



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

Proposal number: 101129405

D9.1 Health data quality enhancement toolbox: 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
Maria	Papaioannou	Cyprus	University of Cyprus (UCY)	Affiliated Entity (AE)
Nicoletta	Prentza	Cyprus	UCY	AE
Ioannis	Rodosthenous	Cyprus	UCY	AE
Waqar Aziz	Sulaiman	Cyprus	UCY	AE
Constantinos	Pattichis	Cyprus	UCY	AE

Reviewed by

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

Abbreviations

AE	Affiliated Entity
CA	Competent Authority
CP	Central Platform
CPL	Central Platform Layer
CY-EHDS-2nd	European Health Data Space for Secondary Use @CY
DAAMs	Data Access Applications Management solution
DBC	Digital Business Capabilities
DQET	Dataset Quality Enhancement Toolbox
DQUL	Dataset Quality and Utility Label
DQV	Data Quality Vocabulary
EC	European Commission
EHDS	European Health Data Space
EU	European Union
FAIR	Findability Accessibility Interoperability and Reuse of data
HDAB	Health Data Access Body
HealthDCAT-AP	Data Catalog Vocabulary Application Profile for health
ID	Identification
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
NeHA	National eHealth Authority of Cyprus
nHDsC	National Dataset Catalogue for health data
SHACL	Shapes Constraint Language
SPE	Secure processing environment
TEHDAS2	Second Joint Action Towards the European Health Data Space
UCY	University of Cyprus
WP	Work Package

Executive Summary

In a time when Europe is embracing the transformative promise of health data, the value of that data is measured not only by its volume, but by its meaning. A dataset without clarity, structure or trust can be a missed opportunity instead of a valuable resource. The European Health Data Space for Secondary Use @CY (CY-EHDS-2nd)¹ project represents Cyprus commitment to ensuring that health data becomes a usable, meaningful and responsible asset for the benefit (mainly) of every citizen.

This deliverable defines the Dataset Quality Enhancement Toolbox (DQET) which is a foundational module in the national infrastructure for secondary use of health data. Developed under Work Package (WP) 9, the DQET enables the creation, review and publication of structured quality and utility labels for health datasets. These labels allow dataset providers to communicate clearly and transparently the characteristics of their data: its completeness, structure, governance, accessibility, and applicability to real-world challenges.

The importance of this module is amplified by Cyprus starting point: a European Member State with limited experience in structured health data reuse. The provision of the DQET module in the CY-EHDS-2nd framework reflects the belief that if we are to build a trustworthy ecosystem for secondary health data use, we must pay particular attention in safeguarding and promoting data quality. It is a mechanism to shift culture, enabling institutions to think not only about what data they hold but how that data might serve citizens, researchers, innovators, and future generations.

Structured in alignment with Article 78 of Regulation European Union (EU) 2025/327² (European Health Data Space (EHDS) Regulation) and based on the model developed in the QUANTUM³ project, the DQET module defines a national implementation that is semantically interoperable, legally compliant and ready for integration with the HealthData@EU⁴ infrastructure.

This deliverable sets out:

- Its role within the CY-EHDS-2nd architecture
- The regulatory landscape
- The alignment of DQET with QUANTUM project
- The architecture and structure of the label
- The functional requirements of the DQET module

By establishing the DQET module at this early stage of national infrastructure development, we are laying the foundation for what is envisioned to become a compass for all stakeholders in the CY-EHDS-2nd ecosystem. More than a technical component, DQET is intended to guide and align practices around quality in every dimension, from dataset creation and collection to publication, validation and ultimately, reuse. As the ecosystem matures, DQET will help embed a shared culture of responsibility and transparency, ensuring that data is not only available, but valuable, and that the insights derived from secondary use meaningfully serve citizens' health and well-being.

¹ <https://www.cy-ehds-2nd.eu/>

² <https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng>

³ <https://quantumproject.eu>

⁴ <https://acceptance.data.health.europa.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	D9.1 Introduction	9
1.3	Objective of the Deliverable	11
1.4	Connection with other Work Packages.....	11
1.5	Interdependencies within CY-EHDS-2nd infrastructure	12
2	LEGAL FRAMEWORK	12
2.1	National Legal Framework.....	13
3	ALIGNMENT WITH QUANTUM PROJECT.....	13
3.1	Use of the QUANTUM Label Structure.....	14
3.2	Integration of QUANTUM Data Holder Workflow	14
3.3	Alignment with Identified Tools and Validation Mechanisms.....	14
3.4	Support for FAIR, DQV and HealthDCAT-AP Vocabularies	15
3.5	Future Alignment with QUANTUM and HealthData@EU	15
4	DATASET QUALITY ENHANCEMENT TOOLBOX ARCHITECTURE	15
4.1	Actors and Responsibilities	15
4.2	Label Lifecycle and Workflow	16
4.2.1	Overview	16
4.2.2	Workflow Stages	17
4.3	Label Format and Contents.....	18
4.4	Internal Modules.....	19
4.5	External Interfaces and Interoperability	21
5	FUNCTIONAL REQUIREMENTS	22
6	CONCLUSION AND NEXT STEPS.....	29
6.1	Next Steps.....	30
7	References.....	30
	APPENDIX I – EHDS Reg. Art. 78 & CY-EHDS-2nd DQET Legal Table Mapping	32

List of Figures

Figure 1: CY-EHDS-2nd DQET module (as implemented in WP9) 10

List of Tables

Table 1: Actors of the CY-EHDS-2nd DQET module as defined in *D5.1 Annex D: Common Actor Definitions – CY-EHDS-2nd (v1.0)*..... 16

Table 2: Estimated Lifecycle States of the Label 18

Table 3: DQUL Structure as Defined in QUANTUM D1.1 [10]..... 18

Table 4: CY-EHDS-2nd DQET Functional Requirements 23

Table 5: Legal Mapping Table: EHDS Regulation Article 78 vs. CY-EHDS-2nd DQET Features 32

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 WPs, 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, National Dataset Catalogue for health data (nHDS), Data Access Applications Management solution (DAAMs), Secure Processing Environment (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 Second Joint Action Towards the European Health Data Space (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 (DBCs) that Health Data Access Bodies (HDABs) are expected to implement, as defined by the European Commission (EC) 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 Community of Practice 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 (CP). 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 D9.1 Introduction

The CY-EHDS-2nd project aims to build the national infrastructure required to enable the secondary use of health data in Cyprus, both at national level and as part of the broader EHDS. Its purpose is to unlock the potential of health datasets for research, innovation, industry development, policymaking and evidence-based public health, by enabling secure, structured and legally governed access to high-value data.

However, Cyprus is starting from a position where such infrastructure, and the associated mindset and practices for quality-driven data reuse are still being established. One of the most immediate and critical challenges is the need to ensure that the datasets made available for secondary use are not only technically accessible, but also trustworthy, interpretable, and fit-for-purpose. Without attention to data quality and utility from the outset, the national ecosystem risks generating fragmented, low-impact results that undermine the transformative potential of health data reuse.

In this context, the DQET, developed under WP9, plays a foundational role. It introduces a structured, reviewer-validated mechanism through which data holders can assess and communicate the key characteristics of their datasets — including completeness, governance, coverage and timeliness. This helps ensure that data users across the secondary use ecosystem can understand the relevance and reliability of the datasets they intend to use, resulting in more meaningful analysis and ultimately more impactful outcomes for citizens' health.

