ENVRI-Hub Design and Architecture White Paper

Author(s)	Andreas Petzold, Ana Rita Gomes, Ulrich Bundke, Chris Schleiermacher, Angeliki Adamaki, Alex Vermeulen, Zhiming Zhao, Markus Stocker, Cathrine Lund Myhre, Damien Boulanger, Anca Hienola, Daniele Bailo, and the ENVRI-FAIR Task Force 6 Team
Lead partner	FZJ
Work Package	WP5
Status	Final
Date	16 June 2023

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

TABLE OF CONTENTS

Rationale

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). The high-impact ambition of ENVRI-FAIR is to establish the technical preconditions for the successful implementation of a virtual, federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRIs. This interface is called the ENVRI-Hub. Full integration of services across RIs and even subdomains is continuously progressing at ENVRIs with focus on environmental data and scientific research objectives. Each RI is specialized in a specific number of parameters related to its specific competences. For users that require a broader or the full spectrum of environmental parameters, the ENVRI-Hub will offer a platform that reflects the complexity and diversity of the ENVRI landscape, while preserving their specific structures and addressing the requirements they were designed for.

The ENVRI-Hub will be a federated system of harmonized subdomain/RI-specific systems of data policies and management, access platforms and virtual research environments. The system will be completely open source, modular and scalable and build on the experience available in the consortium and already operational systems. Figure 1 shows the key features of the ENVRI-Hub.

The ENVRI-Hub community metadata and data store is foreseen to be based on semantic web technology, ontologies, and open linked data, allowing integration of the vocabularies and metadata standards developed in the implementation work packages of ENVRI-FAIR. The cross-subdomain development will create large benefits in efficiency and robustness of the ENVRI-hub system. It will enable true interoperability of access to metadata and data objects across the RIs and subdomains and thus facilitate the development of (joint) higher-level services. By following a strict modular design, the developments can be (re-)used in the different RIs. All components should contain couplers to the EOSC services, like AAI, data storage solutions and catalogue of services, to enable the seamless integration of ENVRIs data and services into EOSC.

Figure 1. Key features of the ENVRI-Hub

1 Drivers of ENVRI-Hub Design and Architecture

Sharing open data and related data-based and science-based services is one of the major purposes of EOSC, and Science Clusters are key providers of content to EOSC.
User and stakeholder requirements determine the ENVRI-Hub functionalities and architecture.
In its final stage, ENVRI-Hub functionalities and content will include:

- discovery of data provided by ENVRIs,
- discovery of services and products provided by ENVRIs,
- facilities for cross-domain search,
- discovery of and access to thematic services of research infrastructures,
- the ENVRI Training Platform,
- the ENVRI Science Demonstrators, and
- an ENVRI Virtual Research Environment.
- ENVRI-Hub will serve as a common portal to RI digital assets.
- ENVRI-Hub will serve as a complementary access path to the RI data products and services, with direct RI access pathways kept in place and functioning.
- ENVRI-Hub swill become part of the EOSC Thematic Service Catalogue, Thematic Data Source Catalogue and Thematic Research Data Catalogue.
- ENVRI-Hub will be connected via the EOSC Interoperability Framework.

Requirements for the ENVRI-Hub design and architecture are as follows:

- Technically, the architecture and functionalities of the ENVRI-Hub will be driven by the applications, or science demonstrators and user requirements, respectively.
- Interoperability on (ideally) all levels is necessary.
- Science demonstrators drive the design of the ENVRI-Hub. The defined demonstrators are linked to the activities in the EOSC Future project.
- Science demonstrators are designed along cross-subdomain science questions, e.g., climate change, air pollution, interaction of air pollution with climate change, spreading of diseases, changes in biodiversity, and others.

2 User Requirements

A survey of user requirements provided a series of generic use cases to be covered by the ENVRI-Hub functionalities and architecture. These use cases include (*Figure 2* to *Figure 6*):

