ENVRI-Hub NEXT

D7.5

Report on the Fair Data Point Implementation for ECVs

Status: Under EC Review

Dissemination Level: Public

Funded by the European Union

Abstract	
Keywords	FAIR, Metadata, ECV, ENVRI

This deliverable presents the concept of FAIR Data Point (FDP) and describes the most relevant functionalities of the FDP demonstrator that was deployed for the purpose of the ENVRI-Hub NEXT (EHN) project, including how the concept of Essential Climate Variable (ECV) with I-ADOPT annotations was implemented in the FDP demonstrator.

The conclusions drawn from the first half of the EHN project are that the FDP can play a useful role in the Hub architecture to describe individual datasets, but a number of issues still need to be addressed in the second part of the project:

- Find an agreement among all participating research infrastructures (RIs) about the granularity of datasets that should be included in the FDP;
- Find technical solutions for all RIs to add metadata about all relevant datasets to the FDP;
- Fix some issues in the FDP reference implementation (which is used for the EHN FDP).

Revision History			
Version	**Date**	**Description**	**Author/Reviewer**
V 0.1	11/07/2025	ToC and first draft	Jonathan Schenk (ULUND/ICOS)
V 0.2	15/07/2025	External Review	Marta Gutierrez (EGI) Peter Thijsse (MARIS) Delphine Dobler (Euro-Argo ERIC)
V 0.3	16/07/2025	Document Review	Jonathan Schenk (ULUND/ICOS)
V 1.0	**21/07/2025**	**Final Version**	Jonathan Schenk (ULUND/ICOS)

Document Description			
D7.5 - Report on the Fair Data Point Implementation for ECVs			
Work Package Number 7			
Document Type	Deliverable		
Document Status	Under EC Review	**Version**	1.0
Dissemination Level	Public		
Copyright Status	 This material by Parties of the ENVRI-Hub NEXT Consortium is licensed under a Creative Commons Attribution 4.0 International License.		
Lead partner	ULUND		
Document Link	https://documents.egi.eu/document/4042		
DOI	https://zenodo.org/records/16572548		
Author(s)	• Jonathan Schenk (ULUND/ICOS)		
Reviewers	• Marta Gutierrez (EGI) • Peter Thijsse (MARIS)		
Moderated by:	• Matteo Agati (EGI)		
Approved by:	• Delphine Dobler (Euro-Argo ERIC)		

Terminology / Acronyms	
Term/Acronym	**Definition**
API	Application Programming Interface
CoS	Catalogue of Services
DCAT	Data Catalog Vocabulary
EHN	ENVRI-Hub NEXT
ECV	Essential Climate Variable
FAIR	Findable Accessible Interoperable Reusable
FDP	FAIR Data Point
GUI	Graphical User Interface
I-ADOPT	Interoperable Descriptions of Observable Property Terminology
ICOS	Integrated Carbon Observation System
JSON-LD	JavaScript Object Notation for Linked Data
NERC	Natural Environment Research Council
NVS	NERC Vocabulary Server
RDF	Resource Description Framework
RI	Research Infrastructure
SHACL	Shapes Constraint Language
VRE	Virtual Research Environment
W3C	World Wide Web Consortium

Useful Reference: https://doi.org/10.5281/zenodo.14794634

Table of Contents

Table of Figures

Executive Summary

This deliverable presents the concept of FAIR Data Point (FDP) and describes the most relevant functionalities of the FDP demonstrator that was deployed for the purpose of the ENVRI-Hub NEXT (EHN) project, including how the concept of Essential Climate Variable (ECV) with I-ADOPT annotations was implemented in the FDP demonstrator.

The conclusions drawn from the first half of the EHN project are that the FDP can play a useful role in the Hub architecture to describe individual datasets, but a number of issues still need to be addressed in the second part of the project:

- Find an agreement among all participating research infrastructures (RIs) about the granularity of datasets that should be included in the FDP;
- Find technical solutions for all RIs to add metadata about all relevant datasets to the FDP;
- Fix some issues in the FDP reference implementation (which is used for the EHN FDP).

