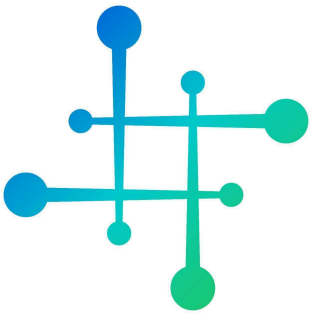


EHDS

HealthData@EU Pilot

Milestone M6.1

Report on the landscape analysis of available metadata catalogues and the metadata standards in use

Pascal Derycke, Irène Kesisoglou, Truls Korsgaard, Hans Aage Huru, Charles Andrew Vande Catsyne

30 March 2023

Co-funded by
the European Union

Disclaimer: 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.

Table of Contents

Introduction

Multiple studies have shown that one of the main barriers[1] to the re-use of health data has been the lack of findability, the lack of an inventory listing all kinds of datasets that can be made available for a different purpose than the one they were initially collected for. Such an inventory can be a metadata catalogue. A catalogue composed of metadata records, one for each dataset describing information about the dataset without giving access to the actual data.

Considering the above finding, the need to have a common metadata catalogue at European level, presenting the health-related data that can be made available for secondary use, the European Commission drafted the legislation for the European Health Data Space (EHDS) and mentioned this in the following articles:

Article 38:
(i) a national dataset catalogue[2] that shall include details about the source and nature of electronic health data, in accordance with Articles 56 (Data quality and utility label) and 58 (minimum dataset specifications/implementing act), and the conditions for making electronic health data available. The national dataset catalogue shall also be made available to single information points under Article 8 of Regulation […] [Data Governance Act COM/2020/767 final];

Article 55: Dataset description
1. The health data access bodies shall inform the data users about the available datasets and their characteristics through a metadata catalogue. Each dataset shall include information concerning the source, the scope, the main characteristics, nature of electronic health data and conditions for making electronic health data available.
2. The Commission shall, by means of implementing acts, set out the minimum information elements data holders are to provide for datasets and their characteristics. Those implementing acts shall be adopted in accordance with the advisory procedure referred to in Article 68(2).

Article 57: EU Datasets Catalogue
1. The Commission shall establish an EU Datasets Catalogue connecting the national catalogues of datasets established by the health data access bodies and other authorised participants in HealthData@EU.
2. The EU Datasets Catalogue and the national datasets catalogues shall be made publicly available.[3]

After highlighting the importance and legal mandate to start developing metadata catalogues or dataset catalogues, we would like to introduce the methodology of how to do so. When creating a metadata record, a description, for the dataset that can be made available, it is important to consider the software, standard and template one can use to develop that. This is essential in order to make the record interoperable; machine-to-machine readable, so that the metadata record can be accessed remotely through many metadata catalogues or other sources, such as

[1] https://tehdas.eu/results/tehdas-identifies-barriers-to-data-sharing/
[2] Definition: (ac) 'dataset catalogue' means a collection of datasets descriptions, which is arranged in a systematic manner and consists of a user-oriented public part, where information concerning individual dataset parameters is accessible by electronic means through an online portal;
[3] https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en

metadata brokers[4] or overarching metadata catalogues (e.g. data.europa.eu).

Within this pilot project, which aims to design, develop and pilot the infrastructure for the EHDS for secondary use, the HealthData@EU infrastructure, WP6 is responsible for the development of a standardised metadata template and an online tool to support the development of local, national and European metadata catalogues and ensure a seamless interconnectivity between them. To start the design of this standardised metadata template where data providers can input information about their datasets, we started by exploring the already available metadata standards and catalogues.

A recent study performed by the Joint Action TEHDAS[5], revealed that there are multiple metadata standards currently in use by stakeholders, such as MIABIS[6], DCAT-AP[7], CESSDA CMM[8] and DDI[9]. The metadata standard used by the data.europa.eu, which harvests metadata records from various sectors, such as environmental, health, education, justice and transport, is the DCAT-AP. Therefore, in order to ensure that the national metadata catalogues describing the health-related datasets that can be made available for secondary use within the EHDS are compliant with and can be harvested by the data.europa.eu, we decided to use the DCAT-AP as the baseline metadata standard.

DCAT-AP has been used to describe datasets in various fields. When a field needs more specific properties, the DCAT-AP community develops extensions to this standard. For example, there is currently an extension for metadata records describing datasets on statistics, StatDCAT-AP, and an extension for metadata records describing geospatial datasets, GeoDCAT-AP.

At the moment the health-related metadata records found through data.europa.eu or other metadata catalogues using DCAT-AP don't address the specificity of health data enough and this has an impact in the findability and reusability of these health-related datasets. There are properties in the current version of DCAT-AP that are not relevant for health-related datasets or data registries and hence data providers end-up leaving most of these properties empty resulting in a reduced amount of available information about their datasets. Then, some properties specific to health-related data and essential to improve the findability and reusability of the data by researchers and policy makers might be missing, such as the granularity of the data (aggregated/individual), topic of the health data (e.g. disease etc.), the age range and the sex and the quality of the data (e.g. which semantic standard are they structured with). Furthermore, throughout this landscape analysis we noticed that in the drop-down menu of the property fields in the current DCAT-AP version, there are controlled vocabularies missing and others that need modification.

Therefore, for the HealthData@EU pilot project, we decided to design the Health extension to the DCAT-AP to involve properties specific to health-related datasets and address the

[4] Definition: the purpose of a metadata broker is to provide a central point of control for metadata, enabling users to search, retrieve, and manage metadata from different sources and formats. It is like an overarching metadata catalogue, as it harvests metadata records from other metadata catalogues that are DCAT-AP compliant.
[5] https://tehdas.eu/results/tehdas-assesses-data-interoperability-standards/
[6] https://github.com/BBMRI-ERIC/miabis
[7] https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en/
[8] https://datacatalogue.cessda.eu/
[9] https://ddialliance.org/

aforementioned issues.

WP6 scope and objectives

The aim of WP6 is to design, develop and implement a standardised descriptive metadata template for health-related datasets respecting the FAIR principles (Findable, Accessible, Interoperable, Reusable) and user needs of the EHDS. For the reasons described above, the machine readable structured descriptive metadata template will be based on the Health DCAT-AP standard.

Objectives of WP6:

1) To carry out a landscape analysis of the descriptive metadata standards currently in use by the nodes of the HealthData@EU pilot project consortium
2) To customise a metadata template fine-tuned to potential EHDS users, which includes the provision of an extension to DCAT-AP: HealthDCAT-AP
3) To support the implementation of the HealthDCAT-AP in the EHDS nodes by supporting the transition from existing metadata templates. To evaluate the feasibility and functionality of HealthDCAT-AP.
4) Management of the development of the central search portal (HealthDCAT-AP metadata catalogue) of the central service
5) To provide recommendations on further development and deployment for possible EU-wide uptake of the descriptive metadata template and ensure its compatibility with European data portals across sectors.

This milestone MS6.1 presents the results of the landscape analysis, to address the first objective of WP6. In this task we collected information about the available metadata catalogues, presented the descriptive metadata standards currently used by the consortium nodes (whether DCAT-AP is being used or a different standard) and the properties described within each metadata template.

Methodology

The sandbox

In January 2023, WP6 published an online tool to support the activity of the work package on collecting information for the landscape analysis, enabling health data providers to create a metadata record for their datasets and for designing the Health DCAT-AP extension. This tool is accessible at the following URL: http://search.healthdataportal.eu/. It offers a safe space (i.e.: sandbox) for the WP6 Technical Working Group (TWG) to later on design and pilot the Health DCAT-AP extension. The sandbox is composed of a front-end (FAIR DATA POINT Web client) and a back-end (FAIR DATA POINT server) and has been conceived as a validation testing framework for the Health DCAT-AP extension. In order to collect information from the health data community, three forms were developed using this online tool.

