



European Health Data Space for Secondary Use @CY - CY-EHDS-2ND

Proposal number: 101129405

D6.1 nHDsC Requirements and Specifications

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

Author(s)

Name	Surname	Country	Organisation	Role of the Organisation
Nicoletta	Prentza (Leader)	Cyprus	University of Cyprus (UCY)	Affiliated Entity (AE)
Maria	Papaioannou	Cyprus	UCY	AE
Ioannis	Rodosthenous	Cyprus	UCY	AE
Constantinos	Pattichis	Cyprus	UCY	AE

Reviewed by

Name	Surname	Country	Organisation	Role of the Organisation
Ioannis	Konstantinou	Cyprus	UCY	AE
Erietta	Iouliauou	Cyprus	NeHA	Competent Authority (CA)
Antonis	Stylianides	Cyprus	NeHA	CA
Christos	Schizas	Cyprus	NeHA	CA
Theodoros	Kyprianou	Cyprus	NeHA	CA

Abbreviations

AE	Affiliated Entity
CRUD	Create, read, update, and delete
CP	Central Platform
CoP	Community of Practice
CY-EHDS-2nd	European Health Data Space for Secondary Use @CY
CY-HDAB Portal	Cypriot Health Data Access Body Portal
DGA	Data Governance Act
DBC	Digital Business Capabilities
DG DIGIT	Directorate-General for Informatics
DQUL	Data Quality and Utility Label
EHDS	European Health Data Space
EC	European Commission
EU	European Union
FAIR	Findability, Accessibility, Interoperability and Reusability
HDAB	Health Data Access Body
Health DCAT-AP	Health Data Catalogue Vocabulary Application Profile
IAA	Identification, authentication, and authorization
MFA	multifactor authentication
MS	Member States
NCP	National Contact Point
NeHA	National eHealth Authority
nHDsC	National Health Dataset Catalogue
TEHDAS2	Second Joint Action Towards the European Health Data Space
RBAC	role-based access control
UCY	University of Cyprus
WP	Work Package

Executive Summary

The National Health Dataset Catalogue (nHDsC) is one of the core modules of the European Health Data Space for Secondary Use @CY (CY-EHDS-2nd) Central Platform (CP). nHDsC is being developed to fulfil the obligations established under Articles 60 and 79 of the European Health Data Space (EHDS) Regulation European Union (EU) 2025/327¹. The system is a critical component of the broader EHDS Regulation infrastructure, serving as a central metadata registry for health datasets available for secondary use within the Republic of Cyprus jurisdiction. Its core purpose is to ensure transparency, discoverability and interoperability with the HealthData@EU dataset catalogue, thereby enabling secure, lawful and efficient secondary use of health data across the EU.

The scope of this deliverable is to specify the requirements for the Health Data Access Body (HDAB) of Cyprus National eHealth Authority (NeHA) to develop and deploy the nHDsC module. It includes use cases, functional requirements, business logic and some considerations about the technical solution. This work built upon the outputs of the Analysis & Design of HealthData@EU infrastructure², as well as the Second Joint Action Towards the European Health Data Space (TEHDAS2)³, to ensure alignment and interoperability with the other HDABs. Any not explicitly specified aspects will be defined or revised in the upcoming deliverable D6.2 'nHDsC: Implementation and piloting'.

¹ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202500327

² <https://acceptance.data.health.europa.eu/>

³ <https://tehdas.eu/>

Table of contents

1	INTRODUCTION	7
1.1	Guidance for Interpreting the CY-EHDS-2nd Requirements and Specifications.....	7
1.1.1	Annexes Applicable Across All Requirement and Specifications (deliverables D5.1-D9.1).....	7
1.1.2	Recommended Reading Order	8
1.2	D6.1 Introduction	8
1.3	nHDsC & HealthData@EU CP	10
1.4	Key terminology	10
1.5	Interdependencies	10
1.6	Overview	11
2	ALIGNMENT WITH HEALTHDATA@EU CENTRAL PLATFORM.....	11
3	NHDSC OVERVIEW	12
3.1	Overview	12
3.2	Actors	13
3.3	Business uses cases	13
3.3.1	Dataset creation	13
3.3.2	Dataset publication and update	14
3.3.3	Searching for a dataset.....	15
4	REQUIREMENTS.....	16
4.1	Functional requirements	16
4.1.1	User Identification, authentication and authorization	16
4.1.2	Dataset creation	16
4.1.3	Data publication and update	21
4.1.4	Searching for a dataset.....	24
4.1.5	Interoperability of health data (HealthDCAT-AP)	26
4.1.6	Communication	27
5	SPECIFICATIONS.....	28
5.1	nHDsC Processes	28
5.1.1	Dataset creation	28
5.1.2	Dataset description updates	29
5.1.3	nHDsC & HealthData@EU	29
5.2	nHDsC module of CY-EHDS-2nd CP.....	30
5.2.1	Core features	30
5.2.2	Proposed services	31
5.2.3	Integrations	31
5.3	Health DCAT-AP	32

5.3.1	Health DCAT-AP Editor	32
5.3.2	Health DCAT-AP Validator	33
6	REFERENCES	34

List of Figures

Figure 1:	CY-EHDS-2nd nHDsC module (as implemented in Work Package (WP6)	9
Figure 2:	National Dataset Catalogue, from Community of Practice (CoP)	10
Figure 3:	nHDsC overview (a revised version of DBC blueprint for nHDsC).....	12
Figure 4:	Dataset creation Use case.....	14
Figure 5:	Dataset publication and update Use case.....	15
Figure 6:	Searching for a dataset Use case	15
Figure 7:	Dataset creation process.	28
Figure 8:	Dataset description updates process flow.....	29
Figure 9:	Schematic overview of the information flow described by DG SANTE.....	30
Figure 10:	HealthDCAT-AP editor.....	33
Figure 11:	Health DCAT-AP Validator (source: HealthData@EU CP Open-source R3, deliverables).....	34

List of Tables

Table 1:	CY-EHDS-2nd nHDsC Actors as defined in <i>D5.1 Annex D Actor Definitions – CY-EHDS-2nd (v1.0)</i>	13
Table 1:	Integrated communication use cases	27
Table 2:	Cross-border/National Connector communication use cases	27
Table 3:	nHDsC module core features	30
Table 4:	CY-EHDS-2nd CP Modules Integration Overview	31

1 INTRODUCTION

1.1 Guidance for Interpreting the CY-EHDS-2nd Requirements and Specifications

This guidance text is intended to help the reader navigate in a meaningful and structured manner through the first series of technical deliverables concerning the requirements and specifications of the CY-EHDS-2nd national infrastructure. It provides essential context for understanding how the various components fit together and how supporting annexes serve as shared reference points across modules. This same text is included identically at the start of each deliverable in the requirements and specifications series (D5.1 to D9.1) to ensure consistency in interpretation and approach.

This deliverable is part of a coordinated set of Requirements and Specifications documents (D5.1 to D9.1) that together define the design and implementation logic for the national infrastructure supporting the secondary use of health data in the Republic of Cyprus, under the CY-EHDS-2nd project. Each module addresses a distinct part of the infrastructure, yet all share a unified architectural foundation, legal interpretation, and design methodology.

To support this shared structure and ensure full coherence across the Work Packages, a series of five common annexes (Annex A to Annex E) has been developed. These annexes are referenced throughout this and all other deliverables and are essential for a complete and accurate understanding of the specifications that follow.

The reader is strongly advised to begin by reading Annex A of D5.1. This annex provides the architectural and conceptual entry point for understanding the CY-EHDS-2nd infrastructure and explicitly references all other annexes. Only once Annex A is read, should the reader proceed to the main content of this or any other Requirements and Specifications deliverable.

1.1.1 Annexes Applicable Across All Requirement and Specifications (deliverables D5.1 - D9.1)

The following annexes provide cross-cutting foundations that support the interpretation, alignment and drafting of all CY-EHDS-2nd requirements and specifications deliverables (D5.1 to D9.1). Each annex focuses on a specific dimension of the project, from technical architecture to digital business capabilities, semantic mapping, actor roles and terminology. These annexes are shared across the series to ensure coherence, eliminate duplication and align the national infrastructure design with the HealthData@EU infrastructure model.

All annexes listed below are formally included as annexes of the D5.1 deliverable and are referenced as needed across the remaining deliverables. Each annex should be understood as integral, not supplemental, to the specification content.

1.1.1.1 Annex A – CY-EHDS-2nd Architecture and Central Platform Infrastructure

This annex presents the overall architecture of the CY-EHDS-2nd infrastructure and introduces its main functions and capabilities. It provides a high-level vision of the system's structural design and its role within the broader EU eHealth ecosystem for secondary use. It explains how the five core components, nHDsC, DAAMs, SPE, National Connector and Quality Toolbox, are interconnected through a central coordination layer (the Central Platform Layer, CPL), which ensures secure, interoperable and legally compliant access to health data for secondary use. The annex also outlines the long-term modularity and scalability of the platform, highlighting its capacity to integrate future EU deliverables, including outputs from TEHDAS2, QUANTUM, and the upcoming EHDS implementing acts.

1.1.1.2 Annex B – HDAB Digital Business Capabilities

This annex contains a structured overview of the Digital Business Capabilities (DBC) that HDABs are expected to implement, as defined by the European Commission in the context of the EHDS. While these capabilities are not themselves regulatory, they are based on and support compliance with the EHDS Regulation and provide critical guidance for implementation planning. Each capability is listed with its code, description, interdependencies, and its

applicability within CY-EHDS-2nd. These capabilities were further clarified by the EC during the 2nd EHDS2 Community of Practice General Assembly (February 2025) and are formally recorded in the EHDS2 CoP Confluence.

1.1.1.3 Annex C – Indicative Mapping to HealthData@EU Central Platform Infrastructure

This annex provides a basic, high-level mapping between selected components of the CY-EHDS-2nd national infrastructure and the HealthData@EU Central Platform. The mapping focuses on core modules that are most relevant for interoperability and cross-border integration. While the alignment is not exhaustive or granular, it offers a conceptual view of how Cyprus national solution reuses, adapts, or complements elements of the EU reference platform. This annex supports early-stage validation of design choices and highlights areas of future alignment and integration.

