



Deliverable 5.3
OD4RD Dedicated APIs
January 2023

OD4RD

Orphanet Data For Rare Diseases



**Co-funded by
the European Union**

Co-Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Content summary

Context	3
Objective	3
Methodology	3
Implementation and tools	5
Datavisualisation service	5
Nomenclature Pack API	7
ORPHADATA WEBSITE	8
ORPHADATA API	10
ORPHADATA FAIRDATAPOINT	11
API MANAGEMENT GATEWAY	12
Discussion	14

OD4RD
Orphanet Data For Rare Diseases

Context

As Rare Diseases (RD) are poorly represented in existing medical terminologies in use such as ICD10, the RD field across Europe benefits from the knowledge and dedicated domain-specific Orphanet nomenclature of rare diseases (ORPHAcodes). This nomenclature and classifications system was developed and is continuously maintained thanks to European support after the recognition as a priority, in the Council Recommendation of 8 June 2009 on an action in the field of rare diseases, of the improvement in codification of rare diseases. Moreover, the Orphanet Rare Diseases nomenclature is aligned with several non-RD specific terminologies, allowing semantic interoperability in a context of large heterogeneity of coding systems in Europe. Therefore, in the context of OD4RD, the WP5 aims, across tasks 5.2 and 5.3, to provide technical support, implementation and tools dedicated to ease the process of the ORPHAcodes integration within health IT systems, in a standardised way. This deliverable describes the provided tools and solutions.

Objective

The main objectives addressed by WP5 are :

- To put in place the necessary mechanisms to deliver the Orphanet nomenclature and knowledge base and its updates (as reference body of data).
- To develop the exchange workflow and tools allowing access to the nomenclature, classifications and alignments, with the possibility to include other relevant data according to end-user's needs (could also include disease-phenotype relationships (using HPO), gene-disease relationships, and other data necessary to support datasets annotations and usages)
- To develop dedicated APIs (Application Programme Interface), following Open API standard specifications and ensure the setup of the technical infrastructure. (subtask 5.3.1)
- To permit the management and control access to the several APIs through an API Gateway (API management) based on Azure solution. Define the "client" typology, threshold and "API Keys" management system. (subtask 5.3.2)

Methodology

During the previous project RD-CODE (826607 RD-CODE), web-based applications development by Orphanet was foreseen. The aim was to facilitate the informatic access to Orphanet nomenclature data and allow flexible implementation into the various IT systems in use in the different countries and/or settings.

An API (<https://api.orphacode.org/>) was firstly released as a beta version in December 2019 (API for visualisation of Orphanet Nomenclature – RD-CODE Deliverable 4.2) and then the official release and upgrade was made in June 2020 (Facility for ORPHA nomenclature correspondence - RD-CODE Deliverable 4.5) . The Orphanet Data visualisation tool was delivered in December 2020. (<https://dataviz.orphacode.org/>) based on the API.

Based on this previous work, we have proceeded to major updates both on the content and IT infrastructure.

The “Nomenclature pack” was updated in July 2022 and therefore the API has been upgraded to reflect the July 2022 content (including mappings to ICD 11 as a new feature).

A new set of APIs has been developed to ease the access to Orphanet knowledge base (<https://api.orphadata.com>)

An API management gateway was setup, allowing access keys generation and thresholds configuration.

A new version of the Orphadata website have been made available. This new website (<https://www.orphadata.com>) allows access both to the “coding” and “scientific content” of the Orphanet’s knowledge base but also to the other services proposed (including abovementioned APIs).

A FAIR (Findable Accessible Interoperable & Reusable) datapoint layer above Orphadata has been setup (<https://fairdatapoint.orphadata.com>) which ensures machine readability and machine accessibility to the content.

Moreover, a set of new tools, based on the APIs, are proposed to ease the use of the ORPHAcodes.

The IT architecture of the infrastructure relies on an extraction workflow from the Orphanet main knowledge base to an ElasticSearch cloud platform. The ElasticSearch cloud platform serves the datasets to the different APIs and grant access through the API management gateway to stakeholders who are using the APIs for their own purposes. The tools provided are using also the dedicated APIs. (fig.1)

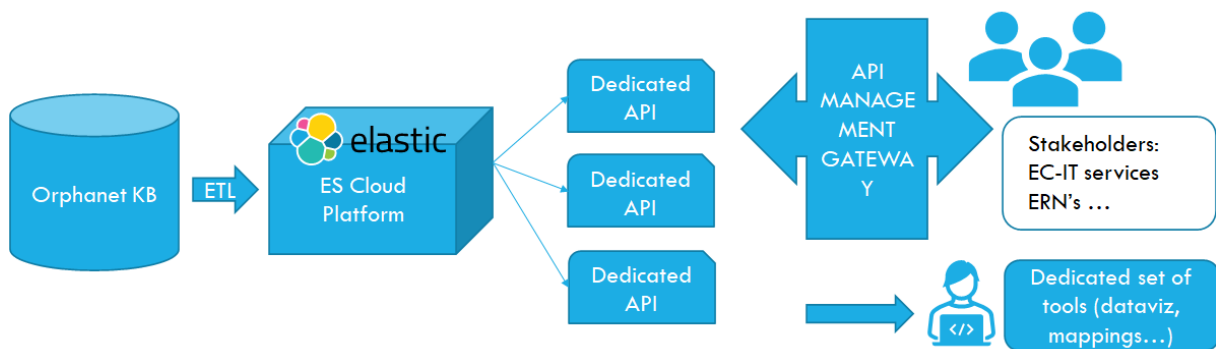


Figure 1: the data workflow from the Orphanet knowledge base to the different services and implementation.