2.1 Basic Data Discovery

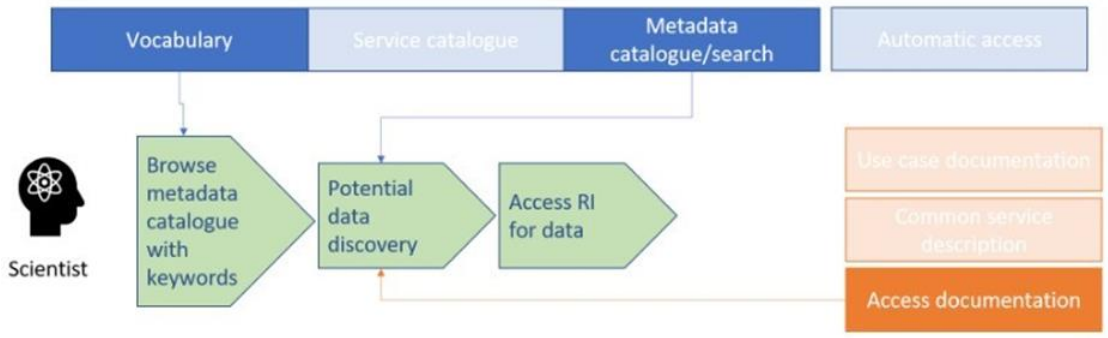


Figure 2. Basic Data Discovery

2.2 Basic Service Discovery

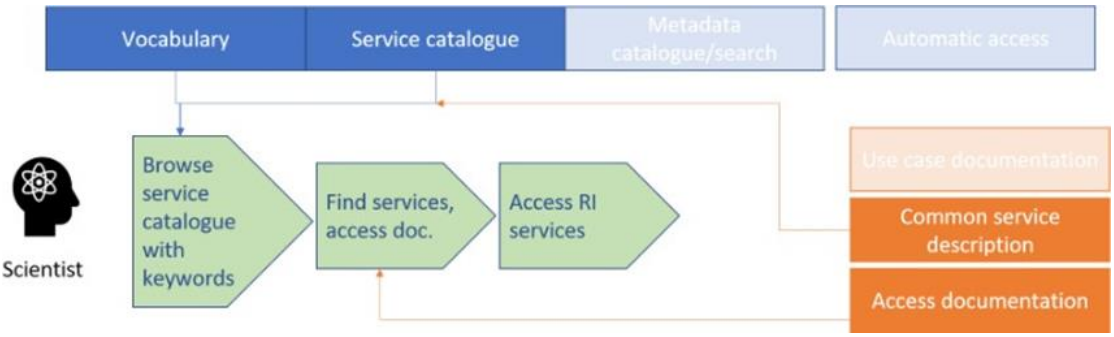


Figure 3. Basic Service Discovery

2.3 Advanced Data Discovery and Use

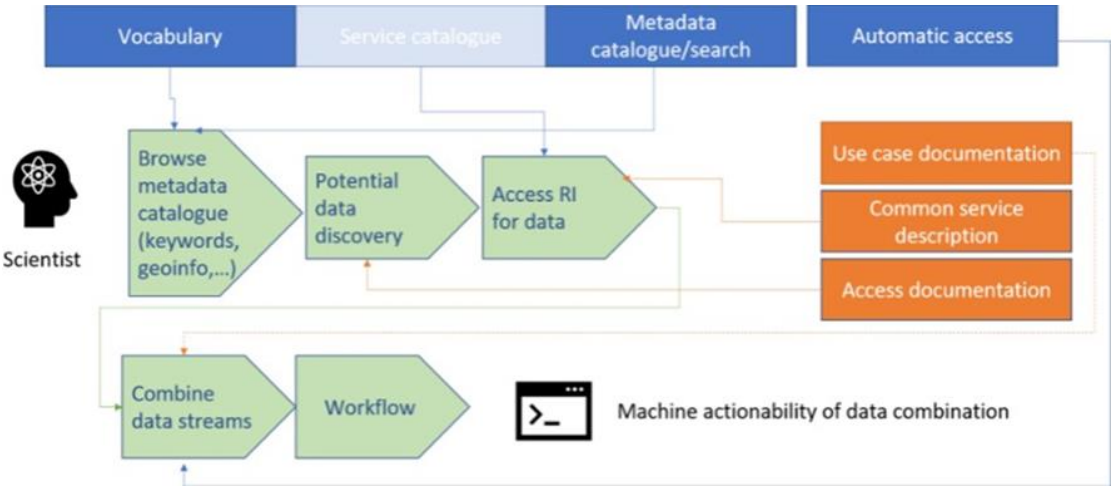


Figure 4. Advanced Data Discovery and Use

2.4 Advanced Service Discovery and Use

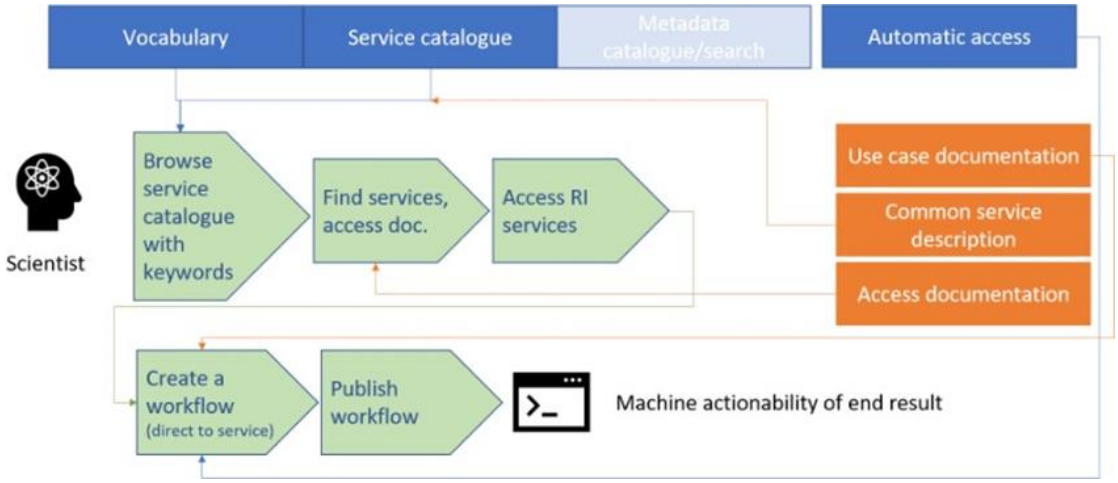


Figure 5. Advanced Service Discovery and Use

2.5 Very Advanced Service Discovery and Use

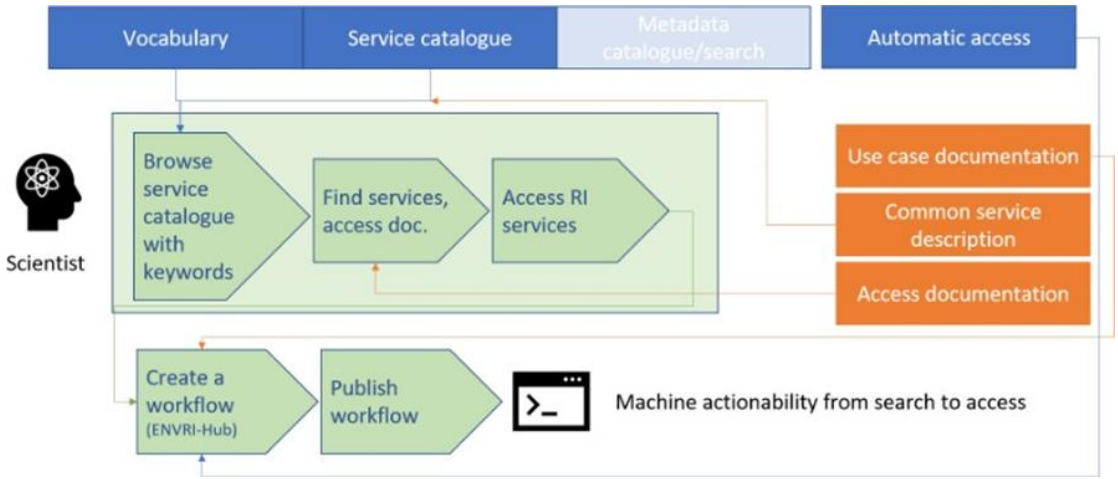


Figure 6. Advanced Service Discovery and Use with data through the ENVRI-Hub

3 ENVRI-Hub components

The ENVRI-Hub should be built from four main pillars:

- The ENVRI Knowledge Base as the human interface to the ENVRI ecosystem;
- The ENVRI Catalogue as the machine actionable interface to the ENVRI ecosystem;
- The ENVRI Training Catalogue as a base for the ENVRI training resources;
- Sub-domain and cross-domain Science Demonstrators showing the capabilities of service provision among ENVRIs and across Science Clusters.

Additionally, the ENVRI-Hub aims to provide Services, e.g., virtual research environments, workflows, or customized pipelines for automated access to data.

3.1 ENVRI Knowledge Base

The ENVRI Knowledge Base aims at sharing technical practices, identifying common data and service requirements and design patterns, and facilitating search and analysis of existing RI solutions for interoperability challenges that are shared among environmental RIs. ENVRI-FAIR Deliverable D7.3 on the ENVRI-FAIR Knowledge Base V1 [https://doi.org/10.5281/zenodo.4311047] introduces the building of the Knowledge Base in the ENVRI-FAIR context, describes the chosen approach for the knowledge construction, its support for sharing technical practices, identifying common problems and solutions, searching existing solutions for interoperability challenges among environmental RIs, and knowledge-based decisions.
The ENVRI Knowledge Base serves as interface to the ENVRI ecosystem for:

- quick discovery of data, knowledge, services and assets;
- sharing of engineering practices, technologies and knowledge;
- enhancing collaboration among communities with customisable knowledge subscription and publishing pipelines.

The components of the ENVRI Knowledge Base are illustrated in *Figure 7*:

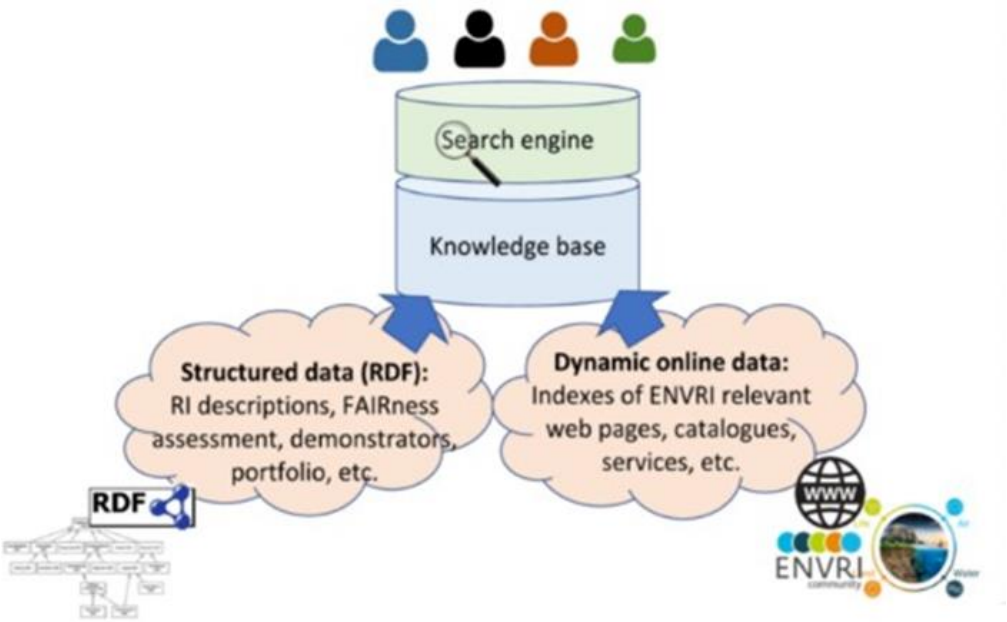


Figure 7. Knowledge Base content components: a Knowledge Base search engine is provided for different end users to search different contents

Key functionalities of the Knowledge Base include:

Quick information discovery:

- Dialogue interface with the Knowledge Base.
- Obtain machine readable information for scientific workflows.

Sharing engineering practices:

- Publish valuable community knowledge (e.g., FAIRness status, new technologies, training materials, services, etc.) to the Knowledge Base.
- Allow users to discover the newly published knowledge via a coherent interface.
- Notify relevant users of new updates.

