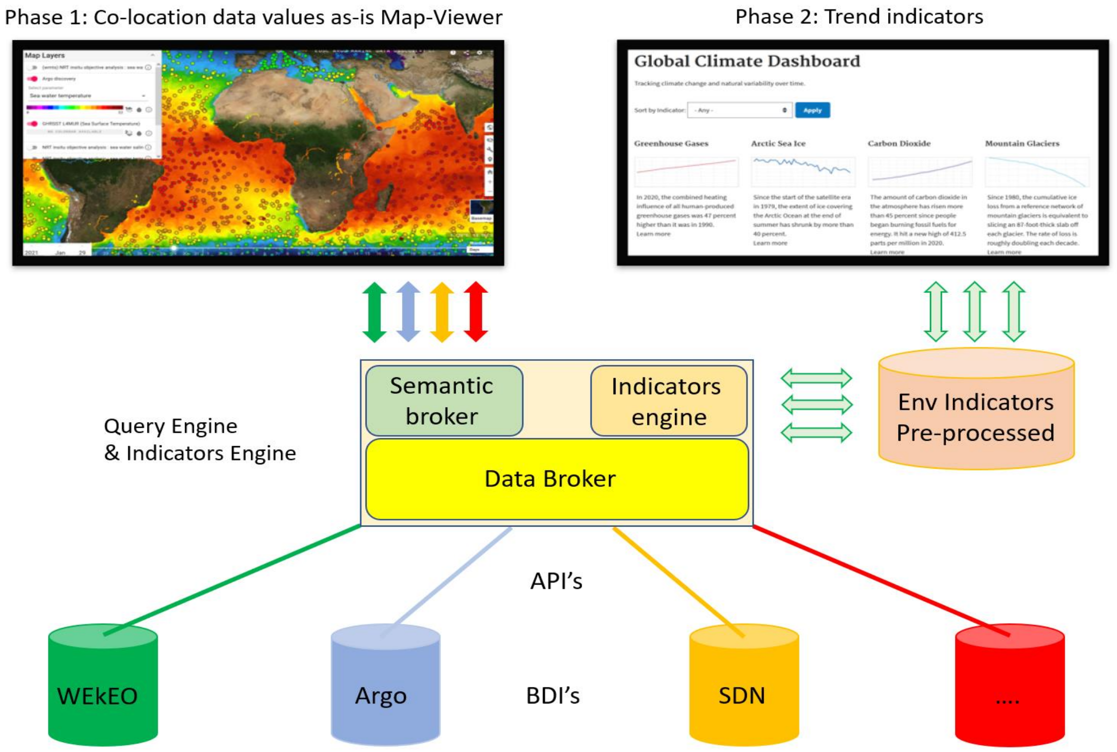
EOSC-Future ENVRI SP – Dashboard of the state of the environment: Ocean component – V5 – 03-04-2023

As part of the EU EOSC Future project, ENVRI-FAIR partners are involved in WP6 for developing two Science Projects, one about Invasive species, led by LifeWatch, and one about a Dashboard of the state of the environment, led by ICOS. This dashboard should provide easy means to users to determine the state and follow trends of our Earth system for a selected number of parameters. The dashboard concerns the Earth components Atmosphere, Ocean, and Biodiversity. This document describes how the ocean component is developed and deployed.

Framework

The framework below shows the different building blocks of the ocean component, which consists of two phases, **1) <u>the co-location of data values as-is Map Viewer</u>** and **2) <u>trend indicators</u>**, with a focus on the Essential Ocean Variables (EOVs) **Temperature, Oxygen, Nitrate + Nitrite, Silicate, Phosphate and PH**.



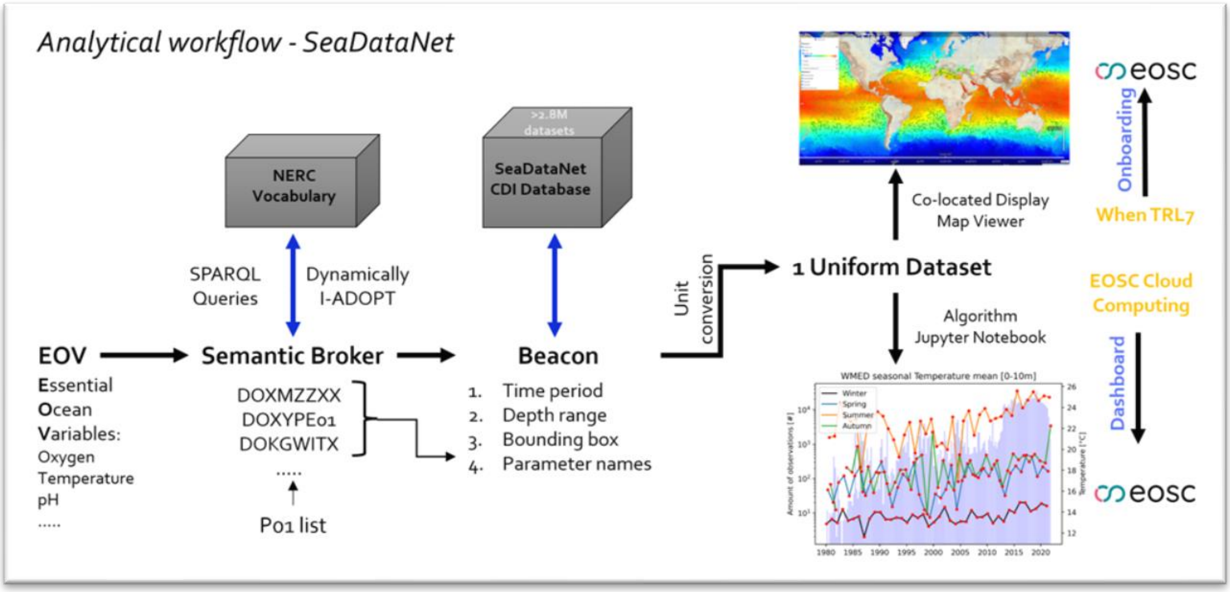
In the first phase (M18), the ocean component focused on **<u>the co-location of data values as-is Map Viewer</u>** that displays measurement values of selected parameters as retrieved from the available Blue Data Infrastructures (BDIs). The Map Viewer includes a user interface designed for *citizen scientists* with the possibility to interact with the data by using sliders for time, depth, and space.

