

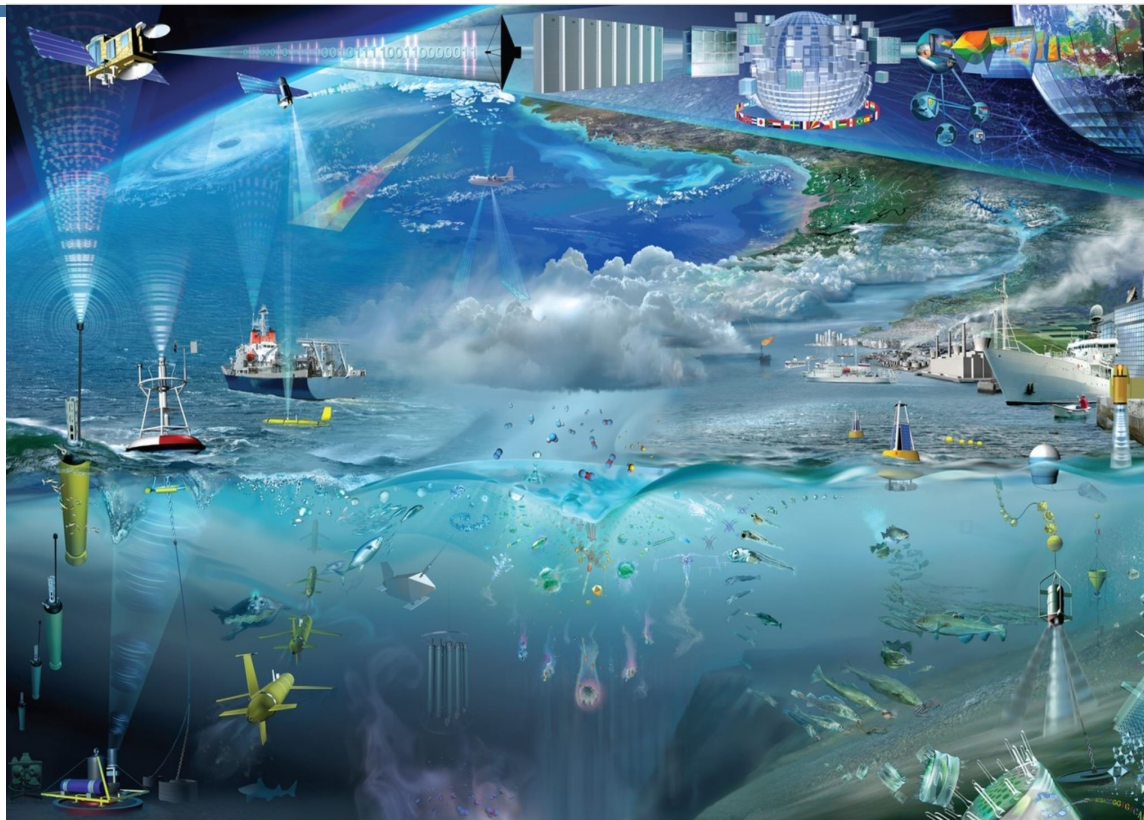
# Blue-Cloud: Exploring and demonstrating the potential of Open Science for ocean sustainability



Dick Schaap, MARIS, The Netherlands  
Massimiliano Assante PhD, CNR, Italy  
Pasquale Pagano, CNR, Italy  
Leonardo Candela, CNR, Italy



# Acquisition of marine and ocean data



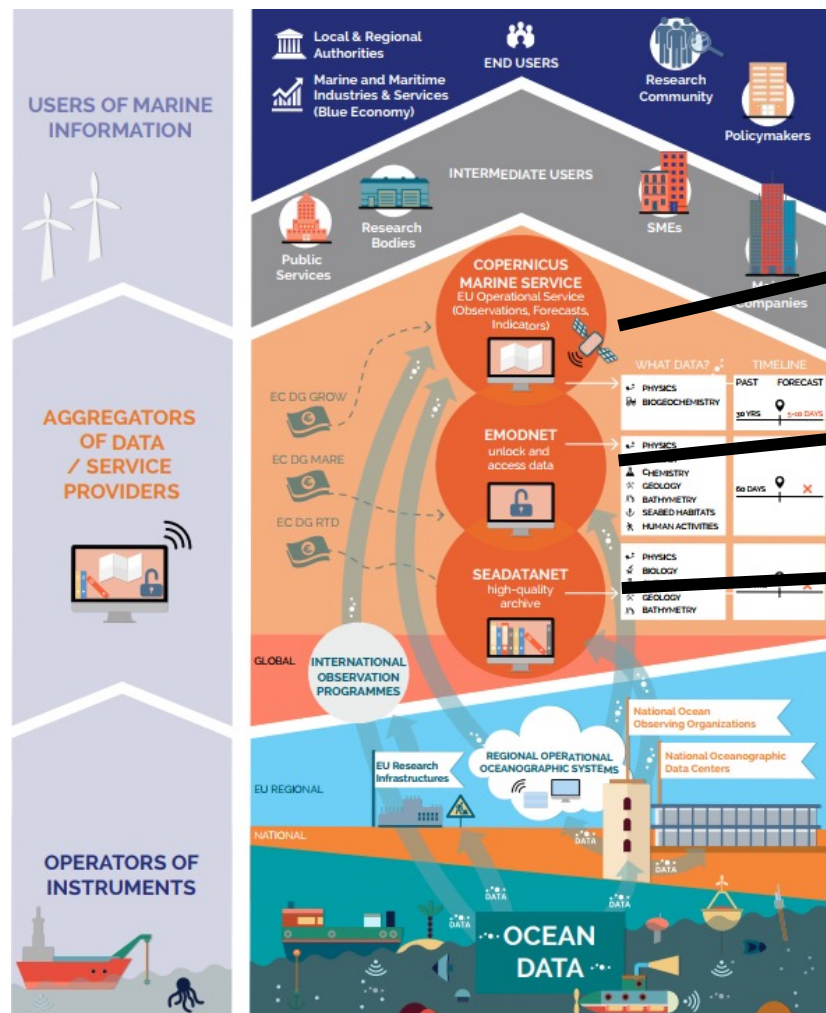
**In Europe we spent circa 1.4 Billion Euro a year  
in marine data acquisition**

- Scientific Research to gain knowledge and insight
- Modelling (including hindcast, nowcast, forecast)
- Economic activities: shipping, offshore industry, dredging industry, fisheries, tourism, engineering ..
- Environmental Management: monitoring and assessment (water quality, climate status, stock assessment)
- Marine Conventions and Directives, in Europe: Water Framework Directive (WFD), Marine Strategy (MSFD), Marine Spatial Planning (MSP), Coastal Zone Management
- EU Strategies, such as Green Deal, Blue Environment, Blue Economy





# European landscape marine data management



**Data aggregators and providers of data products and services**



# EU H2020 Blue-Cloud project

Marine thematic contribution to European Open  
Science Cloud (EOSC)

October 2019 – March 2023; 20 partners

## Blue-Cloud Mission

**“to promote the sharing of *data, processes and research findings* in the marine domain by delivering a collaborative web-based environment that enables *Open Science*;**

