Project Title	Blue-Cloud 2026: A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters
Project Acronym	Blue-Cloud 2026
Project Number	101094227
Type of project	RIA – Research and Innovation Action
Topics	HORIZON-INFRA-2022-EOSC-01
Starting date of Project	01 January 2023
Duration of the project	42 months
Website	www.blue-cloud.org

D4.2 – VLabs Implementation Guidelines

Work Package	WP4	Blue-Cloud Virtual Labs for demonstrating cross-domain web-based open science
Task	T4.1	WP4 coordination
Lead author	Patricia Cabrera (VLIZ), Massimiliano Assante (CNR-ISTI)
Contributors	Alexander Barth (ULiège), Charles Tropin (ULiège), Francesco Palermo (CMCC), Joao Vitorino (IH), Julien Barde (IRD), Bastien Grasset (IRD), Steven Pint (VLIZ)
Peer reviewers	Dick Schaap (MARIS), Stephane Pesant (EBI)
Version	V1.0
Due Date	30/07/2023
Submission Date	26/07/2023

Dissemination Level

X	PU: Public
	CO: Confidential, only for members of the consortium (including the Commission)
	EU-RES. Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
	EU-CON. Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
	EU-SEC. Classified Information: SECRET UE (Commission Decision 2005/444/EC)

Version History

Revision	Date	Editors	Comments
0.1	15/05/2023	Patricia Cabrera (VLIZ)	Table of Content
0.2	22/05/2023	Patricia Cabrera (VLIZ)	Introduction, WP4 Summary
0.3	9/06/2023	Patricia Cabrera (VLIZ)	Blue-Cloud Virtual Labs VRE Implementation guidelines
0.4	16/06/2023	Patricia Cabrera (VLIZ)	First draft
0.5	30/06/2023	Massimiliano Assante (CNR-ISTI), Enrico Baglione (INGV), Simona Simoncelli (INGV), Francesco Palermo (CMCC), Bastien Grasset (IRD)	Second draft
0.6	17/07/2023	Dick Schaap (MARIS)	1st review
0.7	21/07/2023	Stephane Pesant (EBI)	2nd review
1.0	26/07/2023	Patricia Cabrera (VLIZ), Sara Pittonet (Trust-IT)	Quality check and final version

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 2 of 19

Glossary of terms

Item	Description
API	Application Programming Interface
CMEMS	Copernicus Marine Environment Monitoring Service
CNR	Italian National Research Council
CMS	Content Management System
CVS	Code Versioning System
DD&AS	Data Discovery & Access Service
D4Science	Data Infrastructure promoting Open Science (managed by CNR)
ECMWF	European Centre for Medium-Range Weather Forecasts
EMODnet	European Marine Observation and Data Network
EOSC	European Open Science Cloud
EOVs	Essential Ocean Variables
GRSF	Global Record of Stocks and Fisheries
JSON	JavaScript Object Notation
IDE	Integrated Development Environment
OGC	Open Geospatial Consortium
RFMO	Regional Fisheries Management Organisations
RIA	Rich Internet Application
TBD	To Be Defined
TBA	To Be Arranged
VLab	Virtual Laboratory
VRE	Virtual Research Environment

Keywords

EOSC; Virtual Labs; Big Data; Virtual Research Environment; Data infrastructures

Disclaimer

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 3 of 19

EXECUTIVE SUMMARY

Blue-Cloud 2026 builds upon the pilot Blue-Cloud project that established a pilot cyber platform, providing researchers access to multi-disciplinary datasets from observations, analytical services, and computing facilities essential for blue science. The Blue-Cloud Open Science Platform developed a collaborative environment where different services are available. A **Data Discovery and Access Service** (DD&AS) was developed as an overarching service to facilitate federated queries & retrieval, and smart sharing of multi-disciplinary datasets with human and machine users. The Blue-Cloud **Virtual Research Environment (VRE)** orchestrates the computing and analytical services in specific integrated and managed applications exploiting federated Blue-Cloud data resources as well as external data resources**.** The Blue-Cloud innovation potential was explored and unlocked by **five dedicated demonstrators** as Virtual Labs (VLabs) co-designed with top-level marine researchers to demonstrate the power of the Blue-Cloud Open Science platform.

In Blue-Cloud 2026, three of these VLabs are expanded and two new ones are created. The objective is to deploy a wide range of custom methods and technologies, such as software and algorithms, using cloud-based services in the Blue-Cloud VRE.

