eosc | 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

Content

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

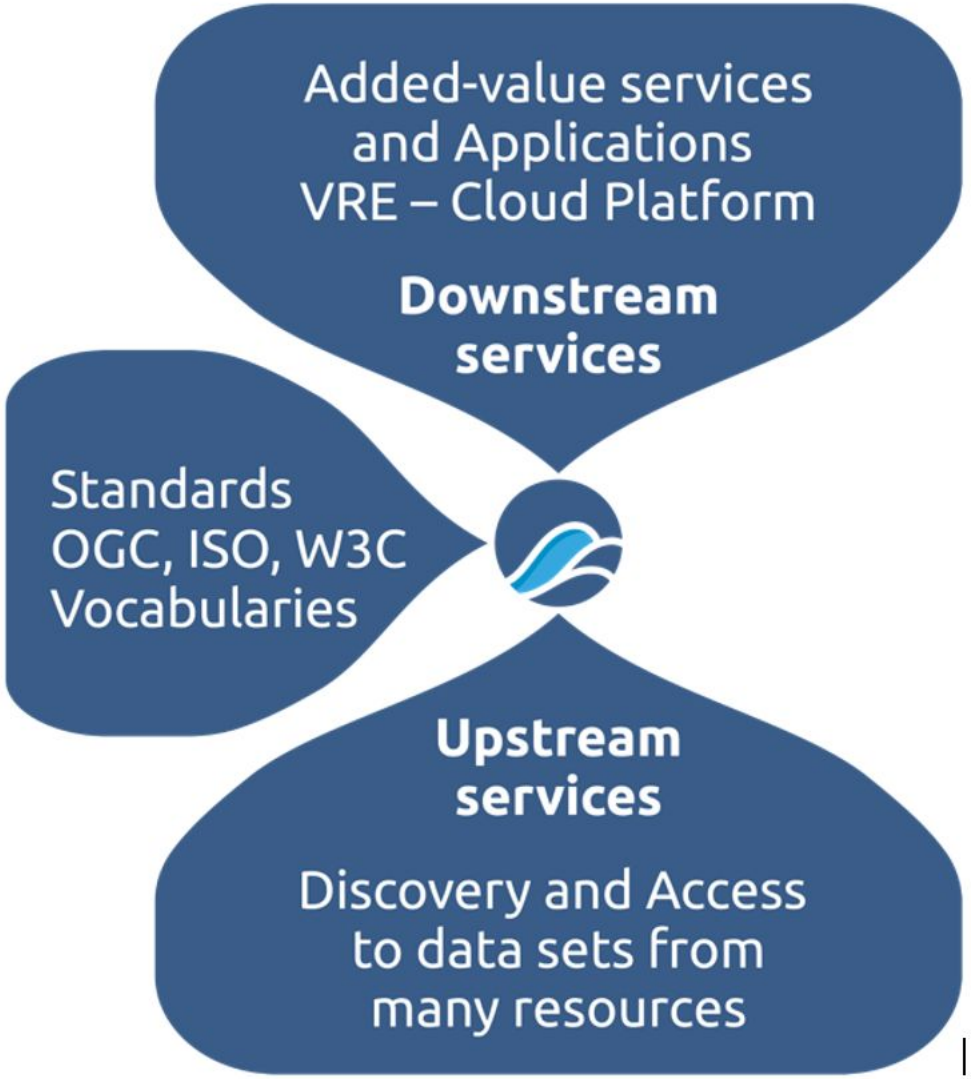
The Blue-Cloud concept and core services