1.1.1.4 Annex D – Common Actor Definitions – CY-EHDS-2nd (v1.0)

This annex defines all functional actors involved in the implementation and operation of CY-EHDS-2nd across WPs 5 to 9. Each actor is described according to its formal role, regulatory grounding and operational scope. It is aligned with the final EHDS Regulation (EU) 2025/327, Scale-Up documentation (Dec 2024), and HealthData@EU CP deliverables (Mar 2025). To ensure consistency and avoid ambiguity, any actor referenced in this deliverable must correspond exactly to the naming and definition in this annex. Updates or additions to actor roles must be formally proposed and approved at the project level.

1.1.1.5 Annex E – CY-EHDS-2nd Key Terminology

Annex E provides a project-wide glossary of terms used in D5.1 to D9.1. Each term is defined based on its actual use across the deliverables, and its meaning is aligned with the HealthData@EU CP documentation, TEHDAS2 outputs (up to May 2025), and the EHDS Regulation (EU) 2025/913. This glossary ensures conceptual clarity and harmonised understanding of terminology across the entire CY-EHDS-2nd national implementation effort.

1.1.2 Recommended Reading Order

To ensure a coherent understanding of the CY-EHDS-2nd architecture and its module-specific implementations, the following reading order is recommended:

1. D5.1 – Annex A: CY-EHDS-2nd Architecture and CP Infrastructure (**mandatory starting point**)
2. D5.1 – All other annexes, especially Annex D: Common Actor Definitions – CY-EHDS-2nd (v1.0)
3. D5.1 – DAAMs Requirements, Specifications and Prototype
4. D6.1 – nHDsC Requirements and Specifications
5. D9.1 – Health Data Quality Enhancement Toolbox Requirements and Specifications
6. D7.1 – SPE: Requirements and Specifications
7. D8.1 – National Connector for Cross-border gateway Requirements and Specifications

1.2 D6.1 Introduction

The purpose of this document is to define the requirements and specifications for the nHDsC, that will be implemented and deployed as a core module of the CY-HDAB-2nd CP, in alignment with Chapter IV of the EHDS Regulation [8], and broader EU data governance frameworks, such as the Data Governance Act (DGA) (Figure 1).

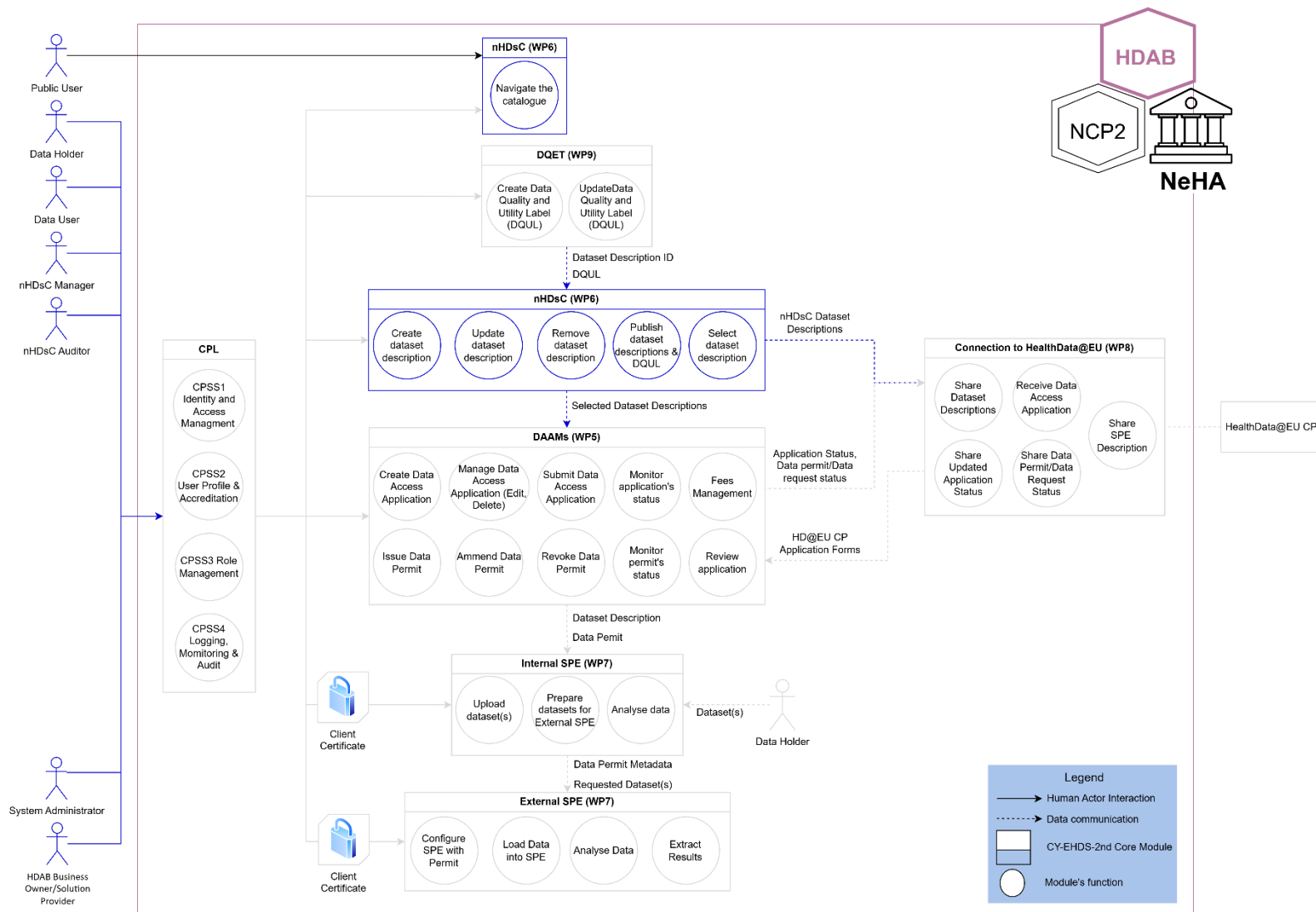


Figure 1: CY-EHDS-2nd nHDsC module (as implemented in Work Package (WP6))

1.3 nHDsC & HealthData@EU CP

A key objective of the EHDS Regulation 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. To achieve this, Member States (MS) need to develop a national dataset catalogue that can federate seamlessly into the EU dataset catalogue, through the adoption of the Health Data Catalogue Vocabulary Application Profile (Health DCAT-AP) standard (Figure 2).



Figure 2: National Dataset Catalogue, from Community of Practice (CoP)⁴

nHDsC minimum set of technical and operational capabilities:

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

nHDsC will serve as a cornerstone for facilitating the secondary use of health data, ensuring their Findability, Accessibility, Interoperability and Reusability (FAIR)⁵ principles across the EU.

1.4 Key terminology

- “National dataset catalogue”: refers to the metadata catalogue managed by the HDAB at the MS 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.
- “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 MS.

1.5 Interdependencies

- CY-EHDS-2nd Architecture and CP Infrastructure (D5.1, ANNEX A).
- CY-EHDS-2nd WP5: DAAMs (D5.1).
- CY-EHDS-2nd WP8: National Connector & Cross-border Gateway (D8.1).

⁴ <https://webgate.ec.europa.eu/fpfis/wikis/spaces/EHDS/pages/1818400440/National+Dataset+Catalogue+-+implementation+blueprint>

⁵ <https://www.go-fair.org/fair-principles/>

1.6 Overview

The rest of the document is organized as follows:

- Section 2: Alignment with HealthData@EU CP
- Section 3: nHDsC Overview
- Section 4: Requirements
- Section 5: Specifications

2 ALIGNMENT WITH HEALTHDATA@EU CENTRAL PLATFORM

Cyprus, through the CY-EHDS-2nd grant agreement, is establishing its national infrastructure for the secondary use of health data entirely from scratch. Given the absence of an existing national framework, the project's strategic approach is to align fully with the HealthData@EU infrastructure and the CP model developed by the EC [1-2].

To achieve this, for the nHDsC component of the CY-EHDS-2nd CP:

- Bases where applicable its business requirements and technical specifications on the:
 - o HealthData@EU Central Services Release 3 deliverables [1-2].
 - o ANALYSIS & DESIGN of HealthData@EU Infrastructure deliverables [3-7].
 - o TEHDAS2 M5.3 Draft technical specifications on the national metadata catalogue: guidelines and technical specifications for national HDABs to receive and transform dataset descriptions provided by data holders at the national level into a specification, as part of a national dataset catalogue [9].
- Adopts the HDAB Digital Business Capabilities (DBC) blueprint [10].
- Intends to adapt the European solution as fully as possible, thereby avoiding duplication of design efforts and ensuring regulatory and technical consistency.

The Analysis & Design Deliverables and Release 3 documentation has proven highly suitable to the needs and constraints of nHDsC. As a result, they will play a vital role in guiding both the drafting and refinement of national business requirements, not only during the design phase but also during pilot implementation, which will serve to validate interoperability with the EU infrastructure and prepare for subsequent CP releases.