This deliverable provides guidelines on how to implement the VLabs in the VRE. The VLabs technical requirements are assessed by CNR-ISTI, partner responsible for the Blue-Cloud VRE (WP5), to ensure they meet the platform requirements and further develop the services provided by Blue-Cloud. This deliverable is the outcome of discussions held during the VRE Hands-on workshop that took place in Heraklion, Greece in June 2023. First a summary of the activities of WP4 is presented, followed by an overview of the VRE services, as presented during the workshop. VRE services offer a comprehensive array of features, fostering collaboration, facilitating data analytics, enabling result dissemination, and ensuring seamless integration with external systems. In fact, the VRE caters to the entirety of the research lifecycle.

The document is structured as follows: Section 1 briefly introduces the role of this deliverable for the development and delivery of the VLab implementations. Section 2 summarises the WP4 objectives and composition. Section 3 describes the VRE services by giving a detailed overview and focuses on the solution and technical requirements for each of the 5 VLabs. The last section concludes the report summarising the guidelines and the solutions provided.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 4 of 19

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 5 of 19

1. Introduction

This deliverable provides guidelines to implement the five VLabs (Virtual Laboratories) within the Blue-Cloud Virtual Research Environment (VRE) in the most efficient way. The recommendations are based on the technical requirements collected by and from each of the five VLabs in the first deliverable D4.1 VLbas Technical Requirements, and the challenges and outcomes from the VRE Hands-on workshop that took place in June 2023, Heraklion, Greece. D4.1 focused on describing the methodologies and the input datasets to be used and the computing and storage needs, whereas D4.2 analyses these to provide solutions. During the workshop, VLab representatives and developers learned how the VRE services are evolving in Blue-Cloud2026, and how their VLab can benefit from them.

The VRE services play a pivotal role in promoting Open Science practices within VLabs, enabling researchers to leverage the benefits of advanced e-infrastructures. These services encompass a wide range of functionalities, facilitating collaboration, data analytics, result publication, and seamless integration with external systems. These services also provide interfacing with the Data Discovery & Access Service (DDAS) and the interfacing with the WEkEO service of Copernicus Marine, so as to enable researchers to leverage external resources and data sources, expanding the capabilities and reach of their research endeavours. These integrations contribute to the comprehensive and interconnected nature of the Blue-Cloud ecosystem. It is important to note that these VRE services are built on the D4Science infrastructure and leverage the gCube open-source technology. This infrastructure provides a reliable and scalable foundation for the VRE services, ensuring their stability and performance. The exploitation of the gCube open-source technology further enhances the flexibility and extensibility of the services, allowing for continuous innovation and improvement. To ensure accessibility, the VRE services are deployed in the Blue Cloud gateway, accessible at https://blue-cloud.d4science.org. This gateway serves as a centralised hub, providing researchers with easy and convenient access to the services and Virtual Laboratories.

The objective of this deliverable is that VLabs can apply the best technical solutions to adapt their technologies and tools to the Blue-Cloud VRE. This work will benefit both WP4 and WP3 (Workbenches) for the above-mentioned reasons, but also WP5 'Blue-Cloud VRE platform evolution and integration with EOSC resources and services' to expand/ upgrade their current VRE functionalities.

2. WP4 status summary

WP4 *"Blue-Cloud Virtual Labs for demonstrating cross-domain web-based open science"* designs, manages and runs the creation of five VLabs in the Blue-Cloud Virtual VRE.

- VLab #1 Coastal oceans observations along Europe (led by IH)
- VLab #2 Coastal currents from observations (led by ULiège)
- VLab #3 Carbon-Plankton dynamics (led by VLIZ)
- VLab #4 Marine Environmental Indicators (led by CMCC)
- VLab #5 Global Fisheries Atlas (led by IRD)

The VLabs cover a set of marine multidisciplinary data from the BDIs and other marine Research Infrastructures (RIs), like JERICO-RI. These data may be from very different types and origins and are usually not gathered together (coastal and open sea data, plankton, hydrodynamics and fisheries data). The VLabs innovative services, in the form of data products and/or analytical tools, demonstrate the added-value of web-based open science as promoted by European Ocean Science Cloud (EOSC). VLabs #1 and #2 are newly developed in Blue-Cloud 2026, while VLabs #3 to #5 were first developed in the pilot Blue-Cloud project and will be expanded and upgraded during this new phase.