Blue-Cloud overarching concept

- 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

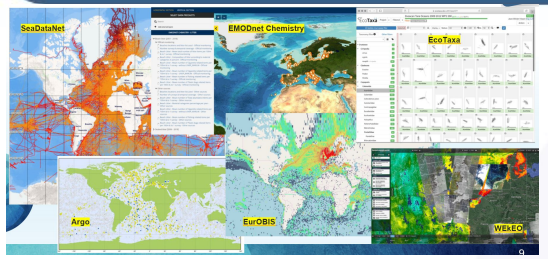


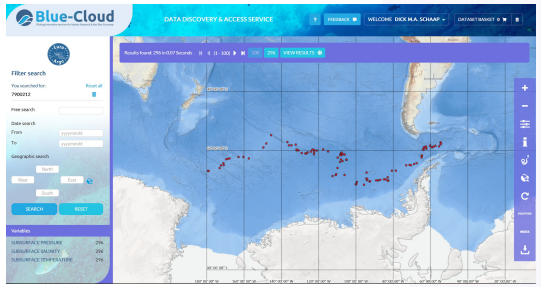
Blue-Cloud federation of major infrastructures

Blue Data infrastructures

E-infrastructures

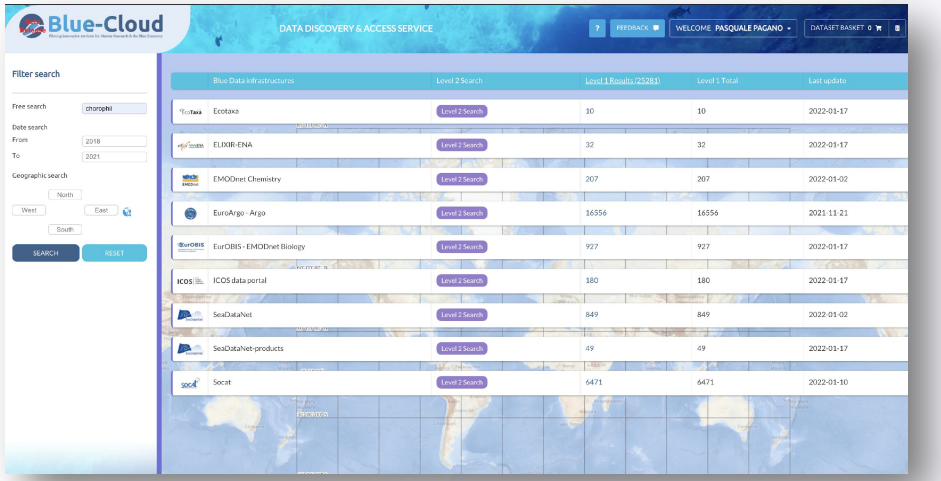
Blue Cloud Discovery and Access service

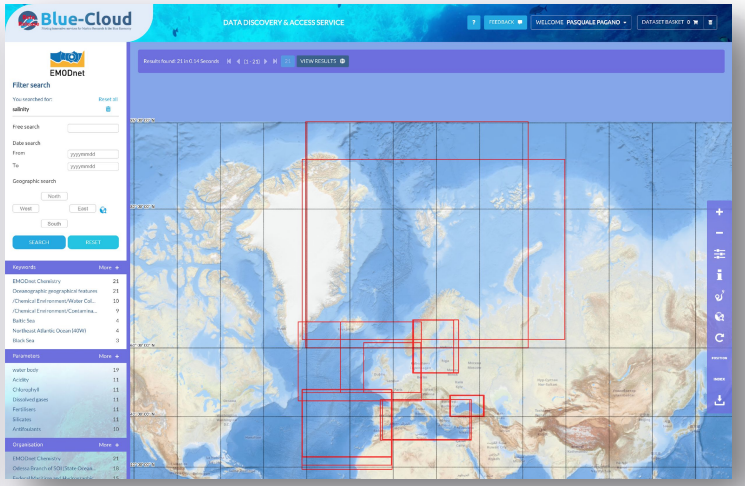


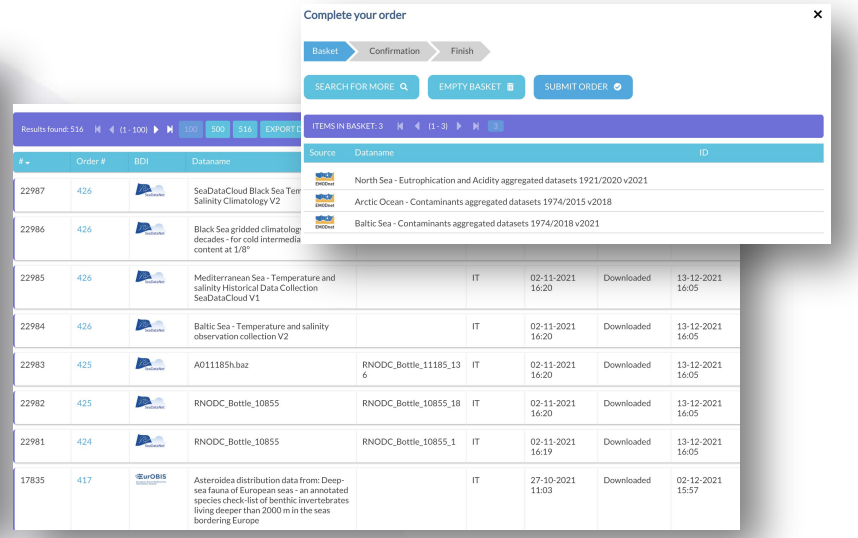


- **Facilitates users**:
  - Federated search for discovering interesting data sets (currently more than 10 million) in a two step approach
  - Federated retrieval of identified data sets using a shopping basket mechanism
  - Download of data sets or push to Blue-Cloud VRE
- **Facilitates managers of 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

