ENVRI-Hub NEXT

D11.2 Metadata and Vocabularies Harmonisation

Status: Under EC Review

Dissemination Level: Public

Funded by the European Union

Abstract

Keywords	Metadata Interoperability, I-ADOPT compliant Vocabularies, Catalogue of Services, Fair Data Points, Essential Variables, EOSC Integration

This deliverable, produced within the ENVRI-Hub NEXT project, addresses the harmonisation of metadata schemas and vocabularies across European Environmental Research Infrastructures (RIs). Its primary objective is to enhance the interoperability, discoverability, and reusability of environmental data by promoting the adoption of standardised metadata practices and semantic frameworks.

The document presents a landscape analysis involving 14 RIs, revealing a moderate convergence toward widely accepted metadata standards such as ISO 19115 and DCAT-AP, and metadata services including CSW, OAI-PMH, and SPARQL. Vocabulary usage varies significantly, with some RIs employing domain-specific controlled vocabularies and others relying on free-text or partially curated lists.

The deliverable highlights the role of the ENVRI-Hub Catalogue of Services in enabling federated, standards-compliant metadata publication in DCAT metadata format. It also introduces the I-ADOPT framework as a concrete approach for achieving semantic interoperability through structured descriptions of observed variables.

Key recommendations include adopting multiple metadata schemas and services, incrementally aligning vocabularies with I-ADOPT, and ensuring metadata readiness for integration with the European Open Science Cloud (EOSC). The document concludes that while foundational steps have been taken, further harmonisation is needed to support scalable, cross-domain environmental research and data reuse.

Revision History			
Version	**Date**	**Description**	**Author/Reviewer**
V 0.1	31/03/2025	ToC and first draft	Claudio Dema (CNR) Markus Fiebig (NILU)
V 0.2	12/05/2025	Available for External Review	Claudio Dema (CNR) Markus Fiebig (NILU) Shridhar Jawak (NILU) Alex Vermeulen (ICOS ERIC) Alessandro Turco (EPOS ERIC) Marta Gutierrez (EGI) Peter Thijsse (MARIS) Jan Bumberger (UFZ) Giuseppe D'Amico (CNR) Ermann Ripepi (CNR) Francesco Izzi (CNR) Giuseppe La Scaleia (CNR)
V 0.3	21/05/2025	Document Review	Valerio Vinciarelli (EPOS ERIC) Tjerk Krijger (MARIS)
V 1.0	**28/05/2025**	**Final**	Claudio Dema (CNR) Alessandro Turco (EPOS ERIC) Marta Gutierrez (EGI) Peter Thijsse (MARIS)

Document Description			
D11.2 Metadata and Vocabularies Harmonisation			
Work Package Number 11			
Document Type	Deliverable		
Document Status	Under EC Review	**Version**	1.0
Dissemination Level	Public		
Copyright Status	CC BY This material by Parties of the ENVRI-Hub NEXT Consortium is licensed under a Creative Commons Attribution 4.0 International License.		
Lead partner	CNR		
Document Link	https://documents.egi.eu/document/4052		
DOI	https://zenodo.org/uploads/15555563		
Author(s)	• Claudio Dema (CNR) • Markus Fiebig (NILU) • Shridhar Jawak (NILU) • Alex Vermeulen (ICOS ERIC) • Alessandro Turco (EPOS ERIC) • Marta Gutierrez (EGI) • Peter Thijsse (MARIS) • Jan Bumberger (UFZ) • Giuseppe D'Amico (CNR) • Ermann Ripepi (CNR) • Francesco Izzi (CNR) • Giuseppe La Scaleia (CNR)		
Reviewers	• Valerio Vinciarelli (EPOS ERIC) • Tjerk Krijger (MARIS)		
Moderated by:	• Matteo Agati (EGI)		
Approved by:	Development Steering Board (DSB)		

Terminology / Acronyms	
Term/Acronym	**Definition**
RI	Research Infrastructure
ENVRIs	European Environmental Research Infrastructures
CoS	ENVRI-Hub Catalogue of Services
FDP	Fair Data Point instance
EXV	Essential Variable
ECV	Essential Climate Variable
EOV	Essential Ocean Variable
EBV	Essential Biodiversity Variable
EGV	Essential Geodetic Variable

Useful Reference: **ENVRI Glossary**

Table of Contents

Table of Figures

Table of Tables

Executive Summary

This deliverable, "*Metadata and Vocabularies Harmonisation*", produced within the framework of the ENVRI-Hub NEXT project, outlines the current landscape, challenges, and strategic recommendations for harmonising metadata and vocabularies across European Environmental Research Infrastructures (ENVRIs). Its primary objective is to enhance interoperability, discoverability, and reusability of environmental data by promoting the adoption of standardised metadata practices and semantic frameworks.

Purpose

The document aims to:

- Assess the current state of metadata and vocabulary usage across participating RIs.
- Recommend standard solutions for **metadata schemas and services**;
- Promote the adoption of **well-structured vocabularies** for observed variables aligned with the **I-ADOPT framework** to support semantic interoperability;
- Facilitate integration with the **ENVRI-Hub Catalogue of Services (CoS)** for a smooth RDF DCAT metadata publication.

Key Findings

- A landscape analysis involving 14 RIs revealed a moderate convergence towards standard metadata schemas (e.g., ISO 19115: 78%, DCAT-AP: 36%) and services (e.g., SPARQL: 35%, CSW: 28%, OAI-PMH: 21%);
- Vocabulary usage varies significantly: some RIs use domain-specific controlled vocabularies, while others rely on free-text or partially curated lists;
- Adoption of the I-ADOPT framework is still limited but growing, with early adopters like ACTRIS and SeaDataNet demonstrating its potential for cross-domain data discovery.

ENVRI-Hub NEXT Contributions

- The **Catalogue of Services (CoS)** provides a federated, standards-compliant infrastructure for metadata publication and discovery of services;
- A **demo Fair Data Point (FDP) instance** was developed to showcase integration of Essential Variables (EXVs) using I-ADOPT, enabling semantic search and machine-actionable metadata of datasets;
- The **EXV Task Force** initiated pilot implementations using Python notebooks to demonstrate EXV-related observation data access and harmonisation workflows across Research Infrastructures.

Recommendations

1. **Metadata Schemas**: Adopt and expose metadata in multiple formats, prioritising ISO 19115 family and DCAT-AP for compatibility and integration;

2. **Metadata Services**: Use standard services (e.g., CSW, OAI-PMH, SPARQL, CKAN API) to ensure discoverability and machine-to-machine interoperability;
3. **Semantic Interoperability**: Incrementally adopt the I-ADOPT framework for observed parameters, starting with high-priority variables, and align local vocabularies through mappings and shared registries;
4. **Integration with EOSC**: Ensure metadata and services meet the requirements for onboarding into the European Open Science Cloud (EOSC).

Final Remarks

The harmonisation of metadata and vocabularies used to describe datasets, data products, and services is essential for enabling scalable, cross-domain environmental research. While progress has been made, particularly in developing metadata minimum requirements and standardisation, and in promoting collaborative efforts among Research Infrastructures, further efforts are needed to standardise vocabulary usage, streamline data access procedures and harmonise data and metadata contents. The deliverable sets the foundation for continued harmonisation in the next phase of the project, supporting the broader goals of FAIR data principles and open science and advancing to the final goal of enabling data federation.

1 Introduction

1.1 Goals

The overarching goal of ENVRI-Hub NEXT is to create a robust conceptual and technical framework that will empower the ENVRI Science Cluster to provide interdisciplinary science-based services. Through this framework, ENVRI-Hub NEXT is facilitating the integration of the environmental sciences community into the European Open Science Cloud, guided by the science-based framework of Essential Variables (EXVs).

To transform the challenging task of integrated Earth observation into a concept towards a global climate observation system, the World Meteorological Organisation (WMO) has specified a set of Essential Variables (including Essential Climate Variables [1], Essential Ocean Variables [2], Essential Biodiversity Variables [3], and Essential Geodetic Variables [4]) relevant for the continuous monitoring of the state of the climate and the environment.

It is, therefore, a major priority for the Research Infrastructures (RIs) to observe or derive these variables and make this data available to the scientific community.

Discovering data with required parameters from distributed data sources, which often use diverse metadata schemas or vocabularies, is very difficult. It requires enormous human effort to go through different systems and infrastructures, check the existence of specific observations, select the relevant data sets for computing a specific EXV and aggregate the data from the various sources. So part of the challenge lies in the services, the other part in the metadata and data.

At the service level:

Accessing data from multi-sources if they do not share a common interface and protocol and access controls, which hampers the automation of the data processing workflow via direct machine-to-machine interactions, is challenging. Data is often behind service interfaces, and their diverse application programming interface (API) and locations require customised integration of that data in the workflow applications, which makes the reusability of the workflow for new data services very low, if without heavy code-level adaptation.

At (meta)data level:

Traceability, transparency, and information on the data quality through the whole value chain are essential for using this data in relevant societal evaluations and decision support systems. Also, in the provenance metadata of all datasets derived from EXV-related observation data, proper attribution to the data providers is needed to support the impact analysis for the RIs and their observational efforts towards their funders and organisations, supporting their sustainability. In the metadata and datafiles it is important to be able to find the observed parameter and units, with clear semantics, in order to be able to aggregate/process the data into EXV datasets.

The ENVRI-Hub is designed to be the single access platform to interdisciplinary services that enable cross-RI exploitation of data, guided by the science-based framework of EXVs. It will provide a centralised metadata catalogue that harmonises the metadata of services and datasets from individual RIs, extending interoperability and making datasets and their metadata searchable/accessible for EXVs. The ENVRI catalogue will map the metadata from individual services and data catalogues provided by different RIs onto a centrally agreed high-level rich metadata and vocabulary (compliant with the I-ADOPT framework), which will facilitate the search much more efficiently and precisely across multi-sources using metadata facets.

1.2 Metadata Standardisation

In the realm of environmental research, the significance of metadata cannot be overstated and is a key element for achieving FAIR data and services. Metadata, often described as "data about data"(but could also describe other objects like services, organisations, etc), provides essential context and descriptive information that enhances the usability, accessibility, and longevity of research data. It encompasses details about the content, format, source, accuracy, quality, and provenance of the data, answering critical questions such as who collected the data, what the data represents, why it was collected, where and when it was gathered, and how it can be accessed and utilized [5].

The utility of metadata extends beyond mere documentation; it plays a pivotal role in data stewardship, governance, and integration. This is particularly crucial in environmental research, where understanding complex systems often requires synthesizing data from multiple sources and scales.

Metadata is an indispensable tool in environmental research, fostering efficient data management, promoting collaborative efforts, and advancing our understanding of environmental phenomena. Its importance lies not only in the immediate utility it provides but also in its capacity to extend the life and impact of research data, ultimately contributing to the broader goals of environmental science and sustainability.

In this context, the adoption of standard metadata practices is crucial for ensuring the quality, accessibility, and longevity of data. By adhering to standardised metadata practices, researchers can ensure that data is consistently described, easily discoverable, and readily usable by others.

This consistency is vital for enabling interoperability, which allows data from different sources and disciplines to be integrated and compared. For example, when studying climate change, researchers might need to combine data on temperature, precipitation, and biodiversity from various studies. Using standard data formats, vocabularies, and units, together with standardized metadata, ensures that these datasets can be seamlessly merged, facilitating comprehensive analyses and more robust conclusions.

Moreover, standard metadata practices enhance data discoverability. Well-documented metadata makes it easier for researchers to find relevant datasets, reducing the time and effort required to locate and understand data. This is particularly important in environmental research, where data is often collected over long periods and across vast geographic areas. By making data more accessible, standardized metadata practices support collaboration and data sharing, fostering a more connected and efficient research community.

In addition to improving discoverability, standard metadata practices also contribute to data preservation. Properly documented metadata ensures that data remains understandable and usable over time, even as technologies and research priorities evolve. This long-term usability is essential for tracking environmental changes and trends, which often require decades of data.

In summary, standard metadata practices are a cornerstone of effective environmental research. They enable interoperability, enhance discoverability, and ensure the long-term usability of data. By adopting these practices, researchers can maximize the value of their data, support collaborative efforts, and contribute to a deeper understanding of our environment.

Standard metadata practices involve using:

- Common schema and formats;
- Standard metadata services for metadata access;

- Controlled vocabularies to describe data.

1.2.1 Use of Standard Metadata Schema

A standard metadata schema establishes a common way of structuring and understanding data and includes principles and implementation issues for using the standard. It defines how metadata should be organized and described. The standard contains specifications for the minimum information that should be collected about research data for it to be reused. It includes predefined categories and formats that ensure consistency across datasets. This schema provides a common language for documenting data [6].

Many research fields have worked together to define what kind of metadata should be collected to describe certain types of data. Standards describe the information that needs to be collected and the format in which it is expected.

Metadata has specific standards, and these will depend on the type of data collected, the intended audience, and specific requirements of the field itself.

Key components of a standard metadata schema are:

1. **Metadata Elements**: These are the specific pieces of information that describe the data. Common elements include the title, creator, date, geographic location, and methods used to collect the data. Each element provides essential context that helps researchers understand the data's origin and purpose.
2. **Controlled Vocabularies**: These are standardized terms and definitions used within the metadata schema. Controlled vocabularies ensure that everyone uses the same terms to describe similar concepts, which enhances clarity and reduces ambiguity.
3. **Data Content Standards**: These guidelines specify how to input values into metadata elements. For example, they might dictate the format for dates or geographic coordinates. Consistent data content standards ensure that metadata is uniform and easily interpretable.
4. **Data Exchange Standards**: These specifications define how metadata, and in some cases also the described dataset, should be encoded for sharing and storage. Common formats include XML and JSON, which are widely used for their readability and compatibility with various systems.

1.2.2 Use of Standard Metadata Services

A standard metadata service is a system or platform that provides tools and protocols for creating, managing, and accessing metadata. These services facilitate the documentation of data, ensuring that it is well-described and easily discoverable. Metadata services often include features such as metadata templates, controlled vocabularies, and automated validation tools to help researchers maintain high standards of data quality and consistency.

Key functions of standard metadata services include:

1. **Creation and Management**: Metadata services provide standardized templates and guidelines for documenting data. This ensures that metadata is consistent and comprehensive, covering essential aspects such as the origin, context, and structure of the data;
2. **Validation and Quality Control**: These services often include tools for validating metadata, checking for completeness, accuracy, and adherence to standards. This helps maintain the integrity of the metadata and ensures that it meets the required quality standards;

3. **Access and Discoverability**: Metadata services enhance the discoverability of data by making metadata searchable and accessible through various platforms. This allows researchers to easily find and utilize relevant datasets, fostering collaboration and data sharing.

Standard metadata services are particularly important in environmental research due to the complexity and diversity of the data involved. Some key benefits are:

1. **Interoperability**: Standard metadata services ensure that data from different sources can be integrated and compared seamlessly. Standardized metadata facilitates this integration, enabling comprehensive analyses;
2. **Efficiency and Collaboration**: By providing standardized templates and tools, metadata services reduce the time and effort required to document and manage data. This efficiency supports collaboration among researchers, allowing them to share and reuse data more effectively. Enhanced discoverability through metadata services also means that researchers can quickly locate relevant datasets, further promoting collaborative efforts;
3. **Long-term Usability and Preservation**: Properly documented metadata ensures that data remains understandable and usable over time, even as technologies and research priorities evolve, ensuring that valuable research findings remain available for future use.

