



M5.5 Draft guideline for data user navigating the catalogue

Step-by-step guide for navigating the dataset catalogue as a data user

TEHDAS2 – Second Joint Action Towards the European Health Data Space

21 April 2026

Co-funded by
the European Union



0 Document info

Disclaimer

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.

0.1 Authors

Author(s)	Organisation
Anna Niemeyer	Technology and Methods Platform for Networked Medical Research e.V., Germany
Maria Ahlsén	Karolinska Institutet, Sweden
Beatriz Barros	Sciensano, Belgium
Alexander Leander Knudsen	Danish Health Data Authority, Denmark
Ana Martin-Moreno	Ministry of Health, Spain
Antonis Stylianides	NeHA, Cyprus

Acknowledgements

The authors would like to express their sincere gratitude to Matthias Löbe (IMISE, University of Leipzig, Germany) his comprehensive and valuable contributions, as well as to him and the entire review board for their careful and thorough editorial review of this user manual. Special thanks go to all external interview partners for their invaluable contributions to the interviews conducted in the course of this project. Particular thanks are likewise due to Zdenek Gütter (Ministry of Health, Czech Republic), Danielle Welter (National Data Service, Luxembourg) and Nienke Schutte (Sciensano, Belgium), whose critical observations and constructive suggestions have substantially improved both the rigour and the clarity of the present work. The authors remain, of course, solely responsible for any errors or omissions that may persist.

0.2 Keywords

Keywords	TEHDAS2, Joint Action, Health Data, European Health Data Space
-----------------	--

0.3 Document history

Date	Version	Editor	Change	Status
17/12/2025	0.1	Anna Niemeyer	Initial document creation	Draft
06/02/2026	0.2	Anna Niemeyer and all contributors	First draft for review board	Draft
14/03/2026	0.3	Anna Niemeyer and all contributors	Draft for consortium review	Draft
17/04/2026	0.4	Anna Niemeyer	Draft for PSG approval	Draft
21/04/2026	0.5	Anna Niemeyer	Document to be submitted for public consultation	

Accepted in Project Steering Group on 21.04.2026.

Copyright Notice

Copyright © 2024 TEHDAS2 Consortium Partners. All rights reserved. For more information on the project, please see www.tehdas.eu.

Contents

1 Executive summary.....	3
2 List of abbreviations.....	4
3 Introduction.....	5
3.1 User Journey.....	7
3.1.1 HealthData@EU Central Platform: Functions and significance.....	8
3.2 What can be seen on the HealthData@EU Central Platform website?.....	9
4 What is the catalogue?.....	13
4.1 Nature and function of the catalogues.....	13
4.2 What is not in the EU dataset catalogue?.....	14
4.3 How to access the catalogue?.....	14
4.3.1 The main menu of the HealthData@EU Central Platform.....	15
5 How to search for datasets?.....	16
5.1 Search Functionality.....	16
5.1.1 Basic keyword search.....	16
5.1.2 Filtering.....	16
5.1.3 Sorting.....	18
5.1.4 SPARQL queries (advanced search).....	18
6 How to understand a Dataset record?.....	19
6.1 Accessing Dataset Details.....	20
6.2 Understanding the Dataset Information.....	21
6.2.1 What is this dataset about?.....	21
6.2.2 Who and what does the data cover?.....	22
6.2.3 Who provides and manages the dataset?.....	22
6.2.4 What standards and classifications are used?.....	23
6.2.5 How current and maintained is the dataset?.....	23
6.2.6 How is this dataset related to other resources?.....	23
6.2.7 What are the conditions for access and use?.....	24
6.2.8 Where can I find additional technical or supporting information?.....	24
6.2.9 What information is available about the variables?.....	25
6.3 Next steps: how do I request access to the dataset?.....	25
7 Basket Functionality.....	26
7.1 Adding Datasets.....	26
7.2 Basket Management.....	26
7.3 Next Steps.....	27
8 Troubleshooting.....	27
8.1 Common search issues.....	27
8.2 Browser compatibility problems.....	27
8.3 Support contact information.....	28
9 References.....	28
10 Annexes.....	29
10.1 Methodology.....	29
10.2 User Journey.....	31
10.3 Use case scenarios.....	32
10.3.1 Use Case - (Academic) Healthcare Professionals.....	32
10.3.2 Use Case - Policy Makers/Institutional Users.....	33
10.3.3 Use Case - Academic Researchers / Data Scientists.....	33
10.3.4 Use Case - Industry/Business.....	34

1 Executive summary

This guideline, formulated within the TEHDAS2 framework, provides a comprehensive orientation for navigating the HealthData@EU Central Platform's EU Dataset Catalogue. Its primary objective is to facilitate the cross-border discovery and appraisal of health data, thereby supporting research, innovation, and evidence-based policymaking as envisaged under the European Health Data Space (EHDS) Regulation.

The document focuses specifically on the pre-authentication workflow, delineating a systematic user journey for identifying and evaluating datasets prior to formal login. Central to this infrastructure is the use of standardised metadata compliant with the HealthDCAT-AP model, which ensures interoperability and comparability across the federated network of national catalogues.

To ensure efficient data discovery, the guideline explicates several search modalities:

- **Search Mechanisms:** Users are introduced to basic keyword retrieval for initial exploration and advanced, property-based SPARQL queries for technical users requiring precise, repeatable results.
- **Iterative Refinement:** These search tools are augmented by dynamic filtering mechanisms—mapped directly to EHDS data categories, population characteristics, and coding systems—alongside multi-faceted sorting options to prioritise relevance and currency.

Furthermore, the guideline serves as a manual for interpreting dataset records as structured decision aids. It enables potential data users to evaluate essential metadata parameters, including:

- **Provenance and Governance:** Identifying the organisations responsible for data creation and the relevant Health Data Access Bodies (HDABs) overseeing the application process.
- **Content and Coverage:** Assessing geographic scope, temporal resolution, and the specific categories of electronic health data present.
- **Technical Interoperability:** Reviewing the standards, terminologies (such as ICD-10 or SNOMED CT), and data models employed within the dataset.
- **Variable-Level Insight:** Utilizing non-disclosive artefacts, such as data dictionaries and proxy datasets, to assess the feasibility of research hypotheses without necessitating access to primary health data.

To bridge the transition from discovery to formal application, the 'basket' functionality is presented as a preparatory shortlisting mechanism. This feature allows for the consolidation and comparative analysis of candidate datasets prior to authentication, without initiating formal interactions with data holders.

The development of this guidance adhered to a rigorous user-centred methodology, incorporating semi-structured expert interviews with diverse stakeholder groups and iterative review cycles. This evidence-informed approach ensures that the platform's technical complexities are translated into accessible, operational instructions for a broad professional audience.

2 List of abbreviations

Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
Art	Article
CDM	Common Data Model
DICOM	Digital Imaging and Communications in Medicine
DPA	Data Processing Agreement
DRG	Diagnosis Related Groups
DSA	Data Sharing Agreement
DUA	Data Use Agreement
EEHRxF	European Electronic Health Record Exchange Format
EHD	Electronic Health Data
EHDS / EHDSR	European Health Data Space / European Health Data Space Regulation
eHDSI	eHealth Digital Service Infrastructure
EMA	European Medicines Agency
EU	European Union
ETL	Extract, Transform, and Load
GDPR	General Data Protection Regulation
HDAB	Health Data Access Body
HDL	High-Density Lipoprotein
HL7® FHIR®	Health Level Seven, Fast Healthcare Interoperability Resources standard
IDMP	Identification of medicinal products
ICD	International Classification of Diseases
ISO/TC	International Organization for Standardization /Technical committees
JCA	Joint Controllership Agreement
LLM	Large Language Model
LOINC	Logical Observation Identifiers Names and Codes
MedDRA	Medical Dictionary for Regulatory Activities
NCP	National Contact Point for secondary use
OMOP	Observational Medical Outcomes Partnership
Rec	Recital
RWD	Real World Data
RWE	Real World Evidence
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms
SPARQL	SPARQL Protocol and RDF Query Language
SPE	Secure Processing Environment
TEHDAS2	Joint Action Towards a European Health Data Space 2

3 Introduction

Advancing health data use in the European Health Union

As part of the European Health Union, the European Union (EU) is advancing the use of health data for secondary purposes, including research, innovation and policymaking. Smooth and secure access to data will drive the development of new treatments and medicines and optimise resource utilisation - all with the overarching goal of improving the health of citizens across Europe.

TEHDAS2, the second joint action Towards the European Health Data Space, represents a significant step forward in this vision. The project will develop guidelines and technical specifications to facilitate smooth cross-border use of health data, and support data holders, data users and the new HDABs (health data access bodies) in fulfilling their responsibilities and obligations outlined in the European Health Data Space (EHDS) regulation.

TEHDAS2 focuses on several critical aspects of health data use.

- **Data discovery:** findability and availability of health data to ensure that it is accessible for secondary purposes.
- **Data access:** developing harmonised access procedures and establishing standardised approaches for granting data access across Member States.
- **Secure processing environment:** defining technical specifications for environments where sensitive health data can be processed safely.
- **Citizen-centric obligations:** providing guidance on fulfilling obligations to citizens, such as communicating significant research findings that impact their health, informing them about research outcomes and ensuring transparency in how their data is used.
- **Collaboration models:** developing guidance on collaboration and guidelines on fees and penalties as well as third country and international access to data.

