HealthData@IE – setting up health data access body services in Ireland

Deliverable 6.1

National health dataset catalogue: Requirements and Specifications Report

May 2026

Có-mhaoinithe ag an Aontas Eorpach
Co-funded by the European Union

Health Information and Quality Authority
An tÚdarás Um Fhaisnéis agus Cáilíocht Sláinte

An Roinn Sláinte
Department of Health

HRB
An Bord Taighde Sláinte
Health Research Board

Work Package 6: National Health Dataset Catalogue Requirements and Specifications

Deliverable number and name	6.1 – National Health Dataset Catalogue-Requirements and Specifications Report
Work Package number	Work Package 6
Lead Beneficiary	HIQA - Health Information and Quality Authority
Type	R – Document, report
Dissemination Level	Public
Due Date	Month 24 - October 2025
Contact email	healthinformation@hiqa.ie

HealthData@IE: setting up health data access body services in Ireland 2023-2027

Project number: 101129357 – HDAB IE

Call: EU4H-2022-DGA-MS-IBA2

Table of Contents

Acknowledgments

HIQA would like to thank all of the individuals and organisations who provided their time, advice and information in support of this report.

Particular thanks are due to the project Steering Group chaired by the Department of Health, see **Appendix 1** for a list of the Steering group members and the Working Group, see **Appendix 2** for Working Group members list. Your input was valuable in informing the development of the nHDsC Requirements and Specification report.

Members of the Health Information and Standards Directorate Team:

This report was led by staff in HIQA's Health Information Quality Team.

The following individuals contributed to the management, technical writing or dissemination of this report:

Rachel Flynn, Barbara Foley, Suzanne Barror, Kirthan Rama Chandra, Joanne Fallon, Clare O'Donnell.

Conflicts of Interest

None reported.

1. Introduction

1.1 Overview of the European Health Data Space

The European Health Data Space (EHDS) is the first common data space to emerge from the European data strategy. The EHDS Regulation was published in the Official Journal of the European Union on 5 March 2025 and entered into force on 26 March 2025.(1)

The EHDS will empower individuals to take control of their own health data and make it easier to access and exchange health data across EU Member States, both to support health and social care delivery (known as primary use of data) and to facilitate other uses of the data, including research and policy-making (known as secondary use of data).(2)

Under the EHDS regulation, each Member State is required to establish Health Data Access Body (HDAB) services by March 2029. A HDAB is a designated national entity that supports the secondary use of health data by enabling data users such as researchers and policymakers to discover and access datasets. Dataset discovery will be facilitated through a National Health Dataset Catalogue (nHDsC), while access to data will be managed through a structured Data Access Application Management System (DAAMS). Once a request is reviewed and approved, the HDAB will issue a data permit that defines the conditions for access, including strict data processing requirements and usage within a highly Secure Processing Environment (SPE) under supervision.

Article 77 of the EHDS Regulation mandates HDABs to provide metadata about available datasets that fall under the 17 categories of health data set out in the Article 51 of the EHDS Regulation.(3) The metadata must detail the source, scope, main characteristics, data nature and access conditions of the datasets.(1) Categories of health data include electronic health records, data impacting on health, pathogen genomics data, and healthcare-related administrative data, among others. The majority of datasets that fall within these data categories must be made discoverable by 26 March 2029, which is four years after the EHDS Regulation's entry into force. Certain sensitive data types, such as genetic data, have an extended timeline and must be made discoverable by 26 March 2031, six years after the regulation becomes effective (for a full list of data categories please see Appendix 3).(4) Implementation of these obligations under the EHDS Regulation will be supported in Ireland by the development of a nHDsC.

1.2 National Health Dataset Catalogue

A National Health Dataset Catalogue (nHDsC) is a centralised digital registry designed to support the discovery of health and related datasets across the nation. It plays a vital role in enabling both data users and data holders to participate in a unified, transparent, and standardised metadata environment. Its primary function is to make datasets findable and understandable, thereby supporting the use of health data for secondary purposes such as research, innovation, service planning, and policy development.(5)

By offering a consistent structure for metadata publication, the nHDsC allows data users to efficiently locate, evaluate, and understand relevant datasets, while supporting data holders in improving the visibility, governance, and usability of the datasets they maintain.