1.2.3 Well-structured Vocabularies

When defining a vocabulary for describing data, it is important to specify the target audience. While it is possible to define vocabulary for a core group of experts on a given data type, it is good practice to design the vocabulary understandable to a wide range of the educated public. For enabling use across scientific domains, this type of vocabulary design is of paramount importance.

In understanding a vocabulary, it is useful to contrast it with a glossary. A glossary establishes a common understanding of concepts by defining them, but adding a few restrictions. While not good practice, it is possible in a glossary to assign shifting definitions to parts of concepts depending on the context, and to use shifting grammar. A well-structured vocabulary puts much stricter constraints on the structure of concepts, their definitions, and the relation between them:

- Establishes a common understanding of how to call concepts (same as glossary);
- Internally consistent definitions, i.e., definitions of concepts or parts of concepts, are constant and independent of context;
- Concepts follow a uniform grammar;
- Concepts have a persistent identifier (PID);
- Relations and hierarchy between concepts are defined;
- Is machine-readable;
- Reuses concepts established in existing, domain-relevant vocabularies;
- Links to existing vocabularies;
- Has a well-defined governance and update procedure, and is otherwise stable.

In describing data, vocabulary for the observed variables is of highest importance. However, other aspects where vocabulary is needed include data sources (e.g., instruments, models, etc.), data provision and use, manufacturers, quality assurance, quality control, geometry, timeliness, and licences.

1.2.3.1 Use of well-structured Vocabularies

Within ENVRI-Hub NEXT, well-structured vocabularies support:

- Semantic clarity: By standardising terms, RIs avoid the pitfalls of inconsistent naming conventions, ensuring that "temperature," "temp," or "T" all map to a single, well-defined concept;
- Machine-actionable metadata: Vocabularies with stable URIs and robust documentation allow automated systems, such as catalogue aggregators or semantic search engines, to parse and interpret metadata uniformly;
- FAIR data compliance: Aligning with recognized controlled vocabularies helps meet the *Interoperable* aspect of the FAIR (Findable, Accessible, Interoperable, Reusable) principles, facilitating data integration across diverse domains such as marine, atmospheric, terrestrial, and solid Earth research;
- Future-proofing: As new observation methods or emerging phenomena arise, RIs can extend existing vocabularies or create compatible concepts, preserving semantic consistency and minimizing disruption to existing data systems;
- Being understandable across domain boundaries;
- Being understandable for artificial intelligence as a key emerging technology.

1.2.3.2 The I-ADOPT Framework

One promising route to achieving greater consistency in variable definitions is the I-ADOPT (InteroperAble Descriptions of Observable Property Terminology) framework. I-ADOPT provides a community-driven grammar for constructing structured descriptions of "what" is observed or measured, "how," and "under which context." This approach standardises the way properties (e.g., "air temperature at 2 m above ground") are assembled from smaller conceptual building blocks (observable, height, medium, etc.), so that even domain-specific terms can be compared and interlinked more effectively.

- **Improving cross-domain searches:** When multiple RIs adopt I-ADOPT-aligned descriptions of their variables, it becomes simpler to identify which datasets measure the same (or closely related) phenomena. This is especially advantageous for queries spanning multiple domains (e.g., marine and atmospheric) or focusing on specific EXVs (e.g., Essential Climate Variables);
- **Building a common conceptual foundation:** I-ADOPT's grammar helps break down observations into smaller semantic components (e.g., property, measured medium, context). Aligning these components across RIs not only reduces terminology conflicts but also simplifies the creation of bridging or mapping solutions;
- **Facilitating incremental adoption:** RIs can begin with a subset of high-priority variables and gradually expand I-ADOPT coverage. This incremental approach mitigates the learning curve and allows technical staff to refine processes and tools over time;
- **Path to semantic interoperability:** By embracing I-ADOPT, ENVRI-Hub NEXT stakeholders create fertile ground for advanced analytics and integrated data products that rely on precise, unambiguous definitions of environmental measurements;
- Overall, adopting well-structured vocabularies - including partial or full alignment with the I-ADOPT framework - signals a significant step toward **data interoperability within**

ENVRI-Hub NEXT. It paves the way for more efficient cross-domain discovery, deeper data integration, and the capacity to address complex environmental questions spanning multiple scientific communities.

2 The ENVRI Landscape

2.1 Metadata Standardisation in the ENVRI Landscape

The use of standard metadata practices plays a key role in the achievement of the ENVRI-Hub NEXT project objectives.

To provide the basis and relevance to the recommendations for standard metadata practices, the ENVRI-Hub NEXT Work Package 11 (WP11) team decided to collect results from a landscape analysis on metadata schema and services across Environmental Research Infrastructures (ENVRI RIs), both inside and outside the ENVRI-Hub NEXT consortium. The analysis, whose results are summarised below, also helped to have a clear picture of the current status and to estimate and organise the further metadata harmonisation work.

A total of 14 RIs (including 9 in the ENVRI-Hub NEXT consortium) responded to the landscape analysis survey (see Table 1).

Table 1 - Environmental Research Infrastructures involved in the landscape analysis.

Environmental Research Infrastructures involved in the Landscape Analysis	
RI Name	**ENVRI-Hub NEXT Consortium**
ACTRIS ERIC	Yes
AnaEE	Yes
eLTER	Yes
EPOS ERIC	Yes
Euro-Argo ERIC	Yes
IAGOS	Yes
ICOS	Yes
LifeWatch ERIC	Yes
SeaDataNet	Yes
DANUBIUS RI	No
DiSSCo	No
EISCAT	No
EMSO ERIC	No
SIOS	No

First of all, the analysis reveals that all the RIs enrich their data with metadata.

There are three categories of metadata [7]:

- **Descriptive metadata:** Describes a resource for purposes such as discovery, identification, and reuse. E.g. title, subject, date(s), location, keywords, PIDs, creator(s), publication information, provenance and quality processes;
- **Administrative metadata:** Gives information that helps manage the assets; it will inform users of any instructions, rules, or restrictions placed on the assets. E.g., licences and intended usage, authentication and authorisation needs, long-term curation;
- **Structural metadata:** Gives information about a certain object or resource itself and indicates how it may be sorted. E.g., file format, file size, and file dimensions.

While all the RIs provide exhaustive descriptive metadata, about a third of them acknowledge that the provision of administrative and structural metadata could be made more complete, as shown in Figure 1.

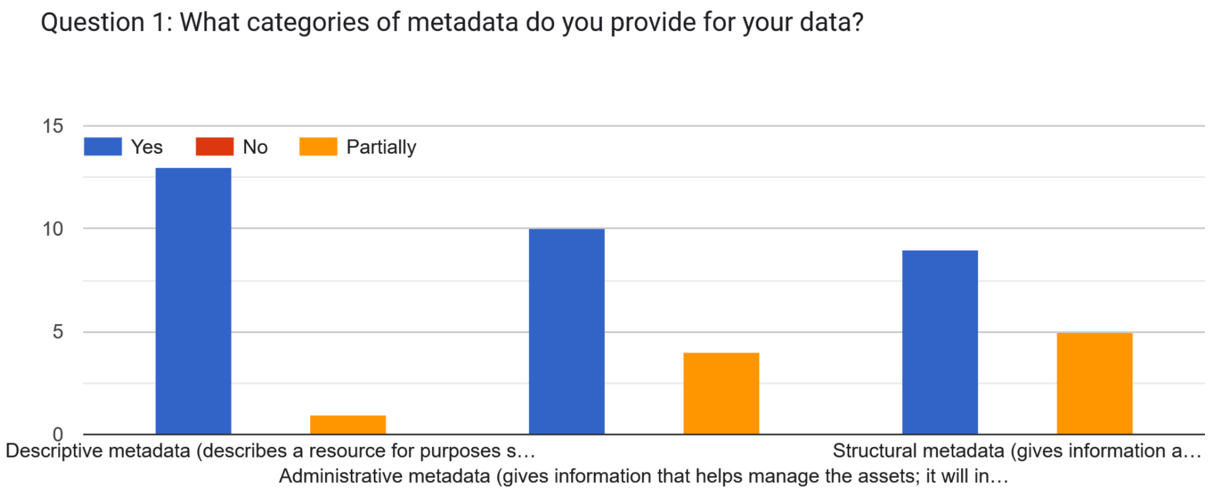

Figure 1 - Categories of Metadata provided by the ENVRI RIs.

2.1.1 Metadata Schema

The analysis of RI metadata schema usage shows several noteworthy aspects.

A first aspect that should be mentioned is the good practice, used by half of the infrastructures, of exposing their metadata in multiple formats and according to multiple metadata schemas. This is very important because it promotes cross-RI interoperability and integration.

It is also possible to observe a moderate convergence towards the use of some of the most commonly used metadata standards in the field of environmental research:

- 78% expose their metadata using schemes that are implementations or extensions of the ISO 19115 standard;
- 36% of the infrastructures, on the other hand, use the DCAT schema or an extension of it.

In most cases, the extension of a schema is due to the need to expose metadata considered relevant in the specific context but not provided for by the standard. Even when the schema used is the same, the metadata content can vary. In addition to that, several infrastructures use a different and domain-specific metadata schema.

A brief description of the most common metadata schema standards follows.

2.1.1.1 The ISO 19115 Standard and its Implementations

ISO 19115 is an international standard for describing geographic information and services by means of metadata. It provides a comprehensive schema for documenting the identification, extent, quality, spatial and temporal aspects, content, spatial reference, portrayal, distribution, and other properties of digital geographic data and services. This standard is crucial for ensuring that geographic data can be easily discovered, accessed, and used effectively.

ISO 19115-3 is at the moment the last implementation specification that defines the XML schema for the fundamental concepts outlined in ISO 19115-1. It integrates the metadata elements from ISO 19115-1 and ISO 19115-2, providing a structured XML format for encoding metadata. This implementation ensures that metadata can be validated and exchanged seamlessly across different systems, promoting interoperability and data integration.

ISO 19139 is another implementation specification that provides an XML schema for encoding the metadata defined in ISO 19115. It focuses on the practical aspects of implementing the metadata standard in XML format, allowing for efficient data exchange and validation. ISO 19139 has been widely adopted in various geographic information systems (GIS) to ensure consistent and accurate metadata documentation.

These implementations are essential for facilitating the use and sharing of geographic data in environmental research, as they ensure that metadata is standardized, interoperable, and easily accessible.

Standards like ISO 19115 and its implementations are regularly updated to incorporate new technologies and methodologies. For example, at the time of writing this document, ISO 19115-4 and ISO 19115-5 are under development. The former will provide a JSON schema implementation of metadata fundamentals [8], while the latter will implement a DCAT mapping [9].

This ensures that metadata practices remain relevant and effective over time.

2.1.1.2 The DCAT Standard

DCAT (Data Catalog Vocabulary) is an RDF (Resource Description Framework) vocabulary designed to facilitate interoperability between data catalogs published on the web. It provides a standard model and vocabulary for describing datasets and (data)services in a catalog, which helps in the consumption and aggregation of metadata from multiple catalogs.

DCAT-AP (Data Catalog Vocabulary Application Profile) is a specification based on the DCAT standard, designed specifically for describing public sector datasets in Europe. It aims to enhance the interoperability and discoverability of datasets across European data portals by providing a common framework for metadata documentation.

DCAT-AP includes detailed implementation guidelines that describe the mandatory, recommended, and optional metadata elements for cataloging datasets. These guidelines help data publishers ensure that their metadata is consistent, comprehensive, and compliant with the DCAT-AP specification.

DCAT-AP is widely used by European public sector data portals to enhance the discoverability and interoperability of datasets. It supports various scenarios, such as enabling researchers to find datasets from different countries and facilitating the reuse of public sector information.

This is an essential tool for promoting the effective use and sharing of public sector data in environmental research and other domains.

The EPOS-DCAT-AP is an extension of the DCAT Application Profile (DCAT-AP) specifically designed for the European Plate Observing System (EPOS). This schema aims to standardize the metadata for datasets related to solid Earth sciences, facilitating interoperability and data sharing across various research infrastructures. It builds upon the DCAT-AP, which is a profile of the W3C Data Catalog Vocabulary (DCAT). It ensures compatibility with existing standards for data catalogues. The schema incorporates additional vocabularies such as Schema.org and Hydra to capture the diverse and heterogeneous nature of EPOS assets, and it is designed to support cross-disciplinary collaboration by representing core concepts in solid Earth sciences.

By adhering to the DCAT-AP standards, EPOS-DCAT-AP ensures that metadata can be exchanged seamlessly between different systems and platforms within the European Union.

This extension is among the main candidates to be used as a metadata schema for exposing the metadata of the ENVRI-Hub externally.

2.1.2 Metadata Services

Even in the case of metadata services, it is noteworthy that the common practice among different RIs is to expose their metadata through multiple services. 78% of these services are public and accessible.

The analysis shows a moderate level of convergence between the RIs, even in terms of the services used for publishing metadata:

- SPARQL endpoints (35%);
- OGC CSW (28%);
- OAI-PMH (21%);
- CKAN API (21%).

In addition to that, each RI has a different approach on how to make the datasets discoverable and downloadable.

The output of the services is different. Some services need first a metadata search, and then a second service to retrieve the data files, while some RIs also offer direct data access (to datasets and/or data products) and subsetting services.

A brief description of the most common metadata service standards follows.

2.1.2.1 SPARQL Endpoints

A SPARQL endpoint is a web service that allows users to query RDF (Resource Description Framework) data using the SPARQL query language.

SPARQL is an acronym for SPARQL Protocol and RDF Query Language [10]. It is a language used to query and manipulate semantic data, which is a way of organizing information in a way that machines can understand. Semantic data is based on linked data principles, which means that data is interconnected and linked together in a way that promotes reusability, extensibility, and interoperability. The idea behind linked data is that data should be structured in a way that makes it easy to connect and share. By using open standards and technologies, semantic data is designed to work smoothly and easily together, no matter where it is located.

This service is particularly suitable for managing and querying large datasets.

SPARQL endpoints enable researchers to query RDF data, which is structured as triples (subject-predicate-object). This is ideal for representing complex relationships between environmental entities.

SPARQL endpoints are accessible via a web service interface, allowing users to send queries over the internet. This interface typically supports HTTP GET and POST methods. SPARQL endpoints process queries by parsing the SPARQL syntax, executing the query against the RDF data, and returning the results in various formats such as XML, JSON, or CSV.

The ability to perform complex queries allows researchers to uncover patterns and relationships in environmental data that might not be evident through simpler querying methods.

2.1.2.2 OGC CSW

The OGC Catalogue Service for the Web (CSW) [11] standard is designed to support the publication and search of collections of metadata for geospatial data, services, and related resources.

