D5.3 Technical specification for Health Data Access Bodies on the national metadata catalogue

TEHDAS2 – Second Joint Action Towards the European Health Data Space

23 June 2025

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

Lead Author(s)	Lead organisation
Irina-Afrodita Balaur	The Luxembourg National Data Service, Luxembourg
Beatriz Barros	Sciensano, Belgium
Emilie Courtois	The French Health Data Hub, France
Serge Eifes	The Luxembourg National Data Service, Luxembourg
Markus Englund	Uppsala university, Sweden
Ann Gustafsson	The Swedish e-Health Agency, Sweden
Truls Korsgaard	The Norwegian Institute of Public Health, Norway
Michael Peolsson	The Swedish e-Health Agency, Sweden
Östen Petersson	The Swedish e-Health Agency, Sweden
Danielle Welter	The Luxembourg National Data Service, Luxembourg
Per Wiklander	The Swedish Research Council, Sweden

0.2 Keywords

Keywords	TEHDAS2, Joint Action, Health Data, European Health Data Space, Dataset catalogue, Metadata, Dataset description

0.3 Document history

Date	Version	Editor	Change	Status
25/10/2024	0.1	Ann Gustafsson	First draft	Draft
28/11/2024	0.2	Ann Gustafsson	Second draft	Draft
19/12/2024	1.0	Ann Gustafsson	Milestone	Milestone 5.3
12/01/2025	1.1	Ann Gustafsson	Reviewed Milestone	Milestone 5.3
07/05/2025	1.2	Ann Gustafsson, Gabriella Jansson, Michael Peolsson	Draft D5.3	Draft
03/06/2025	2.0	Ann Gustafsson, Gabriella Jansson, Michael Peolsson	Final draft	Deliverable 5.3
22/06/2025	2.0	Ann Gustafsson, Gabriella Jansson, Michael Peolsson	Editorial from comments	Deliverable 5.3

Accepted in Project Steering Group on 24 June 2025.

Contents

1 Executive summary

This technical specification provides technical guidance for Health Data Access Bodies (HDABs), data holders and other stakeholders on how to implement and maintain a national metadata catalogue for datasets intended for secondary use under the European Health Data Space (EHDS) regulation. The catalogue is a key component of the broader HealthData@EU infrastructure, supporting the discoverability and interoperability of electronic health data across member states.

Although the EHDS regulation formally entered into force on 26 March 2025, detailed operational guidance — particularly concerning HDABs — will be developed in forthcoming implementing acts. This document provides preparatory technical guidance based on the EHDS regulation and the expected scope of forthcoming implementing acts under Article 77(4).

The national metadata catalogue serves as the authoritative registry for all dataset descriptions. Article 51 defines data categories; the authoritative registry aspect is established through Articles 60(3), 77(1) and 79(1). By exposing standardised records through a public interface and via the National Contact Point (NCP), the catalogue operationalises the FAIR principles (findable, accessible, interoperable and reusable) and supports metadata discovery across member states.

The document is primarily addressed to HDABs, which serve a pivotal function in facilitating the secure and lawful access to health data. It also engages, where appropriate, with health data holders — such as hospitals, research institutions and health authorities — as well as health data users.

This technical specification outlines guidance implementing such catalogues, including metadata ingestion (as received from data holders), management, output and access.

This specification outlines:

- **Functional requirements** for the national metadata catalogue, covering metadata ingestion, validation, versioning, publication, synchronisation with the EU-level catalogue, and user search and retrieval. These requirements are derived directly from the EHDS regulation (notably Articles 57, 73, and 77).
- **Non-functional recommendations**, such as minimum availability and performance targets, system scalability, and security considerations, to ensure robust and reliable catalogue operation.
- A set of **user stories** that illustrate the perspectives of key stakeholders — HDAB staff, health data holders, and data users — and demonstrate how the catalogue should support their roles.
- **Proposed solutions** for implementation, including modular architectures, metadata model alignment (e.g., HealthDCAT-AP), and alignment with the General Data Infrastructure (GDI) and FAIR principles.

- **Good practices** and examples from national and European initiatives, such as Health-RI (Netherlands), Healthdata.be (Belgium), and the European Open Data Portal, which illustrate how catalogues can be effectively developed.
- A **risk analysis** that identifies potential challenges — such as variability in national implementation, technical capacity gaps, and security risks — and outlines strategies for mitigation.
- An overview of **existing solutions and case studies**, such as the HealthData@EU Central Platform, that can support national catalogue development.

2 Introduction

2.1 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 health data holders, data users and the new health data access bodies (HDAB) in fulfilling their responsibilities and obligations outlined in the EHDS regulation.

TEHDAS2 focuses on several critical aspects of health data use.

- Data discovery: findability and availability of health data, ensuring 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 are 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.

2.2 The National dataset catalogue: a key to secondary use of health data

Under Article 55(1) of the EHDS regulation, member states shall designate one or more HDABs. According to Article 57(j), HDABs are responsible for maintaining a national dataset catalogue describing the available electronic health data. As a consequence, the HDAB is expected to maintain a national dataset catalogue — referred to in this report's title as the national metadata catalogue. This catalogue provides structured, accessible dataset descriptions for secondary use of health data. Each description contains key metadata, such as content, structure, origin, access conditions and intended use.

The TEHDAS[1] joint action (February 2021 – July 2023) provided the conceptual and governance groundwork for the EHDS, issuing guidance on how member states should structure secondary-use access bodies, adopt common metadata standards and protect citizens' rights. Building on that blueprint, the two-year HealthData@EU Pilot — co-financed under EU4Health and launched in 2023 — delivered working prototypes of the key building blocks, including a dataset catalogue, cross-border messaging infrastructure and interoperability tools. Its Deliverable 6.3, "Recommendations on a Data-Quality Framework for the EHDS"[2] (26 September 2023), fed directly into the formation of the HDAB Community of Practice, where member state teams now test cross-border discovery and permit workflows ahead of full-scale rollout.

For more detailed explanations of the term's national dataset catalogue and dataset description, see the Key Terminology section.

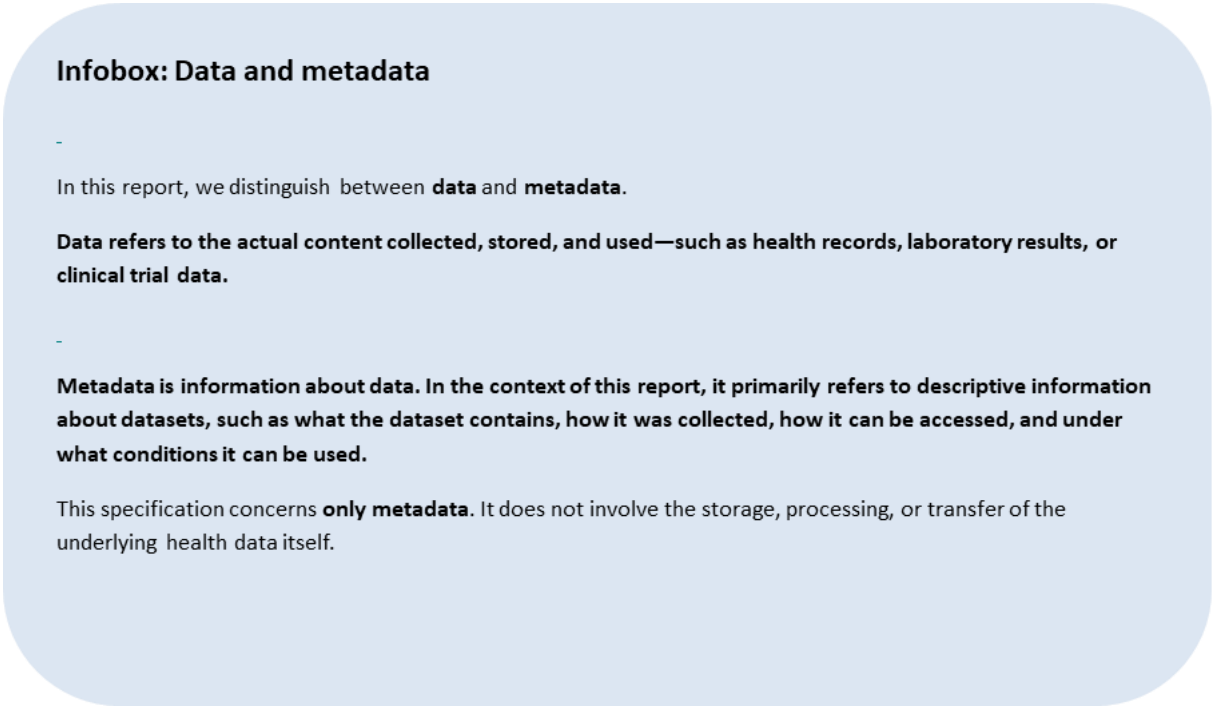
The catalogue serves as a central reference point for dataset descriptions submitted by various health data holders, including healthcare providers, research infrastructures and national authorities. It supports the collection, validation, curation and publication of metadata in a harmonised and interoperable format, aligned with the FAIR principles and the requirements of Regulation (EU) 2025/327.

It is essential to distinguish between metadata (information about datasets) and the actual health data itself. The national dataset catalogue handles only metadata — not patient-level or raw health data (see Figure 1 Infobox). The metadata managed in the national dataset catalogue correspond to the 'dataset descriptions' defined in Article 2(2)(y), submitted by health data holders under Article 60(3), and structured according to Article 77(1) and (4). Throughout this report, the terms 'dataset descriptions' and 'metadata' are used interchangeably depending on the context, in reference to the structured records required under Articles 60 and 77 of the EHDS regulation.