Enhancing collaboration (planned):

- A user can subscribe to changes of multiple knowledge sources; the Knowledge Base will automatically notify the user of new updates.
- An expert can register as an expert of the Knowledge Base.
- A user can post questions via the Knowledge Base, and the Knowledge Base will in- vite registered experts to answer the questions; an incentive model will be provided.

Version one of the Knowledge Base focuses on the indexing and searching functionality.

Version two (in preparation) will enrich the content of the Knowledge Base, fix bugs, and add new features requested from the community.

The enterprise view shown in *Figure 8* highlights the different key stakeholders and their interaction scenarios. The numbered circles indicate the possible orders of the interactions.

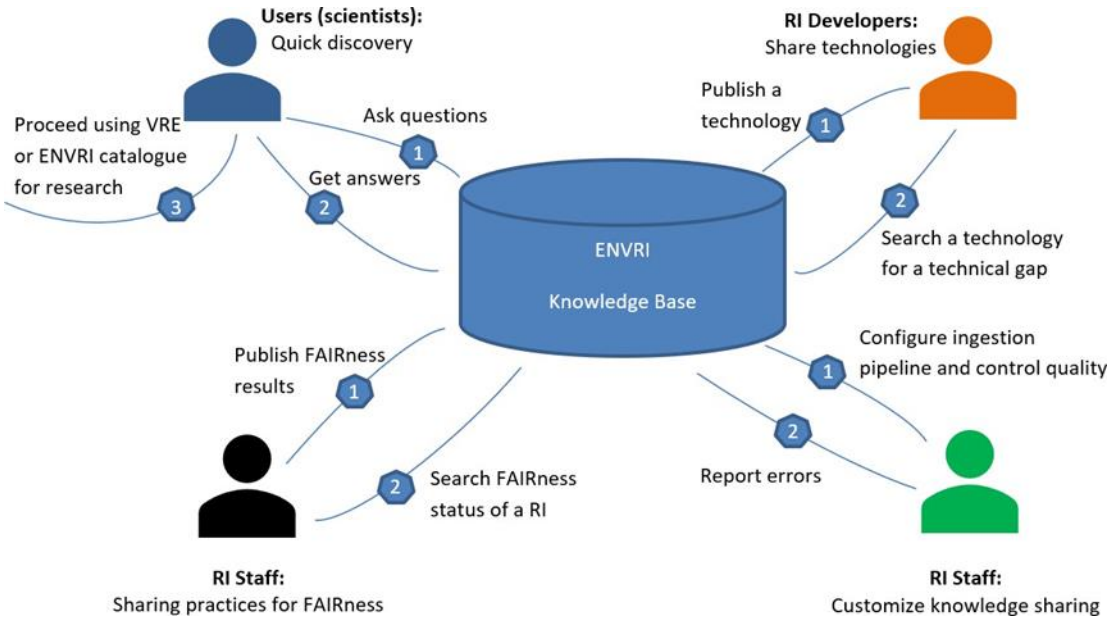


Figure 8. An enterprise view of the ENVRI Knowledge Base: Key stakeholders and their interaction scenarios, numbered circles indicate the possible orders of the interactions

3.2 ENVRI Knowledge Base's integration with the ENVRI-Hub

As referred in section 3.1, the Knowledge Base will be one of the ENVRI-Hub's pillars acting as an enabler interface to the whole ENVRI ecosystem.

For the integration of the Knowledge Base in the ENVRI-Hub, the architecture will focus on the Dynamic Online Data, which is searchable via the Search Engine, and the Structured Data, stored and served by a Triple Store.

Dynamic Online Data: This data is indexed and ranked, and becomes discoverable by the Search Engine, allowing users to perform a search on the online resources, by providing a search criterion; see Figure 9 for illustration.
It is foreseen that the ENVRI-Hub will adopt and integrate the Search Engine, therefore providing a means for the user to have this **search functionality at the ENVRI-Hub level**.
The user will be able, thus, to use the dynamic online data search functionality by accessing for example a search widget in the ENVRI-Hub and perform a query about a desired resource.

Structured Data (RDF): As opposed to the data that feeds the Search Engine, the Structured Data is not dynamic, and comprises topics like RI descriptions, FAIRness assessment, demonstrators, etc. This data is persisted in a Triple Store and exposed both via SPARQL endpoint and via REST service, due to the nature of *Apache Jena Fuseki* SPARQL server; see Figure 10.
The ENVRI-Hub is also running an instance of *Jena Fuseki* on its side, therefore being able to query the Knowledge Base's Triple Store, persist the data, and display it to the user.
It is not yet defined how the Structured Data will be displayed to the user – some interesting examples like FAIRness assessments can be done, instead of displaying the content like a "textbook".

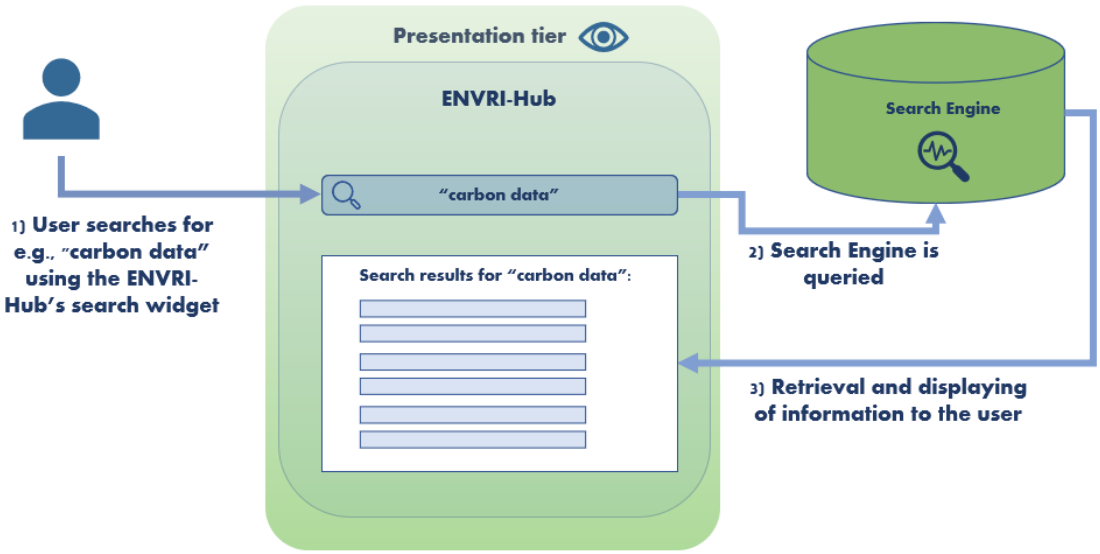


Figure 9. Integration of the Search Engine within the ENVRI-Hub. The user searches for a desired resource in the ENVRI-Hub's GUI and the Search Engine is queried, retrieving the corresponding result set

