

10 – 11 March 2025

3rd International Ocean Data Conference

The data we need for the ocean we want

Santa Marta, Colombia

Blue-Cloud workbenches for Essential Ocean Variables (EOVs)

Presented by

Catalina Reyes, OGS (Italy) on behalf of the BlueCloud2026 project



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

- Blue-Cloud 2026 **builds upon the pilot Blue-Cloud project** to further evolve its **pilot ecosystem into a Federated European Ecosystem to deliver FAIR & Open data, analytical services, instrumental for deepening research of oceans, EU seas, coastal & inland waters.**
- It develops a **thematic marine extension to EOSC (European Open Science Cloud)** for open web-based science, serving the needs of the EU Blue Economy, Marine Environment and Marine Knowledge agendas.
- All in all, Blue-Cloud 2026 will expand the federated approach of the previous Blue-Cloud, involving more aquatic data stakeholders, and interacting with EOSC developments, in support of the EU Green Deal, UN SDG, EU Destination Earth, and the EU Mission Starfish on healthy oceans, seas, coastal and inland waters, ultimately to provide a core data service for the European Digital Twin of the Ocean (EDITO).

Project Information

Blue-Cloud 2026

Grant agreement ID: 101094227

DOI

[10.3030/101094227](https://doi.org/10.3030/101094227)

EC signature date

1 December 2022

Start date

1 January 2023

End date

30 June 2026

Funded under

Research infrastructures

Total cost

€ 8 845 420,00

EU contribution

€ 8 845 420,00

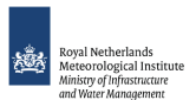
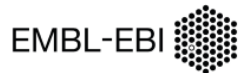
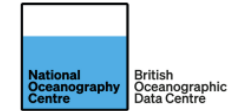
Coordinated by

CONSIGLIO NAZIONALE DELLE RICERCHE

 Italy



PROJECT COORDINATION OFFICE



A solid, multidisciplinary, committed team of 40 partners from 14 EU countries








- Developing and deploying **Virtual Research Environments (VREs)** with an array of services for configuring and running specific analytical workflows, use cases and demonstrators.
- Applying **common standards and interoperability** solutions to provide harmonized metadata and data.
- Developing and deploying **harmonized discovery and access services to established EU marine data management** and processing infrastructures.

Services

-  **Data Discovery & Access Service**
-  **Virtual Research Environment**
-  **Blue-Cloud Catalogue**

Thematic Virtual Labs

-  Coastal Ocean observations along Europe
-  Coastal currents from observations
-  Carbon-Plankton Dynamics
-  Marine Environmental Indicators
-  Global Fisheries Atlas