[1] Bernal-Delgado, E., Estupiñán-Romero, F., Martínez-Lizaga, N., Doupi, P., Mäkinen, M., Mulkholm, M., Lundbye, A. N., Gonzalez-García, J., & Telleria-Orriols, C. (2023). Deliverable 6.3 – Recommendations on a data quality framework for the European Health Data Space for secondary use. TEHDAS.

[2] Derycke, P., Korsgaard, T., Vande Catsyne, C. A., Huru, H. A., & Schutte, N. (2024). Deliverable 6.2 HealthDCAT-AP – A DCAT application profile for the description of health datasets: Recommendations on further development and deployment for possible EU-wide uptake. HealthData@EU Pilot.

Figure 1 Infobox on data and metadata



Once dataset descriptions are curated at the national level, they are transferred to the National Contact Point (NCP), (Article 77.1) which is responsible for synchronising them with the EU central dataset catalogue, part of the HealthData@EU platform.

Figure 2 illustrates the overall flow of dataset descriptions from health data holders to the EU-level catalogue:

a) A basic scenario where a health data holder submits metadata directly to the national dataset catalogue, which then synchronises with the EU catalogue via the NCP.
b) A more complex case where dataset descriptions from multiple data holders are first aggregated in a separate metadata catalogue — such as a thematic or regional catalogue — before being transferred to the national dataset catalogue and onwards to the EU level.

Figure 2 The journey of dataset descriptions from health data holders to the EU dataset catalogue.

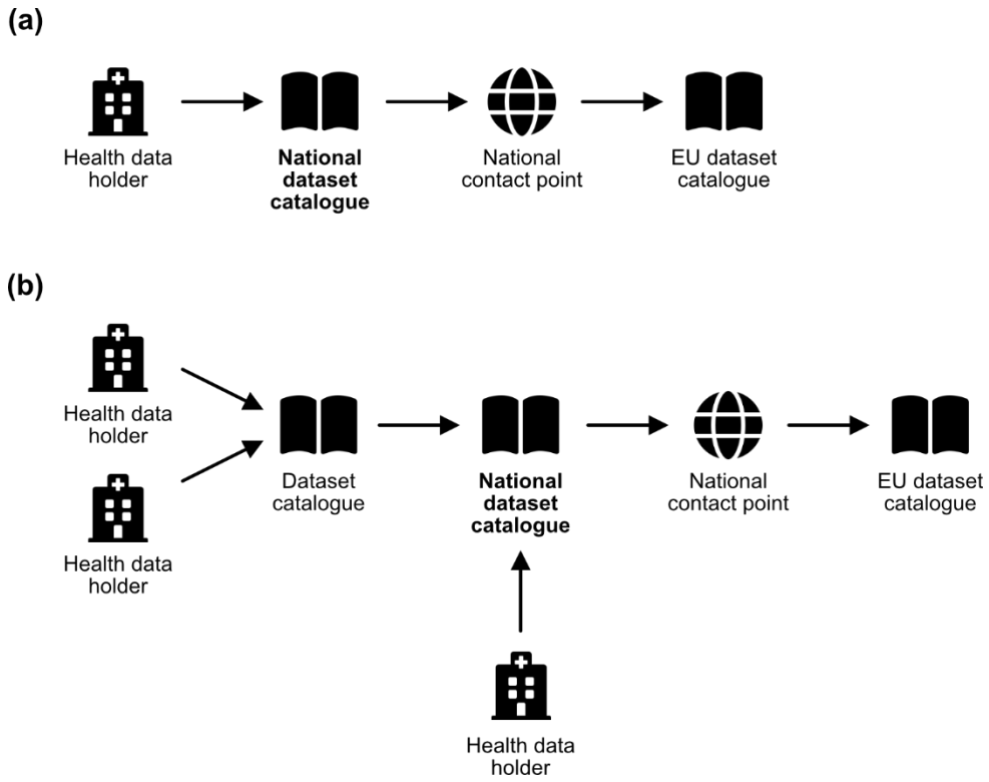
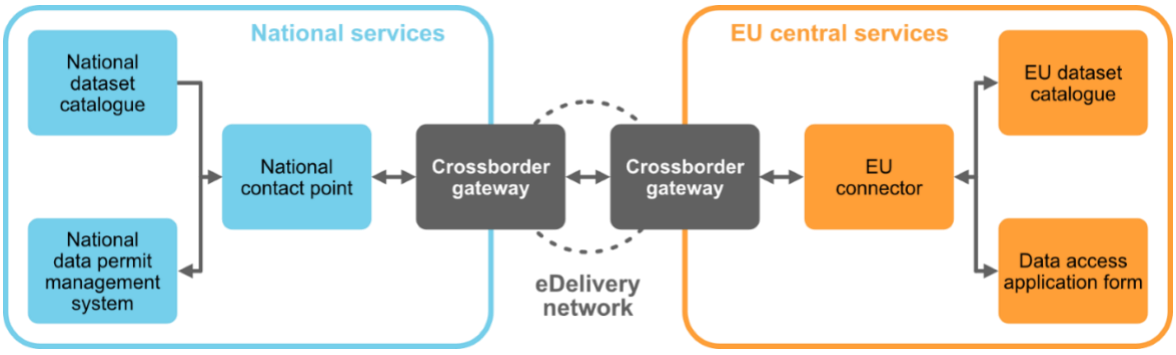


Figure 3 shows how the national metadata infrastructure interfaces with the HealthData@EU platform via the NCP. It illustrates how national infrastructures connect to the EU-level services through the eDelivery network. On the national side, the national dataset catalogue and the national data access application management system provide metadata and access services. These connect to their respective counterparts at the EU level — the EU dataset catalogue and the EU data access application form — enabling cross-border discovery and request of health data in line with the EHDS framework.

Figure 3 Connection between national and EU infrastructures for secondary use. Adopted from EHDS2 pilot.



Several key roles are involved in ensuring the national dataset catalogue functions effectively. The HDAB is the designated authority responsible for maintaining the national dataset catalogue if there is only one HDAB in the member state. In cases where multiple HDABs exist, a coordinating HDAB should be appointed to take on this responsibility. Health data holders — such as hospitals, public health authorities, or research infrastructures — are responsible for providing accurate and up-to-date dataset descriptions to the catalogue. The NCP is tasked with transmitting curated dataset descriptions from the national to the EU-level catalogue. Finally, health data users — including researchers, policymakers and public authorities — use the catalogue to identify datasets suitable for secondary use.

2.3 Purpose and goals of the technical specification

The technical specification defines the requirements and guidance for developing national dataset catalogues that can interoperate and federate seamlessly with a European-level dataset catalogue as part of EHDS. It aims to support member states in building catalogues that allow structured, standardised and secure access to dataset descriptions, regardless of where the underlying data are located in Europe. The dataset descriptions referred to in this specification correspond to the definition set out in Article 2(2)(y) of the EHDS Regulation, and are provided by data holders in accordance with Article 60(3).

A key outcome is to enable a central EU-level search portal that aggregates dataset descriptions from national catalogues, allowing researchers and policymakers to discover and explore information about available health datasets, while the actual data remains securely held at national level.

The specification covers multiple aspects, including metadata standards, validation workflows, interoperability requirements and governance considerations. One important component of this is the use of the metadata specification HealthDCAT-AP.

The specification builds on the results of the HealthData@EU Pilot and integrates reusable documentation of the open-source components developed under the HealthData@EU Central Platform[3], which are publicly available for national implementation. Task 5.2.2 of the TEHDAS2 project leads the groundwork for interoperable EHDS-compliant national catalogues.

While promoting alignment with EHDS requirements for interoperability, data governance and security, the specification allows member states to implement national-specific extensions or adjustments in parallel, to reflect their legal and organisational contexts.

A major challenge – and necessity – is for all member states to adopt a minimum set of capabilities. While national extensions are possible, they must be built on a common baseline. Harmonising this baseline set of capabilities is essential for enabling cross-border metadata exchange and supporting the EHDS infrastructure.

[3] European Commission: Directorate-General for Health and Food Safety, *HealthData@EU central platform – Open-source release 3 – Architecture model and Technical Specification*, Publications Office of the European Union, 2025, https://data.europa.eu/doi/10.2875/6472234

2.4 Target audience

This specification is intended for individuals and teams involved in the design, implementation and governance of national dataset catalogues, including:

Core implementers

- **Technical architects and developers** within national HDABs or related agencies, responsible for building or maintaining catalogue platforms.
- **Metadata officers and data stewards** who work on creating, validating and curating dataset descriptions in alignment with EHDS requirements.
- **National EHDS coordinators or policy leads** involved in planning the implementation of EHDS infrastructure and governance frameworks at national level.
- **Legal and compliance experts** responsible for interpreting and applying regulatory requirements related to data governance and metadata exchange.

External stakeholders

- **Representatives of the European Commission or EU-level bodies** contributing to the development of implementing acts and alignment across member states.
- **Health data holders and institutional IT managers** tasked with preparing and maintaining metadata to submit to national catalogues.

3 Scope

This document provides guidance for HDABs on the high-level technical capabilities required to implement a national catalogue of dataset descriptions (metadata). The catalogue must support the ingestion, storage, maintenance and publication of metadata, ensuring that dataset descriptions are publicly searchable and technically interoperable in line with relevant FAIR principles (Findable, Accessible, Interoperable and Reusable). The objective is to facilitate the secondary use of health data in accordance with Regulation (EU) 2025/327 on the EHDS.

The document focuses on functional requirements derived from the relevant articles of the EHDS regulation. It also describes the process for which the HDAB is responsible: from receiving dataset descriptions from health data holders to making them available to the NCP, which manages synchronisation with the EU central dataset catalogue.

