



Blue-Cloud2026

Process and developments to achieve marine data FAIRness to support Virtual Research

BC2026 webinar 1, 26 September 2023

Peter Thijsse (MARIS)

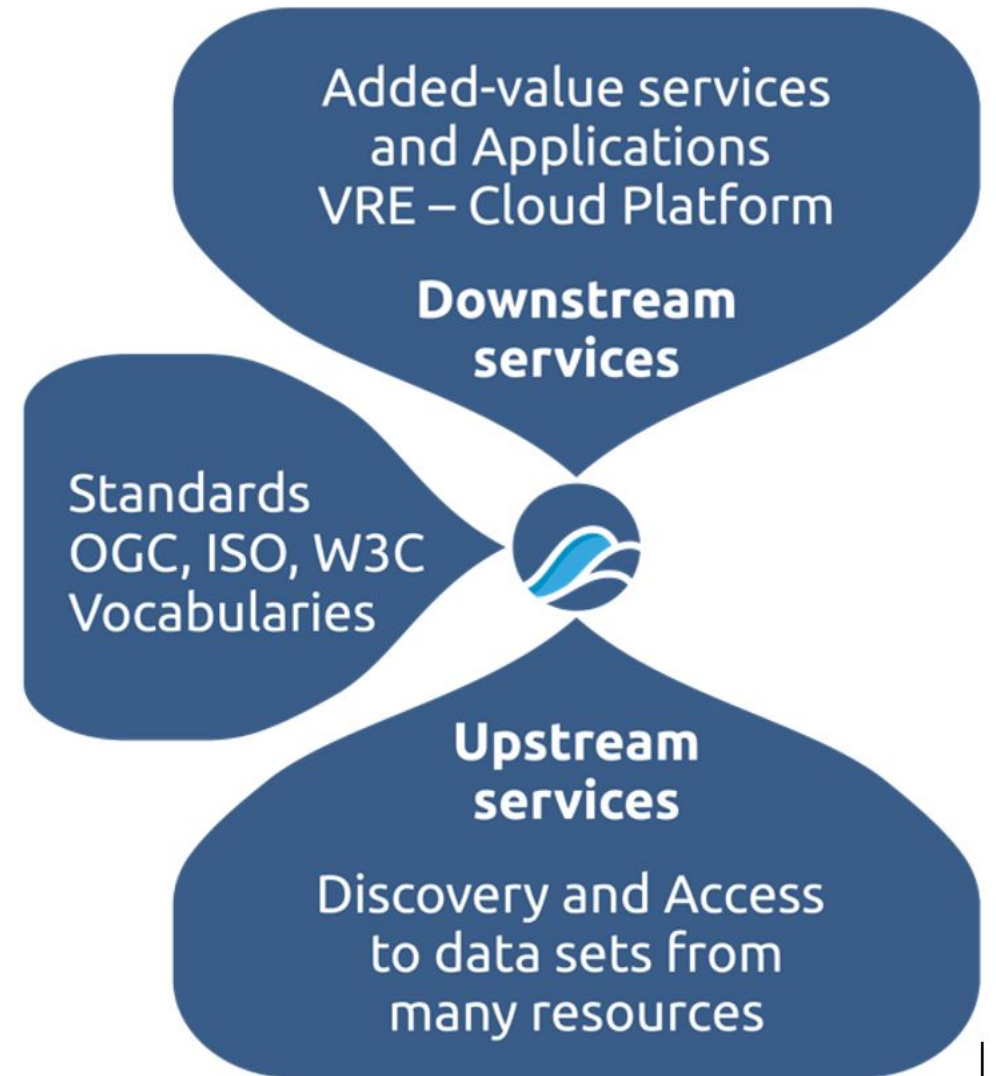


Funded by
the European Union

- Introduction
- Blue-Cloud core services and role data discovery and access
- Steps to improve data FAIRness
- FAIR data, FAIR software, ... FAIR services??