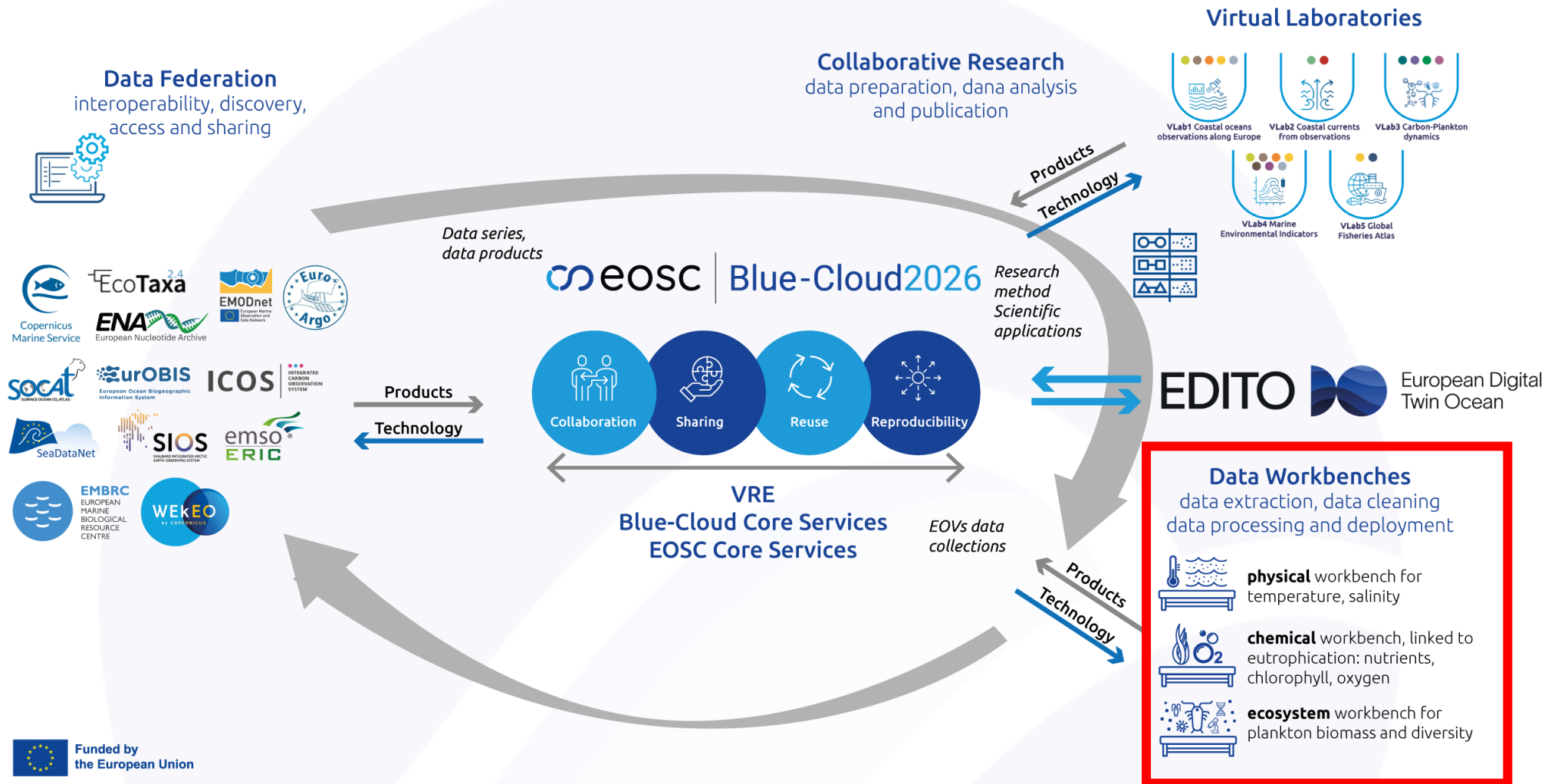
Essential Ocean Variables Workbenches

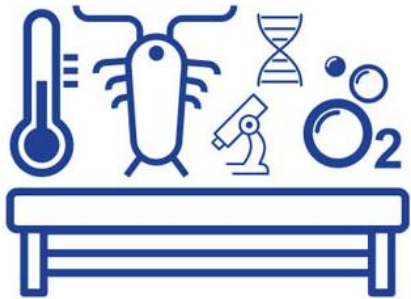
-  Physics: temperature & salinity
-  Eutrophication: chlorophyll, nutrients, oxygen
-  Ecosystem-level EOVs



Scan the QR Code to visit our website

Blue-Cloud open science platform





Essential Ocean Variables Workbenches

Blue-Cloud is working on three big data processing Workbenches (WB). These will facilitate the generation of **validated and harmonized data collections for a selection of Essential Ocean Variables (EOVs) in physics, chemistry and ecosystems** filling the existing gaps by integrating several datasets from different EU and non-EU Blue Data Infrastructures (BDIs) and providing EOVs datasets and workflows to EU operational services and the Digital Twin of the Ocean (DTO).



Ecosystem-level EOVs

Enhancing plankton data quality & accessibility via EMODnet, EurOBIS & ELIXIR for better insights.



Eutrophication: chlorophyll, nutrients, oxygen

Creating an efficient workflow to merge data from Copernicus, EMODnet, World Ocean Database & key EU RIs for precise eutrophication EOv datasets.



Physics: temperature & salinity

Cloud-based workflow generates customisable EOv data for Mediterranean from EU & non-EU sources.

BEACON: High performance data lake for sub-setting of big datasets

The new Beacon (©MARIS) data lake technology has been implemented to allow data sub-setting and extraction services for the BC26 WB, VLabs and beyond.