This deliverable defines the requirements and specifications for the DQET, developed under WP9 of the CY-EHDS-2nd project. The DQET module enables the creation, validation, and publication of DQULs⁵ within the national infrastructure for secondary health data use in Cyprus (Figure 1).

This is the first in a series of three deliverables for WP9:

- D9.1 (this document) – Requirements and Specifications
- D9.2 – Pilot Implementation
- D9.3 – Production Implementation

It lays the foundation for implementation by defining the system's functional requirements, semantic structure, legal alignment and architectural components.

The DQET module provides a structured interface and governance mechanism for data holders to self-assess the quality and utility of health datasets. Labels are completed by authorized users and reviewed by the national HDAB, enabling transparent, versioned and machine-readable quality indicators to be published alongside dataset descriptions.

The module directly implements Article 78 of the EHDS Regulation (EU) 2025/327, which introduces the concept of DQULs as part of the HealthData@EU framework. The structure of the label is based on the model presented in the Deliverable 1.1 "Specification of the data sets' quality and utility label" of the QUANTUM project (hereafter referred to as QUANTUM D1.1) and supports semantic interoperability through alignment with Data Catalog Vocabulary Application Profile for health (HealthDCAT-AP) and the Data Quality Vocabulary (DQV).

⁵ For the purposes of this document, the term "label" refers to the dataset quality and utility label as defined in Article 78 of the EHDS Regulation

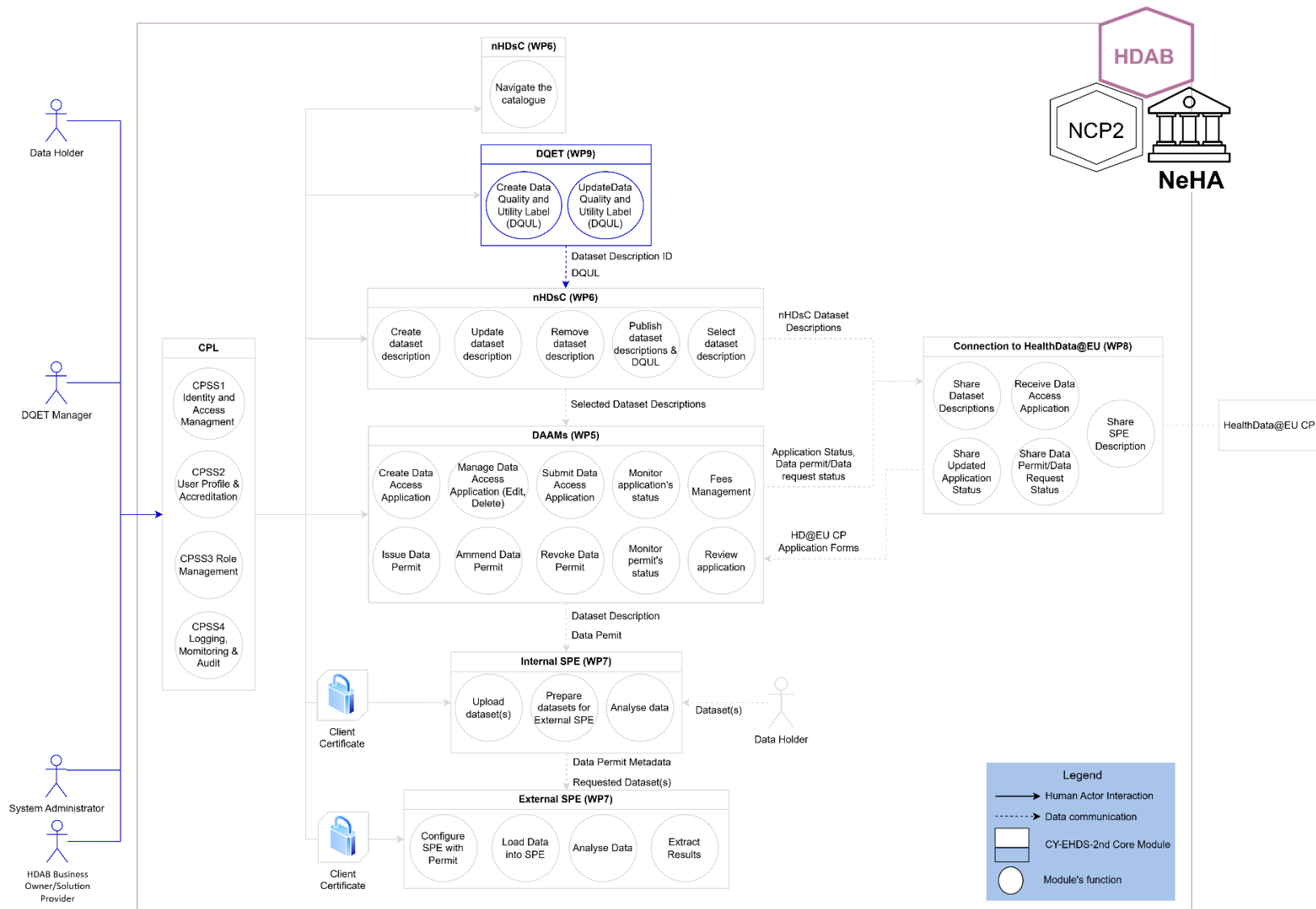


Figure 1: CY-EHDS-2nd DQET module (as implemented in WP9)

1.3 Objective of the Deliverable

The main objective of this deliverable is to lay the foundation for a national mechanism that systematically promotes and validates the quality and utility of health datasets intended for secondary use. Through the definition of the CY-EHDS-2nd DQET, this document aims to operationalize the principles of transparency, governance and reusability as introduced by the EHDS and to position Cyprus at the forefront of data quality maturity in health data infrastructures.

By establishing the DQET structure, requirements, and governance logic, the deliverable seeks to ensure that dataset providers are equipped to self-assess and improve the quality of their data, and that the HDAB National eHealth Authority of Cyprus (NeHA⁶) can apply structured, transparent and traceable validation processes. It further aims to foster a culture of quality awareness and stewardship, encouraging best practices in documentation, data management and utility communication across all actors in the national eHealth ecosystem.

Beyond defining the quality and utility label, this deliverable initiates the creation of a core toolbox that will serve as the hosting environment for future quality promotion tools and validation services. Starting with the label as the primary instrument, the DQET module will provide a framework through which the importance of data quality can be communicated to stakeholders across the secondary use ecosystem. The long-term goal is to ensure that health data is not only technically interoperable, but also trustworthy, meaningful and ultimately beneficial to public health outcomes as well as for personalized healthcare serving citizens' health at a personal level. By raising quality standards at the point of data publication, the infrastructure enables better reuse, more accurate analysis and improved evidence for decision-making.

1.4 Connection with other Work Packages

The DQET module, developed under WP9 is functionally and technically linked to several other CY-EHDS-2nd WPs that manage metadata, publication and interoperability.