The development of the VLabs includes several tasks and phases, from data collection and processing to the development of analytical services, and to the publication of the results in the Blue-Cloud Data and Service Catalogue and EOSC Catalogues. This work is divided in four phases:

- **Phase 1 [M1-M3]: Description of the workflows.** Identification of datasets, tools, and technical requirements to implement/upgrade the VLabs in the VRE.
- **Phase 2 [M4-M24]: Implementation.** Beta versions of VLabs developed & implemented in VRE.
- **Phase 3 [M25-M36]: Testing & fine tuning.** Beta version fully matured and integrated in the VRE, making use of VRE services (DD&AS, and other services developed under WP2-3-5).
- **Phase 4 [M37-M42]:** Dissemination and user support: In collaboration with WP6-7, VLabs are presented at different events and exploited by users (e.g. hackathon). Feedback is used for service updates.

At the moment, phase 1 has been successfully completed and phase 2 has started. All VLabs have been created, including all the core VRE services, and the first implementations are foreseen during the next months.

**Funded by
the European Union**

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 7 of 19

3. Blue-Cloud Virtual Labs VRE Implementation guidelines

3.1. VRE services overview

The VRE services are instrumental in advancing Open Science practices within VLabs, empowering researchers to harness the advantages of state-of-the-art e-infrastructures. By leveraging these services, researchers can capitalise on the power of the Cloud and of e-infrastructures, driving scientific progress and enabling collaborative research efforts within the realm of Open Science.

VRE services offer a comprehensive array of features, fostering collaboration, facilitating data analytics, enabling result dissemination, and ensuring seamless integration with external systems. In fact, the VRE caters to the entirety of the research lifecycle, providing diverse services. Specifically, it offers access to three service frameworks:

(i) The Collaborative Storage Framework fosters collaboration among users. The Workspace provides a platform for VRE members to collaborate, share resources, and work together on projects. This collaborative environment enhances the efficiency and effectiveness of research activities within VLabs.

(ii) The Analytics Engine Framework empowers VLabs with powerful tools and resources. The DataMiner and the Cloud Computing Platform (CCP) enable sophisticated data analysis, allowing researchers to extract valuable insights from complex datasets coming from Blue Data Infrastructures. Additionally, VRE services offer specialised solutions such as the RStudio-based, the JupyterHub-based, the ShinyProxy-based for ShinyApps, and the solution for Dockerized Applications. These tools cater to different analytical needs and preferences, ensuring VLab members to have the flexibility and diversity of options to conduct their analyses effectively.

(iii) The Publishing Framework facilitates the dissemination of research outcomes by means of the VRE Catalogue and the Spatial Data Catalogue, that provide a means to organise and publish research results, making them accessible to the wider scientific community. This framework ensures transparency, reproducibility, and the sharing of valuable knowledge generated within VLabs.