Blue Cloud Discovery and Access service





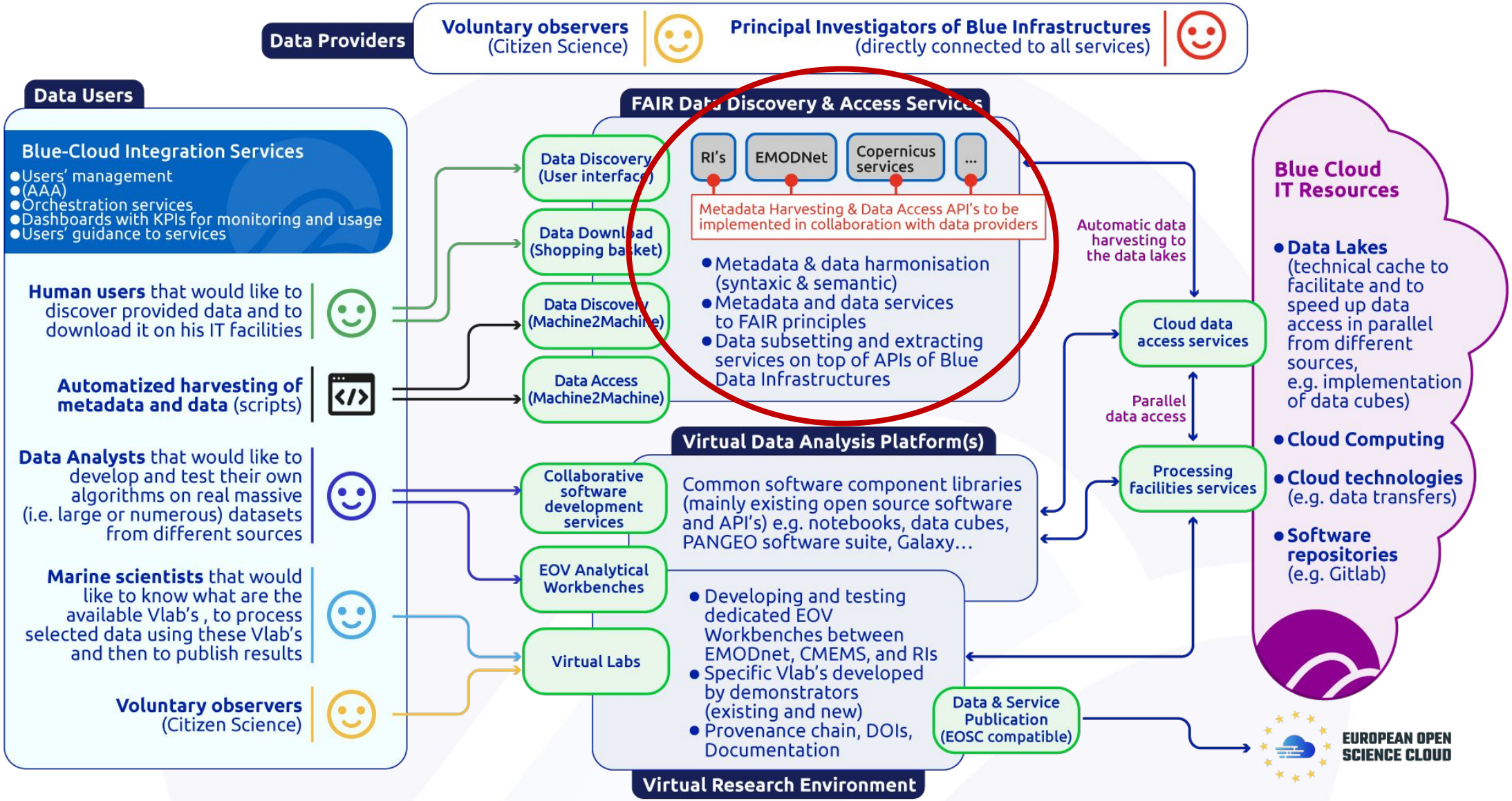


Compose and submit shopping request at the granule level

Retrieve the datasets by downloading from the Dashboard

Push datasets to the Blue-Cloud VRE Data Pool

Blue-Cloud 2026 services core: different “plugins” to enable data access



Steps in improving data FAIRness

ENVRI's valuable input to Blue-Cloud

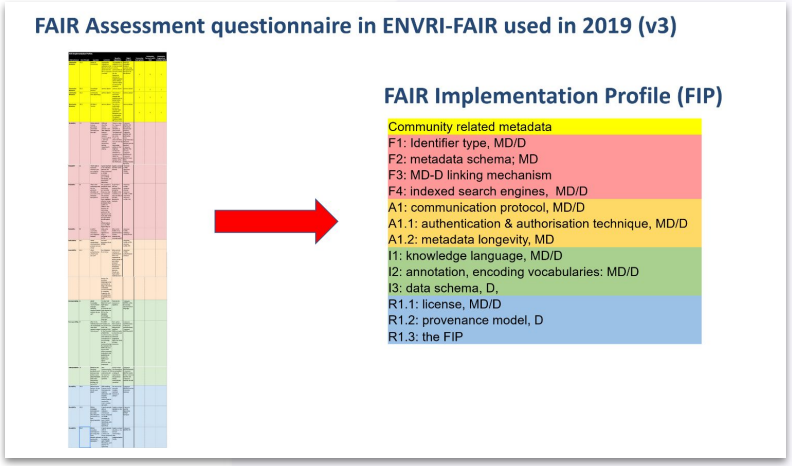