The ElasticSearch cloud platform is managed by the editor of the ElasticSearch solution. The hosting service is AWS (Amazon Web Services), the data center is based in Europe (Ireland).

(N.B.: no data extracted from Orphanet knowledge base requires GDPR compliance, as there is no personal data neither patient information).

The APIs and tools instances are hosted in France, by a third-party service (GANDI).

The API management gateway is provided by Microsoft Azure (Gateway), hosting services based in Europe.

Implementation and tools

Datavisualisation service

Once the content of the “Nomenclature Pack” was released in July 2022, an upgrade of the “data visualisation tools” has been performed.

The code was updated to Github (code management tool) (fig.2)

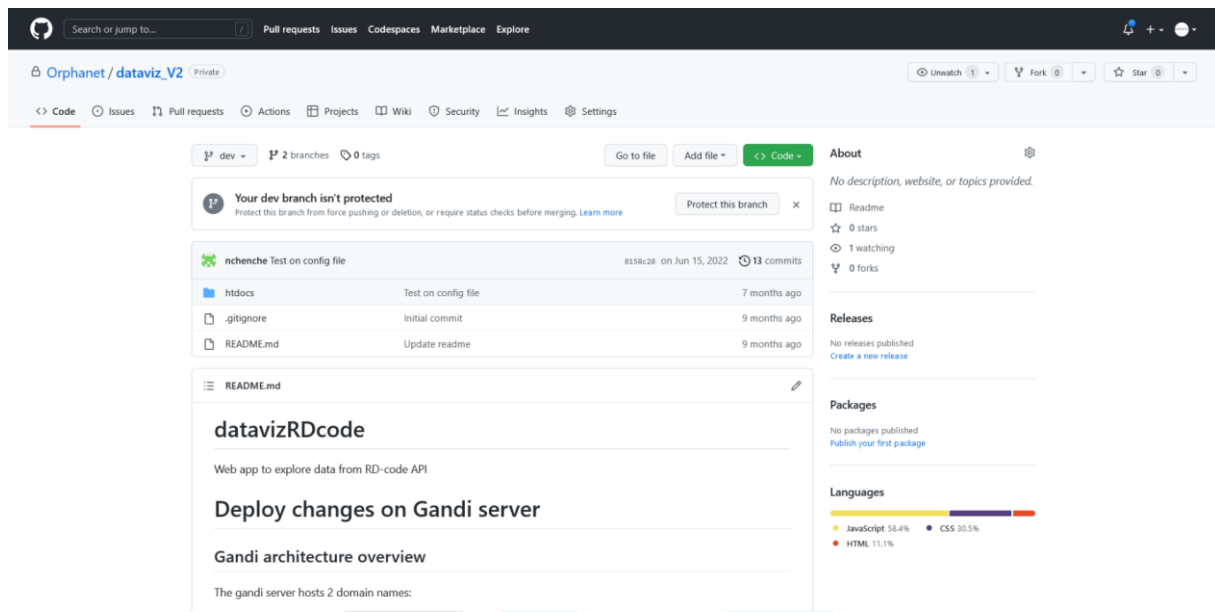


Figure 2: Github repository for dataviz. https://github.com/Orphanet/dataviz_V2 (the repository is in “private” mode and therefore may be not accessible without proper credentials).

And deployed online : <https://dataviz.orphacode.org> (fig.3)