1. The first form aimed at answering the first objective of WP6, the landscape analysis and hence collect information on already available metadata catalogues and their specifications.

   “Landscape analysis” form has the following fields to be filled: [Metadata catalogue's name], [Country], [URL], [Metadata standards], [Metadata for machines], [Contact

email].

2. The second form is a DCAT-AP (v2) metadata template allowing data providers to create a metadata record for their datasets that will be by default DCAT-AP compliant. Data providers can also provide their feedback under each property in this form to describe whether, for example, they think it is important to maintain the property in the Health extension.

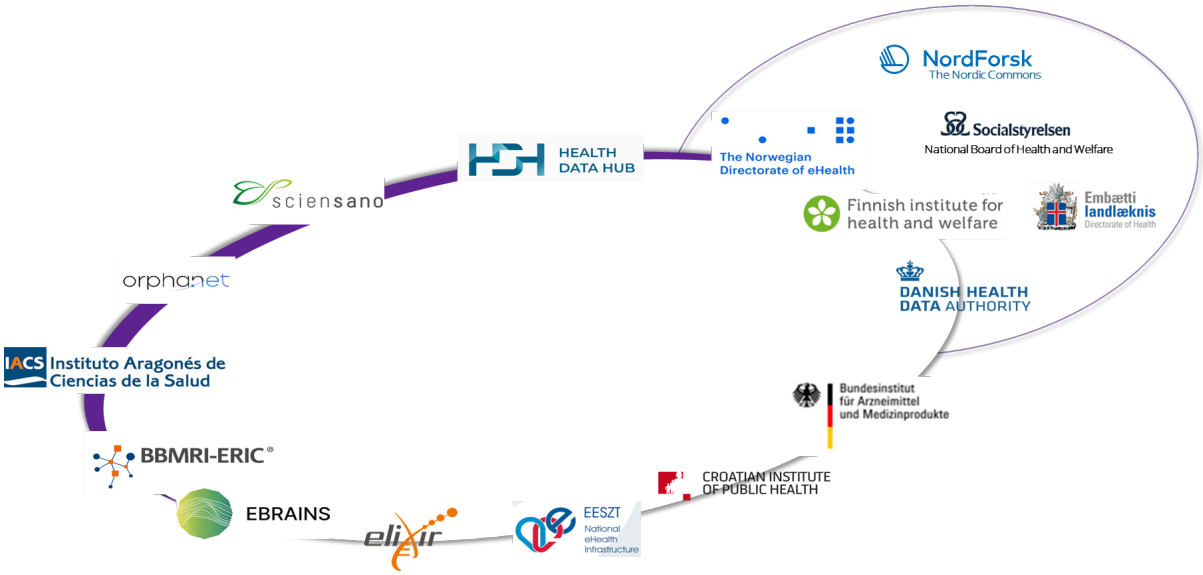
   "Create DCAT-AP metadata record (Use cases)" form provides a field to fill in for each of the DCAT-AP properties according to the official DCAP-AP's classes [dataset] and [distribution].

3. The third form is where data providers or metadata catalogue owners can propose a new property to be added to the DCAT-AP standard and hence contribute to the design of its "Health" extension. This third form has 7 fields. Data holders can describe what they consider as a missing and useful new DCAT-AP property by providing a title, a summary, a controlled vocabulary to use, a use case and specifying which of the FAIR principles it could be linked to. The data holders have the possibility to suggest as many new DCAT-AP candidate properties as they want.

Dissemination of the forms

To collect this information we disseminated the three forms described above, to the different nodes within the consortium of the HealthData@EU pilot project and to the use case leaders that had to distribute them to the nodes participating in their use case.

Figure with the consortium nodes that filled in the forms:



In parallel with the HealthData@EU pilot project, the Nordic Commons (NC) [10] includes ongoing work on metadata with more or less similar objectives. The workplan of the NC Task Force

[10] https://www.nordforsk.org/nordic-commons

"Metadata, data quality and semantics" is very well aligned with the work in the HealthData@EU/WP6, and is a kind of "pilot in the pilot". This ensures, among others, quality in the deliverables and shared experiences when it comes to the need for high level anchoring, governance, capacity building, training, infrastructure, technical and semantic interoperability.

The Nordic countries have a lot of high quality data sources for research, a long history when it comes to research collaboration and management of national metadata catalogues for health.[11] Another advantage of this cooperation is that both Iceland and Sweden are closely following the work and "training" in the HealthData@EU and are willing to contribute to the work performed in WP6 similarly to Norway, Finland and Denmark.Therefore, the above mentioned forms were also disseminated to the countries participating in the Nordic Commons that are not part of the consortium of this pilot project.

Analysis of the information collected

The data collected from the first form, the "Landscape analysis" form, are analysed and presented below. It is important to note that the factsheets present not only the responses to the 6 questions of the form but also information that resulted from a more in depth investigation of the metadata catalogue URL provided. Each fact sheet has been complemented (if information available) with a description of the structure of the catalogue, list of sub-catalogues, list of datasets, list of properties, list of health domains covered and additional comments.

Results

We started the landscape analysis by examining what metadata catalogues and health related metadata records are already possible to access/query through the data.europa.eu portal. There are two ways to perform this analysis; either through the web interface using the filters provided by the search portal or through SPARQL queries[12], which is a standard query language and protocol for Linked Open Data and Resource Description Framework (RDF) databases. Having been designed to query a great variety of data, it can efficiently extract information hidden in non-uniform data and stored in various formats and sources. SPARQL brings capability to navigate relationships in RDF graph data through graph pattern matching. In this process, simple patterns can be combined into more complex ones, which explore more elaborate relationships in the data.

Therefore, we started by performing the following SPARQL queries from the SPARQL search tool interface (https://data.europa.eu/data/sparql) which provides a machine readable interface.

To find the number of metadata catalogues already compliant with DCAT-AP and queryable through the SPARQL endpoint we performed the following query:

[11] https://www.nordforsk.org/2019/vision-nordic-secure-digital-infrastructure-health-data-nordic-commons

[12] https://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/#:~:text=The%20greatest%20strength%20of%20SPARQL,elaborate%20relationships%20in%20the%20data.

https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT (COUNT(*) as ?count)
WHERE {
?x a dcat:Catalog ;
   dcterms:title ?catalogue.
}
```

Query result :

179

- List of all metadata catalogues available through the portal :

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dc: <http://purl.org/dc/terms/>
SELECT distinct ?catalogue_name
WHERE {
?x a dcat:Catalog .
?x dc:title ?catalogue_name
}
```

Query result : (only first results shown)

“data.bev.gv.a”; “inspire.gv.at”; “National Data Catalog of Norway”; [...]