The **Blue-Cloud Data Discovery & Access service** federates services of Blue Data Infrastructures, of which many are "powered by ENVRI RI's": SeaDataNet, EMODnet, Euro-Argo, EurOBIS, LifeWatch, ICOS, but also others like ELIXIR, ENA, SIOS, ..

- 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 EOV datasets to come out of the workbenches

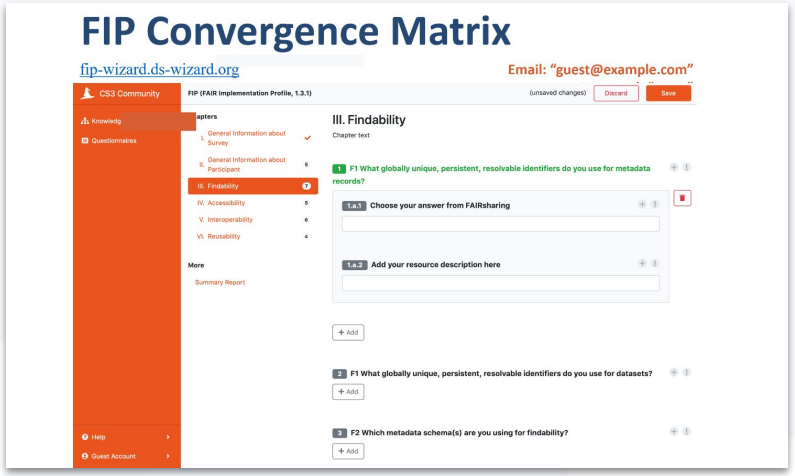
FIP process in ENVRIFAIR in the subdomains