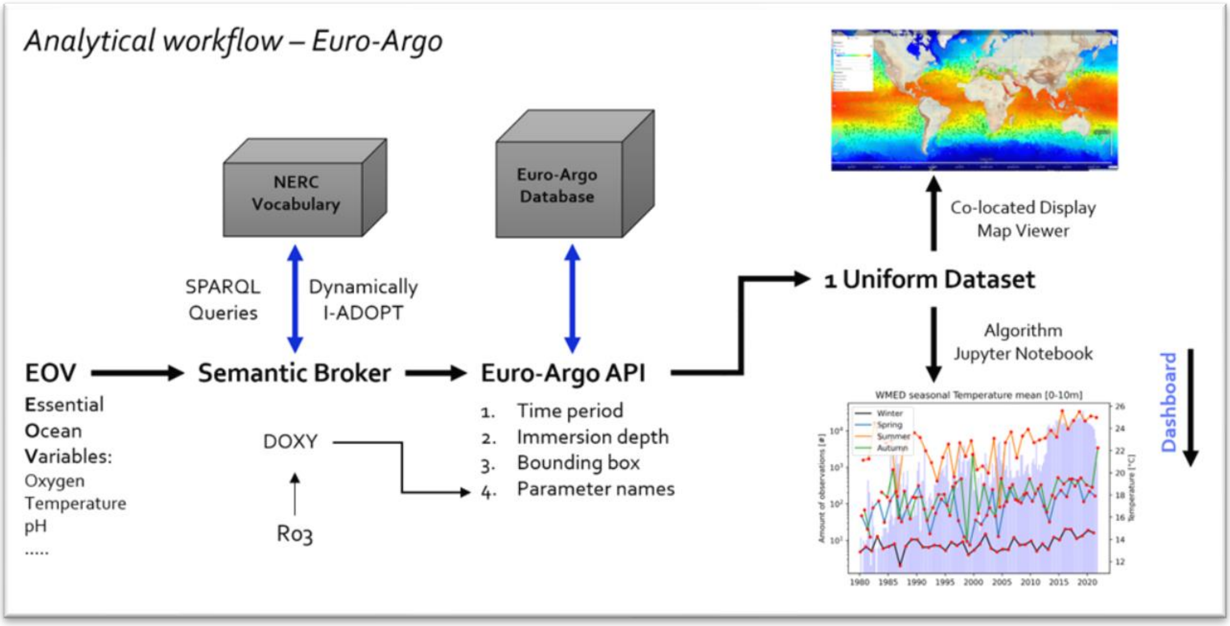
In the second phase (M29), the analytical workflows per parameter for determining scientifically justified aggregated values from the data as held in the BDIs will be deployed. These should facilitate to feed dynamic **<u>trend indicators</u>** at the 1st level of the dashboard, with some configuration options.

Finally, the goal is to have the environmental dashboard (level 1) include a number of **<u>trend indicators</u>** for the chosen EOVs, providing a trend in time for the designated areas. While, users can then click on such an indicator guiding them to the Map Viewer (level 2) to browse deeper into details and the actual data sets and timeseries that facilitate the trends.

Technical Framework

The frameworks below show a more elaborate overview of the different steps required for the map viewer and trend indicators, for the Blue Data Infrastructures SeaDataNet and Euro-Argo. The different components are explained in more detail below. Afterwards the first phase including the map viewer is discussed





Essential Ocean Variables (EOVs)

The focus within this Science Project lies on the Essential Ocean Variables (EOVs) Oxygen, Temperature, Nutrients and pH, where the nutrients are divided into Silicate, Phosphate and Nitrate. Each of these **six EOVs** will follow the technical framework individually.