- Number of datasets in the category HEALTH of the European Data Portal :

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT (COUNT(*) as ?count)
WHERE {
?entity dcat:theme ?theme.
VALUES ?theme {<http://publications.europa.eu/resource/authority/data-theme/HEAL>}
}
```

Query result :

18966

Similar results have been obtained from the official search tool of the European Data Portal (https://data.europa.eu/data/datasets) by performing manual search using filters available.

The following fact sheet reports information on the European Data Portal metadata broker or overarching European metadata catalogue:

Catalogue : **EU Data portal**			Coverage : **International** Language : **EN**	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://data.europa.eu/en	CKAN + CKAN-EDP extension (for mapping DCAT) + Virtuoso (triplestore) for storing semantic metadata in RDF	Hybrid approach : CDS (CKAN data schema) and DCAT	Sparql query: https://data.europa.eu/sparql CKAN-JSON	The Publications Office of the European Union

Description :

Structure:

The **harvester** service supports multiple protocols (CKAN-API, SPARQL, RSS, OpenDataSoft API..). It collects metadata from data providers and transforms them into DCAT-AP compliant data structure through defined transformation rules. They define how the source serialisation is converted into the target one. Metadata are then pushed into the metadata registry.

The **metadata registry** is the core component of the EDP and is the primary access point for users. Metadata are accessible through web content, CKAN-JSON and from SPARQL query (SPARQL manager). It is based on CKAN. Metadata are stored in CKAN Data schema (CDS format) into PostgreSQL relational database (for syntactic metadata) but also in RDF format into Virtuoso =Triplestore database (for semantic metadata) because CKAN doesn't natively support RDF/DCAT.
They developed a CKAN-EDP-extension to bring new functionalities including compatibility of CKAN with DCAT via mapping rules. By using this hybrid approach, they provide support for already established non linked Data and also for linked data compliant for DCAT. But it implicates complex management due to the need of replication of metadata into two distinct database systems, and because metadata has to be represented in two formats : in CKAN-JSON and in DCAT-AP with bi-directional conversion.

The last main component is the **metadata quality assurance** service. It checks metadata stored into the metadata registry according to two levels : it checks the format correctness of the metadata (missing properties, wrong data types…) but also the content of specific metadata properties (accessibility, most used data formats, most used licenses…). There is also a full validation of the entire data pool once every month.

Finally, there are external services connected to the metadata registry (External vocabulary, eTranslation service, EU login(CAS)..).

All of these services communicate via a RESTful-API.

List of catalogues :
176 catalogues from 36 different countries.
List of metadata records :
1.606.841 records.

Properties :

Following DCAT-AP properties (cfr documentation).
https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/211

Health domains :

- ☑ Health Registries
- ☑ Diseases Registries
- ☑ Drugs Registries
- ☑ Biobanks
- ☑ OMICS
- ☑ Laboratory
- ☑ Health Surveys, Clinical Trials
- ☑ Tests Diagnostic & Clinical
- ☑ Health Care Services, Facilities or Quality
- ☑ Health Networks
- ☑ Population Data
- ☑ Epidemiology And Surveillance

Comments :

It is actually a **metadata broke**r or an overarching metadata catalogue for open data from different sectors (health, environment, transport,...).

Responses to the landscape analysis forms

The information received from the "Landscape analysis" form allows us to reflect on the current state and maturity level of available health related metadata catalogues in Europe. 18 forms have been received at the moment of the drafting of this milestone. Based on the information

collected, WP6 leads have visited each metadata catalogue and performed an in depth assessment of their structure, information and services. For each catalogue, a fact sheet has been created to provide, in a structured way, information about the URL, the software used to develop the metadata catalogue, the metadata standard in use, the structure of the catalogue, the domain(s) of the catalogue, the metadata properties in use to describe datasets, the technologies supporting the catalogue and the contact information of the responsible entity for the metadata catalogue.

Catalogue : **EMA EU metadata catalogue of data sources and studies**			Coverage : EU Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://data-catalogue.molgeniscloud.org/catalogue/catalogue/#/networks-catalogue URL derived from info provided in API field/ The public side of the catalogue is developed in Drupal and will be a domain such as ema.europa.eu/catalogues The MOLGENIS component of the solution will ensure interoperability with other catalogues.	MOLGENIS + Drupal https://molgenis.gitbook.io/molgenis/#why-molgenis https://molgenis.org/	Dedicated metadata schema Using Molgenis, they have designed their own metadata schema. References for interoperability of MOLGENIS: https://elixir-europe.org/platforms/interoperability/rirs metadata standard: a.o. https://www.w3.org/TR/vocab-dcat-2/	https://data-catalogue.molgeniscloud.org/catalogue/api/graphql Doc : https://molgenis.gitbook.io/molgenis/interoperability There is also a Drupal API for which we do not have references yet. Final references to the exposed APIs will be made available in Q4 2023.	European Medicines Agency metadata@ema.europa.eu Paolo.Alcini@ema.europa.eu

Description :

Under development.

At a first stage they intend to migrate the data from the existing catalogue here: Resources Database (encepp.eu) Also, to add metadata of other data sources that expressed interest to be included in the new catalogue or were found useful in the preparation stage. After go-live the Drupal interface will allow data entry (and maintenance) directly via a web interface for data source holders. The MOLGENIS component is meant to ensure interoperability with other catalogues and a potential future automation where possible.

EMA/787647/2022 Good practice guide for the use of the Metadata catalogue of real-world data sources https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf

Structure:

- Data sources
  - Administrative details (16 – ex: Data source type: Hospital discharge records)
  - Data elements collected (8 – ex: Percentage of the population covered by the data source in the catchment area)
  - Quantitative descriptors (6 – ex: Population size by age)
  - Data flows and management (18 – ex: CDM mapping: Is the data source ETL-ed to a CDM?)
  - Data source – Vocabularies (10 – ex: Cause of death vocabulary:<select multiple: Not captured | ICPC | ICD9| ICD10| ICD1|)
- Study metadata

- Administrative details
- Methodological aspects
- Data management

- Institution metadata
- Network metadata

Properties :

For more information about properties (data sources): EMA/563896/2022 List of metadata for Real World Data catalogues
https://www.ema.europa.eu/en/documents/other/list-metadata-real-world-data-catalogues_en.pdf

Health domains :

- ☑ Health Registries
- ☑ Diseases Registries
- ☑ Drugs Registries
- ☑ Biobanks
- ☑ OMICS
- ☑ Laboratory
- ☑ Health Surveys, Clinical Trials
- ☑ Tests Diagnostic & Clinical
- ☑ Health Care Services, Facilities or Quality
- ☑ Health Networks[13]
- ☑ Population Data
- ☐ Epidemiology And Surveillance

Comments :

Under development.
There is planned work with Molgenis in Q2 2023 to investigate making EMA's catalogues available in a DCAT-AP compliant format.

Catalogue : **European Joint Programme Rare diseases Virtual Platform**			Coverage : EU Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
VP Portal web interface: https://vp.ejprarediseases.org		https://www.w3.org/TR/vocab-dcat-2/	https://vp.ejprarediseases.org/discovery/docs/apiDocs.html Serialisation: JSON	EJP RD INSERM Hôpital tala.haddad@ejprd-project.eu
FDP index : https://index.vp.ejprarediseases.org/	FAIR DATA POINT	https://www.w3.org/TR/vocab-dcat-2/	https://fairdatapoint.readthedocs.io/en/latest/usage/api-usage.html SPARQL can be used : https://index.vp.ejprarediseases.org/search?isSparql=true	

Description :

It is not a metadata catalogue, it is a **metadata broker**, they link different metadata catalogues.