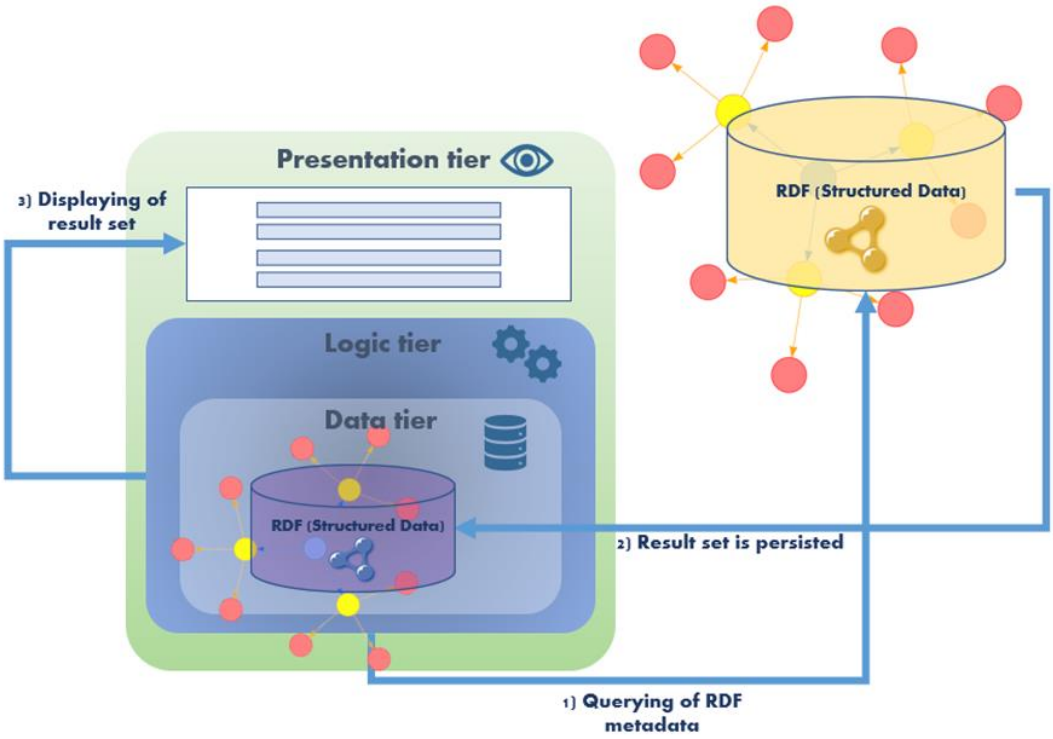


Figure 10. Querying of RDF data by the ENVRI-Hub. 1) ENVRI-Hub queries RDF data on the Knowledge Base's Triple Store; 2) The Knowledge Base retrieves the requested data, and the result set is persisted in the ENVRI-Hub's Triple Store; 3) The result set is displayed to the user

3.3 ENVRI Catalogue

ENVRI-FAIR has two major objectives concerning the ENVRI RIs that have assets of interest to share: (1) to make those assets available FAIR and (2) to make those assets visible to EOSC users and deployable to EOSC.
The canonical ENVRI catalogue of assets (starting with services for EOSC compatibility) is thus an essential component of the ENVRI-Hub architecture (and hence the relationship of the ENVRI-Hub to the ENVRI RI e-Infrastructures) allowing interoperation among ENVRI RIs and interoperation between ENVRI and EOSC. ENVRI-Hub will serve as the machine actionable interface to the ENVRI ecosystem and is characterised by the following functionalities:

- cataloguing all RIs in the ENV domain;
- catalogue should not map all metadata, but describe services that provide access to data;
- starting point for accessing RIs datasets via metadata search – no direct access to data;
- interface to the European Open Science Cloud and other users (e.g., Copernicus).

The design of the service catalogue is described in the ENVRI-FAIR Milestone Report MS18 [https://zenodo.org/record/4024173] and its schematic design is shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**. The core element is the mapping ofmetadata which requires agreement on a harmonised metadata schema and a metadata exchange protocol (Figure 12). This agreement needs to be reached and implemented across the RIs; it is where the challenge lies.

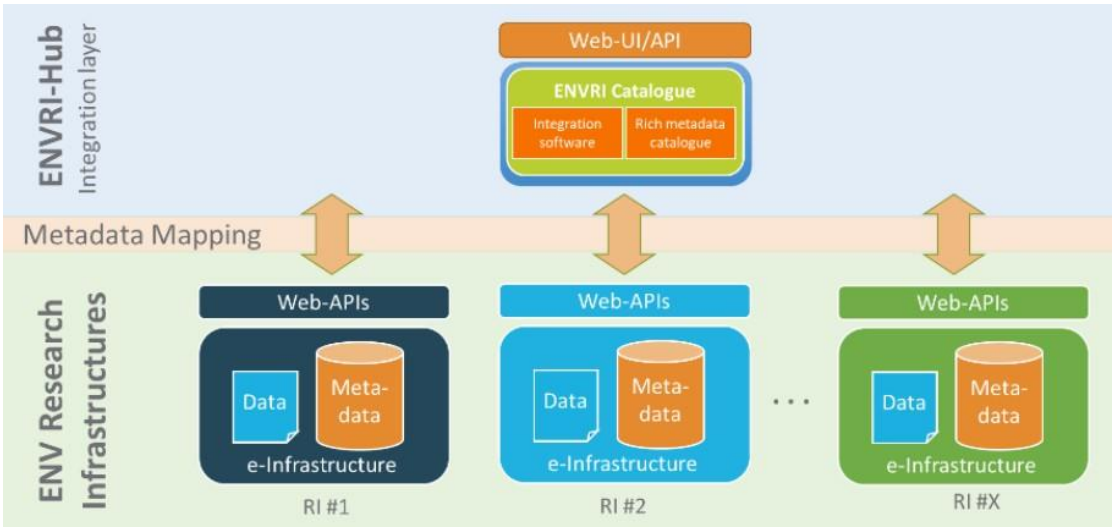


Figure 11. Schematic design of the ENVRI Catalogue

It has to be noted that:

- The integration layer does not mean that EOSC has to go through this layer to access RIs;
- The integration layer contains a catalogue of services where, for each service, information about how to access and consume the data is presented by means of Web APIs.
- Once EOSC (or any other external consumer) has this information, then it can access the services from the RIs directly.
- The catalogue responds to the problem: "how can I, as an external agent, know how to access the RIs assets in a machine-readable way?"

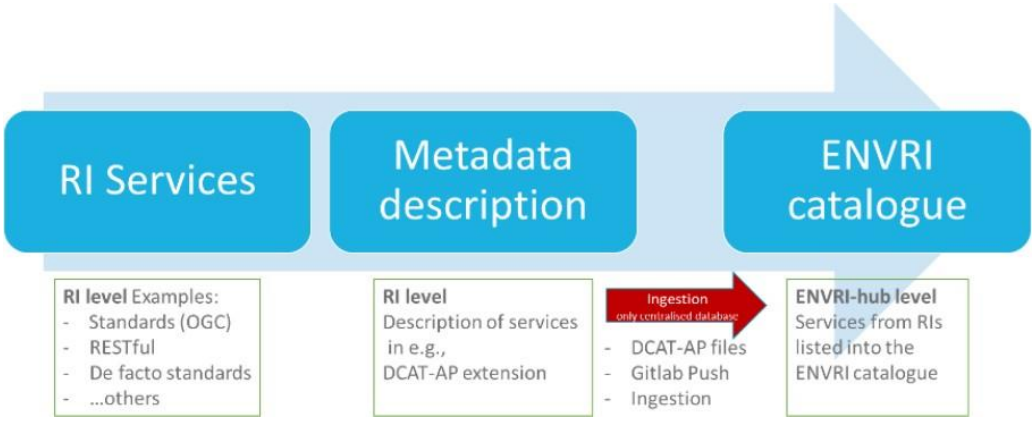


Figure 12. Schematic ENVRI Catalogue of Services proof of concept

3.4 ENVRI Catalogue's integration with the ENVRI-Hub

Acknowledging the ENVRI Catalogue as another pillar of the ENVRI-Hub, its integration is a vital part of this project.
It is clear at this stage of development that the ENVRI-Hub will query the ENVRI Catalogue's metadata and display it to the user.
Is it not yet defined whether the ENVRI Catalogue will have its own GUI, so how to display the metadata in the ENVRI-Hub is still under discussion.

A proposed solution was to adopt the interface of the EPOS' Portal as an interface for the ENVRI Catalogue in the ENVRI-Hub.
In order to access the Catalogue's machine actionable interface, the ENVRI-Hub will query a PostgreSQL database, where the Catalogue's resources are persisted. A PostgreSQL instance is also running at the ENVRI-Hub level, to make this connection have the least possible overhead.
Interactions between both entities – ENVRI Catalogue and ENVRI-Hub – are illustrated in Figure 13.

