# Interoperable services as foundation for data access and processing

Alexandra Kokkinaki[1] (alexk@noc.ac.uk), Peter Thijsse[2], peter@maris.nl), Tjerk Krijger[2] (tjerk@maris.nl), Marc Portier[3] (marc.portier@VLIZ.be), Gwen Moncoiffé[1] (gmon@noc.ac.uk), Paul Weerheim[2] (paul@maris.nl)

[1]British Oceanographic Data Centre - BODC (UK)
[2]MARIS (The Netherlands)
[3]VLIZ, Flanders Marine Institute (Belgium)

**The problem in context**

One of the aims of the Blue-Cloud 2026 and FAIR-EASE projects is to offer users Virtual Labs and a Virtual Research Environment (VRE) in which they can process and visualise marine- and other domain datasets from a wide range of data infrastructures. A key component of both projects is the Data Discovery and Access service (DDAS), which provides easy access to diverse distributed datasets offered by Blue Cloud Data Infrastructures (BDIs) and other data providers relevant to the FAIR-EASE project. These datasets have been harmonised to adhere to the ISO19115 standard, thanks to the geoDAB broker, which is an essential component of DDAS. The geoDAB broker connects to a variety of heterogeneous data services, that comply with standards like the Catalogue Service for the Web (CSW), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and ISO 19115/19139) and collects metadata records heterogeneously encoded and harmonises them to meet the ISO19115 metadata model. Although the web services comply with the above-mentioned standards, machine to machine access is not yet enabled, as their descriptions are not harmonised nor are they machine actionable. This hinders web service discoverability, accessibility interoperability and reusability.

**Subsetting services**

A similar situation exists with subsetting services. Both FAIR-EASE and Blue-Cloud 2026 follow user requirements for more direct data access and data processing, offering services to facilitate access to subsets of data, where possible. Often original datafiles are too large or too many, and contain too many parameters the user may not be interested in. The advancement in subsetting software has led to an increase in BDIs providing subsetting services. In order to facilitate harmonised, machine-actionable, and ultimately scalable access to subsetting services, it's imperative that these services also provide harmonised and machine-actionable descriptions, aligned with the above-mentioned data services.

**Research question**

In the FAIR-EASE and Blue-Cloud projects we encountered a variety of machine-to-machine data access services that require harmonisation. This led us to focus on the following main research question:

"Define a schema to describe web services offered by a diverse range of service providers using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple providers."

This will lead to more FAIR services as it will increase their discoverability and automatic utilisation. Consequently, software applications capable of universal discovery and consumption of compliant services will emerge, simplifying development efforts. This question is equally relevant to the interoperability of EOSC services where no generally accepted solution exists.

An initial analysis of existing ontologies and gaps could lead to a set of recommendations. This will be presented during the session.

**Service semantic artefacts**

In order to achieve FAIR data access services, these must be semantically described using FAIR ontologies/vocabularies in a standardised manner, so that the following information is made available: what the service does; how it works; how to access it. This must not be mistaken with FAIRsFAIR's FAIR Assessment Framework for Data Services (1) which is a set of guidelines used to assess on a high level how well data services support FAIR data, and not the FAIRness of the actual services.

A preliminary list of available vocabularies, ontologies and standards for describing services has been drafted and includes OWL-S[9], OpenAPI Specification[10], smartAPI[11] project, Dublin Core Metadata Initiative[12] (DCMI), Data Catalog 3 (DCAT 3)[13], Hydra[14] and schema.org[15]. Schema.org guidance on service descriptions is provided by both the Earth Science Information Partners (ESIP) science-on-schema.org (SOSO)[16] cluster and Ocean Data Information System (ODIS)[17] although they cannot yet cover our needs for machine to machine (M2M) interoperability and actionability.

**Application in the frame of FAIR-EASE**

As part of the wider FAIR-EASE asset catalogue, the aim is to compile a list of data access services with standardised descriptions that will not only help developers navigate the API documentation and software libraries, but also enable machines (generic clients) to effectively select and execute the targeted data-requests. In a subsequent phase, this process might be further automated so that a generic request from a Virtual Research Environment (VRE) can select an appropriate data access service along with instructions on how to access it, resulting in a list of specific subset URLs for seamless retrieval and further processing.

**References**

Koers H., Herterich P., Hooft R., Gruenpeter M., & Aalto T., (2020). *M2.10 Report on basic framework on FAIRness of services (1.0)*. Zenodo. https://doi.org/10.5281/zenodo.4292599

Poster session

[9] https://www.w3.org/submissions/OWL-S/

[10] https://spec.openapis.org/oas/v3.1.0

[11] https://smart-api.info/

[12] https://www.dublincore.org/

[13] https://www.w3.org/TR/vocab-dcat-3/

[14] https://www.hydra-cg.com/spec/latest/core/

[15] https://schema.org/

[16] https://github.com/ESIPFed/science-on-schema.org

[17] https://oceaninfohub.org/odis/