D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

EJP RD
European Joint Programme on Rare Diseases

H2020-SC1-2018-Single-Stage-RTD
SC1-BHC-04-2018
Rare Disease European Joint Programme Cofund

Grant agreement number 825575

Del 11.10
Fourth update Virtual platform of RD resources annotated with EJP ontological model

Organisation name of lead beneficiary for this deliverable:
Partner 01-b – INSERM-Orphanet

Contributors: AMC, BBMRI-ERIC, ELIXIR/EMBL-EBI, GUF, LUMC, UMCG, UPM

Due date of deliverable: month 48

Dissemination level:
Public

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

Table of Contents

1. Introduction

During the 4 years of the project, subsequent implementations of the EJP RD metadata model to capture descriptions of a growing number of resources included in the Virtual Platform (VP) as a network were produced, introducing adaptations based on the evolution of the metadata model itself, from a first version released in Y1 to a DCAT-based, evolutive version, released from Y2 on. Common models and methodologies to capture and make discoverable resources' description as first layer of discoverability were piloted from Y1, as a federated model based on a deployed "FAIR Data Point" at sources which feed a Virtual Platform FAIR Data Point Index. This main FAIR Data Point Index (VP Index) ensures a minimal gatekeeping point and serve services such as the Virtual Platform Portal (VP Portal https://vp.ejprarediseases.org/). The VP Portal allows end users to search the Virtual Platform network resources at once in real time to find those of interest.

In order to add functionalities to the Virtual Platform network and especially to the Virtual Portal, several webservices were progressively developed in order to implement the Orphanet nomenclature of rare diseases (ORPHAcodes) by the resources and by the human-readable interface (VP portal).

This deliverable presents the updates and upgrades performed in Y5 to implement the VP as a network based on the evolutions of the metadata model, the incorporation of new resources and the implementation of a machine-readable and human-readable federated ecosystem.

During year 5 (Y5), according to sustainability considerations and maintainability over the project, and based on existing community services such EJP RD GitHub, work foci (the Pillar 2 thematic working groups) also improved changes tracking and processing around ticketing system and decision-making process proposals, to reach an admitted overall governance methodology.

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

As technical assets further webservices were updated and upgraded to fit the needs of functionalities in the VP ecosystem and especially the human-readable VP Portal which have been through a deep refactoring process in order to release a completely new version. Beyond the update of the previous webservices provided by Orphanet (mapping services, classifications, genes to/from diseases, we also switched the previous query API mechanism to the Orphanet's catalogue of registries and biobanks to a new API Endpoint, based on the decision to use GA4GH beacon like framework to serve as an underlying API query mechanism.

Eventually, the VP Portal demonstrates the implementation of this mechanism in its backend. Moreover, the VP portal backend has been adapted to implement also new versions of the VP API specifications (https://github.com/ejp-rd-vp/vp-api-specs) and accordingly the front end has been adapted to reflect the new functionalities (search by genes, search by "group" of disorders...). A user interface working group was set up to design the new user interface. https://vp.ejprarediseases.org/

2. Approach

During Year 5 (Y5) and following the fifth Annual Retreat decisions (May 2023), a complete revised process to develop a new version of the VP Portal has been setup with a dedicated team.

Moreover, based on feedback by the "onboarding" teams and technical teams in charge of specifications, we have adapted the different available webservices. Resources which provide "Individuals" and "Catalogues" Endpoint and the proper metadata description of those services in compliance with the federated FAIR Data Point and DCAT modelisation are enlisted into the VP Portal.

2.1. The Virtual Platform Index and the FAIR-complaint approach for resource onboarding to the Virtual Platform

The FAIR Data Point (FDP) is a metadata service that, from one side, allows digital objects owners/publishers to expose the metadata of their digital objects following the FAIR principles and, for another side, allows digital objects' consumers to discover information (metadata) about the offered digital objects. Commonly, the FAIR Data Point is used to expose metadata of datasets but metadata of other types of digital objects can also be exposed such as biobanks, patient registries, tools, etc.