3 NHDSC OVERVIEW

3.1 Overview

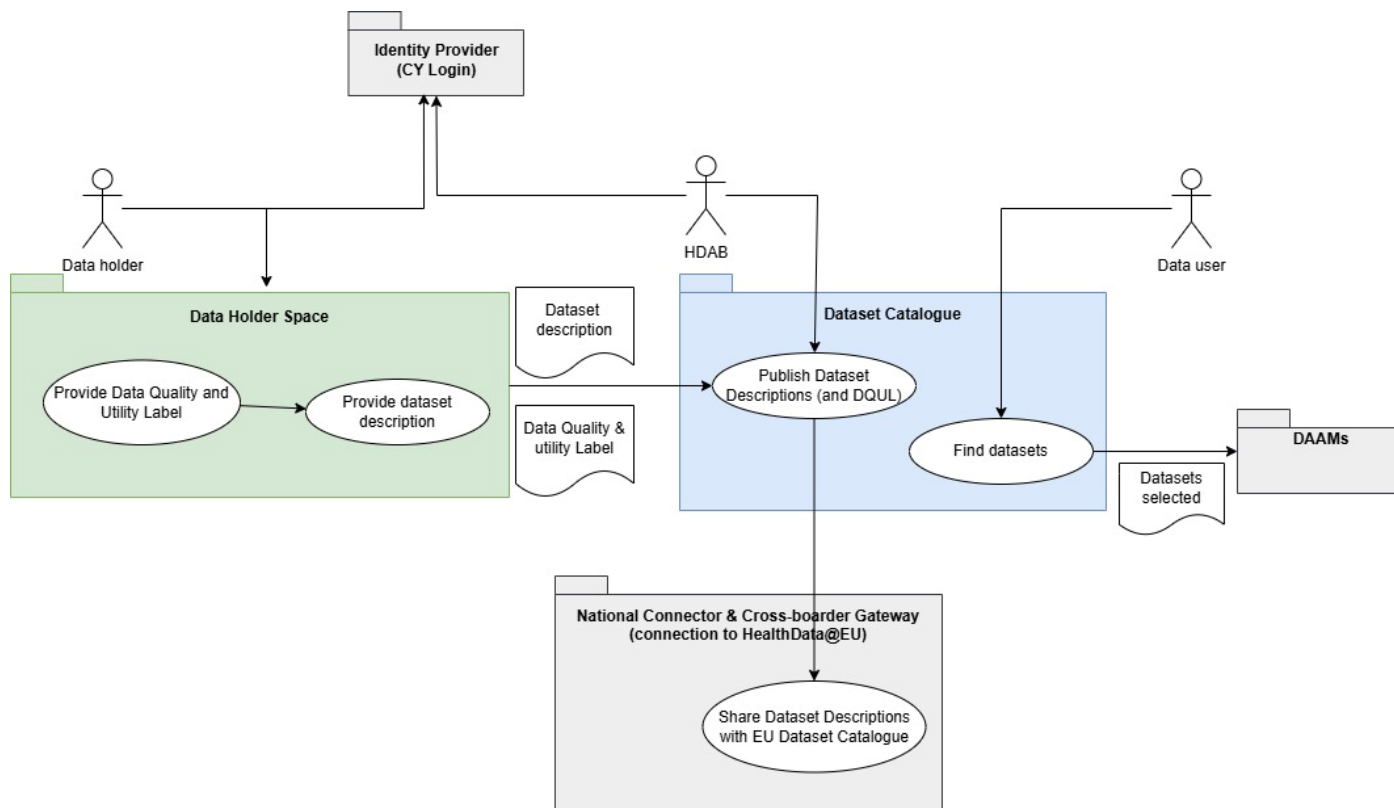


Figure 3: nHDsC overview (a revised version of DBC blueprint for nHDsC).

Gray shaded areas refer to external components.

Before a user can search for a dataset, the data holder space is a concept in which the description of the dataset is sent by the data holder to the HDAB (for review and publication). Additionally, some data holders must present a Data Quality and Utility Label (DQUL) (QUANTUM⁶ project, WP9). The data holder must update the dataset description on an annual basis. A data user can explore the national dataset catalogue to find and select datasets for a data access or request application (to be created and managed under the scope of the DAAM system, WP5) (Figure 3).

The nHDsC requirements specification focus on four core capabilities:

1. **Metadata input:** enabling data holders to submit dataset descriptions. It is optional for MS 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 Regulation 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 user-friendly search tools, facilitating discoverability by researchers and policymakers.

⁶ <https://quantumproject.eu/>

4. Metadata access: a user-centric interface will enable stakeholders to easily locate datasets through a national and EU-wide integrated platform.

nHDsC builds on top of the CY-EHDS-2nd CP leveraging the offered shared services, such as the user authentication, role-based permissions and accreditation status, and the user interface. nHDsC also interfaces with the NCP to synchronize relevant metadata records with the HealthData@EU CP.

3.2 Actors

The complete set of actors and their responsibilities across the CY-EHDS-2nd infrastructure is defined in *D5.1 – Annex D: Common Actor Definitions – CY-EHDS-2nd (v1.0)*. That annex serves as the single authoritative reference for roles appearing in WP 5 to 9, ensuring alignment with the EHDS Regulation and the implementation scope agreed by NeHA.

Table 1 below presents only the actors relevant to this module. These have been selected from the common list based on their participation in the workflows and responsibilities specific to this WP. In the functional requirements section, Actors Involved may refer to 'HDAB users which includes A7 and A8 actors.

Table 1: CY-EHDS-2nd nHDsC Actors as defined in *D5.1 Annex D Actor Definitions – CY-EHDS-2nd (v1.0)*.

Actor No.	Role
A1	General Public Citizen
A2	Data User
A3	Data Holder
A7	nHDsC Manager
A8	nHDsC Auditor

3.3 Business uses cases

3.3.1 Dataset creation

The nHDsC system will provide comprehensive capabilities for managing dataset descriptions in alignment with the Health DCAT-AP. Data holders will be able to create, view, edit, and remove dataset descriptions with ease, while also enriching them with DQULs. To ensure consistency and usability, the system will enforce a minimum set of required elements for each dataset entry. In addition to managing metadata, authorized HDAB users will have the ability to review and validate both the descriptions and their associated quality indicators before publication, helping to ensure clarity, completeness, and reliability across the dataset catalog. Published dataset descriptions will be sent to the HealthData@EU CP via the NCP for publication in the EU dataset catalogue (Figure 4).

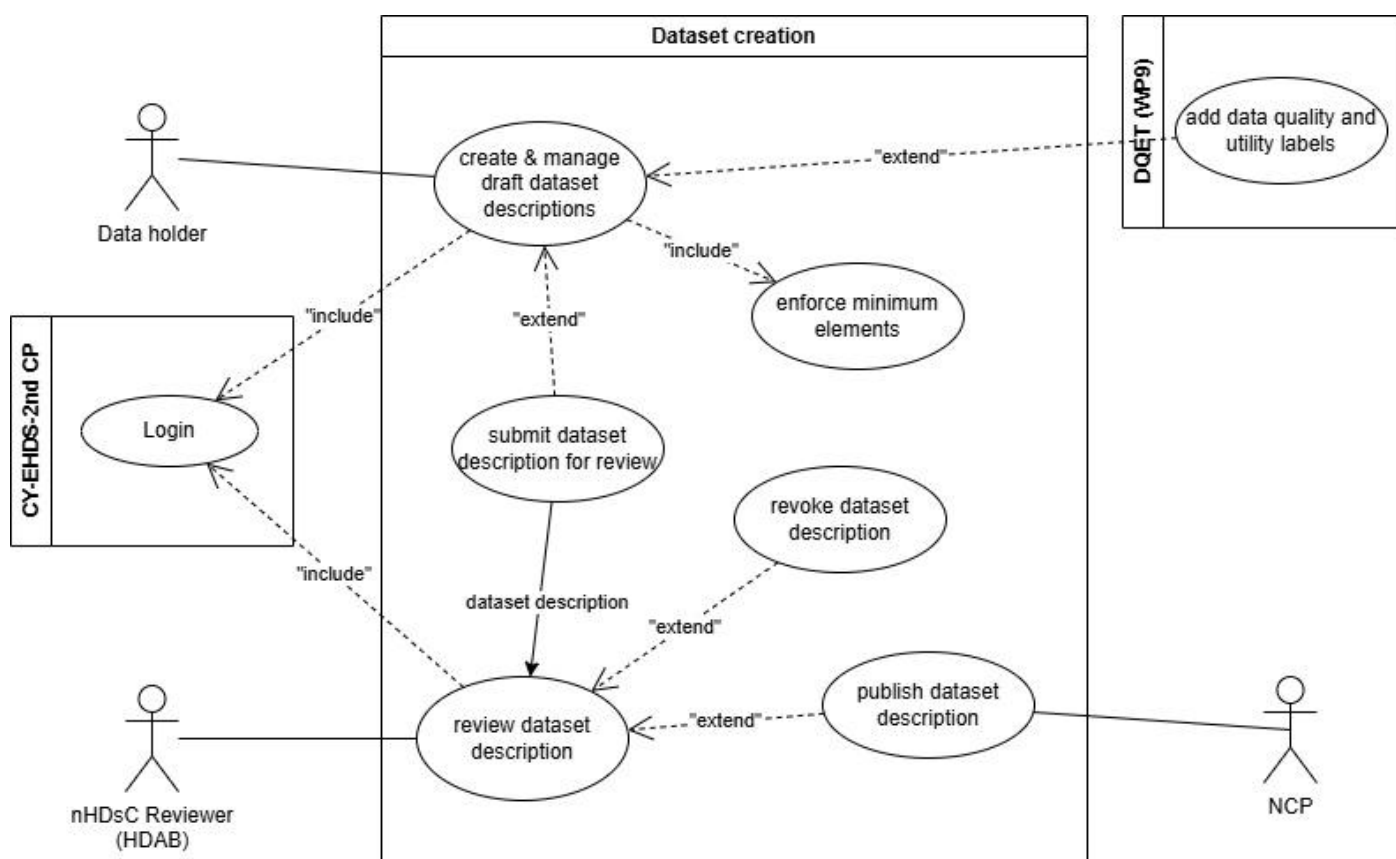


Figure 4: Dataset creation Use case

3.3.2 Dataset publication and update

The nHDsC system is built to support the full lifecycle of dataset descriptions, with particular emphasis on maintaining accurate and up-to-date metadata. In alignment with the Health DCAT-AP, users can publish and revise dataset descriptions as needed, with all changes tracked through a maintained history of updates. To ensure metadata remains current and relevant, the system enables automated scheduling of periodic reviews and audits, complemented by proactive notifications to data holders. Additionally, all actions related to dataset quality and metadata management are logged in detailed audit trails, reinforcing transparency and control throughout the update process (Figure 5).