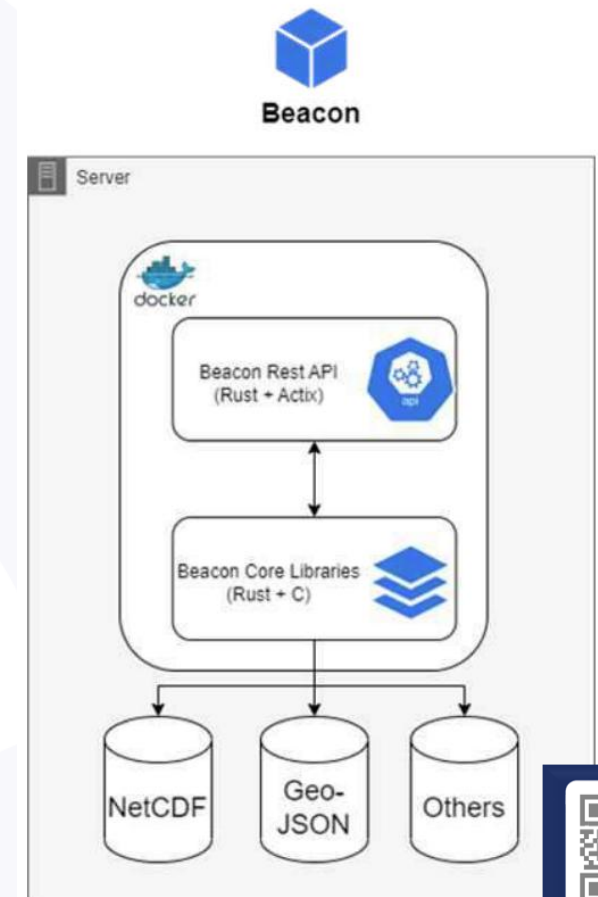
Beacon software system comes with a unique indexing system that can, on the fly, extract specific data based on the user request from millions of observations data files. Beacon exposes a REST API so that clients can query data via simple JSON request, integrated in jupyter notebooks.

The system returns one single harmonized file as output allowing extraction of subsets and exporting these in one coherent file within seconds.

Beacon Instances

Two main use cases have been formulated:

- A number of monolithic Beacon instances, each for selected data from a specific BDI, and made available for all Blue-Cloud VRE users.
- Two integrated Beacon instances for Workbench 1 and Workbench 2, merging and harmonizing selected data from multiple BDIs, and with regular control.



Scan to visit
beacon.maris.nl

<https://github.com/maris-development/beacon-blue-cloud/tree/main/notebook-examples>

The screenshot shows a Jupyter Notebook environment. On the left is a file explorer with a sidebar containing a file tree. The tree shows a 'main' branch with a '1.0.0 (latest)' folder, which contains several notebook files. The selected file is 'WB2 - Eutrophication - Merged Demo Notebook 1.0.0 - Arrow Quicker.ipynb'. The main area displays the notebook's code cell 'In [8]:' and its output 'Out[8]:'. The code cell contains Python code for reading a feather file, creating a directory, and processing a DataFrame. The output is a DataFrame with columns: COMMON_PHOSPHATE, Unit, COMMON_PHOSPHATE_qf, COMMON_PHOSPHATE_P01, COMMON_PHOSPHATE_P06, and DEP. The DataFrame is sorted by datetime.

```
In [8]:
regionname = f"[{minlat},{minlon}],[{maxlat},{maxlon}]"

if not os.path.exists("./Beacon_V1.0.0_Output"):
    os.makedirs("Beacon_V1.0.0_Output")

open(
    f"./Beacon_V1.0.0_Output/EWB_merged_{parameter}_{regionname}_{mindate}-{maxdate}_{mindepth}-{maxdepth}m.arrow", "wb")
df = pd.read_feather(
    f"./Beacon_V1.0.0_Output/EWB_merged_{parameter}_{regionname}_{mindate}-{maxdate}_{mindepth}-{maxdepth}m.arrow")
df = df.assign(datetime=pd.to_datetime(
    df["datetime"])).set_index("datetime").sort_index()
df
```

```
Out[8]:
```

