Milestone 5.3 Draft technical specification on the national metadata catalogue

TEHDAS2 – Second Joint Action Towards the European Health Data Space

20 January 2025

This project has been co-funded by the 4th EU Health Programme (2021–2027) under Grant Agreement no 101176773.

0 Document info

Disclaimer
Views and opinions expressed in this deliverable represent 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
Michael Peolsson	The Swedish e-Health Agency, Sweden
Beatriz Barros	Sciensano, Belgium
Jérôme Daumas	The French Health Data Hub, France
Pascal Derycke	Sciensano, Belgium
Markus Englund	Uppsala University, Sweden
Truls Korsgaard	The Norwegian Institute of Public Health, Norway
Östen Petersson	The Swedish e-Health Agency, Sweden
Nienke Schutte	Sciensano, Belgium
Per Wiklander	The Swedish Research Council, Sweden
Danielle Welter	The Luxembourg National Data Service, Luxembourg

0.2 Keywords

Keywords	TEHDAS2, Joint Action, Health Data, Health Data Space, National Metadata Catalogue, National Dataset Catalogue

0.3 Document history

Date	Version	Editor	Change	Status
25/10/2024	0.5	Ann Gustafsson	First draft	Draft
28/11/2024	0.7	Ann Gustafsson	Second draft	Draft
19/12/2024	1.0	Ann Gustafsson	Milestone	Milestone 5.3

Accepted in Project Steering Group on 14 January 2025.

Contents

1 Executive summary

As part of the European Health Union, the European Union (EU) is advancing the secondary use of health data to drive research, innovation and evidence-based policymaking. A key initiative in this effort is TEHDAS2 (second joint action Towards the European Health Data Space), which supports the implementation of the European Health Data Space (EHDS) Regulation through detailed guidelines and technical specifications. Additionally, TEHDAS2 facilitates the development of a shared health data infrastructure by linking national health data catalogues, promoting interoperability and improving data discovery.

This deliverable aims to provide the technical specifications for the development of national dataset catalogues in alignment with Chapter IV of the European Health Data Space (EHDS) Regulation.

These catalogues will serve as a cornerstone for facilitating the secondary use of health data, ensuring their findability, accessibility, interoperability and reusability (FAIR principles) across the European Union. The specifications aim to:

1. define a standardised approach for metadata collection, management and dissemination at the national level;
2. enable seamless integration of national catalogues into the EU-wide HealthData@EU platform through the adoption of the HealthDCAT-AP standard, a health-specific extension of DCAT-AP, tailored for describing datasets in the healthcare domain;
3. provide guidance to Member States on implementing catalogues that comply with legal and technical requirements, particularly Articles 57 and 77 of the EHDS Regulation.

1.1 Why HealthDCAT-AP?

HealthDCAT-AP builds upon the established DCAT-AP standard, widely used across Europe, by extending its capabilities to specifically address the requirements of health data.

By leveraging this standard:

- interoperability between Member States' dataset catalogues and the EU dataset catalogue is guaranteed;
- consistency in dataset descriptions is achieved, supporting compliance with the EHDS Regulation;
- ease of implementation is ensured by reusing familiar metadata standards while adapting them for health-specific contexts, facilitating interoperability with other data spaces.

1.2 Benefits for Stakeholders

By implementing these technical specifications:

- Member States can reduce fragmentation, enhance data discoverability and foster collaboration across borders;
- researchers, policymakers and data users will gain streamlined access to high-quality metadata for datasets, accelerating research and evidence-based decision-making;
- the EU achieves a unified approach to managing health data catalogues, paving the way for effective EHDS implementation.

The specifications focus on four core capabilities.

1. Metadata input: enabling data holders to submit dataset descriptions. It is optional for Member States to provide tools to support data holders in creating dataset descriptions.
2. Metadata management: metadata shall be stored, validated and maintained to ensure consistency with EHDS legal requirements, helping data holders to meet their obligations efficiently. This includes the ability of data holders to update and improve dataset descriptions and to document their annual audits.
3. Metadata output: dataset descriptions shall be made available as HealthDCAT-AP via public APIs and user-friendly search tools, facilitating discoverability by researchers and policymakers.
4. Metadata access: a user-centric interface will enable stakeholders to easily locate datasets through a national and EU-wide integrated platform.

The document specifies functional and performance requirements for these capabilities, identifying mandatory elements and offering recommendations. The specifications are presented at a high level to ensure flexibility by avoiding specific technology mandates, enabling Member States to tailor their systems to national needs while achieving the overarching goals of discoverability, interoperability and usability. The focus remains on the design and operation of the dataset catalogue itself, rather than the data described within.

2 List of abbreviations

Name	Abbreviation
Application Programming Interface	API
Community of Practice	CoP
Data Catalogue Vocabulary Application Profile	DCAT-AP
Data Governance Act	DGA
Directorate-General	DG
European Health Data Space	EHDS
European Union	EU
General Data Protection Regulation	GDPR
Geospatial Data Catalogue Application Profile	GeoDCAT-AP
Graphical User Interface	GUI
Health Data Access Body	HDAB
Health Data Catalogue Vocabulary Application Profile	HealthDCAT-AP
Joint Action	JA
Minimum Viable Product	MVP
National Contact Point	NCP
Statistical Data Catalogue Vocabulary Application Profile	StatDCAT-AP
The Finnish Innovation Fund	Sitra
Towards the European Health Data Space	TEHDAS
Work Package	WP

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 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, 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 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.

TEHDAS 2 primarily follows Chapter IV of the EHDS Regulation, establishing a national dataset catalogue for sharing health data, particularly through the national health data access body (HDAB), may require additional legal considerations. Work package (WP) 5, task 5.2.2 of TEHDAS2 is titled "Guidelines on technical specifications of the national metadata catalogues and metadata catalogue maintenance".

This task focuses on developing technical specifications for national metadata catalogues, enabling the discovery and interoperability of health datasets across EU. This task especially focuses on article 57 and 77 in the EDHS regulation aiming at dataset descriptions and dataset catalogues.

3.1 Key terminology

For consistency, the following terms will be used throughout the text (Annex 2 Key terminology – Glossary).

- "National dataset catalogue" refers to the metadata catalogue managed by the health data access body (HDAB) at the Member State level. This catalogue contains machine readable descriptions (metadata) of datasets relevant for secondary use under the EHDS Regulation.
- "EU dataset catalogue" refers to the centralised metadata repository within the HealthData@EU platform. This catalogue does not store datasets but aggregates and organises dataset descriptions (metadata) received from national dataset catalogues via the national contact point (NCP). It enables stakeholders to locate metadata across Member States.
- "Dataset description" refers to the structured metadata describing a dataset within the national dataset catalogue. This includes information such as the dataset's source, scope, main characteristics and conditions for access, as defined in Article 77 of the EHDS Regulation. It should not be confused with the "description" property in DCAT-AP or HealthDCAT-AP metadata standard.