Metadata records are typically encoded in XML and include core fields such as title, format, type, bounding box, coordinate reference system, and associations

CSW enables efficient discovery of geospatial data and services by providing standardized interfaces for querying metadata. Key operations defined by the CSW standard include:

- **GetCapabilities:** Retrieves service metadata from a server;
- **DescribeRecord:** Allows a client to discover elements of the information model supported by the target catalogue service;
- **GetRecords:** Searches for records, returning record IDs;
- **GetRecordById:** Retrieves the default representation of catalogue records using their identifier.

This helps users find relevant resources quickly.

Moreover, CSW helps in organizing and managing large collections of geospatial data by providing tools for publishing, searching, and retrieving metadata.

2.1.2.3 OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [12] is a protocol designed to facilitate the harvesting of metadata descriptions from various repositories. The protocol uses six verbs (services) that are invoked within HTTP:

- Identify;
- ListMetadataFormats;
- ListSets;
- ListIdentifiers;
- ListRecords;
- GetRecord.

Metadata is exchanged using XML over HTTP, ensuring a standardized format for data transfer.

While OAI-PMH must support Dublin Core, it can also accommodate additional metadata formats.

By exposing metadata through OAI-PMH, repositories can increase the visibility and accessibility of their resources, which can be harvested by search engines and discovery services.

2.1.2.4 CKAN API

CKAN (Comprehensive Knowledge Archive Network) [13] is an open-source data management system designed to make data accessible and usable. It provides tools for publishing, sharing, finding, and using data. The platform is highly customizable, allowing users to tailor it to match their specific research needs and branding.

CKAN provides a comprehensive set of metadata for each dataset, including title, unique identifier, description, tags, and more. It supports displaying a revision history for datasets, which is useful for tracking changes and updates. A rich search experience with keyword search and filtering, making it easier to find relevant datasets, is offered.

Advanced geospatial capabilities allow for data preview, search, and discovery, which is particularly useful for environmental data.

The platform provides an API [14] that allows access to every metadata field of the dataset and enables modifications. The CKAN API also automates data exchange and integration with other tools. This can include linking with databases, data visualization tools, or other research platforms.

CKAN supports various metadata standards, including Dublin Core and custom schemas.

2.2 Use of I-ADOPT Compliant Vocabularies in the ENVRI Landscape

The adoption of I-ADOPT-compliant vocabularies in the standardised metadata as well as datafiles across the ENVRI landscape significantly enhances cross-domain discovery, enabling researchers to identify and compare, and aggregate similar types of observations in different disciplines. By providing a structured method for describing observable properties, I-ADOPT helps unify terminologies within and across Research Infrastructures (RIs). This section provides an overview of how well-structured vocabularies are currently used by participating RIs (Section 2.2.1) and the extent to which these vocabularies align with I-ADOPT principles (Section 2.2.2).

2.2.1 Well-structured Vocabularies

Results from the ENVRI-Hub NEXT WP11 landscape analysis indicate that while a few RIs (ACTRIS, SeaDataNet) already employ robust domain-specific controlled vocabularies (e.g., SeaDataNet Parameter Discovery Vocabulary "P01" in marine science), many still rely on free-text variable descriptions, inconsistent naming conventions, or partially curated keyword lists (See Fig.2, Fig.3, Fig.4, Fig.5 and Fig.6). Such heterogeneity presents barriers to cross-domain data searching and automated indexing; it also complicates aligning dataset descriptions with Essential Variables (EXVs) relevant to climate, marine, terrestrial, and atmospheric research represented by various RIs.

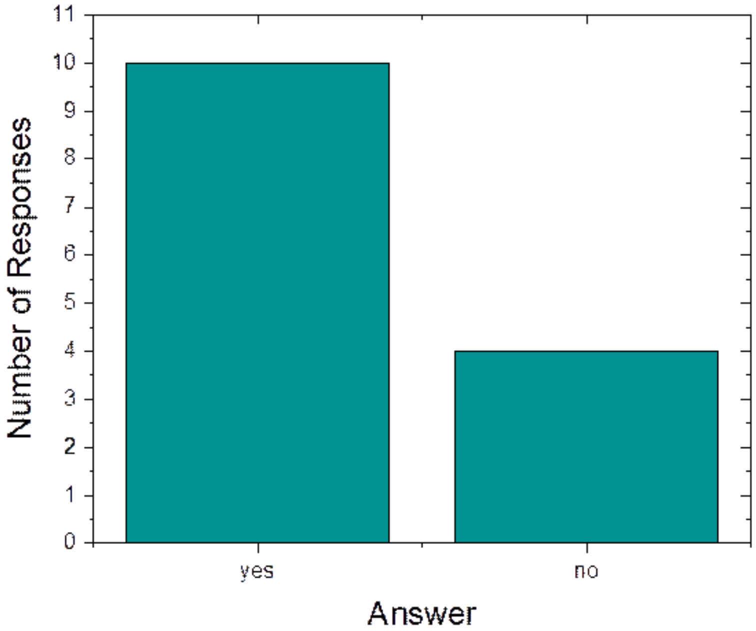

Figure 2 - Use of vocabulary by RIs within ENVRI-Hub NEXT.

- **Diverse current practices**: Some RIs integrate well-known vocabularies maintained by external organisations (e.g., CF Standard names in atmospheric science or SWEET in Earth and environmental science), whereas others have developed internal lists to meet their community's needs;
- **Challenges**: Vocabularies maintained in-house may lack thorough documentation, multilingual support, or robust versioning. Moreover, as revealed by the WP11 questionnaire, only a small subset of RIs systematically track updates or maintain cross-references between different controlled term sets;
- **Opportunity for harmonisation**: By mapping existing vocabularies to a common framework, RIs can retain domain-specific distinctions while enhancing interoperability. In the near term, incremental steps, such as adopting stable URIs for variables, can substantially improve machine-to-machine data discovery.

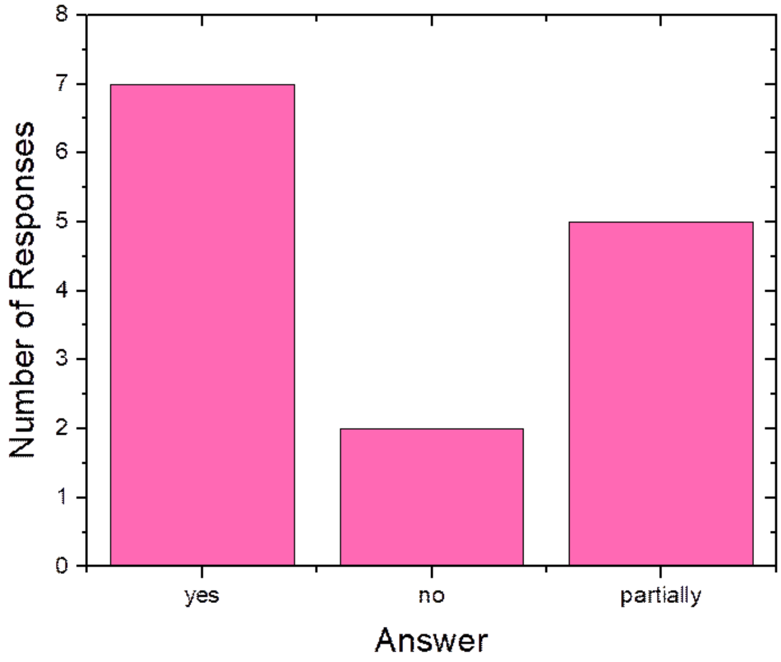

Figure 3 - Use of well-defined grammar for vocabulary usage by RIs within ENVRI-Hub NEXT.

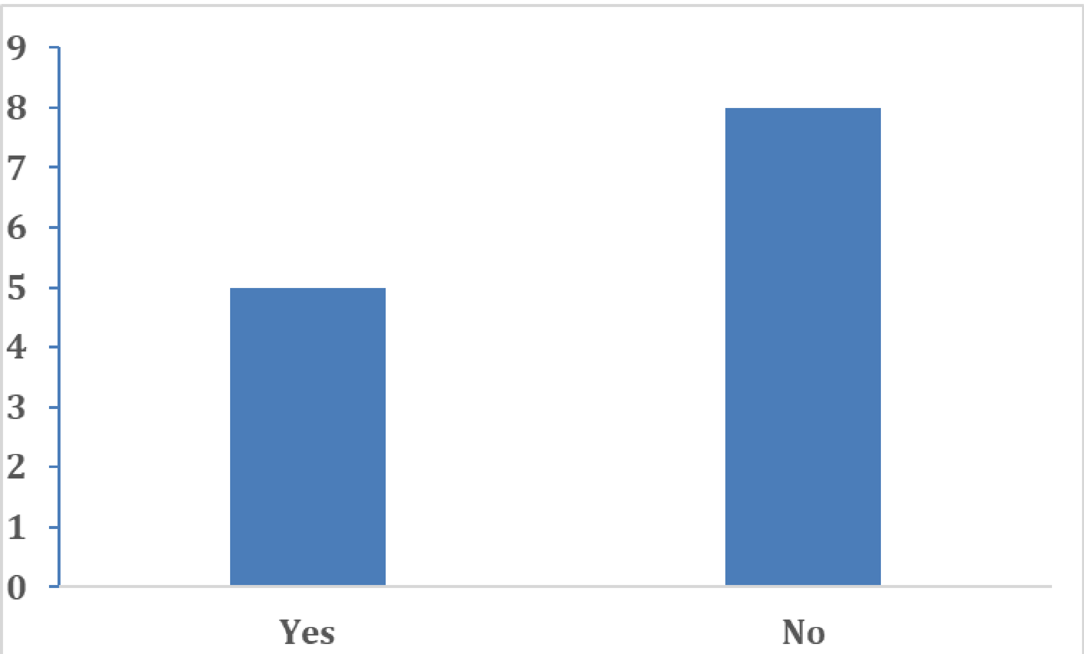

Figure 4 - Use different vocabulary for variables at data annotation/data use level, as opposed to discovery level by various RIs within the ENVRI-Hub NEXT.

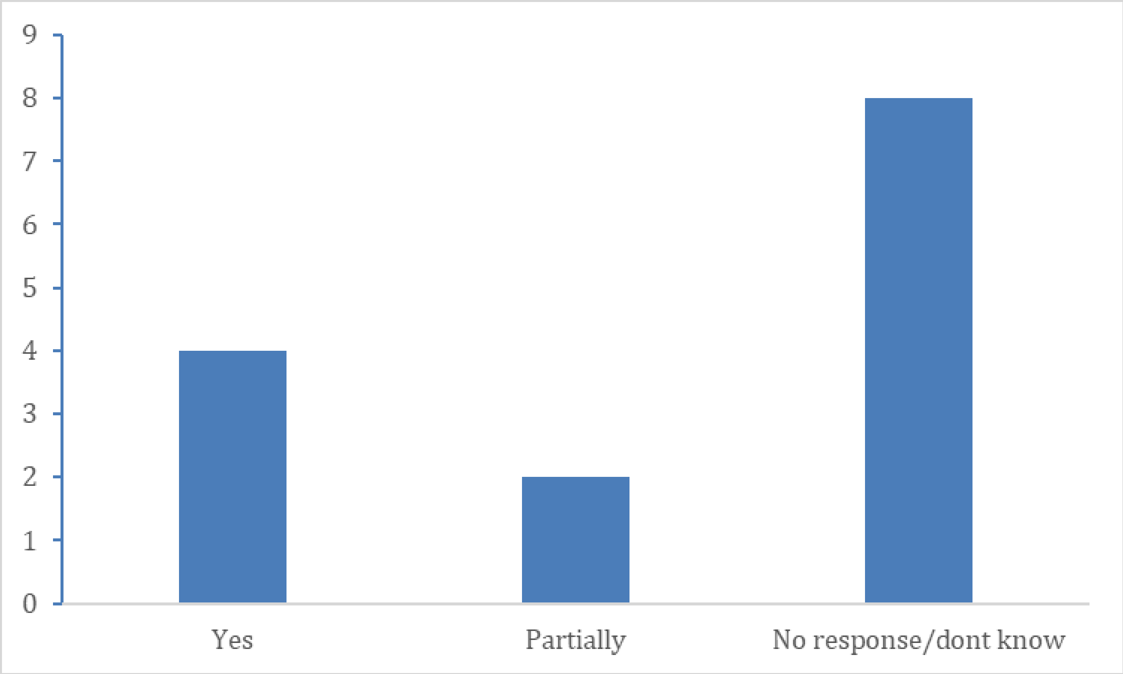

Figure 5 - Use of well-defined grammar in vocabulary usage for annotation variable by RIs within ENVRI-Hub NEXT.

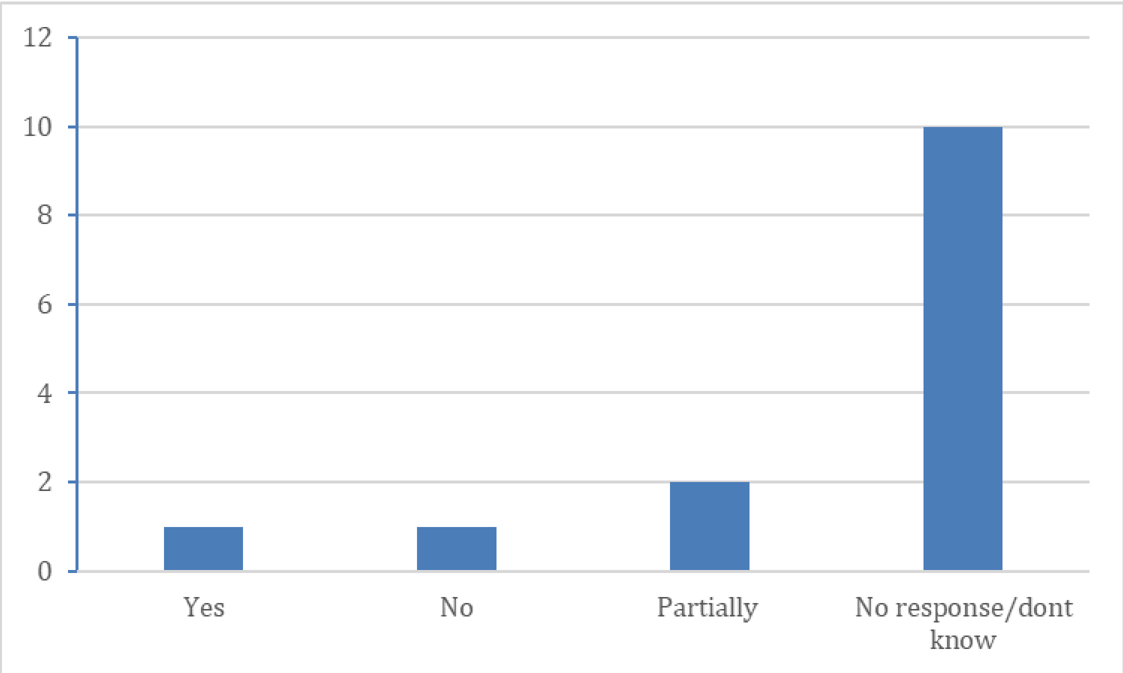

Figure 6 - Use of well-defined grammar in vocabulary usage for annotating variables and their adherence to the I-ADOPT framework.