While providing metadata in a FAIR way contributes to making the described resources more findable, accessible, interoperable and reusable, if no one knows where these metadata are been offered does not help much. A fundamental part of the approach is to follow the FAIR principle F4 by "*registering or indexing the metadata in a searchable resource*". This indexing is the responsibility of the VP Index.

The VP Index works by being notified by willing FDPs of their existence and harvesting and indexing the available metadata from these metadata sources. We have adopted a distributed approach where the control and management of the metadata information remains at the source and the VP Index only keeps an updated index of these metadata records. In other words, the resource provider has total control over which information they would like to provide about their offered resources at any time.

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

Naturally, the resource provider can use their FAIR Data Point to advertise metadata about whatever resource they would like. To differentiate resources that should be available for general purposes to the ones that the resource provider would like to make available to the RD-VP, and to indicate the connection level available to the platform, we adopted a metadata-based approach. For this to work, the resource provider should provide:

- For connection level 1 (resource metadata discovery): Use the property vpConnection with the value VPDiscoverable as defined in the [EJP-RD ontology](#) and [metadata schema](#).

- For connection level 2 (resource content discovery): In the metadata of the Data Service (as defined in the DCAT vocabulary), use:
    o Property type (defined in the Dublin Core Terms) with the value of either VPBeacon2_Individuals or VPBeacon2_Catalog (defined in the EJP-RD ontology);
    o Property endpointURL (defined in DCAT) with the value the URL of the endpoint of the service API related to the type of service made available;
    o Property endpointDescription (defined in DCAT) with the value the URL of the documentation for the type of endpoint.

The information related to connection level 1 is required for all resources that are part of the RD-VP. This information is the explicit declaration from the part of the resource provider that they are offering that resource for the RD community.

The information related to connection level 2 is optional and should only be provided by the resource provider if they have adopted the Beacon 2 approach.

Applications such as the VP Portal make use of the metadata gathered by the VP Index. The Portal uses the connection level 1 information to provide a list of the resources available in the RD-VP. The connection level 2 information is used by the VP Portal to identify which resources offer Beacon 2 individuals or catalogue interfaces, and to know the URL of their API endpoints so that the portal can send them requests and display their responses to the portal users.

Originally, the FDP Index automatically accepted connections from any FAIR Data Point. This made the FDP Index an open service that is available to anyone without prejudice. However, during several discussions in EJP-RD Pillar 2, has been decided that a certain control over which resources should be visible in the VP Index. Because of this decision, a new feature has been implemented in the FDP Index software to be used in the RD-VP where any connection from new FAIR Data Points is directly put in a "waiting list" instead of being directly registered and indexed. A new user interface has also been implemented where a controller is able to check the new candidate resources and approve their inclusion as part of the RD-VP. The approval process is currently under elaboration by the governance teams and should include criteria such as relevance for the RD community, quality of the resource, among others.

2.2.   Integration of semantic resources into the EJP RD Virtual Platform

The first version (V1.0) of [The Virtual Platform Portal (VP Portal)](#) was further developed during Y5 and released in August 2023. The following marked the key advancements in the process to enhance the execution of queries and the display of relevant the results:

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

- The modification of the API (https://github.com/ejp-rd-vp/vp-api-specs) at the backend of the VP to ensure compliance with the new Beacon-2 framework specifications for Level2 discovery and querying of resources and of records; which started in Y4 was completed in the first quarter of Y5.