3.2 Purpose and goals

The main goal is to establish technical specifications for national dataset catalogues that can federate seamlessly into an EU dataset catalogue. A key objective is to establish a central search portal for descriptions of datasets via the EU dataset catalogue, providing researchers and policymakers with comprehensive access to information, regardless of the data's location in Europe. Achieving this requires a standardised dataset specification. The proposed specification, HealthDCAT-AP, builds on the DCAT-AP standard and is tailored specifically for describing health datasets in a European context. Drawing from the outputs of the HealthData@EU Pilot and utilising open-source tools from the HealthData@EU Central Platform, task 5.2.2 seeks to lay a foundation for interoperable national dataset catalogues.

The project targets two main audiences.

1. Member States: It supports the development of national dataset catalogues managed by HDABs, which may involve one or multiple HDABs per Member State, with a coordinating function if necessary.
2. The European Commission: It informs the preparatory work for the implementation acts under the EHDS Regulation.

The technical specifications aim to guide national HDABs in receiving, validating and managing dataset descriptions provided by data holders. These specifications emphasise compliance with EHDS Regulation requirements for interoperability, data governance and security.

It also allows Member States to implement in addition or in parallel specificities related to their national legal frameworks.

A significant challenge is to ensure that all Member States adopt a minimum set of capabilities to achieve interoperability across national dataset catalogues. This harmonisation is crucial for the seamless exchange of health data for secondary use and for enabling a unified European Health Data Space.

3.3 Relationship between the HealthData@EU Pilot and TEHDAS2

Task 5.2.2 builds on the foundational work carried out in the HealthData@EU Pilot[1], particularly the development of a health-specific extension of the DCAT-AP specification for exchanging descriptions of health datasets. This task focuses on developing a technical specification for dataset catalogues at the national level, ensuring alignment with cross-border processes for the discovery of existing health datasets across Europe.

As a critical component of the HealthData@EU infrastructure, national dataset catalogues focus on managing and making metadata about datasets available in a human and machine-readable format. These catalogues:

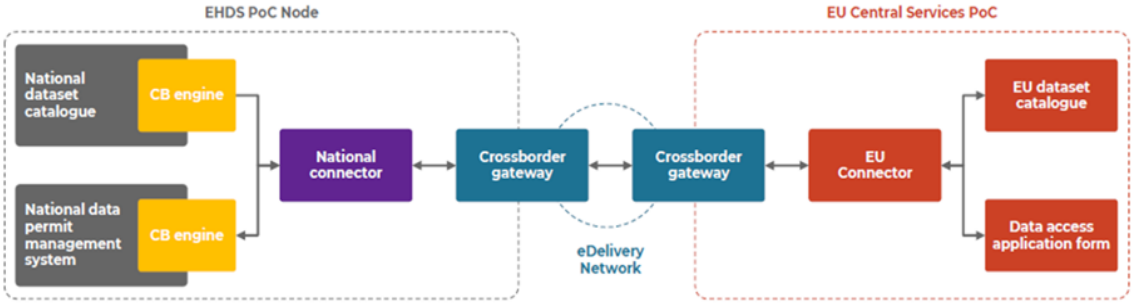
- enable data holders and HDABs to provide consistent and structured metadata about health datasets;
- facilitate discoverability of datasets at both national and EU levels by potential data users and the general public.

This work is grounded in legal and technical frameworks, notably the Data Governance Act (DGA)[2] and the EHDS Regulation. They emphasise creating a metadata management process, connecting data holders, national HDABs and/or national coordinating HDABs and EU portals, ensuring seamless data discoverability and usability for secondary purposes such as research and policymaking.

The tasks of HDABs are stipulated in/follow from the EHDS regulation, particularly article 57. One of these tasks, that is relevant for this work package is that the HDAB through electronic means shall make public a national dataset catalogue. The catalogue should fulfill the requirements set out in articles 77, 78 and 80. The catalogue should also include the conditions for making electronic health data available. Article 77 follows that the catalogue shall be standardised and machine-readable. HDAB shall also provide a description of the available datasets and their characteristics. The descriptions shall be made in the form of metadata. The description of each dataset shall include information about the source, scope, main characteristics, nature of the electronic health dataset and the conditions for making those data available. By making national dataset catalogues interoperable and compliant with legal framework, task 5.2.2 ensures the establishment of a robust infrastructure that enhances the discoverability, accessibility and interoperability of health data for secondary use across Europe.

[1] Healthdata@EU Pilot, https://ehds2pilot.se
[2] Data Governance Act, Regulation - 2022/868 - EN - EUR-Lex

Figure 1: The general context for data sharing described in the HealthData@EU Pilot



3.4 The metadata management process

The metadata management process involves collecting, validating and publishing metadata (dataset descriptions) to ensure their integration into the EU dataset catalogue. This process enables the following steps to be achieved.

- Metadata collection: Dataset descriptions provided by data holders at the national level are collected and validated to ensure accuracy and compliance with the HealthDCAT-AP application profile.
- Metadata management: The metadata repository, managed by the HDAB, organises and maintains the dataset descriptions, including tracking updates and versioning.
- Metadata publication: Validated metadata are formatted in the HealthDCAT-AP specification and made available via the NCP integration into the EU dataset catalogue, supporting discoverability across Member States.

4 Scope

4.1 Introduction

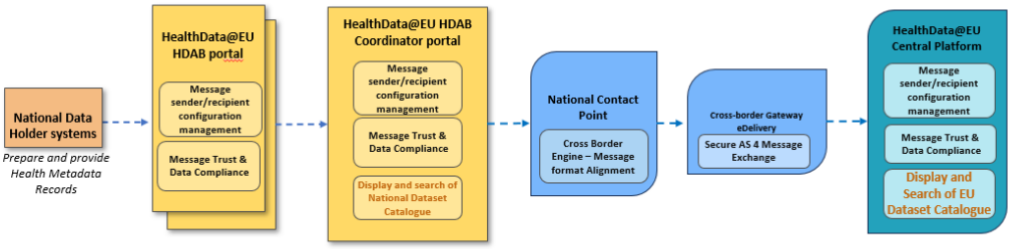
The framework for this technical specification is defined by legal provisions that outline the requirements for national dataset catalogues, with a specific focus on health data under the EHDS Regulation. While the EHDS provides the primary legal and operational foundation for health metadata catalogues, its implementation is supported by broader EU data governance frameworks, such as the Data Governance Act (DGA) - see Box 1.

BOX 1 - Relation between the Data Governance Act and EHDS Regulation

The DGA provides the legal context and broader framework by underpinning the interoperability, accessibility and discoverability through single information points of metadata catalogues across various domains, including health. The EHDS builds further on the DGA to create a specialised structure for health data management within the EU. Aligning the EHDS requirements with the DGA's principles ensures consistency with overarching EU data governance policies. In this context, it is important to distinguish between the legal mandates of the EHDS and the broader scope of the DGA. The EHDS focuses on health-specific requirements, defining the duties of HDABs to establish and maintain catalogues that adhere to the FAIR principles. The alignment to the two legal frameworks not only supports the mandatory requirements of a MVP for national HDAB portals but also facilitates the integration of additional functionalities that enhance the catalogues' overall utility.