The FAIR data point (FDP) index (https://index.vp.ejprarediseases.org/) which is used to federated the FDP provided by each resources on-boarded into the VP. This is used by the VP portal (https://vp.ejprarediseases.org) to list the resources.

Structure: Patient registries, Biobanks, Knowledge bases

They have added filters. This filter should work for some of them but not all resources. The query works with ORPHAcodes and ICD-10 for now.
Filters of search tool: male/female + Age of patient (min/max) + country

[13] Examples of health networks:Expert Centers Networks, Patient Organisations Networks, Multinational Research Projects Networks, Multinational Clinical Trials Networks, Patients Registries Networks

Example of Patient registries, Biobanks, knowledge bases connected to the search tool: (federated queries?)

- The European Registries for Rare Endocrine Conditions. (https://www448.lamp.le.ac.uk/e-REC/)
- The European Rare Kidney Disease Registry. ()
- The BBMRI-Eric catalogue of rare disease registries and biobanks. (https://directory.bbmri-eric.eu/#/)
- The Orphanet catalogue of rare disease registries and biobanks. (https://www.orpha.net/consor/cgi-bin/Disease.php)
- Cellosaurus (https://www.cellosaurus.org/)
- Duchenne Data Platform (https://duchenne.nl/duchenne-data-platform-english/)
- …

Orphanet ICD-10 Coding Rules for Rare Diseases:
https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rules_R1_Nom_ICD_EP_06.pdf

List of catalogues:
Accessible from https://vp.ejprarediseases.org/discovery/# (connected resources) Please note that the “on-boarding” of new rare diseases resource is on-going.

Properties :
DCAT-AP properties

Health domains :

- [x] Health Registries
- [ ] Diseases Registries
- [ ] Drugs Registries
- [x] Biobanks
- [ ] OMICS
- [x] Laboratory
- [ ] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [ ] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [ ] Population Data
- [ ] Epidemiology And Surveillance

Comments :
Metadata broker under development

Catalogue : **Orphanet**

Coverage : International
Language : EN/FR/IT/DE/NL/ES/CS/PL/PT

URL	Software	Metadata standard	API	Contact information
https://www.orpha.net		Bioschemas (on rare diseases pages) Bioschemas - Marking up clinical entities in the Rare Diseases community	https://api.orphacode.org/ (RD Code)	INSERM US14 - Orphanet data.orphanet@inserm.fr

Description :
https://www.orpha.net

List of catalogues :
Rare diseases, disabilities, phenotypes, Encyclopedia (guidelines), Expert Centers (including ERNs), Laboratories and Diagnostic tests, Orphan Drugs, Genes, Research projects, Clinical trials, patient organisations and federations, platorms, registries, biobanks, variants databases

Properties :
List of properties:
Huge list of metadata, depending on type of resources.

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [x] Biobanks
- [x] OMICS
- [x] Laboratory
- [x] Health Surveys, Clinical Trials
- [x] Tests Diagnostic & Clinical
- [ ] Health Care Services, Facilities or Quality
- [x] Health Networks
- [ ] Population Data
- [x] Epidemiology And Surveillance

Comments :
Health networks: Expert Centers Networks, Patient Organisations Networks, Multinational Research Projects Networks, Multinational Clinical Trials Networks, Patients Registries Networks

Catalogue : **Orphadata**

Coverage : EU
Language : EN

URL	Software	Metadata standard	API	Contact information
https://www.orphadata.com	FAIR DATA POINT	https://www.w3.org/RDF/	https://api.orphadata.com/ http://fairdatapoint.orphadata.com Serialisation: JSON	INSERM US14 - Orphanet data.orphanet@inserm.fr

Description :
Orphadata Fair data point : https://www.orphadata.com/ (access to datasets about rare diseases- ex: Classifications of rare diseases)
Please note that www.orphadata.com is a subset of the Orphanet Knowledge base, allowing downloads of large datasets.

List of catalogues :
Classifications of rare diseases, Cross-referencing of rare diseases, Genes associated with rare diseases, Linearisation of rare diseases, Natural history of rare diseases, Phenotypes associated with rare diseases, Rare disease epidemiology, Rare diseases and functional consequences

Properties :
List of properties:
Title, Responsible party, Processor, Purpose, Content, Period, Source of data, Registration details, Comparable or related datasets, Processing and publication, History

Example of a metadata records: Rare disease epidemiology in French language

Health domains :

- [ ] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [ ] Biobanks
- [x] OMICS
- [ ] Laboratory
- [ ] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [ ] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [ ] Population Data
- [x] Epidemiology And Surveillance

Comments :
No correct DCAT implementation: This catalogue is actually a dataset and datasets are distributions of the same dataset. They need to add the actual datasets in the class of distribution. ex: http://fairdatapoint.orphadata.com/catalog/eb9d621f-f2ba-4f4e-bcbe-e9a7b1e5d127

Catalogue : **Lists of variables for health registers**

Coverage : Iceland

			Language : Icelandic	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://island.is/en/gagnasofn-embaettis-landlaeknis/eydublod-breytulistar		HTML (Webpage). The Icelandic Causes of Death Register is mapped to DCAT as a PoC in The Nordic Commons project. The DCAT-based metadata is shared as a JSON-files on the GitHub that belongs to the Norwegian Directorate of e-health.		Office of the National Medical Examiner gudrun.kr.gudfinnsdottir@landlaeknir.is sigridur.haraldsd.elinardottir@landlaeknir.is mottaka@landlaeknir.is

Description :

Databases of the Directorate of Health in Iceland. See https://island.is/gagnasofn

Very similar to a metadata catalogue but no DCAT-AP compliant technology behind. Very easy to be mapped to DCAT-AP though.

Current progress:Mapping

List of metadata records:

Vaccination Register, Causes of Death Register, Birth Register, Register of Nursing Home Pre-Admission Assessments, Health and Wellbeing in Iceland, Coronary Event Register, Cancer Register, Prescription Medicines Register, Register of Sterilizations (no longer updated), Register of interRAI Assessments, Register of Primary Health Care Contacts, Register of Contacts with Medical Specialists in Private Practice, Cancer Screening Register, Accident Register (no longer updated), Register of Communicable Diseases, Register of Licensed Healthcare Practitioners, Register of Biobanks Opt-Outs, Hospital Discharge Register, Register of Terminations of Pregnancy (no longer updated), Register of Health Care Operators.

Properties :

List of properties:

Title, Responsible party, Processor, Purpose, Content, Period, Source of data, Registration details, Comparable or related datasets, Processing and publication, History

List of variables available as downloads (xls files).

Ex: The Icelandic Causes of Death Register (https://island.is/gagnasofn/eydublod-breytulistar)

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [ ] Biobanks
- [ ] OMICS
- [ ] Laboratory
- [x] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [x] Population Data
- [ ] Epidemiology And Surveillance

Comments :

Part of the Nordic Commons task force for metadata, data quality and semantics.

The project started in 2022 and will end in 2024.

In addition to training and enabling the participating countries in structuring, standardization and sharing metadata about their health data sources for research, the objective is as well to set up an infrastructure with endpoints for sharing machine readable metadata between the national nodes, and with EHDS.

The scope is both standardized metadata about the data source (data set), metadata about the belonging variable, and the code lists/value sets the relevant variables are based on.

Catalogue : **Metaregistar**			Coverage : Croatia Language : Croatian	
URL	**Software**	**Metadata standard**	**API**	**Contact information**

https://metaregistar.gov.hr/metareg/html/javno_pocetna.xhtml This is the address of the actual Metaregistry, which is outdated and no longer up-to-date, neither in terms of data nor software.	SDURDD is currently developing Metaregistar on a new software platform (cfr description)	Currently, a semantic model of the Metaregister is being developed, which is based on several entities, that the European Interoperability Framework also recognizes as the fundamental entities of registers, such as persons, vehicles, economic entities, and others https://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_3&format=PDF	Once the Metaregistar platform is completed, the API and documentation for the legal download of data from authoritative sources will be available.	Republic of Croatia and Croatian Institute of Public Health ured@rdd.hr pero.ivanko@hzjz.hr pero.ivanko@hzjz.hr