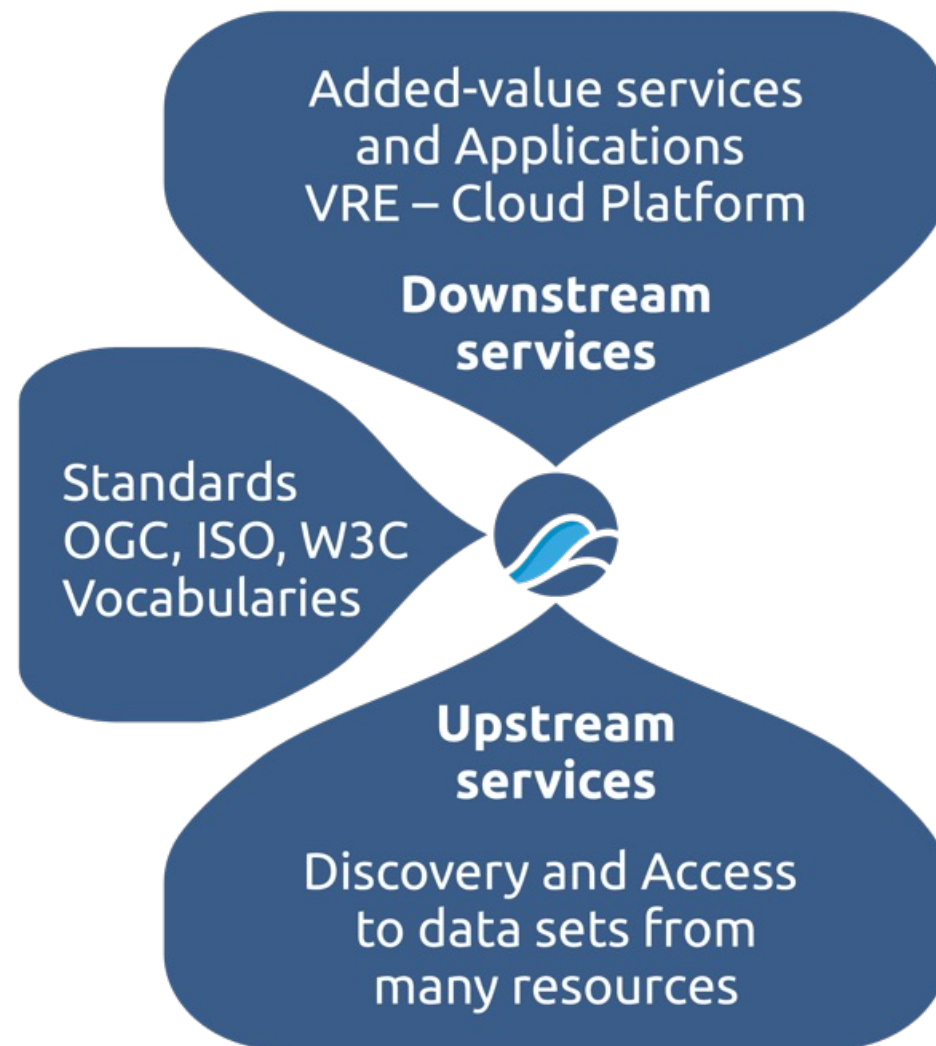
**underpinned by simplified access to an unprecedented wealth of marine data resources and interoperable added-value services and products”**





# Blue-Cloud overarching concept

- Developing and deploying Virtual Research Environment (VRE) with an array of services for configuring and running Virtual Labs for specific analytical workflows, use cases, and demonstrators
- Applying common standards and interoperability solutions for providing harmonized metadata and data
- Developing and deploying harmonized discovery and access to established European marine data management and processing infrastructures





Blue-Cloud

# Blue-Cloud federation of major infrastructures

**urOBIS**

European Ocean Biogeographic  
Information System



**EMODnet**

European Marine  
Observation and  
Data Network



European Nucleotide Archive



**ICOS**

Ocean  
Thematic  
Centre

**EcoTaxa** BETA

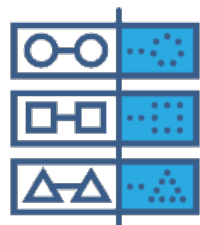
Blue Data infrastructures



E-infrastructures



# Key products and services



- **Blue-Cloud Data Discovery & Access service**, federating key European data management infrastructures, to facilitate users in finding and retrieving multi-disciplinary datasets from multiple repositories
- **Blue-Cloud Virtual Research Environment infrastructure** to provide a range of services and to facilitate orchestration of computing and analytical services for constructing, hosting and operating Virtual Labs for specific applications
- **Blue-Cloud Virtual Labs**, configured with specific analytical workflows to serve as **Demonstrators**, which can be adopted and adapted for other inputs and analyses



# Blue Cloud Discovery and Access service

Built and operated by MARIS, CNR-IIA, and CINECA (EUDAT)



# Blue Cloud Discovery and Access service



- **Facilitates users:**

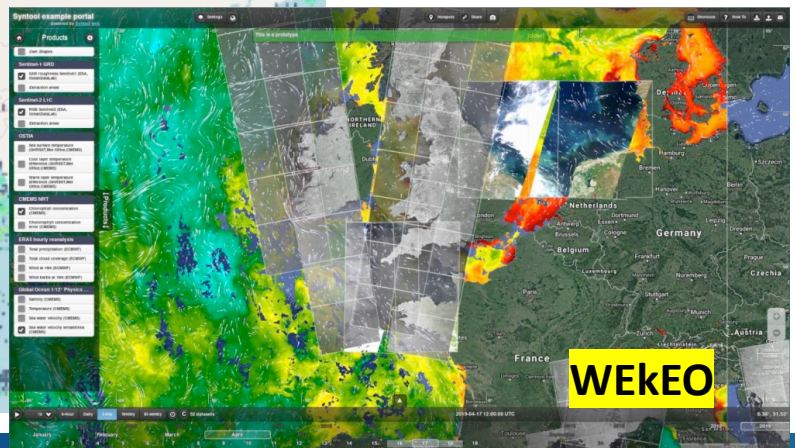
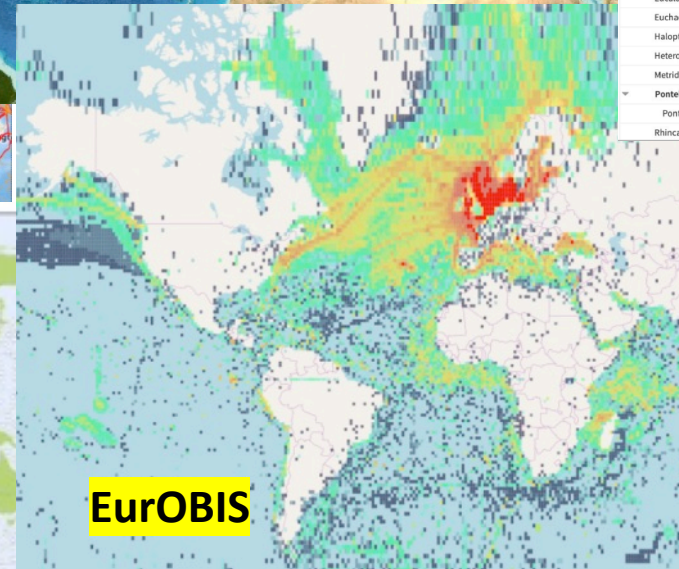
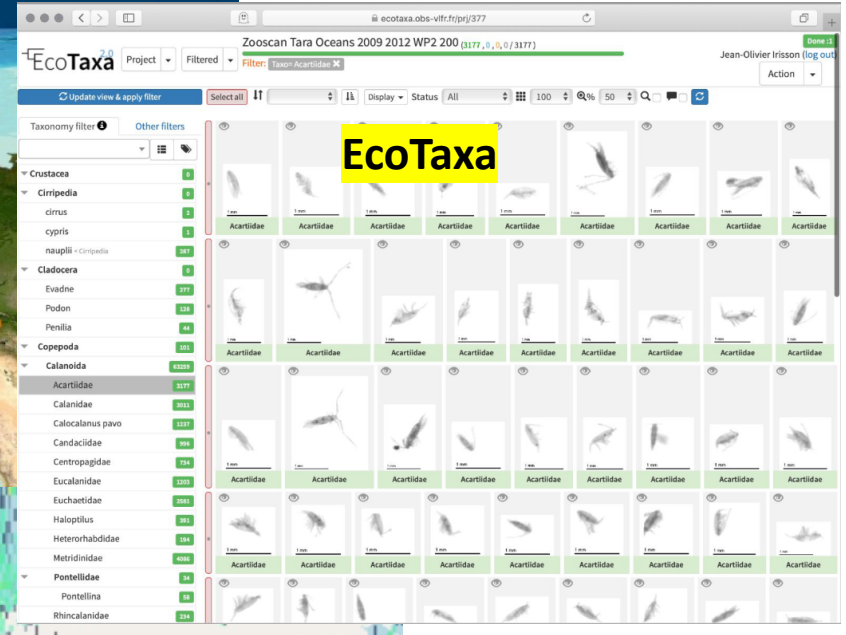
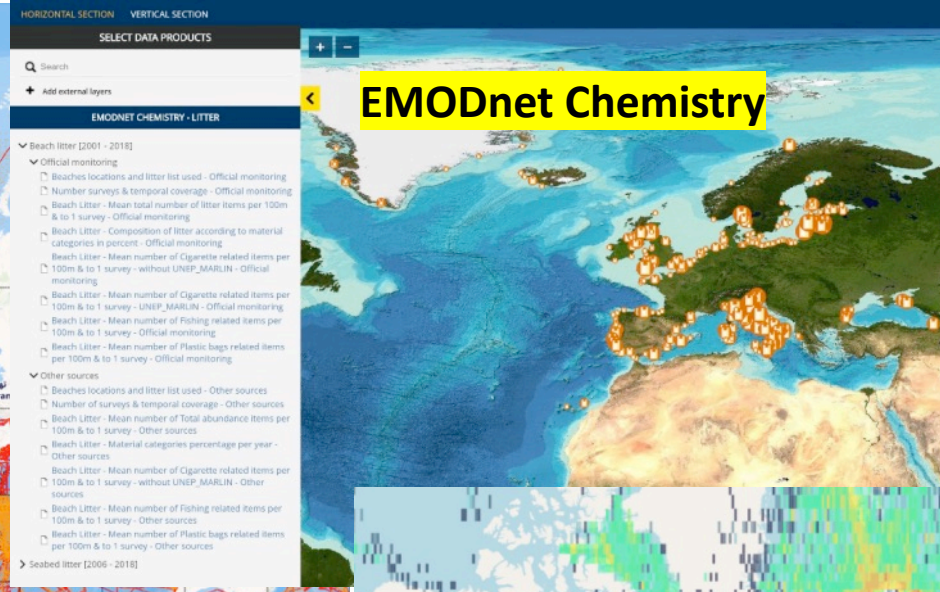
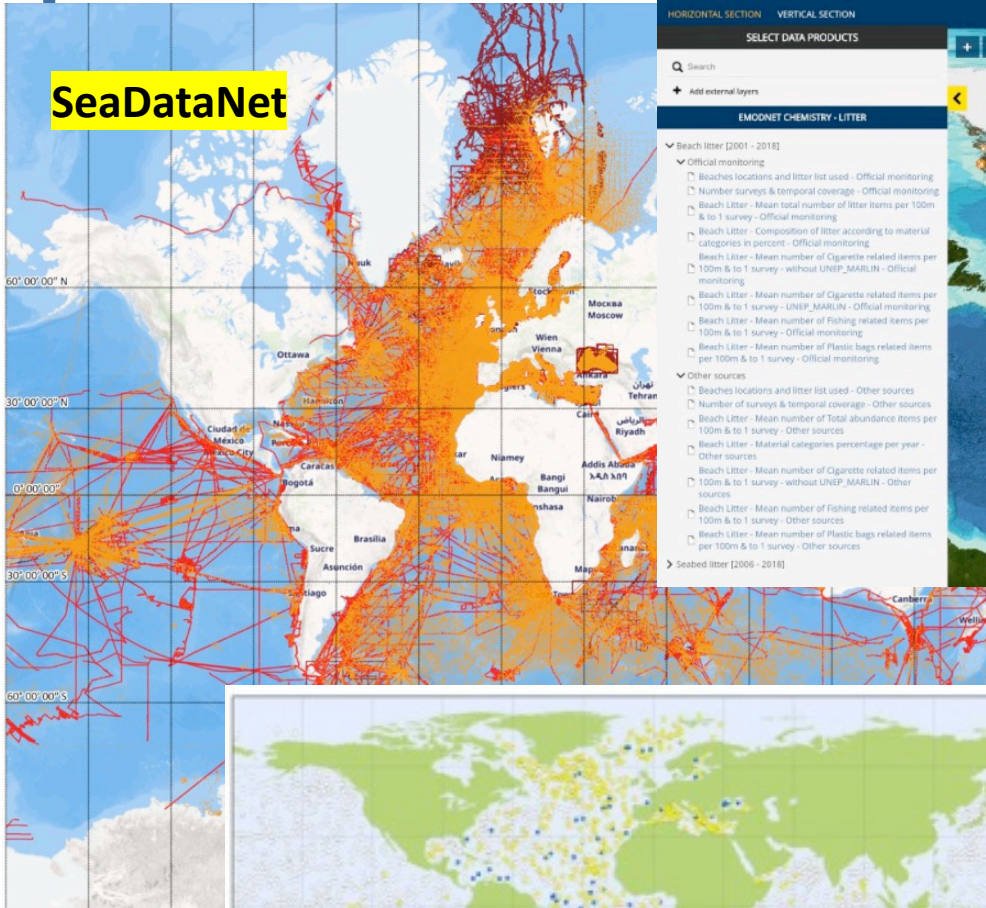
- Federated search for discovering interesting data sets (currently more than 10 million) in a common way
- Federated retrieval of identified data sets using a shopping basket mechanism
- Download of data sets or push to Blue-Cloud VRE