TEHDAS2 will contribute to harmonised implementation of the EHDS regulation through the concrete guidelines and technical specifications. Some of these documents and resources will also provide input to implementing acts of the regulation. Hence, the joint action will increase the preparedness for the EHDS implementation and lead to better coordination of Member States' joint efforts towards the secondary use of health data, while also reducing fragmentation in policies and practices related to secondary use.

This document should be understood as an expert opinion and guidance document developed within the TEHDAS2 framework, reflecting technical and expert input from the project partners. It is not legally binding and does not constitute a formal guideline or technical specification under the European Health Data Space.

This document does not represent the position of the European Commission.

Legally binding and enforceable requirements under the European Health Data Space are laid down in Regulation (EU) 2025/327 and, where applicable, in Implementing Acts adopted by the European Commission, within the limits of the empowerments provided by the Regulation.

Purpose and scope of the user guide

This document serves as a high-level orientation aid designed to assist users in navigating the EU central catalogue and interpreting available information without necessitating prior technical expertise or specialist knowledge in metadata. It is structured around three primary objectives: providing first-time users with an initial orientation regarding the platform's organisation and the nature of discoverable information; facilitating efficient searching by explaining how to use keywords, filters, and SPARQL queries to locate and compare datasets; and supporting the assessment and preparation phase.

By following these instructions, users can effectively determine if a dataset is suitable for specific research, innovation, or policy projects and prepare for necessary follow-up actions (such as starting a data access application or data request through the appropriate process).

The scope of this manual is explicitly restricted to the public search area (pre-authentication workflow), covering all functionalities accessible to the public without a prior login.

Specifically, it provides instructions on navigating the catalogue, utilising search and filtering mechanisms, evaluating metadata on dataset cards, and employing the "add to basket" function for shortlisting.

The guide deliberately excludes technical or restricted processes, such as the EU Login authentication procedure, the management of data descriptions by data holders, and any aspects of actual data processing.

Further references

For guidance on actions following the initial discovery phase—such as formal data access applications and regulatory compliance—please refer to separate accompanying documents as the "[D6.2 Guideline for data users on good application and access practice](#)". This guideline provides systematic instructions for good application and access practice, fulfilling regulatory obligations, and complying with data access and usage requirements. Another important resource is the official [HealthData@EU central platform user manual of the European Commission](#); [HealthData@EU central platform user manual of the European Commission](#) and the [HealthData@EU documentation](#).

Finally, to have more information on the metadata model HealthDCAT-AP, we refer to the TEHDAS2 guideline on dataset descriptions.¹

Understanding the framework and the infrastructure of the EHDS

The EHDS regulation establishes a framework for the secondary use of health data, requiring that datasets falling within the categories specified in Article 51 be appropriately described by their respective data holders. These descriptions consist of structured metadata—rather than the raw health data itself—designed to make resources findable and understandable for users assessing relevance prior to initiating access applications.

To ensure European-wide consistency and comparability, all dataset records, i.e. the structured metadata, are expected to align with the HealthDCAT Application Profile (**HealthDCAT-AP**) metadata model. The infrastructure follows a federated approach,

¹ Available in: [Results - Tehdas](#)

meaning that while the actual health data remains securely managed in its country of origin, descriptive metadata is made accessible across the Union via a single interface: the **HealthData@EU Central Platform**.

The ecosystem accessible via the platform is organised into two hierarchical levels:

- Each member state is required to maintain a **national dataset catalogue** (article 77). These national catalogues collect and publish dataset descriptions from national data holders, such as public authorities, registries, and other organisations holding health data. They provide a structured overview of what datasets exist at national level and act as the main entry point for compiling dataset records. National data catalogues are linked to the EU central data catalogue through the relevant **national contact points** (NCPs). This guideline is focused on guidance on how to navigate the EU dataset catalogue only as described below. For guidance on navigating national dataset catalogue please refer to the provider of such service at national level.
- (1) The **EU dataset catalogue** contains structured and harmonised metadata descriptions of health-related datasets made available by health data holders across the EU. These metadata are provided by member states via their coordinating HDAB and the NCP, EU institutions and bodies, authorised third countries, and authorized participants, such as European research infrastructures.

None of the catalogues stores the underlying health data itself; instead, they provide detailed information that enables users to understand what data is available, what characteristics it has, where and by whom it is stored, and under what conditions it can be accessed.

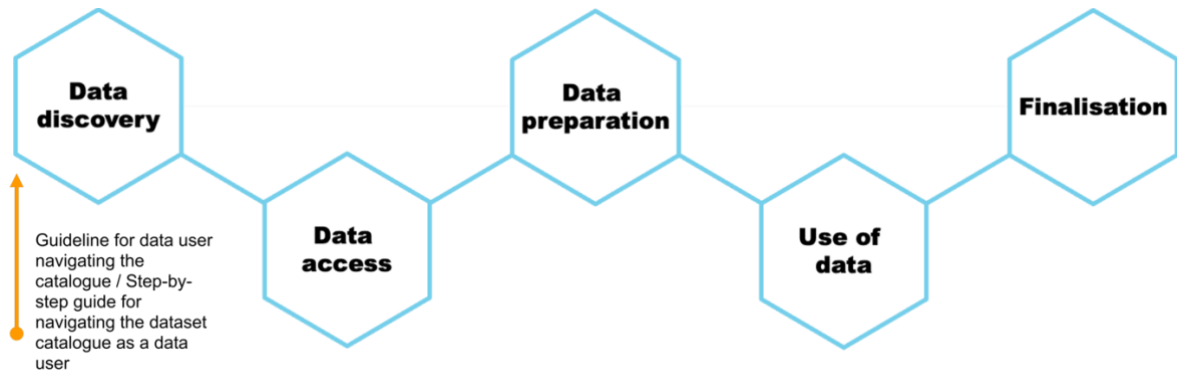
The federated governance model

HealthData@EU infrastructure operates on a federated principle: health data remains securely decentralised within the respective member states. Conversely, descriptive metadata is aggregated via NCPs and synchronised with the EU dataset catalogue. This architecture ensures cross-border discoverability through a unified interface while strictly maintaining national sovereignty over data governance and security.

3.1 User Journey

EHDS user journey consists of five main phases: data discovery, data access, data preparation, use of data and finalisation. The focus of this guideline is exclusively on the first part of the user journey—the discovery of data—as indicated in **Error! Reference source not found.** by the orange arrow.

Figure 1: EHDS user journey and positioning of the user guide (this document)



For a detailed description of the user journey, see the appendix User Journey10.2

3.1.1 HealthData@EU Central Platform: Functions and significance

The [HealthData@EU Central Platform](#) helps to make health datasets for secondary use more visible and comparable across countries and supports cross-border use by providing a single point of discovery.

- **Discovery Tool:** This EU-wide catalogue lets you find datasets useful for research, innovation, or policy. With a single interface, it enables cross-border searches to locate health datasets not listed in national catalogues. You can also use the SPARQL search tool or download documentation as needed.
- **Metadata focus:** The HealthData@EU Central Platform is designed to publish and display harmonised dataset descriptions (metadata). It presents structured information about datasets, such as their content, coverage, responsible organisation, and high-level access conditions. This approach supports transparency and findability while respecting data protection and governance requirements.
- **Central access:** In addition to help you find datasets; the platform supports the next steps by providing information that can be used to prepare and guide a data access request (see chapter 6.2.7 for more information). The information contained in a dataset description contributes to a more coordinated and efficient process for cross-border data use, by making it easier to identify who is responsible for a dataset and how access can be requested through the relevant mechanisms.

When you use the navigation process, you'll get: a filtered and sorted list of datasets, a metadata record for each dataset, a list of variables from the sample distribution if available, and resources you can export like documentation or SPARQL endpoints.

In addition to the EU Dataset Catalogue, the Healthdata@EU Central Platform, provides several tools, accessible through the main navigation, that enhance catalogue navigation and support users in understanding dataset content. See chapter 4.3.1 for a detailed description.