The most direct and critical integration is with WP6 (nHDsC). DQET extends the catalogue's functionality by enabling data holders to create structured DQULs linked to their dataset metadata records. These labels are initiated and submitted within the nHDsC environment and are reviewed independently by the HDAB. In addition, WP6 manages the dataset descriptions validation and publication process. If the dataset is identified during WP6 review as falling under a public or EU funding obligation, the system can enforce mandatory label submission prior to publication. Thus, the governance and technical enforcement of labelling rules are coordinated through WP6. More information on the nHDsC can be found in CY-EHDS-2nd *D6.1 nHDsC Requirements and Specifications*.

Looking ahead, DQET is designed to be compatible with future export mechanisms developed under WP8 (National Connector for Cross-Border Gateway). While no active interface currently exists, the labels generated by WP9 are prepared in semantically structured formats that align with HealthDCAT-AP, enabling their inclusion in cross-border dataset catalogues once interoperability is activated under WP8 (requirements and specifications of WP8 are provided in CY-EHDS-2nd *D8.1 National Connector for Cross-border gateway Requirements and Specifications*).

⁶ <https://www.neha.org.cy/>

1.5 Interdependencies within CY-EHDS-2nd infrastructure

In addition to its alignment with other WPs, the implementation of the DQET module depends on several critical infrastructural and service-level components within the CY-EHDS-2nd ecosystem.

DQET integrates closely with:

- **CPL Shared Services:** DQET relies on three key CPL shared services to enforce identity, access, and audit requirements:
 - CPSS1 - Identity and Access Management: All users interacting with DQET, including data holders and HDAB reviewers, must authenticate through the national identity layer. CPSS1 ensures session control and consistent access across modules.
 - CPSS2 - Profile and Accreditation: Only organizations and users who are properly registered and accredited as dataset holders can access the label creation functions in DQET. CPSS2 manages this eligibility by maintaining user affiliations and entity profiles. It acts as a prerequisite for granting DQET access rights.
 - CPSS3 - Role Management: Role-specific permissions within DQET, such as the ability to create, review, or revoke labels, are enforced through CPSS3. This ensures that only authorized actors, as defined in the governance framework, can perform critical actions.
 - CPSS4 - Logging and Monitoring: All label-related events (e.g. creation, submission, validation, publication, revocation) are recorded via CPSS4. These logs are immutable, timestamped, and traceable to individual users.
 - The CY-EHDS-2nd CPL is thoroughly described in *D5.1 Annex A: CY-EHDS-2nd Architecture and CP Infrastructure*.
- **nHDsC module:** The DQET module is tightly integrated with the nHDsC module, developed under WP6. DQET labels are linked to dataset descriptions through persistent identifiers and are created, reviewed and published in coordination with WP6 lifecycle logic. The nHDsC Manager is also responsible for identifying cases in which label submission is mandatory (e.g. publicly funded datasets). This interdependency governs both the workflow conditions and the publication state logic shared between the two modules. More information on the nHDsC can be found in CY-EHDS-2nd *D6.1 nHDsC Requirements and Specifications*.

2 LEGAL FRAMEWORK

The DQET is implemented within the legal framework defined by the EHDS Regulation (EU) 2025/327) [8] and its supporting Union instruments. It specifically operationalizes the requirements and policy intentions related to DQUL for the secondary use of health data.

The core legal basis for the CY-EHDS-2nd DQET module is established in Article 78 of EHDS Regulation which introduces a common EU data quality and utility label for electronic health datasets made available through HDABs. The regulation defines:

- The conditions under which the label is mandatory or optional
- The structural elements of the label, including six content categories
- The validation and governance role of HDABs

- The possibility for the EC to define technical specifications and adopt delegated and implementing acts

In particular:

- Article 78(1) establishes that dataset descriptions made available through HDABs may have a Union data quality and utility label, provided by data holders.
- Under Article 78(2), datasets collected or processed with national, or EU public funding must include a data quality and utility label.
- Article 78(3) specifies the required components of the label, including data documentation, assessment of technical quality, data quality management processes, assessment of coverage, information on access and provision and information on data modifications.
- Article 78(4) assigns responsibility to HDABs to assess, approve, or revoke labels, ensuring trust and accountability.
- Article 78(6) requires the Commission to adopt implementing acts defining the visual and technical specifications of the label by 26 March 2027.

The DQET module is fully aligned with these obligations and anticipates the evolution of this legal framework by maintaining technical flexibility and semantic compliance with forthcoming EU-level standards. For a structured cross-reference between the EHDS Art. 78 and the CY-EHDS-2nd implementation, see *Appendix I - EHDS Reg. Art. 78 & CY-EHDS-2nd DQET Legal Mapping Table*, which maps Article 78 provisions to their corresponding features in the DQET module.

2.1 National Legal Framework

As of May 2025, there is no provision in the current Cyprus national eHealth law (L. 59(I)/2019) that explicitly addresses DQULs or defines obligations for data holders or the HDAB regarding such labels. However, the forthcoming amendment to the national eHealth law, expected by 2026, will incorporate new provisions aligned with the EHDS Regulation, including Article 78.

The CY-EHDS-2nd architecture and the DQET module are designed to anticipate and accommodate this legal evolution. Once the amended national law enters into force, the DQET governance and validation mechanisms will be updated accordingly to ensure full legal alignment at both national and Union level.

3 ALIGNMENT WITH QUANTUM PROJECT

The design and implementation of the CY-EHDS-2nd DQET is directly informed by the findings, reference model, and technical recommendations of the QUANTUM project, an EC-funded initiative (2023–2025) supporting the quality and utility labelling of health datasets for secondary use. QUANTUM provides the conceptual and technical foundation for the Union-wide data quality and utility label, as introduced in Article 78 of EHDS regulation.

The DQET module reflects and operationalizes QUANTUM core components, including the label structure, the label creation and validation workflow, and the approach to Findability Accessibility Interoperability and Reuse of data (FAIR)⁷ alignment.

⁷ <https://www.go-fair.org/fair-principles/>

3.1 Use of the QUANTUM Label Structure

CY-EHDS-2nd fully adopts the label template and dimensional model defined in QUANTUM Deliverable D1.1 [10], which is structured around six core sections:

1. Data Documentation
2. Technical Quality
3. Quality Management Processes
4. Coverage
5. Access and Provision
6. Data Modifications

Each section contains specific dimensions that describe dataset characteristics in qualitative or quantitative terms (presented in Chapter 4). These sections correspond directly to the requirements outlined in EHDS Article 78(3). The self-assessment form implemented in DQET mirrors the layout, content and terminology used in QUANTUM D1.1 to ensure semantic compatibility and avoid national divergence from the evolving EU standard.

3.2 Integration of QUANTUM Data Holder Workflow

The data holder journey within the DQET module replicates the process identified in QUANTUM end-user studies and workflow design to ensure procedural compatibility between CY-EHDS-2nd infrastructure and QUANTUM design principles and provides the necessary safeguards for maintaining label trustworthiness as envisaged in the HealthData@EU infrastructure.