- The following functionalities were implemented using Orphanet services (**detailed information on each of the services is provided in section 3**):

    o Hierarchical search of rare diseases; enabling the end user to search groups of diseases and expand the search upward or downward in the hierarchy using the Classification API of Orphanet (see **Figure 1**)
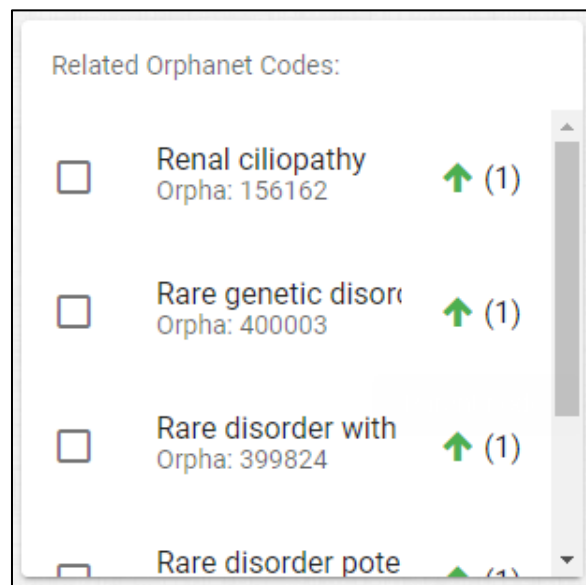

Figure 1. Display of expansion of search by hierarchy

    o Ability to search using other common disease codification systems: for Version 1 of the VP Portal, Orphanet mapping services will be used to enable search by ORPHAcodes, OMIM numbers, and ICD-10 codes.

    o Ability to search by gene functionality; which allows the end user to use a gene symbol in the search field resulting in a list of resources associated with that gene (specifically, associated with the diseases of ORPHAcodes mapped to those genes). The details of this gene-disease link are further described in the section 3.3.

- Connecting the Virtual Platform to the VP Index described in section 2.1; which enabled the resources that are indexed to be dynamically discoverable via the VP Portal at level 1 (resources discovery through their metadata descriptors). The connection was established in December 2022 and will undergo further testing and minor interface improvements to ensure semantically-enhanced display of discoverable resources on the landing page of the Portal (**see figure 2**).
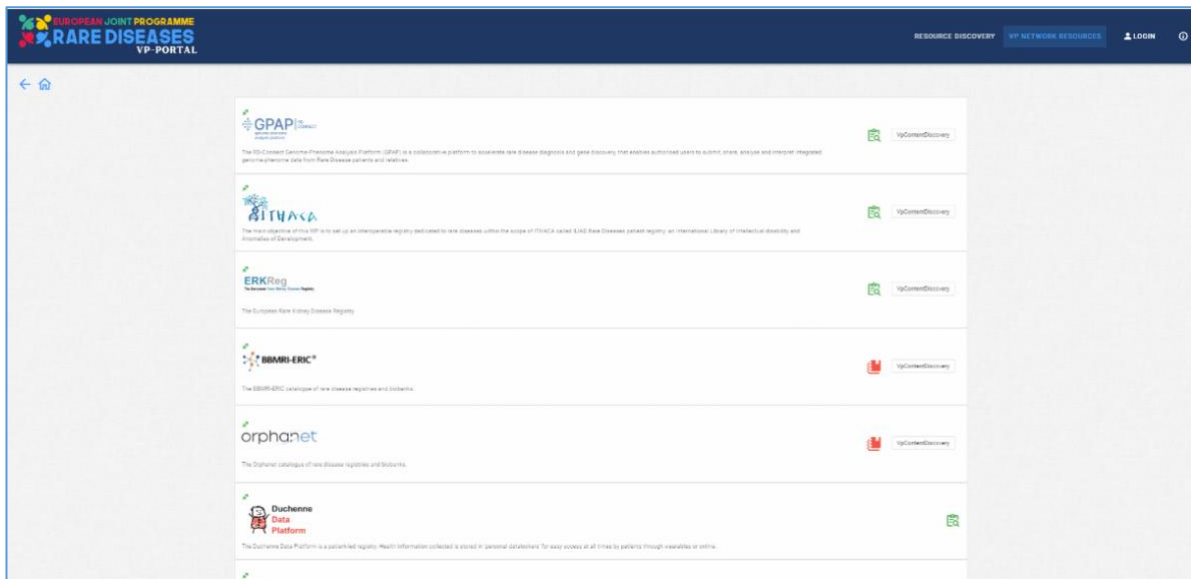
D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