Important notice

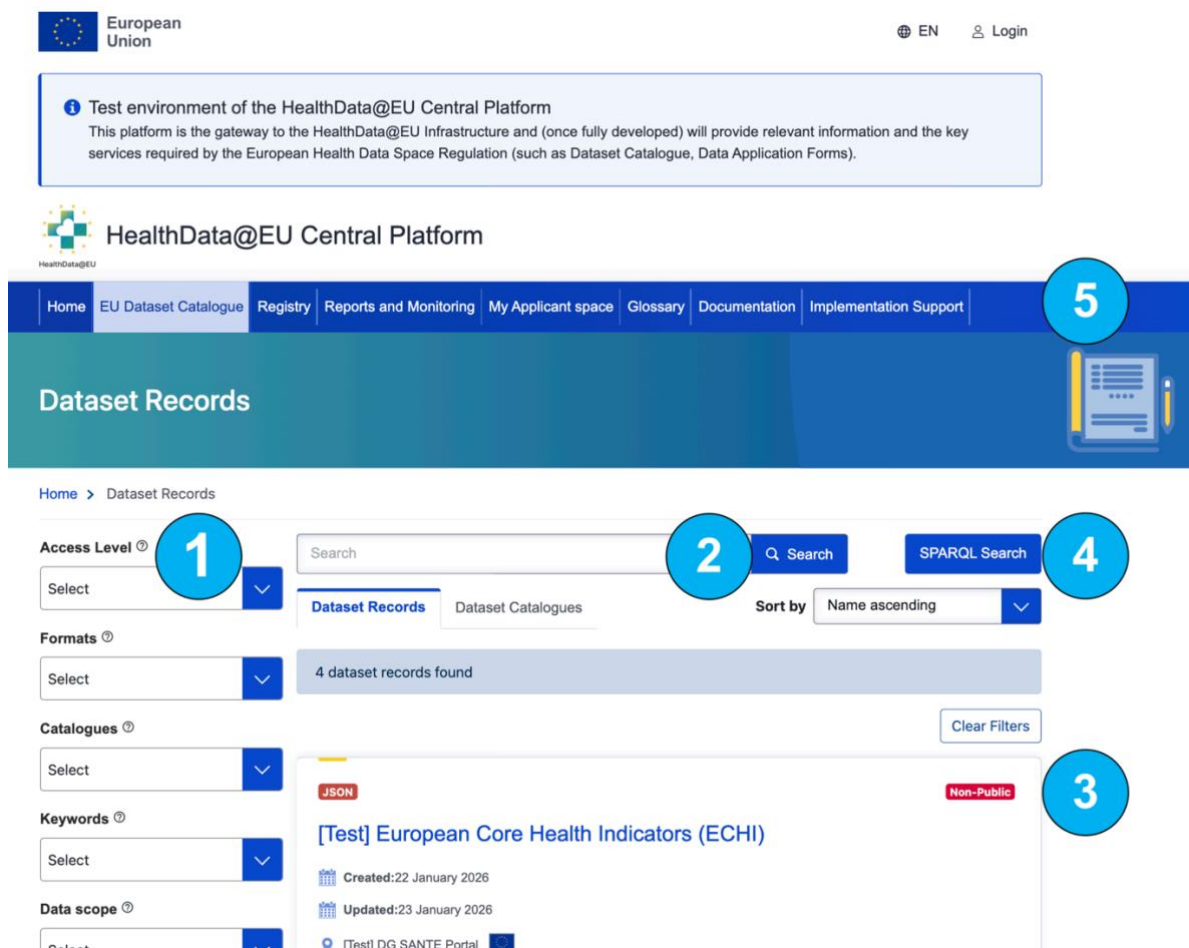
Healthdata@EU Central Platform and the EHDS catalogue continues to evolve in terms of data description elements and other functionalities in 2026 and 2027 (further refinements are planned until go live in 2029), in line with the release plan (delivery timeline) provided on the

HealthData@EU central platform website (see URL in chapter 4.3) and may also reflect new inputs from users testing and commenting on it. While every care is taken in order to ensure the correlation between this guideline and the implemented elements provided in the Central Platform website, there could be some discrepancies due to delay in updating this guideline.

3.2 What can be seen on the HealthData@EU Central Platform website?

This section contains screenshots of the most important functions and views of the Healthdata@EU Central Platform, release 6, as of January 2026. Some elements or areas of the screenshots are numbered, with each number labelled and briefly described below. The functions and mechanisms of the portal are explained in the rest of this user guide, with reference to these numbers for illustration purposes.

Figure 2: Start page of the EU Dataset Catalogue



Nr.	Name	Where is it described?
1	Filter, derived from the HealthDCAT-AP metadata model	See chapter 5.1.2
2	Search Field	See chapter 5.1.1

Nr.	Name	Where is it described?
3	Dataset Record	See chapter 6
4	SPARQL Search	See chapter 5.1.4
5	Implementation Support	See chapter 4.3

Figure 3: Dataset record detail page

Home > Dataset Records > [Test] European Core Health Indicators (ECHI)

Dataset Record Theme

[Test] European Core Health Indicators (ECHI)

[Test] DG SANTE Portal Updated: 23 January 2026 Access Level: Non-Public

7

Keywords (9) Theme (1) Categories (4) Health Themes (2) Health codes (1) Personal Data (0)

- Description +
- Purpose +
- Provenance +
- Population Coverage +

8

Dataset Record feed Export Dataset Record Cite Embed

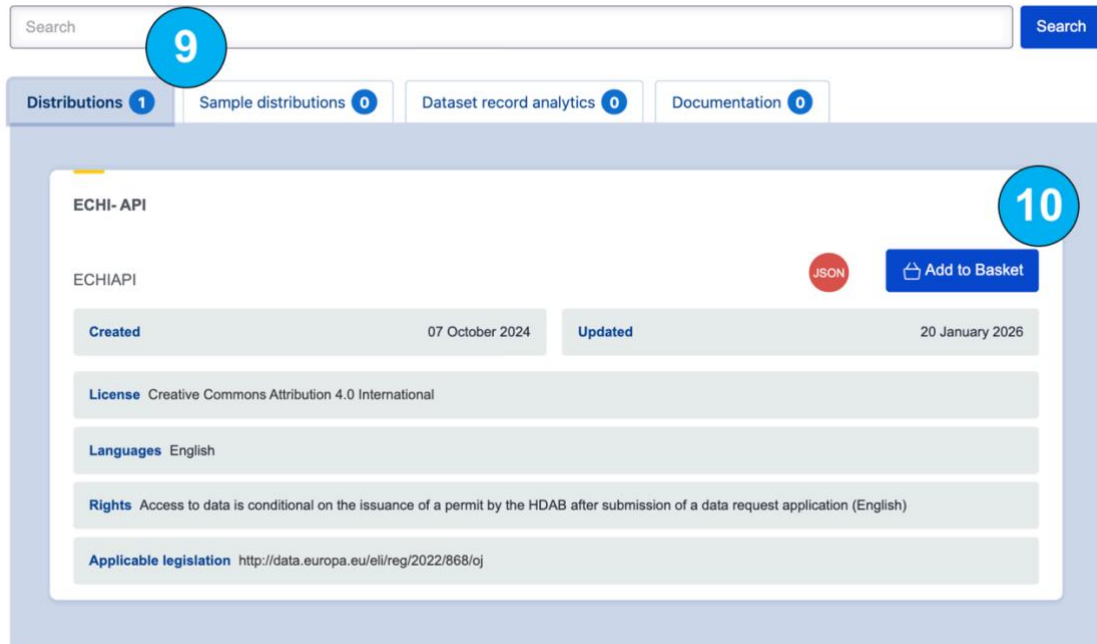
Created: 22 January 2026 Updated: 23 January 2026

Acronym: above urls Retention period: 01 January 2026 - 31 December 2034

Nr.	Name	Where is it described?
6	Access Level	See chapter 6.2.7
7	Dataset Record details on content	See chapter 6.2.1

Nr.	Name	Where is it described?
8	Information on the status of the dataset record and other functions	See chapter 6.2.5

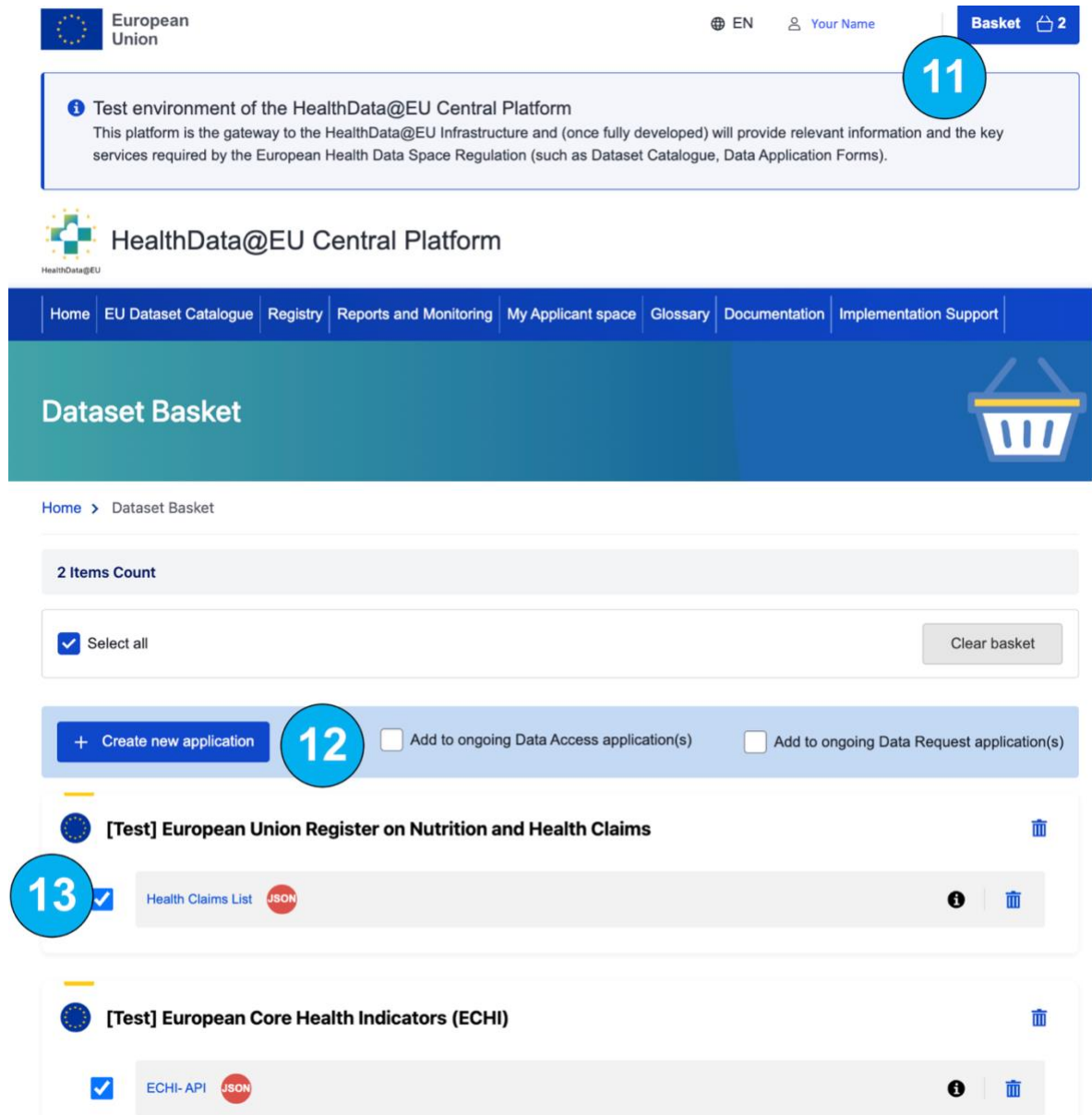
Figure 4: Dataset Information on distributions and basket



Nr.	Name	Where is it described?
9	Distributions and sample distributions	See chapter 7.1
10	Add to basket	See chapter 7.2

You will need to authenticate and successfully log in via [EU Login](#), the European Commission's user authentication service, to the Healthdata@EU Central Platform before you can see the basket and the selected dataset records. The following screenshot shows what to expect.

