D9.8
Marine subdomain EOV product version 2

Work Package	WP9
Lead partner	IFREMER
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30-04-2023
Submission date	03-04-2023

Deliverable abstract

Marine subdomain EOV product version 2 is available publicly.

The overarching goal of ENVRI-FAIR is that all participating research infrastructures (RIs) will improve their FAIRness and become ready for connection of their data repositories and services to the European Open Science Cloud (EOSC). Deliverable D9.1 has reported on the roadmap of the RIs in the marine subdomain towards improving their FAIRness. It presented the approach of using FAIR questionnaires (together with WP5) to identify the strengths and weaknesses of each RI and a first indicative set of activities to improve identified weaknesses or gaps. After formulation in Deliverable D9.2 of implementation plans for mitigating these gaps during the next phase of the ENVRI-FAIR project, the RIs from the marine subdomain have specified in Deliverable D9.3 the technical services and interfaces to be implemented at RI level and have undertaken the implementation. The RI vocabularies and erddap services were presented in Deliverable D9.4 and the operational list of FAIR services presented in Deliverable D9.5.

The present Deliverable D9.8 is linked to Task 9.8 going from M24 to M48 and aiming to demonstrate the marine subdomain FAIRness. A first version of the ENVRI-FAIR Marine semantic Essential Ocean Variables (EOV) data broker was already presented in Deliverable D9.7, and this deliverable D9.8 presents the version 2 software, which includes new technologies for performance enhancement. This version 2 is now available on GitHub and preserved on Zenodo:

- https://github.com/envri-fair/marine-eov-broker
- https://doi.org/10.5281/zenodo.7871643

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Thierry Carval	IFREMER	10-04-2023
Contributing Authors	Guillaume Alviset	IFREMER	13-04-2023
Reviewer(s)	Delphine Dobler	Euro-Argo	28-04-2023
Approver	Andreas Petzold	FZJ	03-04-2023

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	10-04-2023	Draft version	Thierry Carval
V1.0	26-04-2023	Finalised version	Thierry Carval

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY

A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at http://doi.org/10.5281/zenodo.4471374.

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

D9.8 - Marine subdmain EOV product version 2

1 Rationale

The Task 9.8 "Marine Essential Ocean Variable - EOV" (started end of 2020) is a use case based on the data and metadata services set up by the Research Infrastructures (RIs). It will provide interoperable access to RI data to end users, in particular the VIP users Copernicus Marine In Situ, SeaDataNet or EMODnet involved in the project.

Each RI involved in WP9 has developed an implementation plan [D9.2] that addresses the results of their FAIRness self-analysis, with the shared objective to improve the FAIRness at the Marine subdomain level as illustrated in Figure 1.
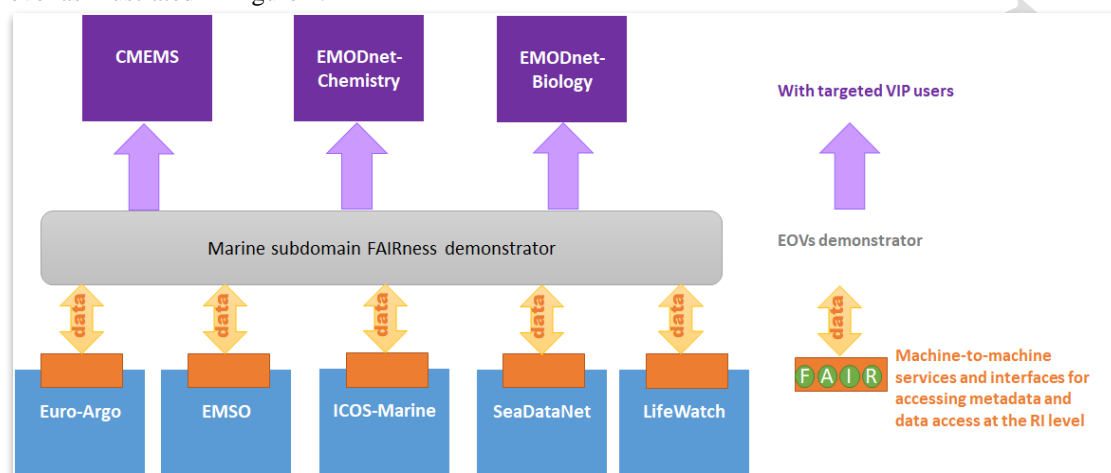