QUANTUM Data Holder Journey

- A label is created after the dataset description is created
- The data holder completes a structured self-assessment
- Submission is allowed only once the dataset description has been submitted to the nHDsC
- Labels are subject to validation by an authorized reviewer
- Once validated, the label is linked and published with the dataset description

3.3 Alignment with Identified Tools and Validation Mechanisms

The QUANTUM project conducted a comprehensive landscape analysis of data quality tools and frameworks relevant to the health sector. While the current DQET implementation focuses on manual validation workflows, it is designed to integrate with emerging automated validation solutions proposed by QUANTUM. In particular:

- The label structure is implemented in structured format JavaScript Object Notation for Linked Data (JSON-LD) and can be validated using schema-based technologies such as Shapes Constraint Language (SHACL), consistent with QUANTUM vision for semi-automated quality validation
- The backend modules support version control and audit logs, enabling future adoption of automatic trust scoring or label reputation models.

Where possible, the CY-EHDS-2nd DQET system will reuse open-source tools and validation pipelines identified by QUANTUM (e.g. SHACL validators), ensuring interoperability and extensibility.

3.4 Support for FAIR, DQV and HealthDCAT-AP Vocabularies

QUANTUM emphasizes the role of semantic interoperability and metadata alignment in supporting trustworthy labels. The CY-EHDS-2nd implementation adheres to the same interoperability principles, ensuring that all label records:

- Are described using HealthDCAT-AP for dataset descriptions
- Use standardized vocabularies, including DQV for expressing label attributes
- Support machine-readable export for integration with the HealthData@EU CP⁸ infrastructure

This ensures that the national infrastructure for secondary use of health data in Cyprus remains interoperable with the HealthData@EU CP, which will incorporate labelled dataset descriptions into a EU-wide catalogue.

3.5 Future Alignment with QUANTUM and HealthData@EU

QUANTUM findings directly inform the governance design of the DQET module. The assignment of the DQET Manager role to authorised HDAB personnel reflects the QUANTUM recommendation to entrust review responsibilities to recognised entities, while ensuring traceability and procedural transparency.

Moreover, the modular architecture of DQET is intentionally designed to evolve alongside QUANTUM outputs. Notably:

- The visual and technical specifications of the label to be defined by means of implementing acts by the EC (EHDS Article 78(6))
- The eventual adoption of EU-wide validation tools as part of the HealthData@EU service ecosystem.

In this framework, the CY-EHDS-2nd project commits to monitoring and integrating future deliverables of the QUANTUM project to ensure continued alignment with the Union label framework.

4 DATASET QUALITY ENHANCEMENT TOOLBOX ARCHITECTURE

The DQET module is implemented as a modular component within the national CY-EHDS-2nd infrastructure. It follows a service-oriented architecture pattern and interacts closely with both internal CY-EHDS-2nd components and shared platform services provided by the CY-EHDS-2nd CPL. While full deployment and system-level interface specifications will be delivered in D9.2 (Pilot Implementation), this section presents the conceptual system architecture and defines the key technical modules and their interactions.

4.1 Actors and Responsibilities

The DQET operates within a defined governance and publication workflow involving three primary actors: data holders, the HDAB and the nHDsC Manager (WP6). Their roles reflect regulatory obligations set by the EHDS regulation and are embedded in the functional architecture of the CY-EHDS-2nd infrastructure and are defined in *D5.1 Annex D: Common Actor Definitions – CY-EHDS-2nd (v1.0)*. The DQET actors are outlined in Table 1.

⁸ <https://acceptance.data.health.europa.eu/Release3/documentation>

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

Actor No.	Actor	Role
A2	Data User	This role does not participate in the labelling process itself but relies on the label for evaluation and dataset selection.
A3	Data Holder	responsible for initiating the data quality and utility labelling process by completing a structured self-assessment.
A7	nHDsC Manager	ensures the publication and visibility of validated dataset records and associated labels.
A9	DQET Manager	ensures that all submitted labels are accurate, complete, and in line with national and European standards before publication.

All role-based actions are traceable, and accountability is supported by access controls defined in the CY-EHDS-2nd architecture. Audit and logging services are managed centrally via CPL infrastructure and are referenced but not replicated within DQET.

4.2 Label Lifecycle and Workflow

The DQET module operates in coordination with the nHDsC module to support the structured creation, review, and publication of DQULs. The process is based on a modular architecture where the dataset description and the label follow distinct but interrelated validation paths. This section presents the updated lifecycle and the interdependencies between the actors and components involved.

4.2.1 Overview

The label is an optional but structured enhancement that may be initiated by a data holder after they have created a dataset description. While labels are linked to dataset description records from their inception, their submission and publication follow specific conditions. Both the dataset metadata and the label are reviewed separately, but they are shown together in the data holder's personal dashboard within the nHDsC space.

The key principles of the lifecycle are:

- A dataset description must be created before the label can be created (even in draft status).
- The label creation is optional and initiated by the data holder from the data holder's dashboard in the nHDsC space.
- A label cannot be submitted for review unless the associated dataset description has been submitted.
- Once submitted, the label enters a review state visible to both the data holder and the HDAB (DQET Manager).
- The dataset description and the label are published only after separate approvals.
- The label cannot be published unless the dataset description has already been published (or is published simultaneously).
- A dataset description can be published without a label.
- A dataset description can proceed through all its lifecycle stages without any dependency on the status of the label.

4.2.2 Workflow Stages

1. Dataset Description Creation

The data holder uses the HealthDCAT-AP Editor (WP6) to create the dataset metadata record. The record appears in the data holder's dashboard within the nHDsC interface and may remain in draft or be submitted for review.

2. Enabling of Label Creation

As soon as the dataset description is created, even in draft status, the system activates a "Create Label" action for that specific dataset. This allows the data holder to proceed to the DQET self-assessment interface to begin preparing a label.

3. Label Creation and Submission

The data holder completes the label using the structured self-assessment form provided by DQET component. However, the system does not allow submission of the label for review unless the associated dataset description has already been submitted. Once both conditions are met, the label can be submitted and registered in the system appearing in the dashboard next to the associated dataset.

4. Parallel Review Process

The dataset description is reviewed by the nHDsC Auditor, while the label is reviewed independently by the DQET Manager. Each reviewer performs their own validation in accordance with governance rules. Their decisions are logged and timestamped.

5. Publication

- If the dataset description is approved, the nHDsC Manager publishes it in the national catalogue.
- If the label is also approved, it becomes visible together with the dataset metadata.
- If the label is still under review or rejected, the dataset can still be published without it.
- A label cannot be published unless the dataset description is already published or is published simultaneously.
- If the dataset description is rejected, the label is not eligible for publication.

6. Dataset Description's Updates

If a dataset description is updated by the data holder after it has already been published in the nHDsC, the previously approved DQUL is automatically invalidated. This is to ensure that labels remain accurate and reflect the current state of the dataset metadata. Upon update, the label is marked as "Expired" in the system, and the data holder is notified to initiate a new label submission reflecting the updated dataset. No label may remain publicly linked to an outdated dataset description record.

The lifecycle of the DQUL follows a defined sequence of states that reflect its creation, submission, validation, and potential publication. While the formal state model will be finalized as part of the ongoing work in the QUANTUM project, Table 2 presents the current estimated state structure to be adopted in the CY-EHDS-2nd implementation.