Figure 5: Dataset Basket (only available after EU Login)



Nr.	Name	Where is it described?
11	Basket containing selected records (after login)	See chapter 7
12	List of selected dataset records and functions for requesting or applying for data	See chapter 7.1
13	Add more datasets to the application	See chapter 7.2

4 What is the catalogue?

4.1 Nature and function of the catalogues

Essential to navigating the infrastructure is the distinction between a 'dataset'—defined as a structured collection of health data—and a 'dataset record', which represents the specific metadata entry within the catalogue describing that resource.

Conceptual Framework: Datasets and Records

It is imperative for users to distinguish between a 'dataset' and a 'dataset record'. A dataset constitutes a structured aggregate of electronic health data, whereas a dataset record represents a discrete metadata entry within the catalogue that describes a specific dataset. Analogous to a database architecture, the dataset serves as the overarching table, whilst the record functions as an individual entry detailing its provenance, scope, and access conditions

EU dataset catalogue

As mentioned above, the [EU dataset catalogue](#) forms part of the central HealthData@EU Central Platform, supporting the secondary use of health data in the EU through a common metadata catalogue and an EU-wide application mechanism. The catalogue does not store data itself but provides descriptions of datasets and access conditions. For a description of the platform's functions and role, see chapter 3.1.1 .

National dataset catalogue

The national dataset catalogue is a mandatory digital business capability under the EHDS framework and plays a central role in enabling the secondary use of health data. It serves as the primary national-level mechanism through which health datasets can be identified, described, and discovered for reuse for permitted secondary use purposes, such as research, innovation, public health, and policymaking.

The national dataset catalogue contains standardised dataset descriptions of health-related datasets made available by health data holders, including public authorities, healthcare providers, registries, and other authorised organisations. While the catalogue does not store the health data itself, it provides comprehensive information on the existence, characteristics, and access conditions of datasets that may be reused under EHDS rules.

The national dataset catalogue is intended to function as an authoritative national source of metadata and to ensure transparency and consistency in how health data availability is presented within each member state. Importantly, it is designed to interoperate with the EU dataset catalogue, using one metadata model (HealthDCAT-AP) and technical specifications defined at EU level.

Each member state's national catalogue will feed metadata into the EU datasets catalogue, enabling cross-border discoverability of datasets through a single union-wide interface.

Article 51 of the EHDS regulation determines *the categories of data eligible for secondary use, which are expected to be described through dataset metadata* for further explanation see [D5.1 Guideline for data holders on data description](#).

4.2 What is not in the EU dataset catalogue?

The EU Dataset Catalogue is strictly a metadata directory and does not function as a data repository, access platform, or secure processing environment. Its primary utility lies in providing standardised records that describe available datasets and their respective access conditions, rather than hosting the datasets themselves.

To ensure clarity for the user, it is essential to distinguish between the discovery tool and the broader ecosystem. The catalogue does **not** encompass the following:

- **Raw Health Data:** Neither personal, pseudonymised, nor anonymised health data are stored within the catalogue.
- **Direct Data Access:** Functionalities such as data downloads, API access to raw data, or integrated analytical tools are excluded.
- **Application Management:** The formal assessment, processing, and administrative storage of data access applications, contracts, or decisions occur outside the catalogue's remit.
- **Research Dissemination:** The publication of research results, statistical outputs, or derived datasets is not handled within this interface.
- **Quality and Legal Services:** The catalogue does not provide quality assurance for the underlying data content, nor does it offer binding legal interpretations of the EHDS Regulation or the GDPR.

While these functionalities are excluded from the catalogue, most operational tasks—including application tracking, secure data processing, and the publication of secondary use results—are facilitated through the secure sections of the overarching **HealthData@EU Central Platform**. Thus, the catalogue serves as a navigational signpost, whereas the platform provides the infrastructure for subsequent data utilisation.

4.3 How to access the catalogue?

When you navigate to the EHDS official entry point via the URL:

<https://acceptance.data.health.europa.eu>.

From the platform, you can choose to:

- Navigate the **HealthData@EU Central Platform**, which supports secure cross-border access to health data for research, innovation, and policymaking.
- Access the **EU database for the registration of EHR systems and wellness applications**, which provides structured electronic health records to support improved clinical data exchange. Please see chapter III of the [EHDS regulation](#) for further details.

Alternatively, you can use the direct link to the HealthData@EU Central Platform:

<https://acceptance.data.health.europa.eu/healthdata-central-platform>

You may select one of the 27 available interface languages using the language switcher icon:

As a user you begin by entering the [platform's homepage](#), where the [EU dataset catalogue](#) is accessible through the main navigation menu. The catalogue serves as the central entry point for discovering, filtering and understanding datasets made available under the EHDS framework. It provides a structured overview of available dataset records, including metadata, descriptions, and links to related resources. See **Error! Reference source not found.**

4.3.1 The main menu of the HealthData@EU Central Platform

On the homepage of the HealthData@EU Central Platform, the main menu provides the following options:

- [Home](#) – displays the HealthData@EU Central Platform homepage, including the development timeline and an overview of key features and functionalities implemented across platform releases.
- [EU Dataset Catalogue](#) – presents the published dataset catalogues and their associated dataset records. You will encounter several key components that guide your navigation and dataset discovery:
 - **Filters panel** – allows refinement of dataset results using free text and metadata fields. You may filter by domain, publisher, temporal coverage, or other metadata attributes.
 - **Sorting options** – datasets can be sorted by title (ascending or descending), relevance, creation date, or modification date, enabling users to prioritise the most relevant or recently updated datasets.
 - **Pagination controls** – you can adjust the number of items displayed per page and navigate through multiple pages of results.
- [Registry](#) – includes sections presenting data access and request applications with their respective status, measures taken against data holders or users for non-compliance, and results from the secondary use of data:
 - **Data Access Applications:** Displays a list of all data access applications submitted to date. It also displays a list of data permits issued and data access applications that have been refused.
 - **Data Request Applications:** Displays a list of all data request applications submitted to date. It also displays a list of data request approved and data request applications that have been refused.
 - **Non-compliance measures:** Shows a list of all measures taken against data holders or data users due to non-compliance.
 - **Results from secondary use:** Shows results derived from secondary use of data.
- [Reports and Monitoring](#) – provides insights into catalogue growth, dataset distribution, and metadata completeness. This section presents statistics in a graphical dashboard as well as the biennial reports published on the HealthData@EU Central Platform, including detailed information.
- [My Applicant space](#) – logged in users (for further information on how to get an EU-Login see [EU Login User Guide](#)) can take part of their active and draft applications, as well as final decisions (positive or negative) transmitted by NCPs on behalf of HDABs. After

logging in via EU Login, you will find a list of your own applications and requests, along with the corresponding decisions.

- [Glossary](#) – provides definitions of key terms used in the HealthData@EU Central platform and concepts related to EHDS which are used in the creation of data access applications
- [Documentation](#) – provides a list of topics offering a comprehensive overview of the HealthData@EU Central Platform and its underlying infrastructure.
- [Implementation Support](#) – the tools Dataset Description Assistant and Metadata model Explorer to be used by data holders.
 - **Dataset Description Assistant:** Provides a web-based assistant that enables users to create new dataset descriptions in accordance with the HealthDCAT-AP specification.
 - **Metadata Model Explorer:** Displays a glossary of the main HealthDCAT-AP classes and their associated properties.

5 How to search for datasets?

5.1 Search Functionality

The EHDS central catalogue provides different ways for you to search for datasets. For most users, a simple keyword search is the best starting point. For advanced users and technical teams, the catalogue provides a [SPARQL](#) query option, which enables more structured searches across the catalogue’s machine-readable metadata. This section explains both approaches and helps you choose the most suitable one for your needs.

5.1.1 Basic keyword search

Users may initiate discovery via a keyword search by inputting terms related to clinical conditions, data categories, or demographics in the search bar. The system subsequently queries metadata fields, including titles and descriptions, to identify relevant records. The results list shows datasets that match your search terms, allowing you to quickly open dataset records and explore them in more detail. If your search returns many results, you can refine your search by adjusting your keywords or by using filters to narrow down the results list. Keyword search is particularly useful when you are exploring a topic for the first time and when you want to quickly identify datasets that may be relevant before reviewing details such as coverage, data holder information, or access conditions.

5.1.2 Filtering

In addition to keyword search, the catalogue provides a set of filtering mechanisms that allow you to narrow down the list of datasets based on specific characteristics. Filters are available in the dedicated filtering panel alongside the search results and are designed to help you progressively refine your search as your needs become clearer. Filters correspond directly to metadata fields available in dataset records. This means that each filter reflects information that data holders have formally described as part of the dataset metadata, supporting consistency and comparability across datasets.