Within a VLab, these service frameworks are conveniently organised and accessible through dedicated menu items displayed on the VLab landing page, as depicted in Figure 1.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227
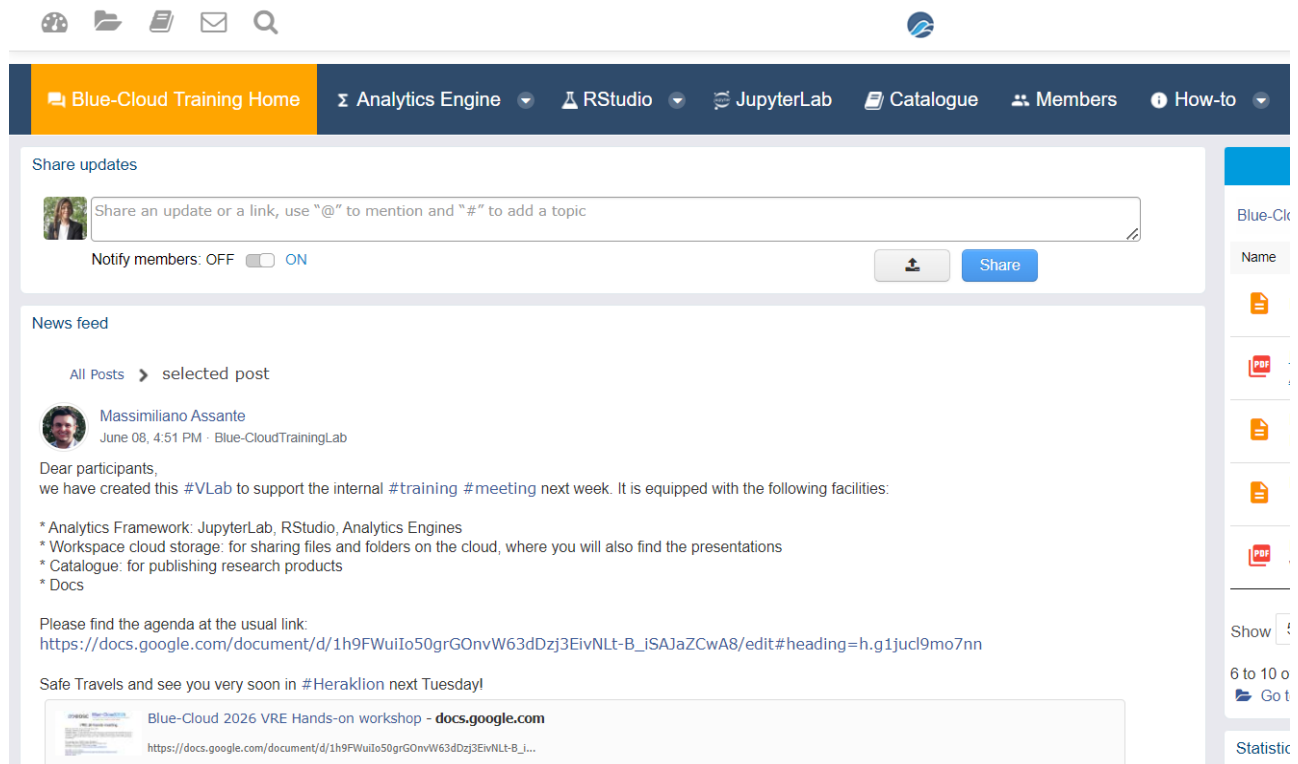
Page 8 of 19

Figure 1: Screenshot of the Blue Cloud training VLab set up for the VRE Hands on workshop.

These services are made available either by default or through specific requests to WP5. Every VLab can be equipped with:

- **Communication area:** for collaborative and open discussions on any topic and disseminating information of interest for the community, for example, the availability of a research outcome.
- **Administration area:** User and groups Management dashboard for managing membership and roles.
- **Analytics Computing Framework area:** a set of data analytical services can be requested and customised in the VLabs:
  - **Analytics engine:** Blue-Cloud2026 features 2 engines as described in the following.
    - *DataMiner service*: the engine that was used in Blue-Cloud and it will be maintained in Blue-Cloud 2026. It comprises a rich array of ready to use analytical methods ranging from data clustering methods to geospatial data analytics, occurrence data management, and species distribution maps generation.

■ **Cloud Computing Platform (CCP)**: a newer engine sharing same implementation principles and same main features, that will evolve in Blue-Cloud 2026, and was developed to improve its functionality exploiting the advancements in IT and Software engineering of the last decade.

Table 1: Synoptic table comparing the DataMiner and CCP engines features

Features	Data Miner	CCP
Virtualisation	Virtual Machines	Containers, isolated runtime
Execution Infrastructures	Per VLab, selectable by VLab mgmt.	Multiple, selectable by users
Standard protocol	OGC WPS (XML)	OGC API Processes (JSON)
Method Integrations tool	Software Importer (SAI), Github support	Method Importer / Your IDE, any CVS support
Parallel Executions	Yes, also Map Reduce	Basic, more support will be added
out-of-the-box methods availability	Many	Few
JupyterLab and Rstudio integrations	Yes	Yes, Easier (Code generation tool)
Front-end technology	RIA	Web Component based, exportable in multiple CMS

○ **RStudio** allows users to perform online statistical analyses. RStudio is no longer shared between several users but a dedicated instance of it is activated upon request. Moreover, its storage space is now persistent. It offers a predefined list of R packages, and each VLab can define its RStudio server's configuration, Standard (4 cores/8GB RAM) and Large (8 cores/32GB RAM).

○ **JupyterLab**, a web-based interactive development environment for Jupyter notebooks, code, and data. It allows users to configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.

JupyterLab comes with a set of preconfigured kernels and several hardware resources options to cover a set of user requirements.