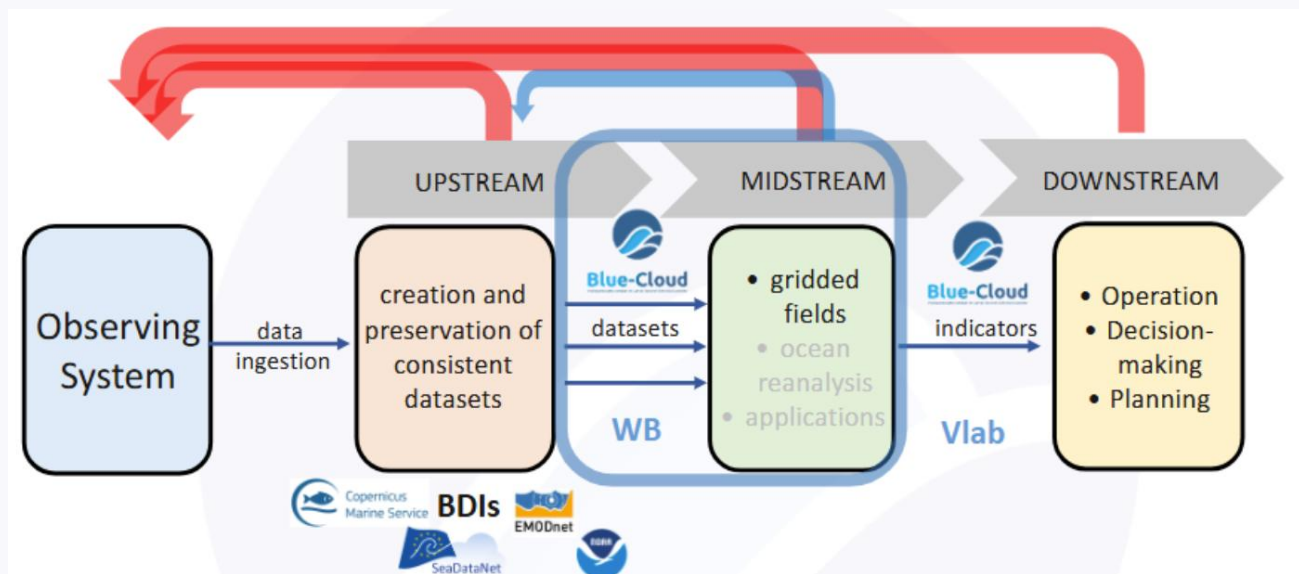
datetime	COMMON_PHOSPHATE	Unit	COMMON_PHOSPHATE_qf	COMMON_PHOSPHATE_P01	COMMON_PHOSPHATE_P06	DEP
2010-01-04 00:00:00	2.390000	mmol m-3	1.0	SDN:P01::PHOSZZXX	SDN:P06::UPOX	1.0000
2010-01-04 00:00:00	3.100000	mmol m-3	1.0	SDN:P01::PHOSZZXX	SDN:P06::UPOX	1.0000
2010-01-04 00:00:00	4.000000	mmol m-3	1.0	SDN:P01::PHOSZZXX	SDN:P06::UPOX	1.0000
2010-01-04 07:35:59	0.400000	umol/kg	1	SDN:P01::MDMAP906	SDN:P06::KGUM	0.0000
2010-01-04 07:35:59	0.390244	umol/kg	1	SDN:P01::MDMAP906	SDN:P06::KGUM	9.9061

Semantic harmonization

Semantic harmonization of the incoming data is implemented within Beacon based on mappings hosted on the NVS (NERC Vocabulary Server) or derived by NOC-BODC, assisted by the newly developed Semantic Analyzer software. In a first instance, the primary focus was harmonization of platforms, instruments, parameters and units.

Visualization, QC & Duplicates

The webODV service enables data exploration and visualization, which is useful for the initial validation of the merged beacon datasets. Duplicates are analyzed using webODV and clone-wars, a web application developed to analyze duplicates and for further quality control.