The filtering interface is dynamic; selected criteria automatically ascend to the top of the panel to facilitate efficient oversight and subsequent adjustment of the active parameters.

Examples of commonly used filters

The catalogue offers a wide range of filters, and the exact set may evolve over time. You do not need to use all filters. Instead, you can select those that best match the characteristics you are looking for. The following examples illustrate some of the most commonly used filters and how they can support your search.

- **Categories (of health data):** The filter logic based on *Categories*—as defined in Article 51 of the EHDS Regulation—allows you to refine datasets based on the type of health data. This helps you identify whether a dataset contains, for example, electronic health records, registry data, administrative data, or other predefined categories. Using this filter is particularly useful when you already know which type of data is relevant for your use case.
- **Formats and distribution types:** The *Formats* filter indicates how a dataset is made available, such as through specific file formats (for example CSV or JSON) or services. This can help you assess whether the dataset is technically compatible with your intended use or processing environment.
- **Coding systems and data models:** Filters such as *Coding systems* and *Data models* allow you to identify datasets that use specific standards, such as ICD-10, SNOMED CT, or particular data models. These filters are especially relevant if you need datasets that align with established classification systems or interoperability standards used in your organisation or project.
- **Population characteristics and size:** Filters related to population characteristics, such as *Age range* or *Population size*, help you understand the scale and scope of a dataset. For example, population size can give an early indication of how many records or individuals are included, which can be useful when assessing whether a dataset is suitable for statistical analysis or large-scale studies.
- **License and access levels:** Filters related to access level, licensing, or the responsible HDAB allow you to narrow results based on how datasets can be accessed and under which legal or organisational conditions. This can help you focus on datasets that match your eligibility or intended use.
- **Keywords:** Keywords are a short list of specific words or phrases assigned to a dataset to describe its main content and topics. The Keywords filter allows you to refine your search using these predefined terms already used in dataset records. You can either select keywords from the available list or enter text to match relevant terms. Using keywords can help you quickly identify datasets related to specific subjects, especially when different datasets use consistent terminology to describe similar content.

Multi-filter application and logic

Filters are intended to be combined. You can start with a broad search and progressively apply filters as you learn more about the available datasets. Each time you add a filter, the catalogue narrows the results to dataset descriptions that meet all selected criteria at the same time. In practical terms, this means that filters work cumulatively: a dataset description must match one filter and the other to remain in the results list. As you apply filters, the

interface shows which criteria are currently active. This allows you to easily review, adjust, or remove individual filters at any point during your search. If your search returns very few or no results, you may consider removing one or more filters to broaden the scope again. Conversely, if your search returns too many dataset descriptions, adding additional filters can help you focus on datasets that better match your needs.

The catalogue does not require you to use any specific filter. Instead, filters are meant to support an iterative exploration process, helping you balance between too broad and too narrow results. The set of available filters may expand or change as the catalogue evolves, so the filtering panel should be seen as a flexible tool for refining and tailoring your search rather than a fixed checklist.

5.1.3 Sorting

When searching through keyword or filters, the catalogue allows you to sort the list of dataset descriptions to control the order in which results are displayed. Sorting can help you review results more efficiently and find the datasets that are most relevant to your needs.

Available sorting options include:

- **Name:** Sort datasets alphabetically by title, either in ascending (A–Z) or descending (Z–A) order.
- **Last created:** Sort datasets based on the date they were added to the catalogue, with the newest or oldest first.
- **Last modified:** Sort datasets based on the most recent update to their metadata.
- **Relevance:** Sort results according to how closely they match your search keywords. This is the default option for keyword searches and helps highlight datasets that are most likely to match your query.

Sorting affects how search results are presented, but it does not change the datasets included in the results list. You can combine sorting with filters to refine and prioritise results based on the characteristics most relevant to your research, innovation, or policy purpose. Using sorting effectively can help you quickly access recently added or updated datasets or prioritise datasets most relevant to your search terms.

5.1.4 SPARQL queries (advanced search)

In addition to keyword searching, the catalogue offers the possibility to query dataset records using SPARQL. SPARQL is a query language used to search and retrieve information from machine-readable metadata. Because the EHDS central catalogue is designed to be machine-readable and interoperable, SPARQL queries can be used to perform more precise and systematic searches across dataset records. While the keyword search looks for terms appearing across indexed text fields (such as titles, descriptions, or keywords), SPARQL allows you to query directly on specific metadata properties.

Importantly, SPARQL is not limited to the filtering facets available in the user interface. It can be used to query any metadata property that is part of the dataset record and exposed in the catalogue’s RDF representation. This means that users can target particular properties (for example, geographic coverage, temporal coverage, standards used, access rights, or EHDS data category) rather than searching for a term across all indexed fields.

Although all metadata elements are indexed and searchable through the general search functionality, combining specific properties with precise query conditions in SPARQL often produces more controlled and reproducible results. This is particularly useful when querying across multiple catalogues simultaneously or when constructing automated, machine-to-machine discovery workflows.

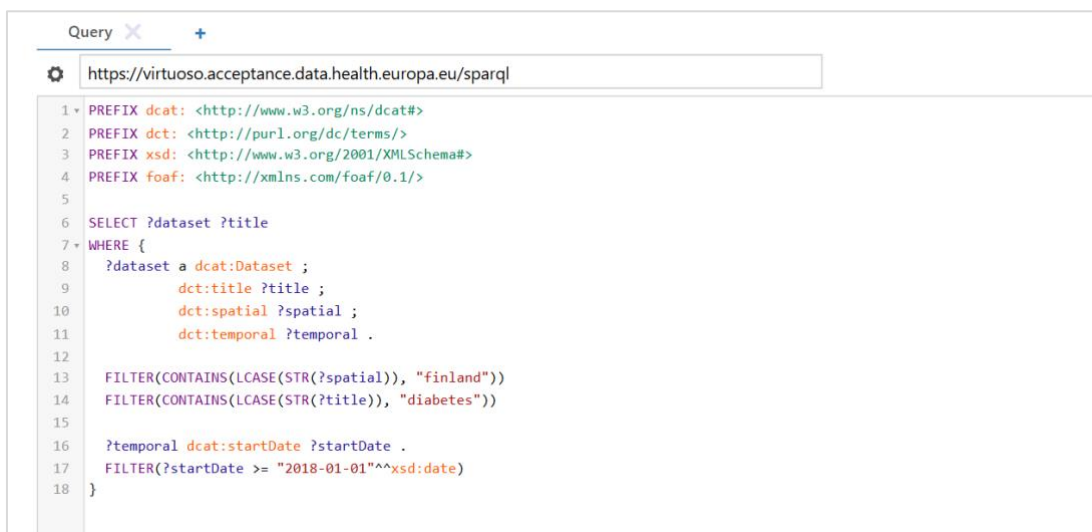
Example

With a basic keyword search, you might type “*diabetes Finland 2018*”. This approach is effective for discovery, but the results depend on how and where those words appear in the dataset record description. Using SPARQL, you can express the same request more explicitly by asking for datasets that:

- cover Finland as geographic coverage,
- include diabetes as a keyword or description and
- have temporal coverage starting from 2018 onwards.

An illustrative SPARQL query expressing this logic could look like the following:

Figure 6: SPARQL Research Example



```

Query X +
https://virtuoso.acceptance.data.health.europa.eu/sparql
1 PREFIX dcat: <http://www.w3.org/ns/dcat#>
2 PREFIX dct: <http://purl.org/dc/terms/>
3 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
4 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
5
6 SELECT ?dataset ?title
7 WHERE {
8   ?dataset a dcat:Dataset ;
9           dct:title ?title ;
10          dct:spatial ?spatial ;
11          dct:temporal ?temporal .
12
13   FILTER(CONTAINS(LCASE(STR(?spatial)), "finland"))
14   FILTER(CONTAINS(LCASE(STR(?title)), "diabetes"))
15
16   ?temporal dcat:startDate ?startDate .
17   FILTER(?startDate >= "2018-01-01"^^xsd:date)
18 }

```

This example shows how SPARQL allows you to define exact conditions on dataset metadata, rather than relying only on word matching. Using SPARQL can support advanced use cases such as:

- Identifying datasets that meet multiple criteria at the same time (for example, specific coverage and dataset characteristics),
- Retrieving selected metadata fields in a structured format, and
- Supporting automated or repeated queries across catalogue records.

6 How to understand a Dataset record?

At this point, it is important to understand how the terms dataset and dataset record are used in the EU and national dataset catalogues. A dataset is a structured collection of electronic

health data, as defined in article 2(2) (ab) of the EHDS regulation. You can think of a dataset as a table that organises related information in a standardised way. A dataset record refers to the metadata description of that dataset within the catalogue. It does not represent an individual data entry or row, but rather a structured description of the dataset as a whole.

The EU dataset catalogue aggregates metadata descriptions of datasets made available through national catalogues and other sources.

Each national dataset catalogue contains descriptions of the national health datasets in that country. In other words, each record in the national catalogue table describes one of the datasets. In the following, we will discuss precisely these records in the national catalogues.

What is the purpose of a dataset record?