Table 2: Estimated Lifecycle States of the Label

State	Description
Draft	The label is being prepared by the data holder
Submitted	The label has been submitted and awaits review
Under Review	The DQET Manager is reviewing the label
Approved	The label is validated and ready for publication, pending dataset approval
Published	The label is published together with the approved dataset description
Rejected	The label has not been approved and feedback is provided
Expired	The label has exceeded its validity period and requires update
Revoked	The label has been revoked by the DQET Manager due to governance non-compliance

4.3 Label Format and Contents

The DQUL used within the CY-EHDS-2nd infrastructure adopts the model defined in the European project QUANTUM. The label is completed by the data holder during the self-assessment process and subsequently reviewed by the DQET Manager prior to publication. It is stored in a structured format and made available to data users through nHDS*C*.

The structure of the label consists of six main sections. Each section contains a defined list of dimensions that the data holder must address. These dimensions reflect documentation quality, technical attributes, data governance processes, population coverage, access characteristics, and any modifications made to the dataset.

Table 3 presents the official structure of the label, using the exact naming conventions and content defined in QUANTUM D1.1⁹ [10].

Table 3: DQUL Structure as Defined in QUANTUM D1.1 [10]

No.	Section Name	Short Description	Dimensions
DQUL-S1	Data Documentation	Describes the metadata, structure, and documentation supporting the dataset, including the source and format used.	1. Metadata 2. Support documentation 3. Data dictionary 4. Format and standards used 5. Provenance 6. Data model
DQUL-S2	Technical Quality	Captures core data quality indicators relating to completeness, validity, accuracy, and overall technical fitness.	1. Completeness 2. Uniqueness 3. Accuracy 4. Validity 5. Timeliness 6. Consistency
DQUL-S3	Data Quality Management Processes	Identifies whether formal processes exist for reviewing, auditing, and managing data quality, including bias detection.	1. Review processes 2. Audit mechanisms 3. Biases examination

⁹ QUANTUM D1.1, Section 4.2 “Final Specification of the Label Structure” (February 2025).

DQUL-S4	Coverage	Indicates the time and population scope of the dataset, including representativity and duration of subject data.	1. Representation of multi-disciplinary electronic health data 2. Representativity of population sampled 3. Average timeframe in which a natural person appears in a dataset
DQUL-S5	Access and Provision	Reflects operational metrics on how quickly data is made available after collection and how long delivery takes.	1. Time between the collection of the electronic health data and their addition to the dataset 2. Time to provide electronic health data following electronic health data access application approval
DQUL-S6	Data Modifications	Describes whether data is merged, linked, or incrementally updated over time.	1. Merging and adding data to an existing dataset 2. Including links with another dataset

Each dimension includes qualitative or quantitative descriptions provided by the data holder, supported by documentation where possible. The DQET interface enforces a structured format for these responses.

The CY-EHDS-2nd implementation will adopt this structure without modification. In line with practices used in the HealthData@EU infrastructure, the label is intended to be implemented in a structured, machine-readable format, such as JSON-LD, and validated using SHACL or an equivalent schema-based mechanism to ensure data integrity and interoperability. This approach will remain flexible and will be revised as needed based on forthcoming guidance from the EC or final outputs of the QUANTUM project.

4.4 Internal Modules

The DQET is designed as a modular system that supports the creation, validation, and governance of DQULs. It operates in coordination with the nHDsC, while ensuring strict separation of responsibilities between data holders, label reviewers, and catalogue reviewers and managers.

This section describes the core internal modules of DQET, focusing on their functional roles and contribution to the label lifecycle. The architecture reflects both the user-facing interface and the backend components required to manage review workflows, status tracking, and data integrity.

1. Self-Assessment Module

This module provides a structured interface for data holders to complete the label, based on the final model defined by the QUANTUM project. It guides the user through all six sections of the label, with structured input fields, drop-down menus where applicable, and support for attached documentation.

The module enforces:

- Required fields per section
- Help text based on each dimension
- Draft saving functionality
- Real-time validation of field formats (e.g. dates, URLs)

The dataset's quality self-assessment module is only accessible only after the creation of the corresponding dataset description even while in draft status.

2. Label Validation

In the current implementation of the CY-EHDS-2nd infrastructure, the validation of DQULs is performed manually by the DQET Manager. Approval by the DQET Manager is a necessary precondition for the label's publication in the nHDsC, where it is linked to the corresponding dataset description. While the creation, submission, and validation of the label can proceed independently with the dataset description's review process, the label cannot be published unless the dataset description has first been approved and published. The DQET Manager is responsible for assessing the completeness, internal consistency, and plausibility of the submitted self-assessment. All review actions and decisions are documented and logged within the DQET Manager's dashboard.

Although the validation process is presently manual, the system is designed to accommodate structured inputs and schema-based field controls where applicable. This supports internal consistency and prepares the infrastructure for the adoption of semi-automated validation in future releases.

The QUANTUM project, in Deliverable D1.1, confirms that future deliverables will define and support automatic tools for label validation and FAIRness checking. As soon as these technical components become available, the CY-EHDS-2nd DQET infrastructure will be updated to integrate the official EU-level validation mechanism, ensuring alignment with evolving European interoperability standards.

3. DQET Manager's Dashboard

This administrative module is accessible only to authorized staff within the HDAB (NeHA). It presents an overview of all submitted labels pending review, including their associated dataset metadata and supporting files. The DQET Manager dashboard only displays labels that are linked to dataset descriptions which have reached the 'submitted' status in the nHDsC module. Labels linked to draft dataset descriptions are not visible for review and cannot be processed until the health data metadata record is formally submitted.

Features include:

- Case-by-case label review and approval interface
- Predefined decision actions (e.g. approve, reject, request clarification, revoke)
- Version history and audit log access
- Deadline tracking and notification system

The dashboard ensures traceability and supports role-based access control aligned with the national HDAB governance framework.

4. Label Store and Version Control

This module manages the persistence of label records, their version history, and associated timestamps. It ensures that:

- Each label is uniquely identified and versioned
- Expired labels are marked and archived
- All reviewer actions are stored in immutable logs
- Every published label can be traced to its corresponding metadata record (dataset description)

Each dataset label submitted through DQET is assigned a version identifier upon creation. When the data holder edits and resubmits the label (without changes to the underlying dataset description), the system generates a new label version. This allows incremental updates (e.g. corrections, improved scoring, added documentation) to be tracked independently of the dataset description lifecycle. Each version is timestamped, logged and made available in the reviewer dashboard. Only the latest approved version is visible in the public catalogue.

5. Catalogue Linkage Component

This backend component ensures the technical and functional linkage between DQUL and its corresponding dataset metadata record in the nHDsC. Once the label has been validated and approved by the DQET Manager, it is made available for publication through the nHDsC module (WP6) provided that the associated dataset description has also been approved and published. The system enforces a rule that a label cannot be published unless the corresponding dataset metadata record has also been approved and published. Labels linked to unpublished datasets will remain in approved and pending for publication state.

The component supports:

- Metadata record identification through a persistent internal Identification (ID)
- Status tracking of the label (e.g. “submitted”, “approved”, “published”, “expired”)
- Synchronization with the user interface of the nHDsC module to ensure accurate presentation and state updates

4.5 External Interfaces and Interoperability

The DQET interacts with external systems and components of the CY-EHDS-2nd infrastructure. These interfaces ensure interoperability, secure data exchange, and the visibility of validated labels within the national and European context. DQET is not designed as a standalone system but as a modular component integrated into the broader health data governance ecosystem.