- **Facilitates managers of Blue Data Infrastructures:**

- Wider outreach to potential users
  - Stay informed about data requests and users for their repository
  - Periodic reporting of downloads from their repository
- Built and operated by MARIS, CNR-IIA, and CINECA (EUDAT)



# Illustrations of data coverage







# Approach

- **Federated discovery and retrieval** of data sets and data products from the Blue Data Infrastructures
- **Concept of two-step search approach:**
  - First step: identifying interesting data collections and products with few criteria
  - Second step: drilling down with more criteria to select specific data at granule level, where possible, otherwise at collection/products level
- Metadata and Data Brokerage services interacting **Machine-to-Machine** with web services and APIs as provided and operated by the Blue Data Infrastructures

## Filter search

Free search

Date search

From

yyyymmdd

To

yyyymmdd

Geographic search

North

West

East

South

SEARCH

RESET

| Blue Data infrastructures                                                           |                           | Level 2 Search | Level 1 Results (25848) | Level 1 Total | Last update |
|-------------------------------------------------------------------------------------|---------------------------|----------------|-------------------------|---------------|-------------|
|    | Ecotaxa                   | Level 2 Search | 10                      | 10            | 2022-06-19  |
|    | ELIXIR-ENA                | Level 2 Search | 32                      | 32            | 2022-06-19  |
|    | EMODnet Chemistry         | Level 2 Search | 210                     | 210           | 2022-06-19  |
|    | EuroArgo - Argo           | Level 2 Search | 16998                   | 16998         | 2022-05-01  |
|    | EurOBIS - EMODnet Biology | Level 2 Search | 1024                    | 1024          | 2022-06-19  |
|    | ICOS data portal          | Level 2 Search | 195                     | 195           | 2022-05-08  |
|    | SeaDataNet                | Level 2 Search | 859                     | 859           | 2022-05-29  |
|    | SeaDataNet-products       | Level 2 Search | 49                      | 49            | 2022-06-19  |
|  | Socat                     | Level 2 Search | 6471                    | 6471          | 2022-05-29  |

<https://data.blue-cloud.org>

## Filter search

You searched for:

Reset all

**EuroArgo - Argo**



Free search

Date search

From

yyyyymmdd

To

yyyyymmdd

Geographic search

North

West

East



South

SEARCH

RESET

← Back to search

Results found: 16998

⏪ ⏩ (1 - 100) ⏪ ⏩

100

500

1000

VIEW MAP



|                                                                                     | ID    | Title                                                | Start date | End date | Level2                          |
|-------------------------------------------------------------------------------------|-------|------------------------------------------------------|------------|----------|---------------------------------|
|    | 78501 | 6461 - 5905972 - Argo UW-SOCCOM                      | 20180830   | 20190928 | <a href="#">Level 2 details</a> |
|    | 78843 | APEX Profiling Float - 4901132 - Argo Canada         | 20090715   | 20140109 | <a href="#">Level 2 details</a> |
|    | 78347 | 0587 - 1902334 - Argo JAMSTEC                        | 20191225   | 20220423 | <a href="#">Level 2 details</a> |
|    | 78625 | SOLO Profiling Float - 4900333 - Argo SIO            | 20030331   | 20060628 | <a href="#">Level 2 details</a> |
|    | 78728 | SOLO Profiling Float SBE - 1901435 - Argo SIO        | 20100117   | 20160426 | <a href="#">Level 2 details</a> |
|    | 79669 | APEX Profiling Float SBE 5389 - 2901303 - Argo INDIA | 20110212   | 20200425 | <a href="#">Level 2 details</a> |
|    | 78812 | NAVIS-A-SBE - 2902536 - Argo JAMSTEC                 | 20150601   | 20200914 | <a href="#">Level 2 details</a> |
|    | 78723 | APEX Profiling Float IR_SBE_5374 - 5903740 - Argo UW | 20120305   | 20171125 | <a href="#">Level 2 details</a> |
|  | 78889 | PROVOR Profiling Float - 6901881 - Argo GREECE       | 20131101   | 20140901 | <a href="#">Level 2 details</a> |
|  | 79762 | SOLO Profiling Float SBE - 5903757 - Argo SIO        | 20110821   | 20160320 | <a href="#">Level 2 details</a> |
|  | 78817 | APEX Profiling Float - 3900500 - Argo UW             | 20051217   | 20200603 | <a href="#">Level 2 details</a> |