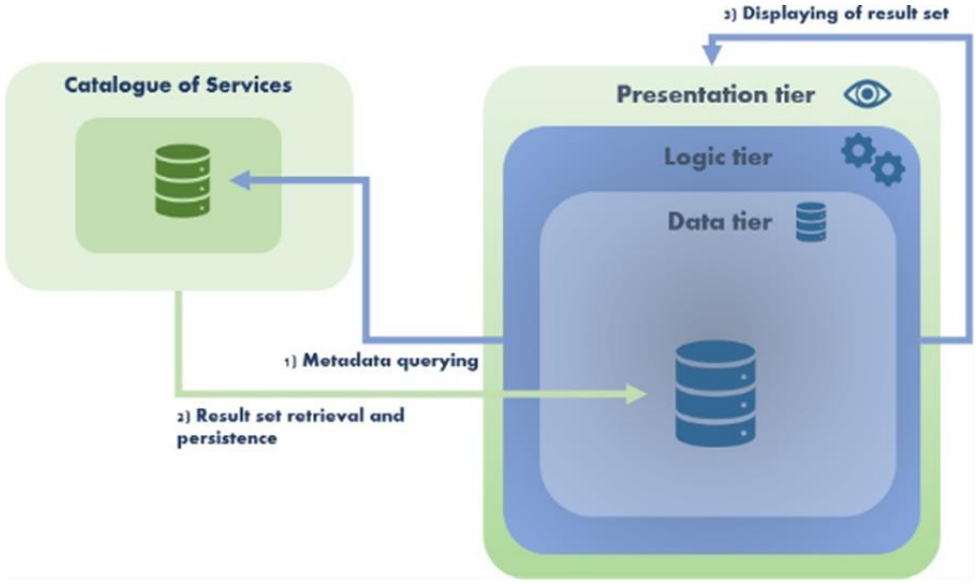


Figure 13. Querying of metadata by the ENVRI-Hub. 1)ENVRI-Hub queries metadata to the ENVRI Catalogue; 2) ENVRI Catalogue retrieves the requested metadata, and the result set is persisted in the ENVRI-Hub; 3) The result set is displayed to the user

3.5 ENVRI Science Demonstrators

The ENVRI Science Demonstrators and the Science Projects in the Horizon 2020 project EOSC Future aim at demonstrating how joint projects can address major challenges for Europe's societies and how research infrastructures can support Horizon Europe's missions within the EOSC. Use cases are designed preferentially across subdomains to foster interoperability of RIs.

Identified scientific demonstration cases are:

- **FAIR ENVRI atmospheric data demonstrator (Atmosphere subdomain):**
  FAIR ENVRI atmospheric data demonstrator: This demonstrator aims to shorten time response in providing scientific analysis of an extreme event and harmonized dataset and tools. It will offer the search for the ATMO-RIs data availability for the period and area of interest with a focus on Essential Climate Variables (ECVs), fetch the selected data including automatic previews of the datasets and provide statistical analysis on the downloaded time series. In addition, access to data from ATMO-RIs is offered as a bundle with automated compilation of provenance information, and a co-location service with satellite data is available.

- **ICOS Carbon Portal Footprint tool (cross subdomains):**
  Users can request atmospheric surface emission influence calculations (a.k.a. footprints) for any 3D location or specific station in the model domain and any period for which meteorological data are available. When ready, the results can be viewed in the tool and directly accessed through our Jupyter notebook service and Python library, to be combined with emission data for concentration calculations at the receptor(s) and comparison with observations retrieved from the RI data repositories.

- **Essential Ocean Variables global product (Marine subdomain, Task 9.8):**
  Essential Ocean Variables broker links queries on EOVs to Marine RI data servers considering RI-specific vocabularies.

- **Soil Water Content:**
  Soil, water content (SWC) is a key environmental variable, important to, e.g., farmers, meteorologists, water resource management and disaster management units. Global soil moisture products (remote sensing, SWC monitoring network (ISMN)) remain limited to the soil surface and/or provide little information on contextual data that drive SWC temporal dynamic. The SWC use case will deliver comprehensive and interoperable data sets including SWC together with context data and metadata from different RIs.

Exposed EOSC Future Science Projects are:

- **Dashboard State of the Environment:**
  Each Research Infrastructure working in the ENVRI-FAIR science cluster gathers and manages a wealth of environmental data and model results, divided over four subdomains and thus the Earth system in its full complexity. Keeping a close watch on environmental boundary conditions and informing society stakeholders on short- to long-term developments therein, will be of imminent economic value. This can be implemented by means of a 'dashboard' with a set of easily understandable real-time indicators which inform public and policy makers on the state of the environment. This dashboard should be mounted as a front-end of the ENVRI-Hub, a virtual common platform for ENVRI-FAIR data and services and integrated in the EOSC portal. The primary objectives are: 1) to develop and launch a dashboard for environmental indicators by setting up analytical workflows for different environmental disciplines and integrating their outputs; 2) to connect the analytical framework to the EOSC portal by means of the ENVRI-Hub in order to mobilise and empower a larger community of researchers and potential data providers; and 3) to demonstrate and promote the benefits and potential of web-based science using EOSC.

- **Impact of Non-Indigenous Invasive Species (ENVRI-FAIR – EOSC-Life):**
  Several reports (e.g., IPCC, UN Global Assessment Study) and conferences (e.g., COP21) demonstrate that climate change poses severe threats to our life supporting system, the biosphere. We need to improve our current knowledge by seeking cross-domain analyses. A promising option is to investigate the rapid increase of Non-Indigenous Invasive Species (NIS) in European ecosystems. These species may not only replace indigenous species but also alter habitats, interact with the changing environment, and eventually severely influence established socio-economic regimes. The challenge here is to adopt a comprehensive approach by considering the bulk of the biotic and abiotic variables and their interactions, which may be even more important than the occurrence of the NIS. Such approaches require access to big

datasets (from genomics to in-situ and satellite borne environmental data) and high computational power, especially for those models with iterative algorithms. The overarching objectives are: 1) to integrate different scientific disciplines in the marine subdomain (e.g. chemistry, physics, biodiversity, ecosystems, genomics, socio-economics) into an analytical frame-work to progress our knowledge about the impact of NIS on European marine bio-diversity and ecosystems; 2) to connect the analytical framework and federate access to relevant data infrastructures at the EOSC portal in order to mobilise and empower a larger community of researchers and data providers; and 3) to demonstrate and promote the benefits and potential of web-based science using EOSC.

The use cases are made accessible as executable demonstrators and demonstrator VREs, e.g., via Jupyter notebooks. Access is hosted by the ENVRI- Hub and Knowledge Base.

3.6 ENVRI Service and User Hierarchy

From the position of ENVRI-FAIR and the other science clusters, the EOSC infrastructure needs to provide the consolidated cross-European core of e-Infrastructure services: AAI, high performance storage, computing, archiving, simulation, and analysis services. The services of this Minimum Viable EOSC Ecosystem will be requested by the science clusters to operate their data infrastructures and will be contributed by the e-Infrastructures at the level of EOSC Core. On the other hand, the science clusters contribute thematic services to the EOSC Ecosystem which will be aggregated at the cluster level and reflect the needs, requirements and considerable digital assets and services of the associated research communities.

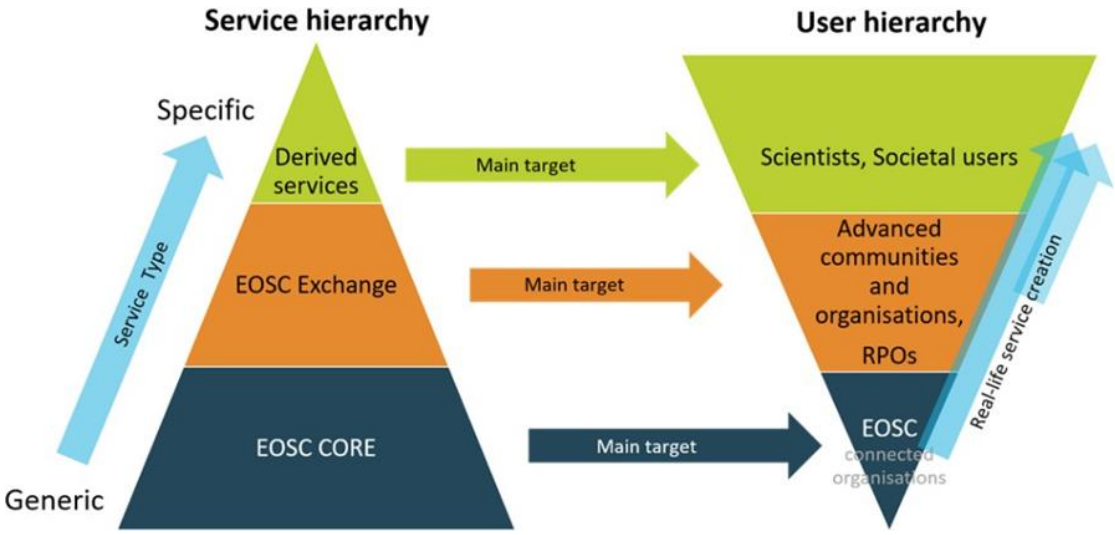


Figure 14. Hierarchy of Services and Users

These thematic services will be further integrated with EOSC Core and horizontal e-Infrastructure services of the EOSC Exchange in the framework of the H2020 project EOSC Future. Expanded horizontal services from the science Clusters will build on work conducted in the science cluster projects and be further expanded in the framework of EOSC Future. The largest fraction of scientific users from the research community will be attracted by the specialized services derived from scientific products provided by the clusters. **Fehler! Verweisquelle konnte nicht gefunden werden.** illustrates the different hierarchies.