3.1 In scope

- **Metadata Ingestion:** Receiving dataset descriptions from health data holders using standardised formats and interfaces.
- **Catalogue Management:** Storing, versioning and maintaining dataset descriptions and their unique identifiers to ensure accuracy, completeness and alignment with EHDS obligations.
- **Metadata Quality and Validation:** Supporting the validation of metadata content against required fields and quality standards, including the handling of quality and utility labels.
- **Dataset Publication:** Publishing validated dataset descriptions in a standardised, machine-readable format; making metadata available through public interfaces, search functionalities and synchronisation with the EU catalogue via the NCP.
- **Interoperability and Access:** Providing open, machine-readable Application Programming Interface (API) for metadata access and ensuring availability through the EU's Single Information Points in line with the Data Governance Act (DGA).

3.2 Out of scope

- Handling or processing of health data – this document concerns dataset descriptions only.
- Detailed technical design or implementation systems for managing of data description.
- The publishing and synchronisation process between the NCP and the EU central dataset catalogue (covered under TEHDAS2 work package 7, task 3).

- Security, authentication and authorisation mechanisms (covered by other TEHDAS2 work packages and decisions at the national level).
- Underlying infrastructure services.
- Choice of metadata standards for dataset descriptions (a topic covered in related work under TEHDAS2 work package 5, notably in deliverable D5.1 *Guidelines for Data Holders on Data Description*, assumes alignment with the expected EHDS metadata standard (e.g. HealthDCAT-AP).)

4 Key terminology

The Table below translates the legal vocabulary — mostly found in Article 2(2) (definitions) and Articles 55, 57, 75, 77–79 — into key elements in this technical specification.

Table 1 Key terminology

Term	Original text
dataset	A structured collection of electronic health data. (EHDS Article 2(2)(w)).
dataset description	A dataset description is metadata provided by a health data holder, describing the content, characteristics, and access conditions of a dataset intended for secondary use. In accordance with **Article 60(3)** of the EHDS regulation, dataset descriptions must be submitted to the Health Data Access Body (HDAB) by the health data holder. The HDAB must then make those descriptions publicly available via a standardised and machine-readable catalogue as specified in **Article 77(1)**.
dataset catalogue	A collection of dataset descriptions, arranged in a systematic manner and including a user-oriented public part, in which information concerning individual dataset parameters is accessible by electronic means through an online portal. (EHDS Article 57 and 2(2)(y)).
national dataset catalogue	Making public, through electronic means: (i) a national dataset catalogue that includes details about the source and nature of electronic health data, in accordance with Articles 77, 78 and 80 and the conditions for making electronic health data available. (EHDS Article 57(1)(j)(i))
EU dataset catalogue	The Commission shall establish an EU dataset catalogue connecting the national dataset catalogues … The EU dataset catalogue, the national dataset catalogues and the dataset catalogues of authorised participants in HealthData@EU shall be made publicly available. (EHDS Article 79(1)-(2))
data quality	Data quality means the degree to which the elements of electronic health data are suitable for their intended primary use and secondary use. (EHDS Article 2(2)(z))
data quality & utility label	Data quality and utility label means a graphic diagram, including a scale, describing the data quality and conditions of use of a dataset. (EHDS Article 2 (2) (aa)).
electronic health data	Personal or non-personal electronic health data. (EHDS Article 2(2c)).
health data access body (HDAB)	Member state designated authority that facilitates the secondary use of electronic health data. HDABs assess the information provided by the health data applicant and decide on health data requests and access applications,

	authorise and issue data permits, obtain data from data holders and make data available in Secure Processing Environments. HDABs systematically track the data request and data access applications received and the data permits issued. As per Article 58 of the EHDS, HDABs are required to publicly list information on the data permits issued. (EHDS Article 55 and Recital 52).
health data holder	Any person, organisation or public body involved in healthcare, care services, health-related products, wellness apps or health(care) research, that has the right to process data for health care provision or for public health purposes, reimbursement, research, policy making, official statistics or patient safety. This includes, for example, hospitals, insurers, research institutes and EU institutions. For a more detailed definition: EHDS regulation, Article 2(2)(t).
health data user	A natural or legal person, including Union institutions, bodies, offices or agencies, which has been granted lawful access to electronic health data for secondary use pursuant to a data permit, a health data request approval or an access approval by an authorised participant in HealthData@EU. (EHDS Article 2(2)(u).
national contact point for secondary use (NCP)	A national contact point for secondary use is the organisational and technical gateway for making electronic health data available for secondary purposes, including research, innovation, policy-making and public health. It plays a crucial role in connecting national data infrastructures to the HealthData@EU platform, enabling secure and efficient data sharing across borders. The NCP is understood as the organisational and technical interface that enables secure cross-border exchange of dataset descriptions (EHDS Article 75 (1) and recital 80).

5 From regulation to implementation: How EHDS shape the national dataset catalogue

The Regulation (EU) 2025/327 on the EHDS establishes a mandatory EU-wide framework that requires health data holders — such as hospitals, biobanks, industry, and public bodies — to provide general descriptions of electronic health datasets for secondary use (e.g. research, innovation, public health and policy-making) via their national HDABs.

To support this, each member state must create and maintain a national dataset catalogue — defined in Article 2(2)(y) as “a collection of dataset descriptions, arranged in a systematic manner and including a user-oriented public part, in which information concerning individual dataset parameters is accessible by electronic means through an online portal”. This catalogue will contain only dataset descriptions, not the data itself.

According to Article 55, member states are required to designate one or more HDABs. Under Article 57(1)(j), these bodies must make the national dataset catalogue publicly accessible.

Health data holders are obliged under Article 60(3) to submit dataset descriptions to the HDAB. These descriptions must include details about the source, scope, main characteristics, and nature of the dataset, as well as the conditions under which the data may be made available (Article 77(1)). Health data holders must also verify the accuracy and completeness of their dataset descriptions at least once per year.

Each dataset description should clearly indicate what the dataset contains, who holds the data, and what types of information it includes. However, the underlying data always remains with the original data holder.

The types of electronic health data that must be described and submitted are listed in Article 51(1)(a)–(q). Therefore, any dataset held by a hospital, registry, research infrastructure or similar entity that falls under these categories must be represented in the national dataset catalogue (see also TEHDAS2 D5.1)

6 Methodology

This section outlines the methodology used in developing the technical specification, with a focus on two key phases of the process: the milestone report and the final report.

Alignment with legal framework

A detailed analysis of EU legal frameworks applicable to health data sharing and governance has been conducted together with DG SANTÉ and with participants in TEHDAS2.

Three related EU initiatives are considered in this work:

- **The HealthData@EU Central Platform** is the EU-level infrastructure that the European Commission is required to establish under Chapter IV of the EHDS regulation (notably Articles 73–79). It will provide key services, including the EU Dataset Catalogue, a metadata aggregation function, and interfaces with National Contact Points (NCPs) to enable discoverability and access across borders.

- **The HealthData@EU Pilot** (2023–2025) is an EU funded pilot project[4] building and road-testing the first cross-border network of national gateways, certified secure-processing environments and a prototype of the EU catalogue. Lessons learned feed into this task and report.

- **The HDAB Community of Practice** is a voluntary coordination platform. It brings together the competent authorities designated by member states in preparation for the implementation of Chapter IV of the EHDS regulation. These authorities exchange experiences, align national approaches, and discuss best practices on key elements such as dataset catalogues, permit workflows, and secure processing environments. By sharing templates, governance models, and technical blueprints, the Community supports convergence towards a coherent, EU-wide infrastructure.

Workshops and discussions

The team of Major contributors has developed the first draft of the technical specification from May 2024 until January 2025 in cooperation with legal experts and the European Commission. It has involved a series of workshops and discussions.

In work package 5 there is a Review board with representatives of five countries which has provided essential feedback throughout the writing process as well as has been responsible for reviewing the content in the specification of the metadata catalogue. The Milestone has also been reviewed by Sitra and DG SANTÉ before the public consultation.

User story collection

User stories were gathered using the Connextra template (“As a [user], I want [feature], so that [benefit]”).

[4] Pilot for a European Health Data Space on secondary use of health data

Although user stories were written with real end users in mind as is considered best practice, the primary focus was to reflect the perspectives of key stakeholders in the context of this report, such as the HDAB and health data holders. This approach was taken to better address the needs of critical actors in relation to the national metadata catalogue.

The user stories also focused on the mandatory requirements for the national metadata catalogue as laid out in the EHDS legislation. Member states may identify many other user stories related to interesting but non-mandatory functionalities or that are applicable in specific national contexts only.

Writing process

Each Major contributor has been responsible for a section of the final report. In a weekly meeting upcoming questions and changes has been discussed. Collaborative tools and methodologies were used to draft the specification. This allowed for real-time input and revisions from all participants, ensuring a transparent and inclusive writing process.