Each dataset entry in the nHDsC includes key metadata elements such as the dataset's title, a brief description, the thematic area it relates to, geographic and temporal coverage, update frequency, data format, and licensing or access terms. This structured approach to metadata not only promotes interoperability and transparency but also aligns with broader national digital transformation strategies and ensures compliance with the European Health Data Space (EHDS) regulation.

Article 79 of the regulation requires the European Commission to develop an EU dataset catalogue that brings together Member State catalogues and those of other authorised participants under HealthData@EU.(3) This catalogue will form part of the overall HealthData@EU infrastructure and will act as a federated system that interconnects nHDsC. It supports cross-border health data discoverability by aggregating health data metadata from Member States via a unified European Catalogue and making it accessible through a secure gateway at the EU level. The integration of Ireland's nHDsC with the EU central platform will be facilitated through a National Contact point (NCP). An NCP manages the automatic transmission (at semantic and format level) of actions from the nHDsC to the HealthData@EU Central Platform, this includes the management of metadata records and other relevant communications.(3) Once an NCP is established, Ireland's nHDsC will be integrated into the European infrastructure. This is essential to ensure that national datasets are not only discoverable domestically but also findable across Europe. This enhances cross-border collaboration in research, policy-making, and evidence-based health interventions while strengthening Europe's collective capacity to respond to health challenges. By aligning with European standards and connecting with the

HealthData@EU Central Platform, Ireland's nHDsC will contribute to a coherent, interoperable, and future-proofed EHDS.

1.3 Standardisation of National Health Dataset Catalogues

A metadata standard is a defined set of guidelines that specify how metadata should be structured, maintained, and exchanged thus making data easier to manage, discover, and use. It ensures consistency, quality, and interoperability across systems. In essence, it acts like a grammatical framework for data, defining how information about datasets (such as their origin, structure, and usage) should be described. Metadata standards are essential for effective data management, supporting data sharing, collaboration, and regulatory compliance.

In order to ensure the standardisation of catalogues across the EHDS, nHDsC must use a metadata standard known as HealthDCAT-AP.(6) This is an extension of the Data Catalogue Application Profile (DCAT-AP), a specification for describing public sector datasets in Europe based on the Data Catalogue vocabulary (DCAT[1]). DCAT-AP is used widely across open data portals in Europe. While DCAT-AP provides a minimal common basis within Europe to share datasets and data services across borders and domains, HealthDCAT-AP introduces a refined RDF (Resource Description Framework) vocabulary to accommodate the unique requirements of electronic health data. By incorporating established health standards (as outlined in the Appendix 4), HealthDCAT-AP offers a comprehensive framework for describing health datasets in a machine-actionable and interoperable way. This ensures that health data, whether clinical, epidemiological, or genomic, can be efficiently exchanged across Member States, supporting secondary uses such as research, innovation and public health analysis.

Of note, DCAT itself does not provide guidance on metadata management. To address this gap, the ISO/IEC 11179 standard can be applied to the nHDsC as a complementary standard as it provides the necessary guidelines for Health Data Access Bodies to manage metadata effectively. Together DCAT and ISO/IEC 11179 can support the creation of interoperable and well-governed health data spaces.(7) The usage of these standards will ensure that data cataloguing and metadata practices are

[1] DCAT is a Resource Description Framework (RDF) vocabulary designed to facilitate interoperability between data catalogues published on the Web. (https://www.w3.org/TR/vocab-dcat-3/). DCAT-AP is a specification for describing public sector datasets in Europe based on the Data Catalogue vocabulary.

harmonised among Member States, enhancing interoperability and accessibility across the EU.[8]

1.4 Development of a National Health Dataset Catalogue for Ireland

1.4.1 HealthData@IE

Work on the development of a nHDsC for Ireland is being carried out as part of the HealthData@IE project. The HealthData@IE project, led by the Department of Health in collaboration with the Health Information and Quality Authority (HIQA), the Health Research Board (HRB), Health Service Executive (HSE), and other key stakeholders, aims to establish the national infrastructure required for EHDS compliance specifically in relation to a HDAB. The HDAB, once established, will serve as the central national authority responsible for facilitating access to health data for secondary purposes, in full compliance with the legal, ethical, and technical requirements outlined in the EHDS regulation.