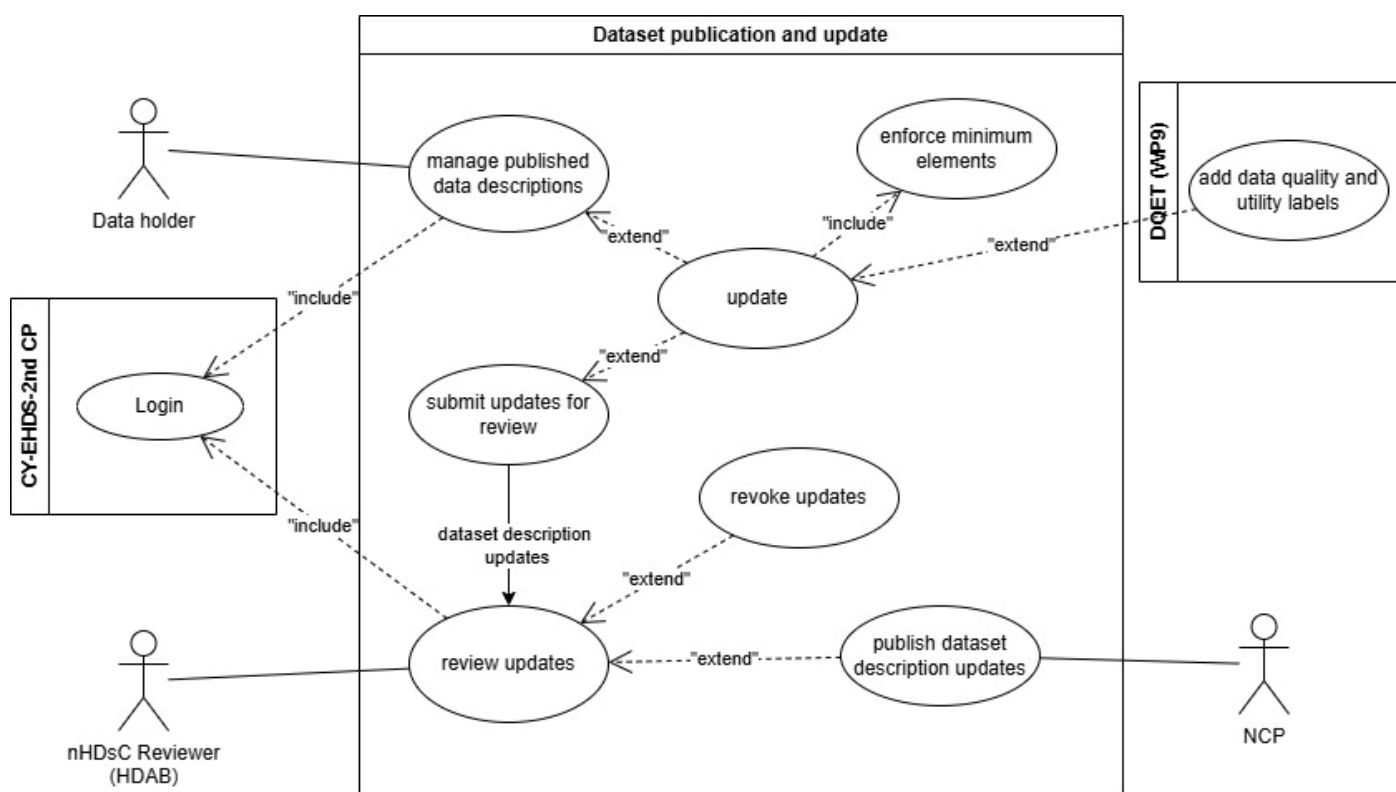


Figure 5: Dataset publication and update Use case

3.3.3 Searching for a dataset

The nHDsC system will provide an intuitive search interface for exploring and displaying dataset catalogue results, enhanced with advanced filtering and sorting options, including key elements from the Health DCAT-AP. The nHDsC search functionality will be integrated with the DAAMs CY-EHDS-2nd module to enable the selection of datasets for the creation of a data access or request application (Figure 6).

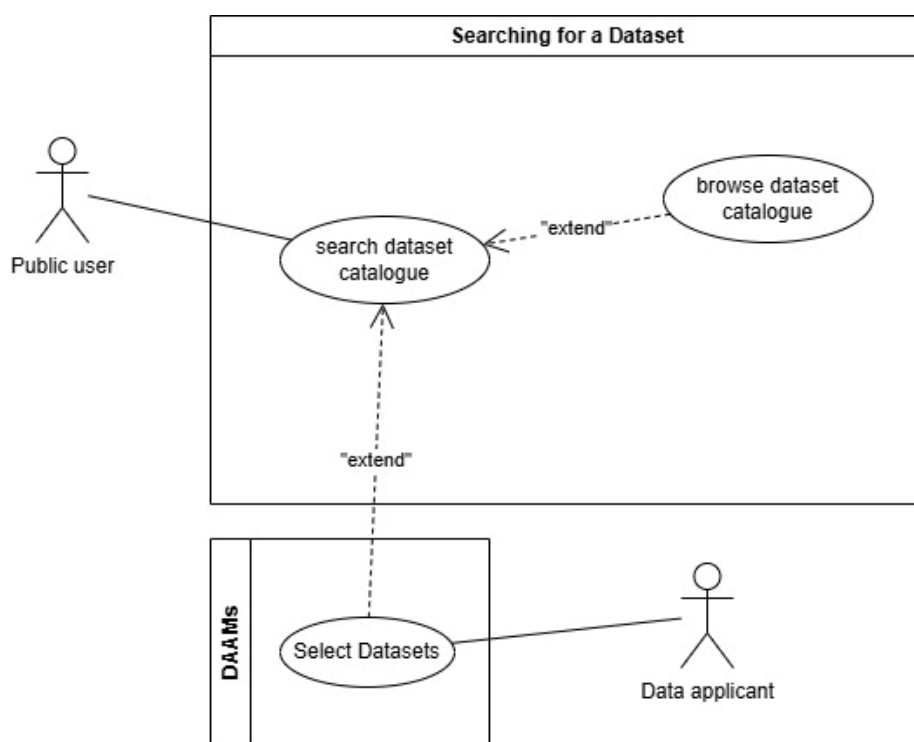


Figure 6: Searching for a dataset Use case

4 REQUIREMENTS

4.1 Functional requirements

We followed the domains suggested in the ANALYSIS & DESIGN of HealthData@EU Infrastructure deliverables to organise and present the requirements for the nHDsC. Identification, Authentication, and Authorization (IAA) and communication requirements are presented separately.

4.1.1 User Identification, authentication and authorization

Requirement code:	nHDsC_2.1
Requirement title:	The system must allow for holder's IAA, including the option to enforce multifactor authentication (MFA).
Actors involved:	Data holders.
Required features:	- Allow to Log in. - Incorporate some IAA mechanisms.
Detailed Description: The systems must include processes for IAA for data holders to provide electronic health data for secondary use (provide datasets).	
Functionality	Description
Mechanisms for identification, authentication and authorisation	- The system use identification and authentication - The system use Authorisation mechanisms to validate the user is related to the organisation on which the obligation to create the dataset is imposed. - o Establish a role-based access control (RBAC) system to assign permissions based on user roles. - o Implement organisation-specific access controls to ensure users are related to the relevant organisation. - o Maintain an up-to-date user directory that maps users to their respective organisations.
Data Holder Log in	- Data holders follow these processes to provide datasets descriptions for secondary use.

The required functionality will be inherited from the CY-EHDS-2nd CP (*D5.1 Annex A: CY-EHDS-2nd Architecture and CP Infrastructure*).

4.1.2 Dataset creation

Data holders must provide initial dataset descriptions using the HealthDCAT-AP through the Cypriot Health Data Access Body Portal (CY-HDAB Portal) and its nHDsC module. Dataset descriptions must include: Key metadata elements, including the dataset purpose, structure, format and access conditions, information on any legal or ethical constraints, including anonymisation or pseudonymisation levels.

HDABs are responsible for validating submitted dataset descriptions to ensure compliance with the EHDS Regulation. This validation involves assessing metadata completeness, data quality, and utility labels. The HDABs may request revisions or revoke inaccurate metadata to ensure the reliability of published datasets.

4.1.2.1 Create, read, update, and delete (CRUD) dataset descriptions using the Health DCAT-AP.

Requirement code:	nHDsC_3.1
Requirement title:	The system must allow to CRUD dataset descriptions using the Health DCAT-AP.
Actors involved:	Data holders.
Required features:	- Provide text edition functionalities to complete the datasets description. - Provide automated mechanisms to check quality and conformity of datasets descriptions (e.g. including validation tools to verify compliance with Health DCAT before submission).
Detailed Description: Health data holders must provide dataset descriptions to the HDAB, for categories outlined in Article 51 of the EHDS Regulation. To support compliance with making specific data categories available for secondary use, the system provides the following functionalities.	
Functionality	Description
Create New Dataset description:	A create a new dataset description option is clearly available on a landing page. For this, two options can be contemplated: - Create a copy from a submitted dataset description. - Create a new dataset description from scratch.
Guidance for Data Holders	- A wizard tool guides data holders through a user-friendly process to determine their obligations under the EHDS Regulation. - o The wizard includes steps and questions based on legal requirements in Article 51. - Based on responses, the wizard provides non-binding guidance on data categories that must be made available for secondary use.
Wizard Steps	- Step 1: Types of Datasets Held: Lists data categories and determines obligations based on responses. - Step 2: Guidance on Creating Dataset Description: Provides instructions on describing datasets (minimum elements and characteristics of dataset descriptions. These elements are expected to align with HealthDCAT to ensure consistency and interoperability across MS), compliance with standardised and machine-readable format, and data protection and security requirements. - Step 3: Finalising Guidance and Feedback: provides a validation tool to verify compliance with HealthDCAT and other standards and rules before submission (see requirement d) Enforce minimum elements for dataset descriptions)
Dataset Description Form	- Includes mandatory and optional fields. - Supports various answer types (controlled vocabularies, multiple choice, numerical, free text). - Features auto-save functionality.

	- Allows highlighting protected parts and categorising datasets as "Highly Sensitive," with justifications (for example to protect IP rights /trade secrets). - Supports file uploads with size and security checks.
Dashboard for Draft Descriptions	- Data holders can save descriptions as drafts and access a dashboard to manage and complete them later.
When the data holder has finalised with the dataset description, it can select a "Submit for review" option/feature that will trigger a process for review of the HDAB. Data holders can contact their respective HDAB for further guidance on their obligations.	

4.1.2.2 Add DQULs

Requirement code:	nHDsC_3.2
Requirement title:	The system must enable the addition of DQULs.
Actors involved:	Data holders.
Required features:	- Allow data holders to add DQULs to their dataset description. - Allow data holders to upload documentation supporting DQULs.
Detailed Description: An option to "Add DQULs" is available during the dataset description creation process, prior to submission for review. This option redirects users to a dedicated space for managing DQULs.	
Functionality	Description
DQULs	- When selected, users can add a label detailing the following information: ○ Data documentation: Information on metadata, format and standards, provenance, and data model. Supporting documentation can be added for each field. ○ Assessment of technical quality: Information on completeness, uniqueness, accuracy, validity, timeliness, and consistency. Supporting documentation can be added for each field. ○ Data quality management processes: Information on the maturity of data quality management processes, including review, audit processes, and bias examination. Supporting documentation can be added for each field. ○ Assessment of coverage: Information on time period, population coverage, representativity, and the average timeframe a person appears in a dataset. Supporting documentation can be added for each field. ○ Access and provision information: Information on the time between data collection and dataset addition, and the time required to provide ○ data after access application approval. Supporting documentation can be added for each field.