1. Introduction

The ENVRI-Hub NEXT (EHN) project aims to provide a centralized access point for discovering and accessing metadata and data that are produced, maintained and/or stored by the participating environmental research infrastructures (RIs). The Catalogue of Services (CoS) and the Analytical Framework are key components of the Hub architecture to reach this goal. However, the harmonization of metadata about individual (observation) datasets is not fully handled by the CoS, which focuses on services, and would be beneficial for the Analytical Framework. This harmonization, and provision of an easy-to-set-up access point for DCAT V3 RDF metadata, can be accomplished by the implementation of a FAIR Data Point (FDP) [**R1**], which serves as an harmonization interface between the metadata schemes of the participating RIs.

Furthermore, the concept of **Essential Climate Variable** (ECV) has been proposed as the science-based framework guiding the discovery and access of metadata and data via the Hub. This concept can be applied at the granularity level of individual datasets and is integrated in the EHN FDP. The I-ADOPT framework ontology [**R2**] is used to describe ECVs in the FDP so as to facilitate mappings to external vocabularies.

2. FAIR Data Point

The FDP [R3] is a metadata provisioning infrastructure that implements the FAIR concepts (metadata stewardship to maximise Findability, Accessibility, Interoperability and Reusability of research assets) [R4] and thus facilitates the creation of FAIR metadata and data repositories that can be accessed both by humans and machines. The core infrastructure of the FDP is defined in the FDP specification [R1], which then needs to be implemented into a concrete tool. The FDP reference implementation [R5] was developed by some of the creators of the FDP concept to provide an out-of-the-box implementation of the FDP infrastructure.

The FDP stores the metadata as RDF triples (Resource Description Framework) in an RDF metadata store. The metadata content of the FDP follows the World Wide Web Consortium (W3C) Data Catalog Vocabulary (DCAT) model [R6] and must be presented at least in the Turtle and JavaScript Object Notation for Linked Data (JSON-LD) syntaxes. The DCAT model, together with its FDP extension, shown in Figure 1, follows a hierarchical ordering of concepts. The FAIR Data Point, which is also a Metadata Service in the extended DCAT model, is the root concept. A basic FAIR Data Point must contain at least one Catalog, which may be related to some Datasets, which may in turn be related to some Distributions. Additional concepts beyond these core DCAT/FDP concepts can be added to enrich the metadata model. The minimum requirement of the FDP is that the metadata records of the Metadata Service/FAIR Data Point are provided and that metadata records of all other types of resources are grouped in at least one Catalog.