Figure 2. Display in the VP Portal of Resources Connected to the VP via the VP Index
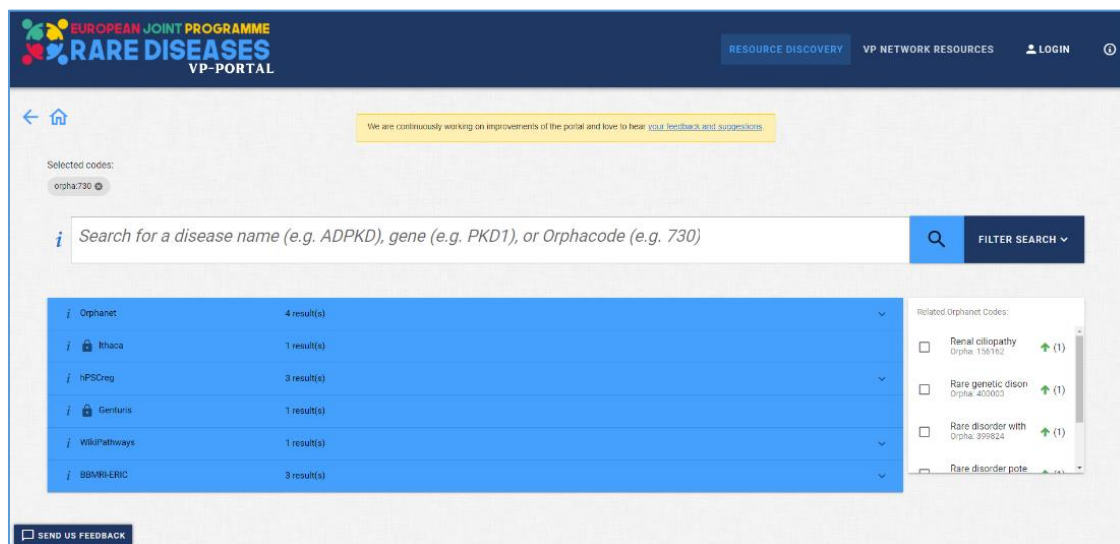

Figure 3. Search results display

- Throughout the execution of the functionalities, the Metadata and Data Modelling Work Focus closely collaborates with the Overall Architecture and L1/L2 Implementation teams to ensure that the resulting design of the user interface accurately reflects the implementation of semantic approaches and the use of existing interoperable tools and services for the enhancement of the VP network of resources as a whole. A workflow on GitHub enabled the teams to raise issues as well as to alert resources and developers on updates in the query facilities or in the metadata schema. The workflow was reviewed and endorsed prior to the VP Governance Workshop hosted in Amsterdam in November 2023.

- Two user requirements analysis workshops for the EJP RD Virtual Platform were conducted in November and December 2023 by the front-end development team, inviting focus groups of end-users to discuss the context of use for various persona.

- The results of the workshops will be used in Y6 to enhance the visualization of filters for L2-connected resources and the overall usability of the VP Portal. **Figure 3** shows the current display of results at the front-end for reference.

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

2.3.   Onboarding of resources to the Virtual Platform

During Y5 the onboarding work focus worked on streamlining the process of onboarding of rare disease resources to the Virtual Platform at Level 1 and Level 2 respectively. A team of FAIRification data stewards, in collaboration with the metadata and data modelling work focus, developed a thorough onboarding guidance document[1], which was disseminated to all the EJP RD partner resources. The documentation contains introductory information about the EJP RD Virtual Platform, its purpose and its functionalities. It additionally includes clear instructions on requirements and processes of onboarding using the various offered tools. It additionally provides an updated thorough explanation of the EJP RD Metadata Schema.

