

ENVRI KNOWLEDGE BASE

An Intelligent System for Discoverability and Collaboration

ENVRI-Hub Services Involved



AI Environmental
Expert



Knowledge
Base

Main Partners Involved



UNIVERSITEIT
VAN AMSTERDAM

Stakeholder Groups

- Research Infrastructures and the wider scientific community
- EOSC and related initiatives
- Policy makers and public authorities
- Environmental and biodiversity data users
- Civil society, NGOs and citizen science initiatives
- Industry and private sector actors

User Types

- Researcher
- Professor
- Data Scientist
- Project Manager
- Lab Manager
- Industry Researcher
- Policymaker

Purpose

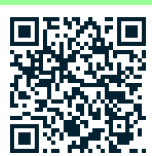
The ENVRI Knowledge Base is designed to provide seamless access to environmental data and services across multiple European Environmental Research Infrastructures (RIs), addressing key challenges in data discovery, access, composition, processing, and usability. The knowledge base enhances the collaboration and interoperability of RIs by offering tools for EOSC-compatible authentication, automated data processing, and search, while supporting virtual research environments (VREs) to facilitate reproducible research.

Task

The task of the project is to implement the FAIR principles end-to-end, the ENVRI-Hub as a VRE, 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 Large Language Model (LLMs).

In summary, the ENVRI-Hub aims to achieve its vision by gaining scientific knowledge with secure and federated access to heterogeneous resources while hiding infrastructure complexity and accelerating reproducible and collaborative environmental science at scale.

Such collaboration is essential for calculating Essential Climate Variables (ECVs) and informing disaster risk reduction, biodiversity management, and water and air quality policies in the face of rapid global environmental change.



Demo

Scan to view our demo
<https://youtu.be/ThbDTB8Mh4g>

Challenge

An early career scientist may want to calculate the ECV of a particular location at a certain point in time. However, this requires multiple steps to achieve. Initially, the early-career scientist has to find the correct dataset. To do that, they have to analyse all the RIs to find the correct data of a specific location at a certain time. Then, afterwards, the researcher has to find a method to calculate the ECV. Finally, they need to find the correct code to read the dataset and calculate the ECV with the method.

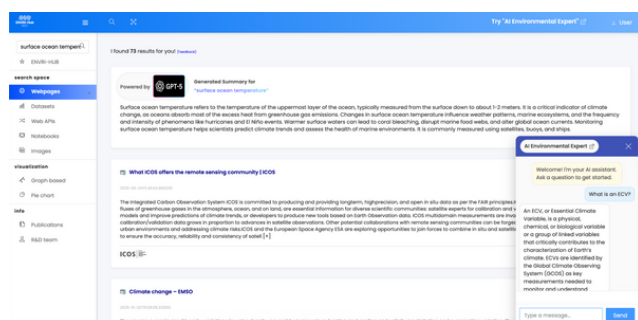


Figure 1: Searching in the Classical Knowledge Base. The AI Assistant on the bottom right helps users with basic questions.

Results

- Easy searching of contents
- Get easy access to metadata
- Easy to find working Notebooks
- Coding assistant from scratch or debugging
- Knowledge enhancement using an intelligent dialogue-based system

Solution

As a Knowledge Base, 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 an LLM, which provides users with a summarised response followed by a list of contents categorised into four classes: i) Web Content, ii) Metadata, iii) APIs, and iv) Jupyter Notebooks. Furthermore, we also offer a dialogue-based search system, which complements the classical search with a dialogue-based style of natural communication, allowing users to discuss and analyse content with follow-ups. Finally, we provide a virtual research environment to run the shared notebooks, which we found from the search system to explore the datasets, metadata, and APIs. Overall, to consolidate all the services, a federated authentication system is implemented that enables RIs to easily and seamlessly index their data while adhering to FAIR principles.

Environmental Expert...

Ask any question about the Environment and Earth Science



Write the code to calculate the ECV "Surface Ocean Temperature" using the data from python ENVRI Libraries.



To calculate the Essential Climate Variable (ECV) "Surface Ocean Temperature" (also known as Sea-Surface Temperature) using data in Python, you would typically leverage data access and analysis functions provided by ENVRI libraries. Although no exact ENVRI code snippet is listed above, here's a generic Python example using popular scientific libraries, which is generally transferable to ENVRI-related data (like from NetCDF, OPeNDAP, or THREDDS endpoints):

```
import xarray as xr # Common for opening NetCDF data from marine observati

# Open a dataset
ds = xr.open_dataset('path_or_url_to_your_surface_ocean_temperature_data.nc')

# Assume 'sst' is the variable name for Sea-Surface Temperature
sst = ds['sst'] # Change variable name based on your dataset

# Calculate basic statistics (mean, min, max, etc.)
mean_sst = sst.mean(dim=['time', 'lat', 'lon'])
print("Mean Sea-Surface Temperature:", mean_sst.values)
```

What questions do you have about Environment?



Figure 2: AI Environmental Expert, a dialogue-based system which can answer any questions related to the environment, including finding datasets/metadata, notebooks and assist in programming.