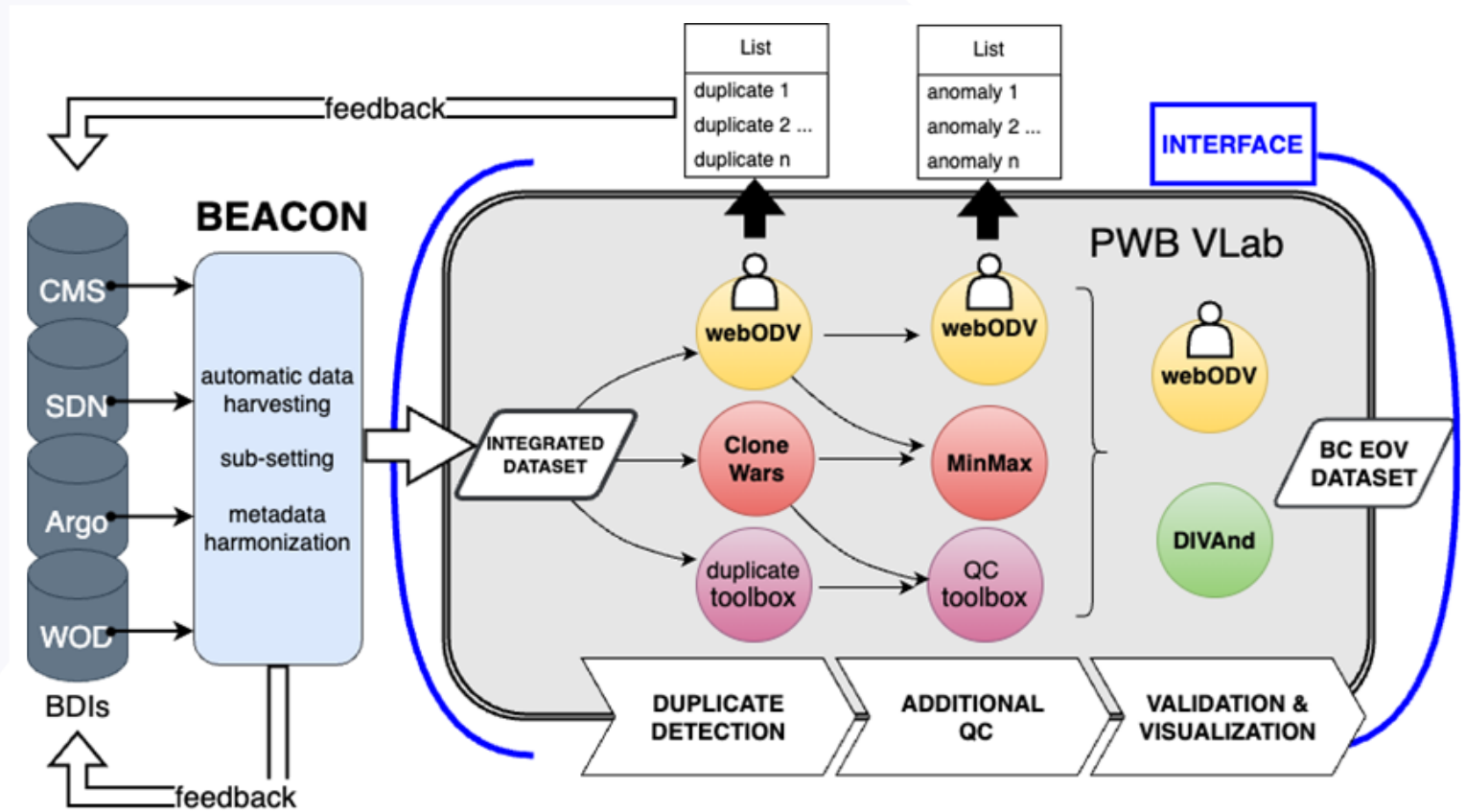




The integration of the data coming from different BDIs relies on the interoperability of the data discovery and access services and the different metadata associated with the Temperature and Salinity observations, which will be mapped to a common metadata schema.