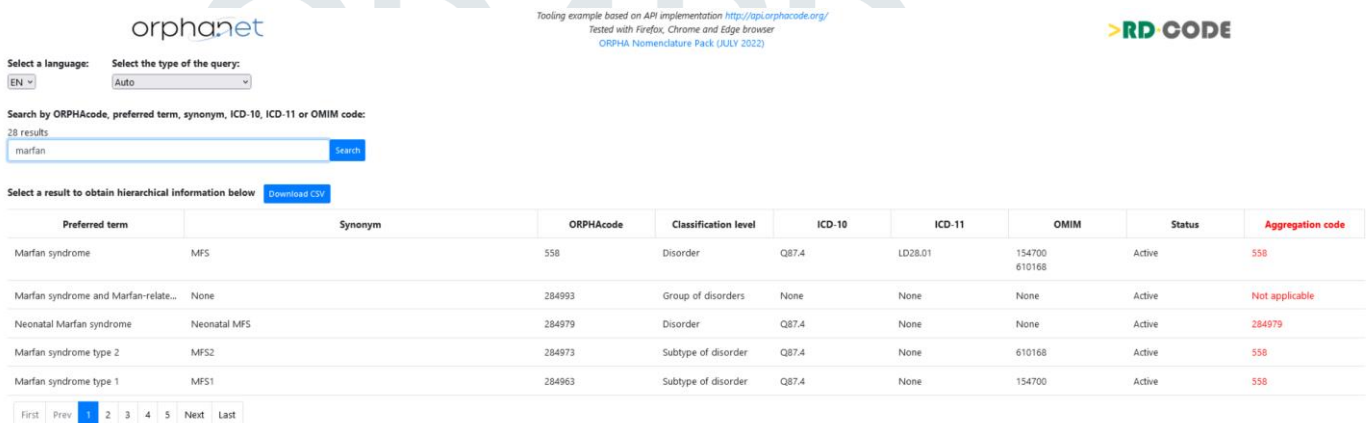


Figure 3: the deployed version of the Data visualisation tool. Including ICD 11.

This tool has been used by around 1,400 unique users (3,357 viewers) during 2022 (fig.4a).

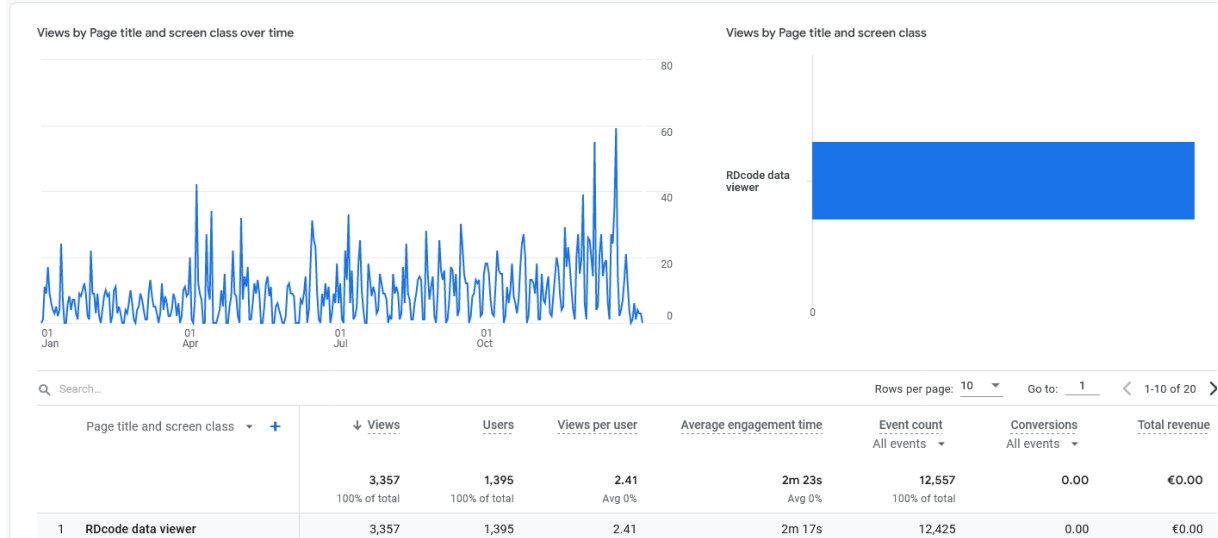
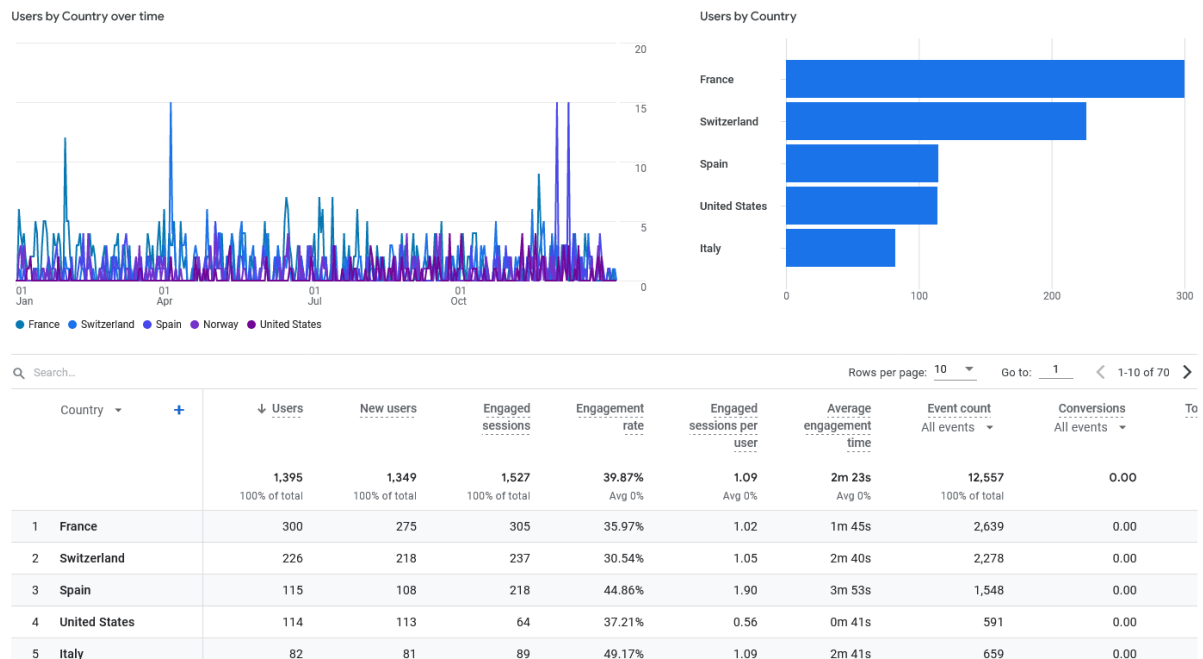