Semantic Broker

The semantic broker uses “smart” mappings: i.e., a SPARQL query that relies on the correspondences between A05, P01 and R03 vocabularies, based on the decomposition of observable properties according to the I-ADOPT ontologies. These vocabularies are listed on the NERC Vocabulary Service (NVS) that gives access to standardised and hierarchically-organized vocabularies. The I-ADOPT properties will be used as query criteria to the NVS to retrieve the P01/R03s related to each EOV. For example, EOV Oxygen has the following properties:

- hasMatrix water body
- hasObjectOfInterest oxygen
- hasProperty Concentration

Summarizing, the SPARQL query gives the relevant P01/R03s depending on SeaDataNet (P01) or Euro-Argo (R03), which are then used as input to Beacon or the Euro-Argo API to retrieve data related to the EOVs. Specific details of the Semantic Broker are given in Appendix A.

Beacon - innovative data lake system accessible via a REST API

At MARIS there have been ongoing developments to create a system called beacon with a unique indexing system that can, on the fly, extract specific data based on the user’s request from millions of observations datafiles containing multiple parameters in diverse units. This system and its data can be accessed via a REST API that is exposed by beacon itself meaning clients can query data via a simple JSON request. The system is built in a way that it returns one single harmonized file as output, regardless of whether the input contains many different datatypes or dimensions. It also allows for converting the units of the original data if parameters are measured in different types of units.

It is important to mention that the system can be applied to different data infrastructures and is not tailor made for one specific type of database. Here, we show the possibilities of this API by applying it to the SeaDataNet CDI database where a query makes it possible to obtain specific data based on the following filter options:

- Parameter selection
- A bounded box with minimum and maximum longitudes and latitudes
- Depth range
- Time period

For the parameter selection, the P01 vocabulary from the NERC Vocabulary Server is used, as the data within the SeaDataNet CDI is mapped with these elements that are identifiers of physical parameters such as Temperature or Oxygen. The data can be retrieved in formats such as one single flattened NETcdf or JSON with dimensions LONGITUDE, LATITUDE, PARAMETER VALUE, TIME and DEPTH.

Euro-Argo API

The Euro-Argo API is connected to the Euro-Argo ERIC where a query makes it possible to obtain specific data based on the following filter options:

- Parameter selection
- A bounded box with minimum and maximum longitudes and latitudes
- Time period
- Immersion depth

The immersion depth is the upper water level of pre-defined depth ranges that can be seen on the map viewer. For example:

- If an Argo float makes multiple observations between the pre-defined [0, 5] m depth range, then the observation closest to the immersion depth 0 m is chosen.
- If there are five Argo floats in the same small region, for the [0, 5] m depth range it would return an observation closest to the upper level for each float individually, hence 5 in total (ex: float1 0.5 decibar, float2 1.5 decibar, …, float5 nothing if the parameter is not observed between 0-5 decibar).

For the parameter selection, the R03 vocabulary from the NVS is used, as data within the Euro-Argo database is mapped with these elements that are given in the Table below with their corresponding harmonized units.

CORIOLIS_ CODE	PARAMETER_CODE	LABEL	UNIT	CF_STANDARD_NAME
35	TEMP	Sea temperature	Degree_Celsius	Sea_water_temperature
64	DOXY	Dissolved Oxygen	µmol/kg	moles_of_oxygen_per_unit_ mass_in_sea_water
124	PH_IN_SITU_TOTAL	pH	-	sea_water_ph_reported_on_ total_scale
381	NITRATE	Nitrate (N03-N)	µmol/kg	moles_of_nitrate_per_unit_ mass_in_sea_water

Unit conversions

SeaDataNet

The data measurements stored within the SeaDataNet CDI have a wide range of different units and in order to display these measurements simultaneously they require the same unit. For each of the six EOVs a preferred unit is selected that corresponds to the CMEMS model background layers used in the Map Viewer (see table below). For example, for EOV Oxygen, measurements can be in µmol/kg, mol/kg, µmol/L, etc.., which then have to be converted to the desired unit: mmol/m3.

Temperature	**Oxygen**	**Phosphate**	**Silicate**	**Nitrate**	**pH**
[°C]	**[mmol/m3]**	**[mmol/m3]**	**[mmol/m3]**	**[mmol/m3]**	**[-]**

In order to obtain the required conversion factors and consequently enable harmonizing of the units, the P24 vocabulary from the NVS is used. Each P24 concept is a classification concept for units of measure based upon the fundamental SI quantities. In the NVS, P24 is linked to BODC's units of measure vocabulary P06 and hence P24 groups units of measure of identical vector dimensions. The relevant P24 dimension vectors are linked to each of the six EOVs listed above, such that all the units of measure can be selected that could be associated with data targeting that EOV including the observable properties defined in P01.

Each unit is also uniquely mapped to a QUDT unit through the owl:sameAs property. So, each P24 concept is associated to a list of valid units in the NVS which are linked to their equivalent in QUDT, giving us access to QUDT's conversion algorithms from the QUDT SPARQL endpoint.

For example, for EOV Oxygen for each P24 dimension it will retrieve the relevant P06 units and using the owl:sameAs property it will provide the associated QUDT units as shown in the diagram below. Having acquired the QUDT unit, we can federate to the QUDT SPARQL endpoint to extract the unit conversions.

Euro-Argo

For the data retrieved from the Euro-Argo API, for Oxygen and Nitrate a conversion factor of 1.025 is applied as the measurements in µmol/kg must be converted to mmol/m3, where we assume a density of sea water of 1025kg/m3.