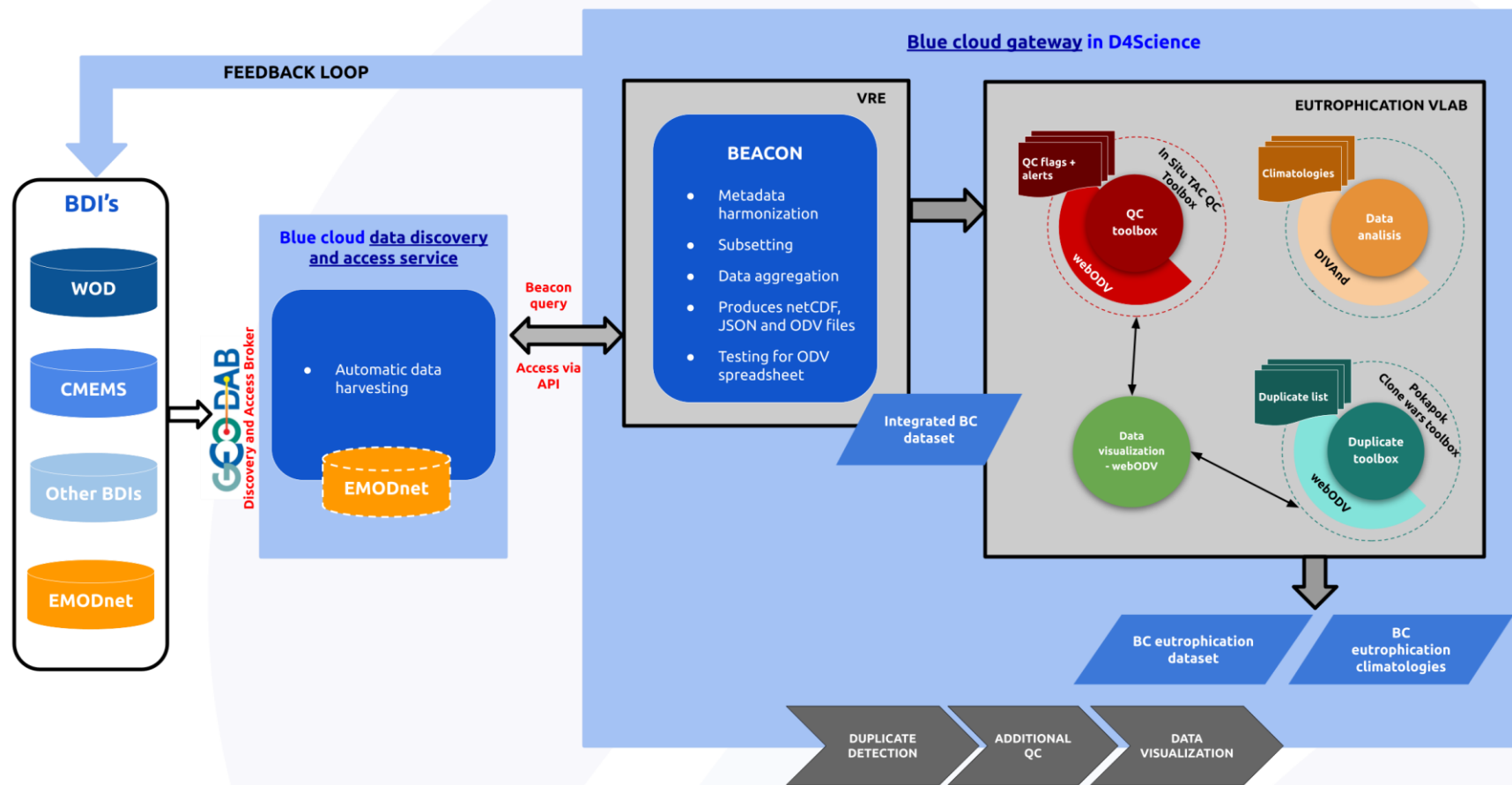
Metadata harmonization allows to identify and handle potential duplicate observations. A further Quality Control procedure will provide an added-value and consistent EOv dataset.

The PWB will be developed and tested for the Mediterranean Sea with the aim of extending it to the global ocean.





Interoperability services of data infrastructures, common vocabularies and brokering services are used to enable the aggregation and harmonisation of eutrophication datasets from **WOD, EMODnet and CMEMS**, with great attention paid to semantic aspects. A specific protocol will be jointly developed to identify and handle possible duplicate observations as well as QC procedures.



Tested for the North-East Atlantic, the aim is to further extend it to the global ocean during the last year of the project.

Go to 19 Nydia Catalina Reyes Suarez ▾

[Administration](#) [CCP](#) [JupyterLab](#) [RStudio](#) [JupyterLab at Google](#) [webODV](#) [Members](#) [How-to](#)

Shared Folder

Eutrophication-Work ... Recent

Name	Owner	Last modified
Catalina_TES ...	me	11 Apr 09:56:24
datasets_dup ...	CW	05 Jul 14:24:24
DIVAnd	me	17 Jan 14:40:25

Show 5 entries Previous 1 Next

1 to 3 of 3 items
[Go to shared workspace](#)

VRE Managers and Groups
[View Managers](#)

Trending Topics
 No Topics found in News Feed

Share updates

Share an update or a link, use "@" to mention and "#" to add a topic

Notify members: OFF ON

[Share](#)

News feed

Show sorted by: newest Post ▾

Nydia Catalina Reyes Suarez
 October 15 2024, 3:43 PM

Dear EutroWB colleagues,

Please find in the following link Beacon monolithic instances:
<https://github.com/maris-development/beacon-blue-cloud/tree/main/notebook-examples>

In case you do not have a token yet, you should write to paul@maris.nl and robin@maris.nl to grant access, after which the token will be available in the DD&AS bluecloud (<https://data.blue-cloud.org/search>).