1. Interface with the nHDsC - WP6

DQET maintains a close technical and functional integration with the nHDsC module developed under WP6. This connection enables the alignment of dataset descriptions and quality labels and ensures that labels are only published in association with validated dataset descriptions.

The main interactions between DQET and nHDsC modules are:

- **Dataset Reference Linkage**

Each label is linked to an existing dataset metadata record through a persistent internal identifier. This linkage is established as soon as the label is created in draft form and persists through all workflow steps (e.g. submission, review, approval, publication, rejection). The linkage is required to ensure that each label is uniquely associated with a specific dataset metadata record within the nHDsC.

- **Dashboard Synchronization**

The data holder’s dashboard in the nHDsC module includes the list of submitted datasets. Once a label is created, it appears alongside the corresponding dataset with its metadata (e.g. label version, creation date, status). This allows contextual navigation between dataset and label.

- **Publication Trigger**

Once the label is approved by the DQET Manager and the dataset description is approved by the nHDsC Auditor, the system triggers publication through the nHDsC Manager interface. The label is then exposed in the public dataset catalogue as an enriched metadata component.

Versioning and linkage between dataset descriptions and associated labels are managed within the HDAB Portal infrastructure (WP6), not by the CPL. Each versioned metadata record is assigned a unique identifier, and labels are linked at creation time through this persistent reference. CPL services support auditability and logging of version changes, but not version control itself.

The integration follows modular principles and does not replicate metadata management. All metadata fields remain under the nHDsC module governance, while DQET module governs only the label content and its creation, submission and review process.

2. Interface with the CPL

DQET interacts with a set of shared infrastructure services provided by the CPL which ensures consistency, security, and traceability across all CY-EHDS-2nd components. Rather than implementing these capabilities independently, DQET references and integrates CPL services for identity management, access control, auditing, and versioning.

The following CPL services are used:

- **Identity and Access Management**

User authentication for both data holders and reviewers are handled via CPSS1 - Identity and Access Management and CPSS3 - Role Management shared services including support for CY Login and role-based access control (e.g. data holder, DQET Manager, nHDsC Auditor). Access rules are enforced uniformly across interfaces between DQET and nHDsC modules.

- **Action Logging and Audit Trail**

All critical user actions in DQET module, including label creation, submission, reviewer decisions (approval, rejection, withdrawal) and publication triggers, are logged through CPSS4 - Logging and Monitoring shared service provided by the CPL. These records are timestamped, attributed to the appropriate user, and stored immutably to support traceability, governance reporting, and auditability in accordance with EHDS Article 78(4).

3. Future Integration with the HealthData@EU Infrastructure

While not directly connected in the current phase, DQET module is designed to remain interoperable with the dataset metadata and publication frameworks defined by the HealthData@EU CP. Labels validated through DQET module and exposed via the nHDsC module will be published using HealthDCAT-AP compatible metadata formats, allowing eventual integration with the HealthData@EU CP catalogue.

5 FUNCTIONAL REQUIREMENTS

The functional requirements of the DQET module developed under the WP9 of the CY-EHDS-2nd project, are provided in this chapter. These requirements define the core functionalities that the DQET module must implement to support the creation, review, validation, and publication of DQULs within the CY-EHDS-2nd infrastructure.

The scope of this chapter is limited strictly to functional behaviors specific to the DQET module. Non-functional requirements (e.g. security, performance, availability) and underlying technical specifications (e.g. audit logging, access control, infrastructure resilience) are covered by the common CY-EHDS-2nd CPL shared services in CY-EHDS-2nd *D5.1 Annex A: CY-EHDS-2nd Architecture and CP Infrastructure*. DQET fully inherits those platform-level properties and does not redefine them here. The requirements listed in this chapter are scoped exclusively to functionality implemented within the DQET module. Features already provided by other CY-EHDS-2nd components, such as dataset metadata versioning (WP6), access control and audit logging (CPL), or DAAMs governance workflows (WP5) are excluded to avoid duplication asd they are documented in different deliverables. The list reflects the current implementation scope of DQET and is expected to evolve in future project phases to incorporate refinements from the final QUANTUM label model and emerging EHDS implementation guidelines.

Each requirement is derived from the detailed architectural and workflow logic presented in Chapter 4 and is cross-validated against the principles defined in the EHDS regulation, the HealthData@EU CP technical documentation [1-7] and the QUANTUM project deliverables, particularly the current version of the quality and utility label specification (D1.1) [10].

Table 4 presents the functional requirements of the DQET module. Each requirement is presented along with a set of use cases which are not exhaustive but serve to ground each requirement in practical user interactions.

Table 4: CY-EHDS-2nd DQET Functional Requirements

Requirement Code	DQET-FR1
Requirement Title	Visibility of Label Creation
Requirement Features	- The "Create Label" action is only visible once a dataset description has been created (even in draft state) - Label creation is triggered from the dataset dashboard in nHDsC - Labels cannot exist without a linked dataset metadata record
Detailed Description	The system must restrict access to the label creation functionality so that it becomes visible only when a dataset metadata record has been created in the nHDsC module. This ensures that every label is structurally linked to a dataset description from the outset. The "Create Label" action must be rendered dynamically in the user interface of the data holder’s dashboard. No labels may be created in isolation.
Use Cases	DQET-UC1: A data holder logs in for the first time and sees no option to create a label, as no dataset has been started DQET-UC2: A data holder begins creating a dataset in the nHDsC; once saved, the “Create Label” button appears next to the draft dataset DQET-UC3: A data holder attempts to initiate label creation without an existing dataset (e.g. via an outdated bookmark or shared link). The system prevents access and prompts the user to first create a dataset in the nHDsC.
Requirement Code	DQET-FR2
Requirement Title	Label Submission Gated by Dataset Description Submission

Requirement Features	- A label cannot be submitted unless the associated dataset description has reached the "submitted" status - The "Submit" button for the label is disabled or hidden if this condition is not met - The UI should notify the data holder when the precondition is not satisfied
Detailed Description	To ensure consistency between dataset and label governance workflows, the system must enforce a gating rule on label submission. Although the data holder may begin creating a label as soon as a dataset description exists (even in draft), the submission of the label for review is only permitted once the dataset description has been formally submitted via the nHDsC module. The DQET module must enforce this both in the user interface and backend service layer to avoid premature validation requests.
Use Cases	DQET-UC4: A data holder completes a label linked to a draft dataset and clicks "Submit." The system blocks the action and returns a message indicating that label submission is only allowed once the associated dataset description has been submitted. DQET-UC5: After the submission of the dataset description in the nHDsC space, the "Submit Label" button becomes active in the DQET interface DQET-UC6: A data holder saves a draft label while the dataset is still unsubmitted. The draft is preserved but cannot be submitted until the dataset status changes
Requirement Code	DQET-FR3
Requirement Title	Review Workflow for Label
Requirement Features	- Labels submitted by data holders must be reviewed independently by a DQET Manager - The reviewer must have access to the submitted label, metadata and any attachments - Review actions must include: Approve, Reject, Revoke (post-approval), Request Clarification - Reviewer comments must be logged and displayed to the data holder - Reviewer interface must show the linked dataset description (read-only)
Detailed Description	The DQET module must implement a dedicated reviewer interface for authorized HDAB personnel (DQET Managers) to evaluate submitted dataset labels. This review is separate and independent from the metadata review handled by the nHDsC Auditor. The DQET Manager must be able to view the full content of the label, issue a decision, and optionally provide feedback. Valid decisions include Approve, Reject, and Revoke. Once a decision is made, the label status is updated accordingly and reflected in both reviewer and data holder dashboards.
Use Cases	DQET-UC7: A DQET Manager accesses the reviewer dashboard, selects a submitted label, reviews the self-assessment content, and approves it DQET-UC8: The reviewer finds inconsistencies in the label and chooses "Reject," entering a comment; the label status is updated and the data holder is notified DQET-UC9: A previously approved label is later found to violate governance rules;