The HealthData@IE project is focusing on the development of national infrastructures needed for data access. The data access infrastructure includes systems identified by the EU Commission as core Digital Business Capabilities for HDABs:[9] Data Discovery, which enables users to locate and understand available datasets through the National Health Dataset Catalogue (nHDsC); Data Access Management, which handles requests, evaluations, and permissions for datasets through a Data Access Application Management System (DAAMS); and Secure Data Processing, which ensures that approved data is accessed in a controlled, privacy-preserving manner within a SPE. In addition, there are programmes of work in progress in the areas of data quality enhancement and dissemination, education and training.

Under the HealthData@IE project, the initial phase of developing Ireland's nHDsC focuses on enabling data users to discover, locate and understand available health datasets from across Irish data holders. This is being achieved through completing three key deliverables as part of the HealthData@IE project:

- D6.1 – A Requirements and Specifications Report;
- D6.2 – A Pilot Report;
- D6.3 – A Publicly available nHDsC.

The emphasis of this report is deliverable 6.1 – the development of requirements and specifications for the design of Ireland's nHDsC.

1.4.2 Current Catalogue of National Health and Social Care Data Collections

In 2010, HIQA published the first Catalogue of National Health and Social Care Data Collections, which was later updated in 2014 and 2022. The aim of this was to outline what health and social care data is available and also to ensure that data is discoverable for data users. National data collections are national repositories of routinely collected health and social care data, offering comprehensive national overviews of specific issues or services. The Catalogue uses a metadata template that is based on the Dublin Core Metadata Initiative's Metadata Terms and follows the principles of good information management as outlined in HIQA's National Standards for Information Management in Health and Social Care,[10] which were mandated by the Minister for Health in 2024.

The Catalogue has helped to establish consistent metadata practices across national health and social care data collections. Building on these foundations, the nHDsC will support the enhancement of Ireland's metadata infrastructure for health data. It will improve dataset discoverability for data users and support data holders in managing and publishing metadata aligned with European requirements and standards.

1.4.3 Methodology for the development of the nHDsC

The development of requirements and specifications for a nHDsC for Ireland has followed a phased approach:

1. A review of European-level initiatives including the Towards a European Health Data Space Joint Actions (TEHDAS and TEHDAS2) and the HealthData@EU Pilot to capture guidance and technical outputs relevant to nHDsC design.
2. A review of existing Irish catalogues — Catalogue of National Health and Social Care Data Collections, the Public Service Data Catalogue, and the Open Data Portal to identify strengths, gaps, and functional requirements.
3. Drafting requirements and specifications supported by a prototype to test functionality and usability.

To support this development process, a Working Group has been established. This group plays a key role in ensuring timely and successful project delivery by offering expert advice, reviewing project outputs, providing structured feedback, and facilitating stakeholder engagement throughout key stages of the project. This role is critical in ensuring that the final nHDsC meets both national needs and European obligations.

1.4.3.1 Review of European Initiatives: TEHDAS, TEHDAS2 and the HealthData@EU Pilot

Ireland's work on the nHDsC is closely aligned with progress at the European level in respect of the EHDS. This alignment ensures that the Irish catalogue will be interoperable with the European Central Platform and compliant with the EHDS Regulation.

The TEHDAS (Towards a European Health Data Space) Joint Action laid the groundwork for creating a European framework for the secondary use of health data.(11) This joint action aimed to help EU Member States and the European Commission to develop concepts and proposals to promote the secondary use of health data to benefit public health and health research and innovation in Europe. The recommendations from the TEHDAS project were used by the European Commission to inform the development of a proposal for regulation of the EHDS in May 2022.