The Blue-Cloud concept and core services

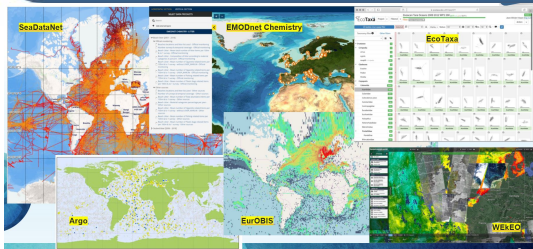
- Developing and **deploying a Virtual Research Environment (VRE) and Virtual Labs**, with an array of services for configuring and running virtual labs for specific analytical workflows, use cases and demonstrators
- **Applying common standards and interoperability solutions** for providing harmonized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





Blue Data infrastructures

E-infrastructures

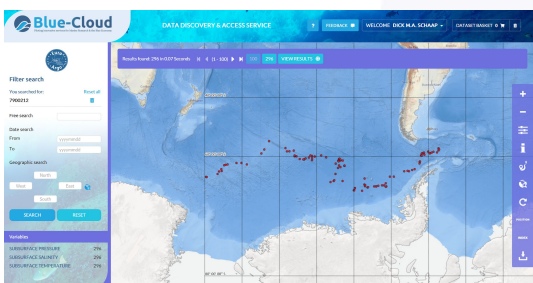
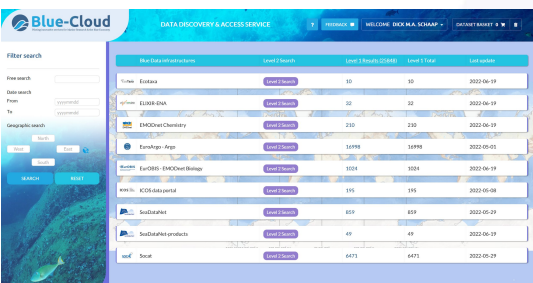


Facilitates users:

- Federated search for discovering interesting data sets (currently more than 10 million) in a two step approach
- 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



#	Order #	BDI	Dataname	Source	Dataname	ID
22987	426	SeaDataCloud Black Sea Temperature and Salinity Climatology V2	IT	North Sea - Eutrophication and Acidity aggregated datasets 1921/2020 v2021		
22986	426	Black Sea gridded climatology decades - for cold intermedia content at 1/8°	IT	Arctic Ocean - Contaminants aggregated datasets 1974/2015 v2018		
22985	426	Mediterranean Sea - Temperature and salinity Historical Data Collection SeaDataCloud V1	IT	Baltic Sea - Contaminants aggregated datasets 1974/2018 v2021		
22984	426	Baltic Sea - Temperature and salinity observation collection V2	IT			
22983	425	A01185hbaz	IT	RNODC_Bottle_1185_13		
22982	425	RNODC_Bottle_10855	IT	RNODC_Bottle_10855_18		
22981	424	RNODC_Bottle_10855	IT	RNODC_Bottle_10855_1		
17835	417	Asterocoida distribution data from: Deep-sea fauna of European seas - an annotated species check-list of benthic invertebrates living deeper than 2000 m in the seas bordering Europe	IT			