	the DQET Manager issues a revocation, and the status changes to “Revoked” in the system
Requirement Code	DQET-FR4
Requirement Title	Conditional Publication of Label
Requirement Features	• A label can only be published if the associated dataset description has been approved and published • Labels that are approved while the dataset is still under review remain in a “pending publication” state • A label and its dataset may be published together if both are approved at the same time • The system prevents publication of any label linked to a rejected or unpublished dataset
Detailed Description	The DQET module must enforce strict sequencing logic to ensure that no label is published independently of its corresponding dataset description. If a label is approved but the dataset is not yet published, the label remains in a pending state. The system must synchronize with the nHDsC module to detect when both conditions are met and trigger the coordinated publication. If the dataset is rejected, any approved label must remain unpublished and is marked as blocked.
Use Cases	DQET-UC10: A label is approved by the DQET Manager, but the dataset description is still under review. The label status should indicate that it is approved but awaiting for dataset publication. DQET-UC11: A dataset and its label are both approved in parallel. The system publishes them together in the catalogue. DQET-UC12: A dataset description is rejected. The linked label remains unpublished and its status is blocked automatically
Requirement Code	DQET-FR5
Requirement Title	Self-Assessment Form Based on QUANTUM Structure
Requirement Features	• The label form must implement the six-section structure defined in QUANTUM D1.1 • Form logic must follow the terminology and structure of the QUANTUM model • The label must be rendered in machine-readable format (e.g. RDF or JSON-LD), compatible with HealthDCAT-AP extensions • The form implementation must remain adaptable to future structural updates from QUANTUM
Detailed Description	The DQET module must provide a structured self-assessment interface for data holders to create a DQUL. The form must reflect the six-part structure defined in QUANTUM D1.1: Dataset Overview, Provenance, Data Quality, Data Utility, Reusability Conditions, and Evaluation Summary. Fields must support both descriptive entries and standardized responses. The system must generate a machine-readable label aligned with HealthDCAT-AP and designed for future

	integration into EU catalogues. Structural changes issued by QUANTUM must be adopted in future releases.
Use Cases	DQET-UC13: A data holder completes the six sections of the self-assessment, including optional attachments, and generates a complete label for review DQET-UC14: A future update to the QUANTUM label structure adds new fields. The system administrator updates the form generator accordingly DQET-UC15: A label is exported in machine-readable format and integrated into a metadata publishing pipeline following HealthDCAT-AP compatibility
Requirement Code	DQET-FR6
Requirement Title	Expiration of Label on Dataset Description Update
Requirement Features	• The system must monitor for updates to published dataset descriptions in the nHDsC • When a dataset is updated, any linked published label is automatically marked as "Expired" • The data holder must be notified and prompted to submit a new label for the updated dataset • Expired labels must remain accessible in the audit trail and reviewer history but not displayed in the nHDsC
Detailed Description	To preserve alignment between metadata and labels, the DQET module must automatically revoke the current label if its linked dataset description is updated after publication. This ensures that labels always reflect the current state of the dataset. The system must integrate with the nHDsC module to detect such updates and trigger label status changes accordingly. The data holder must be notified through the dashboard that a new label is required. Previously published labels should remain accessible internally for auditing but be withdrawn from public display.
Use Cases	DQET-UC16: A dataset is published with an approved label. The data holder later updates the dataset description, triggering automatic expiration of the label DQET-UC17: The system sends a dashboard notification to the data holder indicating that the label must be resubmitted following metadata changes
Requirement Code	DQET-FR7
Requirement Title	Label Lifecycle State Management
Requirement Features	• The system must assign and track a label's lifecycle state throughout its existence • Supported states include: Draft, Submitted, Under Review, Approved, Published, Rejected, Expired, and Revoked • State transitions must be triggered by user actions (e.g. submit, approve) or system events (e.g. dataset update) • The current state must be clearly visible in both the data holder and reviewer dashboards • Historical states must be recorded in the audit trail

Detailed Description	The DQET module must manage and expose a full state model for each DQUL. Each label transitions through clearly defined lifecycle states representing its progress from creation to publication or deprecation. These states must be consistently enforced and updated in real time based on actions taken by the data holder, DQET Manager or the system itself. The current state must be visible in the user interface, and historical states must be stored for governance traceability and auditability.
Use Cases	DQET-UC19: A data holder saves a new label. Its status is marked “Draft” DQET-UC20: After submission, the label appears in the reviewer dashboard as “Under Review” DQET-UC21: A label is approved, but the dataset is not yet published; the label remains in “Approved – Pending Publication” DQET-UC22: A data holder submits a label, and the DQET Manager approves it. Once the associated dataset description is also approved and published by the nHDsC Manager, the system automatically updates the label’s status to “Published” and displays it in the public dataset catalogue alongside the metadata.
Requirement Code	DQET-FR8
Requirement Title	Label Versioning per Dataset Description Record
Requirement Features	• The system must support versioning of labels submitted for the same dataset description record • Each version must have a unique identifier and timestamp • Only the latest approved version may be published or displayed publicly • Past versions must be retained in the reviewer history for audit and governance purposes • Data holders must be able to edit and resubmit labels without losing version traceability
Detailed Description	The DQET module must allow data holders to revise and resubmit labels for the same dataset without overwriting historical records. Each submitted version must be stored with its own version identifier, creation date, and status. When a new version is approved, it replaces the previously published version in the public dataset catalogue. Prior versions must remain accessible in the internal dashboards for the DQET Manager and HDAB governance officers, ensuring transparency and regulatory traceability. Data holders must be able to access, edit, and resubmit their label at any time, including after initial submission or following approval, in order to reflect new information or improvements.
Use Cases	DQET-UC23: A data holder realises an error in a previously submitted label, edits the form, and submits a new version. Both versions remain available in the reviewer dashboard DQET-UC24: A data holder selects to update a published label linked to a dataset description, modifies selected sections based on new information, and submits a