- **Workspace** for storing, organising, and sharing items, such as datasets, scripts, or outputs. This cloud-based solution provides two storage options: the Workspace and the Dataspace. Both can be accessed through RStudio and JupyterHub. The Workspace replicates the data, and it is fault tolerant, resulting in being slower with respect to the Dataspace which is faster with very low latency but does not implement replicas and faults management. The folders in the workspace can be shared, private or public, this is managed by the VLab manager that can provide permissions (read only, write any or write own).

  The computations run in the analytical services can take inputs from the Workspace. Inputs can also come from Workspace folders shared among several users. After a computation is completed, a Workspace folder is created which contains the input, the output, the parameters of the computation, and a provenance document summarising this information. This folder can be shared with other people and used to execute the process again. Thus, the complete information about the execution can be shared and reused.

- **Spatial Data Infrastructure (SDI) services** provides users with the capability to store, discover, access, and manage vectoral and raster georeferenced datasets. The SDI exploits the following technologies: GeoServer equipped with PostgreSQL and PostGIS, GeoNetwork, Thredds.

- **Catalogue Framework** to document and publish any generated research product. Its primary component is the VRE Data Catalogue. The VRE Data Catalogue service is a catalogue service built on open-source technology for data catalogues (CKAN ckan.org). Via this Catalogue users can also search and browse data, products, and resources (posters, deliverables, etc) of interest from the Blue-Cloud community.

3.2. Solutions to technical requirements

The provisional technical requirements of each of the VLabs were provided in the first deliverable D4.1 Technical Requirements earlier this year. These have been advanced and further detailed during the VRE Hands on workshop, where solutions to the challenges presented were provided and they are captured in this document.

3.2.1. VLab#1 Coastal oceans observations along Europe: ICOOE

The ICOOE VLab will widely expand users' ability to access and integrate observations collected along Europe's coastal ocean, building new knowledge on key scientific and societal questions and challenges, supporting researchers, Blue Economy actors, environmental managers, and crisis response. This VLab is

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 11 of 19

divided into three Thematic Services (TS): TS#1 Transboundary Processes and Connectivity, TS#2 Extreme Events, and TS#3 Ocean Glider, along the European coastal margins.

A test case is under development for TS#1 "Transboundary processes and connectivity along the European coastal margins" and was presented during the VRE Hands on workshop. TS#1 explores the potential of data integration, analysis and advanced visualisation in the understanding of transboundary processes along the European Coastal Ocean and mapping their potential impacts (e.g., biological connectivity, contaminants spread).

TS#1 will integrate an interface for users to select a subset of various parameters (surface and subsurface temperature, salinity, currents and sea surface height) from different sources for a requested geographical area within the Iberian margin and a specific time period, as well as specific tools for data pre-processing, exploration, integration and visualisation. The first step in this process will consist in data identification and collection. Data available in the BDI's can be accessed through the DD&AS. Other data that is available for specific time periods can be accessed through the Blue Cloud Vlab.

The test case presented in the workshop focuses on surface current data provided by HF radar and by NEMO model results. The structure presented comprises data access, data pre-processing and data exploration. Python programs for access to the data were presented. For the data pre-processing a number of FORTRAN routines already exist but require further development. For the exploration phase FORTRAN programs providing basic statistics for the surface current fields of HF radars or NEMO and correspondent visualisation were presented.

The test case will allow a simplified approach to the development and implementation of the different key components that integrate TS#1. The user interface is one of such key components. The workshop provided the opportunity to advance in the definition of the solution. The development of this interface will be performed using the Dash web app. The possibility of integration of web apps can be performed through the ShinyProxy solution. Other alternatives available concern Shiny for R and Python.

A second key component discussed in the VRE Hands on workshop was related to the incorporation of routines in different languages. FORTRAN code will be used for the data pre-processing, processing, exploration and integration stages. For this case, as well for Python and Java languages, this has to be done with Docker containers. Code written in different programming languages have to be divided into modules and containerized.

The work for TS#2 Extreme Events, and TS#3 Ocean Glider, along the European coastal margins have not started yet.