Figure 4a: Data visualisation tool usage during year 2022.

The tool was used in several countries, in Europe and worldwide (70 countries listed) (fig.4b).



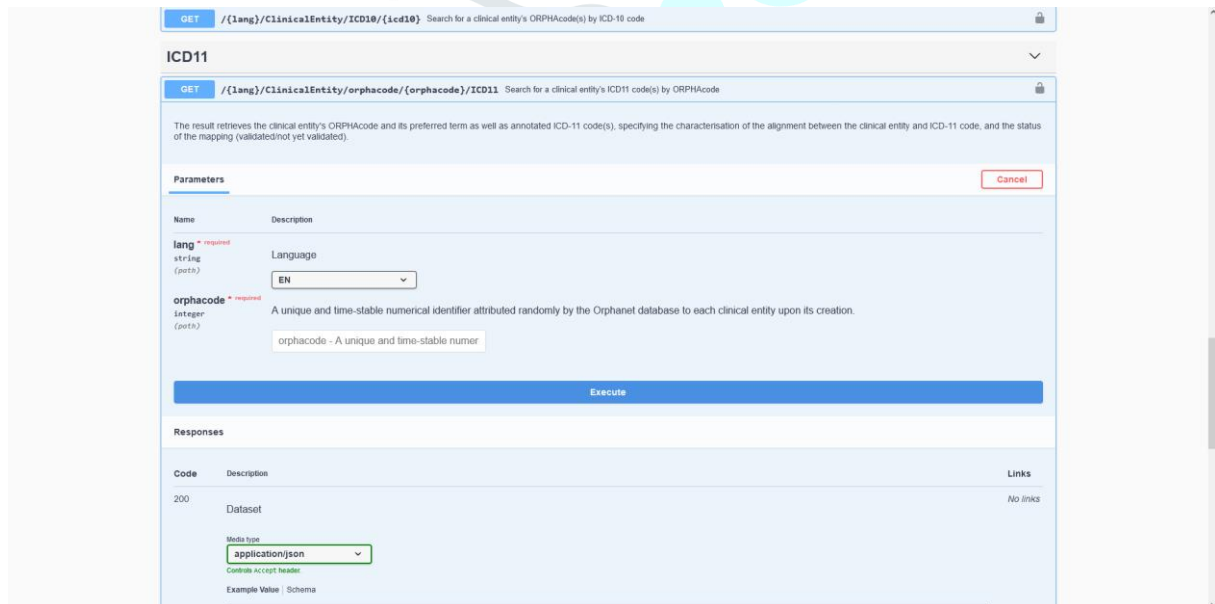
Search... Rows per page: 100 1-70 of 70

Country	↓ Users	New users	Engaged sessions	Engagement rate	Engaged sessions per user	Average engagement time	Event count All events	Conversions All events
	1,395 100% of total	1,349 100% of total	1,527 100% of total	39.87% Avg 0%	1.09 Avg 0%	2m 23s Avg 0%	12,557 100% of total	0.00
1 France	300	275	305	35.97%	1.02	1m 45s	2,639	0.00
2 Switzerland	226	218	237	30.54%	1.05	2m 40s	2,278	0.00
3 Spain	115	108	218	44.86%	1.90	3m 53s	1,548	0.00
4 United States	114	113	64	37.21%	0.56	0m 41s	591	0.00
5 Italy	82	81	89	49.17%	1.09	2m 41s	659	0.00
6 Germany	77	75	72	48.32%	0.94	2m 02s	546	0.00
7 China	66	66	25	29.76%	0.38	0m 21s	285	0.00
8 Netherlands	59	57	57	31.15%	0.97	3m 03s	528	0.00
9 Norway	50	46	153	48.26%	3.06	10m 10s	1,035	0.00
10 Belgium	41	41	52	34.21%	1.27	2m 07s	471	0.00
11 Czechia	22	21	19	55.88%	0.86	1m 34s	117	0.00
12 Sweden	20	20	32	38.55%	1.60	2m 14s	251	0.00
13 United Kingdom	19	17	23	74.19%	1.21	1m 51s	129	0.00
14 Austria	17	16	13	48.15%	0.76	1m 51s	101	0.00
15 Finland	15	15	3	20%	0.20	0m 18s	53	0.00

Figure 4b: Data visualisation tool usage by countries during year 2022

Nomenclature Pack API

The Data visualisation tool is based on the API released also in July 2022, updated with the nomenclature pack content. (<https://api.orphacode.org/>) (fig.5).



