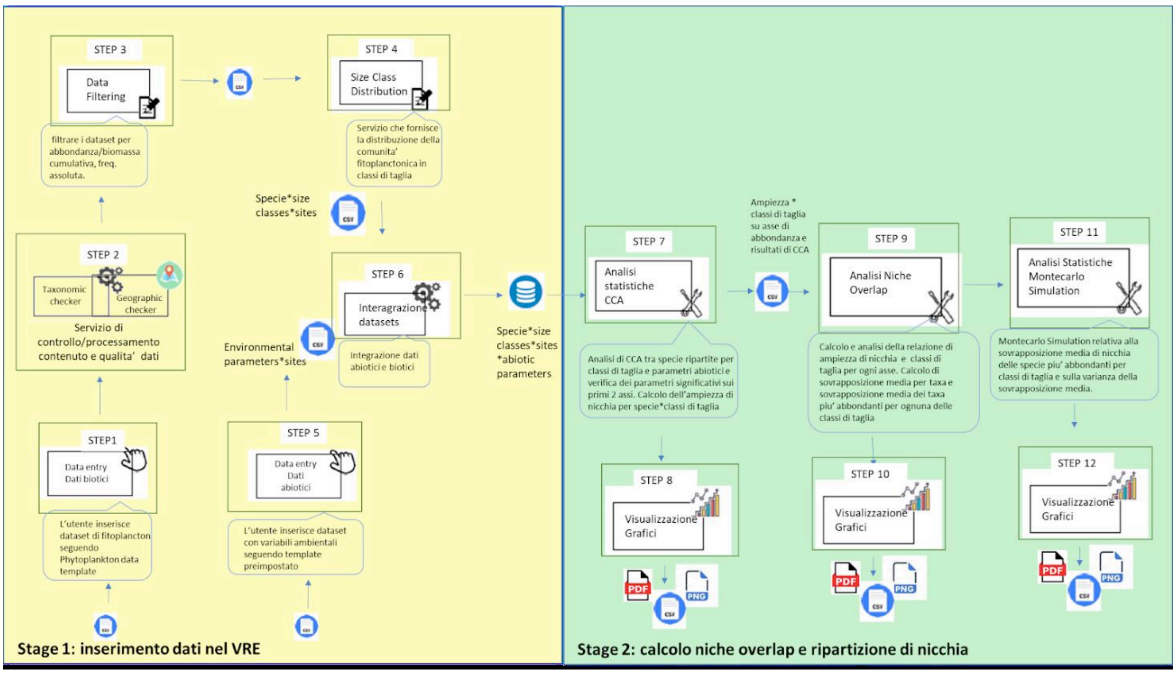


Figure 5 c: Illustrative Diagram of Workflow 3, extrapolated from the *Term of Reference*.

4. ANALYSIS CURRENTLY DEVELOPED AND CASE STUDY

4.1. Workflow 1 development using NaaVRE technologies

The development of Workflow 1 has been conducted utilizing Notebook-as-a-VRE (NaaVRE) technologies (https://naavre.lifewatch.dev/vreapp). Specifically, the Phytoplankton Virtual Research Environment (Phyto VRE), which includes six computational services implemented in R, has been integrated into the NaaVRE prototype (Figure 6a, b).

Starting with the web services developed by LifeWatch Italy and already available online, a workflow was created to connect these pre-existing services into a unified workflow. This was made possible using Jupyter Notebooks, which provide features for containerizing and orchestrating notebook fragments (cells) as workflows, while automating their execution in cloud environments. This workflow allows users to analyze phytoplankton occurrence data and morphological traits (e.g., size and shape) at various levels of data aggregation and scale.

To do so, a dedicated Virtual Lab was created inside the LifeWatch ERIC testbed cluster (https://naavre.lifewatch.dev/vreapp) to test the feasibility of running the Phytoplankton VRE computational and analysis R scripts in the NaaVRE (Figure 6a). After an initial step of authentication through a personal account or with credentials from public identity providers like GitHub or Google, six separated Notebooks, one for each of the Phytoplankton VRE web services, were created and each service was enclosed in a single-cell container (Figure 7 a,b). The connection between the various containerized cells was tested in order to understand if the six web services of the Phytoplankton VRE could be executed in a workflow on the Cloud. The workflow scheme was designed to have the minimum computation time and the maximum possibility of choosing between the services useful to the user, who could choose whether to run all the services or choose only some of them, through an interface designed to be simple and intuitive, relying solely on the linking of cells.