Figure 1: marine sub-domain implementation plan

The demonstrator targeted within Task 9.8 aims to set up a workflow that will serve data files answering to EOV requests, in particular of the VIP users (SeaDataNet, Copernicus Marine In Situ and EMODnet). This will allow them to process data aggregation and product assessment (QA/QC) on the extracted data.

As shown in Figure 2, the Marine EOV demonstrator delivers data from the Marine RIs to users querying an EOV (such as oxygen), on a given spatial and temporal bounding box.

This D9.8 report documents the public access to the version 2 of the Marine EOV product as it was specified in the previous deliverable D9.6. This version 2 is built upon the cross-domain and RI operational services presented in the D9.5 report and the EOV broker set up within Task 9.8. Compared to the version 1 (presented in D9.7), the version 2 of the Marine EOV includes enhanced access to data and metadata by using a SPARQL endpoint combined to the ERDAPP instead of using ERDAPP alone. The SPARQL endpoint is a more efficient way to select specific metadata information, as it provides rich semantic selection criteria. The focused criteria can reduce the volume of data transferred from ERDDAP server and improve response time. It also provides more accurate data queries/answers.
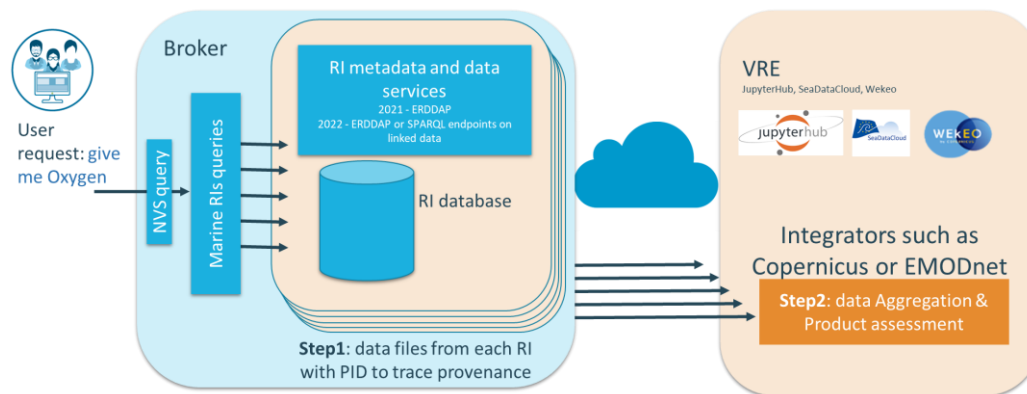
Figure 2: Marine EOV demonstrator