The screenshot shows the Orphacode API interface. At the top, there is a search bar with the endpoint `/{lang}/ClinicalEntity/ICD10/{icd10}`. Below it, the **ICD11** section is expanded, showing the endpoint `/{lang}/ClinicalEntity/orphacode/{orphacode}/ICD11`. A description states: "The result retrieves the clinical entity's ORPHAcode and its preferred term as well as annotated ICD-11 code(s), specifying the characterisation of the alignment between the clinical entity and ICD-11 code, and the status of the mapping (validated/not yet validated)." The **Parameters** section includes:

- lang** (required, string): Language, with a dropdown menu set to "EN".
- orphacode** (required, integer): A unique and time-stable numerical identifier attributed randomly by the Orphanet database to each clinical entity upon its creation. The input field contains "orphacode - A unique and time-stable numer".

 An **Execute** button is visible. The **Responses** section shows a table with columns for Code, Description, and Links. The response code is 200, with a description of "Dataset" and "No links". The media type is set to `application/json`.

Figure 5: Release of the API dedicated to the nomenclature pack, including ICD11 alignments.

The API is still freely available and accessible, under the licence Commons Attribution 4.0 International (CC BY 4.0).

In addition to the web interface availability, which is used to ease the documentation readability, an API (Application Programming Interface) aims to be implemented by stakeholders in their own IT system or software, and not through the webpage. The network monitoring provided by the hosted service is not perfectly precise, nevertheless it shows the access statistics to the API (fig.6).

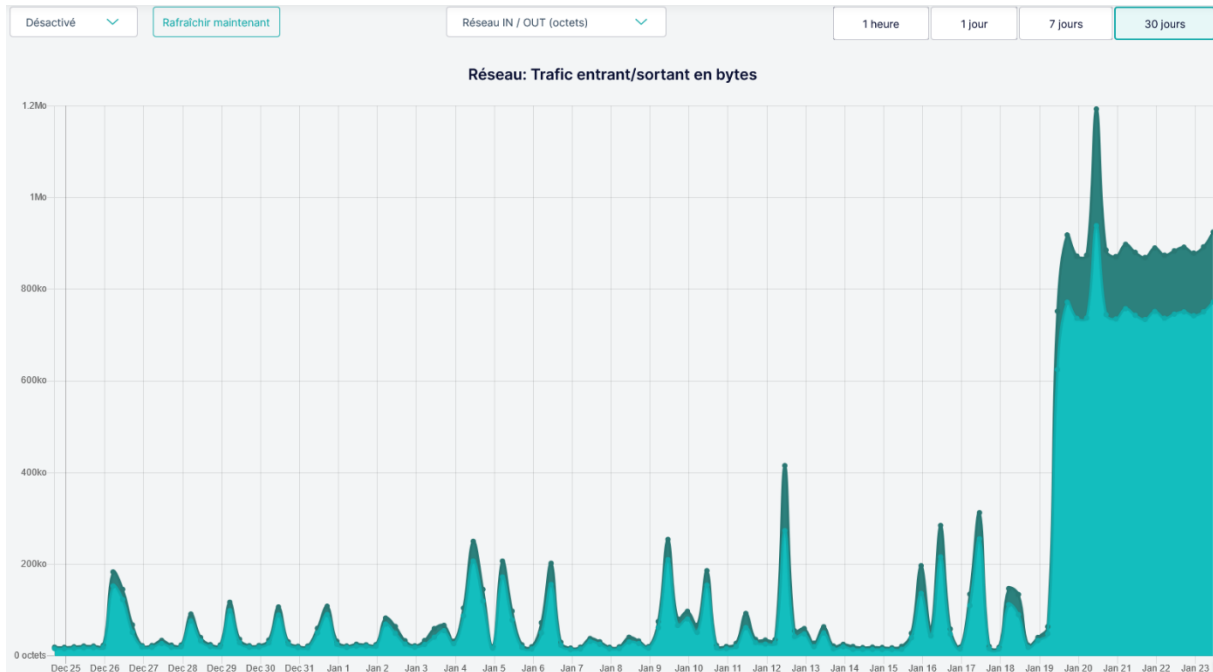


Figure 6: Network “in” and “out” of the API hosted service.

ORPHADATA WEBSITE

During 2022, we have launched a new version of the Orphadata platform website.

<https://www.orphadata.com> (fig.7).