Summary of survey responses on vocabularies

From the WP11 questionnaire (see Table 2) and follow-up discussions, it is evident that RIs vary in the degree to which they employ well-structured or standardised vocabularies:

- **Domain-specific adopters**: A few RIs have embraced domain-specific, well-established vocabularies. For instance, SeaDataNet utilises the P01 Parameter Discovery Vocabulary for marine parameters, while ACTRIS references community-driven lists for atmospheric measurements. These vocabularies typically feature stable URIs for each concept, good versioning practices, and cross-references to related terms;
- **Mixed or partial structures**: Several RIs indicated that they have partially curated lists or rely on a combination of free-text fields and controlled terminology. Often, the controlled terms cover only the most common parameters, leaving niche or newly introduced concepts less standardised;
- **In-house solutions**: In some cases, RIs have developed self-managed vocabularies tailored to their specific data products, but these solutions may lack external references or a broader community maintenance process. This limits discoverability and alignment with other domains;
- **Key barriers**: Survey respondents cited limited resources and a lack of established crosswalks between different vocabularies as primary challenges. They also noted difficulties deciding how frequently to update vocabularies and handle backward compatibility without disrupting existing data systems.

Despite these hurdles, a common trend emerging from the responses is the recognition that consistent variable naming and identification are prerequisites for any meaningful cross-domain search. Many RIs have expressed interest in adopting or mapping to more robust, widely used vocabularies if sufficient support and guidance are provided.

Table 2 - Summary of survey responses on vocabularies.

Question	Total RIs	Yes	No	Partial	No Response
Q3: Does the vocabulary follow a well-defined grammar?	14	3	5	2	4
Q4: Does the vocabulary adhere to the I-ADOPT framework?	14	2	7	1	4
Q6: Different vocabulary for data annotation vs. discovery?	14	4	6	–	4
Q6b: If yes, does the annotation vocabulary have a grammar?	4[1]	1	2	1	0
Q6c: If yes, does it adhere to I-ADOPT?	4[1]	1	3	0	0

[1]Questions 6b and 6c only apply to those RIs answering "*Yes*" to Q6 (i.e., that they use a different vocabulary for annotation).

2.2.2 I-ADOPT Compliance

The I-ADOPT framework proposes a structured grammar for describing observable properties, offering a systematic way to decompose complex measurement concepts (e.g., "concentration of CO_2 in air at 2 m above ground") into smaller semantic units. While this approach is still emerging, the survey indicates:

- **Pioneering RIs**: ACTRIS and SeaDataNet stand out as the early adopters, where core atmospheric variables are defined in a manner broadly compatible with I-ADOPT guidelines. This has already improved the precision and consistency of how their datasets label physical or chemical measurements;
- **Limited implementation**: Apart from these, only a handful of RIs have experimented with formal I-ADOPT elements or plan to do so. Some RIs noted that they had not yet mapped their existing vocabularies to the I-ADOPT grammar, though they see the benefit of doing so in the long term;
- **Common concerns**: Survey participants highlighted the learning curve for staff unfamiliar with semantic technologies, as well as the need for robust tooling or reference examples. The unavailability of domain-specific term mappings within I-ADOPT can also deter immediate uptake;
- **Path to broader adoption**: In discussions, RIs emphasised a willingness to consider I-ADOPT-compliant structures provided there is a clear roadmap, harmonisation support (e.g., from the EXV Task Force), and alignment with existing domain vocabularies. Many see the potential for improved cross-RI analytics and data discovery as a compelling driver for adoption.

Overall, the landscape analysis reveals that, while full I-ADOPT compliance remains limited to a small set of pilot RIs, there is growing awareness of its benefits. As more RIs recognise the advantages of consistent, semantically rich definitions of observational properties, I-ADOPT can serve as a powerful reference point for advancing interoperability across the ENVRI community.

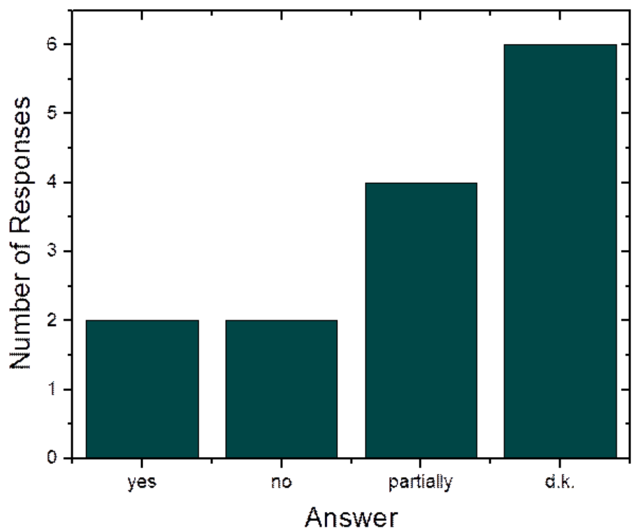

Figure 7 - Adherence of vocabulary usage to the I-ADOPT framework within the ENVRI-Hub NEXT project.

- **Pilot implementations**: The WP11 questionnaire shows that ACTRIS and SeaDataNet, among others, are exploring or already using grammars and standard term sets closely aligned with I-ADOPT principles. These early adopters provide concrete examples of how a more formal structure for variable definitions can improve discoverability and integration across data portals;
- **Recommended path forward**: ENVRI-Hub NEXT encourages RIs to pilot I-ADOPT on selected essential variables or high-priority parameters. Over time, adopting an I-ADOPT-based grammar/vocabulary for all newly introduced variables can substantially enhance cross-RI interoperability, enabling more powerful data exploration tools and facilitating advanced analytics in the ENVRI-Hub ecosystem.

By combining well-structured vocabularies with I-ADOPT's standardised grammar, the ENVRI community can build a more coherent, FAIR-aligned data landscape, ultimately accelerating scientific discovery across environmental domains.

2.3 Essential Variables in the ENVRI Landscape

As explained in paragraph 1.1, Essential Variables (EXVs) are the drivers of all the work the ENVRI-Hub NEXT WP11 team is doing in the metadata field.

Since the goal is to extend interoperability and make datasets and their metadata accessible through Essential Variables, it is crucial to understand which Essential Variables are served by various RIs through their data.

This is another aspect that the WP11 team focused their analysis on.

The analysis specifically targeted the Essential Climate Variables (ECVs) as described by the Global Climate Observing System (GCOS), grouped into the following categories and subcategories:

- Atmosphere
  - Surface
  - Upper-Air
  - Atmospheric Composition
- Land
  - Hydrosphere
  - Cryosphere
  - Biosphere
  - Anthroposphere
- Ocean
  - Physical
  - Biogeochemical
  - Biological/ecosystems.

Additionally, the RIs were asked which other variables they suggested should be made accessible through the ENVRI-Hub.

A significant aspect is that all 55 ECVs are served by at least one of the infrastructures that participated in the analysis.

Additionally, many infrastructures provide data that can be linked to other variables, such as:

- Essential Ocean Variables (EOVs),
- Essential Biodiversity Variables (EBVs),

- Essential Geodetic Variables (EGVs), still under definition.

Therefore, it is important that data related to these variables are also considered and made available/searchable through the ENVRI-Hub. This is why, following the analysis, it was decided to refer to Essential Variables, using a broader term that includes all of them.

3 The Role of ENVRI-Hub NEXT in Metadata and Vocabularies Harmonisation

The role of ENVRI-Hub NEXT in the field of metadata and vocabularies could be summarised as analysing the current state of the art of metadata and vocabulary standards, comparing this to the current state in the RIs and what is needed to cross-disciplinary research in EOSC content, and working towards improvement of the RIs metadata standardisation and data access services (taking into account the financial/organisational limits). ENVRI-Hub NEXT contributes to the upgrades, documents these, and demonstrates the results. In this chapter, the various technical components in scope for ENVRI-Hub NEXT will be presented, as well as the actions to demonstrate the status and updates at the RI as well as the Hub level.

3.1 The ENVRI-Hub: Data Access and Discovery

The ENVRI-Hub is being developed using a service-oriented architecture and microservice-based cloud-native design. A brief description of the ENVRI-Hub components for (meta)data access and discovery follows.

3.1.1 The Catalogue of Services

The ENVRI Catalogue of Services (CoS) represents a central access point for web services providing data from environmental RIs, with a standardised and interoperable interface. The CoS is designed as a microservice-based system with a three-tier architecture:

- A rich metadata catalogue;
- A gateway of APIs for machine-to-machine access;
- A graphical user interface.

The architecture and the underlying code base are derived from the EPOS Open Source project. This guarantees that the approach is valid and the quality and reliability of the software, since it is shared with the operational EPOS Data Platform for Earth Science.

The metadata catalogue stores metadata related to services, data, and products provided by the RI data providers. The metadata schema is based on a refactored version of CERIF, a powerful metadata system that supports formal syntax, semantics, and spatio-temporal tagging, and is fed by metadata described in an extension of the well-known DCAT-AP standard, namely EPOS-DCAT-AP. Recently, EPOS-DCAT-AP has been recognized as a success story by the Interoperable Europe Semantic Interoperability Community (SEMIC) for its contribution to the field.

The API gateway, on top of the catalogue system, will enable external agents to take advantage of the catalogue's advanced functionalities, e.g., for data access, using the most used and known standards, including REST APIs, OAI-PMH, DCAT-related standards, and SPARQL. This layer is fundamental for the development of the Analytical Framework of ENVRI-Hub Next, as highlighted in D5.1.

The top tier provides the graphical user interfaces for accessing data and services in the form of a map-based list of catalogued services and datasets, with semantic-driven search functionalities that enable users to search for appropriate datasets or services using controlled vocabularies of discipline-related keywords. Particular attention has been paid to EXVs with dedicated search fields and templates.

RIs provide their data through interfaces, e.g., RESTful web services, APIs, sub-setting services (for slicing through data collections), or SPARQL endpoints, to ensure machine-to-machine integration across different systems and platforms. Those interfaces often comply with standards in the OGC family (e.g., WFS, WMS), OpenSearch, and proprietary community-based standards (e.g., FDSN for seismology). Some services are, however, targeted from Jupyter notebooks, via Python packages or command line interfaces, and require a more detailed approach to describe. The CoS will FAIRify those services for accessing data with rich metadata based on their different maturity levels.

The number of services integrated into the CoS is monitored as a KPI of the EHN project, and it has steadily increased. Project partners met on different occasions, in person or remotely, for hackathon events where the metadata schema and the ingestion procedure were explained to sustain the KPI growth.

3.1.1.1 FAIR Data Points: an Approach for populating the Catalogue of Services

The **FAIR Data Point (FDP)** is a decentralized architecture designed to support the implementation of the FAIR principles (Findable, Accessible, Interoperable, Reusable) across distributed data repositories. Introduced in 2016 by the Dutch Techcentre for Life Sciences (DTL) and the GO FAIR initiative, the FDP provides a lightweight, standards-based interface for publishing and accessing FAIR metadata.

The **reference implementation**, developed by DTL and TNO, uses REST APIs and RDF-based metadata (e.g., DCAT v3, Dublin Core) to expose structured metadata in a FAIR-compliant way. It follows a hierarchical structure (Catalog → Dataset → Distribution), supports SPARQL queries, persistent identifiers, and Linked Data integration. It is open source and available on GitHub, with easy deployment via Docker.

FDP adopts a **federated architecture** (See Figure 8), allowing each research infrastructure (e.g., ENVRI) to host its own FDP instance. This preserves autonomy while enabling interoperability through shared APIs. Tools like Elasticsearch or GraphQL can federate queries, and metadata harvesters (e.g., OpenAIRE, GEOSS) aggregate metadata without centralizing data.

Key components of the FDP include:

- RESTful API exposing machine-readable metadata (RDF/JSON-LD);
- Metadata model based on DCAT and Dublin Core;
- Optional SPARQL endpoint for advanced querying;
- Web-based, customizable user interface;
- Authentication and authorization support;
- Metadata persistence using RDF triple stores (e.g., RDF4J, GraphDB).

For ENVRIs, FDP enables cross-domain collaboration, scalability, and compliance with EU initiatives like EOSC and INSPIRE. It addresses challenges such as metadata standardization, governance, and technical complexity.

In the ENVRI-Hub project, a customized FDP prototype was developed (Java 21, MongoDB 4.2, RDF4J) with full SPARQL support and a user-friendly UI.

For a detailed description of the FDP demo graphical user interface, see Appendix I.

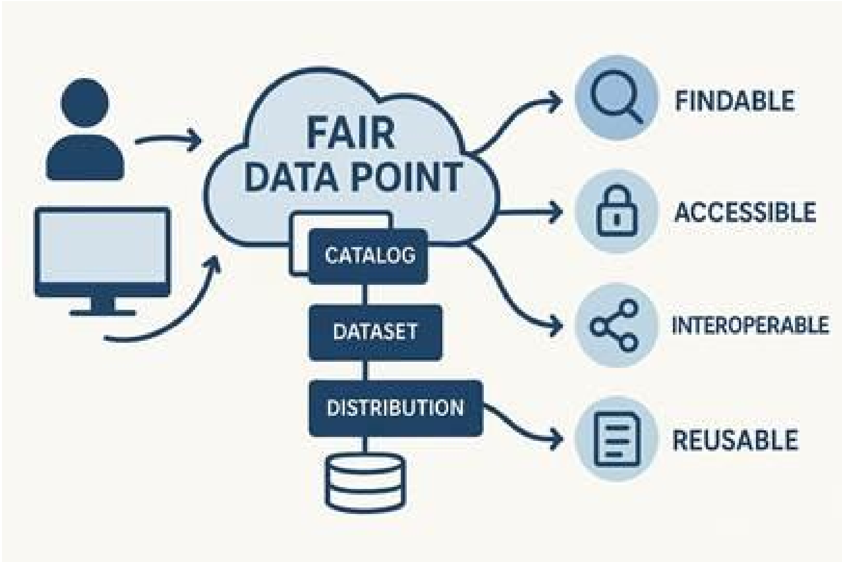

Figure 8 - FDP federated architecture design.

A public demo is available at fdpdemo.envri.eu. Participating infrastructures can define Essential Variables (via I-ADOPT), add datasets, and annotate them semantically.

The ENVRI-Hub FDP prototype demonstrates how a federated, standards-compliant architecture can enable scalable, semantically rich discovery and reuse of environmental data and services across the European research landscape.

The FDP can contribute to the Catalogue of Services (CoS) by exposing datasets through a service interface, such as the protocol used to access them (e.g., HTTP or FTP). In this way, FDP instances can populate the Catalogue of Services by representing datasets as services, acting as interfaces for data access.

Deliverable 7.5 "*Report on the FAIR Data Point implementation for ECVs*" will contain a complete description of the full FDP approach and will explain in detail how the FDP can feed the ENVRI CoS.

3.1.2 The Analytical Framework and the ENVRI-Hub API

The Analytical Framework aims to provide seamless support for users to compose executable analytics tasks or workflows using the data and services provided by the ENVRI-Hub.