<https://data.blue-cloud.org>



## Filter search

You searched for: Reset all

EuroArgo - Argo

Free search

Date search

From

To

Geographic search

North

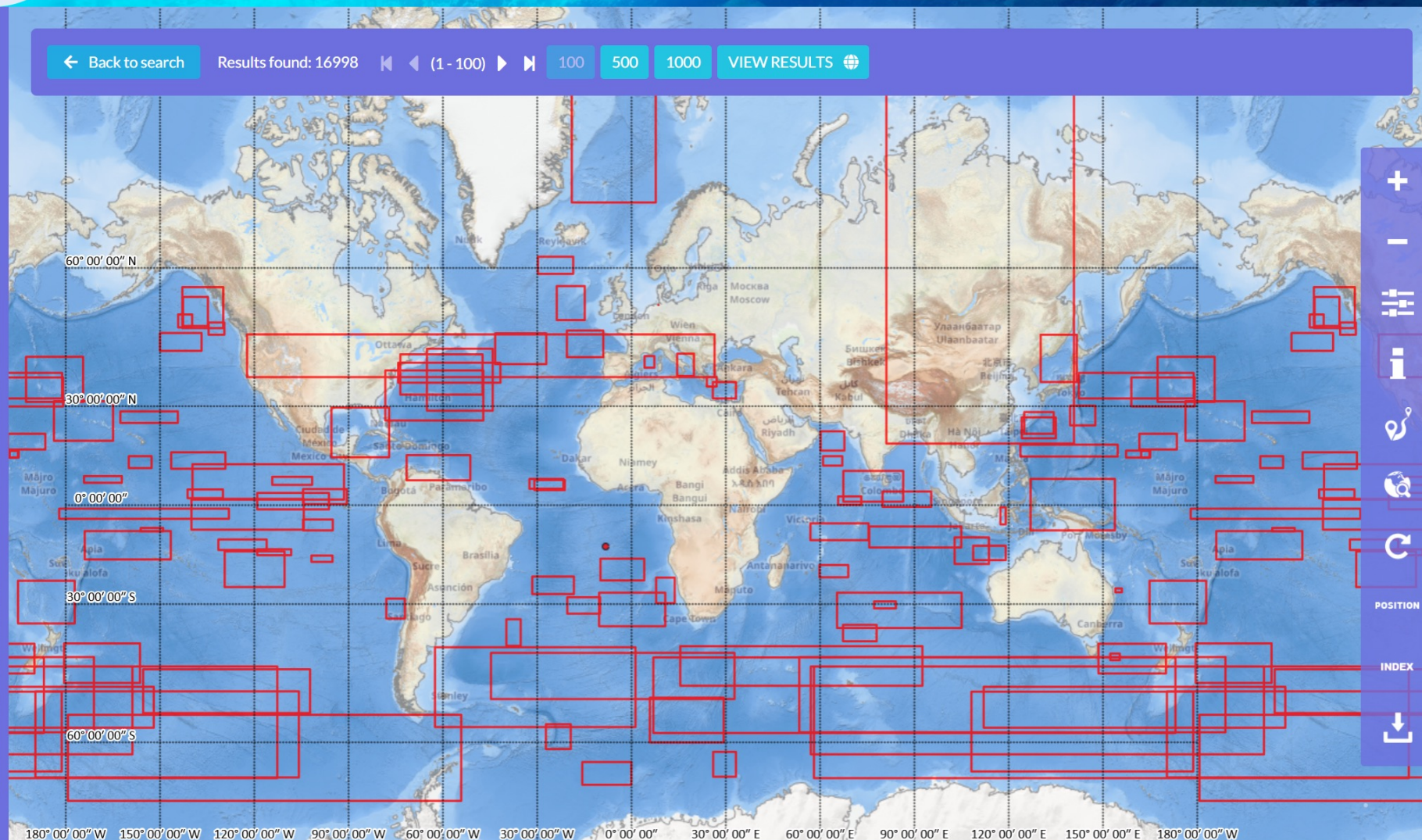
West

East

South

SEARCH

RESET



<https://data.blue-cloud.org>





## Filter search

You searched for:

7900212

Reset all



Free search

Date search

From

yyyymmdd

To

yyyymmdd

Geographic search

North

West

East



South

SEARCH

RESET

## Variables

SUBSURFACE PRESSURE 296

SUBSURFACE SALINITY 296

SUBSURFACE TEMPERATURE 296

Results found: 296 in 0.07 Seconds



(1 - 100)



100

296

VIEW MAP



|                       | Platform     | Platform Code | Transmission System | Cycle | Cycle Start Date         |
|-----------------------|--------------|---------------|---------------------|-------|--------------------------|
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 255   | 2021-08-13T07:27:14.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 194   | 2020-05-04T02:05:25.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 133   | 2018-09-20T00:49:50.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 135   | 2018-10-10T02:49:11.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 241   | 2021-04-24T12:26:02.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 119   | 2018-05-02T12:38:06.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 93    | 2017-08-14T14:20:14.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 33    | 2015-12-22T18:00:27.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 289   | 2022-04-24T07:07:58.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 229   | 2021-01-23T05:32:11.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 211   | 2020-09-12T15:23:41.000Z |
| <input type="radio"/> | SOLO2IR_SBE_ | 7900212       | IRIDIUM             | 73    | 2017-01-26T12:34:37.000Z |

<https://data.blue-cloud.org>





### Filter search

You searched for: **7900212** [Reset all](#)