- Data modifications: Information on merging and adding data to existing datasets, including links with other datasets. Supporting documentation can be added for each field.

Note: The European Commission (EC) may establish the visual characteristics of the label in the future. The DQUL will be created on a separate dedicated page (either within HDAB or elsewhere), and data users will be able to import the DQUL information in a DCAT-AP compatible format.

HDABs should be able to display this information in the catalogue, alongside the dataset description, with an additional "DQUL" tab for dataset descriptions where one was submitted.

The DQUL functionality will be implemented by the DQET module of the CY-EHDS-2nd (D9.1 Health data quality enhancement toolbox: Requirements and Specifications) and be integrated with the nHDsC.

4.1.2.3 Enforce minimum elements for dataset descriptions

Requirement code:	nHDsC_3.3
Requirement title:	The system must enforce minimum elements for dataset descriptions (format).
Actors involved:	Data holders.
Required features:	Provide tools to ensure dataset descriptions comply with the mandatory minimum metadata elements outlined in the EHDS Regulation and align with HealthDCAT-AP metadata standard.
Detailed Description: Before submitting a dataset description, the system will include a format check procedure which will perform an assessment to ensure compliance with minimum common information elements and cross-border data standards.	
Functionality	Description
Automatic control checks	- Automated checks ensure metadata descriptions are valid and complete as well as HealthDCAT compliance.
Error handling for non-compliant descriptions	- If a description is non-compliant, an error message specifies validation errors. - This error message prevents the dataset description submission for review.

4.1.2.4 Review dataset descriptions and DQULs

Requirement code:	nHDsC_3.4
Requirement title:	The system must allow to review dataset descriptions and DQULs.
Actors involved:	HDAB users, Data holders.
Required features:	- Allow to review metadata completeness, data quality, and utility labels added by data holders. - Automatically notify the data holder in case of the revocation of the data quality and utility and utility label, including the justification.

Detailed Description:

Once the data holders have provided a DQUL, the system includes a clear process that allows HDABs to review dataset descriptions and data quality labels (and possibly revoke them if not applicable).

Functionality	Description
New dataset description notification	- nHDsC Manager receives a notification that a dataset description and DQUL has been created.
Review	- nHDsC Manager accesses a dashboard in the system with datasets descriptions for review, assigns nHDsC Auditor. - nHDsC Auditor open the datasets descriptions (metadata completeness), data quality, and utility labels. - The system displays the dataset description and DQULs.
Submission option	- If the nHDsC Auditor establishes that the dataset description is complete and accurate, the auditor selects a "Submit for publication" feature available.
Revocation option	- The system provides a revocation option/feature/functionality for users to revoke descriptions and DQULs.
Justification prompt	- Upon selecting the revocation option/feature/functionality, the system prompts a free text field for providing a justification. - The user must enter a justification for revocation.
Revocation process	- Once the justification is submitted, the system processes the revocation request. - The descriptions and DQULs is systematically marked as revoked, rendering it invalid. - The status of the dataset descriptions and DQULs in the database is updated to indicate it has been revoked.
Automatic notification	- After revocation, the system automatically notifies the data holder. - Notifications are sent through the NCP a secure message challenge. - The notification includes the justification for the revocation.

The reviewing of DQUL will be implemented by the DQET module of the CY-EHDS-2nd (D9.1 *Health data quality enhancement toolbox: Requirements and Specifications*) and be integrated with the nHDsC.

4.1.2.5 Submit dataset description

Requirement code:	nHDsC_3.5
Requirement title:	The system must allow to submit dataset description
Actors involved:	HDAB users, Data holders.
Required features:	- Provide a functionality to submit the dataset description. - Provide a dashboard for data holders with the dataset descriptions that data holders have created.
Detailed Description:	

Once the dataset description is finalized (and validated), the system allows the nHDsC Manager to confirm finalisation through a “Submit for publication” option/ functionality, triggering the following processes:

Functionality	Description
Publish confirmation.	- The system prompts a confirmation that the dataset description is approved. - The dataset description is then published in the national dataset catalogue (publicly available and standardized machine-readable dataset catalogues).
Publish notification.	- The system sends a notification of the published dataset description to the data holder.
Data holder dashboard.	- Each data holder has access to a dashboard displaying the status of their submitted dataset descriptions as “published”. - The dataset description is then removed from the “pending review” dashboard from the system.
Submission to the EU dataset catalogue.	- The system notifies NCP for the submitted dataset description, which in turn sends the description to the HealthData@EU CP for review and publication in the EU dataset catalogue.

4.1.3 Data publication and update

After validating dataset descriptions, the nHDsC publishes dataset descriptions using the HealthDCAT-AP through the CY-HDAB Portal (national dataset catalogues) and in the HealthData@EU CP (EU datasets catalogue). The catalogue must provide detailed metadata for each dataset, including: source of the data, scope and main characteristics of the dataset, nature of the electronic health data included, and conditions for making the dataset available. The catalogue should allow updates to dataset metadata as necessary to ensure accuracy and relevance, with any changes validated by the HDAB prior to publication. To enhance data governance, the system supports automated scheduling of periodic reviews and audits, sending timely notifications to stakeholders to ensure compliance with relevant guidelines and regulations. The catalogue should be made available to the single information points established under Regulation (EU) 2022/868.

Additionally, the CY-HDAB Portal maintains detailed audit trails for all data quality management activities, providing a robust framework for monitoring actions taken on datasets.

4.1.3.1 Publish dataset descriptions using the Health DCAT-AP

Requirement code:	nHDsC_4.1
Requirement title:	The system must allow for the publication of dataset descriptions using the Health DCAT-AP.
Actors involved:	nHDsC Manager, Data holders.
Required features:	- Publish the contents of the applicable dataset descriptions in the national dataset catalogue. - Provide an automated mechanism to associate a code with each dataset for identification and information purposes.

Detailed Description:

After validation, the dataset descriptions are published in the datasets catalogues. These provide information about existing datasets to data applicants.

Functionality	Description
Publish	- The system publishes dataset descriptions after validation of the content and labels in the national dataset catalogue (publicly available and standardised machine-readable dataset catalogue).
Automatic datasets identification processes	- The system generates a unique version numbering for each new dataset description, following a code system that may include references to the country, HDAB, data holder, and a sequential number for published datasets.
Public availability	- The national datasets catalogue is made publicly available through an interface in the nHDsC module of the CY-HDAB Portal (Note: datasets with trade secrets might not be displayed for public view).
Information displayed	- The catalogues display detailed metadata for each dataset, including: source of the data, scope and main characteristics of the dataset, nature of the electronic health data included, and conditions for making the dataset available.

4.1.3.2 Update dataset metadata descriptions and maintain change history

Requirement code:	nHDsC_4.2
Requirement title:	The system must allow for updating dataset metadata descriptions and maintain change history.
Actors involved:	HDAB users, Data holders.
Required features:	- Provide an interface to access their dataset descriptions. - Provide text-edition functionalities to update “published” datasets. - Provide automated mechanism to exchange the new view of the dataset description with the NCP.

Detailed Description:

If a dataset description requires updating or editing, data holders can easily update it through the system.

Functionality	Description
Dashboard overview	- A dashboard displays all published datasets descriptions managed by the data holder, including essential information such as dataset titles, publication dates, current status, and version history. - The dashboard includes search and filter tools to help data holders quickly locate specific datasets.
Editing and updating	- After identifying the dataset, the data holder selects to use an Edit/Update function to modify metadata, content, or supporting documentation.

	- This triggers the process described in requirements nHDsC_3.3, nHDsC_3.4, nHDsC_3.5, nHDsC_4.1 of this domain.
Versioning process	- A new version of the dataset is automatically generated, leaving the previous version intact. - o Each version is assigned a unique identifier to distinguish it from earlier versions, and a prompt is generated for the data holder to provide a comment to explain the change in version. - o The portal logs metadata related to the new version, including details of the data holder who made the updates, the date and time of changes, and a summary of modifications. - o In case of mistake, data holder or HDAB are able access a previous version
Data exchange and synchronisation	- The system exchanges updated dataset descriptions with the NCP, which in turn notifies the HealthData@EU CP.

4.1.3.3 Automate scheduling of periodic reviews and audits

Requirement code:	nHDsC_4.3
Requirement title:	The system must support automated scheduling of periodic reviews and audits for data holders.
Actors involved:	HDAB users, Data holders.
Required features:	- Display the status of the datasets in their dataset catalogue in the dashboard for data holders. - Display a time-tracking functionality with a text indicating the days of the next periodic review and audit (annually since the last audit).
Detailed Description: The system clearly displays the status of each dataset in their catalogues, indicating statuses such as “Published” and “Removed”.	
Functionality	Description
Status Display	- The system displays the current standing of each dataset, such as “Published” and “Removed”.
Time-Tracking Mechanism	- Displays the date of the next review or audit of the dataset by the data holder. - Provides a text-based indicator showing the day for the next review/audit, ensuring data holders are aware of compliance timelines (at least once a year).

4.1.3.4 Notify stakeholders about upcoming reviews

Requirement code:	nHDsC_4.4
Requirement title:	The system must notify data holders about upcoming reviews.
Actors involved:	HDAB users, Data holders.

Required features:	- Send automated notifications to the data holders about upcoming reviews, audit results and required actions.
Detailed Description:	The system automatically notifies data holders of any scheduled reviews or periodic audits.
Functionality	Description
Automatic notifications	The system automatically sends notifications to data holders about scheduled reviews or periodic audits.
Notification details	Notifications include the review/audit date. Specific business rules, such as the number of days before the deadline to send the update, are set up withing the system.