3.2.2. VLab#2 Coastal currents from observations

This VLab provides a service to generate integrated ocean surface current maps from direct and indirect current measurements derived from different sources, High Frequency (HF) radar; drifter data; and geostrophic currents from altimetry data. The outputs are gridded surface currents maps generated in NetCDF format. These will be used to run the MEDSLIK-II oil spill model to simulate the dispersion of an oil spill accident. The operational chain of the MEDSLIK-II model is going to be adapted and deployed in a Docker container. In this way it should be able to run as an autonomous Docker container and then also to be eventually imported in the Cloud Computing Platform (CCP) Analytics Engine. Users will be able to apply the service to a different domain (considering data availability). The integration is performed with DIVAnd (Data Interpolating Variational Analysis in n Dimensions) for HF radar (https://doi.org/10.1007/s10236-020-01432-x. It uses an open source licence (GPL), implemented in Julia: https://github.com/gher-uliege/DIVAnd_HFRadar.jl in a Jupyter notebook. The visualisations of the outputs will be integrated in the Jupyter notebooks. Julia client is used to download CMEMS in situ data: https://github.com/gher-uliege/PhysOcean.jl, and a metadata text file with the bounding box and time information pointing to individual NetCDF files.

Some open questions presented during the workshop dealt with how to handle secrets, such as SSH keys for GitHub or passwords as they should not be stored inside the notebooks. For this, a dedicated space to store them securely and made available when the Jupyter lab environment starts will be provided. To use SSH keys to connect directly to the docker container it is necessary to create a dedicated SSH key and when the container starts it reaches the secret key (this can be a string or a file).

Other issues concerned the connection between the output of the Jupyter notebooks and the MEDSLIK model. For this the use of the shared workspace is recommended.

3.2.3. VLab#3 Carbon Plankton dynamics

This VLab provides a service to analyse the relative contribution of the drivers in phytoplankton dynamics in the Belgium part of the North Sea, using near real-time data. The Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model is built using data from phytoplankton abundances, zooplankton abundances, and abiotic factors (nutrients, photosynthetically active radiation, and temperature). It shows which are the main factors that contribute to the phytoplankton dynamics and its temporal changes. The NPZD model runs in carbon units, allowing to integrate carbon information to infer carbon sequestration as part of the detritus state variable. Additionally, the model is applied with data from the Northern Adriatic Sea improving the reproducibility of the scripts.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 13 of 19

The VLab makes use of the RStudio and RMarkdown to run the model. The challenge presented was the need of heavy computational analyses because the model needs to perform 10,000 iterations (or more) within reasonable time. The solution is to distribute the workload to several "computers" and CPUs to speed up the process. Therefore, more pre-processing power will decrease computing time, allowing to perform sufficient iterations to decrease the uncertainty of the model output. A valuable option could be to use the Analytics Engine to speed up the process, distributing the iterations on several worker nodes, and to invoke these Analytics Engine processes from within RStudio.

3.2.4. VLab#4 Marine Environmental indicators

The Marine Environmental Indicators (MEI) VLab allows users to monitor and assess the environmental status by performing online spatio-temporal analysis with the implemented algorithms, for selected environmental variables. Multiple data sources are exploited in a unique data analysis service, which will allow the online computation of indicators through Jupyter Notebooks or a customised web application that exploits the VRE Analytics Engine services for computing the indicator's outputs.

With respect to the workflow explained in Deliverable D4.1, the expansion of this VLab, that was first developed in the previous Blue-Cloud project, focuses on the implementation of new algorithms to compute environmental indicators and on the enhancement of the user experience of the web application for exploiting other services of the VRE such as the WMS or other services for geospatial data and thus improving the data visualisation and accessibility. The deployment of the new algorithms will exploit the Analytical services JupyterHub and the Cloud Computing Platform (CCP) Analytics Engine.

Since the previous Blue-Cloud project, the Vlab also uses other custom services external to the VRE and based on Docker containers. They are available in two different environments: one for the development and one for the production and thus they are linked to the two VLabs currently deployed for MEI: MarineEnvironmentalIndicators and MarineEnvironmentalIndicatorsDev. The challenge of this VLab concerns if creating a new VLab is needed or if the current one will be updated with the use of the CCP. For now, the work will start using the JupyterHub on the MEI VLab dev.

There is the intention to reduce these two VLabs to only one. But the rules for the users authorization have to allow the separation of users between developers and normal users, and the whole framework used for development (composed by Analytics Engine, WebApp and related services) should not be accessible by the normal users.