Figure 2 below shows how health data is securely shared across borders in the HealthData@EU system. To understand the scope of task 5.2.2 in relationship with other tasks in TEHDAS2, see Annex 4 Relations between tasks in TEHDAS 2).

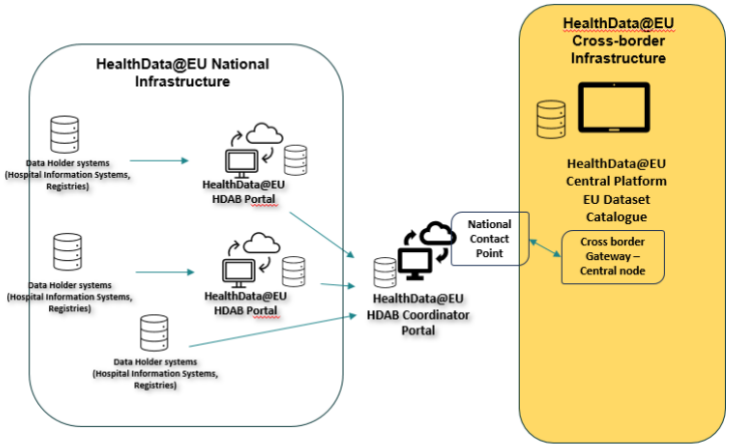
Figure 2: Schematic overview of the information flow described by DG SANTE.[3]



4.2 The European technical infrastructure exchanging health data

The EU Central Platform at HealthData@EU portal works in an ecosystem. The Figures 2, above and 3, below, describe the main components in these systems.

Figure 3. Description of the national eco-system to be connected with the EU ecosystem described by DG SANTE.



This work takes a foundation in the eco-systems described by the HealthData@EU.[4]

4.3 Technical specifications for national metadata catalogues

To meet the aforementioned legal requirements, Member States must establish a minimum set of technical and operational capabilities. These include:

- receiving and validating metadata from data holders;

[3] D03.01 Architecture Artefacts High Level Version
[4] D03.01 Architecture Artefacts High Level Version

- managing and publishing metadata in compliance with interoperability standards (HealthDCAT-AP);
- providing secure and user-friendly tools for stakeholders to access metadata.

While the EHDS Regulation outlines the mandatory outcomes, this deliverable also proposes recommended practices to enhance the efficiency, usability and adaptability of national dataset catalogues.

These recommendations aim to ensure:

- effective governance and maintenance;
- scalability to address future technological and legal developments;
- room of manoeuvre for Member States to make additional developments relevant to their national priorities.

This dual approach — distinguishing between legal obligations and practical recommendations — forms the basis of the technical specifications outlined in this document.

1. These elements form the foundation for creating a national dataset catalogue system that is not only technically robust and compliant with EU regulations, but also flexible enough to take account of national variations. The specifications aim to provide a harmonised yet adaptable framework for building a shared facility for describing national datasets, enabling efficient and secure data discoverability across Europe.

4.4 Framing the scope

This document provides the technical specifications for a national dataset catalogue, focusing on four key capabilities to support the management and discoverability of health datasets in alignment with EHDS. These capabilities are designed to support data holders in making their dataset descriptions accessible while ensuring the catalogue meets the needs of researchers and policymakers in compliance with legal and usability standards.

4.4.1 Metadata input

This capability enables data holders to submit detailed dataset descriptions to the catalogue. It is optional for Member States to provide tools to assist data holders in creating these descriptions; however, such tools are highly recommended to improve consistency and ease of submission. By streamlining the input process, this capability aims to reduce the administrative burden on data holders while enhancing the quality and standardisation of the metadata provided.

4.4.2 Metadata management

Metadata shall be stored, validated and maintained in accordance with EHDS legal requirements. Dataset descriptions may be validated at different stages on their way of becoming accessible. This specification emphasises the importance of ensuring validity of the descriptions throughout the process but leaves it up to each HDAB to decide when and how validation should be done. Included in the metadata management capability is also

functionality for updating information and improving data quality, such as the ability to document the yearly audits conducted by data holders.

4.4.3 Metadata output

Dataset descriptions shall be exported in HealthDCAT-AP, via the NCP to the HealthData@EU central platform. This ensures compliance with interoperability standards and enhances the discoverability of datasets. By making metadata easily accessible to researchers, policymakers and other stakeholders, this capability supports informed decision-making and fosters cross-border collaboration in health data research.

4.4.4 Metadata access

A user-centric interface enables stakeholders to locate and access datasets easily through a national and EU-wide integrated platform. This interface may include advanced search capabilities, filtering options and intuitive navigation to ensure accessibility for a wide range of users. By integrating with the broader EHDS infrastructure, this capability promotes harmonisation and a seamless user experience at both national and EU levels.

For each of these capabilities, the document specifies system capabilities, which describe the functions, features and performance that define what the national dataset catalogue can achieve, which outline the requirements for the catalogue, including mandatory elements and recommendations.

It is important to note that this specification concerns only the national dataset catalogue itself, not the data described by the metadata. The specifications are presented at a high level to ensure flexibility and avoid unnecessarily restricting Member States' choice of technology.

5 Stakeholder and user needs

5.1 User stories

A concise description of a functionality or feature from the perspective of a user.

5.1.1 Data holder

- As a data holder, I want a standardised way of submitting my dataset descriptions, so that my work is easier, and I can get help from others.
- As a data holder, I want to review and update the dataset description in the national dataset catalogue at least annually, so that I can fulfil my obligations according to the EHDS.
- As a data holder, I want to see if, when and who has changed my dataset descriptions in the national dataset catalogue, so that I can fulfil my legal and security needs.
- As a data holder, I want my dataset descriptions to be validated before submitting them to the national dataset catalogue, so that I can be sure that the information is correct.
- As a data holder, I want to receive communication about changes in the system, so that I know if something affects my work with dataset descriptions.
- As a data holder, I want to get statistics of usage and search hits, so that I get to know if data users are interested in my datasets.
- As a data holder, I want to get instructions or get assistance in describing my data sets, so that I can both fulfil my obligations according to the EHDS and provide quality information for data users.

5.1.2 Health data access body

- As an HDAB, I want to see when a dataset record has been revised and/or updated, so that I can hold data holders accountable to their obligations according to the EHDS.
- As an HDAB, I want to validate incoming dataset descriptions, so that I can make sure that the information is complete and valid.

5.1.3 Data user

- As a data user, I want data holders to document the quality of their data sets according to an agreed upon standard, so that I better know if the data is usable for my needs.
- As a data user, I want to find national datasets in the national dataset catalogue, so that I am able to determine if they are useful for my needs.
- As a data user, I want sufficient information to determine if a dataset is useful for my needs, so that I do not apply for access needlessly.