4.1.3.5 Maintain detailed audit trails for all data quality management activities

Requirement code:	nHDsC_4.5
Requirement title:	The system must maintain detailed audit trails for all dataset quality management activities.
Actors involved:	HDAB users
Required features:	- Maintain an audit trail for the whole lifecycle of dataset submitted.
Detailed Description:	The system captures and logs metadata related to.
Functionality	Description
Dataset submitted	Captures and logs dataset ID, dataset name, submission date, data holder's name, and associated DQULs.
Datasets validated	Captures and logs automated validation processes (including timestamp) and results of the validation (e.g. validated, not-validated, corrected, etc).
Datasets published	Captures and logs dataset ID, dataset name, publication date, data holder's name, and the final DQULs applied.
Datasets updated	Captures and logs new version ID, name of the data holder who made the updates, date and time of the changes, and a summary of modifications (e.g. "new data added" or "metadata updated").

4.1.4 Searching for a dataset

CY-HDAB Portals offers a comprehensive search functionality that enables data applicants to efficiently browse and display entries from national datasets catalogues.

Data applicants can perform searches based on various criteria such as dataset source, scope, characteristics, and access conditions. The systems include robust filtering and sorting capabilities based on DCAT and Health DCAT elements, allowing them to refine search results effectively. Data applicants can save their search queries for easy access, receive notifications about updates on their saved datasets, and share these searches with others to enhance collaboration. Additionally, the systems allow data applicants to select and

compare multiple datasets side by side for informed decision-making and provides access to both human-readable and machine-readable versions of the dataset catalogue.

4.1.4.1 Offer a search functionality for browsing and displaying dataset catalogue results

Requirement code:	nHDsC_5.1
Requirement title:	The system must offer a search functionality for browsing and displaying dataset catalogue results.
Actors involved:	Data applicants, public users.
Required features:	- Provide a search function which enables browsing through the national dataset catalogues by data applicants. - Provide automated mechanisms to correct and adjust elements of the search where necessary. - Provide automated mechanisms to detect and alert if not results are found from the search.
Detailed Description: Data applicants must be able to discover existing datasets through a search engine available on all systems.	
Functionality	Description
Accessing the system	Open an internet browser to access the national dataset catalogue
Browsing the catalogue	Browse the entire metadata catalogue, displaying all entries
Search functionality	- Search using an open text field - o Propose commonly used search queries - Create personalised search functions based on predefined filters
Dataset selection	- Display query results on the applicant interface - Select a dataset and display its description
Once the data applicant launches the search, the system functionalities include:	
Matching and results	- Match dataset descriptions to query results - Perform corrections, adjustments, and cancellations of the search
Alerts and information	- Provide alerts if the search yields no results - Display information on the dataset, including metrics measuring the status and use of the dataset

4.1.4.2 Provide filtering and sorting capabilities by DCAT and Health DCAT elements

Requirement code:	nHDsC_5.4
Requirement title:	The system must provide filtering and sorting capabilities (including Health DCAT AP elements).
Actors involved:	Data applicant.
Required features:	- Provide filtering and sorting capabilities alongside the search function.
Detailed Description:	

To foster the discoverability of electronic health data, the systems allow three types of filtering options for search purposes; basic filters based on dataset's characteristics, and queries using Health DCAT-AP elements. Data applicants can combine these filtering options to identify relevant datasets that meet their needs.

Functionality	Description
Basic filtering	- Provides filters based on dataset's characteristics, allowing selection according to criteria such as data holder, keyword, theme, scope, time, geographical coverage, provenance, access, format, DQUL, and AI-suitable data.
Queries using Health DCAT-AP elements:	- Features a applicant interface with a dialogue box that guides applicants through key questions to define the Health DCAT-AP metadata elements matching the applicants search.
Sorting options:	- Alphabetical Order: Sort datasets by name. - Date Submitted: Sort datasets by submission date. - Date Updated: Sort datasets by the most recent updates. - Dataset size: Sort datasets according to their size. - Dataset origin: Sort datasets according to their origin (i.e. Member State, Authorised Participant, etc.).

4.1.5 Interoperability of health data (HealthDCAT-AP)

To ensure semantic and technical interoperability in line with the EHDS Regulation objectives, nHDsC will adopt the HealthDCAT-AP⁷ as a standard for dataset metadata description. Developed under the HealthData@EU Pilot project⁸, HealthDCAT-AP extends the existing DCAT-AP metadata framework to address the specific needs of health data catalogues. Its adoption facilitates harmonized metadata across national and EU-level health data catalogues, improving the findability, comparability and integration of datasets used for secondary purposes such as research, policymaking and innovation.

In alignment with the EHDS Regulation, which underscores the integration of health data with other domains such as environmental and social data, nHDsC must enable unified and consistent data description practices. HealthDCAT-AP serves as a critical enabler for this integration by supporting a common vocabulary and structure for dataset metadata. This standardization minimizes fragmentation, simplifies dataset discovery, and strengthens interoperability between national HDABs and the broader HealthData@EU infrastructure.

Furthermore, nHDsC will ensure compliance with the FAIR data principles by leveraging the HealthDCAT-AP standard. Structured metadata based on HealthDCAT-AP will support the discoverability of datasets, provide clear access protocols, enable system-to-system compatibility, and document provenance and quality for downstream reuse. By embedding HealthDCAT-AP into the nHDsC's metadata management processes, the platform will promote responsible and effective data sharing across European and cross-sectoral contexts.

⁷ <https://www.healthinformationportal.eu/healthdcat-ap>

⁸ www.ehds2pilot.eu

4.1.6 Communication

4.1.6.1 Integrated communication

Integrated communication is a vital part of the nHDsC and the CY-EHDS-2nd CP. It refers to a systematic approach that facilitates seamless interaction among various stakeholders involved in the system. Here are the key components and requirements of integrated communication in the context of nHDsC (Table 2).

Table 2: Integrated communication use cases

Use case	Communication requirement
New dataset description submitted for review.	The system automatically notifies the nHDsC auditor.
After a dataset description revocation.	The system automatically notifies the data holder.
Dataset description published in the national dataset catalogue.	The system automatically notifies the data holder.
A dataset description was rejected by the HealthData@EU.	The system automatically notifies the nHDsC manager for resolution.
A new version of a dataset description submitted for review.	The system automatically notifies the nHDsC auditor.
A new version of a dataset description published in the national catalogue.	The system automatically notifies the data holder.
Automatic notifications about upcoming reviews.	The system automatically notifies the data holder of any scheduled reviews or periodic audits.

All required communication between the data holders and the HDAB can be facilitated via secure email channels or structured communication channels.

4.1.6.2 Cross-border communication

The communication with the HealthData@EU CP is a responsibility of the National Connector and Cross-border Gateway module of the CY-EHDS-2nd CP (D8.1 National Connector for Cross-border gateway Requirements and Specifications). nHDsC integrates with the National Dispatcher component for the synchronization of the dataset descriptions with HealthData@EU CP (Table 3).

Table 3: Cross-border/National Connector communication use cases

Use case	Communication requirement
Dataset description published in the national dataset catalogue	Authorized users can send a notification to the NCP of the published dataset description.
EU dataset catalogue rejects a national dataset description	NPC receives a notification from the HealthData@EU CP
Dataset description updated in the national dataset catalogue.	Authorized users can send a notification to the NCP about the updated dataset description.

5 SPECIFICATIONS

5.1 nHDsC Processes

The section includes visual representations of the main nHDsC processes, for creating and publishing a dataset description, updating and publishing an existing one. The diagrams show how each process works, the actors involved and the overall flow.

5.1.1 Dataset creation

The dataset creation process within the nHDsC system begins with data holders drafting a dataset description. If a DQUL is required, the user is redirected to the ‘add DQUL’ process of the DQET module of the CY-EHDS-2nd CP. As soon as the label is added the user is redirected back to the nHDsC dashboard area. A data holder can add a DQUL to a dataset description any time before its submission for review.

A data description is reviewed by the nHDsC Auditor and upon approval it can be published by the nHDsC Manager. If a label was added it is submitted for review in parallel with the dataset description. Both the description and the label undergo separate review processes—by the nHDsC Auditor and the DQET Reviewer, respectively. Once approved, the description and the label are published. However, if a dataset description is not approved the system proceeds with the revocation process and the label cannot be published. The HealthData@EU CP is notified upon publication (Figure 7).

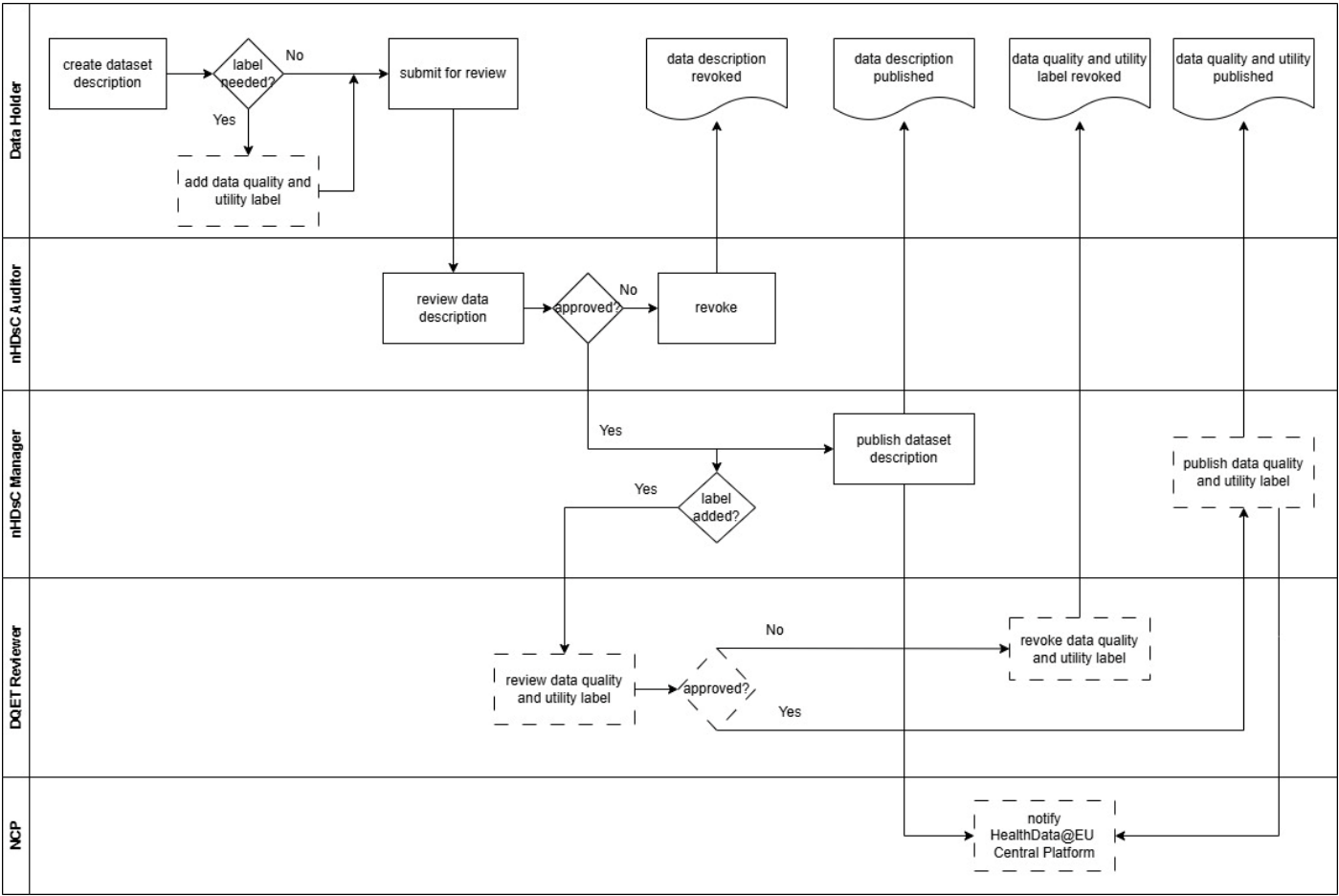


Figure 7: Dataset creation process.