Orphanet Data For Rare Diseases

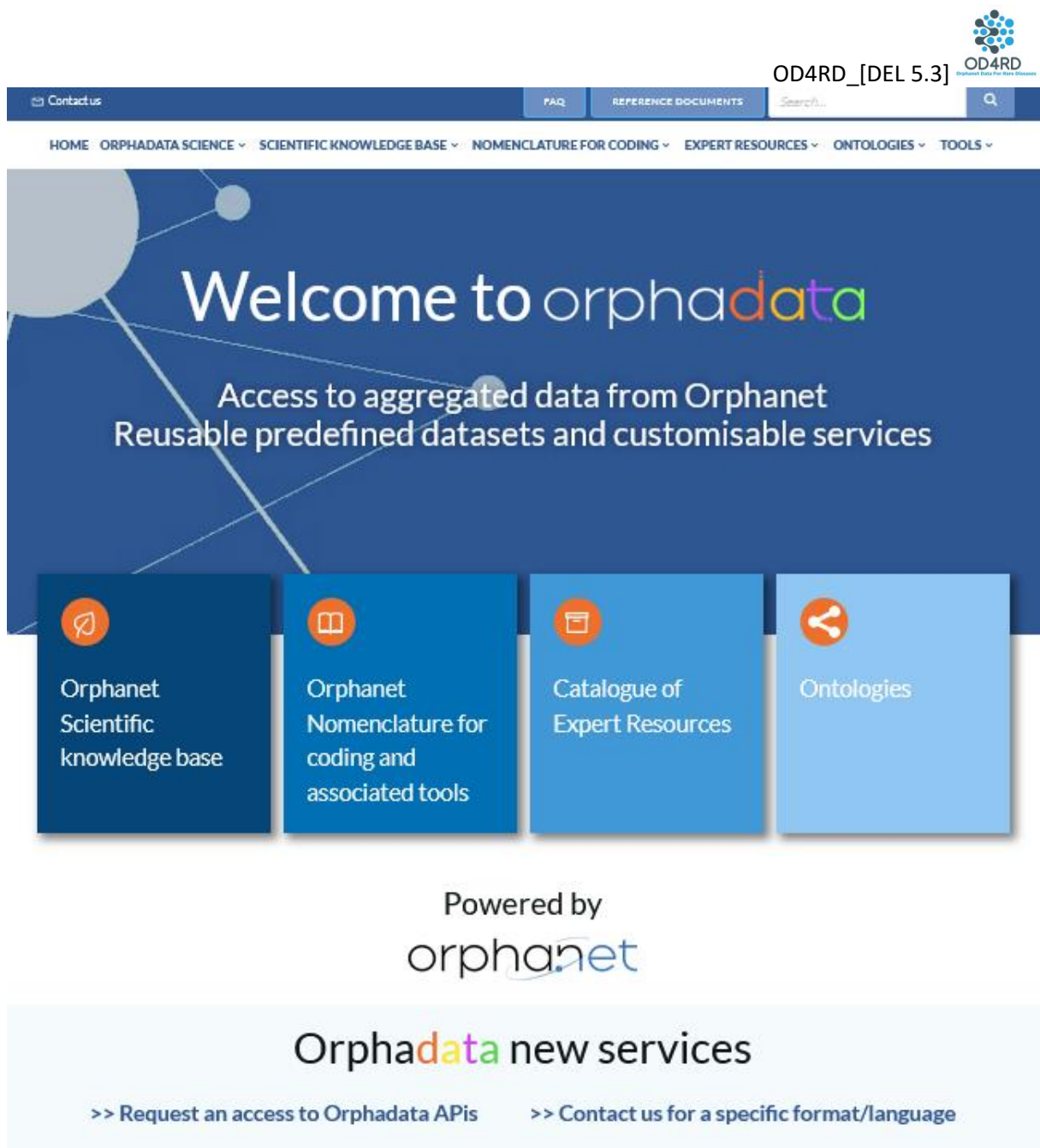
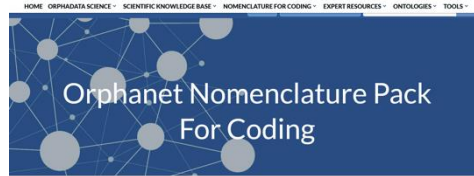


Figure 7: New version of Orphadata

Orphanet Data For Rare Diseases

This version is more user-friendly and facilitates the access to the different datasets and services, including the APIs. The relevant documentation is also accessible and we provide access to the OD4RD Github tracking system (WP4, T4.1 and T4.4) from the Orphanet Nomenclature Pack for Coding page.