The indicator Ocean Heat Content from this VLab is linked to workbench 1 on *Physics: temperature & salinity* from WP3. They both make use of DIVAnd software with Julia. It is necessary that the two of them use the same latest versions, i.e. DIVAnd v2.7.9 and Julia v1.9, to be able to share and reuse outputs.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 14 of 19

These versions need to be updated on the VRE. Other packages that may need an upgrade are: NCDatasets (v0.10.4), PyPlot (v2.9.0), IJulia (v1.21.3). The other indicators are developed with Python based code. Some Python packages could be added to the Vlab but the list is not ready yet and could be provided in a subsequent phase. The Trophic index (TRIX), and other indicators of eutrophication are linked to workbench 2 on *Eutrophication: chlorophyll, nutrients, oxygen*.

Currently the NetCDF outputs generated by the Data Miner are only accessible for 7 days but with the transition to the CCP and to the geospatial services this issue could be resolved. The challenge is how to visualise the map or the time series and how to store them. Possible solutions explored the use of wms services/OpenDap to access the stored outputs. The recommendation is to store the outputs in GeoServer or other WMS services that can handle netCDF data, and the Blue-Cloud VRE Data Catalogue.

Large amounts of data from this VLab relies on the connection with WEkEO and/or the new service called "Marine Data Store" for accessing Copernicus data and based on a new cloud-based data infrastructure. It is planned to be operational with the new advanced data extraction features starting by the end of this year 2023. The MDS is going to be currently available only for beta users. The connection of the VRE with these will improve the data access for this VLab. This connection will be a task in WP5 at a later stage of this project.

3.2.5. VLab#5 Global Fisheries atlas

The objective of this VLab is to provide a hands-on learning experience to explore the use of the Global Tuna Atlas and GRSF catalogue for studying and managing fish populations. The VLab will allow users to navigate and analyse the data provided by these tools, including creating maps and visualisations of tuna distribution and migration patterns. Overall, this VLab will provide a comprehensive and interactive learning experience that will enhance global understanding of tuna biology and management.

This VLab will be made of widely used software like RStudio IDE and usual components of Spatial Data Infrastructures / SDI (e.g. GeoNetwork, GeoServer, and PostgreSQL). RStudio IDE is used to execute the workflow which updates the Global Tuna Atlas and publish data outputs in the Vlab SDI. Java is used to execute the workflow which updates the GRSF and publishes data outputs in a triple store (RDF). New versions of RStudio, GeoNetwork, GeoServer, and PostgreSQL server are expected to be deployed in the coming weeks. In addition, specific software has been created (e.g., R Shiny apps) to explore, map and plot fisheries data.

The deployment of Shiny apps in the Blue-Cloud infrastructure requires establishing a continuous integration workflow. This essential process was facilitated by the training provided during the VRE Hands-on workshop hosted by Forth, with valuable guidance from WP4. The workshop offered insights on

Funded by
the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 15 of 19

seamlessly integrating the Shiny apps with the existing systems. To accomplish the deployment, a two-step approach is followed. Firstly, the apps are containerized using Docker, and subsequently, the Docker images are published to enable continuous integration within the VLab, as per the request of the infrastructure administrators. In collaboration with FAO, demonstrators of GitHub actions have been established to build and publish Docker images of existing Shiny Apps in both Docker Hub and GitHub Container Registry (GHCR). Notably, the initial deployment of a Shiny app from Docker Hub to the training infrastructure has been successful, and other Shiny apps have also been made available through GHCR. Moving forward, additional apps will be deployed using alternative options, such as managing secrets for SQL database authentication. The outcomes of the workshop will be documented and shared, and selected applications will be deployed within the infrastructure. This approach will enable each Vlab to easily deploy comprehensive Shiny apps for exploring and visualising data within the Blue-Cloud 2026 project's infrastructure.

The following steps to continue this work in this VLab are to schedule for software deployments, upgrades, log access and the customization of the Rstudio environment (specific system library installation by updating the Docker image used to deploy RStudio IDE in the infrastructure).

4. Conclusions and next steps

The first part of this document *"VRE Services overview"* explains the core VRE services and how they are evolving in Blue-Cloud2026. The objective is that VLab developers get acquainted with the already existing and new services to apply them to their workflows. The main evolution is the **Cloud Computing Platform** (CCP), and during the VRE Hands-on workshop we learned deeply how to use it and why it will benefit the work. This not only applies to the single VLabs, but also to collaborative shared work (outputs and methods) among them and the workbenches. Another great evolution is RStudio, with the benefit that it is now persistent, and one does not need to install the libraries at the beginning of each session.