A dataset record provides detailed information on certain pre-defined characteristics of a dataset that may be available for secondary use within the EHDS. The dataset record does not contain the data itself. Instead, it provides structured information that helps users understand what the dataset is about, who is responsible for it, and whether it may be relevant for their needs. The main purpose of a dataset record is to support informed decision-making. By reading a dataset record, users should be able to:

- Understand the general content and scope of a dataset;
- Assess whether the dataset could be useful for a specific research, policy or innovation purpose;
- Identify the organisation responsible for the dataset;
- Know where to go next to ask questions or request access.

Importantly, note that dataset descriptions are designed to be read by a wide range of users, including researchers, policy makers, innovators, and other professionals. You do not need technical knowledge or expertise in data standards to understand the information presented. Each dataset description follows a common structure, so that users can easily compare different datasets and find key information in a consistent way across the catalogue. While the level of detail may vary from one dataset to another, the overall goal remains the same: to provide clear and transparent information about datasets available through the EHDS framework. In the following sections, this guide walks you through the different parts of a dataset description, explaining what kind of information you can expect to find and how it can help you decide whether to explore a dataset further.

6.1 Accessing Dataset Details

You can access a dataset record by selecting a dataset from the catalogue overview or from the search results. When you click on a dataset title or dataset card, you are taken to the dataset detail view.

The dataset detail view presents information in a structured and consistent way across the catalogue. You can open and review dataset records freely, without prior EU Login. At any point, you can return to the catalogue overview or your search results to explore other datasets. Dataset records are intended to support iterative exploration, allowing you to review several datasets before deciding on next steps. Once you open a dataset record, you can use the information provided to understand what the dataset is about, who is responsible for

it, what population, time period, and geographic scope it covers, and how to proceed if you want to request access or obtain further information.

The next sections explain the different types of information you can find in a dataset record and how they can help you make informed decisions.

- **Information on dataset history and content:** These elements describe what the dataset is about, how it was created, and what it contains. They help you understand the scope, origin, population, time frame, and general context of the dataset.
- **Contact and responsibility information:** These elements identify the entities responsible for the dataset, including the data holder and relevant contact points. They help you understand who manages the dataset and who to contact for further information or clarification.
- **Information on data standards in use:** These elements describe the standards, classifications, and terminologies used to structure the data. They support transparency and help users assess technical compatibility or potential for reuse.
- **Information on relations with other datasets and resources:** These elements show how the dataset is linked to other datasets, publications, or documentation. They help place the dataset in a broader context and identify related resources.
- **Information on variables and data structure:** These elements provide insight into how the dataset is organised, including which variables are included and how the data is structured. This helps users assess whether the dataset is suitable for their analytical or technical needs.

6.2 Understanding the Dataset Information

When you open a dataset record, you will find different types of information describing the dataset from multiple perspectives. The information presented in a dataset record can be grouped into a small number of logical categories. These categories are aligned with the common metadata model used in the EHDS catalogue and help you focus on the aspects that matter most for your use case. Please note that the categories presented in this document are used to explain the dataset information in logical groupings. They do not necessarily reflect the exact structure or order in which the information appears in the catalogue interface, where metadata elements may be displayed in a different arrangement. In the descriptions below, the exact terms used in the dataset records are highlighted in **bold** to help you locate the corresponding information in the catalogue.

6.2.1 What is this dataset about?

At the top of the dataset record, you will find information that explains the dataset's content, purpose, and background. Together, these elements provide the context needed to correctly understand and interpret the dataset.

You will typically find:

- A **title** that summarises the dataset in a short, descriptive way, often indicating the type of data, geographic scope, and time period.

- A **description** that explains what data is included, how it was collected, and its key characteristics (for example, whether it contains hospital discharge records, prescriptions, registry data, or survey responses).
- Information on the original **purpose** of data collection, such as routine patient care, reimbursement, public health surveillance, or research.
- A **provenance** description, explaining where the data comes from and how it was created. This may include the original data sources (such as electronic health records or registries), the processes involved in producing the dataset, and any major transformations (for example, data cleaning, harmonisation, or calculation of derived variables).
- **Keywords** and **thematic annotations** that highlight the main topics covered by the dataset and support searching and filtering in the catalogue.

Taken together, these elements help you understand not only what the dataset contains, but also why it exists and how it came to be, which is essential for assessing its relevance and appropriate use.

6.2.2 Who and what does the data cover?

Dataset records include information that helps you understand the scope of the data and the population it represents. This section is especially important for determining whether the dataset matches your intended use.

You may find information such as:

- **Geographic coverage (currently referred to as “Spatial”)**, indicating the country or region represented by the data.
- **Temporal coverage**, showing the time period during which data was collected or is available.
- The **EHDS data category**, indicating which category of electronic health data (as defined in the EHDS Regulation in Article 51) is present in the dataset. This helps you understand the type of health data covered by the dataset.
- Information on **personal data**, describing whether the dataset includes personal information and which broad types of personal data are present. This may refer to direct identifiers (such as names), or indirect identifiers and sensitive attributes (such as age, gender, health conditions, or education level) without listing specific variables.
- High-level **indicators** of the dataset’s scale, such as the approximate number of individuals represented or the number of records.
- Information on age ranges, temporal resolution, or spatial resolution, where relevant.

This information helps you assess whether the dataset aligns with your research question, policy analysis, or other secondary use, and gives an early indication of the dataset’s sensitivity and access conditions.

6.2.3 Who provides and manages the dataset?

Provenance and governance metadata identify the entities responsible for data creation and management. Such information ensures transparency and assists users in evaluating the institutional context prior to pursuing formal access:

- The organisation that **created** or collected the data, indicating the dataset's origin.
- The **publisher** or data holder managing the availability and the access of the data.
- The **HDAB**, overseeing the access to the data under the EHDS framework.
- The **contact information** to identify the right person for questions or support before an access request.
- The **custodian**, representing the party that assumes responsibility and accountability for the data and ensures its proper care and maintenance

6.2.4 What standards and classifications are used?

Information about standards and classifications is particularly important if you are looking for data that follows a specific data model or terminology, or if you plan to compare or combine datasets. In the dataset record, you will typically find two key types of information:

- Standards the dataset **conforms to**: This information explains how the dataset is structured or represented. It may indicate that the dataset follows a recognised data model or standard, such as HL7 FHIR, OMOP Common Data Model, or CDISC define.xml.
- **Coding systems** used within the dataset: This information describes which terminologies or classifications are used to encode values inside the dataset according to the data holder. Examples include ICD-10 for disease classification, SNOMED CT for clinical concepts, or DRGs for hospital activity.

Together, these two elements help you assess interoperability and reuse potential: one tells you how the dataset is structured overall, the other tells you how concepts and values are coded within the data.

6.2.5 How current and maintained is the dataset?

Dataset records may include indicators that help you understand whether the dataset is regularly updated or represents a fixed snapshot. You may see:

- Information on **update frequency** (for example, daily, monthly, or annually) of the data in the dataset
- Information on **retention periods**, specifying how long the dataset will remain available for secondary use.
- **Version information**, including version identifiers and notes describing what has changed between versions.
- Dates indicating when the dataset was **created** or **last modified**

This helps you assess whether the dataset meets your needs in terms of timeliness and continuity.

6.2.6 How is this dataset related to other resources?

Dataset records may show how a dataset is connected to other datasets, documents, or online resources. This can include:

- Links to **related datasets** or newer versions.
- References to publications, reports, or studies that use or describe the dataset.
- A landing page with additional background information provided by the data holder.

These links help you place the dataset in context and explore supporting material if needed.

6.2.7 What are the conditions for access and use?

Dataset records help you understand, at an early stage, whether a dataset is openly available or subject to access restrictions, and what kind of process may be required to obtain access. One of the key elements you will see in this section is the access rights indicator. This property uses standardised values to describe the level of accessibility of the dataset. In the catalogue, you may encounter three main access rights categories:

- **Public:** The dataset is openly accessible and can be used without restrictions. Data in this category can typically be accessed directly, without the need for a formal access request. This applies mainly to datasets that do not contain personal or sensitive information. Examples are High-Value Datasets (HVD) according to the EU Open Data Directive.
- **Restricted:** The dataset is not openly available but may be accessed under specific conditions. These conditions can include eligibility requirements, purpose limitations, or procedural steps defined by the data holder or relevant authority. Users may need to submit a request or meet certain criteria before access is granted. Examples are health data protected by intellectual property rights or trade secrets, or confidential statistical data.
- **Non-public:** The dataset is not publicly accessible, and access is subject to formal authorisation. This is the most common access level for datasets containing personal electronic health data under the EHDS Regulation. Access typically requires a data access application reviewed and authorised by the responsible HDAB.

In addition to access rights, the dataset record may include:

- **Applicable legislation**, indicating the legal frameworks governing the dataset (for example, national law, the EHDS Regulation, or the GDPR).
- **Licensing or usage information**, where relevant, describing general conditions for reuse.

This information is intended to support early decision-making. Detailed access procedures, eligibility checks, and authorisation steps are handled outside the catalogue through the formal data access process managed by the relevant HDAB.

6.2.8 Where can I find additional technical or supporting information?

In some cases, dataset records include links to additional resources that help users better understand the dataset without providing direct access to the data. These may include:

- **Dashboards or summaries** derived from the dataset.
- **Technical or methodological reports**, such as data quality or completeness descriptions.

- Indicators of **API availability** or other technical services for advanced users.

6.2.9 What information is available about the variables?

As you scroll further down the dataset record, you may find a section dedicated to providing information about the variables contained in the dataset, including, where applicable, the variable name, definition, data type, coding system or value domain. This section is aimed at helping you understand the dataset's internal structure without needing access to the actual data, by providing a representative overview of how the data is organised across variables and whether it is suitable for your research hypothesis.