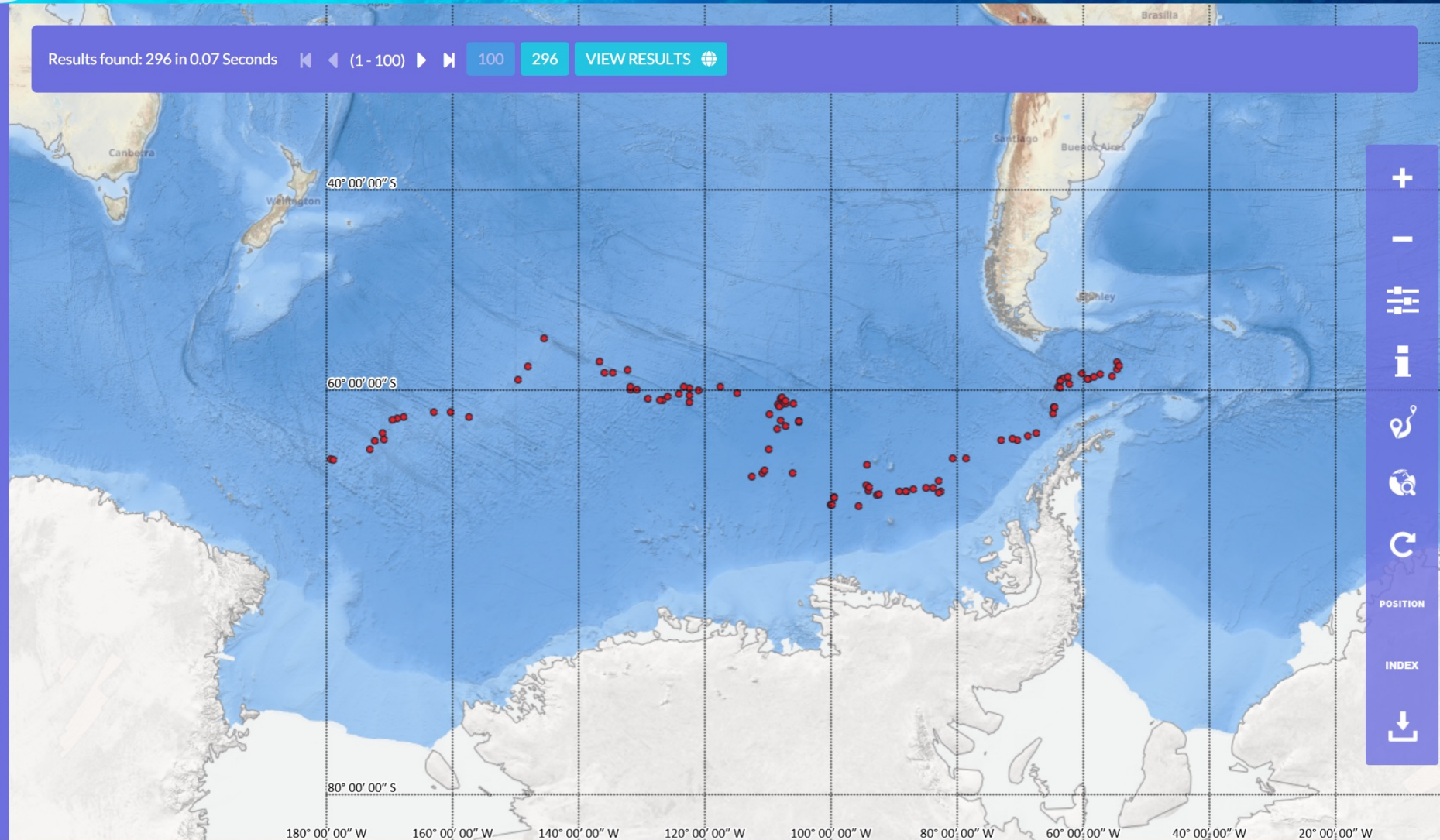
Free search

Date search  
From   
To

Geographic search

Variables

|                        |     |
|------------------------|-----|
| SUBSURFACE PRESSURE    | 296 |
| SUBSURFACE SALINITY    | 296 |
| SUBSURFACE TEMPERATURE | 296 |



<https://data.blue-cloud.org>





# Level 2 Interfacing - Protocols

|                           |                 |  |
|---------------------------|-----------------|--|
| SeaDataNet                | Dedicated API   |  |
| SeaDataNet Products       | OGC CSW service |  |
| EMODnet Chemistry         | OGC CSW service |  |
| EuroArgo - Argo           | Dedicated API   |  |
| EurOBIS – EMODnet Biology | DCAT service    |  |
| Ecotaxa                   | Dedicated API   |  |
| ELIXIR - ENA              | Dedicated API   |  |
| ICOS Marine               | SPARQL service  |  |
| SOCAT                     | ERDDAP service  |  |

# Virtual Research Environment (VRE)

Built and operated by CNR-ISTI (D4SCIENCE)  **D4SCIENCE**





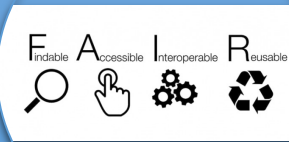
# Blue-Cloud VRE



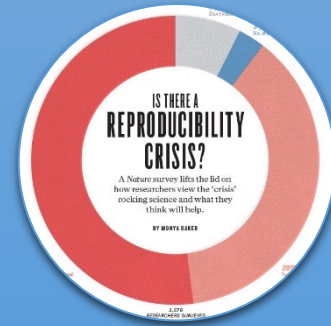
Collaboration



Sharing



Reuse



Reproducibility

Blue Cloud VRE promotes Open Science



# Blue-Cloud VRE

## System of System

Blue-Cloud VRE is built with dedicated services leveraging on existing e-infrastructures and marine infrastructures, EOSC resources and services

## Extensible

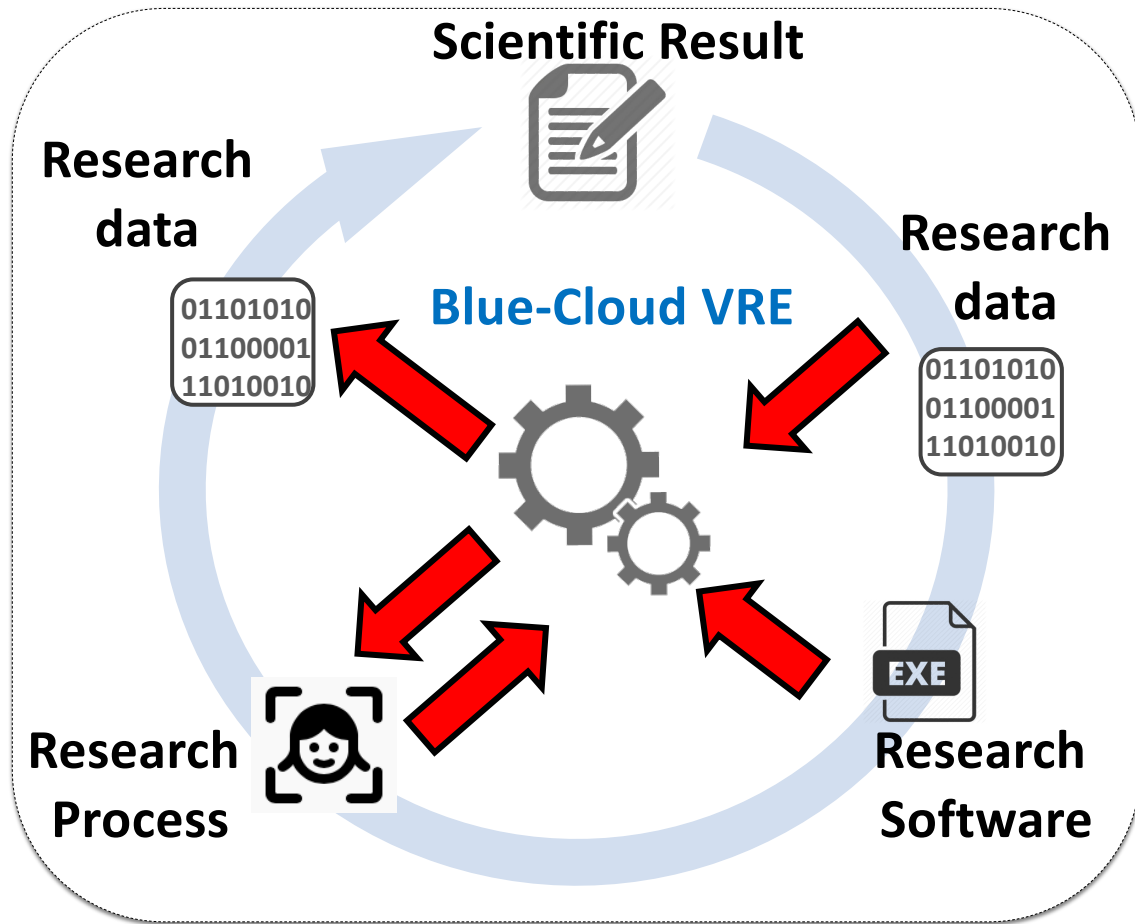
Blue-Cloud VRE can integrate services and resources resulting from existing marine initiatives

## Open

Blue-Cloud VRE promotes open science and practices



# A SoS to support and promote Open Science



## Enable

- Repeat, Reproduce, Reuse, Evaluate
- Active collaboration
- Effective sharing
- Provenance and attribution

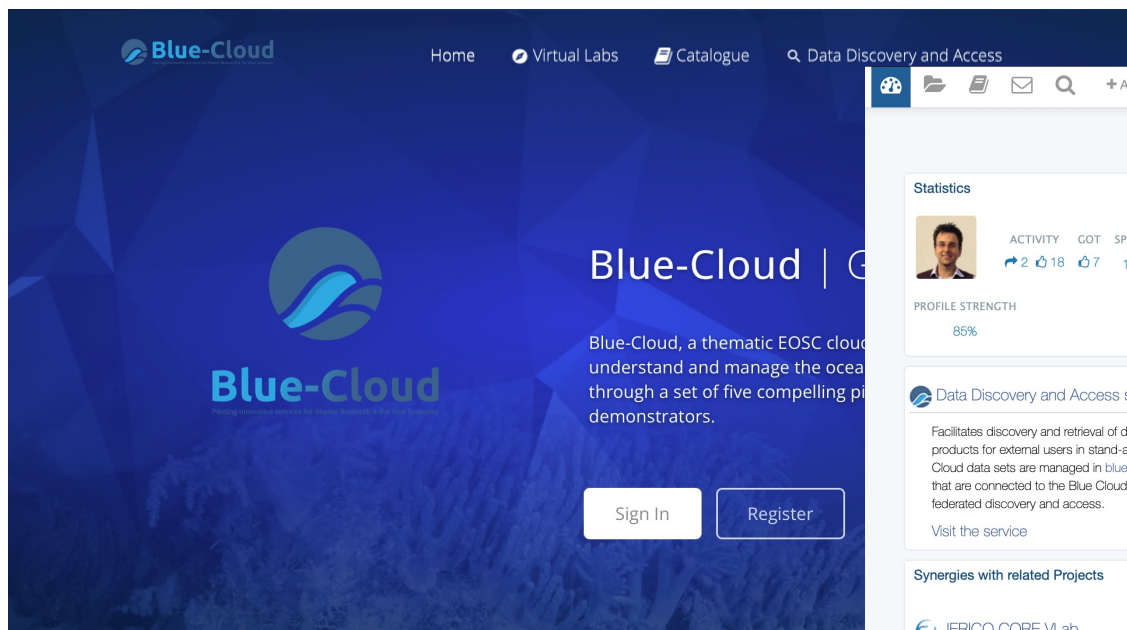
## Adopt

- As-a-service approach
- Standards
- Economy-of-scale to reduce operational costs