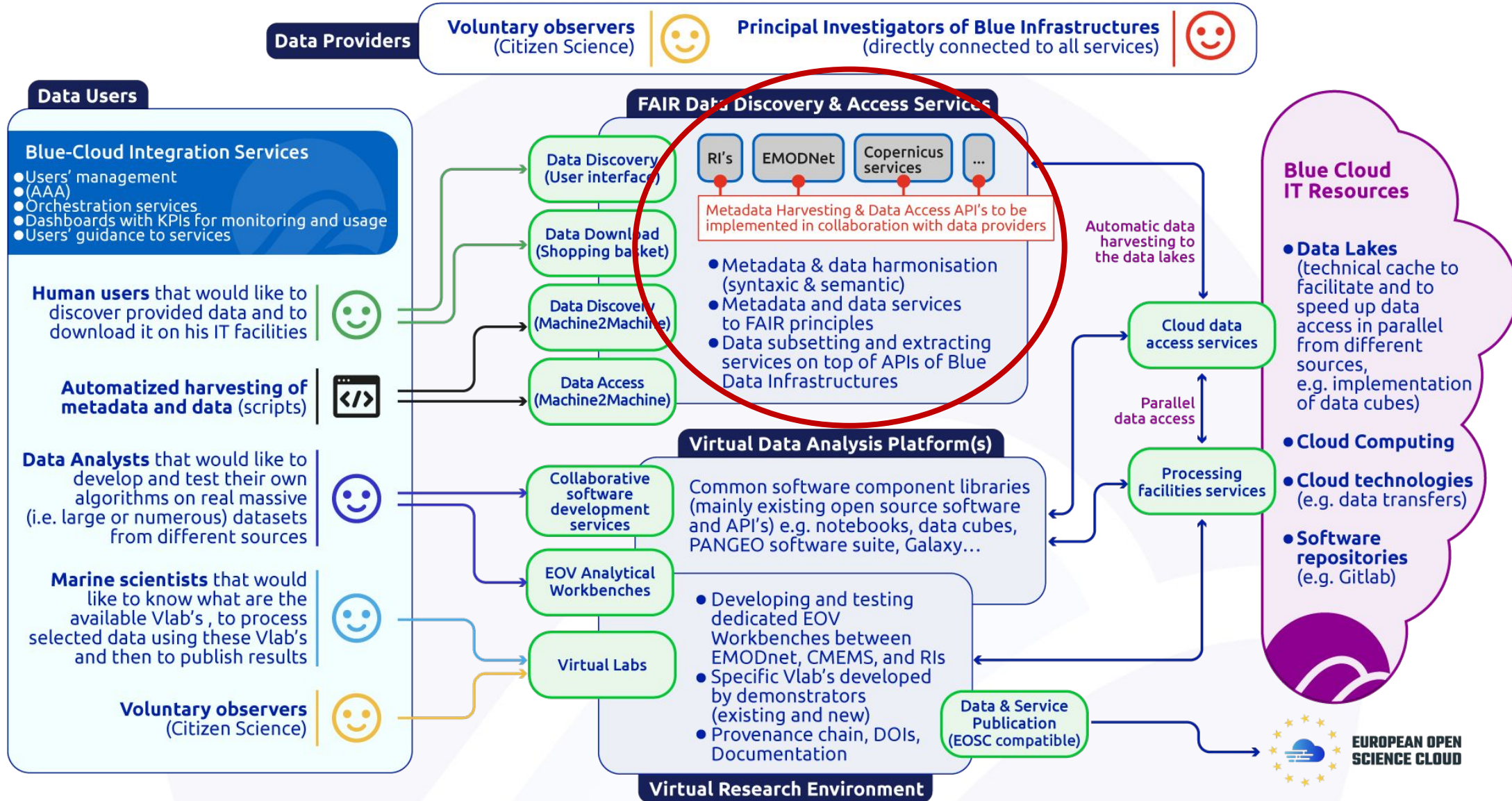
Compose and submit shopping request at the granule level



Retrieve the datasets by downloading from the Dashboard



Push datasets to the Blue-Cloud VRE Data Pool



Steps in improving data FAIRness


The **Blue-Cloud Data Discovery & Access service** federates services of Blue Data Infrastructures, of which many are “powered by ENVRI RI’s”: SeaDataNet, EMODnet, Euro-Argo, EurOBIS, LifeWatch, ICOS, but also others like ELIXIR, ENA, SIOS, ..

- ENVRI-FAIR improvements/best practices have resulted in:
 - **improved machine2machine services** for metadata and data access
 - **Upgraded metadata model** for enhanced FAIRness
 - Expanded vocabularies to support provenance (Re-usability)
- This results in expanded data availability for the VRE, better metadata, and e.g. improved EOVS datasets to come out of the workbenches

The FAIRness assessment and implementation choices have gone through different iterations:

From:

FAIR Assessment questionnaire in ENVRIFAIR used in 2019 (v3)



FAIR Implementation Profile (FIP)

- Community related metadata
- F1: Identifier type, MD/D
- F2: metadata schema; MD
- F3: MD-D linking mechanism
- F4: indexed search engines, MD/D
- A1: communication protocol, MD/D
- A1.1: authentication & authorisation technique, MD/D
- A1.2: metadata longevity, MD
- I1: knowledge language, MD/D
- I2: annotation, encoding vocabularies: MD/D
- I3: data schema, D
- R1.1: license, MD/D
- R1.2: provenance model, D
- R1.3: the FIP

To:

FIP Convergence Matrix

fip-wizard.ds-wizard.org

Email: "guest@example.com"

CS3 Community

FIP (FAIR Implementation Profile, 1.3.1)

General Information about Participant

III. Findability

Chapter text

F1 What globally unique, persistent, resolvable identifiers do you use for metadata records?

1.3.1 Choose your answer from FAIRsharing

1.3.2 Add your resource description here

F1 What globally unique, persistent, resolvable identifiers do you use for datasets?

F2 Which metadata schema(s) are you using for findability?