5.1.2 Dataset description updates

The dataset description update process begins with data holders revising an existing dataset description, resulting in a new version of the description. The review/ publishing follows the same steps as in the 'dataset creation process' (Figure 8).

A new version of a data description is reviewed by the nHDsC Auditor and upon approval it can be published by the nHDsC Manager. Once approved, the new version of the dataset description is published. The HealthData@EU CP is notified upon publication. The new version of the dataset description is revoked if not approved.

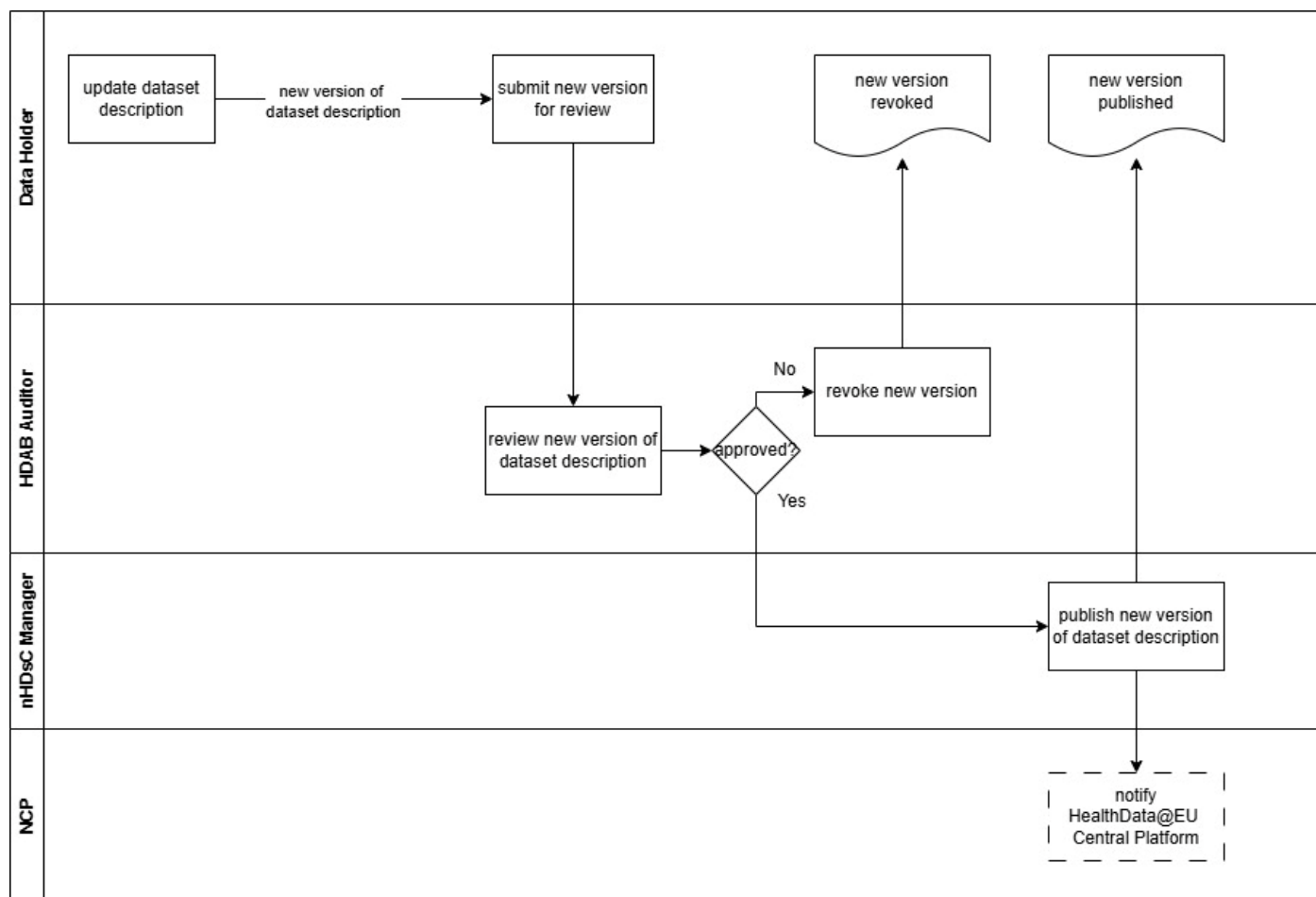


Figure 8: Dataset description updates process flow.

5.1.3 nHDsC & HealthData@EU

Figure 8 shows how health data is securely shared across borders in the HealthData@EU system. It shows how national systems first prepare health data records, which are sent to the HDAB for checks and ensuring trust and standardised formatting for cross-border sharing. Another capability allows searching and viewing dataset information. The data is then sent to a NCP. It travels securely through a gateway and reaches the EU platform for centralised access and management (Figure 9).

As soon as new dataset descriptions or updates of existing ones are published in the national dataset catalogue, the nHDsC Manager can initiate the synchronization process with the HealthData@EU CP through the NCP (D8.1 - National Connector for Cross-border gateway Requirements and Specifications).

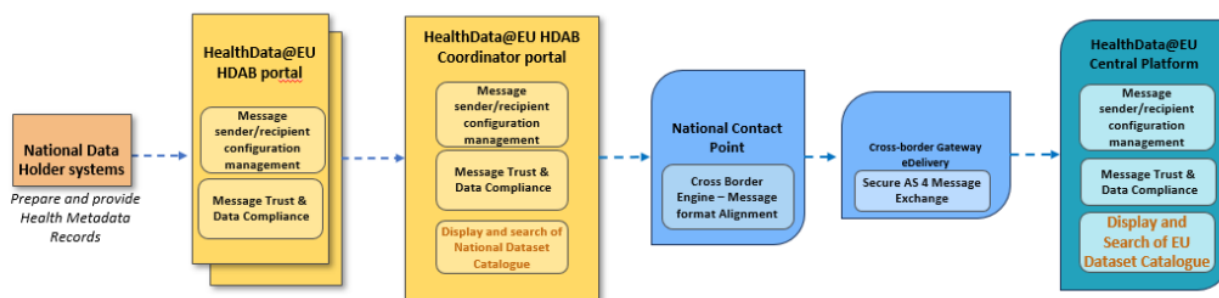


Figure 9: Schematic overview of the information flow described by DG SANTE⁹.

5.2 nHDsC module of CY-EHDS-2nd CP

The nHDsC system is one of the core functional modules of the CY-EHDS-2nd CP. The system will inherit/use core functions of the CY-EHDS-2nd CP infrastructure such as the national identity gateway (CY Login) for IAA mechanisms, and the logging, monitoring, and audit trail capabilities.

The CY-EHDS-2nd CP follows a modular microservices architecture where all components are designed as independent services. The Dataset Catalogue module encompasses three primary processes: dataset creation, publication in national dataset catalogues, and dataset discovery.

This section details the core features for the main nHDsC processes and considers a number of microservices for their implementation. These considerations follow the suggestions in the ANALYSIS & DESIGN of HealthData@EU Infrastructure.

5.2.1 Core features

Table 4: nHDsC module core features

Feature/ service	Description
User IAA	Confirms user's identity and ensures that the user has the right permissions to access and interact with the system. - Verifies user identity and permissions before allowing access to the portal and data access features. - Ensures secure data holder identification and authorisation to create dataset catalogues and publish them. - Ensures secure user authentication to interact with the portal and initiate datasets discovery. - Validates a natural person's identity. Authentication verifies the identity of the user desiring to request access to datasets.
Create dataset description	Allows health data holders to fill in DQLs for their datasets.
Update dataset description	Allows health data holders to update the detailed descriptions of their datasets, at least annually.
Display dataset description	Allows reading dataset descriptions in the national datasets catalogue of the Member State.

⁹ D03.01 Architecture Artefacts High Level Version

Search dataset catalogue	Includes the ability to find and retrieve metadata catalogues and particularly metadata records at the national level.
Assessment of dataset descriptions	Receives from health data holders the description of the dataset, allowing for review of their accuracy.
Provide decision on dataset descriptions	Allows approval or revocation of the dataset description.

5.2.2 Proposed services

Following the CY-EHDS-2nd CP architecture that employs a microservice design pattern, we can consider the following services:

- Dataset description: to provide data holders features for creating and reading dataset descriptors.
- Data validation and approval: to temporarily persist metadata entries for validation before publication in the National dataset catalogue. It prepares, validates and processes datasets, ensuring these datasets meet quality and compliance standards before publishing them for discoverability.
- Dataset publication and synchronization: to publish validated datasets to the national dataset catalogue and synchronizes updates with EU datasets catalogue via the NCP.
- Data standardization: to manage vocabulary, ontologies, ensuring that metadata descriptors are consistent.
- Search and visualization microservice: to discover and select in a user interface catalogues, by searching by specified keywords or advanced criteria fields.