Reusable abstract workflows for specific data products will be documented and demonstrated as Jupyter notebooks, which will be uploaded to a public Git repository as they are authored by ENVRI-Hub Next participants. Said notebooks are distributed under a liberal license, to allow their re-use for new purposes by providing different input data and configurations.

The Analytical Framework includes a Python library (which can be installed and run in any major VRE distribution) that wraps the Catalogue of Services interface, allowing VRE users to access data and metadata resources listed in the CoS in a Pythonic way. The library uses the CoS as an index of services, and whenever data items are not accessible directly through the CoS API, but through a third-party API, the package dynamically builds a dedicated client from its OpenAPI 3.0 specification.

The interactions between the ENVRI-Hub Python library, the ENVRI CoS, and actual data APIs are summarized in Figure 9.

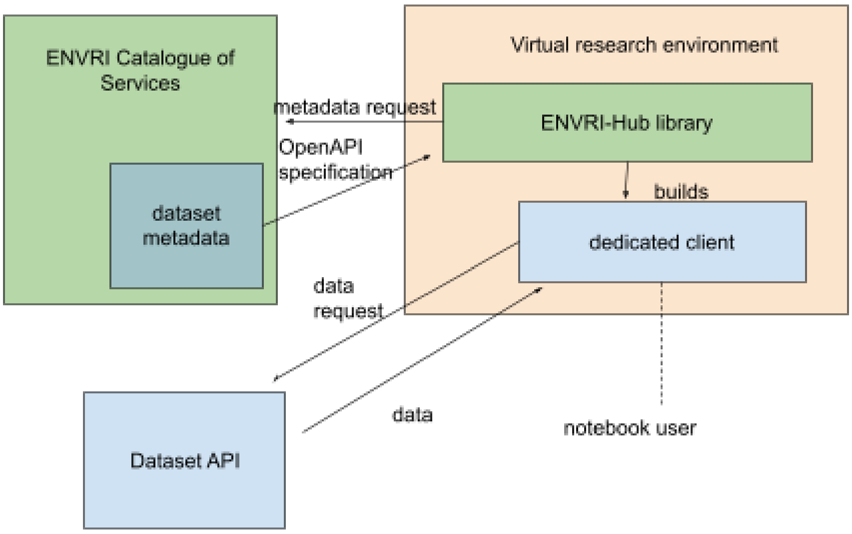

Figure 9: Interactions among the ENVRI-Hub Python library, the ENVRI Catalogue of Services, and Data APIs.

The Analytical Framework, therefore, expects data access services to be documented in the OpenAPI 3.0 specification and said specification to be uploaded into the CoS as data assets are made available to the ENVRI community.

Deliverable 7.2 "*First report on integration of catalogue with the AF*" will offer more details about the connection between the Catalogue of Services and the Analytical Framework.

3.2 RIs (meta)data integration with the Catalogue of Services

The Catalogue of Services allows RIs to share their services on a common catalogue, as long as said services are described by means of an EPOS-DCAT AP file, which is uploaded into the CoS itself. The CoS is agnostic towards what onboarded services are and/or offer, and lets users browse among them, and offers an API to allow access from notebooks and/or other applications.

Figure 10 summarizes this concept.

The Catalogue of Service allows for the onboarding of new services through its back-end (Cos-BE). The requirements for onboarding are:

- The subject requesting the onboarding must either be the service owner or be licensed by the service owner to perform onboarding;
- Service description in EPOS-DCAT-AP format. The documentation of the EPOS-DCAT-AP v1 is available on GitHub [15] and it includes a UML diagram, ontology definition, examples and more details (v1 is used as it has been tested and validated for the EPOS Platform. Version 3, compliant with DCAT-AP v3, is also available and it is undergoing a testing phase).

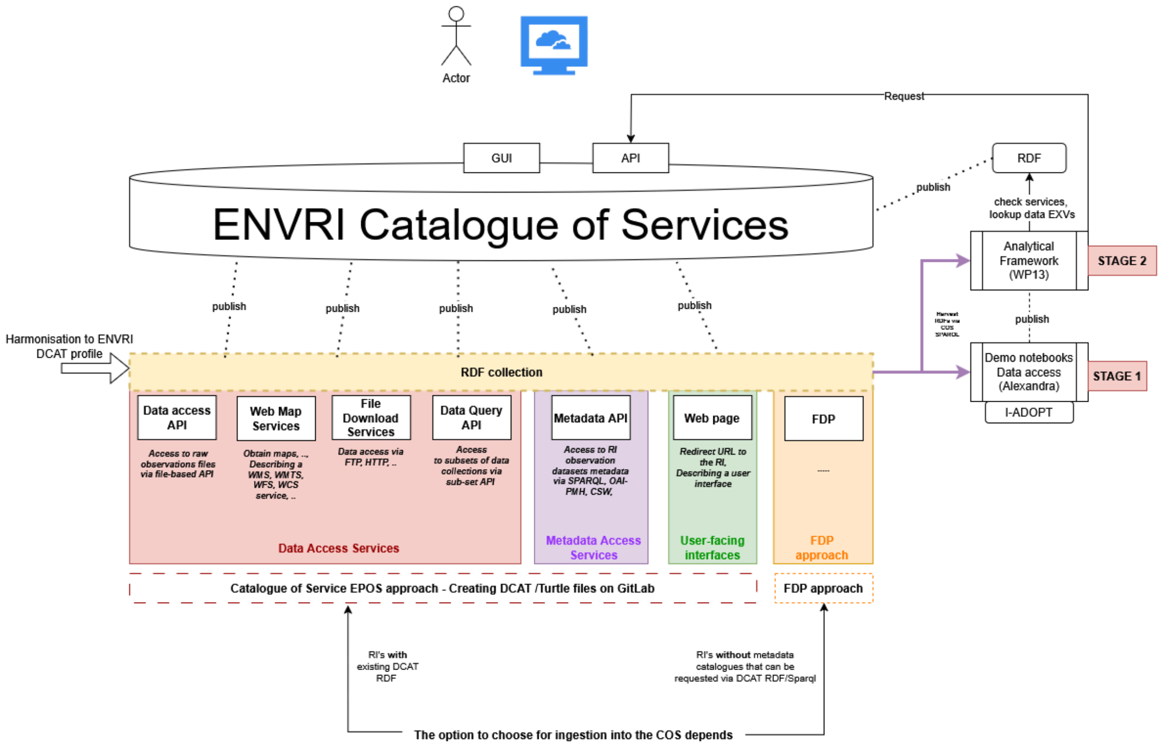

Figure 10 - ENVRI Catalogue of Services architecture.

The onboarding procedure steps are the following:

- RI Services must be described using the EPOS-DCAT-AP and serialized in RDF/Turtle format (.ttl files)- See Appendix II for the metadata service description turtle template;
- EPOS-DCAT-AP enables integrity checks and validation of metadata content by using a set of SHACL constraints;
- Metadata needs to be pushed to the ENVRI-HUB GitLab repository. This will automatically trigger a pipeline for ingestion in a test environment;
- Service owners can check the test environment through the API within GitLab;
- After testing, the service owner can open a merge request to integrate the service into the production environment. This step requires a manual inspection by a dedicated person in the ENVRI-HUB consortium.

Once onboarded on the Catalogue of Services, services will be visible to ENVRI-Hub users through the CoS interface and the VRE-LIB.

3.3 RIs (meta)data integration with the FAIR Data Points

3.3.1 Integration of Federated Catalogs

The FAIR Data Point architecture empowers European ENVRIs to create a **federated, interoperable data ecosystem**. By decentralizing control while enforcing FAIR standards, it enhances data-driven collaboration, supporting large-scale environmental monitoring and policy-making without compromising institutional autonomy.

As the FDP architecture is designed to handle not only datasets but also higher-order catalogs and services, it enables layered, scalable integration of resources across domains. This flexibility is critical for complex ecosystems like environmental research, where infrastructures often manage diverse assets beyond raw data.

Below, we list how the supported catalogs are structured in the FDP reference implementation.

1. **Catalogs of Catalogs**
    - **Hierarchical Aggregation**: Individual FDPs can act as "child" catalogs, while aggregators (e.g., regional or thematic portals) serve as "parent" catalogs, creating a nested structure. For example, a European-wide catalog might integrate sub-catalogs from marine (ICOS, EMODnet), atmospheric (ICOS, ACTRIS and IAGOS), solid earth (EPOS), and ecology and biodiversity (ICOS, ANAEE, eLTER, LifeWatch) infrastructures;
    - **Metadata Propagation**: Metadata from distributed FDPs is harvested and indexed at higher levels (e.g., the European Open Science Cloud), enabling cross-domain discovery without centralizing data storage;
        - **Dynamic Updates**: Changes in local catalogs (e.g., new datasets or services) are propagated to parent catalogs via synchronization protocols (e.g., OAI-PMH, SPARQL Federation).
2. **Catalogs of Services**
    - **Services as First-Class Citizens**: Services (e.g., computational tools, APIs, processing workflows) are treated as FAIR resources, with their own metadata, PIDs, and access protocols. For instance, a climate modeling service from IS-ENES could be registered alongside the datasets it uses;
    - **Service Metadata**: Descriptions include technical details (e.g., input/output formats, endpoints), dependencies (e.g., software libraries), and operational terms (e.g., availability, cost). Standards like **SWORD** or **OpenAPI** ensure machine-actionability;
    - **Integration with Data**: Services can be linked to datasets in metadata (e.g., "this workflow processes dataset X"), enabling automated pipelines. For example, a satellite imagery analysis service in Copernicus might reference Sentinel-2 data;
    - **Federated Queries:** Federated queries can return both datasets and services, allowing users to discover tools alongside data (e.g., "find all ocean acidification datasets *and* visualization services").

By supporting catalogs of catalogs and catalogs of services, the FDP architecture evolves from a data-sharing framework to a unified ecosystem for FAIR resources. This flexibility empowers environmental researchers to navigate Europe's distributed infrastructures seamlessly, bridging gaps between data, tools, and workflows. Such integration accelerates cross-disciplinary collaboration, enabling solutions to grand challenges like climate change and biodiversity loss while adhering to EU open science and interoperability mandates.

3.3.2 Integration of I-ADOPT Variable Metadata Information in FDP Catalogs

The FDP architecture can extend dataset metadata with detailed variable descriptions using the I-ADOPT Framework [16] [17], an ontology designed to standardize the representation of observed or measured variables (e.g., "pH," "soil moisture") across environmental domains.

Implementation Steps:

1. **Variable-Level Metadata Enrichment**:
    - Variables within datasets are described using I-ADOPT's structured vocabulary (e.g., observedProperty, method, unit, resolution). For example:

```
"variable": {
"id":
"[http://example.org/var/sea_surface_temp](http://exampl
e.org/var/sea_surface_temp)",
"label": "Sea Surface Temperature",
"iadopt:observedProperty": "temperature",
"iadopt:method": "satellite infrared sensor",
"iadopt:unit": "degrees Celsius",
"iadopt:temporalResolution": "daily"
}
```

    - Variables are assigned **unique URIs** and linked to their parent datasets in FDP metadata, enabling both human and machine understanding of granular data content.
2. **Semantic Linking**:
    - I-ADOPT terms are mapped to domain-specific ontologies (e.g., NERC NVS, ENVO for environmental terms, QUDT for units) to ensure cross-disciplinary interoperability;
    - Tools like **SPARQL endpoints** or **GraphQL APIs** allow users to query variables directly (e.g., "find all datasets measuring *nitrate concentration* in *river basins*").

This makes sure that the FAIR Compliance is expanded to (Essential) Variables, as Variables will inherit FAIR principles:

- **Findable**: Variables are indexed with their own metadata and linked to datasets;
- **Accessible**: Variable URIs resolve to machine-readable descriptions;
- **Interoperable**: I-ADOPT ensures consistent semantic definitions;
- **Reusable**: Licensing and provenance apply at a variable level (e.g., "CC-BY for temperature data").

3.3.3 Challenges in Search Granularity

Balancing Search Granularity with Collections to prevent users from being overwhelmed by excessive search results while retaining precision, FDP leverages collections as a mechanism to group related resources hierarchically.

To achieve this, collections should be curated by groupings of datasets, variables, or services based on themes (e.g., "Arctic Monitoring"), projects (e.g., "Copernicus Climate Change Service"), or variable types (e.g., "Hydrological Variables").

These collections can be auto-generated using I-ADOPT terms (e.g., "All datasets with variables tagged iadopt:observedProperty=CO2").

In practice, users can start with broad collections (e.g., "Oceanographic Data") and drill down using facets like:

- **Variable type** (via I-ADOPT terms);

- **Spatial/temporal coverage**;
- **Data source** (e.g., sensor type, model).

For example: A search for "*marine biodiversity*" could first return high-level collections (e.g., "*EU Marine Strategy Framework Directive*"), then allow filtering by variables (e.g., "*species richness*," "habitat extent").

Search engines should then prioritize results based on:

- **User context** (e.g., a climate scientist sees model outputs first);
- **Variable usage** (e.g., datasets with variables frequently cited in publications);
- **Machine learning** models can suggest collections based on user behavior.

The FDP collections will be machine-Friendly when collections are exposed as RDF graphs or JSON-LD objects with metadata (e.g., dct:title, dcat:theme), ensuring M2M access. APIs will allow automated workflows to retrieve entire collections (e.g., "*fetch all air_quality variables from Northern Europe*").

While collections simplify human discovery, FDP ensures machines can still access granular data:

- **Dataset-level**: GET /datasets/{id} returns high-level metadata;
- **Variable-level**: GET /datasets/{id}/variables exposes I-ADOPT annotations;
- **Collection-level**: GET /collections/{id}/resources lists all member datasets/services.

By integrating iADOPT for variable-level granularity and collections for layered discovery, the FDP architecture achieves a balance between precision and usability. Researchers avoid information overload while machines retain direct access to fine-grained metadata, enabling scalable, automated workflows. This approach aligns with Europe's vision for a semantically rich, user-centric environmental data ecosystem, bridging the gap between massive data infrastructures and actionable insights.

3.4 I-ADOPT Implementation for Cross-Domain Interoperability

Cross-domain interoperability hinges on the ability to compare and combine datasets describing similar or complementary phenomena, regardless of their disciplinary origin. The **I-ADOPT framework** provides a structured grammar for constructing unambiguous descriptions of what is being measured, observed, or studied, often referred to as "*observable properties*". By breaking these properties into smaller semantic units (e.g., "parameter", "matrix", "procedure", "object of interest"), I-ADOPT establishes a common language that enables more precise data discovery, integration, and reusability.

1. **Motivation for I-ADOPT**
    - **Clarity of observation definitions:** Researchers in different domains may label the same measurement differently. By using a standard grammar and agreed-upon implementation profiles, I-ADOPT reduces ambiguity and promotes consistency;
    - **Enhanced data discovery:** Well-structured, I-ADOPT-compliant variable descriptions allow search engines and catalogue services to infer relationships between similar measurements, even if the naming schemes differ;

- **Futureproofing:** As new measurement techniques or emerging environmental variables appear, RIs can incorporate them into the I-ADOPT schema without losing interoperability with existing concepts.