In this section, you may find:

- A **data dictionary**, also called codebook or variable catalogue, describing the dataset's variables, formats, classifications, taxonomies, or code lists.
- A **proxy dataset**, which simulates the structure and key characteristics of the original dataset but does not contain any personal data or information that could reasonably lead to re-identification.

This section is particularly useful if you want to understand how the dataset is organised before deciding whether to submit a data access application or a data request.

6.3 Next steps: how do I request access to the dataset?

Once you have reviewed a dataset record and determined that the dataset may be relevant for your purpose, the next step is to prepare a data access request. The dataset record is designed to help you reach this decision before entering any formal application process.

At the bottom of the dataset record, you will find the **Distribution** section. This part marks the shift from the discovery phase to getting ready for access.

The distribution indicates how the dataset can be made available if access is granted. From a user perspective, this is also where you can add the dataset to your basket in order to start a data access application. In the distribution section, you may find:

- A **description** of the available distribution(s) for the dataset;
- The **file format** in which the data can be provided (for example, CSV or other structured formats);

This information helps you understand, at a high level, how data would be available (in a SPE) available if your data access application is approved. By adding a distribution to your basket, you signal that you wish to request access to that dataset. This action does not grant access to the data itself but allows you to proceed with the formal data access application process managed by the relevant HDAB.

For step-by-step guidance on what happens after adding a dataset to your basket, including how to complete and submit a data access application, please consult the next chapter of this guide.

7 Basket Functionality

The 'basket' serves as a preliminary shortlisting mechanism during the discovery phase. It facilitates the consolidation of potential datasets for comparative analysis prior to formal authentication, without initiating interactions with data holders or regulatory bodies.

The basket: A preparatory appraisal tool

The 'basket' functionality facilitates the systematic shortlisting and comparative analysis of datasets during the discovery phase. Users may consolidate multiple dataset distributions to evaluate their collective suitability for a specific research objective. Crucially, adding resources to the basket is a non-binding action that does not initiate formal regulatory interactions, thereby allowing for iterative refinement of the research scope after EU Login authentication

7.1 Adding Datasets

Once you logged in you can explore datasets in the EU dataset catalogue and directly add those distributions of interest to your basket. This step is designed to help you organise selection of potential datasets before moving to the next phase of interaction with the platform. The basket is populated using the “Add to basket” button, which is made available within the dataset discovery interface in the dataset distribution tab as shown in the example in **Error! Reference source not found., point 10**, see page 15.

You can add multiple references to datasets to your basket during the same session as shown in **Error! Reference source not found., point 13**. This makes it easier to:

- Compare similar descriptions of datasets across countries or data sources
- Review alternative datasets descriptions covering the same topic
- Keep track of datasets that may be useful for future requests.

7.2 Basket Management

The basket overview is accessible at any time through a dedicated interface element visible during browsing. It provides a consolidated view of all selected datasets.

Within the basket overview, users can review key information associated with each selected dataset, such as titles and high-level metadata. This supports reassessment of relevance and alignment with the user’s objectives prior to authentication.

Users may modify the content of the basket at any time before proceeding further as seen in **Error! Reference source not found..** Typical actions include:

- Removing references to datasets that are no longer relevant
- Refining the overall selection set.

Accessing the basket overview

Open the basket overview to see all the references to datasets you have collected. This view provides a clear snapshot of your selection, showing key information from each dataset so you can compare and reflect without returning to individual dataset pages in the catalogue.

Reviewing selected datasets

Use the basket overview to:

- Recall why each reference to dataset was added
- Compare datasets descriptions conceptually, side by side
- Identify which datasets remain relevant to your goals

This step helps you refine your focus before taking any further action.

Modifying your selection

You can adjust your basket contents at any time:

- Remove datasets records that are no longer relevant
- Make quantity adjustments or version selections where applicable

These changes let your basket evolve as your understanding of the catalogue develops.

7.3 Next Steps

Once your basket reflects the datasets you want to pursue—assuming you are already logged in—you can submit data access requests. These steps involve legal, organisational and technical considerations and are covered in a separate guide.

At this stage, the basket serves as your curated shortlist, helping you approach post-basket workflows with a clear focus and organised selection.

8 Troubleshooting

This chapter provides high-level guidance on common issues that users may encounter while interacting with the HealthData@EU Central Platform, particularly during dataset discovery and search activities. The aim is to support effective navigation and interpretation of search results.

8.1 Common search issues

Users may occasionally experience difficulties when searching for datasets, such as receiving no results, an unexpectedly high number of results, or results that appear only partially relevant. These situations often arise from the use of very broad or very narrow search terms, spelling variations, or the interaction of multiple filters.

It is recommended that users review their search criteria carefully, confirm the terminology used, and progressively refine or broaden queries as needed. The platform's search functionality is designed to support iterative exploration rather than one-step retrieval.

8.2 Browser compatibility problems

The HealthData@EU Central Platform is designed to operate on modern, standards-compliant web browsers. Issues may occur if outdated browser versions are used or if certain browser settings restrict script execution or data storage. Users experiencing display or interaction problems are encouraged to verify that their browser is up to date, disable

restrictive extensions temporarily, refresh the page and confirm that essential browser features are enabled.

8.3 Support contact information

When issues cannot be resolved through basic troubleshooting, users may seek assistance through the official support channels of the HealthData@EU Central Platform. Support contact information is made available directly on the platform interface and in the official user documentation.

Users are encouraged to consult existing guidance materials before contacting support, as many common questions are addressed through published documentation. Support services focus on platform usage and navigation rather than on data access decisions or authorisation outcomes.

For direct support, users may contact the platform support team via:

- Email: SANTE-HEALTHDATA-IT@ec.europa.eu
- Telephone: 00800 6789 1011

9 References

1. European Commission: Directorate-General for Health and Food Safety, HealthData@EU central platform – Open-source release 5 – User manual, Publications Office of the European Union, 2025, <https://data.europa.eu/doi/10.2875/7834692>
2. European Commission: Directorate-General for Health and Food Safety, HealthData@EU central platform – Open-source release 6 – User manual, Publications Office of the European Union, 2026, <https://data.europa.eu/doi/10.2875/7092418>

10 Annexes

Annex number	Annex title
10.1	Methodology
10.2	User journey
10.3	Use case scenarios

10.1 Methodology

The contributors participated according to their promised commitments, ensuring a collaborative and thorough development process. The following steps were conducted:

Expert Interviews

To provide a thorough and clear User Guide, a series of 11 semi-structured expert interviews with representatives from all potential EHDS data-user groups as potential EHDS data users were conducted. The interviews were conducted between September and November 2025 and based on Release 5 of the HealthData@EU Central Platform.

The aim of the interviews was to identify common questions and problems encountered by potential users when using the data portal to search for suitable data sets and to create a user-oriented quick start guide for the portal.

The expert interviews followed a semi-structured interview guide and were designed to last between 30 and 60 minutes. The core element was a "think-aloud" protocol in which the participants operated a test environment of the catalogue via screen sharing. The participants navigated through illustrative data sets, applied filters, examined metadata and simulated the selection of data sets for the shopping basket.

The interviewers conducted the interviews in their respective national languages to avoid interpretation inaccuracies caused by varying levels of English proficiency and recorded them on video.

The interviews were then evaluated by the respective interviewers according to a predefined analysis scheme, and a report was prepared in English for each interview. The reports were then compared and a general review was derived to identify actionable recommendations across four domains: onboarding and guidance, user interface and interaction design, search and filter logic, and metadata and semantics. The analysis was based on the dimensions of discoverability, conceptual clarity, operational clarity and system feedback.

Further activities

- Workshop – As an introduction to the EU dataset catalogue an initial workshop with the experts from DG SANTE were performed in June 2025.
- Working meetings – Regular working meetings were conducted to discuss and outline the key components and structure of the user guide. Initially on a biweekly basis and then weekly from the end of November 2025.

- Responsibilities for the individual thematic sections were assigned based on available expertise. The responsible partners drew up an initial draft, which was then developed and agreed upon by all participants through repeated commentary and discussion and finished by a designated contributor.
- Artificial Intelligence (AI) tools were employed to support the development of this document, particularly Microsoft Copilot and DeepL. Copilot was used for summary and suggestive purposes. All output generated by Copilot was reviewed, edited, and finalised by the contributors. None of the text in the document was independently written by AI or LLM. DeepL was used to improve the English text and to overcome any linguistic barriers. The suggested improvements were reviewed and subject-specific consistency was ensured.
- Consultations with DG SANTE – Four meetings with representatives from DG SANTE were organised to ensure alignment and to gather expert feedback.

Conventions and terminology

To ensure the user-friendliness of this guide, the document is based on established standards and conventions, which are outlined below:

- **Language:** The guide is written in introductory and non-technical language. This ensures that the content is understandable to a wide audience, especially people with language skills at levels B2 to C1 of the Common European Framework of Reference (CEFR). Technical terms and complex concepts are explained clearly and comprehensively to increase accessibility.
- **Address:** The active voice is used throughout the guide, addressing you directly as the user. Phrases such as ‘You can...’ are the norm. This promotes direct communication and facilitates understanding of the processes described.
- **Vocabulary:** Key terms such as ‘dataset’, ‘metadata’ and ‘basket’ are used consistently and uniformly throughout the document. Where necessary, these key terms are also explained in the glossary to avoid misunderstandings and provide clarity.
- **Visual hierarchy:** Information and processes are supported by numbered graphics and screenshots. The focus is on explaining what you, the user, see on the screen. Not every single click is documented; instead, the key visual elements and their meaning are highlighted. This makes it easier to navigate and understand the steps described.