4 ENVRI-Hub Architecture

4.1 Introduction

The architecture of ENVRI-Hub builds on the core element of the ENVRI Catalogue with the EN- VRI Knowledge Base and Training Catalogue, and the Science Demonstrators incorporated into the ENVRI Catalogue design. Within the ENVRI Catalogue framework, it was mentioned that making one more portal to give access to the RI portals is not a priority for all our partners, while making a more complex GUI requires time and resources which are not available in the current project. Given the limited amount of resources, the discussion on the ENVRI catalogue was focusing on web APIs as central access tools, also arguing that those can provide solutions towards a federated environment, while a GUI has been considered a potential future project - a proposed idea would be to reuse EPOS Catalogue code-base, as it presents an elevated stage of maturity and it is a suitable for building an integration front-end for the ENVRI Catalogue into the ENVRI-Hub. See Figure 15 for an overview.
Within the ENVRI-Hub framework and as a result of the users' survey, a potential human in- terface was considered. Designing the Knowledge Base as an interface for humans to access the ENVRI digital assets and the ENVRI Catalogue as a machine actionable interface is seen as an ideal starting point for the ENVRI-Hub demonstrator to meet both requirements. A portal GUI to the ENVRI Catalogue may be developed and implemented in a consecutive phase. Current ENVRI-Hub developments are compliant with user's needs and implement a GUI, which serves as the human interface and single-entry point for the several Hub's pillars.

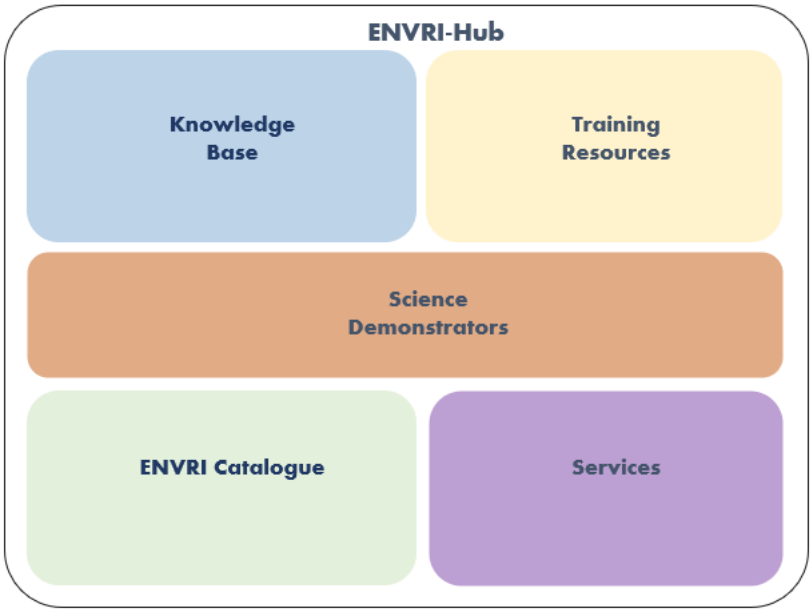


Figure 15. ENVRI-Hub Architecture – 4 pillars of the ENVRI-Hub + Services module

4.2 Architecture Approaches

The ENVRI-Hub architecture needs to consider the design of the ENVRI Catalogue with its core element of mapping the metadata of the individual RIs. In general terms, the ENVRI-Hub can be designed following a centralised approach with the ENVRI-Hub acting as a harvester responding to all incoming requests and supported by a centralised database, or following a decentralised approach with the ENVRI-Hub acting as a broker forwarding all requests to the individual RIs and collecting the responses. In this case, no centralised database is required.

There are certain pros and cons for both design principles:

- The centralised system can be built on the decentralised system of individual RI e-Infrastructures by harvesting the decentralised metadata.
- The centralised system is more demanding concerning sustainability.
- The decentralised system relies on performant decentralised databases.
- Overall, the decentralised system won't perform as well and might be more fragile.

It is worth noting that most of the ESFRIs and EOSC go for a centralised catalogue, and the same is true for commercial large-scale organisations. Given the technical tools available it was agreed to follow a hybrid approach combining the strengths of both designs. In this approach a centralised structure is proposed for all RIs of enough maturity which are capable of providing standardised metadata. For RIs not yet capable of providing standardised metadata, a virtual server can be implemented that the RIs can feed with standardised metadata.

The hybrid approach allows for more flexibility with respect to contributing RIs of different maturity - see Figure 16. For the harmonised metadata schema and exchange protocol, the use of the Data Catalogue Vocabulary (DCAT) [https://www.w3.org/TR/vocab-dcat-2/] was agreed.

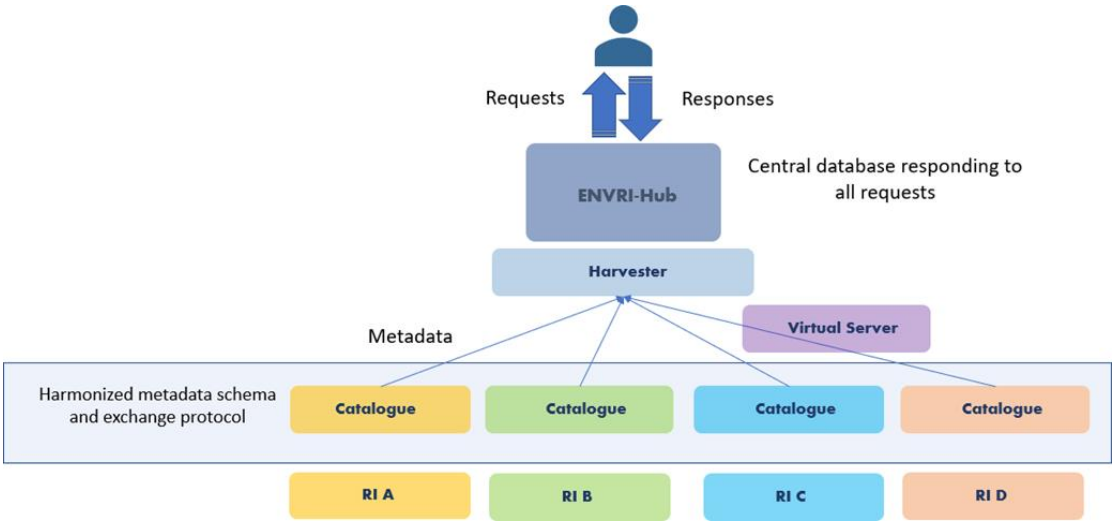


Figure 16. Hybrid ENVRI-Hub architecture approach

Centralised architecture

- Preferred in case all RIs and subdomains provide standardised metadata.
- DCAT was agreed as harmonised metadata schema and exchange protocol, and SPARQL endpoints as search tool whereas an appropriate application profile based on DCAT needs to be agreed upon.

- RIs provide metadata information by creating DCAT descriptions of services - DCAT Profiles.
- Current RI architecture is not disturbed.

Hybrid decentralised approach

- ENVRI-Hub provides for each RI that does not yet have a triple store/SPARQL endpoint, a virtual server that they can feed with metadata in DCAT.
- The virtual server can run in a container from EOSC e-infra, and eventually later be integrated in the RI infrastructure.