- As a data user, I want to know if information about the datasets I have selected has changed since I selected them, so that I have the latest information available before I apply for access to the data.
- As a data user, I want to know how many subjects are available in a dataset for a specific search, so that I know if the dataset contains enough subjects for my needs.
- As a data user, I want to understand the provenance of the data, so that I know how it was created or generated.
- As a data user, I want to understand if there is information about the study design (if a researcher is trying to find specific studies, e.g randomised double blind studies, this information is vital if it will fit the research study to be carried out)
- As a data user, I want to receive notifications about new and relevant datasets, so that I can find datasets that will be available in the future.

5.2 Use cases

Use cases will be written after the public consultation organised during January–February 2025.

6 Functional and non-functional requirements

6.1 Functional requirements

A functional requirement is a specific statement describing a system's functionality or behaviour.

6.1.1 Mandatory

- It is mandatory for the national dataset catalogue to have a repository to store dataset descriptions.
- It is mandatory for the HDAB that all datasets received relevant to secondary use under the EHDS are made available.
- It is mandatory for the national dataset catalogue to be designed to receive dataset descriptions from data holders.
- It is mandatory for the HDAB to validate that the national dataset catalogue records fulfil the minimum elements and their characteristics according to article 77.
- It is mandatory for the HDAB to have processes in place to ensure that received datasets are kept updated.
- It is mandatory for the national dataset catalogue to provide the dataset descriptions to the NCP in HealthDCAT-AP standard.
- It is mandatory for the dataset catalogue to provide public Application Programming Interfaces (API) to access the dataset descriptions stored in the national dataset catalogue.
- It is mandatory for the dataset catalogue to provide a Graphical User Interface (GUI) for dataset search and discovery.
- It is mandatory for the dataset catalogue to provide a Graphical User Interface to maintain dataset descriptions.
- It is mandatory for the dataset catalogue to track changes in the dataset descriptions (Versioning).
- It is mandatory for the dataset catalogue to make previous versions of dataset descriptions accessible.
- It is mandatory for the dataset catalogue to be able to receive metadata from data holders in HealthDCAT-AP standard.
- It is mandatory for HDAB to provide statistics and communicate them with stakeholders.

6.1.2 Recommended

- It is recommended for the HDAB to provide instructions and assistance in case of problems when writing or submitting metadata to the system.
- It is recommended for the HDAB to document the annual revision audit of the dataset descriptions by the data holder.

6.1.3 Optional

- It is optional for Member States to also support other standards for describing datasets (e.g. GeoDCAT-AP, StatDCAT-AP).
- It is optional for Member States to define the formats in which HDABs can receive dataset descriptions (e.g., CSV, XML, etc.), as the regulation does not impose specific format requirements at this level.
- It is optional for the dataset catalogue to make it possible for data users to save their search.
- It is optional for the dataset catalogue to make it possible for data users to receive notifications when tracked dataset descriptions have been changed.

6.2 Non-functional requirements

A non-functional requirement defines the system's quality attributes, performance or constraints, rather than specific functionalities.

Security - The national dataset catalogues must be protected against threats to the confidentiality, integrity and availability of information assets. Security non-functional requirement is defined as the protection of computer software, systems and networks from threats that can lead to unauthorised information disclosure, theft or damage to hardware, software or data as well as from the disruption or misdirection of the services they provide.

Compliance - a number of EU legislations that address this topic exist, and most likely some national laws as well.

Scalability - the systems in the national dataset catalogue must be able to handle variations in load (number of calls per minutes) in an efficient manner.

Availability - usually describing the acceptable downtime for a system in percent (i.e. 99.9% availability is equivalent to 9 hours of downtime per year).

Reliability - the definition of the probability that the system will behave correctly.

Robustness - how well a system handles exposure to incorrect data (i.e. a text in a number field) and usually also the process for managing these problems.

Traceability - comes from security and defines the level of logging of system users and system activities.

7 Proposed solution

The purpose of the national dataset catalogue is to provide an organised and standardised tool for receiving, managing and sharing descriptions of datasets from data holders within the Member State to be exposed at both national and EU levels. A national dataset catalogue must provide an interface that allows searching catalogued datasets. It must allow the connection via the NCP with the HealthData@EU Central Platform.

Figure 3: Description of the capabilities in the national dataset catalogue hosted and maintained by the HDAB.

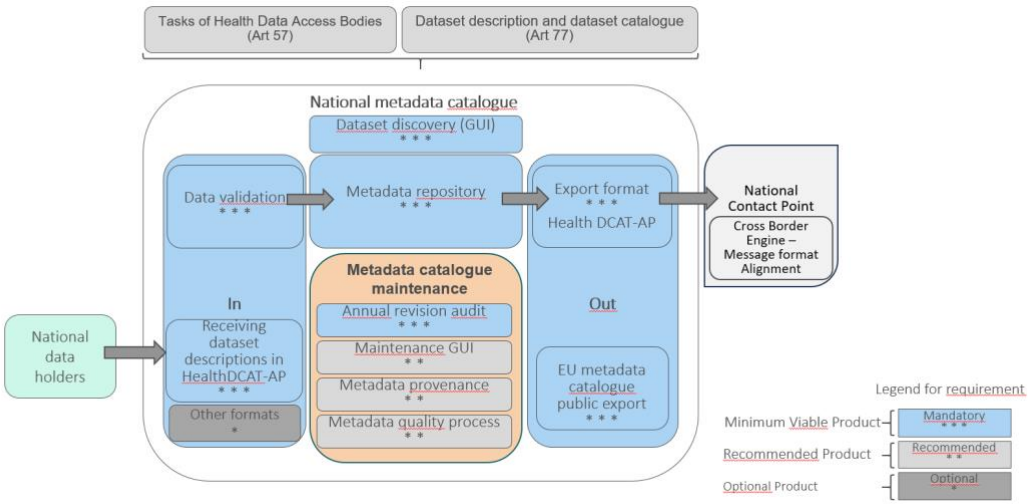


Figure 3 above illustrates the process of receiving, managing and preparing metadata for export to the EU dataset catalogue via the NCP as HealthDCAT-AP. Dataset descriptions are collected from various sources, such as data holders and data inter-mediation entities, and organised through a national dataset catalogue managed by an HDAB. Standardised metadata enables integration with the EU dataset catalogue and other catalogues that support HealthDCAT-AP, promoting data sharing and interoperability at the EU level. Additionally, the national dataset catalogue must make these records accessible through public APIs.

7.1 Describing the main capabilities

The metadata flow process begins with the HDAB receiving dataset descriptions provided by the data holders. It is required that health data holders provide dataset descriptions to the HDAB and verify their accuracy and completeness annually. These descriptions feed into the dataset catalogues to be kept by HDAB. The catalogues must meet the requirements of Article 77, meaning that they must be standardised and machine-readable. Additionally, Article 77(4) empowers the Commission to adopt implementing acts that will define the minimum elements and characteristics of dataset descriptions. These elements are expected to align with the HealthDCAT-AP to ensure consistency and interoperability across Member States.