Uniform Data

At this point of the technical framework, the data that is obtained from the Euro-Argo and SeaDataNet databases should include the preferred unit. This data can then be used as input towards phases **1) the Map-Viewer** and **2) the Ocean Indicators** that are both discussed in more detail below.

Phase 1: Co-location data values as-is Map-Viewer

Current status

In this section, the front-end of the Map Viewer is presented one step at a time. The Map Viewer will show in-situ observations of the EOVs Temperature, Oxygen, Nutrients and pH.

Datalayers

☐ Temperature

☐ Oxygen

☐ Nutrients

☐ Potential Hydrogen (pH)

The EOV nutrients is divided into nitrate & nitrate + nitrite, phosphate and silicate. Whereas the other EOVs do not have sub-parameters.

☑ Nutrients

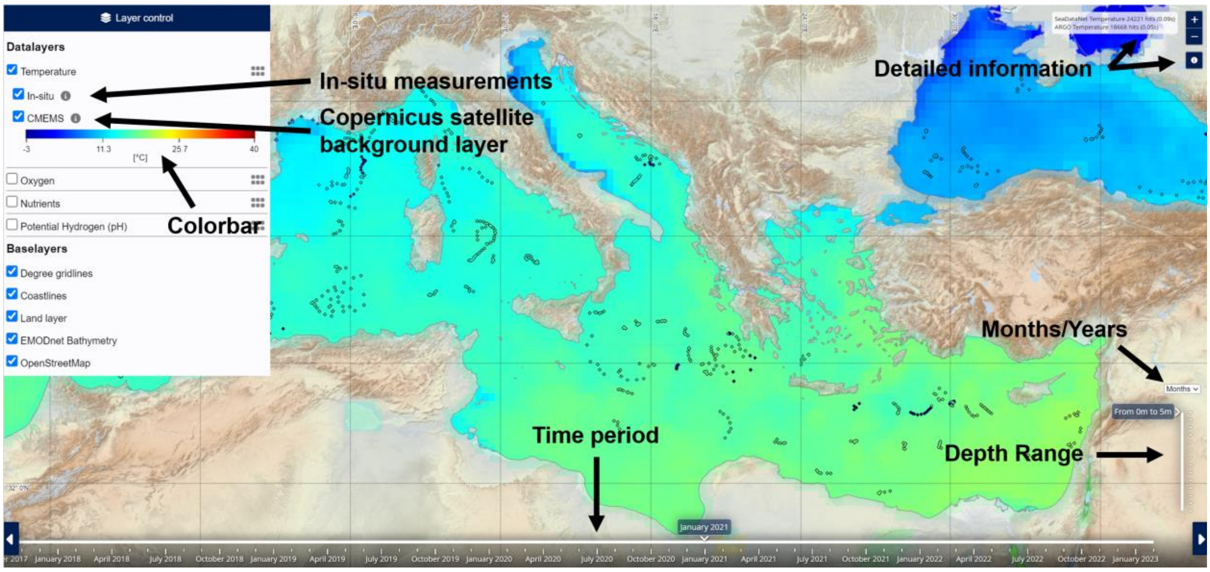
☐ Nitrate & Nitrate+Nitrite

☐ Phosphate

☐ Silicate

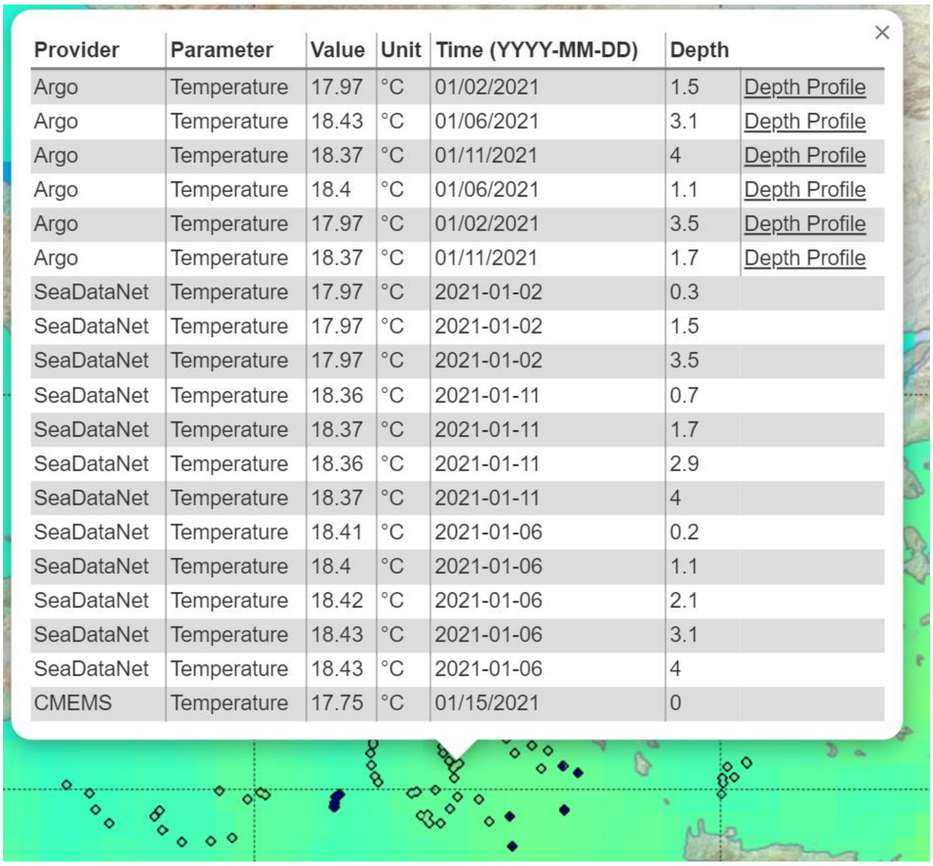
For now, the Map Viewer will include data measurements from SeaDataNet CDI and Euro-Argo ERIC. The related in-situ data sets concern single observations such as profiles, tracks, analysed water samples, etc which are not available as continuous timeseries. Moreover, the spatial coverage varies in time, density, and numbers. The in-situ observations can be plotted on the map by selecting an EOV, time period (years or months) and depth range (e.g., [0, 5] m). The user can also select multiple EOVs at once, creating an aggregated data collection.