Description :

Under development.

SDURDD is currently developing Metaregistar on a new software platform, which will be available through a web browser on a custom tool developed independently in the Shared Services Center in HR. This new platform will provide the basis for all registers used in HR for the exchange of data between state bodies and public authorities for the needs of business processes carried out within the jurisdiction of each individual body. Currently, the new Metaregistry platform is not publicly available and does not have any real-accurate data. This year, we expect the completion of software functionalities and the entry and updating of data on Metaregistries.

List of metadata records:

The scope of data in the Metaregistry will depend on the classification of the data, and the level of restrictions on the availability of data to end users. For example, if an individual register contains restricted data, that data will not be visible in the Metaregister.

Properties :

List of properties (old registry) :

DATA: Registry name, Link, Abbreviation, Register description, Version, Time sequence of, Use, Time sequence until, Form of keeping the register
REGISTRY DATA: Id, The name of the data, Description, Type, Authenticity, Source, A semantic term, Classification, Code book, Length, Maximum value, Cardinality, Management of, Leading to, Personal data, Secrecy
METHOD OF COLLECTION: Description of the method, Collection period
LEGAL BASIS: Legal basis, Species, Reference
THE ROLE: the role, Person, Responsibilities, Date of, Date to
ADMINISTRATIVE AREAS: Administrative areas
JURISDICTION OVER THE REGISTER: Record, Connection type, Jurisdiction of, Jurisdiction to
SERVICES: Record, Connection type, Jurisdiction of, Jurisdiction to

Health domains : Under development. No metadata available.

- ☐ Health Registries
- ☐ Diseases Registries
- ☐ Drugs Registries
- ☐ Biobanks
- ☐ OMICS
- ☐ Laboratory
- ☐ Health Surveys, Clinical Trials
- ☐ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☐ Health Networks
- ☐ Population Data
- ☐ Epidemiology And Surveillance

Comments :

Under development

Catalogue : **Danish national metadata catalogue**			Coverage : Denmark Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
www.datalandkort.rsyd.dk	Dateado	For catalogues and tables → DCAT-AP-DK For variables and metadata → ISO 11179(-7)	Under development	Danish Metadata Map Eva.Sandberg@rsyd.dk

Description :
The Danish Meta Map provides the following services:
Data source explorer from registers, tables and variable explorers. Under development how to create your own variable list.
The data sources are structured with a system to drill down to variable list, where detailed information about data breach, ancestry, collection period, code and value lists e.g. if the information is available from the data supplier.

List of datasets:
selected tables from:
The Danish Registration Register 10
The National Patient Register 5
The Cause of Death Register 2
Health Insurance Register 1
Danish Hip Arthroplasty Register 1
Danish Shoulder arthroplasty register 1

Properties :
Data source, sub sources and data records:
Title, description, data supplier, source, data breach, ancestry, collection period

It is possible to filter on:
Data source, categories, tables, variables, periods

Health domains : Under development. No metadata available.

- [x] Health Registries
- [ ] Diseases Registries
- [ ] Drugs Registries
- [ ] Biobanks
- [ ] OMICS
- [ ] Laboratory
- [ ] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [ ] Population Data
- [ ] Epidemiology And Surveillance

Comments :
Part of the Nordic Commons task force for metadata, data quality and semantics, but not that active due to the deployment of the national health metadata catalogue “Datalkandkortet”
The project started in 2022 and will end in 2024.
In addition to training and enabling the participating countries in structuring, standardisation and sharing metadata about their health data sources for research, the objective is as well to set up an infrastructure with endpoints for sharing machine readable metadata between the national nodes, and with EHDS.
The scope is both standardised metadata about the data source (data set), metadata about the belonging variables, and the code lists/value sets the relevant variables are based on

Catalogue : **RUT - Register Utiliser Tool**			Coverage : Sweden Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://rut.registerforskning.se/		https://statswiki.unece.org/display/gsim/		Swedish Research Council johanna.holm@socialstyrelsen.se

Description :

In the catalogue you get a quick overview of what is in the various registers.
Structure:

- RUT Metadata directory: 39 records.
- RUT Metadata Tool

In the tool, the content of the connected registers is described with metadata in a standardised and detailed manner. You can do advanced searches and compare different variables from several perspectives. To use the tool, you need to log in. Setting up an account is possible only for staff and researchers connected to an organisation performing research, and for staff at data holding authorities.

Properties :

In the metadata catalogue overview https://rut.registerforskning.se/metadatakatalog/register/698318b7-fada-4543-9751-85c65f7d9b4b/:
Title, Record keeping organisation, website, description, purpose, type of registry, connection status, data collection period, metadata in rut (timestamp)

In the metadata tool (which requires log-in):
Metadata on variable level structured according to GSIM; register structure according to the GSIM model, variable name, variable definition, (statistical) object type, population covered by the variable, code list, value domain, measurement scale, period of measurement etc

Health domains :

- [x] Health Registries
- [ ] Diseases Registries
- [ ] Drugs Registries
- [x] Biobanks
- [ ] OMICS
- [ ] Laboratory
- [x] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [ ] Population Data
- [ ] Epidemiology And Surveillance

Comments :

Part of the Nordic Commons task force for metadata, data quality and semantics.
The Nordic commons project started in march 2022 and will end in 2024.
In addition to training and enabling the participating countries in structuring, standardisation and sharing metadata about their health data sources for research, the objective for the metadata task force is as well to set up an infrastructure with endpoints for sharing machine readable metadata between the national nodes.
The scope is both standardised metadata about the data source (data set), metadata about the belonging variables, and the code lists/value sets the relevant variables are based on. Nordic commons is liaised with EHDS through its executive team, for investigating possible synergies of use cases and general work on metadata infrastructure.

Catalogue : **French Health Data Hub**			Coverage : France Language : FR	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://catalogue-metadonnees.health-data-hub.fr/	Ad hoc application based on Jena Apache triple store	DCAT-AP | 2.1.1	Dedicated API Serialisation: JSON, YAML, XML DCAT, SPARQL endpoint	Plateforme des données de santé (Health Data Hub) lara.dirian@health-data-hub.fr

			Swagger interface: https://catalogue-metadonnees.health-data-hub.fr:5001/docs Ex:https://catalogue-metadonnees.health-data-hub.fr:5001/list_documents	arnaud.abreu@health-data-hub.fr

Description :
-> Under development. No metadata available.
https://catalogue-metadonnees.health-data-hub.fr/documentStep1?documentStep1=oscour
https://catalogue-metadonnees.health-data-hub.fr:5001/list_documents

Properties :
Metadata schema mapped to DCAT and extended
https://catalogue-metadonnees.health-data-hub.fr:5001/read_document/oscour
dct_title, skos_altLabel, dct_identifier, dct_description, dcat_keyword, dct_language, dct_creator, dct_publisher, dcat_contactPoint, dct_source, dct_spatial, dct_format, dcat_byteSize, dcat_modified, dct_accrualPeriodicity,

CLASS DATASET :