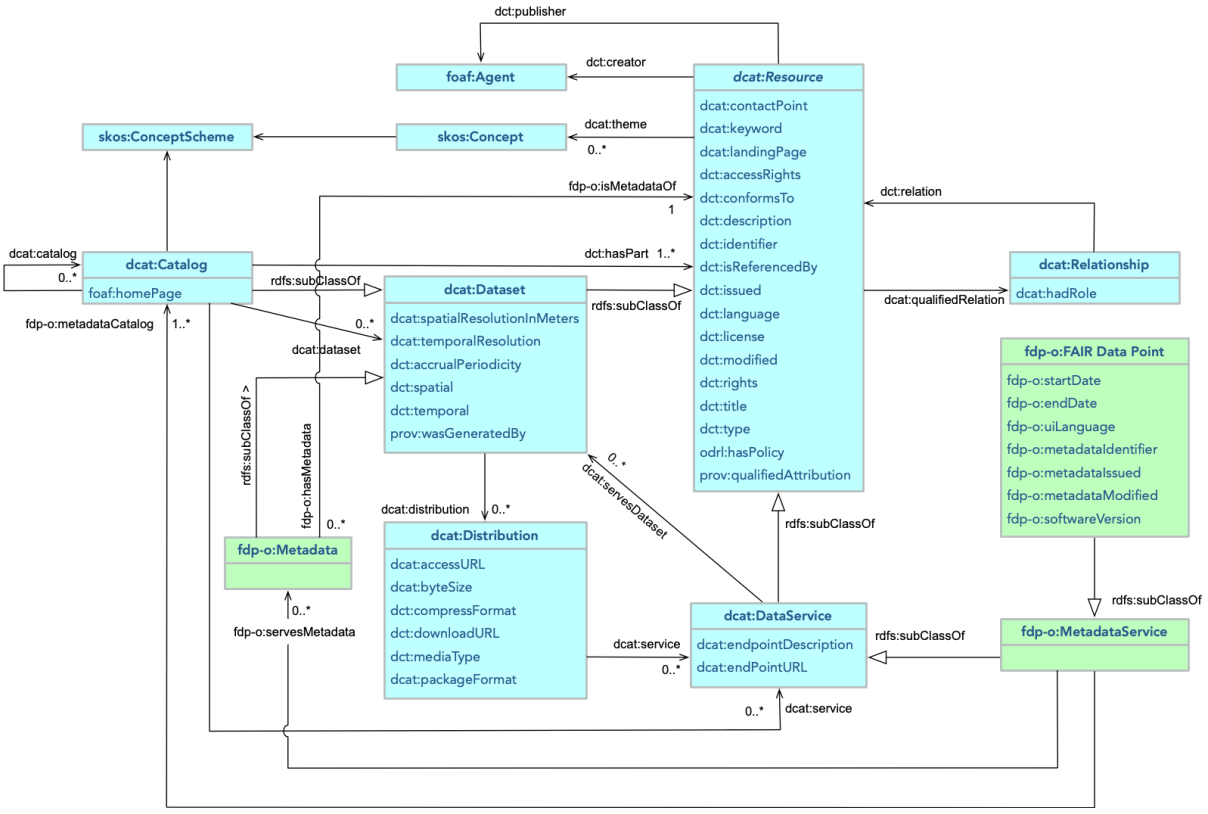


***Figure 1** - FDP Extensions (in green) to the DCAT Model (version 2; in blue). Source: https://specs.fairdatapoint.org/fdp-specs-v1.2.html#metadata*

The FDP uses the Shapes Constraint Language (SHACL) [R7] for defining conditions that new metadata must follow. These conditions, or metadata schemas, are applied at the DCAT Resource type level, i.e. there is a metadata schema for Catalogs, another for Datasets, and so on for Distributions and other user-added Resource types. The metadata schemas guarantee some level of uniformity in the metadata stored in the FDP and allow customisation of the structure of metadata that should be stored in the FDP. The metadata schemas should be referenced by the metadata records of the different types of Resources, so as to facilitate the discovery of the structure of the metadata model by machines and humans.

On a similar note, the navigation through the metadata must be self-discoverable, and it must only be necessary to know the URL of the root of the FDP to understand and discover the rest of the metadata tree. This is enforced by adding triples containing the Linked Data Platform predicates *ldp:contains* and *ldp:hasMemberRelation* to all metadata records that are linked to some other one(s).

The FDP reference implementation ensures that the requirements mentioned above are met. The reference implementation includes both an Application Programming Interface (API) and a Graphical User Interface (GUI) for metadata access and update. The API facilitates the automation of tasks related to metadata uploading and processing, while the GUI provides an intuitive interface for management tasks and to discover and access the metadata stored in the FDP. The FDP API includes a SPARQL endpoint that can be used to send generic SPARQL queries to the RDF metadata store. The SPARQL endpoint that is part of the GUI is more limited in its scope and allows sending pre-formatted queries where only a few fields can be altered by the user.

3. FDP Demonstrator for EHN

An FDP demonstrator for EHN has been deployed as part of D7.4. It is accessible at **https://fdpdemo.envri.eu**.

The procedure to deploy a new FDP is described in detail in [**R5**].

3.1. FDP GUI

The GUI is mostly suitable for tasks related to the management of the FDP, modifying the metadata schemas, adding the description of new variables, adding metadata about services or about a small number of datasets, fixing issues in a few metadata records, using the free-text search functionality of the FDP, etc. Appendix I of **D11.2** [**R8**] presents the GUI of the FDP demonstrator, including instructions on how to complete the tasks mentioned above.