Method of processing the public consultation

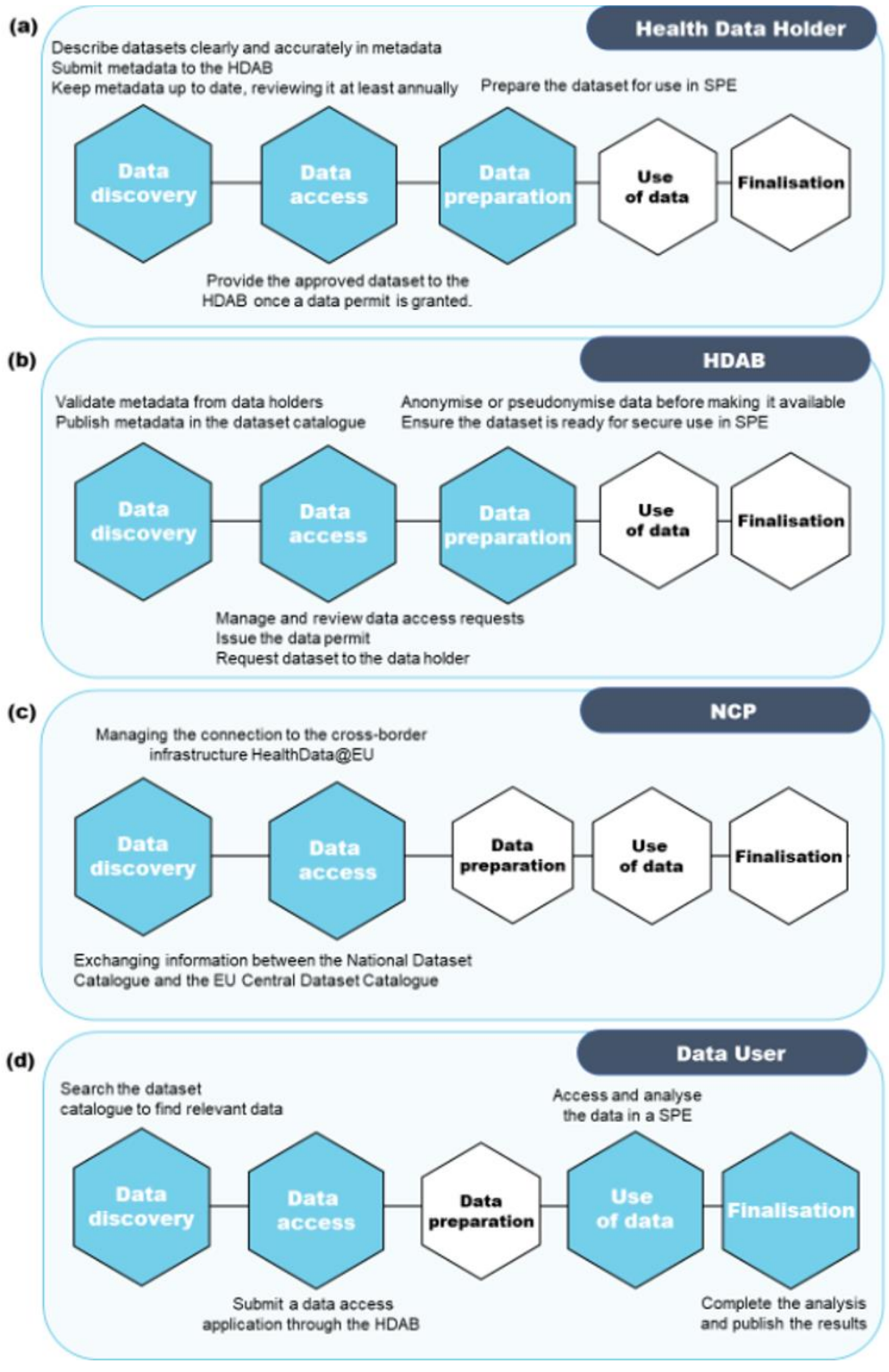
During the consultation review, each Major contributor took charge of a defined set of questions and examined every comment through three lenses: first, if essential details were missing and therefore required additional clarification; second, if any part of the response was factually inaccurate and needed correction; and third, if the point raised fell outside the immediate scope and should instead be earmarked for future discussions. This triage — “Missing,” “Incorrect,” and “Future” — allowed the team to prioritise amendments, fill knowledge gaps and capture forward-looking ideas.

7 Roles and responsibilities in relation to the national dataset catalogue

The national dataset catalogue involves four key stakeholders, each with a different purpose and for most, specific responsibilities in how they interact with it. This section outlines the roles of (a) the health data holder, (b) HDAB, (c) NCP and (d) health data user within the EHDS framework (Figure 4). While the first three have defined obligations, data users/applicants interact with the catalogue primarily as consumers of information. Although this guideline focuses on how these stakeholders engage with the national dataset catalogue, it is important to understand that this interaction represents only one part of the broader EHDS framework for secondary use. To provide a clearer picture of the full data journey, Figure 4 below offers an overview of the key steps and the involvement of each stakeholder across the process. The remainder of this section provides further clarification on the specific roles and responsibilities of each actor in relation to the catalogue.

Figure 4 illustrates the data journey within the EHDS for the secondary use of electronic health data, focusing on the roles and responsibilities of the different stakeholders involved. (a) Health data holders – Responsible for documenting datasets with clear metadata on purpose, methodology and limitations, and for making available the approved data to the HDAB once a data permit is granted; (b) HDAB – Serve as the national intermediaries that support both health data holders and data users throughout the data journey. They validate and publish dataset metadata received from health data holders, maintain the national dataset catalogue and manage data access applications in accordance with the EHDS regulation; (c) Data user – Finds the data in the dataset catalogue that fits the research needs and requests access by submitting a data access application through the HDAB. If a data permit is granted, they access and analyse the data within the Secure Process Environment (SPE) and, upon completing their work, publish the outcomes of their research and analysis.

Figure 4 Data journey within the EHDS for the secondary use of electronic health data, focusing on the roles and responsibilities of the different stakeholders involved.



7.1 Health data holder

The health data holder has the role of providing information about the datasets they hold that fall under Article 51 categories. As mandated by Article 60, health data holders have the legal obligation to 1) communicate to the HDAB a description of the datasets they hold in the form of a metadata record and 2) keep this information accurate and up to date by checking it at least on an annual basis. Article 77 further specifies the content and format of the information that must be submitted. A future implementing act will define the minimum metadata elements required for each dataset and the characteristics of those elements.

These requirements are expected to be defined in a forthcoming implementing act under Article 77(4). As concluded in the EHDS2 pilot project (WP6) and TEHDAS2 Work Package 5 (see Deliverable 5.1), the HealthDCAT-AP specification is expected to serve as the basis for the future implementing act under Article 77(4). While not formally adopted yet, it currently offers the best fit to serve as a common EU metadata framework for dataset descriptions in the context of the EHDS. Accordingly, health data holders will likely be required to describe their datasets using metadata records aligned with this specification. Additionally, if a dataset includes a data quality and utility label, as referred to in Article 78, the health data holder must also provide this label along with sufficient supporting documentation for HDABs to verify the accuracy of the label.

7.2 Health data access body (HDAB)

The HDAB has the role of validating, managing and publishing information about the datasets received from the health data holders. HDABs are responsible for establishing and maintaining the national dataset catalogue where the metadata are made publicly available for data users to consult.

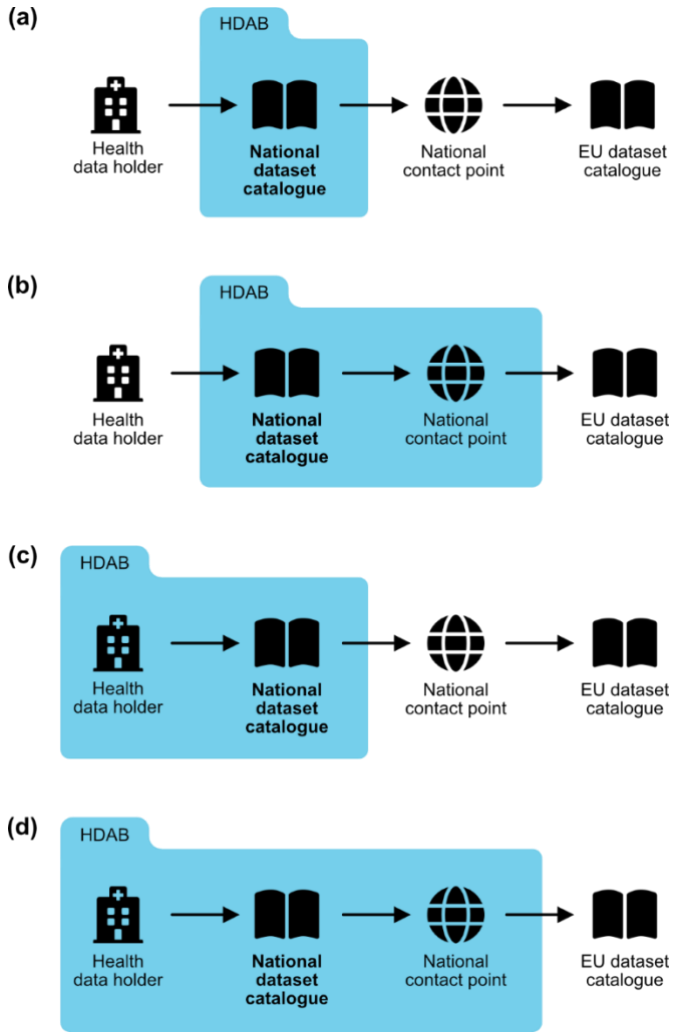
This primarily includes

1. validating that metadata records received from health data holders meet structural requirements and completeness criteria based on HealthDCAT-application profile (verify minimal metadata elements and format).
2. publishing the metadata in the national dataset catalogue in a standardised machine-readable format.

In case of multiple HDABs, the member state shall establish national rules to ensure their coordinated participation. This includes appointing one HDAB to serve as the coordinator, responsible for managing collaboration among the national HDABs.

Figure 5 presents four example configurations of responsibilities that an HDAB — or a coordinating HDAB, if multiple HDABs exist in a member state — may hold: (a) the HDAB is responsible solely for the national dataset catalogue, with no role as a health data holder or national contact point; (b) the HDAB manages both the national dataset catalogue and serves as the national contact point; (c) the HDAB is responsible for the national dataset catalogue and is also a health data holder, but does not act as the national contact point; (d) the HDAB holds all three responsibilities: maintaining the national dataset catalogue, acting as the national contact point and serving as a health data holder.

Figure 5 Example configurations of responsibilities for the HDAB or coordinating HDAB.



7.3 National contact point (NCP)

As defined in Article 75, each member state must designate one NCP for secondary use, that has the role of communicating/sending information (metadata records) from the national dataset catalogue to the EU Central Dataset Catalogue. This role may be fulfilled by the designated coordinating HDAB. The national contact points are responsible for managing the connection to the cross-border infrastructure HealthData@EU, ensuring that any additions, updates or deletions of metadata records at the national dataset catalogue are synchronised with the EU central catalogue.