- Time period, the user can choose to either select a **month** or **year** as seen in the slider in the Figure below.
- Depth range, the user can choose between different depth ranges following the IODE standard depths, including for example the observations between [0, 5] m.



A colorbar is displayed with a clear colour scheme, depicting the unit of the EOV and the corresponding values. Depending on the EOV a linear or log scale is used for visual purposes. There is a layer system in place where the user can choose the desired hierarchy that determines which data should be shown if different parameters are at the same location.

After the user has selected the parameters, the number of measurements retrieved from the SeaDataNet and Euro-Argo databases can be seen in the top right corner. Here, users are also able to click on the information icon in the top right and then click on the map near measurements to get more detailed information as seen below.



Provider	Parameter	Value	Unit	Time (YYYY-MM-DD)	Depth	
Argo	Temperature	17.97	°C	01/02/2021	1.5	Depth Profile
Argo	Temperature	18.43	°C	01/06/2021	3.1	Depth Profile
Argo	Temperature	18.37	°C	01/11/2021	4	Depth Profile
Argo	Temperature	18.4	°C	01/06/2021	1.1	Depth Profile
Argo	Temperature	17.97	°C	01/02/2021	3.5	Depth Profile
Argo	Temperature	18.37	°C	01/11/2021	1.7	Depth Profile
SeaDataNet	Temperature	17.97	°C	2021-01-02	0.3	
SeaDataNet	Temperature	17.97	°C	2021-01-02	1.5	
SeaDataNet	Temperature	17.97	°C	2021-01-02	3.5	
SeaDataNet	Temperature	18.36	°C	2021-01-11	0.7	
SeaDataNet	Temperature	18.37	°C	2021-01-11	1.7	
SeaDataNet	Temperature	18.36	°C	2021-01-11	2.9	
SeaDataNet	Temperature	18.37	°C	2021-01-11	4	
SeaDataNet	Temperature	18.41	°C	2021-01-06	0.2	
SeaDataNet	Temperature	18.4	°C	2021-01-06	1.1	
SeaDataNet	Temperature	18.42	°C	2021-01-06	2.1	
SeaDataNet	Temperature	18.43	°C	2021-01-06	3.1	
SeaDataNet	Temperature	18.43	°C	2021-01-06	4	
CMEMS	Temperature	17.75	°C	01/15/2021	0	

The information is stored in a table, and when the measurements concern Argo floats, the user is able to click on the profile info to get the information of the whole depth profile as seen below.

Float

Country: ITALY

Institution: OGS, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale

Status: Active

Data Center: CORIOLIS

Cycle

Cycle number: 135

Direction: Ascending

Datetime: 02.01.2021 05:52:30 Quality: 1

Number of levels: 104

Data Mode: Delayed-Mode

Position: 36.26593N 20.63666E Quality: 1

Primary sampling: averaged [10 sec sampling 25 dbar average from 700 dbar to 700 dbar; 10 sec sampling 10 dbar average from 700 dbar to 100 dbar; 10 sec sampling 2 dbar average from 100 dbar to 2.0 dbar]

Primary sampling: averaged [10 sec sampling 25 dbar average from 700 dbar to 700 dbar; 10 sec sampling 10 dbar average from 700 dbar to 100 dbar; 10 sec sampling 2 dbar average from 100 dbar to 2.0 dbar]

Primary sampling: averaged [10 sec sampling 25 dbar average from 700 dbar to 700 dbar; 10 sec sampling 10 dbar average from 700 dbar to 100 dbar; 10 sec sampling 2 dbar average from 100 dbar to 2.0 dbar]

The viewer also includes background layers from CMEMS that include modelling or satellite products. This is done by using OGC Web Map Service (WMS), which is a standard protocol providing a simple HTTPs interface for requesting geo-registered map images. The in-situ data can be plotted on top of these products as seen in the figure above, such that a comparison can be made. The CMEMS products used in the viewer are listed below.