The Orphanet nomenclature pack comprises various files (listed below) which provide the computable information necessary to achieve implementation of ORPHAcodes in health information systems, and ensure easier and accurate coding. These files are updated once a year, in 9 different languages: Czech, Dutch, English, French, German, Italian, Polish, Portuguese and Spanish. Excel and PDF files are available in English and are common to all Orphanet nomenclature pack. Each year the Orphanet nomenclature pack includes:

- Orphanet nomenclature file (XML Schema Definitions and JPEG representations for this file).
- Orphanet to ICD-10/ICD-11 mapping files (XML Schema Definitions and JPEG representations for these files).
- Orphanet classifications files by medical specialties (XML Schema Definitions and JPEG representations for these files).
- Linearisation file attributing one preferential medical specialty to every clinical entity (XML Schema Definitions and JPEG representations for this file).
- Master file (Excel file), the minimal set of ORPHAcodes, aligned with ICD-10 codes, that should be used for data sharing and statistical purposes at EU-level. More information about the Master file can be found [here](#).
- Orphanet nomenclature differential file (Excel file) which provides the key changes made to the nomenclature compared to the previous version.
- Description file (PDF file describing the xml files enclosed in the Orphanet nomenclature pack for coding). More information about the Aggregation level can be found [here](#).
- Recommendations and helpdesk for coding are available below in the dropdown menu.

The nomenclature is also available through dedicated APIs and a human-readable view is provided through [DataWiz](#).

All files are available under the [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](#)

These products have been recognised as an [EURO Core Data Resource](#) and a [Global Core Biobank Resource](#)

These products are an [WDORC registered resource](#)

[Nomenclature pack 2022](#)

[Previous nomenclature packs](#)

[Guidance for coding](#)

[Helpdesk for coding](#)

© 2022 Orphadata [Legal Notice](#) [Contact](#) [GDPR](#)  

Fig 8 . Orphanet Nomenclature Pack for Coding page from Orphadata with link to the helpdesk tool (Github tracking system) from WP4.

HOME ORPHADATA SCIENCE ▾ SCIENTIFIC KNOWLEDGE BASE ▾ NOMENCLATURE FOR CODING ▾ EXPERT RESOURCES ▾ ONTOLOGIES ▾ TOOLS ▾

[Guidance for coding](#) ▾

[Helpdesk for coding](#) ▾

You have a question related to the Orphanet nomenclature content and the implementation of ORPHAcodes in Health Information Systems, contact us via [GitHub](#)

ORPHADATA API

To extend the availability of the datasets through API mechanisms, we also started to provide an API access to the Orphadata “scientific content”.

<https://api.orphadata.com/>

The content possibly accessible from the API includes (fig.9):

- Rare diseases aligned with terminologies and databases
- Classification of rare diseases
- Clinical signs and symptoms in rare diseases
- Genes associated with rare diseases
- Medical speciality of rare diseases (Linearisation)
- Epidemiology of rare diseases
- Natural history of rare diseases

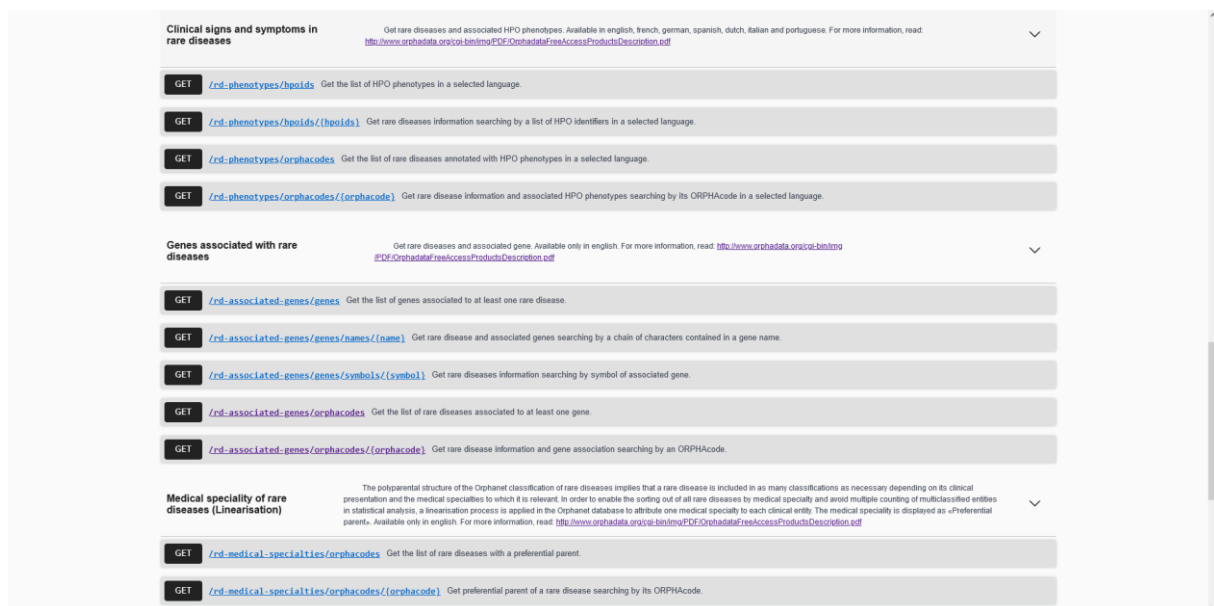


Figure 9: Queries accessible from <https://api.orphadata.com>

This will allow stakeholders to develop or update their software(s) to annotate datasets with content from the Orphanet knowledge base when relevant.