dct:title, dct:description, dcat:keyword, dcat:contactPoint, dct:publisher, dct:spatial, dct:temporal, dcat:theme, dcat:themeTaxonomy (Class catalogue ressource), dct:accessRights, dct:creator, dct:conforms (Class distribution), foaf:page, dct:accrualPeriodicity, dct:hasVersion, dct:identifier, dct:isReferencedBy, dcat:isVersionOf, dcat:landingPage, dct:language, adms:identifier, dct:provenance, prov:qualifiedAttribution, dcat:qualifiedRelation, dct:relation, dct:issued, dct:modified, adms:sample, dct:source, dcat:spatialResolutionInMeters, dcat:temporalResolution, dct:type, owl:versionInfo, adms:versionNotes, prov:wasGeneratedBy
CLASS DISTRIBUTION :
dcat:distribution, dcat:accessURL, dcatap:availability, dct:description, dct:format, dct:license, dcat:accessService, dcat:byteSize, spdx:checksum, dcat:compressFormat, foaf:page, dcat:downloadURL, odrl:hasPolicy, dct:language, dct:conformsTo, dcat:mediaType, dcat:packageFormat, dct:issued, dct:rights, dcat:spatialResolutionInMeters, adms:status, dcat:temporalResolution, dct:title, dct:modified
Extended properties :
hdhrdf_dataProvenance, hdhrdf_dataCollectionProcess, hdhrdf_linkedResourceVariablesDictionnary, hdhrdf_linkedResourceOperatingProgram, hdhrdf_linkedResourceTrainingProgram, hdhrdf_isDataCollectionOngoing, hdhrdf_sndsPaired, hdhrdf_projectFundingType, hdhrdf_medicalField, hdhrdf_legalDataCategory, hdhrdf_spatialGranularity, hdhrdf_sndsPairingType, hdhrdf_publisherDpo, hdhrdf_externalAccessURL, hdhrdf_hasQualityMeasurementSet, hdhrdf_medicalDataType, hdhrdf_itemsCount, hdhrdf_legalStatus

Health domains : Under development. No metadata available.

- ☐ Health Registries
- ☐ Diseases Registries
- ☐ Drugs Registries
- ☐ Biobanks
- ☐ OMICS
- ☐ Laboratory
- ☐ Health Surveys, Clinical Trials
- ☐ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☐ Health Networks
- ☐ Population Data
- ☐ Epidemiology And Surveillance

Comments :
Under development.

Catalogue : **Data resources catalogue Aineistokatalogi (Finnish national health data catalogue)**			Coverage : Finland Language : FI/SV/EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://aineistokatalogi.fi	In-house developed Termieditori	The Generic Statistical Information Model (GSIM) has	Dedicated API	mari.makinen@thl.fi info@aineistokatalogi.fi

		been used behind the metadata model as information model. The metadata specification is mapped to DDI Lifecycle 3.2 in attribute level.	https://aineistokatalogi.fi/swagger-ui.html Serialisation: JSON Ex: https://aineistokatalogi.fi/api/v1/public/datasets/ca8a40b0-1e54-447f-9bd2-70cfe82296f4	

Description :

Metadata catalogue with cross domain datasets. Data resources catalogue Aineistokatalogi is the national health data catalogue in Finland. All data under the legislation for secondary use of health and social data is described in Aineistokatalogi in a way specified in the Regulation of the Health and Social Data Authority: Data content, concepts and data structured for data descriptions. Also some other organisations are describing their data in the catalogue.

Catalogue is maintained in co-operation by the Finnish Institute for Health and Welfare (THL) and Findata - Finnish Social and Health Data Permit Authority. It was originally developed in co-operation between THL, Statistics Finland, the Data Archives and Sitra in a project during years 2016 – 2018.

List of data resources: - https://aineistokatalogi.fi/catalog/studies
Currently there are approximately 615 data resource descriptions by 47 different data holders.

Main functionalities:

- Search data resources by name.
- Filter/browse data resource descriptions by organisation.
- Search variables by variable name, description, keyword or concept variable.
- Download the variable and code list descriptions of a dataset in a CSV file.

Properties :

List of properties:
In Data resource and Dataset level: title/name, alternative name, abbreviation, description, organisation/publisher, contact information, data resource type (ex: registry data, survey data), terms of use (controlled list), terms of use, further information (free text), links (landing page), observation unit type, number of observation units, universe/Population, rReference date (reference period start date, reference period end date), rRegional coverage, tTarget group, sample size, non-response, data resource life cycle phase, keywords, free keywords, series.

In variable level: Name, technical name, description, concept variable, variable group, keywords, free keywords, reference date (reference period start date, reference period end date), domain type (continuous/categorical), label for missing value, code for missing value, information about data quality, variable data source, data type, data format, observation unit type, variable related instance question, code list (references to external code lists: external identifier, name, description / self-defined code lists: name, description, owner, codes: code, value).

Examples:
Causes of death - research data:
https://aineistokatalogi.fi/catalog/studies/778c33bf-aceb-423f-89d9-e5abb5a0585c/datasets/fcc7f03d-3f83-49df-973d-f2d94b3742a6
Cancer registry:
https://aineistokatalogi.fi/catalog/studies/21085403-7be8-4f93-bf05-231518c642a0/datasets/ca8a40b0-1e54-447f-9bd2-70cfe82296f4

List of variables:
https://aineistokatalogi.fi/api/v1/public/studies/778c33bf-aceb-423f-89d9-e5abb5a0585c/datasets/fcc7f03d-3f83-49df-973d-f2d94b3742a6/instanceVariables.csv?lang=fi&encoding=ISO-8859-15

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [x] Biobanks
- [ ] OMICS
- [x] Laboratory
- [x] Health Surveys, Clinical Trials
- [x] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [x] Population Data
- [x] Epidemiology And Surveillance

Comments :
Part of the Nordic Commons task force for metadata, data quality and semantics.
The project started in 2022 and will end in 2024.
In addition to training and enabling the participating countries in structuring, standardisation and sharing metadata about their health data sources for research, the objective is as well to set up an infrastructure with endpoints for sharing machine readable metadata between the national nodes, and with EHDS.
The scope is both standardised metadata about the data source (data set), metadata about the belonging variables, and the code lists/value sets the relevant variables are based on.

Catalogue : **HealthData.be (Sciensano)**

Coverage : Belgium
Language : NL/FR tagged as EN

URL	Software	Metadata standard	API	Contact information
https://fair.healthdata.be/	DKAN	CKAN + RDF DCAT endpoints (powered by ckanext-dcat) https://extensions.ckan.org/extension/dcat/	https://fair.healthdata.be/api (but link down) Serialisation: JSON Swagger interface: https://fair.healthdata.be/dataset/011bdcee-0f1f-40cd-87b5-e5c41d6def97/api	HealthData - Sciensano fair.healthdata@sciensano.be

Description :
278 metadata records (213 registries)
Surveillance Infection Diseases by National Reference Centres humane microbiology: Burkholderia cepacia complex
API : https://fair.healthdata.be/api/1/metastore/schemas/dataset/items/011bdcee-0f1f-40cd-87b5-e5c41d6def97

Properties :
List of properties: https://extensions.ckan.org/extension/dcat/
Title, publisher, description, identifier, last update (modified), licence, contact, contact email, access level, accrual periodicity, spatial, theme

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [x] Biobanks
- [x] OMICS
- [x] Laboratory
- [x] Health Surveys, Clinical Trials
- [x] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [x] Health Networks
- [x] Population Data
- [x] Epidemiology And Surveillance

Comments :
Language : Nl/Fr (merged in dcat:description) Tagged as En
Annexe documents are provided by DCAT distributions. Not a correct DCAT implementation

Catalogue : **Health Information Portal (Sciensano)**			Coverage : EU Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://www.healthinformationportal.eu/	FAIR DATA POINT + DRUPAL	DCAT 2 And dedicated metadata schema (based on DDI) https://doi.org/10.5281/zenodo.6413408	RESTFUL HATEOAS Serialisation: TURTLE, RDF/XML, JSON-LD Swagger interface: https://fair.healthinformationportal.eu/swagger-ui/index.html#/	PHIRI - Sciensano Miriam.Saso@sciensano.be

Description :
Documentation and user guide for the health information portal : Metadata description: https://doi.org/10.5281/zenodo.6413408

Structure:
Population health data catalogue organised by national nodes.

List of metadata records:
Over 300 datasets reported.