RI's have each described their status and plans
Results are captured in the Knowledge Base

- Marine RI's cooperated closely together
- Compare results of FIP responses
- Agree on similar development choices (within marine community and cross RI) leading to a set of **agreed best practices**:
 - Alignment of vocabularies (mainly NERC Vocabulary Service)
 - Improved PID's
 - RDF DCAT-AP metadata models
 - Sparql endpoints
 - Improved API's
 - e.g. ERDDAP services and Restful API's for data access

FAIR data, FAIR software, FAIR services??

- DDAS has a key position in the architecture
- Level 1 metadata level access via DAB (CNR)
- Level 2 data access level to distributed services via metadata to data access services.
- Important: Metadata and data harmonisation
 - metadata model mapping
 - vocabulary mapping (parameters, units, etc)
- But also: For each data access service (level 2) a specific “conversion” has to be implemented
 - how to search datasets
 - how to order
 - how to move from metadata to the data file request
 - difficult and human intervention needed.

FAIR data solutions (e.g. in ENVRI-FAIR) using FIP approach:

- improved machine2machine services for metadata and data access
- Upgraded metadata model for enhanced FAIRness (e.g. quality info)
- Expanded vocabularies to support provenance (Re-usability)

FAIR software (e.g. VLAB) examples:

- Software as a research object
- Publication in Zenodo/Github with sufficient metadata
- version management
- Clear license in metadata
- Software meets community standards

Metrics: FIP, F-UJI

 Findable

- (Meta)data are assigned a globally unique and persistent identifier
- Data are described with rich metadata
- Metadata clearly and explicitly include in the identifier of the data it describes
- (Meta)data are registered or indexed in a searchable resource

 Accessible

- (Meta)data are retrievable by their identifier using a standardized protocol
- The protocol is open, free and universal
- The protocol allows for authentication and authorization, as needed
- Metadata are accessible, even when the data are no longer available

 Interoperable

- (Meta)data use a formal, accessible, shared and broadly applicable language
- (Meta)data use vocabularies that follow FAIR principles
- (Meta)data include qualified references to other (meta)data

 Reusable

- (Meta)data are richly described with a plurality of accurate and relevant attributes
- (Meta)data are released with a clear and accessible data usage licence
- (Meta)data are associated with a detailed provenance
- (Meta)data meet domain-relevant community standards



FAIR data principles - source: CCDC

F: Software, and its associated metadata, is easy for both humans and machines to find.
F1. Software is assigned a globally unique and persistent identifier.
F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.
F1.2. Different versions of the software are assigned distinct identifiers.
F2. Software is described with rich metadata.
F3. Metadata clearly and explicitly include the identifier of the software they describe.
F4. Metadata are FAIR, searchable and indexable.
A: Software, and its metadata, is retrievable via standardised protocols.
A1. Software is retrievable by its identifier using a standardised communications protocol.
A1.1. The protocol is open, free, and universally implementable.
A1.2. The protocol allows for an authentication and authorization procedure, where necessary.
A2. Metadata are accessible, even when the software is no longer available.

I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
I2. Software includes qualified references to other objects.
R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
R1. Software is described with a plurality of accurate and relevant attributes.
R1.1. Software is given a clear and accessible license.
R1.2. Software is associated with detailed provenance.
R2. Software includes qualified references to other software.
R3. Software meets domain-relevant community standards.

FAIR software principles - source: M. Barker et al. Nature 2022

FAIR principles for **data access services** → increase findability, accessibility and interoperability of data access services (machine-2-machine)

This can be achieved by describing the services in a standardized manner, **just an API is not enough**, e.g.:

1. what the service does, what it offers
2. how it works, how to make requests
3. how to access it (authentication?)
4. input/output

A starting point will be **research on currently available ontologies for describing services**

- Several standardized vocabularies and ontologies are available
- Structured and machine-readable descriptions of services, making them more discoverable and interoperable

Existing used models and vocabularies for describing services (but these are in our opinion not yet complete):

- OWL-S
- OpenAPI Specification (formerly Swagger) => most promising candidate, when published as RDF. Needs additional attributes and semantics
- Dublin Core Metadata Initiative (DCMI)
- DCAT Class
- Hydra
- ESIP
- ODIS
- schema.org

- In FAIR-EASE project a working group will focus on **best possible solutions for describing services for m2m access**
- In this work close cooperation with BC2026
- Solutions will be documented and **tested as prototype in FAIR-EASE**
- Starting point is what already exists, building on top of that
- Leading to recommendations and implementation guidelines for BDI services (OBPS?)
- Looking for examples, expertise, in other domains, RDA WG, other infrastructures?
 - Please contact us when interested to share views and experiences

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