7.4 Health data user / applicants

Health data users/applicants have the role of consulting the information in the national dataset catalogue. They rely on the dataset catalogue to identify the datasets that may be relevant to their research question or other secondary use purposes, which requires a user-friendly and accessible catalogue with search functionalities and standardised metadata.

This enables data users to understand the key characteristics of available datasets and assess their suitability for specific use cases.

According to the purposes for secondary use described in Article 53, a health data user can have different roles and search for data with different goals. In the EHDS, health data users can be researchers, public authorities, teachers, hospitals, or companies who use health data to improve healthcare, develop new tools, support policies, or produce health statistics. Some illustrative examples include:

- **Researcher working on early diagnosis of colon cancer** – A researcher interested in colon cancer wants to improve early diagnosis. They study what factors most strongly predict the onset of colon cancer in at-risk individuals. To do this, they explore a health data catalogue, searching for datasets on genetics, lifestyle habits and medical history. They filter the results by topic and format to identify useful data from hospitals and public health agencies.
- **Statistical office preparing national health statistics** – A national statistical institute is preparing an annual report on chronic diseases. To get the latest figures on diabetes, they consult the national health data catalogue. They look for datasets that include hospital admissions, prescriptions and mortality data, using filters for time period and geographic region. This helps them build accurate, up-to-date statistics for policymakers and the public.
- **Public authority shaping cancer prevention policy** – A public health authority is developing a new strategy for cancer prevention. They want to understand how cancer rates vary by region, age and risk factors. Using the national dataset catalogue, they search for cancer registries, screening participation data and environmental exposure records. These datasets help them target prevention campaigns and allocate resources where they are most needed.

8 Stakeholder needs

8.1 User stories related to the dataset catalogue

To ensure the effective implementation of the EHDS, it is essential to understand the practical needs, expectations and challenges faced by the key stakeholders involved in both the national and the cross-border use of health data. This section presents a set of user stories and a typical user journey to illustrate how different stakeholders — health data holders (in particular their metadata curators), data users and HDABs — interact with the EHDS infrastructure.

For example, a health data holder such as a regional hospital may need to respond to a request for anonymised patient data from a research institute in another EU country. A health data user, such as a public health researcher, may seek access to datasets across multiple member states to study the long-term effects of a rare disease. Meanwhile, an HDAB must ensure that the request complies with legal, ethical and technical requirements and facilitate secure data access through the HealthData@EU infrastructure.

While the purpose of user stories is to concretise abstract system requirements into direct examples of how specific users will interact with a system, this report will refer to the organisational roles of “health data holder”, “health data user” and “HDAB” defined above in order to avoid confusion. Examples of typical users within these roles include:

- Health data holder: For example, a metadata curator or data steward employed by a health data holder organisation who is responsible for maintaining dataset descriptions.
- Health data user: As previously described, this group includes a wide range of public and private organisations. Typical users might be researchers, educators, or public sector statistician engaged in secondary use of health data.
- HDAB: In the context of the catalogue, this typically refers to catalogue administrators or other staff within the HDAB who are responsible for validating, managing and publishing metadata records.

8.2 User stories

The user stories are presented here in narrative format, grouped by category of use, to make them easier to read. The same information in table format is available in annex 4 of this report. The following user stories include both EHDS-mandated capabilities and recommended features that may be implemented depending on national priorities or available resources. The user stories are written following industry best practice. They express user needs. They are not requirements in the EHDS regulation. See Annex 3 for cross-references with Functional Requirements.

8.2.1 Catalogue management

US1: As a health data holder, **I want to** interact with the metadata for my datasets through a graphical user interface, **so that** it is user friendly without having to invest in IT-solutions.

US2: As a health data holder, **I want to** update the metadata for my datasets in the catalogue, **so that** I can comply with EHDS periodic update requirements.

US3: As a health data holder, **I want to** receive reminders when the annual update for my dataset is due, **so that I** can easily comply with my legal obligations regarding updating.

US4: As a health data user, **I want to** see the history of a dataset record, **so that I** can understand how this dataset may have changed over time.

US5: As a health data user, **I want to** explore the tables and variables in a dataset in the catalogue, **so that** I can understand what data are available within the dataset.

8.2.2 Dataset publication

US6: As a HDAB, **I want to** export HealthDCAT-AP compliant metadata to the European Health Data Catalogue, **so that** I can comply with the EHDS requirements for HDABs with respect to national catalogues.

US7: As a health data holder, **I want to** view usage statistics about my datasets, **so that** I can evaluate and report on the utility of my data.

US8: As a health data user, **I want to** search datasets recorded in the catalogue, **so that** I can find datasets that might be needed for my work.

US9: As an HDAB, **I want to** explore usage statistics for all datasets in the national catalogue, **so that** I can proactively anticipate which datasets are high-value and may require additional attention.

8.2.3 Metadata ingestion

US10: As a health data holder, **I want to** submit data dictionaries for my datasets alongside descriptive metadata, **so that** I can easily fulfil all my cataloguing-related EHDS obligations and provide relevant details for health data users.

US11: As a health data holder, **I want to** submit metadata about my datasets to the catalogue, **so that** I can fulfil EHDS requirements for listing datasets.

US12: As a health data holder, **I want to** submit metadata via APIs to the catalogue, **so that** I can integrate my metadata submission processes into my existing computational resources.

US13: As an HDAB**, I want to** support existing cataloguing systems used by data holders**, so that I** can reduce duplication effort and facilitate metadata submission.

8.2.4 Metadata quality and validation

US14: As a health data holder, **I want to** receive feedback on the validity of my metadata, **so that I** can ensure that the information provided is sufficient and clear.

9 Functional requirements

To comply with the EHDS regulation, every member state must operate a public, machine-readable catalogue that stores only dataset descriptions. The catalogue's job is to make every dataset that falls within the Article 51 data categories and for which the health data holder has submitted a dataset description in accordance with Article 60(3) instantly findable, both nationally and via HealthData@EU. The functional requirements that follow translate the legal duties into concrete capabilities: how metadata are submitted and validated, how the catalogue is exposed in at least one EU language, how it is maintained and how it synchronises through the NCP for secondary use.

Figure 6 Overview functional requirements for the national dataset catalogue.

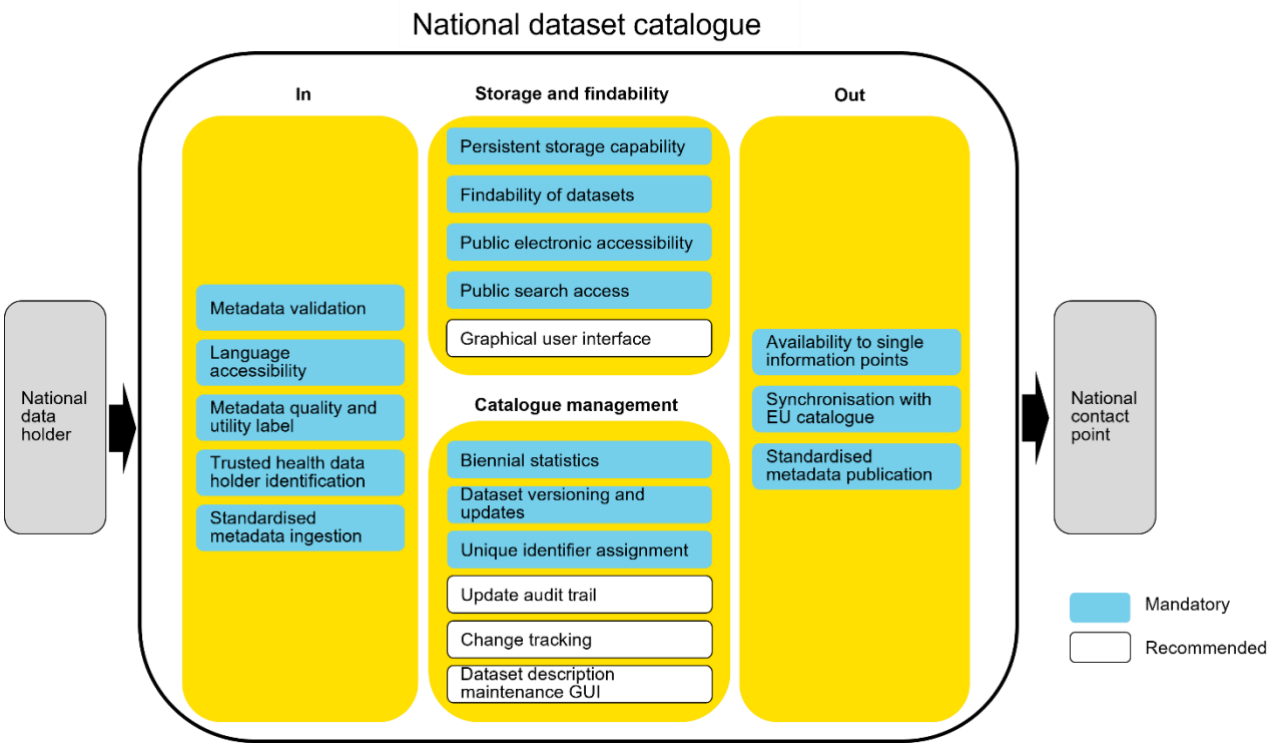


Figure 6 illustrates how data-holders submit dataset descriptions to the national dataset catalogue, which is then published through the NCP. It presents the catalogue in terms of the capabilities required to comply with the relevant EHDS regulation articles. Each capability is flagged as either mandatory or recommended, based on the legal obligation in the cited article. The capabilities are explained in detail below.