10.2 User Journey

When a data user (see explanation below) applies for electronic health data for secondary use purposes, such as research and innovation activities, education, and policy-making, within the EHDS, the user journey consists of several stages (see Figure). Access for certain purposes (public or occupational health, policy-making and regulatory activities, and statistics) is reserved for public sector bodies and Union institutions (see Chapter IV, Art. 53(1) and 53(2)).

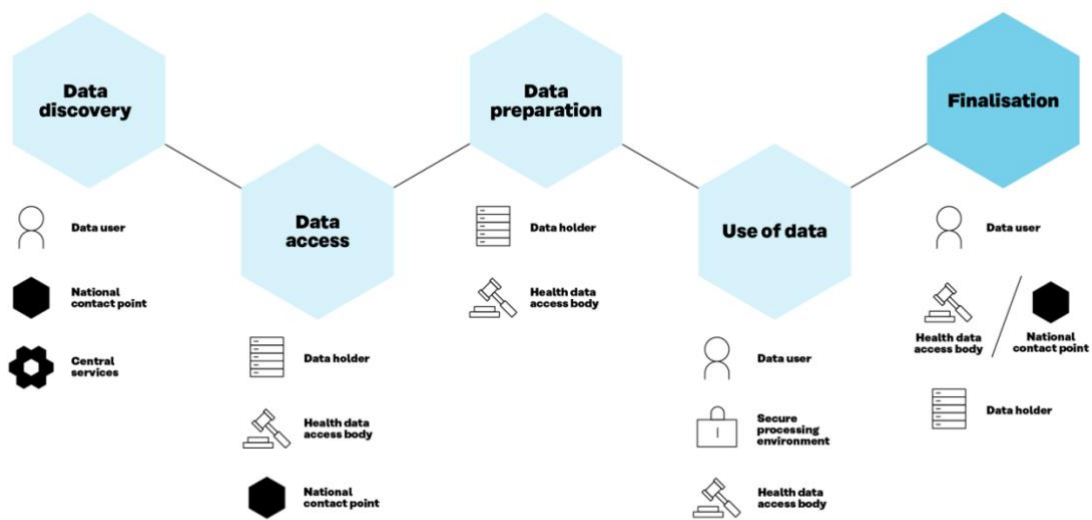


Figure 7: EHDS user journey with roles

Explanation

- i. Data user = a person using electronic health data for a secondary use purpose
- ii. Health data access body (HDAB) = the authority responsible for assessing the information provided by the data user who applies for electronic health data for a secondary use purpose
- iii. Data holder = Any natural or legal person, public authority or other body in the healthcare or the care sectors that has the right or obligation to provide electronic health data for secondary use purposes or the ability to make such data available (see more EHDS Regulation Art. 2 (1t)).
- iv. Secure processing environment = an environment with strong technical and security safeguards in which the data user can process personal level electronic health data

Data discovery

Before being able to use the data, the userⁱ needs to investigate whether the data needed is available, and whether it is available in the necessary format for the secondary use purpose. This phase is called data discovery. Datasets available in the EU can be found in a metadata catalogue at <https://qa.data.health.europa.eu/>. Once the data discovery is completed, the user can begin the process of applying for the data.

Data access

In the data access phase, the user fills in and submits a dedicated and standardised data access application form or a data request to a health data access body (HDAB)ⁱⁱ. The user must complete the information required in the form, upload necessary documents, and provide justifications as needed.

Data access application form is used when the user seeks to use personal level data. **Data request** is for cases when the user wants to apply for anonymised statistical data.

Data preparation

During this phase, the data holder(s)ⁱⁱⁱ deliver(s) the necessary data to the HDAB, which starts to prepare the data for secondary use. Techniques for pseudonymisation, anonymisation, generalisation, suppression, and randomisation of personal data are employed. The data minimisation principle (as per the GDPR) must be respected to ensure privacy.

Use of data

In this phase, the user performs analyses based on the received data for the purpose defined in the application phase. Analysing personal level data must be performed in a secure processing environment^{iv}. The duration of this phase is specified in the Regulation (Art 68(12)).

Finalisation

This last phase of the user journey concerns data user's duties regarding analysis outcomes derived from secondary use of data. Data user must publish the results of secondary use of health data within 18 months of the completion of the data processing in a secure processing environment or of receiving the requested health data. The results should be provided in an anonymous format. The data user must inform the health data access body of the results. In addition, the data user must mention in the output that the results have been obtained by using data in the framework of the EHDS.

10.3 Use case scenarios

Following the target audience (examples)

10.3.1 Use Case - (Academic) Healthcare Professionals

- **Primary need:** Rapid identification of clinically relevant datasets
- **Search behavior:** Likely disease-/condition-specific queries
- **Technical competency:** Variable (C-level executives vs. clinical researchers)
- **Filters:** Data Holder, Health Themes, Coding Systems, Age Range
- **Example:**

Doctor A works in a hospital in the field of rare diseases in paediatric and adolescent medicine. In recent years, many new patients have been admitted with symptoms of pathological exhaustion, shortness of breath and neurocognitive disorders. Doctor A is interested in relevant research in this field.

She uses the EU Dataset Catalogue and searches for dataset descriptions that contain the search term 'Long Covid'. In the results, she filters by Health themes such as MENTAL_HEALTH and IMMUNIZATION_DISEASES. It is important for her to use Coding Systems = ICD-10 to code the diagnoses so that she can compare them with

local findings. She is also only interested in data sets from children, so she selects Age range 0 – 17.

10.3.2 Use Case - Policy Makers/Institutional Users

- **Primary need:** Population health datasets, epidemiological indicators
- **Search behavior:** Geographic and temporal filtering priority
- **Technical competency:** Generally lower; preference for intuitive interfaces
- **Filters:** HDAB, Publisher, Quality Label
- **Example:**

Professor B has taken over the leadership of a group of experts whose goal is to periodic update the S3 guideline on the prevention of cervical cancer. He is interested in which data sets have been added or updated in the last five years. He searches for relevant data sets and receives a large number of results. To ensure that the data sets have been curated, he selects only those that have a Quality and utility label, as curation procedures are specified there. He also wants to focus on epidemiological registry data and selects institutions he is familiar with using the Publisher attribute. He sorts the results by Creation date and Update date. To update the key figures, he later restricts the search to data from France by selecting only the French HDAB.

10.3.3 Use Case - Academic Researchers / Data Scientists

- **Primary need:** Granular metadata examination, interoperability standards
- **Search behavior:** Complex Boolean queries, metadata-driven
- **Technical competency:** High; tolerance for technical terminology
- **Functions:** Citing
- **Filters:** Formats, Catalogue, Keywords, Categories, Data Models, Data Dictionary
- **Example:**

Researcher C wants to conduct a meta-study on whether the administration of aprotinin is beneficial for stopping bleeding during surgery. To this end, she wants to research existing data sets on this topic. Her search terms therefore also include synonyms or more general concepts in order to optimise the sensitivity of the search. She selects “data from clinical trials” as Categories and chooses known study registries as catalogues. She looks at the assigned Keywords and restricts the search to individual keywords on a trial basis in order to get a better overview. Nevertheless, she still gets a lot of results. Since her processing pipeline is based on CDISC standards and subsequent manual transformation is difficult or impossible for third parties, she only selects data sets that offer distribution in CDISC standards as a Data model. In the result set, she carefully searches existing sample data sets or data dictionaries to see if all the necessary data elements are included. She cites the data sets used in her analysis in EU Data Citation Format using the Cite feature when viewing individual data set descriptions, allowing readers to access the data set descriptions directly via the URLs provided.

10.3.4 Use Case - Industry/Business

- **Primary need:** Data licensing terms, coverage assessment, currentness, economic aspects
- **Search behavior:** Use case validation, compliance verification
- **Technical competency:** High; focus on integration feasibility
- **Functions:** SPARQL
- **Filters:** Access Level, Population Size, Update Frequency, Personal data
- **Example:**

Company D manufactures activity trackers that are also used for rehabilitation after a stroke. The trackers remind patients to take their medication, help them plan sporting activities and enable them to analyse their vital signs. Company D. would like to expand the range of applications for its products. Firstly, outdoor activities are to be offered with the help of location data in the vicinity of the user's place of residence. In addition, the influence of the weather on outdoor sports activities is to be investigated. Secondly, local patient profiles are to be compared with data on incapacity for work from the pension insurance provider in order to determine the differences between the groups. Thirdly, the activity trackers in smartwatches are to be connected. To this end, the variables provided by watches and activity trackers are to be compared to determine the extent to which the recorded values are interoperable exchangeable.

Employees of company D search for open data for the first use case by setting the Access level to PUBLIC. Weather data must be updated in real time, which is why they set the Update frequency to 'continuously updated'. For cost reasons, they limit their search to data that has been published under a licence that allows commercial users to process it further.

For the second use case, they need data on incapacity for work. Social data can also be health data within the meaning of the EHDS Regulation. However, it is not openly available but is treated as Restricted. In the course of the Data Governance Act, various institutions provide such data. Due to the resources required to request data sets, it makes sense for employees to limit the results to data sets that have a Population size of at least 100,000 individuals.

Finally, employees specifically search for data sets from Wellness applications, one of the data categories under Article 51 EHDS. In order to compare them with their own data, they need the age of the users, which can be filtered via Personal data equal to Age. In order to be informed when new data sets are added, the employees program a query in SPARQL syntax and send it to the EU Dataset Catalogue every morning.