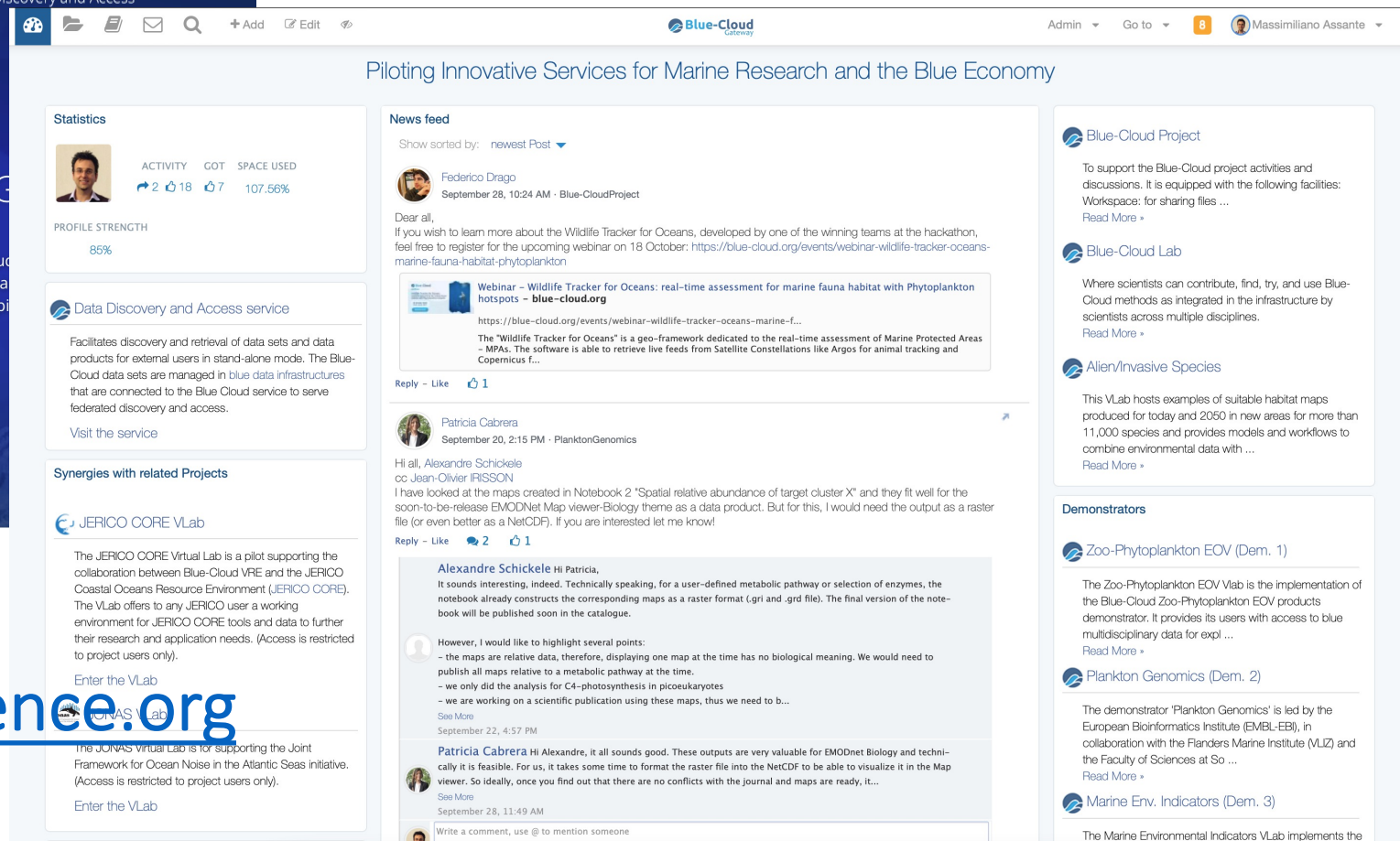


Blue-Cloud

# VRE at-a-glance



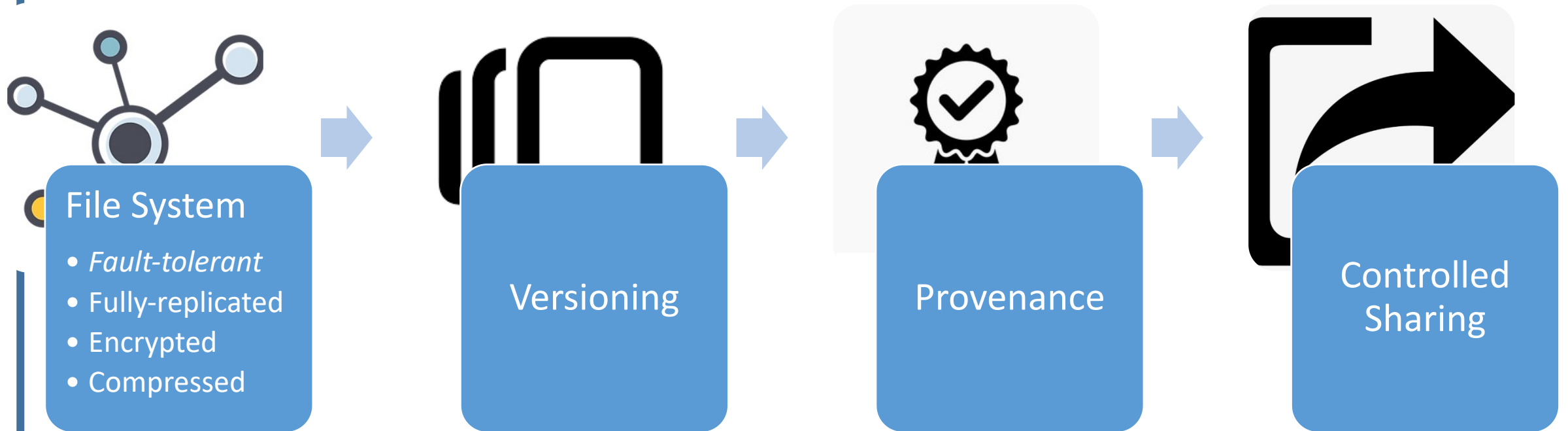
<https://blue-cloud.d4science.org>





# One place for store, share, and preserve datasets

- Common **workspace** and **dataspace** to easily exploit technologies and services not designed to work together