The onboarding team additionally continued the practical support to onboard all the resources that were scheduled to connect to the Virtual Platform for Y5. This resulted in onboarding a total of 22 resources at Level 1. The work continued to ensure that the entirety of the onboarded resources are in complete alignment with the updated metadata schema (refer to D 11.05 for detailed schema enhancements in Y5). The onboarding team additionally worked on the transition from the VP Portal hard-coded onboarding to the live indexing following the integration of the VP Index with the Portal (refer to section 2.1). The process will continue for the first half of Y6 to enable all resources to be dynamically indexed. Furthermore, 8 resources were onboarded at Level 2, with 4 of them containing real data.

Table (1) lists all the resources that are now onboarded into the Virtual Platform.

Table 1. Status of resource onboarding on 31/12/2023

Resource name	Level 1 Connection (VP Index & Hard-coded in portal)	Level 2 Connection
BBMRI-ERIC	✓	✓
RD-Connect GPAP	✓	
hPSCReg	✓	
INFRAFRONTIER	✓	
CTSR	✓	✓
Duchenne Data Platform	✓	
FAIRVASC	✓	
ERN ITHACA	✓	✓
Registry ERN EURO-NMD	✓	
ERN GENTURIS	✓	✓
Registry ERN ERKNet (ERKReg)	✓	✓
Metabolights Catalogue	✓	
ERN Skin	✓	✓
Orphanet Catalogue of Biobanks and Registries	✓	✓

---

[1] EJP RD Virtual Platform Resources Onboarding Manual v1.0. Available from https://www.ejprarediseases.org/wp-content/uploads/2024/02/EJP-RD-Virtual-Platform-Resources-onboarding-manual_v1.0.pdf.

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

Resource name	Level 1 Connection (VP Index & Hard-coded in portal)	Level 2 Connection
WikiPathways	✓	✓
ERN CRANIO	✓	
EGA	✓	
ERN VASCREN	✓	
ERN VASCERN – HHT (ciroco)	✓	
Registry ERN-BOND (EuRR-Bone) and Endo-ERN (EuRRECa) - E-REC Registry	✓	
Registry ERN EURACAN (STARTER)	✓	
Registry ERN eUROGEN	✓	

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

3. ORPHANET webservices and Endpoint

3.1.   Orphanet Mapping services

Rare diseases resources could use different medical terminologies, depending on local constrains or usage. Therefore, a mapping service is needed when querying several resources at once.

Orphanet provides mappings between the Orphanet nomenclature (ORPHAcodes) to other terminologies, including OMIM, ICD-10, MeSH, MedDRA, UMLS. These mappings are released as downloadable files in Orphadata (www.orphadata.com) and are integrated in ORDO (Orphanet Rare Diseases Ontology) twice a year. The mappings are expert-reviewed, using semi-automated mapping tools and are validated by the Orphanet curation team. Each mapping is assessed and identified as an Exact match or not (including Narrower-to-Broader and Broader-to-Narrower) with respect to the respective ORPHAcode(s).

The alignments specify the comparability between terminologies by defining if the concepts are perfectly equivalent (exact mapping) or not.

To ease the implementation of the mapping services within the Virtual Platform network, without having the necessity of using semantic based technologies with the Orphanet ontology , a dedicated webservice has been developed during year 3. The delivered content has been updated during year 4 and year 5.

Description of the service is available on GitHub: https://github.com/ejp-rd-vp/Orphanet-Mapping-API. Useful links to services are provided in table 2 below.

Table 2. Orphanet Mapping and API Parameters

Element	Parameters & Link(s)	Details
API parameters	/mapping/: get: description: "Get information about the API."	Get information about the API
	/mapping?form={origin}&code={code}&to={destination} will return results for a clinical entity (resources), from any terminology listed to another (orphanet, omim, umls, mesh, meddra, icd). Based on Orphanet mappings. "	Get mappings with a query for a clinical entity (resources). From any terminology listed to another (orphanet, omim, umls, mesh, meddra, icd). Based on Orphanet's mappings.
	/mapping?form={/mapping?form={origin}&code={code}&to={destination}	will return results for a clinical entity (resources), from any terminology listed to another (orphanet, omim, umls, mesh, meddra, icd). Based on Orphanet's mappings.
Querying using a different coding system	http://purl.org/orphanetws/mapping?from=icd&code=Q87.4 http://purl.org/orphanetws/mapping?from=omim&code=154700	The webservice is queryable by using different coding systems.