9.1 Metadata ingestion: Mandatory

9.1.1 Standardised metadata ingestion

It is mandatory for the national dataset catalogue to be technically capable of ingesting dataset descriptions provided by health data holders in a standardised and machine-readable format (e.g. HealthDCAT-AP), in line with Article 77(1).

Rationale: While the EHDS regulation does not explicitly require dataset descriptions to be submitted *by health data holders* in a HealthDCAT-AP format, Article 77(4) empowers the Commission to adopt implementing acts that will define the minimum metadata elements and their characteristics. These elements are expected to align with the HealthDCAT-AP specification.

Health data holders are required under Article 60(3) to provide and annually verify accurate and complete dataset descriptions. To ensure that these metadata records can be validated and published in a standardised, machine-readable format (as required by Article 77(1)), the national catalogue must be able to ingest metadata in a HealthDCAT-AP-compliant format.

Therefore, this capability is technically necessary to fulfil the legal obligations imposed on both data holders and HDABs, and to ensure readiness for the upcoming implementing acts under Article 77(4).

EHDS reference: Article 77(1); Article 77(4); Article 60(3) — Article 77(1) mandates standardised catalogue output; Article 77(4) defines health data holder metadata elements; ingestion format not explicitly required but implied.

9.2 Catalogue management: Mandatory

9.2.1 Persistent storage capability

It is mandatory for the national dataset catalogue to include a capability to persistently store and manage dataset descriptions submitted by health data holders.

Rationale: Although the EHDS regulation does not explicitly mandate persistent storage, this capability is technically indispensable to meet the regulation's requirements. Article 77(1) requires HDABs to maintain a publicly available dataset catalogue with structured, machine-readable metadata; Article 60(3) obliges health data holders to verify and update their dataset descriptions annually.

Fulfilling these obligations presupposes that the national dataset catalogue can store, retain, and manage dataset descriptions over time. Without such persistent storage, it would be impossible to ensure continuity, enable updates, or provide reliable synchronisation with the EU catalogue.

This responsibility falls on HDABs as part of their operational duties under Article 55 and 57(1)(j).

EHDS reference: Article 77(1); Article 60(3) — Implied obligation for persistent dataset description management.

9.2.2 Dataset versioning and updates

It is mandatory for HDAB(s) to ensure that the national dataset catalogue supports the registration, update and versioning of dataset descriptions.

Rationale: This ensures that the catalogue reflects current, accurate dataset descriptions, provided by health data holders as required by Article 60(3).

EHDS reference: Article 60(3); Article 77(1) — Article 60(3) requires dataset descriptions to be checked and updated annually by the health data holder; Article 77(1) requires the catalogue to remain accurate, standardised and publicly accessible.

9.2.3 Unique identifier assignment

It is mandatory for the national dataset catalogue to assign a unique identifier to each dataset description record.

Rationale: Unique identifiers for dataset descriptions are not explicitly required in the EHDS but are necessary for synchronisation with the EU catalogue, avoiding duplicates and supporting traceability.

EHDS reference: Articles 77(1), 79(1), Recitals 85–86 — Article 77(1) requires structured dataset descriptions; Article 79(1) requires synchronisation; Recitals 85–86 emphasise discoverability and interoperability.

9.3 Catalogue management: Recommended

9.3.1 Change tracking

It is recommended for the national dataset catalogue to track changes in dataset descriptions over time.

Rationale: This supports transparency, accountability and good governance practices by allowing health data holders and HDABs to review updates, verify compliance with annual check requirements and maintain an audit trail.

EHDS reference: Article 60(3) — While the legal obligation to verify and update metadata lies with the health data holder (Article 60(3)), maintaining an audit trail within the catalogue supports this obligation by allowing both holders and HDABs to review the evolution of metadata over time. This promotes transparency, reproducibility, and accountability.

9.3.2 Update audit trail

It is recommended for the national dataset catalogue to maintain an audit trail of changes made to dataset descriptions over time.

Rationale: While the legal obligation to verify and update metadata lies with the health data holder (Article 60(3)), maintaining an audit trail within the catalogue supports this obligation by allowing both holders and HDABs to review the evolution of metadata over time. This promotes transparency, reproducibility, and accountability.

EHDS reference: Article 60(3) — Although not explicitly required, Article 60(3) implies a need for tracking and verification of updates, which is supported by maintaining an audit trail.

9.4 Metadata quality and validation: Mandatory

9.4.1 Metadata validation

It is mandatory for the national dataset catalogue to validate that each dataset description includes the minimum elements health data holders are to provide for datasets and the characteristics of those elements.

Rationale: Article 77(4) defines the minimum dataset description elements and their characteristics that health holders must provide. The HDAB must validate compliance before publication to ensure catalogue quality.

EHDS reference: Article 77(4); Article 60(3) — Article 77(4) defines minimum elements for health data holders; Article 60(3) requires annual review. Publication format is separate under Article 77(1).

9.4.2 Quality and utility label validation

It is recommended that the national dataset catalogue includes the possibility to store supporting documentation for quality and utility labels, if applied, to facilitate validation by the HDAB upon request.

Rationale: According to Article 60(4), the health data holder must provide such documentation to the HDAB to verify the label's accuracy, when required. This ensures that quality and utility labels are trustworthy and based on verifiable evidence, as required by Article 60(4) of the EHDS regulation.

EHDS reference: Article 60(4) of the EHDS regulation — Article 60(4) requires that, when a dataset carries a quality and utility label, the health data holder must provide documentation enabling the HDAB to verify its accuracy.

9.5 Dataset publication: Mandatory

9.5.1 Findability of datasets

It is mandatory for the national dataset catalogue to make dataset descriptions, corresponding to the minimum data categories listed in the EHDS regulation, findable and discoverable.

Rationale: This ensures that dataset descriptions are accessible as required by the regulation.

EHDS reference: Article 77(1); Annex II — Article 77(1) requires datasets to be described in a publicly available national catalogue; Annex II defines the minimum data categories to which this obligation applies. Together, they establish the requirement to make such datasets findable and discoverable.

9.5.2 Standardised metadata publication

It is mandatory for HDABs to publish dataset descriptions in the national datasets catalogue in a standardised, machine-readable format.

Rationale: This ensures that dataset descriptions made available to external systems, including the EU datasets catalogue, comply with EHDS interoperability requirements. This requirement applies to the catalogue's publication layer and does not restrict how dataset descriptions are initially collected from health data holders.

EHDS reference: Article 77(1); Article 77(4) — While Article 77(4) does not explicitly require a metadata format, the requirement for standardised, machine-readable publication in Article 77(1), along with implementing acts anticipated under Article 77(4), are expected to operationalise this via a profile such as HealthDCAT-AP.

9.5.3 Synchronisation with EU catalogue

It is mandatory for the national dataset catalogue to support synchronisation of dataset descriptions with the EU dataset catalogue via NCPs, in line with Art. 75(1).

Rationale: This ensures that nationally published dataset descriptions are available through the EU dataset catalogue without redundant registration, as required by Article 79(1) and Recital 86.

EHDS reference: Article 79(1); Recital 86 — Article 79(1) requires national dataset catalogues to connect to the EU dataset catalogue; Recital 86 highlights the need for interoperability and minimising administrative burden.

9.5.4 Public search access

It is mandatory for the national dataset catalogue to provide publicly accessible electronic access to dataset descriptions that support search and navigation.

Rationale: This supports discoverability and transparency as required by the EHDS and while a graphical interface is not explicitly required, the public availability and usability goals imply the need for user-accessible online tools.

EHDS reference: Article 77(1); Recital 84 — Article 77(1) requires a publicly available dataset catalogue; Recital 84 emphasises discoverability, which implies the need for accessible presentation of dataset descriptions.

9.5.5 Biennial reporting

It is mandatory for HDABs to report statistical information related to dataset access and usage as part of their biennial activity reports.

Rationale: This supports transparency and accountability by ensuring that the functioning and usage of the national dataset catalogue contributes to the evidence base for EHDS oversight and evaluation, through biennial reporting obligations.

EHDS reference: Article 59(1–2); Article 57(1)(j)(i) — Article 59 requires HDABs to submit biennial activity reports including data usage and outcome statistics; Article 57 requires the operation of a national dataset catalogue as part of their duties, linking catalogue activity to reporting.

9.5.6 Language accessibility

It is mandatory for dataset descriptions in the national dataset catalogue to be available in at least one official language of the Union.

Rationale: This ensures minimum linguistic accessibility and compliance with Article 77(2), which requires dataset descriptions — not just the catalogue interface — to be available in at least one official language of the Union.

EHDS reference: Article 77(2) — Article 77(2) explicitly requires that dataset descriptions in the national dataset catalogue be available in at least one official language of the Union.

9.5.7 Public electronic accessibility

It is mandatory for the national dataset catalogue to be made publicly accessible through electronic means.

Rationale: This fulfils the EHDS requirement for publicly available and machine-readable dataset description access, supporting discoverability, transparency and automation.

EHDS reference: Article 77(1) — Article 77(1) requires the dataset catalogue to be publicly available and standardised in machine-readable form, implying online accessibility through electronic means.

9.5.8 Availability to single information points

It is mandatory for the national dataset catalogue to be made available to the single information points referred to in Article 8 of the DGA (Regulation (EU) 2022/868).

Rationale: This ensures that dataset descriptions can be indexed and made discoverable through the EU's broader data sharing infrastructure, in line with EHDS and DGA obligations.

EHDS reference: Article 77(3) EHDS; Article 8 DGA — Article 77(3) requires national dataset catalogues to be made available to the single information points defined in the DGA.

9.5.9 Trusted health data holder identification

It is mandatory for the national dataset catalogue to indicate which health data holders have been designated as trusted health data holders.

Rationale: This ensures transparency in the simplified access procedure under Article 72 by making trusted entities clearly visible to data users.