3.1.1 FDP Implementation for ECVs

In the EHN FDP, the concept of Variable is defined as one of the types of Resources that are part of the metadata model. It is also a child resource of the FAIR Data Point Resource. As a result, the list of Variables that have been added to the FDP can be found under the "Variables" tab on the root page of the FDP in the GUI and is part of the metadata records of the FAIR Data Point. The metadata model used for Variables supports the use of I-ADOPT annotations. Appendix 1 shows the GUI form used to define a new Variable in the FDP.

It is expected that the metadata records of all datasets included in the FDP refer to the relevant ECV(s) using the *dcat:theme* predicate.

3.2. FDP API

The two most relevant use cases of the FDP API are the batch upload of metadata about a large number of datasets and the use of the SPARQL endpoint that allows generic SPARQL queries to be sent to the FDP. Note that a more detailed documentation of the API (and of other aspects of the FDP client) is available at [**R5**].

3.2.1 Batch upload of Metadata about Datasets

For the FDP to be used as a catalog of data hosting metadata about individual datasets, instead of services, it is necessary to have a mechanism that automates the upload of metadata, since it would be cumbersome to add a large number of metadata records using the GUI. The API allows adding a Dataset to a Catalog by POSTing the metadata of the Dataset, mapped to DCAT and serialized in Turtle, at the dataset endpoint of the FDP. This requires the authorization token obtained as described **here**. Using curl:

```
curl -H "Authorization: Bearer <token>" -H "Content-Type: text/turtle" -d @metadata.ttl https://fdpdemo.envri.eu/dataset
```

The URL of the Catalog to which the Dataset needs to be added must be provided in the metadata (contained in the metadata.ttl file in the curl example above). The Catalog must have been created beforehand, which can be done using the GUI. The newly created Catalog is

assigned an ID (for example, "*866ba6db-ad7b-46de-8d0e-d24abca87fc1*"). An example of a metadata record about an ICOS Dataset, mapped to DCAT and serialized in Turtle, ready to be POSTed to the dataset endpoint of the FDP (assuming that a Catalog with ID "*866ba6db-ad7b-46de-8d0e-d24abca87fc1*" exists) is provided here:

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
<> a dcat:Dataset;
 dct:isPartOf
<https://fdpdemo.envri.eu/catalog/866ba6db-ad7b-46de-8d0e-d24abca87fc1>;
 dct:hasVersion "1.0";
 dcat:version "1.0";
 dct:license <https://data.icos-cp.eu/licence>;
 dcat:contactPoint <https://www.icos-ri.eu>;
 dct:publisher <https://www.icos-cp.eu/>;
 dcat:landingPage
<https://meta.icos-cp.eu/objects/G6XQJolvHywDcMXoAZw7GZMK>;
 dcat:accessURL <https://meta.icos-cp.eu/objects/G6XQJolvHywDcMXoAZw7GZMK>;
 dcat:downloadURL
<https://data.icos-cp.eu/objects/G6XQJolvHywDcMXoAZw7GZMK>;
 dct:title "ICOS ATC CH4 Release, Birkenes (10.0 m)";
 dct:description "ICOS ATC CH4 Release, station Birkenes. Quality-controlled data on
molar fraction of atmospheric CH4";
 dct:issued "2025-07-01T08:22:15.816618482Z"^^xsd:dateTime;
 dct:modified "2025-07-01T08:22:13Z"^^xsd:dateTime;
 dcat:theme
<http://fdpdemo.envri.eu/variable/9150a404-40eb-41ae-b835-4077b989713f>;
 prov:wasAttributedTo <http://meta.icos-cp.eu/resources/stations/AS_BIR>;
 dct:startDate "2020-09-14T00:00:00Z"^^xsd:dateTime;
 dct:endDate "2025-03-31T23:00:00Z"^^xsd:dateTime;
 dcat:byteSize "774465"^^xsd:long;
 dcat:mediaType "zip" .
<https://www.icos-cp.eu/> a foaf:Agent;
 foaf:name "ICOS";
 dcat:landingPage <https://www.icos-cp.eu/> .
```

In the example above, note that the I-ADOPT annotated variable defined in the FDP and relevant for this dataset is referred to with the *dcat:theme* predicate.

Unfortunately, the FDP does not allow uploading metadata about several datasets from a single Turtle file containing metadata about all datasets. It is therefore necessary to POST one Turtle file, such as the one above, for each dataset.

Newly added Datasets are assigned an ID (for example, "79508287-a2a7-4ae2-95b3-3f595e3088cc") and are in a draft state. They need to be published to become visible to the users of the FDP. This is done with a PUT request, such as (using curl):

curl -X PUT -H "Authorization: Bearer <token>" -H "Accept: application/json" -H "Content-Type: application/json" -d '{ "current": "PUBLISHED" }' https://fdpdemo.envri.eu/dataset/79508287-a2a7-4ae2-95b3-3f595e3088cc/meta/state

Adding a Distribution to a Dataset requires the same procedure as for adding a Dataset to a Catalog, just replacing the concepts of Catalog and Dataset, with the concepts of Dataset and Distribution, respectively. The endpoint for Distributions is **https://fdpdemo.envri.eu/distribution**.

3.2.2 SPARQL Endpoint

The SPARQL endpoint that is accessible through the GUI of the FDP poses some restrictions on the type of queries that can be sent. To send a generic SPARQL query, the advanced SPARQL endpoint of the FDP can be used by POSTing the query at **https://fdpdemo.envri.eu/search/sparql**. The returned results are serialized in JSON. This SPARQL endpoint is not part of the FDP reference implementation and was added for the purpose of the EHN project.

The following HTTP header fields need to be set:

- Content-Type: text/plain
- Accept: application/json

For example, using curl:

curl -X 'POST' 'https://fdpdemo.envri.eu/search/sparql' -H 'Content-Type: text/plain' -H 'Accept: application/json' -d 'select * where {?s ?p ?o}'

or

curl -X 'POST' 'https://fdpdemo.envri.eu/search/sparql' -H 'Content-Type: text/plain' -H 'Accept: application/json' -d @query.rq

where "query.rq" is a file containing the SPARQL query.

4. Next Steps

4.1. Scalability and Onboarding more RIs