2. **Practical steps to implementation**
    - **Identify priority variables:** Start with the most frequently used or community-critical variables—often observed variables that are related to Essential Variables (EXVs) and EXVs themselves—since these have the highest impact on cross-domain searches. An example is TEMPPR01, which is a lower-level parameter related to EXV Sea-surface temperature (https://vocab.nerc.ac.uk/collection/EXV/current/EXV017/);
    - **Map existing vocabularies:** Align local or domain-specific vocabularies to I-ADOPT by creating crosswalks for each semantic component (e.g., "property," "medium," "unit");
    - **Leverage community infrastructure:** Collaborate with the I-ADOPT working group or other RIs that have completed full or partial implementations to share and align implementation profiles, lessons learned, tools, and best practices;
    - **Iterative refinement:** As new use cases emerge or additional RIs join, refine the grammar, expand the coverage of terms, and establish versioning to track changes over time.

3. **Challenges and considerations**
    - **Semantic expertise:** Adopting formal grammars often requires training or additional staffing with semantic web or ontology skills;
    - **Tooling and technology:** While a variety of semantic tools exist (e.g., ontology editors, SPARQL endpoints), they must be adapted to each RI's technical stack;
    - **Governance:** Long-term interoperability depends on governance structures for maintaining and updating the vocabulary mappings, including community-driven feedback and version control.

By progressively mapping and atomising their existing concepts to I-ADOPT, RIs can achieve a scalable, modular approach to semantic interoperability. This, in turn, fosters deeper collaboration and more effective data reuse across environmental sciences.

3.5 The EXV Task Force

Within WP7 and WP11, the involved experts and RI representatives have explored in the first year of the project what is possible with the Catalogue of Services, the FDP approach, and I-ADOPT to describe the RI's data access services and available observation datasets related to EXVs. As proof of concept, NOC-BODC created some first notebook experiments using I-ADOPT where possible as a basis for the mapping between EXV and RI parameter requests, and the relevant RI data access services (APIs) to retrieve the datasets from. This has led to a list of the limitations and possibilities. The conclusion was that for providing access to the RI datasets for parameters supporting EXV products:

- A hybrid solution will be needed;
- Depending on the status/availability of data access services at the RI level:
    - Some offer direct subsetting services, observation metadata discovery via DCAT;

  - Other RIs offer more data products that are already harmonised and available via OGC services;
  - Others will work towards using FDP solutions to describe the metadata.

See Figure 9 for the data flow-oriented architecture as an outcome.

To work closer together on this hybrid solution and work towards a first release in M18, the EXV working group has been set up, consisting of I-ADOPT experts, RI data access experts, and led by MARIS.

As can be seen in the above diagram, the main aim is, in the end, to allow users in the analytical framework to process and analyse EXV-related datasets. In order to achieve this, they need to discover RIs datasets via the metadata and data access services offered in the Catalogue of Services, and then access and download these datasets into the Analytic Framework in a smooth machine-to-machine way. There are several hurdles to take before being able to achieve this:

- The metadata and parameters/unit vocabularies are not the same. A mapping is needed, and this will be supported by the I-ADOPT framework and NERC vocabulary service;
- The machine-to-machine services are not similar. Each RI has a different metadata model, different data formats, and a different approach on how to make the datasets discoverable and downloadable;
- The output of the services is different. Some services offer OGC services, others need first a metadata search, and then a second service to retrieve the data files, while some RIs also offer direct data access and subsetting services.

For the first release in M18, the EXV WG will focus on:

- Creating an EXV I-Adopt broker from chosen EXVs to discover the RI's parameter set. This provides information on the degree of compliance of RIs' services and metadata/data with I-adopt for parameters used;
- Accessing the data access services to access observation data with parameters supporting an EXV;
- Each RI develops a Python notebook to use the I-ADOPT approach, find the parameters, and demonstrate access to the data;
- As the last step, create one Python notebook including using all other Python notebooks to request all RI's services using the EXV I-Adopt broker, find the relevant RI data access service and retrieve the datasets;
- All notebooks and the scripts can be used in the Analytical Framework. The main harmonisation achieved lies in the I-ADOPT compliance and the notebook scripts.

We already know that for analytical services, processing and visualisation, individual (and heterogeneous) data files are not sufficiently useful. They need harmonised dataframes ("chunks") for a certain parameter and in similar units, time, geographic system, etc. Not all RIs offer this. For the next period, the WG will work on further harmonisation, upgrading of services where possible, and demonstrate this to AF users.

4 Recommendations for Metadata and Vocabularies Standardisation and Harmonisation

4.1 Recommendations for Metadata Schema

As emerged from the considerations made so far, it is crucial to adopt standard metadata schemas to ensure a smoother integration with the ENVRI-Hub.

The information gathered leads the WP11 team to recommend the practice of making metadata available in multiple formats and according to different schemas.

Among these, the team believes it is important that all infrastructures implement at least one of the **DCAT-AP** and **ISO 19115 family** schemas. These two schemas are already the most commonly used among Environmental Research Infrastructures.

For easy integration with the ENVRI Catalogue of Services, it is essential that the metadata contains key information. The following information should be present at a minimum:

- **Spatial-Temporal Coordinates:** Ensure the metadata includes precise geographic coordinates and the time period covered by the data;
- **Contained Variables:** Clearly describe the observed variables in the data, facilitating understanding and use of the datasets, and the Essential Variable(s) they are related to;
- **Accessibility and Usage:** Provide details on how to access the data and any usage restrictions.

4.2 Recommendations for Metadata Services

In addition to paying attention to the format and schema of the metadata, it is also essential to use standard metadata services.

Even in this case, the WP11 team recommends making metadata available through multiple metadata services.

The results of the landscape analysis have shown that SPARQL endpoints, OGC CSW, OAI-PMH, and CKAN API are the services most commonly used among Environmental Research Infrastructures.

SPARQL endpoints can prove to be the most suitable solution for those infrastructures whose data contains thousands, if not millions, of observation (meta)data files, and is stored in formats to be exposed as triple-stores or linked data.

OGC CSW and **OAI-PMH**, on the other hand, are suitable for exposing metadata of data products and data stored with relational models. Another advantage of these services is that they provide standard interfaces or key operations for querying metadata. This facilitates machine-to-machine interaction for metadata harvesting and reduces the need to develop custom code for integrating the specific RI service into the ENVRI Catalogue of Services.

In any case, the metadata services of each RI should at least provide methods for metadata retrieval based on **spatial-temporal coordinates** and **observed variables**. For proper integration with the ENVRI Catalogue of Services, all this information should be provided in the metadata service description turtle file, possibly specifying also which Essential Variables they are linked to.

4.3 Recommendations for Semantic Cross-Domain Interoperability

Implementing well-structured vocabularies and the I-ADOPT framework across a diverse set of Research Infrastructures requires a roadmap tailored to each domain's starting point—whether an RI is new to controlled variable vocabularies or already using advanced semantic technologies. Below is a suggested phased approach:

1. **Initial assessment and alignment**
    - **Inventory existing practices:** Each RI should document its current variable definitions, observing whether free-text or partial controlled vocabularies are in use or could be used (pref. those that already have I-ADOPT mappings);
    - **Identify overlaps with I-ADOPT:** Match relevant local concepts to I-ADOPT components (e.g., *observable entity*, *parameter*, *matrix*), noting any domain-specific gaps or ambiguities.
2. **Pilot and incremental adoption**
    - **Select key variables (EXVs):** Focus on essential or high-priority variables that are frequently used and have well-established definitions;
    - **Establish mappings and implementation profiles:** Create crosswalks between the local vocabulary and I-ADOPT for these variables, establish templates for how to apply I-ADOPT to variables used in the domain. Validate mappings and profiles via small pilot projects or internal review processes;
    - **Refine governance mechanisms:** Assign roles (e.g., "Vocabulary curator") within each RI to manage ongoing updates, ensure version control, and resolve conflicts or ambiguities.
3. **Scale-out and community integration**
    - **Expand coverage:** After successful pilots, extend I-ADOPT-aligned descriptions to less commonly used variables or newly introduced parameters;
    - **Adopt a shared registry:** Contribute mappings and term definitions to community repositories, enabling broader reuse and consistent versioning across multiple RIs;
    - **Integrate with ENVRI-Hub services:** Ensure that newly defined or revised variable concepts are harvestable by the ENVRI-Hub Catalogue of Services or FDP endpoints, providing cross-RI discoverability.
4. **Iterative improvement and sustainability**
    - **Ongoing maintenance:** Revisit mappings, implementation profiles, and definitions regularly to account for new measurement methodologies, instrument upgrades, or evolving scientific consensus;
    - **Feedback and collaboration:** Collaborate with domain-specific and cross-domain working groups—such as the EXV Task Force—to refine the grammar, incorporate user feedback, and ensure alignment with overarching ENVRI goals;

    - **Alignment with EOSC requirements:** As the European Open Science Cloud evolves, RIs should update semantic infrastructures to comply with emerging EOSC interoperability guidelines, thereby broadening the impact of semantic alignment work.

Domain-specific roadmaps

- **Domains with established vocabularies (e.g., marine, atmospheric):**
    - Already use controlled term sets such as SeaDataNet (P01) or CF Standard Names;
    - Short-term tasks include mapping these existing vocabularies to I-ADOPT grammar constructs and contributing reference examples for others.
- **Domains with partial or proprietary vocabularies (e.g., earth observations, biodiversity):**
    - May require initial standardisation of local implementation profiles before bridging to I-ADOPT;
    - Pilot mappings/profiles on a few well-defined variables, then scale to entire domain coverage.
- **Domains relying primarily on free text (e.g., emergent data science communities):**
    - Begin by identifying the most critical variables, establishing minimal controlled lists, and adopting stable URIs;
    - Ramp up I-ADOPT alignment once a base level of standardisation has been achieved.

In sum, **semantic cross-domain interoperability** is an iterative process that aligns local RIs' vocabularies with a shared grammar such as I-ADOPT. Through systematic assessments, pilot projects, and continuous refinement, ENVRI-Hub NEXT partners can achieve a robust, sustainable approach to describing environmental data, enabling more advanced searches, deeper analytics, and greater scientific collaboration across domains.

4.4 Metadata Requirements for Integration with EOSC

The EOSC Federation aims to provide Europe's researchers with the necessary digital resources to conduct research within and across disciplines and borders.

This is done according to FAIR data principles, open science, in a trustworthy and secure environment driven by the scientific communities, as described by the EOSC Federation Handbook [18].

EOSC Nodes are distributed entities representing organisational and technical structures enabling the EOSC Federation. An EOSC Node provides resources to the Federation following the EOSC Node architecture and a set of EOSC interoperability guidelines, defined according to the type of resources they provide.

This section outlines the data model for publishing resources to the EOSC sandbox environment as delivered by the EOSC Beyond project [19], and subsequently to the EOSC EU Node [20] when the resources are production-ready, regardless of whether they are provided through an EOSC Node. It also describes the types of resources offered by the ENVRI-Hub described in this document (CoS, RDF, FDP), along with the associated metadata requirements necessary for publication.

The resources made available to EOSC are interlinked according to the data model presented in Figure 10.

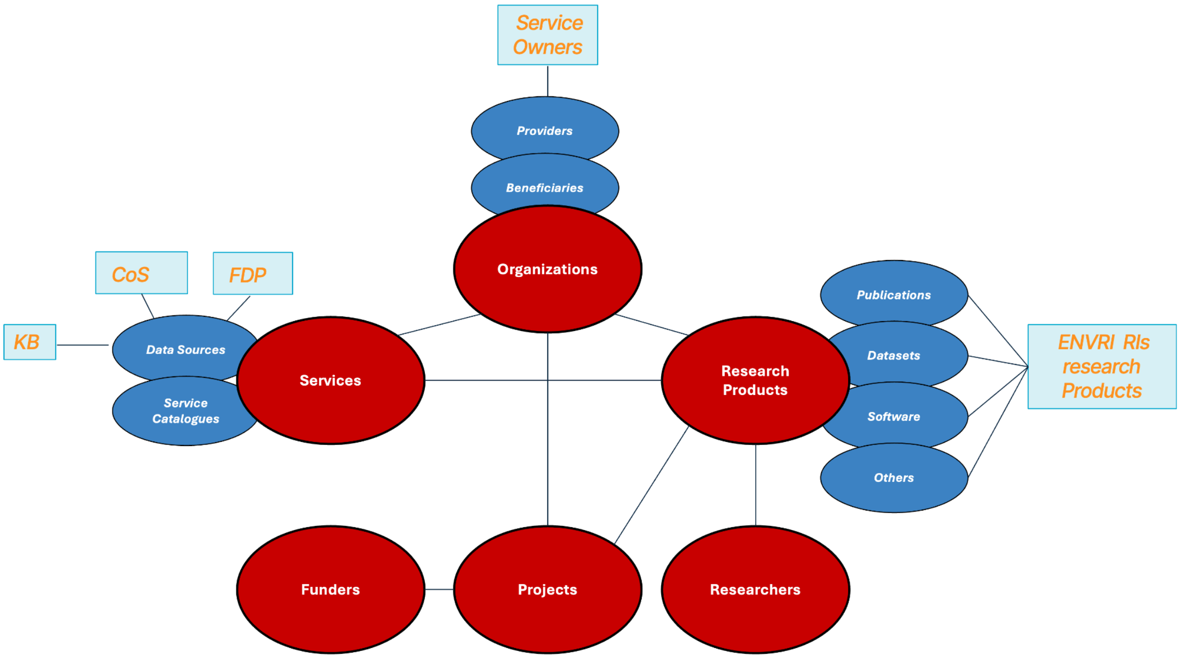

Figure 11 - EOSC resources data model and mapping to ENVRI-Hub resource types.

Following this model, the "data sources" are identified as the entities representing, for example, the ENVRI-Hub Catalogue of Services. Data sources are a type of service representing aggregations of research products, mostly datasets distributed via the ENVRI-Hub CoS. Accordingly, we outline the metadata requirements necessary to onboard a data source and its associated research products.

Providers Profile

Before onboarding any type of services or data sources, in this case, a service owner must be registered as a provider. At the time of writing, the data model is under review to accommodate the integration of EOSC Nodes. Providers are required to complete administrative details as specified in the EOSC Provider Profile v5.0 [21].

Service Profile

Data sources identified as CoS in ENVRI-Hub must be registered in accordance with the EOSC service profile [22].

While the registration of these services in a testing environment may not require a fully rigorous onboarding process, such as the provision of access or privacy policies, these elements are mandatory in operational environments. Therefore, it is recommended to define them already during the staging phase.

Research Products

Individual research products aggregated within a data source must comply with the EOSC-IF Guidelines for Data Sources to onboard Research Products.

These guidelines support the onboarding of research products, including datasets, made available through data sources using the OAI-PMH protocol and following DataCite-based metadata schemas [23].