EHDS reference: Article 72(2) — Article 72 requires HDABs to indicate trusted health data holders in the national dataset catalogue if such designation exists.

9.6 Dataset Publication: Recommended

9.6.1 Graphical user interface

It is recommended for the national dataset catalogue to provide a graphical user interface to manage and review dataset descriptions received from health data holders.

Rationale: This facilitates regular updates and quality improvements by enabling users to interact with and enhance dataset description content, supporting Article 60(3) and the spirit of Article 77(1).

EHDS reference: Article 60(3); Article 77(1) — Article 60(3) requires dataset descriptions to be kept up to date; Article 77(1) requires HDABs to maintain and publish structured dataset descriptions, which is supported by having human-readable management interfaces.

See annex 3 for cross-references with User-stories.

10 Non-functional requirements

While the EHDS regulation primarily focuses on functional capabilities and metadata standards, it is recommended that member states consider a minimal set of non-functional requirements when implementing their national dataset catalogues. These recommendations aim to promote a baseline level of performance, availability, reliability and security across the European Union.

These non-functional requirements are provided as guidance only; member states retain the flexibility to define their own standards and practices in line with national priorities and existing infrastructure. Several related aspects, such as metadata versioning, change tracking and audit trails, are already covered by the functional requirements in Section 9.

10.1 Recommended non-functional requirements

Availability: The national dataset catalogue should aim for at least 99% system uptime, excluding scheduled maintenance periods.

Performance: The catalogue should respond to typical metadata queries within a reasonable time frame (e.g., under 1 second for common search queries).

Reliability: The catalogue should implement mechanisms to detect and recover from failures, ensuring the service is resilient and stable.

Scalability: The system should be designed to handle increasing volumes of dataset descriptions and user access without degradation in performance.

Security and Access Control: While specific security measures are determined at the national level, the catalogue should incorporate appropriate access control, authentication and authorisation mechanisms to protect system integrity, prevent unauthorised changes to dataset descriptions, and safeguard any metadata fields that may include institution-level or commercially sensitive information (e.g. on data availability or usage statistics).

11 Proposed solutions

This section outlines proposed high-level solutions and architectural approaches to meet the functional requirements for the national dataset catalogue, as described in Section 6.

11.1 Catalogue architecture

The national dataset catalogue should be implemented as a modular system comprising:

- A **metadata ingestion interface** that enables health data holders to submit dataset descriptions to the HDAB. Under Article 60(3) of the EHDS regulation, health data holders are legally responsible for providing dataset descriptions that are accurate, complete and compliant with the format and content defined in the implementing act adopted under article 77 (e.g. HealthDCAT-AP). The HDAB may offer a submission interface or metadata generation tool to facilitate this process, but the use of such tools is optional and does not transfer responsibility. Regardless of the technical means used, the data holder remains fully accountable for the validity and compliance of the submitted metadata.
- A **metadata storage and management layer** that enables persistent storage, versioning, change tracking and unique identifier assignment for dataset descriptions.
- A **validation and quality control component** to check that incoming metadata includes required fields and adheres to established rules and structures.
- A **public metadata access layer** that exposes dataset descriptions via an API, enabling machine-readable access and integration by third parties, as well as a human-readable web interface for browsing and discovery.
- A **synchronisation mechanism** that connects the national catalogue to the National Contact Point and, through it, to the EU dataset catalogue, In accordance with the forthcoming implementing act under Article 75 of the EHDS regulation, this synchronisation must be carried out using the **eDelivery infrastructure**, specifically the AS4 protocol, to ensure secure and interoperable metadata exchange.

11.2 Data model alignment

To ensure interoperability, the data model for dataset descriptions should align with the metadata elements defined by the EHDS implementing acts. While the exact format is pending, the design should anticipate conformance with HealthDCAT-AP and FAIR data principles.

11.3 API and access

The catalogue should provide:

- A **public API** for programmatic access to dataset descriptions, supporting standard query operations (e.g. search, filter, retrieve).
- A **user interface** for manual browsing, discovery and management of metadata records.
- **Access mechanisms** that ensure availability through the EU's Single Information Points and support cross-border metadata discovery.

11.4 Data quality controls

The system should include:

- Validation routines to enforce compliance with required metadata fields and structures.
- Tools to receive and validate quality and utility labels submitted by data holders, and ensure supporting documentation is attached where applicable
- Audit trails and versioning mechanisms to support transparency and governance.

11.5 Existing solutions and case studies

To illustrate the practical implementation of the proposed solutions, this section highlights some existing solutions and case studies from existing national or international dataset catalogue projects that align with the principles outlined in this document. This is the current state of these solutions and case studies at the time of publishing.

11.5.1 The Dutch Health-RI Catalogue

The Health-RI initiative in the Netherlands has developed a metadata catalogue based on DCAT-AP v3 and HealthDCAT-AP profiles. This implementation demonstrates how a national platform can provide persistent identifiers, machine-readable metadata publication and a modular architecture supporting interoperability and FAIR principles.

11.5.2 The European Open Data Portal (data.europa.eu)

Although not specific to health data, the European Open Data Portal provides a model for how metadata catalogues can use the DCAT-AP standard to ensure discoverability, multilingual access and cross-border metadata harmonisation.

11.5.3 Belgian Health Data Agency Catalogue

The Belgium Health Data Agency (HDA) offers a national example of a centralised health data catalogue through its publicly available platform (www.catalog.hda.belgium.be). The catalogue complies with the DCAT-AP standard and enables the discovery of health-related data from multiple institutions via a unified interface. It is designed to support metadata

curation and improve data findability for authorised users, with the long-term goal of serving as a comprehensive inventory to foster collaboration. Additionally, HDA is a good example of how to support EHDS actors through the creation of tailored materials. Its Health Data Academy provides e-learning courses, webinars, tutorials and other resources to increase data literacy and equip health data holders, users and citizens with the skills needed to ensure high-quality, accessible and secure health data use.

11.5.4 The Healthdata@EU Pilot

The HealthData@EU Pilot project developed a pilot version of the EHDS infrastructure to enable the secondary use of health data. To demonstrate the feasibility and benefits of cross-border data sharing, the project includes five use cases covering various health topics. Its key outcomes include the creation of IT infrastructure, metadata standards and legal frameworks designed to support the secure, interoperable exchange of health data across EU member states. All project outcomes can be consulted in Results - EHDS2 Pilot.

11.5.5 The European Genomic Data infrastructure (GDI)

The European Genomic Data Infrastructure (GDI) is a major EU initiative aiming to establish a federated infrastructure for secure, cross-border access to genomic and related health data. It supports the goals of the 1+Million Genomes (1+MG) initiative by providing the technical backbone needed to enable data sharing across national boundaries, while respecting legal and ethical requirements.

Although still under development, the infrastructure will include a European-level GDI User Portal that enables data discovery across participating countries. This portal will aggregate metadata from national-level catalogues maintained by each node. To ensure interoperability and alignment with FAIR principles, each GDI node is required to expose its metadata catalogue through a FAIR Data Point (FDP). FDPs provide a standardised, machine-readable interface that supports metadata harvesting, discovery and reuse.

GDI offers a concrete example of how a distributed catalogue architecture can be implemented in practice, with harmonised metadata publication across countries and federation at the European level. This approach is highly relevant to the EHDS context, where national dataset catalogues are expected to interoperate with a central EU catalogue. GDI's model provides valuable lessons on metadata standardisation, decentralised governance and user-centred discovery tools.

11.5.6 HealthData@EU Central Platform

The European Commission's HealthData@EU Central Platform is an open-source solution for EHDS infrastructure, including a dataset catalogue that compiles metadata from member states and European bodies. It uses eDelivery for secure synchronisation with national contact points.

Reference: https://code.europa.eu/healthdataeu

11.5.7 Health-RI metadata schema

Health-RI has developed a metadata schema (CC-BY-4.0 license) combining DCAT-AP v3 and HealthDCAT-AP elements, providing a comprehensive model for describing health datasets in compliance with EU standards.

Reference: https://github.com/Health-RI/health-ri-metadata

11.6 Resources and initiatives under development

In parallel with European-level efforts, several initiatives are emerging, and will continue to emerge, at Member State level to support national implementation of the EHDS. These may include the development of dedicated portals or platforms offering legal, technical, and procedural guidance to data holders, particularly regarding metadata creation and the application of HealthDCAT-AP. For example, the Health Information Portal[5] developed by Sciensano (Belgium) illustrates a national initiative aimed at supporting data holders through accessible documentation and resources. While such tools are not developed or endorsed at EU level, national knowledge-sharing platforms will play an important role in complementing European efforts, supporting consistent training and guidance across Member States, and facilitating the effective participation of data holders in the EHDS ecosystem.

12 Risks and mitigation strategies

This section identifies potential risks associated with implementing the national dataset catalogue in alignment with this technical specification and proposes mitigation strategies to address these challenges.

12.1 Divergence in national implementations

Risk: Different member states implement the national dataset catalogue in inconsistent ways, leading to interoperability challenges.

[5] HealthDCAT AP | European Health Information Portal

Mitigation strategy: Provide clear, harmonised guidance and promote ongoing dialogue between member states and the European Commission to support consistent adoption of standards.

12.2 Delays in implementing acts for metadata standards

Risk: The delayed release of implementing acts leads to uncertainty and delays in catalogue implementation.

Mitigation strategy: Adopt a flexible, modular architecture that allows updates to data models and validation rules once the implementing acts are finalised.

12.3 Lack of technical capacity in some member states

Risk: Limited resources and expertise hinder the effective development of national dataset catalogues.

Mitigation strategy: Facilitate knowledge-sharing, technical assistance and access to open-source solutions and example catalogues.

12.4 Insufficient engagement by health data holders

Risk: Health data holders fail to provide complete and up-to-date dataset descriptions as required by Article 60(3) EHDS regulation.