The second part of this document *"Solutions to technical requirements"* highlights the technical requirements as requested by the VLab developers, and the recommendations and solutions provided (Table 2). Generally, a large part of the challenges presented by each VLab could be addressed by using the CCP, and for this reason a lot of time during the workshop was dedicated to explaining it in detail. Demonstrations with easy examples were provided, a priori the practical sessions, where teams of workshop participants were asked to solve a challenge making use of the new Methods Importer. This was especially helpful to demonstrate with an easy example how to use it, and practical sessions for the VLab developers were organised to test it and provide feedback.

Funded by
the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 16 of 19

Table 2: Summary of recommendations/actions per VLab

VLab#	Recommendation/action
VLab#1 Coastal oceans observations along Europe: ICOOE	TS#1 interface development using the Dash web app through the ShinyProxy solution. Other alternatives available concern Shiny for R and Python. Incorporation of routines (data pre-processing, processing, exploration and integration stages) in different languages: FORTRAN, Python and Java languages using Docker containers. Code written in different programming languages have to be divided into modules and containerized. TS#2 and TS#3 have not started yet.
VLab#2 Coastal currents from observations	Secrets, such as SSH keys for GitHub or passwords, should not be stored inside the notebooks. For this, a dedicated space to store them securely and made available when the Jupyter lab environment starts will be provided. To use SSH keys to connect directly to the docker container it is necessary to create a dedicated SSH key and when the container starts it reaches the secret key (this can be a string or a file). The use of the shared workspace is recommended for the connection between the output of the Jupyter notebooks and the MEDSLIK model.
VLab#3 Carbon Plankton dynamics	Running the model at many iterations requires heavy computation. For this it is recommended the use of the Analytics Engine to speed up the process, distributing the iterations on several worker nodes, and to invoke these Analytics Engine processes from within RStudio.
VLab#4 Marine Environmental indicators	Data visualisation and accessibility should make use of the WMS or other services for geospatial data. The deployment of the new algorithms will exploit the Analytical services JupyterHub and the Cloud Computing Platform (CCP) Analytics Engine. Use of DIVAnd v2.7.9 and Julia v1.9 for the indicator Ocean Heat Content, to be able to share and reuse outputs with workbench 1 from WP3. Recommendation to store the NetCDF outputs in THREDDS, GeoServer or other WMS services that can handle netCDF data.
VLab#5 Global Fisheries atlas	Deployment of R Shiny apps to explore, map and plot fisheries data: - the apps are containerized using Docker - Docker image publication Customization of the Rstudio environment: specific system library installation by updating the Docker image used to deploy RStudio IDE in the infrastructure.

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 17 of 19

On one hand, the feedback provided was generally very positive. Users found that the CCP is more flexible and makes use of good practices, such as git and docker for deploying. The User Interface (UI) is found to be easy to use and the adoption of OGC API processes is appreciated. Other good benefits are the possibility of sharing and duplicating methods, and the efficiency of pushing ShinyApp. On the other hand, for some users it may require specific technical skills, therefore it is recommended to provide an easy step-by-step guide to deploy and execute configuration. Another solution is to provide tips and labels in the UI and debugging feedback hints. External container logs of CCP methods are not available to the user that deploys.

During the next year and a half, *Phase 2: Implementation,* we expect that VLabs largely develop within the VRE. Implementation of their workflows should rely a lot on the use of the VRE services. For this, collaborations between WP4 and WP5 are necessary, and we expect to plan more workshops with practical sessions to advance the work and learn better the new technologies that are emerging rapidly.

References

M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, V. Marioli, P. Pagano, G. Panichi, C. Perciante, F. Sinibaldi (2019a) *The gCube system: Delivering Virtual Research Environments as-a-Service*. Future Gener. Comput. Syst. 95: 445-453 10.1016/j.future.2018.10.035

M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, F. Sinibaldi (2019b) *Enacting open science by D4Science*. Future Gener. Comput. Syst. 101: 555-563 10.1016/j.future.2019.05.063

Funded by the European Union

Blue-Cloud 2026 – A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters has received funding from the European Union's Horizon Europe programme under Grant Agreement no. 101094227

Page 19 of 19