The data request performed by the broker is illustrated on Figure 3: the NVS vocabulary server is queried to provide all the parameter names corresponding to the requested EOV within each RI. The ERDDAP server and SPARQL endpoint of each RI return the corresponding data and metadata. The SPARQL endpoint provides much more metadata than the ERDDAP server.
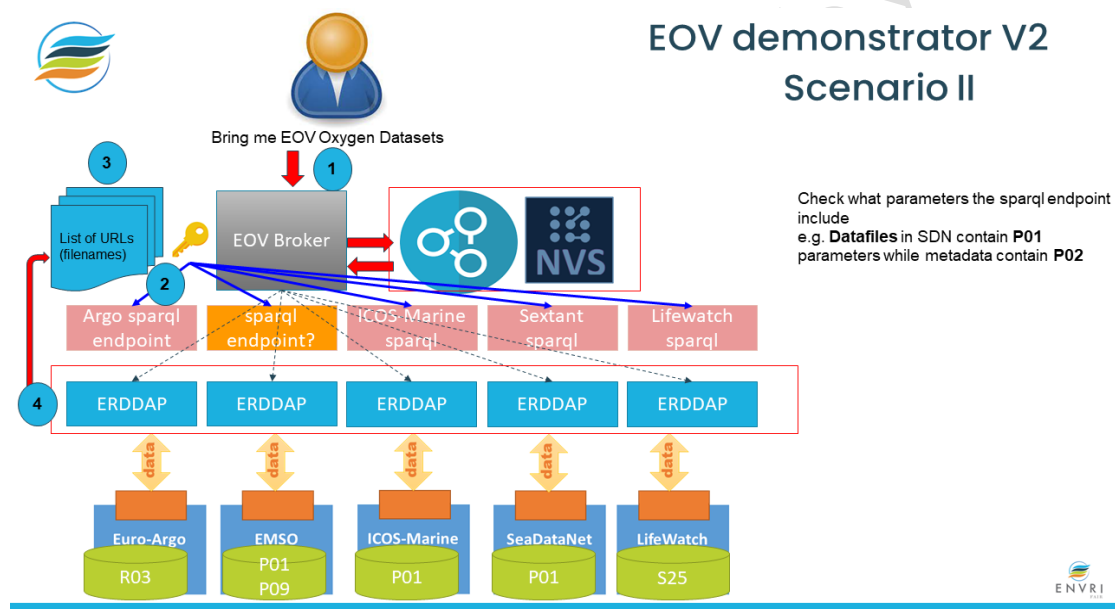

Figure 3: the services and principles of the Marine EOV data broker

The Marine EOV product **is implemented as a Jupyter notebook** (Figure 4) that allows a search on the data/metadata services of the RIs on a specific set of EOVs (oxygen or chlorophyll-A, temperature, salinity, and zooplankton biomass and diversity). An example of Jupyter notebook graphical output is shown on Figure 5.