Temperature

Dataset names:

dataset-armor-3d-rep-monthly (January 1993 – December 2020)

dataset-armor-3d-nrt-monthly (January 2019 – Present)

Variable in datasets: Sea water temperature [°C]

Oxygen

Dataset names:

cmems_mod_glo_bgc_my_0.25_P1M-m (January 1993 – December 2020)

global-analysis-forecast-bio-001-028-monthly (January 2020 – Present)

Variable in datasets: Mole concentration of dissolved molecular oxygen in sea water [mmol/m^3]

Nutrients

Dataset names:

cmems_mod_glo_bgc_my_0.25_P1M-m (January 1993 – December 2020)

global-analysis-forecast-bio-001-028-monthly (January 2020 – Present)

Variables in datasets: Mole concentration of nitrate, silicate, phosphate in sea water [mmol/m^3]

pH

Dataset names:

cmems_mod_glo_bgc_my_0.25_P1M-m (January 1993 – December 2020)

global-analysis-forecast-bio-001-028-monthly (January 2020 – Present)

Variable in datasets: Sea water pH reported on total scale [-]

Phase 2: Trend Indicators

Current status

A methodology is defined and implemented as an analytical workflow for determining scientifically justified aggregated values for feeding the trend indicators as planned at the first level of the Environmental Dashboard. Natural variability in time and space of parameters and the fact that the availability of data sets also varies in time and space, are major challenges to overcome. *The ambition here is to construct the trend indicators based only on the in-situ data that we have (level 1 indicators).*

The focus will lie on the European seas, which is divided into different sub-regions as seen in the figure below. For each of the regions listed below, trend indicators are established based on the in-situ data that is available.

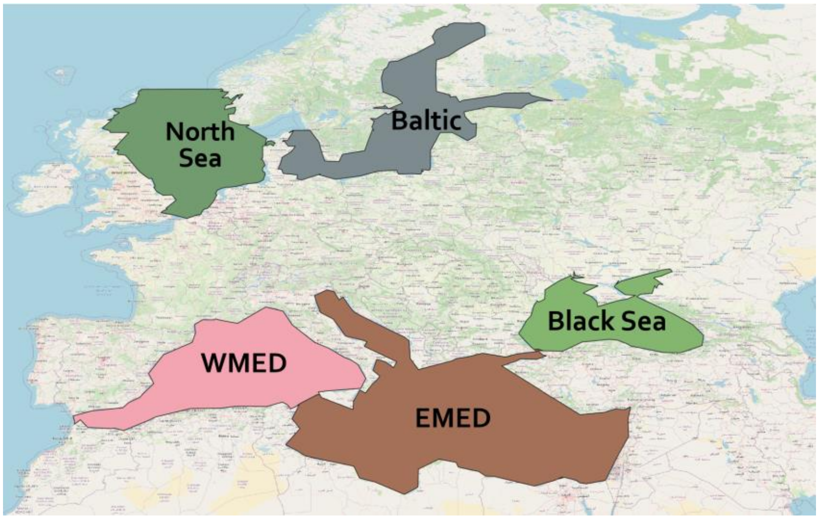


Figure 1 – European sea regions

IPCC's sixth assessment report mentions multiple climate-impact drivers. Some of these drivers can be related to the parameters, temperature, oxygen, nutrients and PH. For each of these drivers there can be opportunities by connecting them to the data in the tool. Below, relevant drivers are listed.

Ocean warming: Sea-surface temperature (SST) observations performed mostly by satellites are routinely used in assessments of the state of the climate. They underpin our understanding of multidecadal to centennial changes and the causes of those changes and are a key metric against which climate change policy decisions are made and progress against international agreements is measured.

Ocean deoxygenation: the loss of oxygen in the ocean, results from ocean warming, through a reduction in oxygen saturation, increased oxygen consumption, increased ocean stratification and ventilation changes. Deoxygenation and acidification often coincide because biological consumption of oxygen produces CO2. Deoxygenation can have a range of detrimental effects on marine organisms and reduce the extent of marine habitats.

Changes in nutrients concentration: the availability of nutrients in the surface ocean often limits primary productivity, with implications for marine food webs and the biological carbon pump. Nitrogen availability tends to limit phytoplankton productivity throughout most of the low-latitude ocean, whereas dissolved iron availability limits productivity in high-nutrient, low-chlorophyll regions, such as in the main upwelling region of the Southern Ocean and the Eastern Equatorial Pacific. Phosphorus, silicon, other micronutrients such as zinc, and vitamins can also co-limit marine phytoplankton productivity in some ocean regions.