Standardised vocabularies for describing metadata

- This topic needs to be further discussed to ensure interoperability across sub-domains.

4.3 Technical Specifications

4.3.1 Chosen architectural pattern

The ENVRI-Hub is developed with a centralized approach. The chosen architectural pattern can be compared to a classical n-tier architecture which divides an application into logical layers and physical tiers – with the difference that the presentation, logic, and data tiers are logically but not all physically separated; see Figure 17.

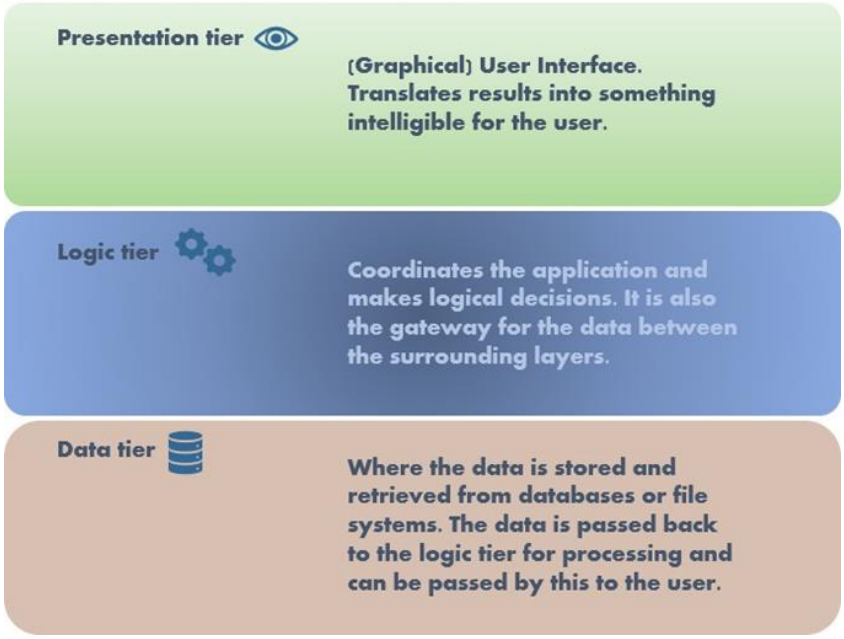


Figure 17. Several layers of the ENVRI-Hub's architecture

4.3.2 DJANGO as a Framework for building the ENVRI-Hub

Django has been established as the framework to build the ENVRI-Hub. Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design.

4.3.3 Metadata Harvesting Schema

Following a Centralised architectural approach, the ENVRI-Hub serves as a central harvester entity, collecting metadata from the ENVRI Knowledge Base and the ENVRI Catalogue - it is not clear at this stage of development if the ENVRI Training Platform metadata will be fetched directly via API or via the Knowledge Base.

- **ENVRI Knowledge Base:** The Static Content (RDF) is persisted in a Triple Store - Apache Jena Fuseki. This Triple Store allows the exposure of resources both via SPARQL endpoint and REST service. Jena Fuseki has also been acknowledged in the ENVRI-Hub as the persistence system to use - mainly to facilitate a compliance between both entities and this Triple Store was also installed in the Hub's side. The ENVRI-Hub will request metadata to the ENVRI Knowledge Base's Triple Store, and this will return either JSON or RDF, which will be persisted on the Hub's Triple Store.

- **ENVRI Catalogue:** Metadata in the ENVRI Catalogue is persisted in a relational database PostgreSQL - and the same technology will be used in the ENVRI-Hub. The Hub's reuse of a Catalogue's already implemented pipeline for verification of updated content has been discussed. The ENVRI-Hub will query the ENVRI Catalogue's database, and the retrieved results will be persisted in the Hub's PostgreSQL database.

- **ENVRI Training Platform:** Training resources are another pillar of the ENVRI-Hub. While it is clear and accepted that they must be a part of the ENVRI-Hub's Key Exploitable Results and will therefore be available at the Hub level, whether this content will be fetched from the Knowledge Base or using the already available Training API in federation with the Knowledge Base's resources is still in discussion. In principle, federation will be needed if the Knowledge Base does not comprise these Training resources – See Figure 19.

- **Science Demonstrators:** Science Demonstrators are being developed by several RIs in parallel. They are the key product to express the ENVRI-Hub's potential regarding easy access to metadata and services, data discovery, as well as the promotion of interoperability in science across sub-domains. Science Demonstrators are built with Jupyter Notebooks - an open-source web application that allows to create and share documents that contain live code, equations, visualizations, and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more. Although they are one of the pillars of the ENVRI-Hub, no metadata flow exists between both entities - the metadata ingestion by these Demonstrators should be provided by the ENVRI Catalogue. The ENVRI-Hub might point to the Science Demonstrators location, providing the user with a URL where to find the resources. More about the Demonstrators is found in section 4.4.

- **Services:** Moreover, the ENVRI-Hub aims to provide a series of resources like time-series synchronisation tools or customised pipelines for metadata access. These resources will be also provided in Jupyter Notebooks and hosted in a Jupyter Hub inside the ENVRI-Hub.

- **Dynamic Online Data via Search Engine:** This data is indexed, ranked, and becomes discoverable via the Search Engine – which will be integrated into the ENVRI-Hub, allowing the user to perform a search into this data at the Hub level.

As previously mentioned, whether the Training resources will be provided by the Knowledge Base or will have to be fetched via their API is not defined at this stage of development.
In the case that they are not provided by the Knowledge Base, the ENVRI-Hub will be responsible for fetching the Training's API (https://trainingcatalogue.envri.eu/site/api).
The corresponding result set can be persisted in the ENVRI-Hub's PostgreSQL database which provides support for JSON data, see Figure 19.

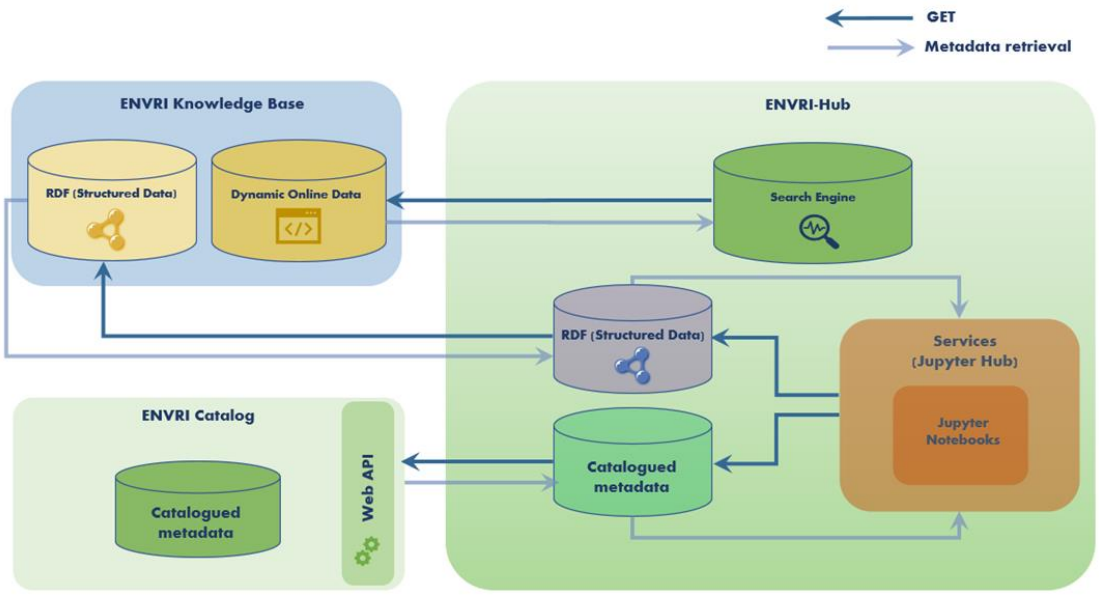


Figure 18. ENVRI-Hub's Metadata Harvesting Schema: Knowledge Base: RDF metadata from the Knowledge Base is fetched by the ENVRI-Hub and replicated in a Triple Store. Dynamic Online Data is queried via the Search Engine; ENVRI Catalogue: Catalogued metadata from RIs is exposed via API by the ENVRI Catalogue, and fetched by the ENVRI-Hub, which will persist it in a PostgreSQL database; Services and workflows: Basically, both structured data and catalogueed metadata can be used by this module. Therefore, both ENVRI-Hub's Triple Store and PostgreSQL database will be queried