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

Element	Parameters & Link(s)	Details
Parallel mapping to more than one coding system at once using the API mapping webservice	http://purl.org/orphanetws/mapping?from=orphanet&code=558&to=omim&to=icd	The API could be used to obtain codes at once

Described in previous Deliverable 1.9, this service has been updated during year 5. Furthermore, in order to ease the next updates, after the end of the project, we have setup an automated process to extract the datasets from the Orphanet's knowledge base and generate the webservice, using "Docker" based scripting.

3.2.   Orphanet Hierarchies service

This Service is dedicated to support implementation and usage of Orphanet's classifications. The API focuses on Rare Diseases classifications based on ORDO (Orphanet Rare Diseases Ontology).

Orphanet clinical entities (groups of disorders, disorders or subtypes of disorder), for which an ORPHAcode exists, are classified in one to several medical domains (classifications) (multi-hierarchical organization), and is one to several groups within a classification (poly-parental organization).

This service help to navigate from an ORPHAcode through the levels in the classifications' trees.

This service is described in the according GitHub repository : https://github.com/ejp-rd-vp/orphanet_hierarchies_service

Described in previous Deliverable 1.9, this service has been updated during year 5. Furthermore, in order to ease the next updates, after the end of the project, we have setup an automated process to extract the datasets from the Orphanet's knowledge base and generate the webservice, using "Docker" based scripting.

3.3.   Orphanet Genes/Diseases Mapper API

This API utilises the Orphanet-curated links between genes and diseases, which offer insight on the nature of association of these genes with their respective rare diseases.
The service is described in GitHub (https://github.com/ejp-rd-vp/Orphanet-GenesDisease-Mapper-API)

Fully described in deliverable 10.9, this service has been updated during Y5. Furthermore, in order to ease the next updates, after the end of the project, we have setup an automated process to extract the datasets from the Orphanet's knowledge base and generate the webservice, using "Docker" based scripting.

3.4.   Orphanet Registries & Biobanks API

The service was initially developed following the specifications of Query Builder previous version (https://github.com/ejp-rd-vp/query_builder_api) but has been reshaped during Y5 accordingly to the updated Beacon-2 framework-based API specification adopted by the

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

VP . Indeed, during Y4, a new set of API query mechanisms has been set up as REST webservices, in compliance with the GA4GH Beacon 2 specifications.

The request and response conform to the Beacon Reference Framework. This Specification defines two types of endpoints -   Informational Endpoints and Query Endpoints.
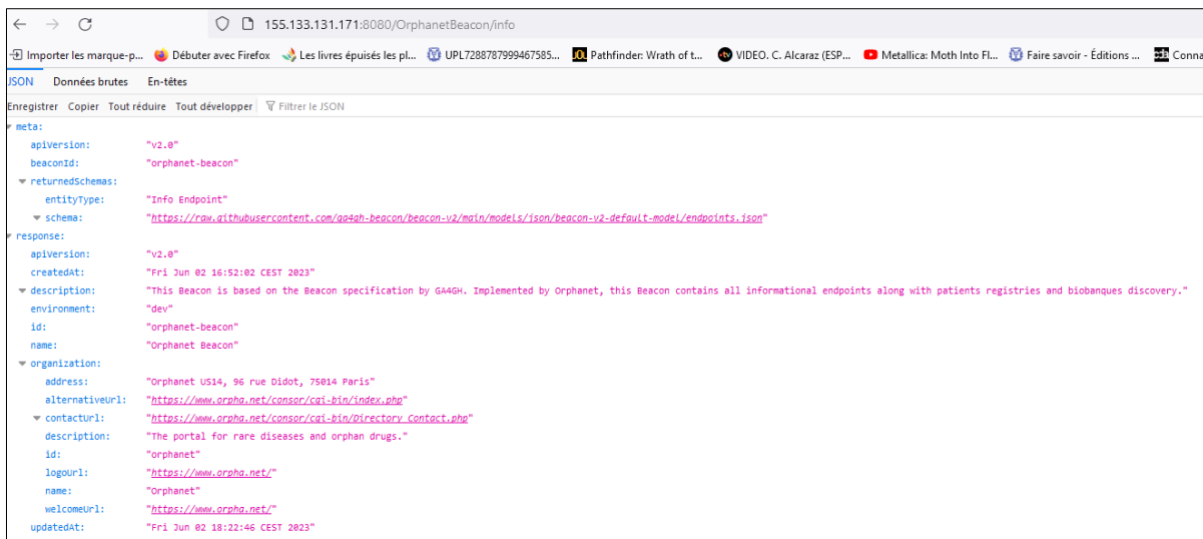
Informational Endpoints allow simple GET requests without needing a request body and respond with information relevant to this Beacon Specification. These are: /info, /configuration, /entry_types, /filtering_terms and /map. A special /service-info endpoint (also a GET request), responds with metadata about this service using the GA4GH Service_Info format.