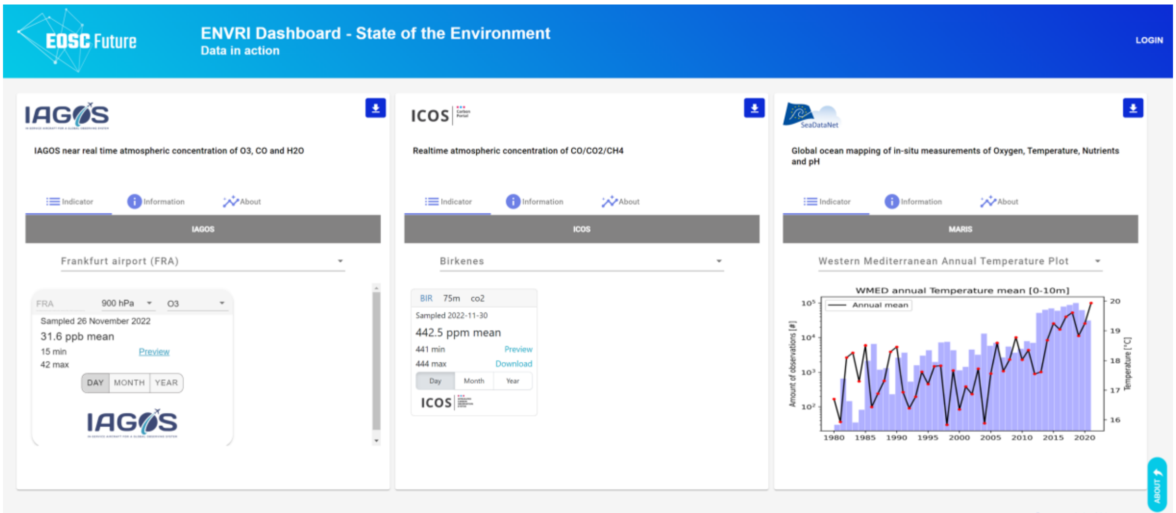
Ocean acidification: the ocean's uptake of anthropogenic carbon affects its chemistry in a process referred to as ocean acidification, which increases the concentrations of aqueous CO2, bicarbonate and hydrogen ions, and decreases pH, carbonate ion concentrations and calcium carbonate mineral saturation states. Ocean acidification affects a variety of biological processes with, for example, lower calcium carbonate saturation states reducing net calcification rates for some shell-forming organisms and higher CO2 concentrations increasing photosynthesis for some phytoplankton and macroalgal species.

Methodology

Trend indicators can be established that substantiate the climate drivers. There is a notebook in development that calculates indicators for the Ocean Variables Temperature, Oxygen, pH, Nitrate + Nitrite, Silicate, Phosphate based solely on in-situ data. More information on this notebook can found in the relevant documentation that is available.

Dashboard state of the environment

The dashboard will concern the Earth components Atmosphere, Ocean, and Biodiversity. The Ocean Indicators will be displayed in a frame per EOV, with some configuration options. The indicators will be pre-processed and updated every month, such that additional data can be used as input. The configuration options will include multiple time periods, depth ranges, sea regions and the type of graph, where from the indicator page, it will be possible to go to the Map Viewer.


ENVRI Dashboard - State of the Environment

Data in action

IAGOS near real time atmospheric concentration of O3, CO and H2O

Frankfurt airport (FRA)

FRA 900 hPa O3

Sampled 26 November 2022

31.6 ppb mean

15 min

42 max

Realtime atmospheric concentration of CO/CO2/CH4

Birkenes

BIR 75m co2

Sampled 2022-11-30

442.5 ppm mean

441 min

444 max

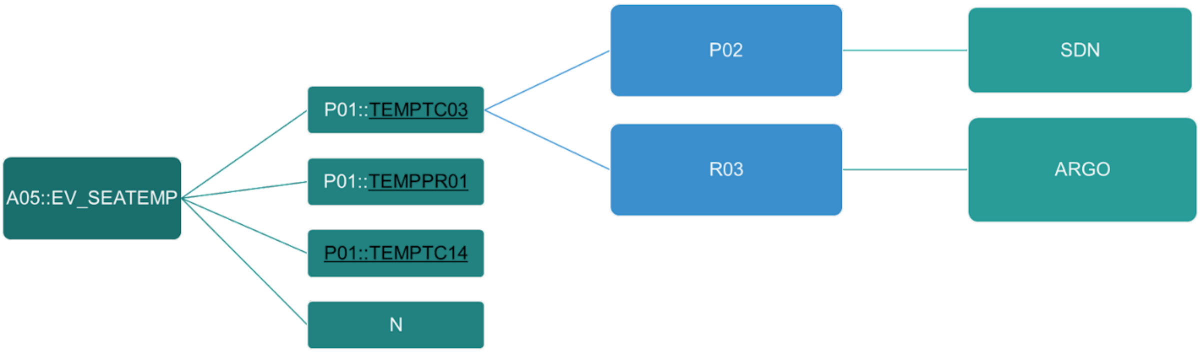
Global ocean mapping of in-situ measurements of Oxygen, Temperature, Nutrients and pH

Western Mediterranean Annual Temperature Plot

Appendix A – Smart Mapping Concept/Semantic Broker & Unit Conversions

The Ocean Component will focus on the following EOV-parameters listed in the A05: AtlantOS Essential Variables vocabulary: **Temperature, Oxygen, Nutrients, pH**

NVS will use the smart mappings approach to discover the observable properties defined in P01 and P07 that are associated with the selected EOV parameters. P01 concepts are mapped to P02 and R03 concepts that SDN and ARGO RIs are using respectively. These observable properties are then used as query criteria to retrieve the relevant datasets from the underlying RIs.



Smart mappings

Smart mappings are dynamically created mappings leveraging the power of SPARQL queries combined with the explicit description of observable properties according to a common ontological framework. For this work, we are using the recently published I-ADOPT framework (InteroperAble Descriptions of Observable Property Terminologies). For the Ocean Component that uses EOV parameters listed in the A05 vocabulary, i-adopt ontology properties are used to explicitly describe A05 concepts i.e., iop:hasObjectOfInterest, iop:hasMatrix, iop:hasProperty.