The description of datasets can be received in other formats. These descriptions are either transformed or validated according to the HealthDCAT-AP standard. The validated dataset description is then organised and managed with versioning and logging. A metadata quality process ensures rich and accurate metadata. A metadata archive stores historical versions of dataset descriptions. Data users can access metadata through GUIs that offer search functionalities.

Member States, as written in the regulation, must have an NCP which is the only system connected to the HealthData@EU Central Platform. The connection happens only through eDelivery Domibus. Domibus functions as a national access point, meaning it sends and receives messages in accordance with the eDelivery standard (e.g. via the AS4 protocol). Member States can use Domibus as their technical solution to connect to the eDelivery network without having to develop their own implementation.[5]

The capabilities of this system are explained below, first in connection to the relevant EHDS articles, and secondly, according to how they support the functions of a coherent system.

The scope of this task is to provide capabilities for an MVP. In this context the main capabilities have been marked as mandatory or recommended in Figure 3.

7.1.1 Data validation: ensuring structural integrity

Once descriptions of datasets have been received, they undergo validation. This ensures that only high-quality, well-structured metadata enter the system.

- The metadata is validated according to a set of rules. Examples of such rules are to verify that:
    - all required fields are present;
    - values in those fields conform to specified formats (data types such as string, integer etc);
    - the structural integrity of the metadata records is intact.
- Issues are identified at the point of entry to ensure the integrity of the metadata before it is ingested into the system. In this context, integrity means that the information entered into the system is accurate, complete and consistent, and that it remains so throughout the process until it is stored or further utilised within the system.

7.1.2 Metadata repository

After the dataset descriptions have been validated, the system stores the information, making it accessible to other parts of the system.

The metadata repository serves as the central hub for all dataset descriptions within the system. This repository stores detailed descriptions of the datasets, including information about the source, contents and attributes.

[5] D03.01 Architecture Artefacts High Level Version

7.1.3 Annual review

Once a year the data holders must review quality and compliance of their dataset description and document the result of the review. The national dataset catalogue should make it possible for data holder to document the annual review.

7.1.4 Maintenance graphical user interface

A maintenance graphical user interface (GUI) provides data holders with an accessible tool to update and maintain dataset descriptions. Key features include:

- updating properties, correcting errors or submitting new records;
- ensuring that metadata remains accurate, complete and aligned with minimum dataset description requirements.

This GUI streamlines metadata management, making it easier for data holders to fulfil their responsibilities under the EHDS Regulation, including keeping metadata up-to-date and compliant.

7.1.5 Metadata quality process

Referring to the maintenance GUI, the dataset description may need further enrichment to complement the automatic validation. It is optional to support to add more detailed annotations, linking external sources or including additional context that increases the metadata's value for research or policymaking.

Quality assurance ensures that metadata remains accurate, consistent and useful over time this provided through a feedback loop from data users in the search processes. This means that data users may discover that data are either missing or incomplete according to the description of the metadata and therefore flagged.

7.1.6 Metadata provenance

The system must log all updates to metadata and maintain a history of previous versions of dataset descriptions. This ensures that stakeholders can track changes to metadata over time (traceability). Previous versions of metadata can be restored if necessary, supporting quality assurance and compliance monitoring (version control). While not explicitly mandated by the EHDS Regulation, metadata provenance is essential for ensuring reliability and accountability in national dataset catalogues.

7.1.7 EU catalogue integration

Member States, as written in the Regulation, must have an NCP which is the only system connected to the HealthData@EU Central Platform. The connection happens only through eDelivery Domibus. Member States send metadata about datasets, and it is updated to the HealthData@EU Central Platform.

It is mandatory to export dataset descriptions in the HealthDCAT-AP standard to the EU central dataset catalogue via the NCP.

This is further described in task 7.3 (see Annex 3 Related legal framework).

7.1.8 Public application programming interface

EHDS Regulation articles 57 and 77 state that Member States should make national dataset catalogues publicly available. This can be accomplished via a public API.

7.1.9 Metadata discovery graphical user interface

This capability provides a user-friendly interface for stakeholders to search, filter and explore metadata (dataset descriptions) within the national dataset catalogue. By enabling intuitive navigation and search functionalities, the GUI enhances discoverability and accessibility of datasets for secondary use. While not explicitly required by the EHDS Regulation, this feature supports the mandate for public access through electronic means and complements programmatic access methods like APIs.

8 Open questions/ unresolved issues

- Alternative terminology to the ones used in this text?
- In a next step, we will consider new published documents, compilation of technical documents from Healthdata@EU Central platform. describing the technical development of the central platform. Especially the work considering dataset descriptions, dataset catalogues, HDAB-function.[6] [7] [8] [9]
- Should we mention the handling and creation of persistent identifiers in this document? This applies both to data holders, the dataset descriptions and HDABs.
- A metadata catalogue can be seen as consisting of three levels: dataset level, content level, and variable level. These levels work together to provide a hierarchical and detailed description of data, enabling effective navigation, analysis and utilisation. Each level serves a distinct purpose, contributing to the comprehensive management of metadata.

  Based on the interpretation of the EHDS Regulation, Articles 51, 57 and 77 collectively outline the requirements for a national dataset catalogue. Article 51 specifies the categories of electronic health data to be made available for secondary use. Article 57 assigns health data access bodies the task of making a national dataset catalogue publicly accessible through electronic means. Article 77 further mandates that such catalogues include dataset descriptions in the form of metadata, detailing the source, scope, main characteristics and nature of the electronic health data in each dataset.

  How do these articles, when considered together, influence the implementation and design of a metadata catalogue that aligns with the EHDS Regulation? Furthermore, how does the combination of these articles affect the choice of national infrastructure for HDABs?

[6] **ICT governance framework and procedures catalogue – scale-up version** (D04.03) Directorate General for Health and Food Safety, 24-11-17 Digital, EU4Health and Health Systems Modernisation (SANTE.C), Version: 1.3, Template Version: 2.5

[7] **D03.03 System Specifications - Scale-Up Version**
Specific Contract n° 022 under Framework Contract DI7925-DI7932– BEACON – Lot 2: Analysis and design of the European Health Data Space infrastructure for secondary use of health data (HealthData@EU). 24-12-10, Version: 1.0, Template Version: 2.5

[8] **D03.03 Architecture Artefacts - Scale-Up Version**
Specific Contract n° 022 under Framework Contract DI7925-DI7932– BEACON – Lot 2: Analysis and design of the European Health Data Space infrastructure for secondary use of health data (HealthData@EU), 24-12-04, Version: 1.0, Template Version: 2.5

[9] Directorate General for Health and Food Safety, 24-11-17 Digital, EU4Health and Health Systems Modernisation (SANTE.C), Requirement catalogue for scale-up-version 24-12-17, version 3.1, template version 2.5