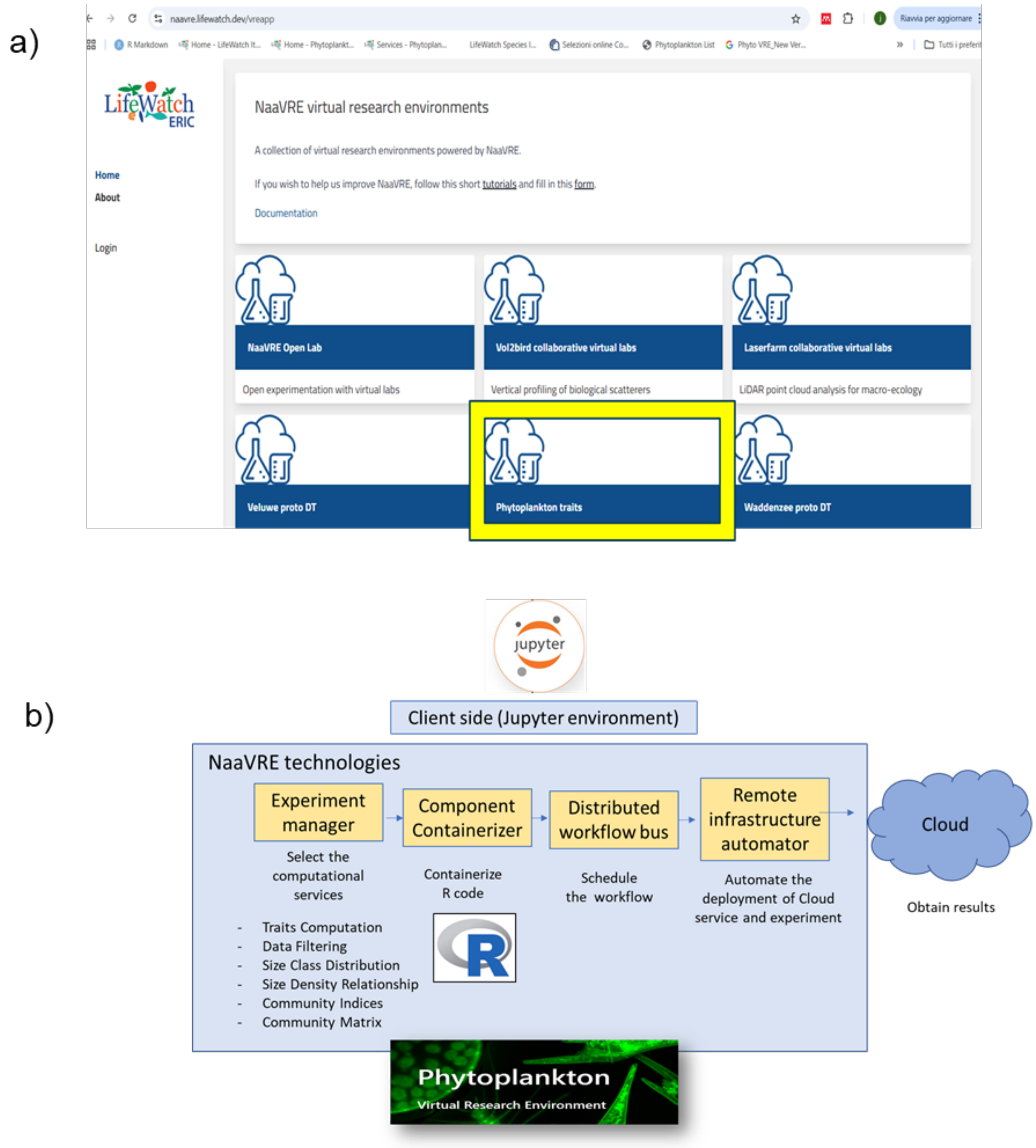


Figure 6 a) The dedicated Virtual Lab “Phytoplankton traits” created within the NaaVRE,
b) Illustrative schema of the integration of Phyto VRE services within the NaaVRE framework