P01 and P07 concepts will also be explicitly expressed using the i-adopt ontology. Commonly expressing different vocabularies, enables interoperability and flexibility and in our case the implementation of the smart mappings approach between A05, P01 and P07. The most important advantage is that real mappings do not need to be created and maintained, making this approach dynamic, maintainable and easily extendable.

Unit conversions

For the EOSC Ocean component deliverable we are focussing on the following parameters and their units of measure:

Temperature	**Oxygen**	**Nutrients**	**pH**
[°C]	**[mmol/m3]**	**[mmol/m3]**	**[-]**

For the unit conversions we did some research on what already exist as we want to align with the international effort and not reinvent the wheel from the main unit of measure authorities and bodies like UCUM: https://ucum.org/trac, QUDT:https://qudt.org/ and UDU units from unidata: https://www.unidata.ucar.edu/software/udunits/. We were also approached by the CODATA Digital Representation of Units of Measurement (DRUM) Task group, to share the unit conversions work with the group in the upcoming Units summit that will be held as part of the SciDataCon 2022 (part of International Data Week IDW)

QUDT was the only body providing unique URIs for the units and a SPARQL endpoint for extracting conversions, which follows the Linked Data standards that NVS also does. QUDT units are also mapped to UCUM and UDU units, so if at any point there is a need to refer to UCUM and UDU, QUDT provides this possibility.

QUDT.org is a 501(c)(3) public charity non-profit organization founded to provide semantic specifications for units of measure, quantity kind, dimensions and data types. QUDT is an advocate for the development and implementation of standards to quantify data expressed in RDF and JSON.

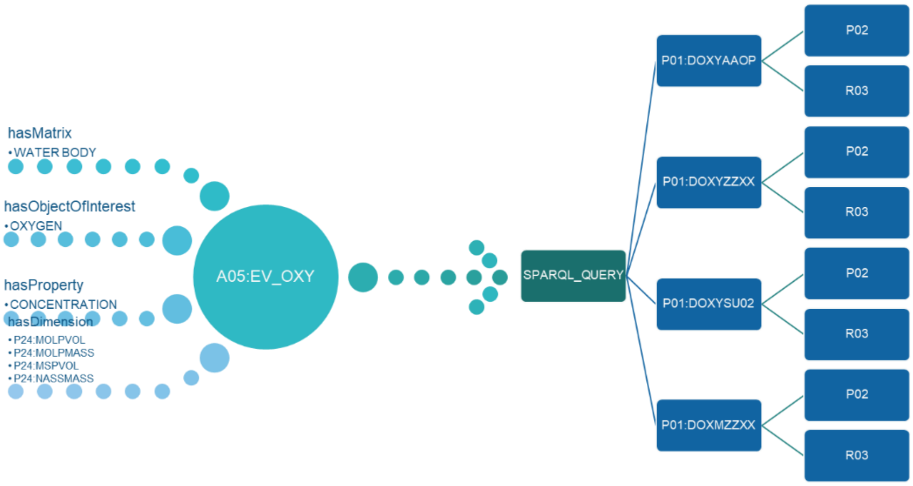
The unit conversion functionality, will showcase the power of Linked Data using federated SPARQL queries: from the NVS SPARQL endpoint, that will provide the unit of measure of the required observable properties (P01, R03 etc), to the QUDT SPARQL endpoint, that will provide the required conversions. Different approaches could be taken to link EOV, observable properties and appropriate unit conversions. Our preferred approach at this stage is to associate each A05 concept to the vector dimension of the quantity kinds that may be associated with the EOV.

In NVS, the P24 collection is the vocabulary for vector dimensions. Each P24 concept is a classification concept for units of measure based upon the fundamental SI quantities. In the NVS, P24 is linked to BODC's units of measure vocabulary (P06) and hence P24 groups units of measure of identical vector dimensions. Thus, by associating each targeted A05 with their relevant P24 dimension vectors we can select all the units of measure that could be associated with data targeting that EOV that have been associated with observable properties defined in P01 and P07.

Each unit is also uniquely mapped to a qudt unit through the owl:sameAs property. So, each P24 concept is associated to a list of valid units in the NVS which are linked to their equivalent in QUDT, giving us access to QUDT's conversion algorithms from the qudt sparql endpoint.

We will use the Oxygen as an example, to describe the process, requiring the transformation of Oxygen from µmol/kg, mol/kg, µmol/L -> To desired unit: mmol/m3.

So, when an EOV parameter is selected from the A05 vocabulary for the oxygen example, A05:EV_OXY its iop properties are used as query criteria to a SPARQL query, to retrieve all the P01 concepts that fulfill the criteria and the mappings to other vocabularies related to different RIs as shown in the diagram.



Another query will retrieve the P24 dimensions associated with the A05:EV_OXY and for each P24 dimension it will retrieve the relevant P06 units and using the owl:sameAs property it will provide the associated qudt units as shown in the diagram below. Having acquired the qudt unit, we can federate to the QUDT SPARQL endpoint to extract the unit conversions.