ORPHADATA FAIRDATAPOINT

A FAIR datapoint has been setup to also ease the access from machines and algorithms. (fig.10).

<http://fairdatapoint.orphadata.com>

Orphadata FAIR Data Point

FAIR Data Point to host free orphanet's dataset(s). These data sets are a partial extraction of the Orphanet data base, freely accessible at www.orpha.net (for consultation purposes only). The data sets are available in nine languages: English, French, German, Italian, Portuguese, Spanish, Dutch, Polish and Czech.

Catalogs

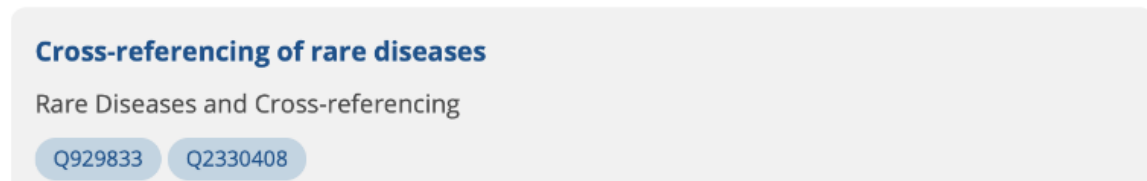


Figure 10: Orphadata FAIR datapoint

The datasets are described in a machine-readable format such as JSON Linked Data (JSON-LD) and RDF (Resource Description Framework)

And deployed online : <https://dataviz.orphacode.org>

API MANAGEMENT GATEWAY

In order to have a better management and sustainable monitoring of APIs usages we have setup an API gateway service.

An API gateway aims to grant access to API for authenticated users, accordingly use thresholds when necessary and manage subscription and authentication process. Authentication is needed, even for users with no fees, which ensures a better adaptation and scalability of the API services. (Detection of abusive usages or simply anticipation of upgrading the services capacities).

Quota limits and rates are easily setup with an API gateway management system. For instance:

Rate limit	10 requests per second	20 requests per second	100 requests per second
Quota limit	5000 requests per week	20000 requests per week	100000 requests per week

Are examples of possible limitations to protect the API services.

A subscription portal has been deployed (fig.11): <https://orphanetapi.developer.azure-api.net/>.

IT describes the different available API and limitations, depending on subscription.

	Free	Standard	Advanced
Price	0€ / month	10€ / month	100€ / month
Requires approval	No	Yes	Yes
Rate limit	10 requests per second	20 requests per second	100 requests per second
Quota limit	5000 requests per week	20000 requests per week	100000 requests per week
Access to Orphacodes API	Yes	Yes	Yes
Access to Orphadata API	Only until 12/30/2022	Yes	Yes

Our APIs



Orphacodes API

The Orphacodes API intends to support Member States in the implementation of the ORPHA nomenclature to rare diseases-specific codification systems.

[Documentation](#)

Orphadata API

The Orphadata API gives access to data from the Orphanet Scientific knowledge

Figure 11: API Gateway Portal

The Microsoft Azure backoffice was configured to manage API key generation, after user sign-in validation (by acceptance of the Use Conditions) (fig.12).

The screenshot shows the Microsoft Azure portal interface for managing API Gateway subscriptions. The main content area displays a table of active subscriptions with the following data:

Nom d'affichage	Clé primaire	Clé secondaire	Étendue	État	Propriétaire	Autoriser le suivi
Built-in all-access subscription	*****	*****	Service	Actif		✓
	*****	*****	Produit: Free	Actif	Administrator	✗
	*****	*****	Produit: Advanced	Actif	Administrator	✗
free	*****	*****	Produit: Free	Actif	jd	✗
testing_orphanet_mrc	*****	*****	Produit: Free	Actif	Marc Test	✗

Figure 12: API Gateway Portal backoffice

Discussion

During the past year, in the context of the Work Package 5 activities, we have extended the Orphanet's capacities to serve the needs of several stakeholders, addressing technical challenges by updating and upgrading the available services.

However, aligning roadmaps with different stakeholders and ensuring production services during the timeframe of the project may represent an issue as new needs arise regularly and therefore, the scalability and adaptation of our services have been and will be performed on regular basis even after the end of the project. This is facilitated by the technical IT workflow conception, which is based on datasets extracted from the Orphanet's knowledge base to populate Elasticsearch Cloud platform and serve dedicated API and services, and it is « agile » by design. This agile design ensures the capacity to provide different tools or even datasets models in a reasonable short time, without touching the core of Orphanet IT system.

Moreover, regarding API access management, the technical solutions are deployed and configured, but the subscription process for massive use of the APIs, in terms of number of rate or quota limitations, should be set up by defining the conditions of use and eventual fees for private sector users, which can contribute to the long-term sustainability of the system. This is being discussed with Inserm's legal department and will lead to ad hoc configuration when specified.