Unlike the Informational Endpoints, the Query Endpoints require the requester to provide a JSON body and send request using the POST method. This document defines three query endpoints to query resources using filters - /individuals, /biosamples and /catalogues. The response is a Beacon Collection response that corresponds to a Resource described by the EJP RD VP Resource Metadata Schema.

As an EJP RD partner resource, Orphanet has developed its own catalogue endpoint, following the specifications, and accordingly its implementation through the VP Portal.

The current schema is available: https://raw.githubusercontent.com/ga4gh-beacon/beacon-v2/main/models/json/beacon-v2-default-model/endpoints.json

The current Orphanet implementation is of version 2 "V2.0"

Information on the Endpoint is available via a series of REST calls listed below, along with a screen capture of the respective information they return:

- http://155.133.131.171:8080/OrphanetBeacon/info
  - Provides information about the Endpoint and its schema

- 155.133.131.171:8080/OrphanetBeacon/service-info

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

o Provides basic information about the Endpoint

- 155.133.131.171:8080/OrphanetBeacon/configuration
  o Provides information about the Endpoint, especially schema, entity type and the compliance to Data Catalog Vocabulary, dcat-v2, https://www.w3.org/TR/vocab-dcat-2/

```
▼ meta:
    apiVersion:               "v2.0"
    beaconId:                 "orphanet-beacon"
    ▼ returnedSchemas:
        ▼ 0:
            entityType:       "catalogs"
            ▼ schema:         "https://github.com/ejp-rd-vp/resource-metadata-schema/blob/master/shapes/catalog.shex"
▼ response:
    ▼ $schema:                "https://raw.githubusercontent.com/ga4gh-beacon/beacon-v2/main/framework/json/configuration/beaconConfigurationSchema.json"
    ▼ entryTypes:
        ▼ catalogs:
            ▼ defaultSchema:
                id:           "dcat-v2"
                name:         "Data Catalog Vocabulary"
                referenceToSchemaDefinition:  "https://www.w3.org/TR/vocab-dcat-2/"
            id:               "Catalogs"
            name:             "Catalogs"
            ▼ ontologyTermForThisType:
                id:           "dcat:Catalog"
            partOfSpecification:  "v2.0"
    ▼ maturityAttributes:
        productionStatus:     "DEV"
    ▼ securityAttributes:
        defaultGranularity:   "count"
```

- http://155.133.131.171:8080/OrphanetBeacon/entry_types
  o Provides information about the expected return schema https://github.com/ejp-rd-vp/resource-metadata-schema/blob/master/shapes/catalog.shex