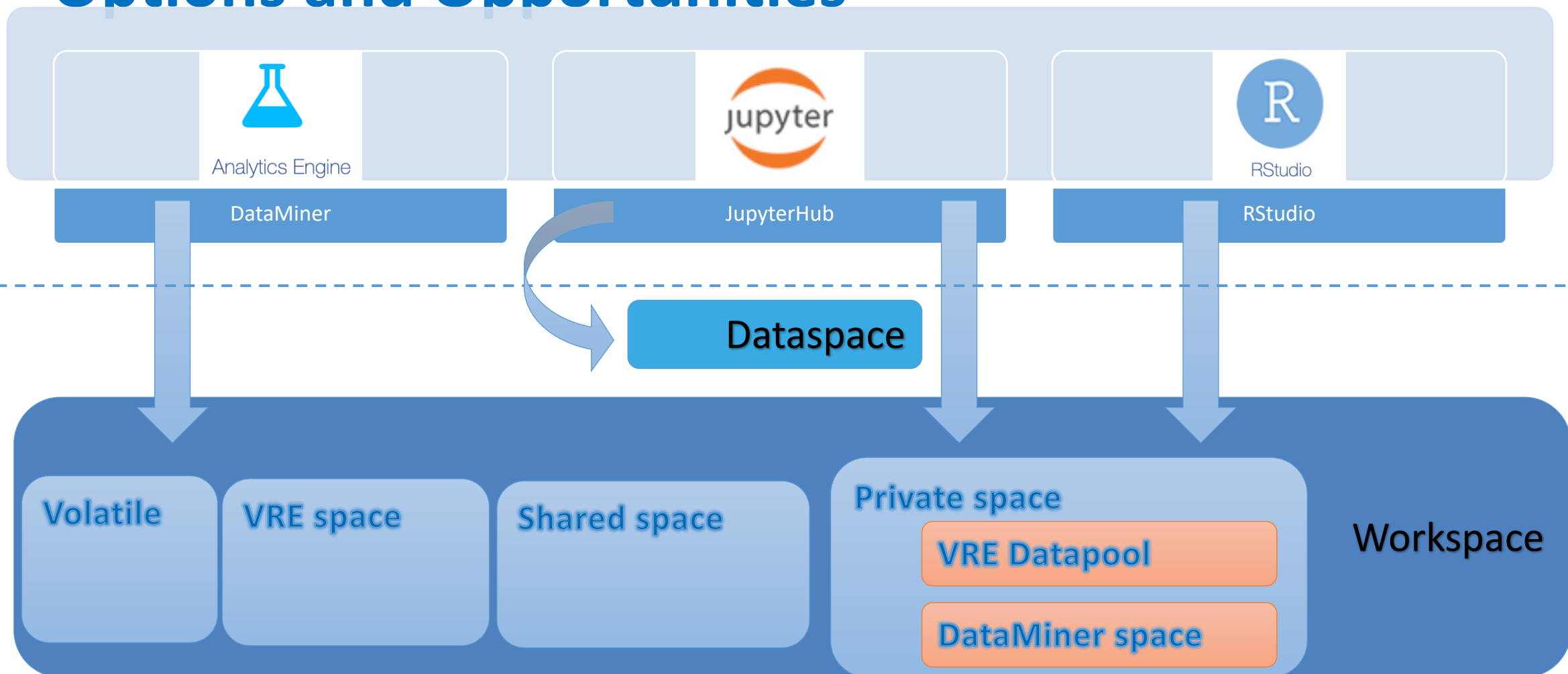




# Data Analytics Options and Opportunities

Computing Layer

Storage Layer





# One place for execute analysis and processes

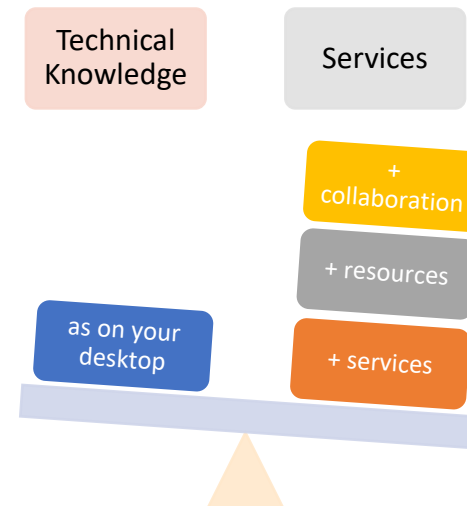
- interactive notebooks via JupyterHub and community-specific applications delivered as a Docker container extend the Analytics framework

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  <p>JupyterHub</p> <p>JupyterHub enables the exploitation of computational environments and resources without burdening users with installation and maintenance tasks. This JupyterHub environment is (i) preconfigured with libraries and packages to ease the execution of common data analytics tasks, and (ii) provides access to the Workspace enabling sharing of resources with other members much easier.</p>                        |  <p>RStudio</p> <p>RStudio provides an integrated development environment for R. It includes a console and a syntax-highlighting editor and it enables code execution. Tools for plotting are also included. This RStudio environment is (i) preconfigured with libraries and packages to ease the execution of common data analytics tasks, and (ii) provides seamless access to the Workspace enabling sharing of resources with other members much easier.</p>                     |
|  <p>Analytics Engine</p> <p>Analytics Engine (DataMiner) permits the execution of an array of analytics methods by transparently relying on distributed computing infrastructure. Executions can run either on multi-core machines or on different computational platforms, such as D4Science and other different private and commercial Cloud providers. New software can be integrated by using the dedicated Software Importer (SAI).</p> |  <p>Catalogue</p> <p>Catalogue contains a list of dataset and products produced by the Blue-Cloud Virtual Laboratories and the methods used to generate such products. All the Catalogue items are accompanied with rich descriptions capturing general attributes, e.g. title and creator(s); accessibility properties; technical properties, e.g. size and format; legal and ethical attributes, e.g. whether containing personal data; intellectual properties, e.g. licences.</p> |

Development and integration environment for R, Python, and other supported software languages

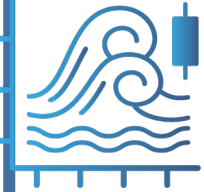


- it is powered by a cluster of DataMiner servers, each with 16 cores and 32 GB RAM.
- It is powered by a cluster of RStudio servers, each with 16 cores and 32 GB RAM.
- It is powered by JupyterHub with a maximum of 8 cores and 32 GB RAM per notebook. Jupyterl