In the current deployment of the EHN FDP, most RIs have only published metadata about a few datasets or services in order to test the metadata upload process via the GUI. A few open questions and some limitations of the FDP reference implementation have been raised and preventing these RIs from uploading metadata about a larger number of datasets. The current issues and open questions include:

1. Differences between what constitutes a dataset from one RI to another (granularity issue);
2. A workaround is required to upload metadata, mapped to DCAT, about many datasets (some RIs have millions), as the FDP API does not allow it to be done in a convenient way;
3. Limitations in the free-text search of the FDP, which also do not always work as intended;
4. FDP being unresponsive when more than a few thousand metadata records are added.

All these issues and open questions should be addressed in WP8. This should allow more RIs to add metadata about a larger number of datasets to the FDP.

The first issue needs to be discussed in more detail directly with the RIs, where the granularity of metadata and data does not allow for conveniently isolating individual datasets and their metadata. A dataset could be considered as the data collected by a specific device over a specific time period, for example, but this type of granularity is not necessarily relevant for all RIs or is not represented in their metadata. Some devices are fixed, while others are moving, affecting the spatial representation of the datasets in the metadata. Some RIs group data points in individual datasets according to some spatial, temporal or other criteria and provide metadata about each individual dataset, while the data services of other RIs allow to retrieve all data points within a given spatial area and time period.

A solution that has been proposed to the granularity issue is to only publish metadata records about groups of datasets in the FDP. However, this could also be done in the CoS by creating dedicated services for each group of datasets if the total number of groups of datasets is low enough, and the FDP would not present much added value in this case. It would also limit what can be done with respect to the filtering of datasets based on spatial extent, period of time or other criteria, despite that it should be one of the main benefits of the FDP approach to storing metadata about datasets. As a result, a finer granularity should be aimed for. The FDP would become a parameterized service with an entry in the CoS (see **D7.1** for details about the CoS), which allows filtering and retrieving datasets according to a range of DCAT properties (including at least ECV, spatial and temporal filters). The API of the FDP could also be one of the data access methods of the VRE library, which is in development for the Hub (see **D7.2**). RIs who already offer services where metadata about individual datasets with adequate granularity are

mapped to DCAT might not need to implement their own FDP as long as the results of their services can be combined with those from the FDP.

The second issue is partially connected to the first one. It was solved for ICOS, and it remains to be seen how the solution that was found, which relies on the ICOS SPARQL endpoint and the granularity of ICOS' metadata, can be applied to other RIs. The requirement is that the metadata service of an RI allows automating the mapping of metadata about individual datasets to DCAT and serializing them in the Turtle format. One Turtle file is needed for each dataset. A discussion with individual RIs will take place regarding this issue.