	revised version. Upon approval, the updated label replaces the previously published version in the catalogue.
Requirement Code	DQET-FR9
Requirement Title	UI Dashboard Synchronisation Between DQET and nHDsC
Requirement Features	- Labels must appear in the data holder's nHDsC dashboard alongside the associated dataset - Each label must display key metadata: status, version, creation date, and review result - Updates in DQET (e.g. submission, approval, expiration) must reflect in the dashboard view in near real-time - Synchronisation is unidirectional: DQET pushes status data to nHDsC
Detailed Description	To ensure clarity and usability, the DQET module must synchronise label status and metadata with the nHDsC dashboard. This allows data holders to view the label's lifecycle alongside the dataset description it is linked to. Each label record must include key details such as version, creation timestamp, current status, and reviewer feedback (if any). The synchronisation is one-way: DQET must notify or push updates to the WP6 module, which then reflects the current label state in the data holder's catalogue interface.
Use Cases	DQET-UC25: A data holder submits a label and sees it appear next to the corresponding dataset description in the nHDsC data holder's dashboard with a "Submitted" status DQET-UC26: The DQET Manager approves a label. The status automatically updates to "Approved" in the data holder's view DQET-UC27: A label is revoked due to a dataset update; the nHDsC dashboard updates the status, and the label is removed from the nHDsC
Requirement Code	DQET-FR10
Requirement Title	DQET Manager Decision Logging and Audit Trail for Label
Requirement Features	- Every review action taken by the DQET Manager must be logged: Approve, Reject, Revoke - Each decision must be timestamped and linked to the reviewer's identity - Reviewer comments or justification must be stored and shown to the data holder - The full decision history must be retained for internal governance and audit purposes - Logs must be immutable and accessible to authorised governance personnel
Detailed Description	The DQET module must maintain a comprehensive audit trail of all review decisions performed on DQUL s. Each decision, whether approval, rejection, or revocation, must be logged with a timestamp and the identity of the reviewer. If reviewer feedback is provided, it must be stored and presented to the data holder on the dashboard. All logged actions must be preserved in an immutable format, forming part of the internal governance archive and enabling traceability in case of audits or disputes.

Use Cases	DQET-UC28: A DQET Manager approves a label. The system records the action with timestamp and reviewer ID DQET-UC29: A rejected label includes reviewer feedback explaining the reasons. This is shown in the data holder's dashboard view DQET-UC30: During a governance audit, an HDAB compliance officer accesses the decision history of a revoked label to confirm procedural integrity
Requirement Code	DQET-FR11
Requirement Title	Upload of Supporting Documentation per Label Section
Requirement Features	- Each label section must allow the data holder to upload one or more supporting documents - Supported formats include PDF, DOCX, CSV, JavaScript Object Notation (JSON), and common image formats (e.g. PNG, JPG) - Uploaded documents must be displayed in the reviewer interface alongside the relevant section - File metadata (name, format, timestamp) must be stored and linked to the label version - All files must be retained with the label for audit and governance purposes
Detailed Description	The DQET module must allow data holders to upload supporting documentation for each section of the label as part of the self-assessment process. This includes, but is not limited to, data dictionaries, cohort definitions, quality reports, and reuse conditions. The uploaded documents must be available to the DQET Manager during the review process, directly associated with the section they support. These files form part of the evidentiary basis for the label assessment and must be retained and versioned with the corresponding label submission.
Use Cases	DQET-UC31: A data holder completes the "Data Quality" section and uploads a PDF report validating the coverage of input data DQET-UC32: A reviewer opens a supporting document attached to the "Provenance" section to confirm the data source lineage claim DQET-UC33: A HDAB officer audits a previously published label and retrieves all uploaded documents linked to the approved version for legal review

6 CONCLUSION AND NEXT STEPS

The functional scope, regulatory alignment, internal architecture, and functional requirements of the DQET module which will be implemented under the CY-EHDS-2nd WP9 is presented in this deliverable. It represents one of the core five functional modules of the national health data infrastructure for secondary use, supporting the creation, validation and lifecycle management of quality and utility labels for structured health datasets.

DQET directly supports the objectives of EHDS Regulation Article 78, by enabling data holders and the HDAB to assess and communicate the value of datasets in a transparent and standardized manner. Its design allows structured quality information to be reviewed, published and eventually integrated into national and cross-border catalogues, increasing trust and enabling more informed dataset discovery and reuse.

Beyond its technical functions, DQET modules is expected to play an essential role in promoting data quality culture and best practices for across Cyprus health data ecosystem. By encouraging transparency in documentation, population coverage, timeliness, and data governance, it helps raise the maturity of participating data holders and aligns national practices with emerging European norms.

6.1 Next Steps

Following this Requirements and Specifications phase, the CY-EHDS-2nd consortium will proceed to the pilot implementation of the DQET module, to be documented in Deliverable D9.2. This next phase will apply a structured pilot implementation to verify functional workflows, reviewer logic, schema structure and user interface design across real-world datasets.

The outcomes of the pilot will allow refinement of:

- Interface usability for both data holders and reviewers
- Validation mechanisms and potential automation via SHACL
- Integration with the nHDsC (WP6) and CPL services
- Semantic export formats for future cross-border label publication

This pilot will serve as a critical checkpoint for ensuring technical robustness and user acceptability prior to production deployment (D9.3).

Throughout the pilot and production phases, the implementation of DQET will remain aligned with evolving European and national requirements, including:

- Final outputs of the QUANTUM project, particularly machine-readable label models and SHACL constraints
- Implementing acts adopted by the EC under EHDS Article 78(6), which may specify mandatory label formats, validation processes, or export models
- Any amendments to the Cyprus eHealth Law (L. 59(I)/2019) or new legal provisions introduced to harmonies national governance with Union-level standards

The CY-EHDS-2nd project will continue to track these developments and integrate them into both the system design and operational governance model of DQET, ensuring long-term sustainability and legal compliance.

7 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.
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. **European Commission (2025).** *Digital Business Capabilities Blueprint.* EHDS2 Community of Practice Confluence Page, January 2025.
10. **QUANTUM Consortium (2025).** *D1.1 – Conceptual Framework and Initial Technical Requirements for Data Quality and Utility Labelling.* Version FINAL, 27 February 2025.

APPENDIX I – EHDS REG. ART. 78 & CY-EHDS-2ND DQET LEGAL TABLE MAPPING

Table 5 maps the key legal provisions under Article 78 of Regulation (EU) 2025/327 (EHDS) to the corresponding technical and governance features implemented in the CY-EHDS-2nd DQET module. It supports traceability and alignment between the regulatory framework and the national infrastructure implementation.

Table 5: Legal Mapping Table: EHDS Regulation Article 78 vs. CY-EHDS-2nd DQET Features

EHDS Article	Correct Legal Content (Title / Summary)	CY-EHDS-2nd DQET Feature Supported
78(1)	<i>Voluntary label by data holders</i>	Optional label creation initiated by data holders through the DQET self-assessment interface
78(2)	<i>Mandatory label for publicly funded datasets</i>	Flag in DQET workflow ensures label is required if funding source indicates public or EU-funded origin
78(3)	<i>Label content elements</i>	DQET label adopts all six sections: data documentation, technical quality, quality management maturity, coverage, access/provision, data modifications
78(4)	<i>Revocation by HDAB if label is inaccurate</i>	DQET Manager (HDAB role) has authority to assess and revoke labels. Any action is logged and versioned
78(5)	<i>Delegated acts may amend label elements (para 3)</i>	DQET schema (JSON-LD, SHACL) is modular and version-controlled to support future amendments to content
78(6)	<i>Commission to adopt implementing acts by 26 March 2027 setting out visual and technical specs</i>	DQET interface and architecture designed to accept prescribed label appearance and technical layout once adopted by the EC