

Using data fragments as the foundation for interoperable data access

Marc Portier¹ (marc.portier@vliz.be), Peter Thijsse² (peter@maris.nl), Alexandra Kokkinaki³ (alexk@noc.ac.uk), Tjerk Krijger² (tjerk@maris.nl), Gwen Moncoiffé³ (gmon@noc.ac.uk), Paul Weerheim² (paul@maris.nl), Erwan Bodere⁴ (erwan.bodere@ifremer.fr)

¹VLIZ, Flanders Marine Institute (Belgium)

²MARIS (The Netherlands)

³British Oceanographic Data Centre - BODC (UK)

⁴Ifremer (France)

FAIR-EASE project context

The FAIR-EASE project aims to provide users with a Virtual Research Environment (VRE) for processing and visualizing datasets from diverse Data Infrastructures across environmental subdomains. Blue Data infrastructures developed within the Blue-Cloud and Blue-Cloud 2026 projects will be utilized to streamline access to data in the VRE. However, for non-marine infrastructures in FAIR-EASE, accessing data poses an additional challenge due to varied data access services and metadata models outside marine domain harmonization efforts. Problem areas include metadata catalogues, subsetting services, and different types of aggregations.

Sustainability of the “interoperability” enterprise

Our FAIR-EASE experience shows that as we grow the scope of our integration initiatives, we also re-asses what harmonisation is required. An optimistic view perceives this as fine-tuning on a guaranteed trajectory, assuming the independent domain efforts follow some universal plan. A more critical perspective, however, acknowledges the trail of successively built and now abandoned systems in the rear-view mirror; and rightfully questions the overall cost and long-term sustainability of such an approach.

Confronting these views compels us to leverage local harmonization, embrace diversity, and explore techniques for greater, domain-overreaching harmonization. Regarding sustainability, this further requires separating spaces prone to change from those needing sustained operation—allowing competitive engineering solutions to compete and replace each other without imposing costly changes in the research space. Success is measured by the effective decoupling of introducing new technologies into research practice with no (or minimal) effect to the sustainable operation.

System interfaces with Uniform semantics

In systems engineering, the need to separate concerns between functional spaces equates to defining interfaces. FAIR-EASE engages in this exercise, providing an architectural map with key dividers—the [[Data Provider]] and [[Data Access]] interfaces. The first captures how to (a) provide data (input side) in an open, yet structured and linkable way; while the second describes how to (b) effectively extract from the assembled pool of data those relevant subsets, slices or fragments (output side) to work on inside the VRE. These interfaces aim to operate without assumptions about specific technical solutions.

While aiming for “no” assumptions is idealistically lowered for practical realism, we agree on safe, low-cost assumptions inspired by and reusing commonly recognised success elements from web-architecture—reusing URI addressing, HTTP(s) protocols, and replacing HTML with RDF. So embracing the idea to publish structured data using the basic rules of the Semantic Web. On that basis, the aim of bridging the demand (data-access) to the distributed, independent,

and cross-domain offer (data-provider) in an effective way is forcing us into designing RDF expressions for both the exposed catalogue of datasets and the requested data-frames, proposing a prototype solution for both from existing vocabularies like DCAT, DCTerms, CSVW, and schema.org.

Research questions

The design effort is distinctively modelling the two viewpoints: (i) the data access (demand) point-of-view, being that of the VRE users; and (ii) the data provider (offer) point-of-view, about what is being made available by the data infrastructures. Key questions include:

- What RDF schemas can bridge the gap between different domains to make data (and their descriptions) from behind a wide range of data access services discoverable and accessible?
- What are the most important metadata elements from a researcher point of view? And how should these be encoded so automated assistance can be provided to them.
- What as-is (meta)data are provided by the data infrastructures? And how should they adapt to support sustainable future rehashing.

During this IMDIS session we plan to explain the approach and the latest results of the research and prototyping work within FAIR-EASE.

List of Abbreviations

CSVW (Comma-Separated Values on the Web)

DCTerms (Dublin Core Terms)

DCAT (Data Catalog Vocabulary)

HTML (Hypertext Markup Language)

HTTP(s) (Hypertext Transfer Protocol Secure)

IMDIS (International Conference on Marine Data and Information Systems)

RDF (Resource Description Framework)

URI (Uniform Resource Identifier)

VRE (Virtual Research Environment)