The EOSC Beyond Sandbox environment offers both web interfaces and APIs to support the bulk onboarding of resources. ENVRI-Hub NEXT will monitor the release of new versions of the profiles to test the interoperability of an ENVRI Node with the EOSC-Interoperability Framework (EOSC-IF).

5 Conclusions

In this document, the results obtained by WP11 in the first phase of the ENVRI-Hub NEXT project have been presented.

The landscape analysis showed a moderate level of convergence among the RIs towards the use of standard metadata schemas and services.

The implementation of the recommendations provided in this document by the infrastructures can constitute a further step towards complete standardisation and can facilitate the onboarding of their metadata into the ENVRI-Hub.

Moreover, the use of standard metadata schemas and services can pave the way towards the ultimate goal of having harmonised metadata and data that can be used, analysed, and combined synergistically.

The work of harmonisation requires further investigation due to several factors. As said, the first is the heterogeneity in metadata approaches and outputs across RIs. Additionally, the maturity level in the use of vocabulary varies significantly. The diversity and complexity in data access procedures further complicate the integration process.

While the EXV Task Force has made initial strides by developing scripts to retrieve data and using I-ADOPT queries, many parts of the code are identical, indicating some progress in harmonisation at the programmatic level.

However, integration and harmonisation of metadata contents and services still need to be thoroughly investigated.

Establishing minimum requirements for metadata content and services is crucial to ensure consistency and interoperability.

Some steps for further harmonisation that have started or been achieved during this work are:

- **Standardisation of Metadata Schemas:** Establishing common standards for metadata schemas to reduce heterogeneity and ensure consistency across different research infrastructures;
- **Developing Minimum Requirements:** Establishing minimum requirements for metadata contents and services (including geospatial, temporal, and EXV-related observed variables) to ensure comprehensive and interoperable data;
- **Programmatic Harmonisation:** Continuing efforts to harmonise metadata at the programmatic level, using scripts and codes to retrieve and standardise data;
- **Collaborative Efforts:** Engaging in collaborative efforts among RIs to share best practices and tools for harmonisation.

The next steps for harmonisation in metadata and data access involve several key actions:

- **Improving Vocabulary Usage:** Enhancing the maturity level in the use of vocabularies by adopting standardised terminologies compliant with I-ADOPT. Semantic cross-domain interoperability is an iterative process that aligns local RIs' vocabularies with a shared grammar such as I-ADOPT. Through systematic assessments, pilot projects, and continuous refinement, ENVRI-Hub NEXT partners can achieve a robust, sustainable approach to describing environmental data, enabling more advanced searches, deeper analytics, and greater scientific collaboration across domains;

- **Simplifying Data Access Procedures:** Streamlining and standardising data access procedures to reduce complexity and improve usability.
    - The ENVRI Catalogue of Services (CoS) represents a central access point for web services providing data from environmental RIs, with a standardised and interoperable interface;
    - The ENVRI FDP prototype is an example of an approach for populating the CoS. It demonstrates how a federated, standards-compliant architecture can be effectively deployed within a complex multi-infrastructure research ecosystem. Through the integration of Essential Variables using I-ADOPT, enhanced SPARQL support, and a customizable user interface, the implementation supports both human usability and machine interoperability;
    - Put the focus on the potential of achieving the next step beyond metadata harmonisation and service descriptions, which is actual Data federation of subsets of data from different RIs. This is the step to go from file-based request to actual subsets (or data cubes) of data, which can be processed in Virtual Research Environments as part of EOSC.
- **Integration and Harmonisation:** Investigating further internal and external (with EOSC) integration and harmonisation of metadata contents, schema, and services.
- **Ongoing Evaluation:** Regularly evaluating the harmonisation process to identify areas for improvement and ensure the effectiveness of implemented standards.

These actions will be the focus of WP11 activities in the second phase of the ENVRI-Hub NEXT project, to promote scalable and semantically rich discovery and reuse of environmental data and services across the European Research Infrastructures landscape.

6 References

Reference	
No	Description/Link
R1	Essential Climate Variables https://gcos.wmo.int/site/global-climate-observing-system-gcos/essential-climate-variables
R2	Essential Ocean Variables https://goosocean.org/what-we-do/framework/essential-ocean-variables/
R3	Essential Biodiversity Variables https://geobon.org/ebvs/what-are-ebvs/
R4	Essential Geodetic Variables https://ggos.org/about/org/bureau/bps/cwg/essential-geodetic-variables/
R5	Introduction to Metadata at NCEI. https://www.ncei.noaa.gov/resources/metadata/introduction-to-metadata-at-ncei
R6	Metadata Standards. https://help.eds.ukri.org/article/5180-metadata-standards
R7	Metadata for Data Management: A Tutorial. https://guides.lib.unc.edu/metadata/definition
R8	ISO 19115-4 https://committee.iso.org/sites/tc211/home/projects/projects---complete-list/iso-19115-4.html
R9	ISO 19115-5 https://committee.iso.org/sites/tc211/home/projects/projects---complete-list/iso-19115-5.html
R10	SPARQL endpoints SPARQL endpoints and how to use them
R11	OGC CSW Catalog Services for the Web (CSW) Standard | OGC
R12	OAI-PMH

	Open Archives Initiative Protocol for Metadata Harvesting
R13	CKAN Features Features
R14	CKAN API Features API Feature Detail Page
R15	Documentation of the EPOS-DCAT-AP v1.0 extension https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/docs/EPOS-DCAT-AP_extension_v1.0.pdf
R16	InteroperAble Descriptions of Observable Property Terminology WG https://i-adopt.github.io/
R17	**InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) WG Outputs and Recommendations.** B. Magagna et al., 2022. DOI: https://zenodo.org/records/6520132#.YnOhMJNBybQ
R18	EOSC Federation Handbook https://zenodo.org/records/14999577
R19	EOSC Innovation Sanbox https://sandbox.eosc-beyond.eu
R20	EOSC EU Node https://open-science-cloud.ec.europa.eu
R21	EOSC Provider profile documentation https://eosc-provider-profile.readthedocs.io
R22	EOSC Service profile documentation https://eosc-service-profile.readthedocs.io/en/5.0/
R23	OpenAIRE Guidelines https://guidelines.openaire.eu/en/latest/

Appendix I: FDP Demo Instance GUI for the ENVRI-Hub

The FDP Demo Implementation includes a generic but customizable user interface (UI), allowing both end users and metadata providers to interact with FDP catalogs at different levels.

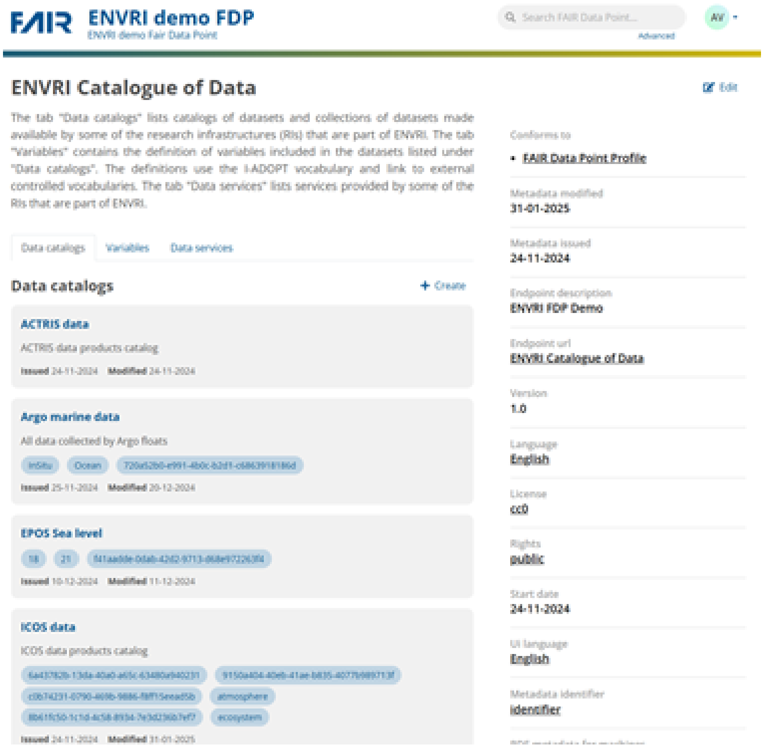

Figure 12 - ENVRI demo FDP top-level catalog.

At the top level, the demo instance provides:

- A **catalog of catalogs**, representing participating RIs,
- A **catalog of datasets** and their respective distributions,
- A **catalog of data services**,
- A **catalog of iAdopt-annotated Essential Variables** (see Figure 2).

All metadata is structured according to DCAT v3, allowing inclusion of geometry, temporal coverage, attribution, rights, license information, and direct access links to RI landing pages or downloadable data endpoints.

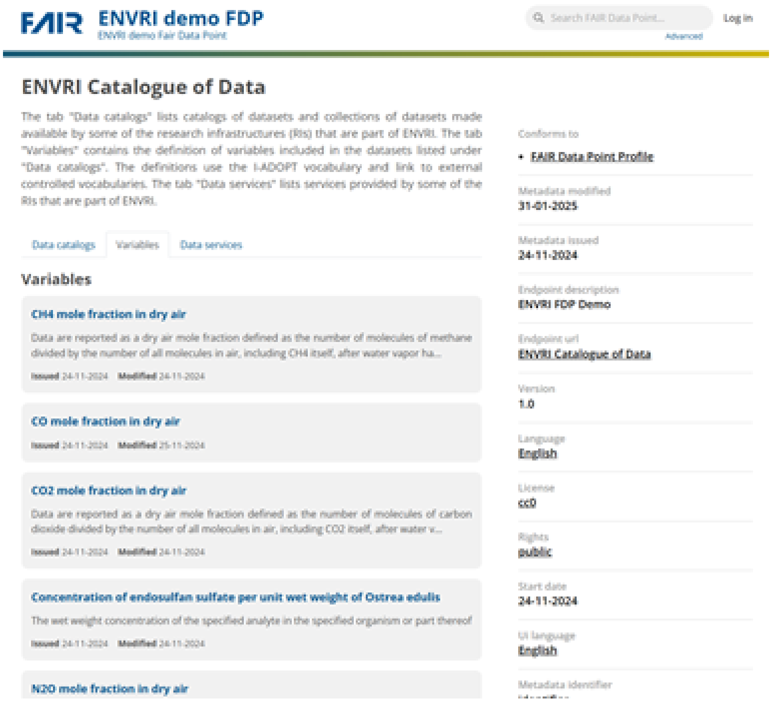

Figure 13 - Essential Variables catalog.

The GUI includes a dedicated interface for browsing and editing Essential Variables annotated using the iAdopt ontology. Variables are assigned URIs and linked to parent datasets via semantic references.

Figure 14 - Variable editing form (example: methane).

Each variable can be semantically linked to:

- Internal RI definitions,
- External vocabularies (e.g., NERC, ENVO),
- Observed properties, methods, units, and temporal resolution.

This ensures interoperability across infrastructures while preserving local semantic context. All metadata is editable through the GUI and also accessible and modifiable via the FDP API (subject to appropriate authorization levels).

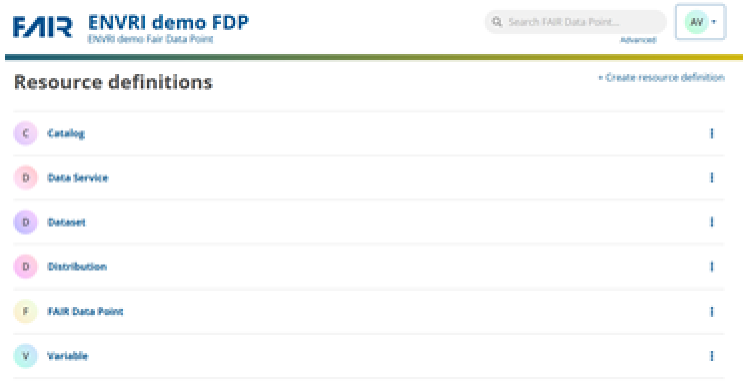

Figure 15 - Metadata element editing interface.

The same editing capabilities apply to metadata for datasets and services. GUI-based forms are used to enter metadata, with structures defined via configurable schemas.

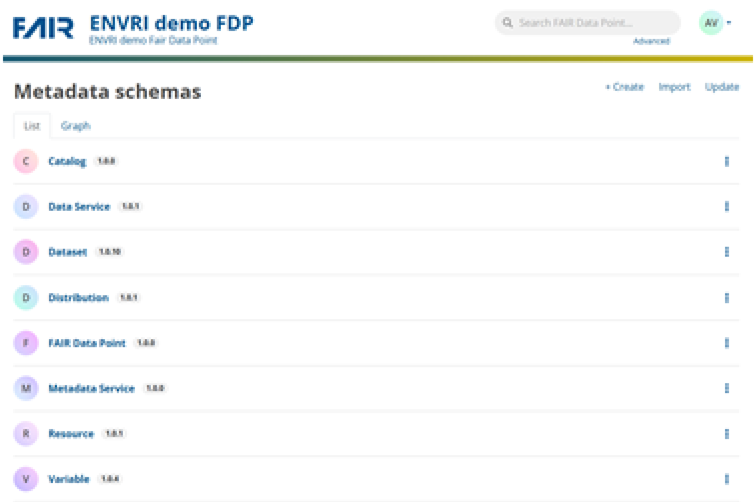

Figure 16 - Metadata schema editor.

The metadata schemas themselves are editable in the GUI, allowing administrators to define:

- Metadata fields,
- Validation rules,
- Display logic.

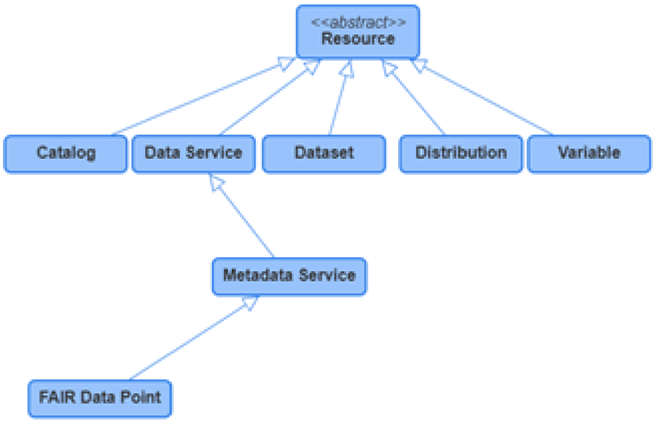

Figure 17 - Schema relations graph.

The GUI provides a visual graph of the relationships between metadata schemas, helping administrators understand schema hierarchies and dependencies.

