

## Enhancing Virtual Research Environments for Environmental Research Using Advanced AI

Virtual Research Environments (VREs) constitute an essential aspect in attaining globally impactful research. With rapid environmental changes worldwide, including global warming, significant challenges that require collaboration between the environmental sub-domains need to be addressed. While workflows from individual Research Infrastructures (RIs) exist, environmental change is a global phenomenon that requires massive collaboration on a global scale. To achieve this level of collaboration and knowledge exchange at scale, federated access to computing resources, shared datasets or metadata, and shared analysis codes are vital assets for research infrastructures.

The ENVRI-Hub NEXT project aims to address these collaboration challenges by providing seamless access to environmental data and services across multiple European Environmental Research Infrastructures and by addressing key challenges in environmental data discovery, access, composition, processing, and usability. It enhances the collaboration and interoperability of RIs by providing tools for secure authentication, automated data processing, and search, while supporting virtual research environments to facilitate reproducing research analyses.

The ENVRI-Hub provides searching capabilities to locate environmental content, including data, metadata, or notebooks, in three different ways. It offers a classical search system powered by a Large Language Model (LLM), which uses content from ENVRI-Hub to provide users with a summarised response followed by content categorised into four classes: i) Web Content, ii) Metadata (supplemented by Catalogue of Services for expert users), iii) APIs, and iv) Jupyter Notebooks. Furthermore, it offers a dialogue-based search system, which complements classical search with a natural, conversational style, allowing users to discuss and analyse content from ENVRI-Hub. Finally, it provides a virtual research environment to run and analyse the notebooks retrieved by the search systems, allowing exploration of datasets, metadata, and APIs. Overall, to consolidate all services, a federated authentication system is implemented, enabling secure authentication for any user and allowing RIs to index their data while adhering to FAIR principles seamlessly.

The ENVRI-Hub primarily provides a unified search feature across 13 RIs. This feature supports researchers affiliated with these research infrastructures and independent scholars in searching for environmental data and metadata. It helps them learn about dataset attributes and obtain downloadable links to datasets and associated notebooks for analysis.

By implementing the FAIR principles end-to-end, the ENVRI-Hub transforms heterogeneous RI assets (datasets, metadata, APIs, web contents, and notebooks) into Findable (enriched indexing across 13 RIs), Accessible (federated AAI), Interoperable (cross-disciplinary index mappings), and Reusable resources (reusable Jupyter notebooks), powered by generative LLMs. This FAIR-by-design enables researchers to seamlessly integrate multimodal content, including text, data, and Jupyter notebooks, thereby enabling reproducible research analyses and tackling environmental challenges globally. In summary, the ENVRI-Hub realises the VRE vision by enabling scientific knowledge through secure, federated access to heterogeneous resources, while hiding infrastructure complexity and accelerating reproducible and collaborative environmental science.

**Primary authors:** ISLAM, Nafis Tanveer (University of Amsterdam); DRAGO, Federico (EGI Foundation); GUTIERREZ, Marta (EGI Foundation); DOLBER, Delphine (Euro-Argo ERIC); BUNDKE, Ulrich (F. Juelich); ZHAO, Zhiming (University of Amsterdam)

**Presenters:** DRAGO, Federico (EGI Foundation); ZHAO, Zhiming (University of Amsterdam)

**Track Classification:** Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)