The integration of the Marine EOV within the ENVRI-hub is completed: it is key component to query data and metadata from the RI services developed or enhanced within ENVRI-FAIR project.
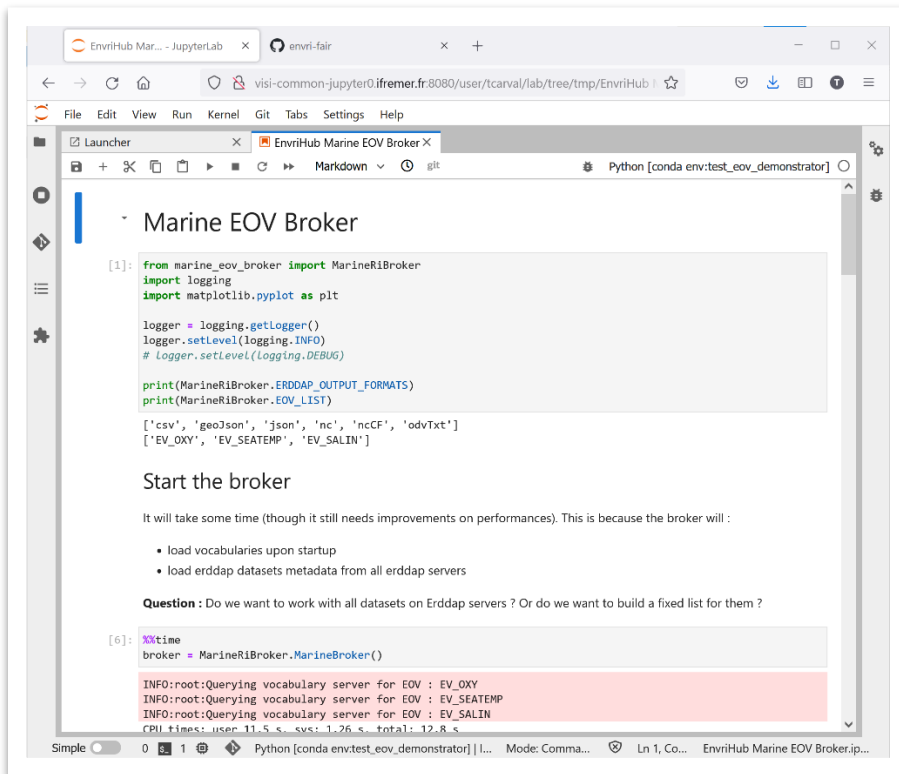
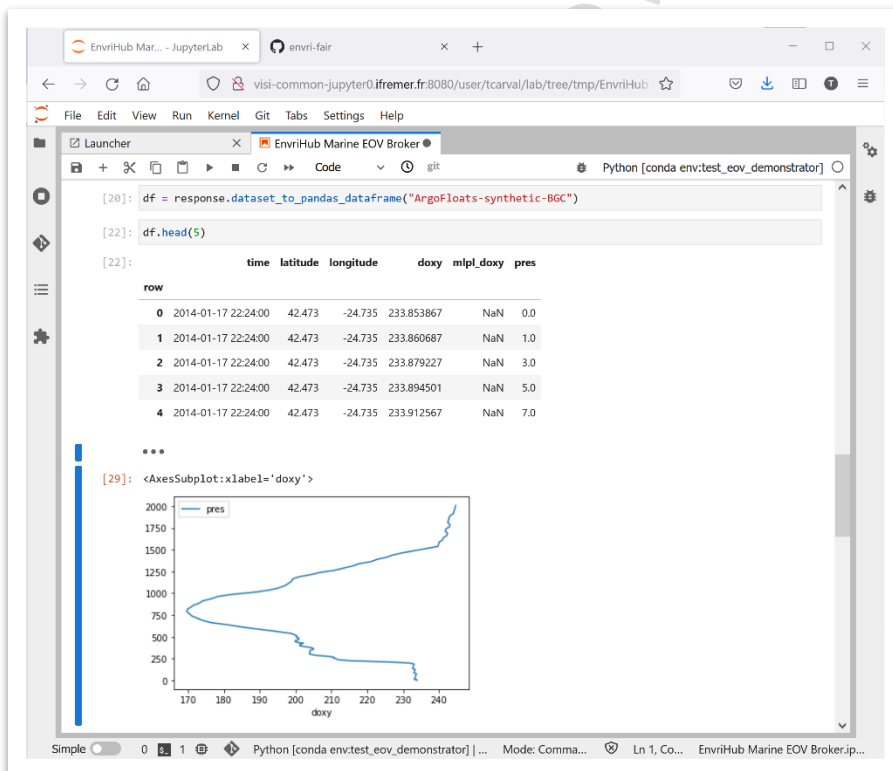Figure 4: Create a query with the Marine EOV Broker in a Jupyter notebook

Figure 5: Explore data results returned by the EOV Marine Broker in a pandas dataframe

2 Marine EOV publicly available

2.1 Marine EOV python code managed on GitHub

The Marine EOV broker is now publicly available from ENVRI-FAIR GitHub (Figure 6):

- https://github.com/envri-fair/marine-eov-broker

The project licence is Creative Commons CC0 (universal public domain).
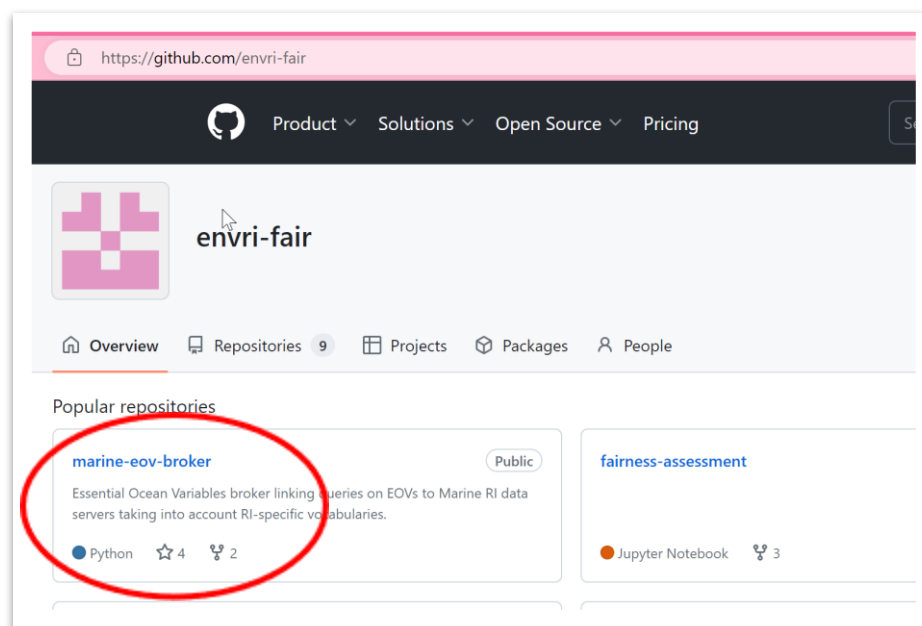