- http://155.133.131.171:8080/OrphanetBeacon/filtering_terms
  o Provides information about the filter compliance (which filters are allowed) within this particular Endpoint

```
▼ meta:
    apiVersion:       "v2.0"
    beaconId:         "orphanet-beacon"
    ▼ returnedSchemas:
        ▼ 0:
            entityType:   "Catalog"
            ▼ schema:     "https://github.com/ejp-rd-vp/resource-metadata-schema/blob/master/shapes/catalog.shex"
▼ response:
    ▼ filteringTerms:
        ▼ 0:
            ▼ id:         "A single value or an array of any of the following values: {PatientRegistryDataset,BiobankDataset}."
            label:        "Resource Types"
            type:         "alphanumeric"
        ▼ 1:
            ▼ id:         "A single value or an array of any of the ISO 3166-1 alpha-2 format string of countries (e.g. IT, FR, NL)."
            label:        "Country"
            type:         "alphanumeric"
        ▼ 2:
            id:           "A single value or an array of orphanet terms."
            label:        "Disease or disorder"
            type:         "ontology"
```

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

- http://155.133.131.171:8080/OrphanetBeacon/map
  - Provides information about retrieving schema and filtering options

```
meta:
    apiVersion:          "v2.0"
    beaconId:            "orphanet-beacon"
    returnedSchemas:
        0:
            entityType:  "catalogs"
            schema:      "https://github.com/ejp-rd-vp/resource-metadata-schema/blob/master/shapes/catalog.shex"
response:
    $schema:             "https://raw.githubusercontent.com/ga4gh-beacon/beacon-v2/main/framework/json/configuration/beaconMapSchema.json"
    endpointSets:
        catalogs:
            entryType:       "Catalogs"
            filteringTermsUrl: "http://155.133.131.171:8080/OrphanetBeacon/filtering_terms"
            rootUrl:         "http://155.133.131.171:8080/OrphanetBeacon/catalogs"
```

The Endpoint is queryable from http://155.133.131.171:8080/OrphanetBeacon/catalogs

- Note that this address is not responding directly in a web browser, as the specification only accept "POST" methods. Accordingly, services such as "Postman" could be used to test it

4. Discussion and next steps

As a whole "technical ecosystem" the Virtual Platform relies on both federated approaches (FAIR Data Points (FDPs), common API mechanisms) and minimal cantered services (VP Index as main FDP, in order to allow for gatekeeping possibilities and curation process at this level). Those approaches are actionable through different services and implementations process (the direct use of federated FDP, the direct use of API mechanisms) and the VP Portal itself demonstrates a controlled way to access and query the resources.

Resources can control the way they expose their metadata, being both discoverable at "high" level or giving more insight by the implementation of more detailed query methods such as Individuals or Catalogues endpoint. Almost all the EJP RD partner resources are now onboarded at level 1, the last ones being completed in year 6, and a growing number of them are onboarded at level 2 of connexion, which is not mandatory.

The main starting point of the ecosystem relies on a proper "DCAT" (Data Catalog Vocabulary) description. DCAT plays an important role not only as a W3C standard but also as a discoverability standard within the future European Health Data Space. The HealthData@EU pilot project, which started in the final quarter of 2022, has been working on a proposed DCAT-AP Health extension. An initial draft was presented in December 2023. Following a review process, the model will be fine-tuned. As a Rare Disease Community, we consider the choice made during the previous 5 years of the project to setup the first layer of metadata description as a DCAT modelisation as relevant to this progress. Our modelling team anticipates the evolution of the DCAT-AP health extension and its implementation within the future European Health Data Space, and we are ensuring that our model remains aligned with major initiatives within the European ecosystem of health data.

To address the challenges, the VP governance should take consideration the impact of any future changes. Complexity of implementation, constraints, compatibility issues, in particular

D11.10
Fourth update Virtual platform of RD resources
annotated with EJP ontological model

with webservices depending on it are to be considered, as should strategic alignment with other data spaces and/or the ability to onboard new types of resources in the future.

With this context, during the last year of the project, we started to design a governance process and setup a way to firstly catch and trace the updates demands through a ticketing system (EJP RD Ticketing GitHub).

The whole VP governance document will be finalized and published during year 6.