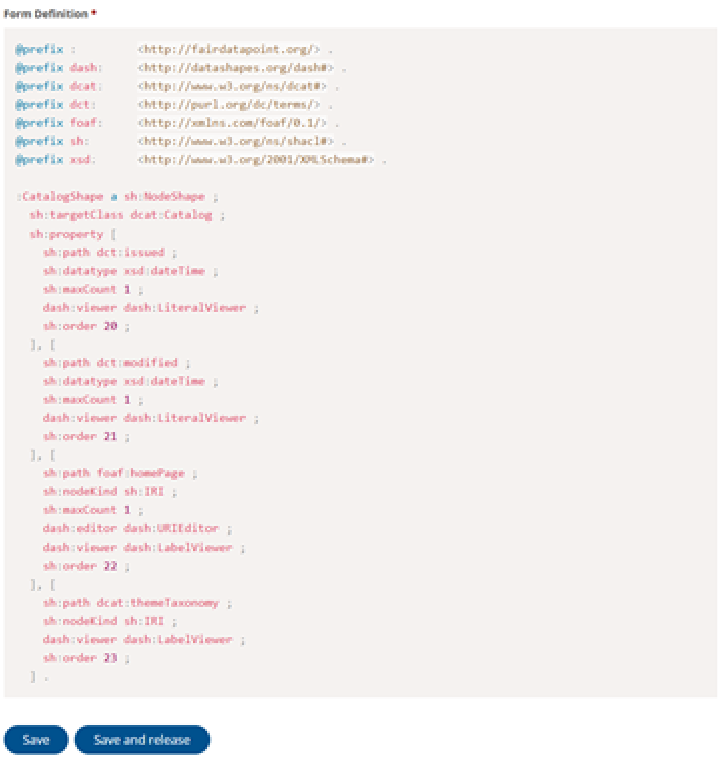

Figure 18 - Form definition using SHACL/DASH.

Form definitions use standards such as SHACL and DASH to validate metadata input. An example is shown for defining a Catalog resource.

Graphical User Interface and Metadata Editing

The user interface allows metadata to be discovered and queried interactively. Variables, datasets, and services can be filtered by themes, spatial coverage, time range, and linked ontologies.

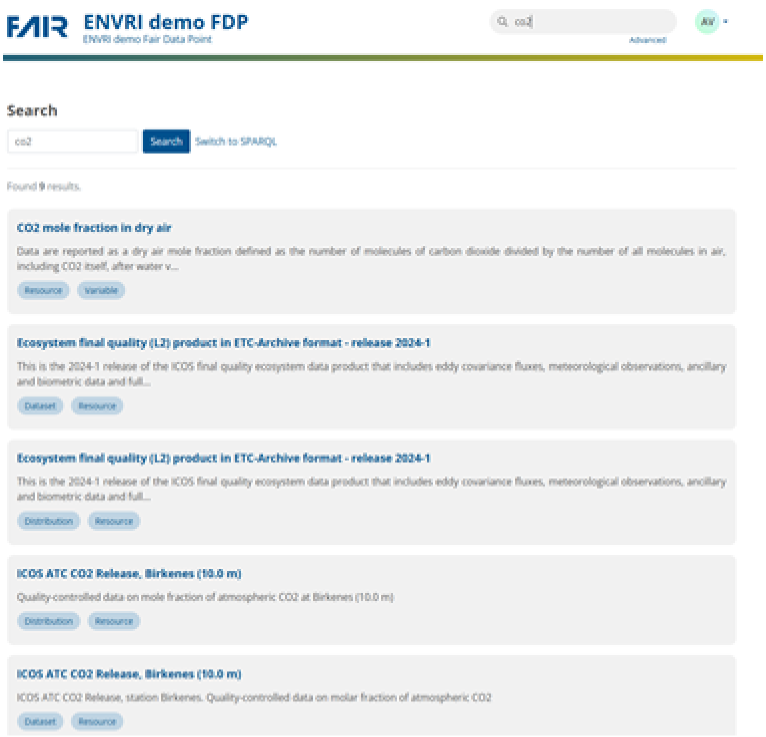

Figure 19 - Example SPARQL query result for a variable search.

The results are displayed in the GUI and are also retrievable in machine-readable formats (JSON, RDF) using content negotiation. The open SPARQL endpoint currently supports JSON results, enabling integration into automated workflows or external applications.

Appendix II: CoS Metadata Service Description Turtle Template

The last version of the metadata service description turtle template is available here.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix epos: <https://www.epos-eu.org/epos-dcat-ap#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix hydra: <http://www.w3.org/ns/hydra/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix schema: <http://schema.org/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix cnt: <http://www.w3.org/2011/content#> .
@prefix locn: <http://www.w3.org/ns/locn#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix http: <http://www.w3.org/2006/http#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix gsp: <http://www.opengis.net/ont/geosparql#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dqv: <http://www.w3.org/ns/dqv#> .
@prefix oa: <http://www.w3.org/ns/oa#> .

#N.B. Properties that are not used can be commented out by adding a '#' at the beginning of the line.

<PIC:000000001> a schema:Organization ;
        schema:identifier [ a schema:PropertyValue ;
        schema:propertyID "PIC" ;
        schema:value "000000001" ;
        ];
        schema:legalName "Institute test1" ;
        schema:address [ a schema:PostalAddress ;
        schema:streetAddress "address, 002" ;
        schema:addressLocality "Rome" ;
        schema:postalCode "00143" ;
        schema:addressCountry "Italy" ;
        ];
    schema:logo "http://www.test.it/logo.png"^^xsd:anyURI ;
    schema:url "http://www.test.it"^^xsd:anyURI ;
    schema:email "test1_institute@email.it" ;
    schema:telephone "+0302206911" ;
```

```
 .

<http://orcid.org/0000-0000-0000-0001> a schema:Person ;
  schema:identifier [ a schema:PropertyValue ;
   schema:propertyID  "orcid" ;
   schema:value   "0000-0000-0000-0001" ;
  ];
  schema:familyName "Surname1" ;
  schema:givenName "Name1" ;
  schema:address [ a schema:PostalAddress ;
                    schema:streetAddress "address, 1" ;
                    schema:addressLocality "Rome" ;
                    schema:postalCode "00143" ;
                    schema:addressCountry "Italy" ;
           ];
  schema:email "test1@private.com" ;
  schema:telephone "+001002003004" ;
  schema:url  "http://orcid.org/0000-0000-0000-0001"^^xsd:anyURI ;
  schema:qualifications "Researcher" ;
  schema:affiliation <PIC:000000001> ;
  schema:contactPoint <http://orcid.org/0000-0000-0000-0001/Contact> ;
 .

<http://orcid.org/0000-0000-0000-0001/Contact> a schema:ContactPoint ;
          schema:email "test1@email.it ";
          schema:availableLanguage "en" ;
          schema:contactType "manager";
 .

<https://envri.eu/example/Dataset/001> a dcat:Dataset ;
    dct:identifier "https://envri.eu/example/Dataset/001" ;
    dct:title "Title of the Dataset" ;
    dct:description "Description of the Dataset" ;

    #adms:identifier [ a adms:Identifier ;
    #   adms:schemeAgency "DOI" ;
    #   skos:notation "00.00000/000" ;
    #];
    # This property links the Dataset to an available Distribution.
    dcat:distribution <https://envri.eu/example/Distribution/001> ;
    dcat:contactPoint <http://orcid.org/0000-0000-0000-0001/Contact> ;
    schema:variableMeasured ""; #EXV(s) https://schema.org/variableMeasured
    dcat:keyword "keyword1","keyword2" ;
         dcat:theme  "http://data.europa.eu/8mn/euroscivoc/909aa237-d251-4262-81c2-fb5e21365f90"^^xsd:anyURI  ;  #Science
domain (URL or SkosConcept)
    dct:publisher <PIC:000000001>;
```

```
    # This property refers to the type of the Dataset.
    # The possible types are here: http://dublincore.org/2012/06/14/dctype
    dct:type "http://purl.org/dc/dcmitype/Collection"^^xsd:anyURI ;
    # This property refers to the frequency at which the Dataset is updated.
    # Example of frequency using a controlled vocabulary: http://purl.org/cld/freq/
    dct:accrualPeriodicity "http://purl.org/cld/freq/continuous"^^xsd:anyURI ;
    # This property refers to a geographic region that is covered by the Dataset.
    dct:spatial [ a dct:Location ;
     locn:geometry "POLYGON((-25 57,37 60,57 22,-22 28,-25 57))"^^gsp:wktLiteral ;
    ];
    # This property refers to a temporal period that the Dataset covers.
    dct:temporal [ a dct:PeriodOfTime ;
        schema:startDate "2020-01-01T00:00:00Z"^^xsd:dateTime ;
        schema:endDate "2020-10-01T00:00:00Z"^^xsd:dateTime ;
    ] ;
    dct:created "2020-01-01T00:00:00Z"^^xsd:dateTime ;
    dct:issued "2020-02-01T00:00:00Z"^^xsd:dateTime ;
    dct:modified "2020-03-01T00:00:00Z"^^xsd:dateTime ;
    owl:versionInfo "1.0.0" ;

    dcat:theme  <category:ENVRI_conc> ;
  .

<https://envri.eu/example/Distribution/001> a dcat:Distribution ;
    dct:identifier "https://envri.eu/example/Distribution/001" ;
    dct:title "Title of the Distribution" ;
          dct:description "Description of the Distribution" ;
    # This property refers to the type of the Distribution. The possible types are
    # (http://publications.europa.eu/resource/authority/distribution-type/WEB_SERVICE
    # OR
    # http://publications.europa.eu/resource/authority/distribution-type/DOWNLOADABLE_FILE)
          dct:type "http://publications.europa.eu/resource/authority/distribution-type/WEB_SERVICE"^^xsd:anyURI ;


    # This property contains a URL that gives access to a Distribution of the Dataset.
    # The resource at the access URL may contain # information about how to get the Dataset.
      # If the type of Distribution is WEB SERVICE, this property refers to the Operation of the Web Service to which the Distribution
conforms.
      # If the type of Distribution is DOWNLOADABLE FILE this property contains the URL that is a direct link to a downloadable file in
a given format.
                                        dcat:accessURL <https://envri.eu/example/Operation/001> ; #dcat:accessURL
"https://test/aaaa/files/file_2018.zip"^^xsd:anyURI ;


      # If the type of Distribution is WEB SERVICE, this property refers to the Web Service that gives access to a Distribution of the
Dataset.
    # Otherwise, this property is not necessary.
          dct:conformsTo <https://envri.eu/example/WebService/001> ;
```

```
    # If the type of Distribution is DOWNLOADABLE FILE, this property contains a URL that is a direct link to a downloadable file in a given format.
    # Otherwise, this property is not necessary.
    # dcat:downloadURL "TestURL"^^xsd:anyURI ;

    # This property refers to the file format of the Distribution.
    # The possible types are here: http://publications.europa.eu/mdr/resource/authority/file-type/html/filetypes-eng.html
    dct:format "http://publications.europa.eu/resource/authority/file-type/XML"^^xsd:anyURI ;

    # This property refers to the licence under which the Distribution is made available.
    dct:license "http://creativecommons.org/licenses/by/4.0/"^^xsd:anyURI ;

    dct:issued "2020-01-01"^^xsd:date ;
    dct:modified "2020-03-01"^^xsd:date ;
 .

<https://envri.eu/example/WebService/001> a epos:WebService ;
 schema:identifier "https://envri.eu/example/WebService/001" ;
 schema:description "Description of Web service" ;
 schema:name "Name of Web Service" ;
 dcat:contactPoint <http://orcid.org/0000-0000-0000-0001/Contact> ;
 dcat:theme "http://data.europa.eu/8mn/euroscivoc/909aa237-d251-4262-81c2-fb5e21365f90"^^xsd:anyURI ; #Service type (URL or SkosConcept)
 # This property refers to the API definitions (e.g., WSDL, WADL)
 hydra:entrypoint "https://envri.eu/example/WebService/001/application.wadl"^^xsd:anyURI ;
 schema:provider <PIC:000000001>;
 schema:datePublished "2020-01-01T00:00:00Z"^^xsd:dateTime ;
 schema:dateModified "2020-02-01T00:00:00Z"^^xsd:dateTime ;

 hydra:supportedOperation <https://envri.eu/example/Operation/001>;
 schema:keywords "keyword1", "keyword2" ;
 # This property refers to a temporal period that the Dataset covers.
 dct:temporal [ a dct:PeriodOfTime ;
     schema:startDate "2020-01-01T00:00:00Z"^^xsd:dateTime ;
     schema:endDate "2020-11-01T00:00:00Z"^^xsd:dateTime ;
 ] ;
 # This property refers to the licence under which the Web Service can be used or reused.
 dct:license "http://creativecommons.org/licenses/by/4.0/"^^xsd:anyURI ;

 # This property refers to the API documentation. (Optional)
 dct:conformsTo <https://envri.eu/example/WebService/001/APIDocumentation> ;
 .

<https://envri.eu/example/WebService/001/APIDocumentation> a hydra:ApiDocumentation ;
   hydra:title "documentation title" ;
```

```
    hydra:description "some description of API Documentation" ;
    hydra:entrypoint "API-Documentation-url"^^xsd:anyURI ;
 .

<https://envri.eu/example/Operation/001> a hydra:Operation;
  hydra:method "GET";

 # This property is used to specify the output format of the Operation.
 # The possible values are listed here:  https://www.iana.org/assignments/media-types/media-types.xhtml
  hydra:returns "application/xml" ;
 hydra:property[ a hydra:IriTemplate ;
             hydra:template "http://www.test.org/{?param1,param2,param3}"^^xsd:string ;
           hydra:mapping[ a hydra:IriTemplateMapping ;
                # This property contains the name of the parameter as required by web service specifications
                hydra:variable "param1"^^xsd:string ;
                               rdfs:label "Label of param1" ;
                # This property contains true if the property is required, false otherwise.
                hydra:required "true"^^xsd:boolean ;
                # This property contains the type of parameter.
                                  # The possible values for this property are: "xsd:string" "xsd:boolean" "xsd:date" " xsd:dateTime"
"xsd:decimal" "xsd:double" "xsd:float" "xsd:int" "xsd:integer" "xsd:long";
                rdfs:range "xsd:string" ;
                # This property contains the default value of the parameter
                schema:defaultValue "defaultvalue" ;
                  ];
           hydra:mapping[ a hydra:IriTemplateMapping ;
                # This property contains the name of the parameter as required by web service specifications
                hydra:variable "param2"^^xsd:string ;
                 rdfs:label "Label of param2" ;
                # This property contains true if the property is required, false otherwise.
                hydra:required "false"^^xsd:boolean ;
                # This property contains the type of parameter.
                      # The possible values for this property are: "xsd:string" "xsd:boolean" "xsd:date" " xsd:dateTime" "xsd:decimal"
"xsd:double" "xsd:float" "xsd:int" "xsd:integer" "xsd:long";
                rdfs:range "xsd:double" ;
                # This property contains the default value of the parameter
                schema:defaultValue "10.9" ;
                # The minimum value of the parameter
                schema:minValue "10.1" ;
                # The maximum value of the parameter
                schema:maxValue "13.5" ;
                ];
           hydra:mapping[ a hydra:IriTemplateMapping;
                                        hydra:variable "param3"^^xsd:string;
                                        rdfs:range "xsd:string";
```

```
                        rdfs:label "Label of param3";
                        http:paramValue "value1";
                        http:paramValue "value2";
                        http:paramValue "value3";
                        schema:defaultValue "value1";
                        hydra:required "false"^^xsd:boolean;
                ];
];
```