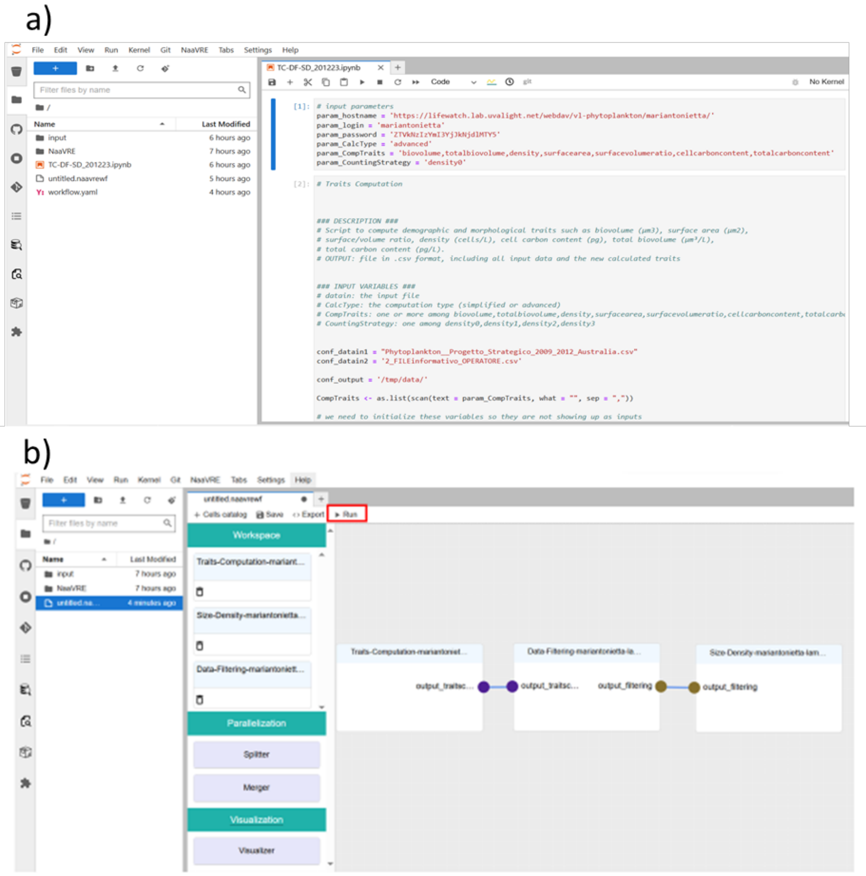


Figure 7. a) Schematics of the workflow 1 within the NaaVRE framework and b) containerization process.

Currently, workflow 1 in NaaVRE is available to developers and members of the working group. Once the testing phase is completed, the workflow will be officially released and made open access. For a comprehensive understanding of how NaaVRE operates, including its features, architecture, and applications, the full documentation is available at the following link: https://naavre.net/docs/about/.

The feasibility of deploying and executing the services provided by the Phytoplankton VRE as a cloud-based workflow using the NaaVRE framework has been evaluated through a dedicated case study.

4.2. Workflow 1 Case Study: Analysis on phytoplankton guilds at different levels of data aggregation and scale, from local to global.

Once the development of the workflow 1 has been performed using the NaaVRE technologies, a case study focused on investigating the role of dimension and taxonomy, in relation to the latitudinal gradient in phytoplankton was used in order to develop and test the workflow 1 prototype. We used an existing phytoplankton dataset with 127311 records from PROGETTO STRATEGICO 2012-2014 (DOI:10.1038/s41597-023-02785-w) from 24 transitional waters from five different biogeographical regions: the South-Western Pacific Ocean (Australia), the South-Western Atlantic Ocean (Brazil), the Mediterranean Sea (Greece and Turkey), the Indo-Pacific Ocean (Maldives) and the Northern Atlantic Ocean (Scotland).

The management of the input dataset was developed by giving the user the possibility to use either their own dataset or to use the datasets on the LifeWatch Italy DataPortal, through direct input of the web page link (https://dataportal.lifewatchitaly.eu/data). The following services were included in the workflow 1 and used for the analysis:

- Traits computation
- Size Class Distribution
- Size Density Relationship
- Community Indices

The results potentially achievable from the analytical workflow 1 are:

A quick evaluation of the species distribution (based on occurrences), common species and most representative shapes (Figure 8 a,b)

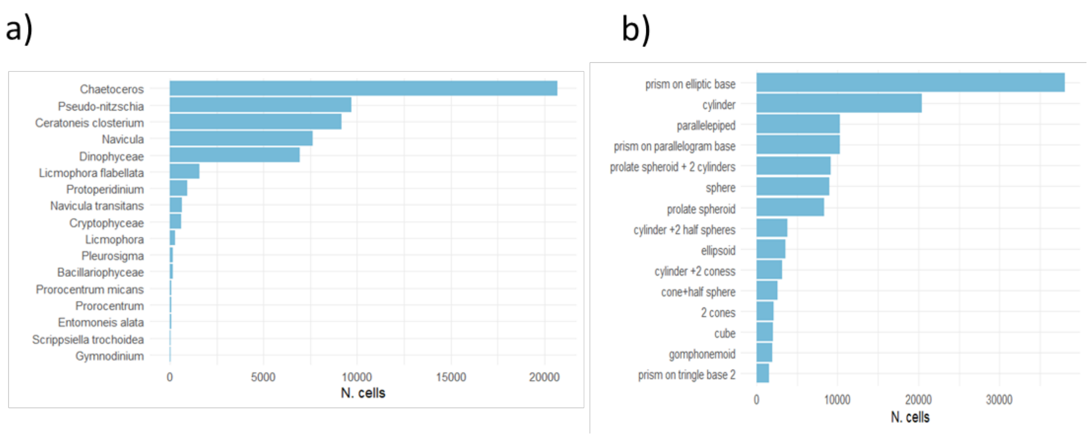


Figure 8: a) number of cells sampled as a function of scientific name, b) shapes most commonly found within the entire dataset.