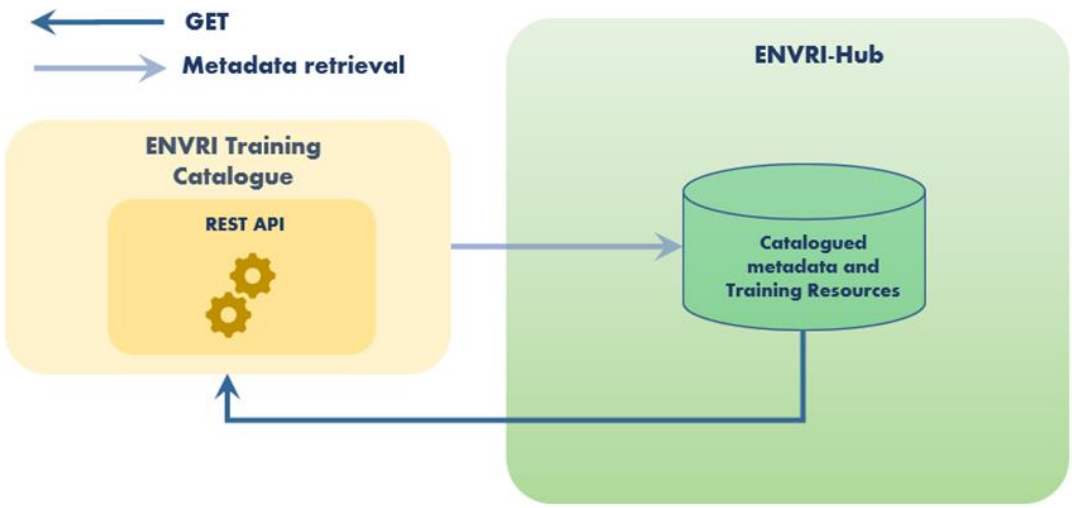
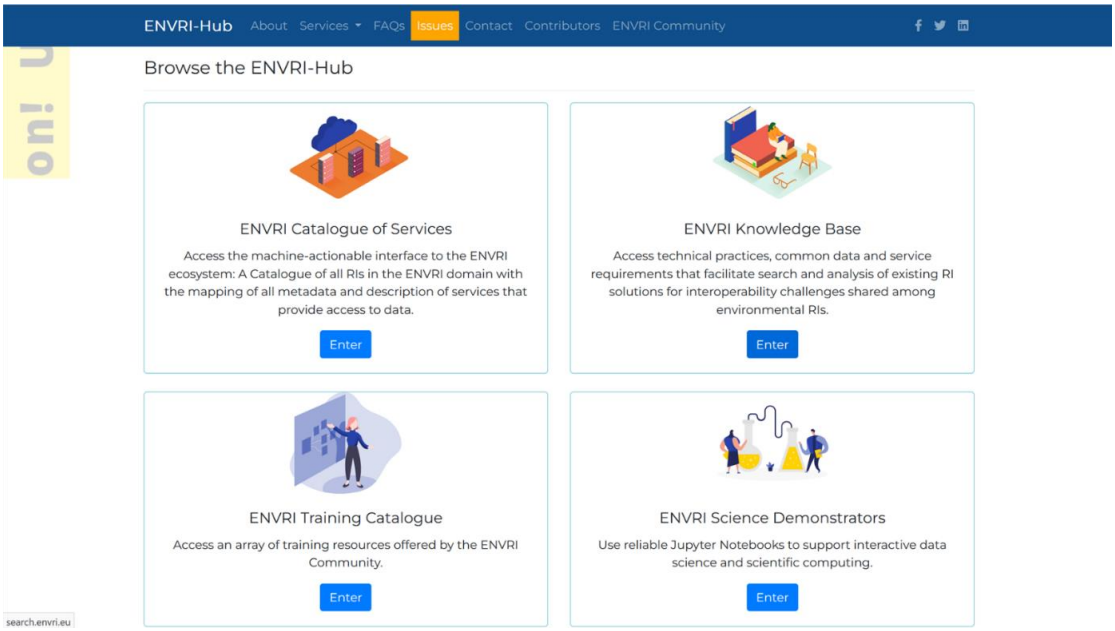
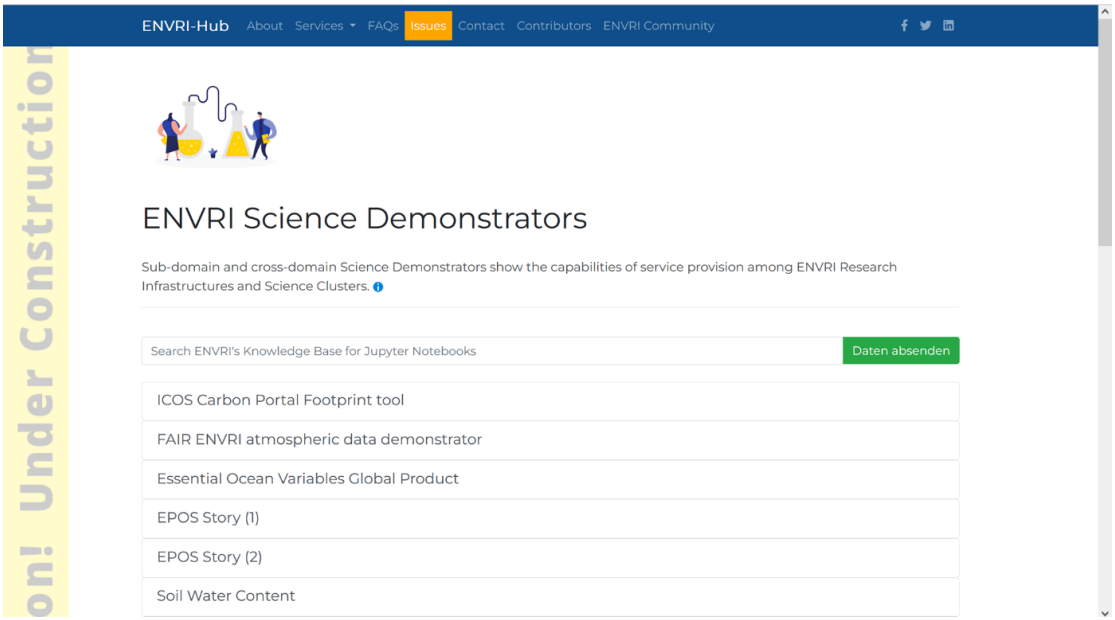


Figure 19. ENVRI-Hub's Metadata Harvesting Schema detail - If the Training Catalogue resources are not comprised in the Knowledge Base, this content will be fetched from the Training's API and persisted into the ENVRI-Hub's relational database – PostgreSQL

4.4 ENVRI-Hub Implementation

A test version of the ENVRI-Hub demonstrator is accessible via https://hubtest.envri-fair.eu/.





Offered Services include:

- ENVRI Catalogue of Services
  - SPARQL Search (rdflib)
  - SPARQL Search (Fuseki)
  - Check availability
- ENVRI Training Gateway (API Search)
- ENVRI Knowledge Base
- ENVRI Science Demonstrators

4.5 Future Functionalities

The preliminary concept of the ENVRI-Hub comprises facilitated and direct access to meta data. Regarding time constraints, it is accepted that the ENVRI-Hub, supported by its pillars, will provide the different users with metadata – that will facilitate the discovery and access of data and services.

Given that the current architecture is adequately modular and scalable to allow the implementation of new functionalities, a desired feature is the establishment of access to data.

Topics like RI homogenization at data delivery and authentication levels must mature for RIs to be able to establish a standardized mean of data provision, and for the ENVRI-Hub to harvest.

A social component has been briefly discussed – Providing a *helpdesk* or even a *forum* platform to promote the engagement of the users with the ENVRI community.