The third issue is associated with the FDP reference implementation, whose maintenance and development have slowed down significantly since the end of 2023 (due to the ending of projects that were funding its development and staffing issues), although it partially restarted in 2025. The free-text search currently uses the more limited SPARQL endpoint, which is also available via the GUI. As a consequence, it only searches for terms matching the terms in the request in literal objects of triples whose subject's metadata includes a triple containing the predicate *dct:title*. Also, it can only return a maximum of 50 results for each search. Furthermore, those results are counted before being filtered for duplicates, which implies that fewer than 50 results can be returned even in cases when more than 50 distinct results were found. The available alternatives to address the limitations of the free-text search of the FDP will be evaluated.

The fourth issue needs to be investigated further to isolate its cause. The possible solutions will be evaluated once the issue is better defined.

4.2. Distributed FDPs

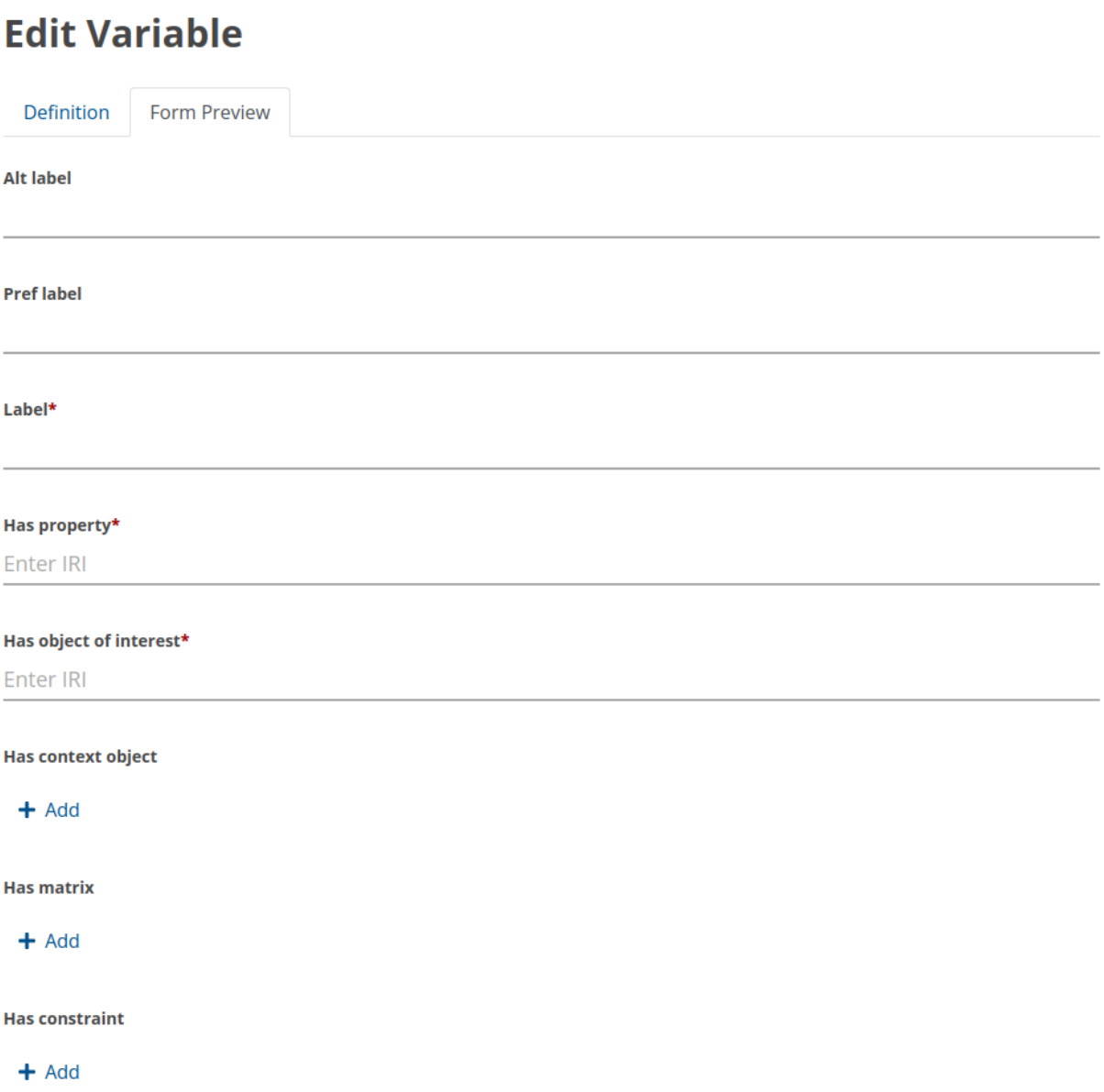
Once more RIs start using the FDP to publish metadata about their datasets, it will be useful to shift from the current situation where all metadata is stored in one central FDP to a distributed and federated system. In such a system, each RI would be responsible for maintaining an independent FDP containing metadata about their datasets, while a FDP containing the metadata about the ECVs would be used as a central node that can fetch metadata from the other FDPs. Defining the ECVs in the central FDP and referring to them in the metadata records in the other FDPs would guarantee that the same definitions of the ECVs are used across all RIs. It would also provide some flexibility and the possibility to customize the definitions of the ECVs according to the needs of the FDP users. The EXV table on the NERC Vocabulary Server (NVS) is an alternative to defining ECVs directly in the FDP. Using the EXV table on the NVS would make the source of information about ECVs more broadly compatible with external vocabularies and easier to reference outside the FDP.

In a distributed and federated FDP, the free-text search (once the issues mentioned earlier have been addressed) and the SPARQL endpoint of the central FDP would be the starting points for discovering metadata and individual datasets from all RIs by humans (free-text search and SPARQL endpoint for advanced users) and by machines (SPARQL endpoint).

5. References

References	
No	**Description/Link**
R1	https://specs.fairdatapoint.org/fdp-specs-v1.2.html
R2	https://i-adopt.github.io/ontology/index.html
R3	**FAIR Data Point: A FAIR-Oriented approach for metadata publication**. da Silva Santos, L.O.B., Burger, K., Kaliyaperumal, R., et al. Data Intelligence 5(1), 163-183 (2023). DOI: 10.1162/dint_a_00160
R4	**The FAIR Guiding Principles for scientific data management and stewardship**. Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. Scientific Data 3, 160018 (2016). DOI: 10.1038/sdata.2016.18
R5	https://docs.fairdatapoint.org/en/latest/index.html
R6	https://www.w3.org/TR/vocab-dcat-3/
R7	https://www.w3.org/TR/shacl/
R8	ENVRI-Hub NEXT_D11.2 Metadata and Vocabularies Harmonisation (**Under EC Review**). Dema, C., Fiebig, M., Shridhar, J., Vermeulen, A., Turco, A., Gutierrez, M., Thijsse, P., Bumberger, J., D'Amico, G., Ripepi, E., Izzi, F., & La Scaleia, G. (2025). Zenodo.

Appendix I: GUI form to define a new Variable in the FDP



Constrains

+ Add

Exact match

+ Add

Close match

+ Add

Title*

Enter a literal

Description

Enter a literal

Publisher*

Name*

Enter a literal

Version*

Enter a literal

Language

Enter IRI

License

Enter IRI

Rights

Enter IRI

ENVRI demo FDP · About