Community indices and diversity trend investigated along the latitudinal gradient, comparing eco regions located at similar latitudes with them and with those located at different latitudes (Figure 9).

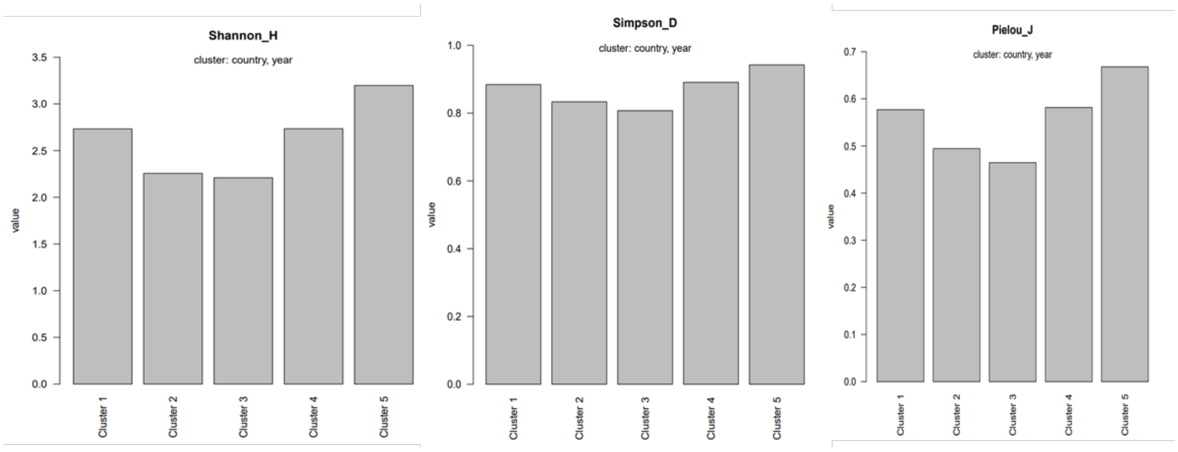


Figure 9: Community diversity indices.

Among the different analysis the workflow calculated the size distribution of the sampled cells in relation to biovolume (Figure 10) highlighting a predominant concentration of cells with intermediate size classes (considering the log2 biovolume). This analysis can be repeated for each ecoregion or to make comparisons among ecoregions.

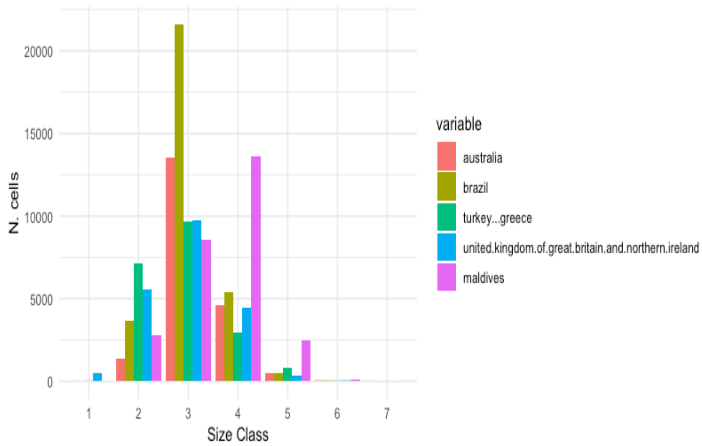


Figure 10: Barplot comparing size distributions, based on biovolume, in relation to different countries.

Overall, the work carried out to develop several analysis in a unique workflow allow the user to quickly develop different analysis on trait-based and diversity and ecological aspect of phytoplankton community on the one side. Thus, facilitating the scientific community to share and reuse their results. The final goal of the workflow was to obtain FAIR data that can be used and shared for forecasting models and investigations on phytoplankton ecology.

In conclusion, what has been included in this deliverable is the full procedure to run experiments or research project in a Virtual Research Environment (VRE) based on NaaVRE technologies that LifeWatch Italy is developing in collaboration with LifeWatch ERIC within the ITINERIS project. All workflows and web services, together with guidelines for training materials, will be available in the LifeWatch ERIC Metadata Catalogue and will be completed and fully operational by the end of the ITINERIS project, providing advanced tools for research and scientific innovation in phytoplankton ecology.