Ex: Austrian Social Health Insurance Data: https://www.healthinformationportal.eu/health-information-sources/austrian-social-health-insurance-data

Properties :
List of properties for the API FAIR DATA POINT:
dct:title @en, dct:title @national language, prov:wasGeneratedBy, dcat:theme, dcat:landingPage, dcat:keyword, dct:description @en, dct:temporal, dct:spatial, dct:language, dct:creator, dct:accrualPeriodicity, dct:accessRights, dct: identifier

List of variables:
Title, alternative title, acronym, type of information, topics, url of the data source, data source url status, free keywords, description, data collection period, governance and legal framework, funding, geo coverage, country, sex, target population, access information, age range from, language, term of data access url, data owners, age range to, updating periodicity, sample size, personal identifier, level of aggregation, contact indo address, linkage possible, terms of data access, contact name, contact email, contact phone number, permanent id of the data source, regulations for data sharing

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [x] Biobanks
- [ ] OMICS
- [ ] Laboratory
- [x] Health Surveys, Clinical Trials
- [ ] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [x] Population Data
- [x] Epidemiology And Surveillance

Comments :
Under development.

Catalogue : **BBMRI-ERIC Directory**			Coverage : World Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://directory.bbmri-eric.eu/	MOLGENIS	MIABIS core 2.0	https://directory.bbmri-eric.eu/api/ (access controlled)	BBMRI-ERIC

	https://molgenis.gitbook.io/molgenis/#why-molgenis https://molgenis.org/	https://github.com/BBMRI-ERIC/miabis		petr.holub@bbmri-eric.eu

Description :
Catalogue of Biobanks
Structure:

- Organisations
- Collections
- Subcollections

Filters: *Diagnosis available, Countries, Collection type, Categories, Material type, Collaboration type, Biobank services, Quality labels, Network, Data type (ex: imaging data, national registries, …)*

Properties :
List of properties:
Title, id, PID, description, collection types, juridical person, contact information, networks

Ex: Biobank-University Hospitals Leuven
https://directory.bbmri-eric.eu/menu/main/app-molgenis-app-biobank-explorer#/biobank/bbmri-eric:ID:BE_B0383

List of variables (collections):
title, id, size, type, sex, materials, data (list of data types), diagnosis

Health domains :

- ☐ Health Registries
- ☑ Diseases Registries
- ☐ Drugs Registries
- ☑ Biobanks
- ☑ OMICS
- ☐ Laboratory
- ☑ Health Surveys, Clinical Trials
- ☑ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☑ Health Networks
- ☐ Population Data
- ☐ Epidemiology And Surveillance

Comments :

Catalogue : **eBrains knowledge graph**			Coverage : EU Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://search.kg.ebrains.eu https://query.kg.ebrains.eu		For the Human Brain Project, we're using the openMINDSschemes to structure the metadata. https://github.com/HumanBrainProject/openMINDS However, the technology of the EBRAINS KG is model-agnostic and can be used to handle any kind of	Custom RestAPI incl. a unique, descriptive query mechanism Serialisation: JSON-LD as its default representation of metadata and is stored in a multimodal database (ArangoDB)	? thibault.griez@ebrains.eu, martin.telefont@ebrains.eu

		linked metadata - either schema-less or schema-oriented.		

Description :

Documentation: https://docs.kg.ebrains.eu/
908 metadata records (free access 747, under embargo 127, controlled access 32, restricted access 2) :
https://search.kg.ebrains.eu/?category=Dataset

Structure: Project, dataset, model, (meta)Data model, Software, Contributor

Filter: *modality (ex: anatomy, species (ex: Homo spaciens), data accessibility (ex: restricted access), methods (ex: spatial registration), Keywords (ex: brain mapping), content types (image/tiff)*

Properties :

List of properties:
title (+version), ethics assessment, project, custodians, description, contact, homepage, experimental approach, technique, keywords, data descriptor (documentation), how to cite, get data, publications, specimen

Ex:
Large prospective, epidemiological-based neuroimaging study of children and adolescents followed from pre-natal life
https://search.kg.ebrains.eu/instances/35883bd4-46f2-4dc7-9f9b-283c5b0e217e

Health domains :

- ☐ Health Registries
- ☑ Diseases Registries
- ☑ Drugs Registries
- ☐ Biobanks
- ☐ OMICS
- ☑ Laboratory
- ☑ Health Surveys, Clinical Trials
- ☑ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☐ Health Networks
- ☐ Population Data
- ☐ Epidemiology And Surveillance

Comments :

Catalogue : **Norwegian health metadata catalogue**			Coverage : Norway Language : EN/NO	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://helsedata.no/en/		Dedicated metadata schema Nasjonal spesifikasjon for metadata om helsedata (Mapped to DCAT)	RDF file on git-hub: https://github.com/Directorate-for-E-Health/Metadata-Felles-datakatalog/blob/main/distribution_prod_rdf_to_fdk.txt	Norwegian Directorate of eHealth Truls.Korsgaard@ehelse.no

Description :

Services: Data source explorer, Variable explorer, Overview open data, Create your own variable list, Application service/form,
Structure: Data sources with drill down to variable lists with quite rich metadata, codeLists/value sets and statistics.

Content:
Administrative registries: 2
National health registries: 17

National medical quality registries: 49
National and regional health studies: 15
Other registries: 2

Explore data sources: https://helsedata.no/en/data-sources/?page=1&sort=0
Filters: Access level, Type of data source,, Category, Data holders, Data source
Example: Cancer Registry of Norway: https://helsedata.no/en/forvaltere/cancer-registry-of-norway/cancer-registry-of-norway/
Drill down to variable explorer

Properties :

Data source, sub sources and data records:
List of properties:
title, publisher, theme, temporal coverage, variables, open data (description), criteria for data access, apply for access to data, contact, landing page
Filters:
Type of data source, Data source, Category, different variable themes/categories, Period for data collection.

"Drill down" to detailed metadata on variable level (lists of variables with metadata) on 28 data sources.
Variable explorer: https://helsedata.no/en/variables/?page=search
Example: Cancer Registry of Norway: https://helsedata.no/en/variables/?datakilde=K_KREG&page=search (312 variables)

In the variable explorer you can create variable lists, share them, and download them and use them in applications.

Metadata shared with:
The Norwegian Cross Sectorial Metadata Catalogue: https://data.norge.no/
Example: https://data.norge.no/datasets/caf3d50c-1d5c-3909-860b-0f491b24177b

The official portal for European data: https://data.europa.eu/en
Example: https://data.europa.eu/data/datasets/https-helsedata-no-no-forvaltere-kreftregisteret-kreftregisteret-?locale=en

Example: Cancer Registry of Norway: https://helsedata.no/en/variables/?datakilde=K_KREG&page=search (312 variables)
"*Variables: Here you can create variable lists, download them and use them in applications. Information about variables is currently only available in Norwegian.*"

Health domains :

- [x] Health Registries
- [x] Diseases Registries
- [x] Drugs Registries
- [x] Biobanks
- [ ] OMICS
- [x] Laboratory
- [x] Health Surveys, Clinical Trials
- [x] Tests Diagnostic & Clinical
- [x] Health Care Services, Facilities or Quality
- [ ] Health Networks
- [x] Population Data
- [ ] Epidemiology And Surveillance

Comments :

Part of the Nordic Commons task force for metadata, data quality and semantics.
The project started in 2022 and will end in 2024.
In addition to training and enabling the participating countries in structuring, standardisation and sharing metadata about their health data sources for research, the objective is as well to set up an infrastructure with endpoints for sharing machine readable metadata between the national nodes, and with EHDS.
The scope is both standardised metadata about the data source (data set), metadata about the belonging variables, and the code lists/value sets the relevant variables are based on.

Catalogue : **European Centre for Disease Prevention and Control (ECDC)**			Coverage : International Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://geoportal.ecdc.europa.eu/geocatalogue/	ASP.NET Core ESRI ArcGIS Server	Dedicated standard	https://gis.ecdc.europa.eu/portal/sharing/rest	Manager institution of the national catalogue ionut.sava@ecdc.europa.eu