Figure 6: the Marine EOV broker V2 is published on ENVRI-FAIR GitHub public repository.

2.2 Marine EOV python code version 2.0.1 preserved on Zenodo

The Marine EOV software is a python code managed on GitHub. The major releases such as the version 2.0.1 documented in this document are preserved on Zenodo as:
https://doi.org/10.5281/zenodo.7871643

The publication of software code from GitHub into Zenodo is a good practice for reproducibility of scientific data (faiR).
GitHub is the FAIR place to manage software development but does not guarantee the long-term preservation of the successive Marine broker releases.
Zenodo is the FAIR place to address the long-term preservation of Marine broker releases.

3 References

Ref	Title	Version / Date
D9.1	Marine subdomain FAIRness roadmap https://doi.org/10.5281/zenodo.3885296	V2.0 August 31, 2019
D9.2	Marine subdomain implementation plan https://doi.org/10.5281/zenodo.3885327	V1.0 November 30, 2019
D9.3	RIs technical specification https://doi.org/10.5281/zenodo.3885330	V1.0 May 19, 2020
D9.4	RIs implementation https://doi.org/10.5281/zenodo.4738541	V0.1 April 29, 2021
D9.5	RIs operational FAIR services https://doi.org/10.5281/zenodo.5843412	V1.0 January 12, 2022
D9.6	Marine EOV product specification https://doi.org/10.5281/zenodo.4766796	V1.0 May 11, 2021
D9.7	Marine EOV broker Version 1 https://doi.org/10.5281/zenodo.6410538	V1.0 March 10, 2022

ENVRI
FAIR

4   Appendix A

AAI	Authentication and Authorisation Infrastructure
ACDD	Attribute Convention for Data Discovery
API	Application Programming Interface
CDI	Common Data Index (metadata format and data access system by SeaDataNet)
CF	Climate and Forecast (semantics for NetCDF)
CMEMS	Copernicus Marine Environment Monitoring Service
COPERNICUS	A major earth observation programme run by European Commission and European Space Agency
CP	Carbon Portal
DwC-A	Darwin Core Archive file format
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENVRI	1) An environmental RI cluster FP7 project 2) Environment research infrastructures (in ESFRI level or upcoming) as a community
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable(s)
ERDDAP	NOAA developed science data server technology
ERIC	European Research Infrastructure Consortium (legal entity type)
ESFRI	European Strategy Forum on Research Infrastructures
EuOBIS	European OBIS
FAIR	Findable Accessible Interoperable Reusable
GUI	Graphical User Interface
ICOS	Integrated Carbon Observation System
IPT	Integrated Publishing Toolkit
M	Month
NetCDF	Network Common Data Format
NVS	NERC Vocabulary Serveur
NOAA	US National Oceanic and Atmospheric Administration
OBIS	Ocean Biogeographic Information System
OGC	Open Geospatial Consortium
PID	Persistent Identifiers
QA/QC	Quality Assurance/Quality Control
RDF	Resource Description Framework
RI	Research Infrastructure
SDN	SeaDataNet pan-European infrastructure for marine data management
SPARQL	The SPARQL protocol performs semantic queries on RDF triple stores
SPARQL endpoint	The API (Application Programming Interface) for semantic queries on information systems
TF	Task Force

ENVRI
FAIR