[beacon-blue-cloud/notebook-examples at main · maris-development/beacon-blue-cloud - github.com](https://github.com/maris-development/beacon-blue-cloud/tree/main/notebook-examples)
<https://github.com/maris-development/beacon-blue-cloud/tree/main/notebook-examples>

Reply - Like

Nydia Catalina Reyes Suarez
 May 13 2024, 10:01 AM

Dear EutroWB colleagues,

Please find in the following link the

- EMODnet Chemistry metadata profile report: <https://dabreporting.s3.amazonaws.com/BlueCloud/emodnet-chemistry.html> and
 - EuroARGO updated metadata profile report: <https://dabreporting.s3.amazonaws.com/BlueCloud/argo.html>

Enrico's presentation is available at the following link: <https://data.d4science.net/qe4g>
 as well as Gwen and Alessandra's: <https://data.d4science.net/oQ9V>

Reply - Like 1

About Eutrophication-Workbench

This Workbench will implement a cloud-based workflow to generate harmonised, validated and customisable EOv data collections for Chlorophyll, nutrients and oxygen integrating several datasets released from different EU and non-EU data infrastructures.

[Edit this text](#)

[Other options ...](#)

Personal Token

Current context: Eutrophication-Workbench

[Get Token](#)

Instructions

Obtain your personal OAuth2 Access, to be used for programmatic interaction with the Resources

Do not use this token to authenticate services but only for personal access ref.

- Accessing Resources - how to
- OAuth2: Making Authenticated Requests

Please note: if you need long lasting access tokens you may obtain a service account by opening a request

<https://blue-cloud.d4science.org/group/eutrophication-workbench/eutrophication-workbench>

INSTRUMENT Values Counts

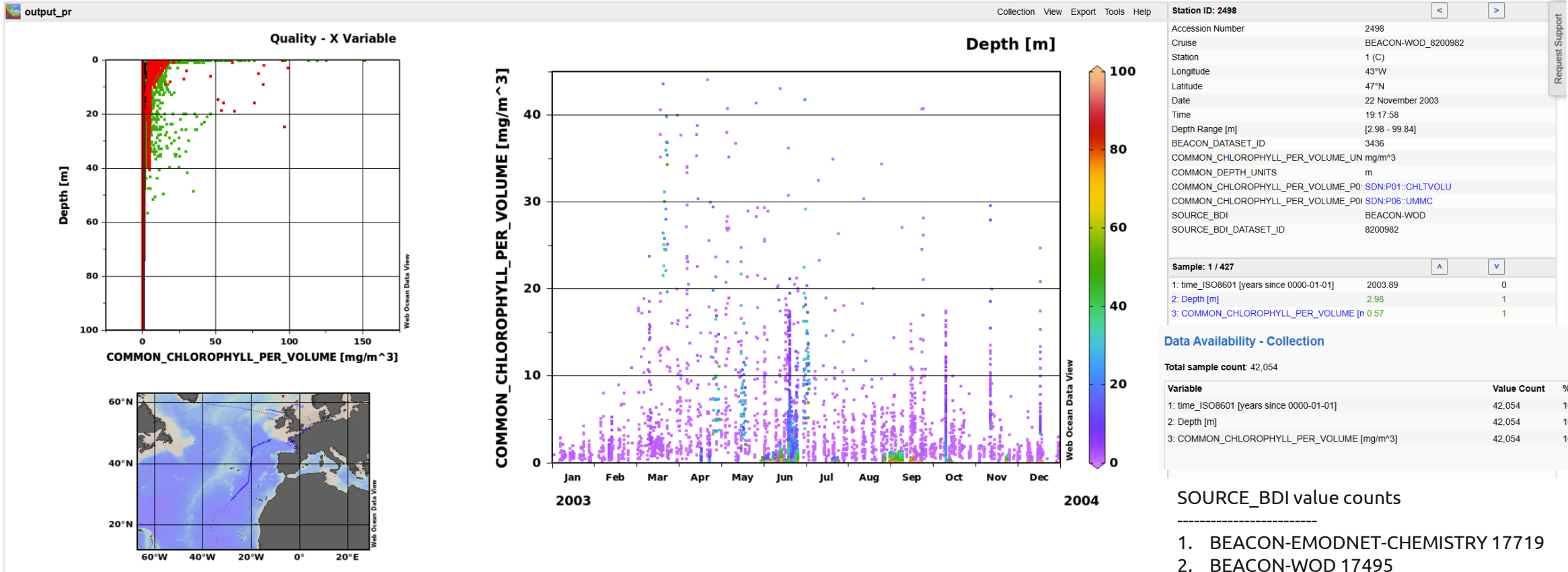
- 1: (empty) 961 → **Argo**
- 2: CTD 409
- 3: Expendable bathythermographs 81
- 4: bathythermographs 129
- 5: discrete water samplers 50
- 6: discrete water samplers;dissolved gas sensors;observers;pH sensors;salinity sensor;water temperature sensor 130
- 7: discrete water samplers;dissolved gas sensors;pH sensors;salinity sensor;water temperature sensor 14
- 8: fluorometers;thermosalinographs 118
- 9: salinity sensor;satellite tracking system;water temperature sensor 17
- 10: salinity sensor;water temperature sensor 1151
- 11: thermosalinographs 22548
- 12: water temperature sensor 1