Description :

The catalogue is structured around disease health topics to present a graphical representation of the geographical distribution of the information.

The ECDC Geoportal is envisioned as a one-stop shop for GIS-related health data for internal as well as external users. At its core, it is a catalogue with map making functionality used for discovering GIS resources providing a uniform way of access to the underlying services such as data sets, already prepared layer or applications.

The current scope of ECDC Geoportal includes:
Web application: ECDC Map Maker (EMMa). An application helping users visualising a dataset containing geospatial information on a map.

Web application: Administration interface based on the Esri Portal for ArcGIS server out-of-the box func-tionality

A tool for power users to manage the ECDC Geoportal's catalogue

Hybrid layer for spatial data (supported by Microsoft SQL Server database)

SQL based geodatabase source or reference spatial data fetched into the Map Maker application through the map service and used for joining it with non-spatial data when creating a map layer from an imported dataset

Web services: Library of ArcGIS Map Services

Properties :

Extensive list of properties depending of the disease or the disease health topic.

Health domains : No catalogue available. Based on metadata records available from website.

- ☐ Health Registries
- ☑ Diseases Registries
- ☐ Drugs Registries
- ☐ Biobanks
- ☐ OMICS
- ☐ Laboratory
- ☐ Health Surveys, Clinical Trials
- ☐ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☐ Health Networks
- ☑ Population Data
- ☑ Epidemiology And Surveillance

Comments :

Catalogue : **Federal Institute for drugs and medical devices (BFARM)**			Coverage : DE Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://www.healthdatalab.de	/	/	/	katharina.schneider@bfarm.de

Description :
No catalogue available, yet.

Health domains :

- ☐ Health Registries
- ☐ Diseases Registries
- ☐ Drugs Registries
- ☐ Biobanks
- ☐ OMICS
- ☐ Laboratory
- ☐ Health Surveys, Clinical Trials
- ☐ Tests Diagnostic & Clinical
- ☐ Health Care Services, Facilities or Quality
- ☐ Health Networks
- ☐ Population Data
- ☐ Epidemiology And Surveillance

Comments :

Other countries outside the consortium:

Catalogue : **HDRUK gateway**			Coverage : UK Language : EN	
URL	**Software**	**Metadata standard**	**API**	**Contact information**
https://healthdatagateway.org		Dedicated metadata schema	Serialisation: JSON Swagger: https://api.www.healthdatagateway.org/api-docs/	Health Data Research UK

Description :
Descriptive Metadata specification: https://zenodo.org/record/5902743#.Y_4UHC-l2gR

List of metadata records: 801 records
Ex: National Joint Registry - Revision Hip Replacement dataset
https://web.www.healthdatagateway.org/dataset/4b8286e2-c4e4-4db4-8fd9-5f1c8b4e56de
Filters: Publisher, Keywords, Phenotype, Coverage, Provenance, Access, Format and standards, Data utility, Technical Metadata, Commercial use

Structure: datasets, tools, data uses, courses, papers, dataset requests, registered users, searches last month

Properties :
List of properties:
https://github.com/HDRUK/schemata
https://github.com/HDRUK/schemata/blob/master/docs/dataset/2.0.0/distribution/Descriptive%20Metadata%20Specification%20v2.0.0%2012.8.2020%20.pdf

Health domains :

- ☑ Health Registries
- ☑ Diseases Registries
- ☑ Drugs Registries
- ☑ Biobanks
- ☑ OMICS
- ☑ Laboratory
- ☑ Health Surveys, Clinical Trials
- ☑ Tests Diagnostic & Clinical
- ☑ Health Care Services, Facilities or Quality
- ☑ Health Networks
- ☑ Population Data
- ☑ Epidemiology And Surveillance

Comments :
Outside of the consortium.

Summary of the results

Catalogue Name	URL	Coverage	Software	Metadata Standard	API
European Data Portal	https://data.europa.eu/en	International	CKAN + CKAN-EDP extension + Virtuoso (Triplestore)	CDS + DCAT	Yes
EMA EU metadata catalogue of data sources and studies	https://data-catalogue.molgeniscloud.org/catalogue/catalogue/#/networks-catalogue	EU	Molgenis + Drupal	Dedicated standard	Yes
European Joint Programme Rare diseases Virtual Platform	https://vp.ejprarediseases.org https://index.vp.ejprarediseases.org/	EU	Fair Data Point	DCAT	Yes
Orphanet	https://www.orpha.net	International		Bioschemas	Yes
Orphadata	https://www.orphadata.com	EU	Fair Data Point	DCAT	Yes
Lists of variables for health registers	https://island.is/en/gagnasofn-embaettis-landlaeknis/eydublod-breytulistar	Iceland		Mapped to DCAT	
Metaregistar	https://metaregistar.gov.hr/metareg/html/javno_pocetna.xhtml	Croatia	SDURDD (Under development)	Dedicated standard Under development	
Danish national metadata catalogue	www.datalandkort.rsyd.dk	Denmark	Dateado	For catalogues and tables → DCAT-AP-DK For variables and metadata → ISO 11179(-7)	Yes
RUT - Register utiliser tool	https://rut.registerforskning.se/	Sweden		GSIM	
French Health Data Hub metadata catalogue	https://catalogue-metadonnees.health-data-hub.fr/	France	Jena Apache Triple Store	DCAT	Yes
Data resources catalogue Aineistokatalogi	https://aineistokatalogi.fi	Finland		GSIM/DDI Lifecycle	Yes

Catalogue Name	URL	Coverage	Software	Metadata Standard	API
(Finnish national health data catalogue)					
Sciensano HealthData.be	https://fair.healthdata.be/	BE	DKAN	CKAN+RDF DCAT	Yes
Health Information Portal	https://www.healthinformationportal.eu/	EU	Fair Data Point + Drupal	DCAT2	Yes
BBMRI-ERIC Directory	https://directory.bbmri-eric.eu/	World	Molgenis	MIABIS core 2.0	Yes
eBrains knowledge graph	https://docs.kg.ebrains.eu/	EU		openMINDS	Yes
Norwegian health metadata catalogue	https://helsedata.no/en/	Norway		Dedicated standard mapped to DCAT	Yes
European Centre for Disease Prevention and Control	https://geoportal.ecdc.europa.eu/geocatalogue/	International	ASP.NET Core ESRI ArcGIS Server	Dedicated standard	Yes
Federal Institute for drugs and medical devices (BFARM)	https://www.healthdatalab.de	DE			
HDRUK gateway	https://healthdatagateway.org	UK		Dedicated standard	Yes

Next steps

According to the results of the landscape analysis presented in the above factsheets we can conclude that 7 out of 18 metadata catalogues are using, mapping or are compliant with the DCAT-AP metadata standard. The rest of the catalogues are either using MIABIS, DDI, GSIM or have their own dedicated metadata standard. However, in most of the cases where we see the use of DCAT-AP, it is clear that there is a need to complement this standard by adding dedicated health data properties and controlled vocabularies in a collaborative way. Therefore, we decided to start designing the Health extension to the DCAT-AP metadata standard.

Technical Working Groups to design the Health DCAT-AP extension

As mentioned in the introduction, the input received from the second and third form that we created and disseminated using the sandbox, will support the design of this extension. Data providers that created a metadata record for their datasets using the second form and proposed new properties for the Health extension, will be invited in recurrent technical working groups.

The result of these working groups will be the specifications for the Health extension to the DCAT-AP metadata standard. These specifications will determine the “required”, “optional” and “recommended” properties that will need to be filled in in a metadata record template.

The data collected from the above forms will also allow the WP6 TWG to design a validation testing plan to address them before publishing the extension. Based on the decisions made by the TWG, a SCHACL[14] shape file will be uploaded in the FAIR DATA POINT, HealthDCAT-AP records will be generated for the pilot use cases and the Health DCAT-AP extension will be tested with real search queries. Moreover, based on a series of defined use cases, SPARQL queries could illustrate the benefit of the Health DCAT-AP extension.

The future Health DCAT-AP will provide a common specification for describing health-related datasets in Europe. At a first stage, it will enable the exchange of descriptions of datasets among the data catalogues participating in the pilot HealthData@EU project and the central services platform established by the European Commission.The health DCAT-AP will enable metadata records to meet the specific discoverability needs of sensitive health data and will define a common language between the National catalogues and the central services.

[14] https://www.itb.ec.europa.eu/shacl/shacl/upload