9 List of Annexes

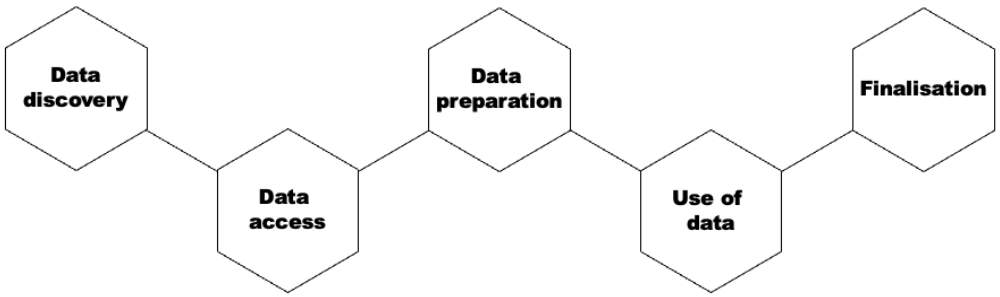
Annex	Title
Annex 1	Use journey
Annex 2	Key terminology - Glossary
Annex 3	Related legal framework
Annex 4	Relations between tasks in TEHDAS 2
Annex 5	Methodology

9.1 Annex 1 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 policymaking, within the European Health Data Space (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 5: 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 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 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. 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.

9.2 Annex 2 Key terminology – Glossary

Table 1: Glossary according to TEHDAS2. It will be aligned across all TEHDAS2 deliverables in next step.

Name	Description
Access	Processing by a data user of data that has been provided by a data holder, in accordance with specific technical, legal or organisational requirements, without necessarily implying the transmission or downloading of such data. Ref. Regulation (EU) 2022/868
Aggregated data	Data combined or collected together in summary or other form such that the data cannot be identified with any individual. Ref. Legal Information Institute, Cornell University
Anonymisation	The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject. Removing personally identifiable information, so as to definitively not allow the identification of the data subjects. The methods used to anonymise the data are context dependent. Ref. DGA COM(2020) 767 final https:// eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN
Anonymous data	Information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. Anonymisation is permanent and irreversible. Principles of data protection do not apply to anonymous information. Ref. GDPR Recital 26
Capability	Is the ability to execute a specified course of action or to achieve a desired outcome. We use this term to avoid specifying how Member States implement functionalities. Ref. Wikipedia
Data	Any digital representation of acts, facts or information and any compilation of such acts, facts or information, Ref. DGA COM (2020) 767 final
Data catalogue	Organised inventory of data assets in the organisation. It uses metadata to help organisations manage their data. It also helps data professionals collect, organise, access and enrich metadata to support data discovery and governance. Ref. Oracle

Data concerning health	Personal data related to the physical or mental health of a natural person, including the provision of health care) services, which reveal information about his or her health status Ref. Article 4(15) of the GDPR and EHDS article 57
Data Governance Act (DGA)	Legislative proposal of the European Commission that aims to create a framework which will facilitate data-sharing. Ref. Proposal for a regulation of the European Parliament and of the Council on European data governance on European data governance, (Data Governance Act) COM (2020) 767 final
Data holder	Legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control. Ref. DGA COM (2020) 767 final
Data model	A data model is an abstract framework that organises the data managed by a computer system. It illustrates the entities involved, their properties, and the relationships between them. Ref. Data model - Wikipedia
Data processing	Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction Ref. Article 4(2) of the GDPR
Data quality	Comprehensive view of usefulness of data to support decision making. Data quality is defined as “fitness for use” for users’ needs. The OECD views quality in terms of seven dimensions: relevance, accuracy, credibility, timeliness, accessibility, interpretability and coherence. Cost-efficiency is included in the Quality Framework as the eight items. Ref. HIMSS Dictionary of Healthcare Information Technology, OECD
Data repository	The data repository is a large database infrastructure, several databases, that collect, manage, and store data sets for data analysis, sharing and reporting. Ref. Digital Guardian
Data request	Discovery procedure in which the requesting party asks another person for specified information or requests the production of documents. Ref. Art. 69 EHDS procedure

Dataset	Any organised collection of data. Ref. OECD
"Dataset catalogue" in the EHDS	'Dataset catalogue' is a synonym to "metadata catalogue" and means a collection of datasets descriptions, which is arranged in a systematic manner and consists of a user-oriented public part, where information concerning individual dataset parameters is accessible by electronic means through an online portal. Ref. European Parliament legislative resolution of 24 April 2024 on the proposal for a regulation establishing the European Health Data Space
Data sharing	Provision by a data holder of data to a data user for the purpose of joint or individual use of the shared data, based on voluntary agreements, directly or through an intermediary. Ref. OECD
Dataset description	"Dataset description" refers to the structured metadata describing a dataset within the national dataset catalogue. This includes information such as the dataset's source, scope, main characteristics and conditions for access, as defined in Article 77 of the EHDS Regulation. It should not be confused with the "description" property in DCAT-AP or HealthDCAT-AP metadata standards
Data source	Specific data set, metadata set, database or metadata repository from where data or metadata are available.
Data subject	Natural person whose personal data is processed by a data controller or processor. Ref. GDPR
Data user	Natural or legal person who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or non-commercial purpose. Ref. EHDS Regulation
EU dataset catalogue	"EU dataset catalogue" refers to the centralised metadata repository within the HealthData@EU platform. This catalogue does not store datasets but aggregates and organises dataset descriptions (metadata) received from national dataset catalogues via the NCP. It enables stakeholders to locate metadata across Member States.
European Health Data Space (EHDS)	The European Health Data Space (EHDS) is an EU initiative aimed at creating a unified health data ecosystem with common rules, standards, and governance. It seeks to provide a secure framework for using anonymised health data for research, innovation, and policymaking. The EHDS builds on existing regulations like the GDPR, Digital Governance Act and is a cornerstone of the European Health Union. It facilitates better coordination among Member States to improve healthcare delivery, prevention, and response to health crises.

	Ref. EHDS Regulation
Functional and non-functional requirements	Functional requirements specify what a system should do and are typically phrased as "the system must do [requirement].” In contrast, non-functional requirements describe the system's qualities or characteristics and are often expressed as "the system shall be [requirement]. Ref.
FAIR principles	The FAIR principles are guidelines aimed at enhancing the management and sharing of scientific data, promoting its utility and transparency. They stand for: Findable: Data should be easy for humans and computers to locate, with richly described searchable metadata. Accessible: Data should be retrievable through a clear protocol, ideally open and free, while adhering to legal and ethical requirements. Interoperable: Data should be in formats that allow integration with other data sources and usability across various tools and platforms. Reusable: Data should be well-documented and described to support replication and further research, including information on provenance and usage licenses. The goal is to maximise the value and impact of scientific data by making it more accessible, usable and transparent.
GDPR	General Data Protection Regulation (GDPR) is a regulation in EU law strengthening and harmonising EU/EEA procedures concerning the collection, storage, processing, access, use, transfer and erasure of personal data. Ref. The General Data Protection Regulation (EU) 2016/679 (GDPR)
Health data access body (HDAB)	An entity designated by an EU Member State responsible for managing and overseeing access to health data within the framework of the European Health Data Space (EHDS). The HDAB's role includes ensuring that health data is accessed and shared in compliance with legal and ethical standards, facilitating the secondary use of health data for purposes such as research, policymaking and innovation. Additionally, the HDAB manages the national metadata catalogue, supports interoperability and ensures that data sharing aligns with EU and national regulations. Ref.
Health	State of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Ref. WHO
HealthData@EU	Infrastructure for secondary use of health data
HealthData@EU Central Platform	The central EU metadata catalogue for health