Mitigation strategy: Develop clear communication strategies, templates and tools for health data holders; provide training and support to ensure compliance with metadata requirements.

12.5 Inadequate funding or resources

Risk: Limited funding delays or restricts the development and long-term sustainability of the catalogue.

Mitigation strategy: Advocate for national and EU-level funding to support catalogue development and ensure ongoing maintenance and enhancement.

12.6 Security and privacy concerns

Risk: Metadata exposure may raise privacy or security concerns, including unauthorised access or misuse of information.

Mitigation strategy: Ensure appropriate security measures are in place (aligned with national frameworks), including access controls, audit logs and compliance with data protection rules.

12.7 Performance and scalability issues

Risk: The system cannot handle increasing volumes of dataset descriptions or user queries, resulting in degraded performance.

Mitigation strategy: Design the catalogue with scalability in mind, use efficient data structures and indexing strategies and regularly monitor system performance.

12.8 Resistance to change

Risk: Stakeholders are hesitant to adopt new catalogue processes and tools.

Mitigation strategy: Engage stakeholders early, through a range of routes such as the HDAB Community of Practice, demonstrate benefits through pilot projects and build trust through transparency and co-creation processes.

Annex 1 Summary of feedback through public consultations

A draft version of this document was in public consultation between the 20th of January and 4th of March 2025. The Milestone 5.3 received 89 replies in total on the questions in the public consultation. The number of responses may include duplicates, as the survey did not require individual identification and verification. Three responses were found to be identical. Some respondents have also responded from the perspective of both health data holders and data users. The responses came from 19 different countries across the EU and the European Economic Area. Responses from Eastern European countries and international organisations were largely missing. The respondents were primarily from three main types of organisations, listed in order of prevalence: Public organisations (21), Academic/Research organisations (17) and Private organisations (17).

The questions from the public consultation were primarily related to the content of the Milestone 5.3. In the general feedback, some respondents thought that the Milestone was too detailed, while others felt it too high-level.

The feedback from the public consultation highlighted the need to clearly define the roles and responsibilities of key stakeholders – such as HDABs, NCPs and health data holders — particularly in relation to data preparation and metadata management. User stories should be improved to better reflect how different actors interacts with the national dataset catalogue, including health data holders, HDABs and health data users. The feedback also highlighted that it should be clarified when the use of Health DCAT-AP is mandatory within the health data journey. The current narrative describing the data value chain (from health data holder → national catalogue → HDAB control → EU catalogue) is overly complex and difficult to follow; a simplified sequence diagram is therefore recommended.

The language used in the document should furthermore be simplified and technical terms should be used according to consistent terminology aligned with relevant EU definitions.

There is a broad consensus that tools to support health data holders in creating dataset descriptions should be mandatory rather than optional, to ensure consistency and facilitate compliance.

The classification of requirements (mandatory, recommended, optional) should be clarified, especially within figures, to ensure consistent interpretation.

A need for additional support was also expressed, including documentation, training and tools for metadata generation. While these elements fall outside the scope of the current technical specification, they are essential for understanding the needs of health data holders and shaping future initiatives.

Summary of feedback and further development of the guideline

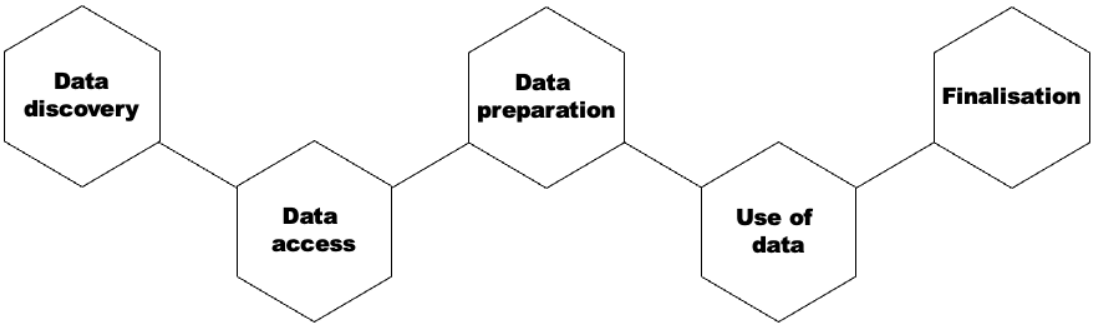
The working party received several valuable comments, many of which have been incorporated where possible. However, many of the issues raised relate to interpretations of the EHDS regulation that remain unresolved. It will be up to the member states, in collaboration with the European Commission and the EHDS Board, to address these gaps during implementation. As such, this guideline cannot answer all questions or reflect every comment received.

Annex 2 User journey

User journey

When a data user 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 7). 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 consists of five main phases: data discovery, data access, data preparation, use of data and finalisation.



Data discovery

Before being able to use the health data, the user 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 an HDAB. 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 health data holder(s) 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. The duration of this phase is specified in the Regulation (Art 68(12)).

Finalisation

This last phase of the user journey concerns health data user's duties regarding analysis outcomes derived from secondary use of data. Health 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 health data user must inform the HDAB 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.

Annex 3 Cross-references table – User stories and Functional requirements

Table 2 Cross-references

User Story	User Story Title	Functional Requirement(s)	Requirement Title(s)	How They Are Related
8.2.1	Catalogue management			
US1	Interact with metadata via GUI	9.6.1	Graphical user interface	The GUI allows metadata curators to manage dataset descriptions without requiring technical expertise or specialised tools.
US2	Update metadata in the catalogue	9.2.2	Dataset versioning and updates	Enables metadata curators to update dataset descriptions in line with EHDS requirements.
US3	Receive reminders for annual updates	9.2.2	Dataset versioning and updates	Supports compliance with the EHDS requirement for annual updates of metadata.
US4	See the history of a dataset record	9.3.1; 9.3.2	Change tracking; Update audit trail	Enables users to review the history of dataset changes, supporting transparency.
US5	Explore tables and variables in a dataset	9.5.4	Public search access	Enables researchers to explore dataset

User Story	User Story Title	Functional Requirement(s)	Requirement Title(s)	How They Are Related
				structures via public interfaces.
8.2.2	Dataset publication			
US6	Export HealthDCAT-AP metadata to EU Catalogue	9.5.3	Synchronisation with EU catalogue	Ensures synchronisation of national metadata to the EU catalogue in a standardised format.
US7	View usage statistics about datasets	9.5.5	Biennial reporting	Supports evaluation of dataset utility and promotes transparency via statistical reporting.
US8	Search datasets in the catalogue	9.5.4	Public search access	Provides functionality for data users to search and discover datasets in the national catalogue.
US9	Explore usage statistics for all datasets	9.5.5	Biennial reporting	Enables catalogue administrators to understand dataset usage patterns.
8.2.3	Metadata ingestion			

User Story	User Story Title	Functional Requirement(s)	Requirement Title(s)	How They Are Related
US10	Submit data dictionaries alongside metadata	9.1.1	Standardised metadata ingestion	Supports rich metadata submission by health data holders, enhancing discoverability and usability.
US11	Submit metadata about datasets to the catalogue	9.1.1	Standardised metadata ingestion	Fulfils data holder obligations to submit dataset descriptions for inclusion in the national catalogue.
US12	Submit metadata programmatically	9.1.1	Standardised metadata ingestion	Enables API-based, automated metadata submission by health data holders.
US13	Import metadata from upstream catalogues	9.1.1; 9.5.3	Standardised metadata ingestion; Synchronisation with EU catalogue	Supports reusing metadata from other catalogues, avoiding duplication and promoting consistency.
8.2.4	Metadata quality and validation			
US14	Receive feedback on metadata validity	9.4.1	Metadata validation	Provides feedback mechanisms to ensure submitted metadata meets EHDS standards.

Annex 4 Table of user stories

Table 3 User stories

#	Category	As a	I want to	In order to be able to
US1	Catalogue Management	Health data holder	interact with the metadata for my datasets through a graphical user interface	avoid having to invest in IT-solutions I may not have
US2	Catalogue Management	Health data holder	update the metadata for my datasets in the catalogue	comply with EHDS periodic update requirements
US3	Catalogue Management	Health data holder	receive reminders when the annual update for my dataset is due	easily comply with my legal obligations regarding updating
US4	Catalogue Management	Health data user	see the history of a dataset record	understand how this dataset may have changed over time
US5	Catalogue Management	Health data user	explore the tables and variables in a dataset in the catalogue	understand what data are available within the dataset
US6	Dataset Publication	Health data access body	export HealthDCAT-AP compliant metadata to the European Health Data Catalogue	comply with the EHDS requirements for HDABs with respect to national catalogues
US7	Dataset Publication	Health data holder	view usage statistics about my datasets	evaluate and report on the utility of my data
US8	Dataset Publication	Health data user	search datasets recorded in the catalogue	find datasets that might be needed for my work
US9	Dataset Publication	Health data access body	explore usage statistics for all datasets in the national catalogue	proactively anticipate which datasets are high-value and may require additional attention
US10	Metadata Ingestion	Health data holder	submit data dictionaries for my datasets alongside	easily fulfil all my cataloguing-related EHDS

			descriptive metadata	obligations and provide relevant details for health data user
US11	Metadata Ingestion	Health data holder	submit metadata about my datasets to the catalogue	fulfil EHDS requirements for listing datasets
US12	Metadata Ingestion	Health data holder	submit metadata programmatically to the catalogue	integrate my metadata submission processes into my existing computational resources
US13	Metadata Ingestion	Health data access body	import metadata from upstream catalogues	can reduce duplication effort and facilitate metadata submission
US14	Metadata Quality and Validation	Health data holder	receive feedback on the validity of my metadata	ensure that the information provided is sufficient and clear