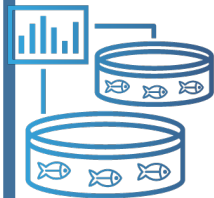
# 5 Virtual Labs at the VRE



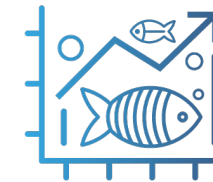
**Marine Environmental Indicators**



**Zoo & Phytoplankton EOVS products**



**Aquaculture Monitor**



**Fish a matter of scales**

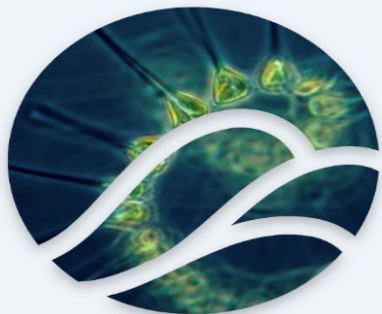


**Plankton Genomics**





# Vlabs for different domains



Biodiversity

Zoo and Phytoplankton EOVI products



Genomics

Plankton Genomics



Environment

Marine Environmental Indicators



Fisheries

Global Record of Stocks and Fisheries



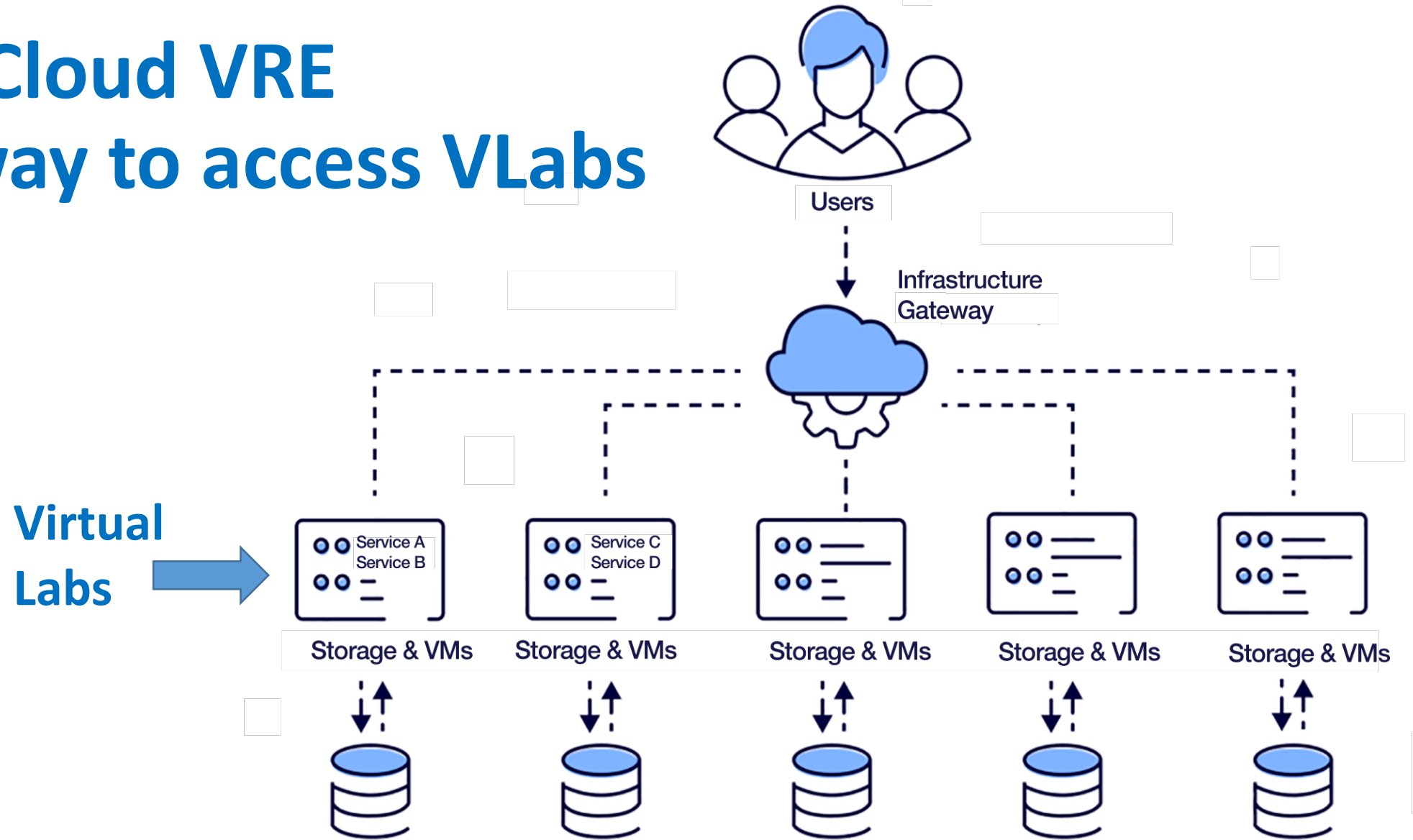
Aquaculture

Aquaculture Monitor



Blue-Cloud

# Blue-Cloud VRE gateway to access V Labs





Blue-Cloud

# VLabs at-a-glance

The collage displays several Blue-Cloud VLab interfaces:

- Use cases interface:** Shows a sidebar with categories like 'greece', 'malta', 'South\_sulawesi', 'Coastal Ponds / Rice Paddy by Sentinel-1/2 (2014-2017)', 'Coastal Ponds by Landsat 5 (1994-1997)', and 'SAR'. The main area shows a map of South Sulawesi with various data layers.
- PlanktonGenomics interface:** Shows a 'JupyterHub' tab with a notebook titled 'Demonstrator 2 - Notebook 1.2. - Creating protein functional clusters for Notebook 2'. The notebook content includes an introduction and a section for Notebook 1.2.
- Blue-Cloud Lab landing page:** Features a grid of icons for Jupyter, RStudio, and Analytics Engine, along with a 'Catalogue' section.
- Marine Environmental Indicators Home:** Shows a map of the Mediterranean Sea with a sidebar for data generation, including fields for 'Data source', 'Type', 'Environmental field', 'Start time', 'End time', 'Area', and 'Depth'.

<https://blue-cloud.d4science.org>



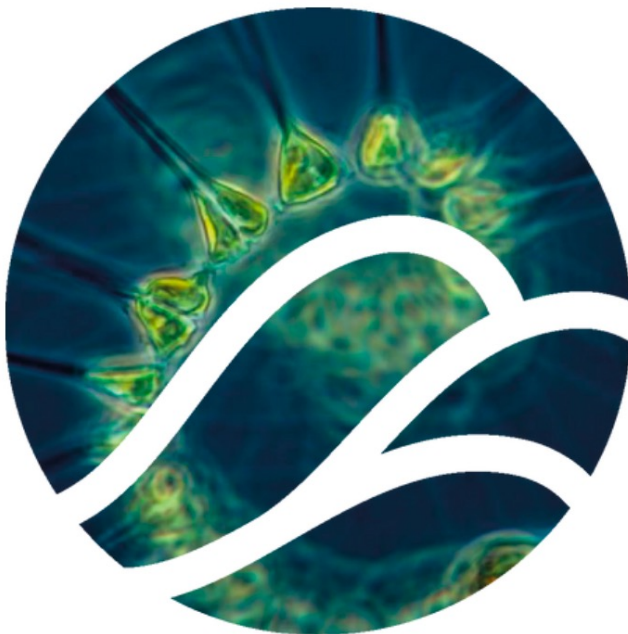
MetroSea  
OCTOBER 3-5, 2022







Blue-Cloud



# Zoo- and Phytoplankton Essential Ocean Variable products

Partners:



Data sources through Blue-Cloud:

EuroBIS, EMODnet Biology, LifeWatch, GEBCO, SeaDataNet, World Ocean Atlas, NOAA, Copernicus Marine Service, Argo GDAC, GlobColour

Test the VLab now!



## Main target users:

Plankton researchers, ocean modellers, data product developers and Blue Data infrastructures, for their data products catalogues and as use cases.

## Services introduction:

The Vlab offers three independent services that consist of the combination of different data types (biological, physical and environmental data) to then apply models that generate an output. These are offered in a working space where data and scripts are accessible and reusable.

## UN SDGs addressed



## SERVICES

### Zooplankton Essential Ocean Variable

Zooplankton EOVS generates zooplankton gridded maps of six zooplankton species in the North East Atlantic. The workflow uses the DIVAnd software tool (Data Interpolating Variational Analysis in n dimensions) that allows to interpolate sparse in situ measurements onto a regular

### Phytoplankton Essential Ocean Variable

Phytoplankton EOVS generates global open ocean 3D gridded products of (1) chlorophyll a concentration (Chl<sub>a</sub>), which is a proxy of the total phytoplankton biomass, and (2) Phytoplankton Functional Types (PFT), as a proxy for phytoplankton diversity, based on temperature and salinity in situ data.

### Modelling phyto & zooplankton interactions

Modelling phyto and zooplankton interactions enables users to calculate the relative contribution that limits the growth of phytoplankton by the drivers: nutrients, phosphates, silicates, light and zooplankton grazing.

Milazzo 2022



# Successor project: Blue-Cloud 2026



## Blue-Cloud 2026

A Federated European FAIR and Open Research Ecosystem  
for oceans, seas, coastal and inland waters

**MISSION:** To develop further the European federation of marine and inland water data management infrastructures & high quality services



### A1. DD&AS

A FAIR compliant Data Discovery & Access Services > access to 10+ million open data sets & products by 13 major BDIs



### A2. VRE

An Open Science Virtual Research Environment (VRE) federating multiple e-infrastructures > supporting Analytical Big Data Workbenches & VLABs



### A3. EOVS

3 EOVS Workbenches for highly qualified data collections

3.000 DATA ANALYTICS SESSIONS PER MONTH - 5.000 HTC DATA ANALYTICS JOBS PER MONTH

### A4. VLABs - FIVE DOMAIN-BASED VIRTUAL LABS



Coastal Ocean observations along Europe



Coastal currents from observations



Carbon-Plankton Dynamics



Marine Environmental Indicators



Global Fisheries Atlas



### A7. COMMUNITY

- All EU countries engaged
- 3k+ engaged Blue-Cloud community users
- 5k+ followers across all the platforms
- 10+ External Stakeholders



### OUTREACH

- 1 Blue-Cloud Hackathon
- 1 Blue-Cloud TV
- 18 Newsletter issues
- 11 Webinars on Blue-Cloud VRE, DDAS & EOVS Workbenches
- 3 Blue-Cloud Annual Impact Events
- 3 Ocean Literacy Webinars
- Videos & Interviews



### A6. TRAINING ACADEMY & CATALOGUE

- 3 Online training course on Best Practices for FAIR data principles
- 3 Info session & course on the EOVS Workbenches
- 2 online webinars dedicated to the BlueCloud VRE
- 2 dedicated to the DDAS and the innovations introduced
- A series of training sessions on how to use the VLABs



### POLICY

- Scientific papers & articles
- Restoring healthy oceans, seas, coastal & inland waters in Europe
- Strategic Roadmap 2030 **A5. ROADMAP**
- Cross-domain expansion factsheets
- Sustainability Business model



DTO Task Force



# Thank you

Massimiliano Assante, CNR, Italy  
[massimiliano.Assante@cnr.it](mailto:massimiliano.Assante@cnr.it)  
Twitter: @maxassante