HealthDCAT-AP	An extension of the DCAT-AP (Data Catalogue Vocabulary Application Profile for Data Portals in Europe) standard, specifically tailored for the health sector. It provides guidelines for describing and sharing metadata related to health data in a standardised way across the EU. By using HealthDCAT-AP, health data catalogues can ensure interoperability and harmonisation, allowing health data to be easily discovered, accessed and used across different systems and national borders. The standard supports the European Health Data Space (EHDS) by facilitating the secondary use of health data while adhering to legal and technical requirements.
Joint action (JA)	Joint action constitutes an operational action by the Member States. Ref. European Commission
Metadata	Data collected on any activity of a natural or legal person for the purposes of the provision of a data sharing service, including the date, time and geolocation data, duration of activity, connections to other natural or legal persons established by the person who uses the service.
Metadata catalogue	definition 1. Metadata catalogues describe the available data collections in a repository or hub Ref. Glossary for HealthyCloud) definition 2. “A metadata catalogue is an organized collection of metadata designed to support data discovery, management, and interoperability by providing a structured repository of information about datasets, resources, and their attributes." Ref. Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9. DOI: 10.1038/sdata.2016.18
Metadata record	The information about a dataset, stored in the catalogue
Metadata repository	A place to store metadata
Minimum viable product (MVP)	The simplest version of a product that includes just enough features to satisfy early adopters and gather feedback for further development. The idea is to launch quickly with a basic set of functionalities that address the core problem or need, allowing an organisation to test its assumptions, learn from user behaviour and make informed improvements
National dataset catalogue	National dataset catalogue” refers to the metadata catalogue managed by the Health Data Access Body (HDAB) at the Member State level. This catalogue contains machine readable descriptions (metadata) of datasets relevant for secondary use under the EHDS Regulation. “Dataset description” refers to the structured metadata describing a dataset within the national dataset catalogue.

	This includes information such as the dataset's source, scope, main characteristics and conditions for access, as defined in Article 77 of the EHDS Regulation. It should not be confused with the "description" property in DCAT-AP or HealthDCAT-AP metadata standards
Non-personal data	Data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679. Ref. https://www.european-health-data-space.com/
Open data	Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike. Ref. Open Data Handbook
Personal data	Any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Ref. Article 4(1) of the GDPR
Primary use of data	Use of personal health information by the organisation or entity that produced or acquired these data in the process of providing real-time, direct care of an individual. Ref. C. Safran et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan-Feb. 14(1): 1–9. doi: 10.1197/jamia.M2273
Processing	Processing covers a wide range of operations performed on personal data, including by manual or automated means. It includes the collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of personal data. Ref. European Commission
Pseudonymisation	The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person Ref. https://www.healthinformationportal.eu/about-platform
Re-use	"Use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data

		between public sector bodies purely in pursuit of their public tasks." Ref. DGA COM (2020) 767 final
Secondary purpose of personal data		Processing of personal data for a purpose other than the primary purpose referred to in primary purpose of personal data. Ref. Act on the Secondary Use of Health and Social Data. Finland. 552/2019
Secondary source of data		Organisation or individual other than those responsible for the collection and aggregation of data from their initial source. Secondary sources may redistribute information received from the primary source either in their initial form or after some transformation including further aggregation, reclassification or other manipulation such as seasonal adjustment. Ref. OECD
Secondary use of data		Secondary use of data occurs when data is used for a purpose different from the purpose for which the data was initially collected. (Note: secondary use of data is not the same as re-use of data.) Ref. Code of Practice on Secondary Use of Medical Data in Scientific Research Projects - 27 Aug 2014 Final Draft. Innovative Medicines Initiative
Systems architecture		Systems architecture is a number of architectural views, used to describe a software system. Ref. Wikipedia
User story		User story is an informal, natural language description of features of a software system. They are written from the perspective of an end user or user of a system, Ref. Wikipedia

9.3 Annex 3 Related legal framework

Table 2: Related legislative articles

Related legislative article applicable to national technical specification for national health data access body				
1. Data Governance Act (DGA) Article 8				
Single information points	“Member States shall ensure that all relevant information concerning the application of Articles 5 and 6 is available and easily accessible through a single information point. Member States shall establish a new body or designate an existing body or structure as the single information point. The single information point may be linked to sectoral, regional or local information points. The functions of the single information point may be automated provided that the public sector body ensures adequate support. 3.6.2022 EN Official Journal of the European Union L 152/25 ”	“The single information point shall be competent to receive enquiries or requests for the re-use of the categories of data referred to in Article 3(1) and shall transmit them, where possible and appropriate by automated means, to the competent public sector bodies, or the competent bodies referred to in Article 7(1), where relevant. The single information point shall make available by electronic means a searchable asset list containing an overview of all available data resources including, where relevant, those data resources that are available at sectoral, regional or local information points, with relevant information describing the available data, including at least the data format and size and the conditions for their re-use.”	“The single information point may establish a separate, simplified and well-documented information channel for SMEs and start-ups, addressing their needs and capabilities in requesting the re-use of the categories of data referred to in Article 3(1).”	“The Commission shall establish a European single access point offering a searchable electronic register of data available in the national single information points and further information on how to request data via those national single information points.”
2. EHDS - final version December 2024				

EHDS Art 57: Tasks of health data access bodies	***Health data access bodies shall carry out the following tasks:***			
	(j) 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;	(l) The national dataset catalogue referred to in point (j)(i) of this paragraph shall also be made available to single information points under Article 8 of Regulation (EU) 2022/868."		
EHDS Art 60: Duties of health data holders	"3. The health data holder shall communicate to the health data access body a description of the dataset it holds in accordance with Article 77. The health data holder shall, at a minimum on an annual basis, check that its dataset description in the national dataset catalogue is accurate and up to date.	4. Where a data quality and utility label accompany the dataset pursuant to Article 78, the health data holder shall provide sufficient documentation to the health data access body for that body to verify the accuracy of the label."		
EHDS Art 77: Dataset description *and dataset catalogue*	1. Health data access bodies shall, through a publicly available and standardised machine-readable dataset catalogue, provide a description in the form of metadata of the available datasets and their characteristics. The description of each dataset shall include information concerning the source, scope, main characteristics, and nature of the electronic health data in the dataset and the conditions for making those data available.	2. The dataset descriptions in the national dataset catalogue shall be available in at least one official language of the Union. The dataset catalogue for Union institutions, bodies, offices and agencies provided by the Union health data access service shall be available in all official languages of the Union.	3. The dataset catalogue shall be made available to single information points established or designated under Article 8 of Regulation (EU) 2022/868.	4. "By means of implementing acts, set out the minimum elements health data holders are to provide for datasets and the characteristics of those elements."