5.2.3 Integrations

Table 5: CY-EHDS-2nd CP Modules Integration Overview

CY-EHDS-2nd CP service/ module	Integration details
Identity and Access Management service	The service authenticates and authorises the users of the system, enabling them to access the dataset descriptions management functions provided by nHDsC. More information can be found in <i>D5.1 Annex A: CY-EHDS-2nd Architecture and CP Infrastructure</i> .
DQET module	The DQET component provides DQULs on metadata records registered in nHDsC. These labels are associated with the respective records and displayed within nHDsC interface. More information can be found in <i>D9.1 Health data quality enhancement toolbox: Requirements and Specifications</i> .
DAAMs module	A user navigates the national dataset catalogue to select dataset descriptions to create a data access or request applications. The DAAMs module responsible for these applications receives the relevant metadata record ID from the nHDsC. This ensures precise dataset referencing and linkage between the data access/requests granted and the registered data assets. More information can be found in <i>D5.1 DAAMs Requirements, Specifications and Prototype</i> .

National Connector for Cross-border Gateway	For cross-border dataset metadata records exchanges, nHDsC provides the required structured data to the NCP. More information can be found in <i>D8.1 National Connector for Cross-border gateway Requirements and Specifications</i> .
---	---

5.3 Health DCAT-AP

HealthData EU Central Services will share with the MS the integration of the HealthDCAT-AP editor and validator into the HealthData@EU CP. The nHDsC's approach is to align fully with the HealthData@EU infrastructure and the CP model developed by the EC. The technical specifications and implementation details of these integrations will be provided in the upcoming deliverables.

5.3.1 Health DCAT-AP Editor

The HealthDCAT-AP Editor is an intuitive, web-based tool designed to help data holders create metadata records that comply with the HealthDCAT-AP specification, supporting the EHDS Regulation initiative. It simplifies the process of generating FAIR (Findable, Accessible, Interoperable, Reusable) metadata for health datasets, even for users without technical expertise. The editor automatically formats entries into RDF, applies correct language tags, and validates records against EHDS Regulation standards. Users can select dataset access rights—Open, Protected, or Sensitive—which dynamically adjusts required metadata fields to align with relevant regulations such as the DGA and the EHDS Regulation. The platform supports multilingual input, session backups, and publishing to a public sandbox catalogue (credentials required). By streamlining metadata creation, the HealthDCAT-AP Editor enhances dataset discoverability and interoperability across the EU's health data infrastructure (Figure 10).

HealthDCAT-AP editor v3.9

Customise Your Editor

Dataset Access Rights ?

Open Data

Protected Data

Sensitive Data

Recover Your Session ?

Load Backup

Track Your Progress

Mandatory

Recommended

Optional

0/5

0/8

0/39

Validate

★

★

★

★

★

★

Publish

Dataset Discovery

Contacts

Documentation

Categorisation

Data Access

Technical Metadata

Metadata Quality

Dataset Discovery

Mandatory

Recommended

Optional

0/2

0/5

0/8

A set of metadata properties that describe essential attributes of the dataset, facilitating its intelligibility, relevance, and usability for a variety of purposes. These properties provide insights into the content, scope, context, etc.

Mandatory

Title DCAT

Definition: A name given to the dataset.

Usage: This property can be repeated for parallel language versions of the name.

Example: Linking of registers for COVID-19 vaccine surveillance

English

Enter a title

+

M

Title

Acronym

Keywords

Description

Provenance

Purpose

Release/Modified Dates

Population Coverage

Geographical Coverage

Temporal Coverage

Language

Retention Period

Update Frequency

Spatial Resolution

Temporal Resolution

Figure 10: HealthDCAT-AP editor¹⁰.

5.3.2 Health DCAT-AP Validator

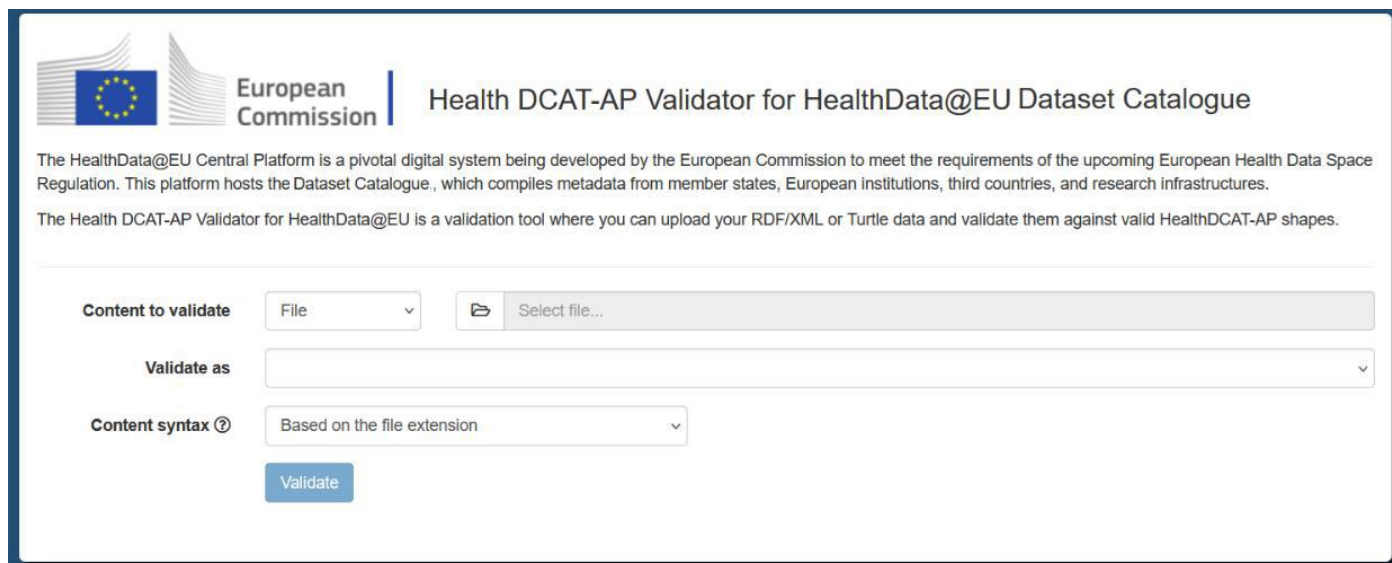
The HealthDCAT-AP Validator for HealthData@EU is an online tool developed by the EC to ensure that RDF metadata records comply with the HealthDCAT-AP specification, a key component of the EHDS Regulation initiative. Accessible via a user-friendly web interface, the validator allows users to upload RDF data in formats such as RDF/XML or Turtle or provide data via URI or direct input. It validates the content against the latest HealthDCAT-AP SHACL shapes, ensuring adherence to EHDS Regulation metadata standards. This service is powered by the Interoperability Test Bed, a conformance testing platform provided by the EC's

¹⁰ <https://ehds.healthdataportal.eu/editor2/>

Co-funded by the European Union

33

Directorate-General for Informatics (DG DIGIT), supporting the delivery of cross-border public services (Figure 11).



The screenshot shows the 'Health DCAT-AP Validator for HealthData@EU Dataset Catalogue' interface. It features the European Commission logo and a title bar. Below the title, there is a descriptive paragraph about the HealthData@EU Central Platform and the validator. The main form area includes three dropdown menus: 'Content to validate' (set to 'File'), 'Validate as' (empty), and 'Content syntax' (set to 'Based on the file extension'). A 'Select file...' button is next to the first dropdown, and a 'Validate' button is at the bottom.

Figure 11: Health DCAT-AP Validator¹¹ (source: HealthData@EU CP Open-source R3, deliverables).

6 REFERENCES

- 1 **European Commission, Directorate-General for Health and Food Safety (2025).** *HealthData@EU Central Platform – Open-source Release 3: Objective, Business Requirements and Use Cases*. Luxembourg: Publications Office of the European Union, March 2025. ISBN: 978-92-68-26028-9. DOI: 10.2875/5437350.
- 2 **European Commission, Directorate-General for Health and Food Safety (2025).** *HealthData@EU Central Platform – Open-source Release 3: Architecture Model and Technical Specification*. Luxembourg: Publications Office of the European Union, March 2025. ISBN: 978-92-68-26029-6. DOI: 10.2875/6472234.
- 3 **PwC EU Services (for DG SANTE) (2024).** *Requirements Catalogue for Scale-Up Version (Deliverable D02.03)*. Specific Contract No. 022 under Framework Contract SLG.AVT.DI07926 – BEACON, Lot 2. European Commission, December 2024.
- 4 **PwC EU Services (for DG SANTE) (2024).** *Architecture Artefacts – Scale-Up Version (Deliverable D03.03, Part 1: Architecture)*. Specific Contract No. 022 under Framework Contract SLG.AVT.DI07926 – BEACON, Lot 2. European Commission, December 2024.
- 5 **PwC EU Services (for DG SANTE) (2024).** *System Specifications – Scale-Up Version (Deliverable D03.03, Part 2: System Specifications)*. Specific Contract No. 022 under Framework Contract SLG.AVT.DI07926 – BEACON, Lot 2. European Commission, December 2024.
- 6 **PwC EU Services (for DG SANTE) (2024).** *Testing Framework – Scale-Up Version (Deliverable D03.03, Part 3: Testing Framework)*. Specific Contract No. 022 under Framework Contract SLG.AVT.DI07926 – BEACON, Lot 2. European Commission, December 2024.
- 7 **PwC EU Services (for DG SANTE) (2024).** *ICT Governance Framework and Procedures Catalogue – Scale-Up Version (Deliverable D04.03)*. Specific Contract No. 022 under Framework Contract SLG.AVT.DI07926 – BEACON, Lot 2. European Commission, November 2024.

¹¹ <https://health-data-itb-rdf-validator.acceptance.data.health.europa.eu/shacl/ehds/upload>

- 8 **European Parliament and Council of the European Union (2025).** *Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. Official Journal of the European Union, L 2025/327, 5 March 2025.* [ELI: http://data.europa.eu/eli/reg/2025/327/oj](http://data.europa.eu/eli/reg/2025/327/oj)
- 9 **TEHDAS2 (2025).** *D5.3 Draft technical specification on the national metadata catalogue.* TEHDAS2 Project, May 2025.
- 10 **European Commission (2025).** *Digital Business Capabilities Blueprint.* EHDS2 Community of Practice Confluence Page, January 2025.