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

From:



To:



RI's have each described their status and plans

Results are captured in the Knowledge Base

FAIR implementation choices

- 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, FAIR software

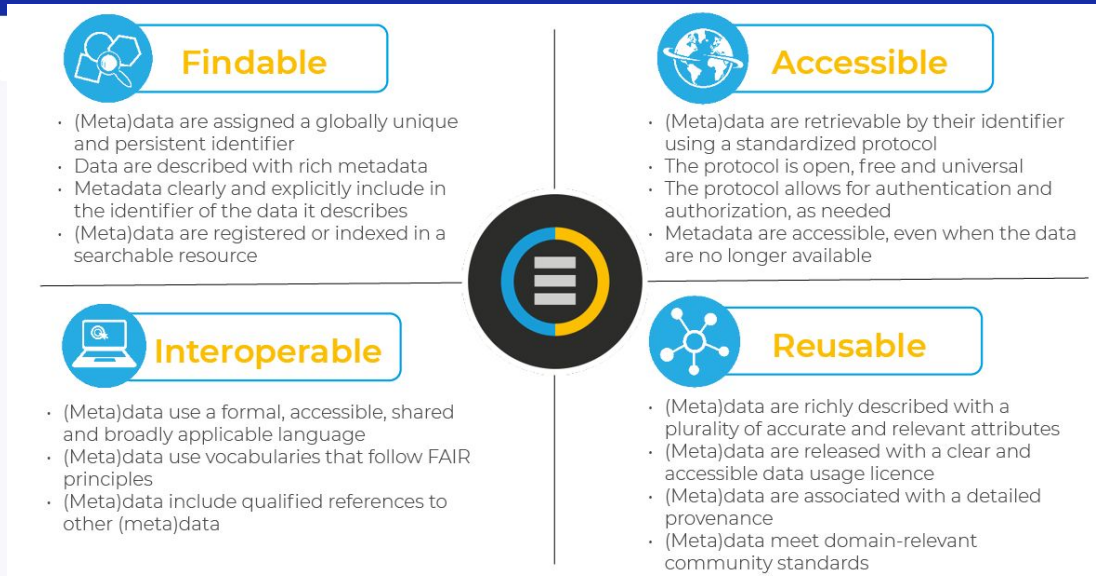
FAIR data solutions (e.g. in ENVRI-FAIR) using FIP approach:

- 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



FAIR data principles - source: CCDC

F: Software, and its associated metadata, is easy for both humans and machines to find.	**I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.**
F1. Software is assigned a globally unique and persistent identifier.	I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.
F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.	I2. Software includes qualified references to other objects.
F1.2. Different versions of the software are assigned distinct identifiers.	**R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).**
F2. Software is described with rich metadata.	R1. Software is described with a plurality of accurate and relevant attributes.
F3. Metadata clearly and explicitly include the identifier of the software they describe.	R1.1. Software is given a clear and accessible license.
F4. Metadata are FAIR, searchable and indexable.	R1.2. Software is associated with detailed provenance.
A: Software, and its metadata, is retrievable via standardised protocols.	R2. Software includes qualified references to other software.
A1. Software is retrievable by its identifier using a standardised communications protocol.	R3. Software meets domain-relevant community standards.
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.	

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

Next step: FAIR services!

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

How to proceed?

- 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

blue-cloud org

Funded by
the European Union