EHDS Art 78: Data quality and utility label	*"3. The data quality and utility label shall cover the following elements, where applicable:*			
	(a) for data documentation: metadata, support documentation, the data dictionary, the format and standards used, the source of the data and, where applicable, the data model;	(b) for assessment of technical quality: completeness, uniqueness, accuracy, validity, timeliness and consistency of the data;	(c) for data quality management processes: the level of maturity of the data quality management processes, including review and audit processes, and bias examination;	(f) for information on data modifications: merging and adding data to an existing dataset, including links with other datasets."
	*"6. B*y means of implementing acts, set out the visual characteristics and technical specifications of the data quality and utility label, based on the elements referred to in paragraph 3 of this Article."			
EHDS Art 79: EU dataset catalogue	"1. The Commission shall establish an EU dataset catalogue connecting the national dataset catalogues established by the health data access bodies in each Member State as well as the dataset catalogues of authorised participants in HealthData@EU.	2. The EU dataset catalogue, the national dataset catalogues and the dataset catalogues of authorised participants in HealthData@EU shall be made publicly available."		
EHDS Art 80: Minimum specifications for datasets of high impact	"The Commission may, by means of implementing acts, determine the minimum specifications for datasets of high impact for secondary use, taking into account existing Union infrastructures, standards, guidelines and recommendations."			

9.4 Annex 4 Relations between tasks in TEHDAS 2

Table 3: Relations between tasks

Work package 5			**Work package 6**	**Work package 7**
Task 5.1.1 Data holders	**Task 5.2.1 HDAB**	**Task 5.3 HDAB**	**Task 6.1 Data holders**	**Task7.2**
Guideline on data description categorising and describing datasets according to Art 33	Guideline on conditions, purposes and categories of data permissible to guide limitations on further processing	Guideline on enrichment of health datasets	Guidelines for data holders making health data available	Technical specifications for data minimisation and de-identification
Task 5.1.2 Data holders	**Task 5.2.2 HDAB**			**Task 7.3**
Guideline for data holders to use HealthDCAT-AP to describe their datasets.	Technical specification on the national metadata catalogue and its maintenance.			Technical specification Implementation of a common IT infrastructure between the NCP and the EU@healthdata Central Platform

9.5 Annex 5 Methodology

The method creating the Milestone report is based on five steps.

Review of literature and legislative proposals

EU legal acts: This stage involved a detailed analysis of EU legal frameworks applicable to health data sharing and governance. The purpose was to ensure the technical specification aligns with current EU regulations and can accommodate future legal updates. (Annex 2 Related legal framework)

TEHDAS2 – project plan[10]: The review of the TEHDAS2 project plan was crucial to understand the scope, objectives and deliverables expected from the project, ensuring that the technical specification supports these goals.

Deliverable template by Sitra[11]: The deliverable template provided by Sitra, has been used to structure the specification in a manner that aligns with established guidelines and facilitates clear, actionable outputs.

HealthData@EU Platform IT-Architecture Report[12]: The IT architecture report has been examined to incorporate technical considerations, data flow and constraints into the specification, ensuring compatibility with the IT infrastructure between national and EU level.

HealthData@EU Pilot Final Report 6.2[13]: The draft reports from pilot projects have been analysed to get insights from practical challenges and feedback, which informed adjustments and improvements in the technical specifications. The reports also provided draft HealthDCAT-AP standard and system infrastructure.

Community of Practice HDAB Forum: Discussions in the HDAB forum have been used to incorporate diverse perspectives on metadata catalogue level and experiences from various EU Member States, enhancing the robustness and applicability of the specification.

Workshops and Discussions

The development of the technical specifications involved a series of workshops and discussions. Internally, the task and working group leads focused on drafting the technical solution. The technical architecture group worked on aligning the system's infrastructure with the requirements for managing and sharing health data, while the legal group ensured compliance with legal EHDS regulation.

Key contributors from Belgium, France, Luxembourg, Norway and Sweden collaborated to review and refine the specification. A review board with representatives from Belgium, the Czech Republic, Ireland, Portugal, Slovenia, Sweden, Germany and the Netherlands provided feedback to enhance clarity, precision and overall quality. Additionally,

[10] Projectplan Sitras homepage
[11] Templates Sitras homepage
[12] D03.01 Architecture Artefacts High Level Version

[13] Healthdata@EU Pilot, https://ehds2pilot.se

consultations with DG SANTE ensured the specifications were compatible with current and planned EU-wide IT infrastructures, making them practical and feasible for implementation.

Collaborative Discussions and External Contributions

- **Major Contributors (Task 5.2.2 of TEHDAS2)**: Collaborative discussions with contributors from Belgium (Sciensano), France (Health Data Hub), Luxembourg (National Data Service), Norway (Norwegian Institute of Public Health), and Sweden (Swedish e-Health Agency) covered sections of the technical specification. These discussions supported the drafting of the Milestone report, ensuring comprehensive coverage of necessary aspects.
- **Review Board (Task 5.2.2 of TEHDAS2)**: The review board, including representatives from Belgium (Health Data Authority), the Czech Republic (Ministry of Health), Ireland (Department of Health), Portugal (Shared Services for the Ministry of Health), Slovenia (National Institute of Public Health), Sweden (Region of Skåne and Uppsala University), Germany (Gematik) and the Netherlands (Nationaal ICT-instituut in de Zorg), provided critical oversight and feedback. Their contributions enhanced the clarity of the final Milestone document.
- **DG SANTE/HealthData@EU IT Infrastructure**: Consultations with DG SANTE discussed the scope of the technical specifications aligned with existing and planned IT infrastructures for health data across the EU, enhancing feasibility and integration potential.

User story collection

User stories were gathered using the Connextra template (“As a [user], I want [feature], so that [benefit]”)[14]

Although writing user stories with real end users in mind is considered best practice, some stories reflecting the perspectives of key stakeholders, such as the HDAB and data holders, were included. This approach was taken to better address the needs of critical actors in relation to the national metadata catalogue.

Writing process

Collaborative writing process: 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.

Anchoring and review process: Each section of the document underwent an anchoring and review process, where initial drafts were critiqued and refined by subject matter experts to ensure accuracy and comprehensiveness.

Review process

[14] https://en.wikipedia.org/wiki/User_story

Review board: Representatives of five countries (see above) provided essential feedback throughout the writing process and were responsible for reviewing the content in the specification of the metadata catalogue.

Sitra: Sitra played a key role in the final review, offering insights from their experience in managing joint actions and ensuring the specification meets high standards of quality and utility.

DG SANTE: Final approval and feedback from DG SANTE were crucial in ensuring the specification meets all EU requirements and is poised for successful implementation.