SOURCE BDI Values Counts

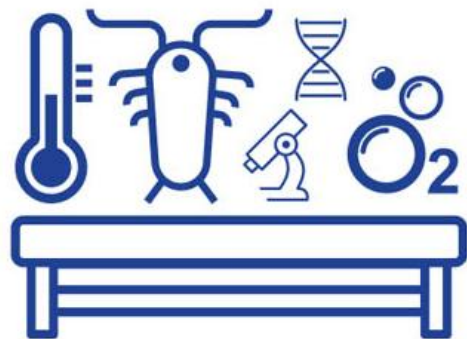
- 1: BEACON-ARGO 961
 - 2: BEACON-SEADATANET-TS 24648
-



Selected dataset: eut-wb -> output_pr.odv



- One of the goals is to develop, validate and document new analytical big data workbenches to produce a set of harmonized and validated data collections for a selection of EOVs in the fields of physics, chemistry and biology.
- Beacon (<https://beacon.maris.nl/>) provides a high-performance data lake solution for storing, subsetting and rapidly querying large amounts of data.
- The workbenches aim to integrate and harmonize data and metadata from different blue data infrastructures (Copernicus Marine Service, CMEMS; European Marine Observation and Data Network, EMODnet-Chemistry; World Ocean Database, WOD; ARGO, SeaDataNet,...) by combining Beacon with semantic harmonization and mappings.
- The webODV service enables data exploration and visualization, which is useful for the initial validation of the merged beacon datasets. Duplicates are analyzed using webODV and clone-wars, a web application developed to analyze duplicates and for further quality control.
- The workbenches cover a wide range of topics related to the marine environment, ecosystems and data types (physical, biological, chemical, in-situ, remote and from autonomous vehicles data, etc.) relevant for assessing the state of the oceans and generating knowledge beyond marine science.



Essential Ocean Variables Workbenches in physics and chemistry working group



Alessandra Giorgetti

Simona Simoncelli

Julie Gatti

Sissy Iona

Tanguy Szekely

Robin Kooyman

Sebastian Mieruch

Catalina Reyes

Marine Vernet

Christine Coatanoan

Virginie Racape

Gwenaelle Moncoiffe

Delphine Leroy

Megan Anne French

WP Leader, EWB, PWB, webODV & BEACON

THANKS FOR YOUR ATTENTION

eosc | Blue-Cloud2026



blue-cloud.org



[@bluecloudeu](https://twitter.com/bluecloudeu)



[blue-cloud org](https://www.linkedin.com/company/blue-cloud-org)



Funded by
the European Union

3rd International Ocean Data Conference

Santa Marta, Colombia

10 – 11 March 2025

Blue-Cloud workbenches for Essential Ocean Variables (EOVs)

Presented by Catalina Reyes (OGS, Italy) on behalf of the BlueCloud2026 project