Following on from the TEHDAS joint action, the HealthData@EU Pilot set out to build a pilot version of the EHDS infrastructure for the secondary use of health data, referred to as HealthData@EU. This project ran until December 2024 and developed a network infrastructure alongside services to support data users.(12) It also provided guidelines for data standards, quality, security and transfers to support the EHDS infrastructure. The pilot demonstrated how nHDsC can interconnect through the European infrastructure, providing practical insights into interoperability, dataset exchange, and cross-border discoverability. A key enabler is the HealthData@EU Central Platform, released by the European Commission as open-source software.(13) It provides reusable components, including a central metadata catalogue, which Member States can adopt to ensure interoperability with EU infrastructure.

In addition, a further EU joint action, TEHDAS2, commenced in 2024 with partners from 29 countries with the aim of developing common guidelines and technical specifications to facilitate secure access to health data and strengthen European collaboration in using data efficiently.(14) TEHDAS2 has refined HealthDCAT-AP, a metadata specification designed for health data discoverability and reuse. Draft guidelines outline how HealthDCAT-AP extends the EU's DCAT-AP standard to meet the specific needs of health datasets, with a focus on metadata harmonisation, quality, and interoperability. TEHDAS2 also published a technical specification for nHDsC, setting out minimum requirements for metadata input, management, export, and access, while allowing for national customisation.

These EU-level outputs and tools have been leveraged in the development of Ireland's catalogue to avoid duplication of effort, accelerate national development, and ensure

that the nHDsC supports cross-border data exchange and can be fully integrated into the HealthData@EU infrastructure.

1.4.3.2 Review of Existing Data Catalogues in Ireland

To support the development of the nHDsC, a review of existing catalogues in Ireland was undertaken. The aim was to identify current practices, assess how catalogues manage metadata, and identify opportunities for alignment with the nHDsC. The main catalogues reviewed are described briefly below:

- HIQA Catalogue of National Health and Social Care Data Collections[15]
  As described previously in section 1.4.2, this catalogue through its metadata offers a structured view of national health and social care datasets. As part of the review, HealthDCAT-AP fields were mapped against catalogue content. This mapping highlighted the areas of alignment and gaps. While the nHDsC, based on HealthDCAT-AP, represents a more technical and precise model, it follows the same principle as HIQA's catalogue to enable data users to discover available data.

- The Irish Public Service Data Catalogue[16]
  This catalogue provides an overview of data held across the public service in Ireland, including personal, business, spatial, and sensitive information critical to policy and service delivery. As part of the review, we engaged with the organisation managing this catalogue to understand the end-to-end dataset metadata lifecycle, including submission, validation and approval process, available features, and any Ireland-specific considerations relevant to the development of nHDsC.

- Ireland's Open Data Portal[17]
  The national open data portal provides access to non-personal governmental datasets, promoting transparency, innovation, and reuse. As per the Open Data Technical Framework, metadata is standardised by public bodies using the W3C Data Catalogue Vocabulary (DCAT), and more specifically, the DCAT Application Profile for European Data Portals (DCAT-AP). DCAT-AP was used to ensure interoperability between datasets published by public bodies, at both national and international levels.[18] This portal was analysed to better understand the process of dataset metadata addition by publishers,[19] the available functionalities, and to map HealthDCAT-AP fields against the catalogue content.

This review provided valuable insights into catalogue structures, user interfaces, data vocabularies, and helped to ensure that the development of the end-to-end dataset metadata lifecycle for the nHDsC has incorporated functionalities tailored to Ireland's specific data landscape and policy context.

1.4.3.3 Draft requirements and specifications and prototype

The draft requirements and specifications for the nHDsC were developed based on the findings from the review of European initiatives (Section 1.4.3.1) and the review of existing catalogues in Ireland (Section 1.4.3.2). These reviews provided the basis for defining the essential features, metadata standards, and interoperability requirements of the catalogue.

The initial draft of the requirements and specification was used to develop visuals of the catalogue application pages. These visuals served as an early prototype, providing a practical demonstration of how the requirements could be implemented. They were presented to Working Group members to gather feedback, enabling them to evaluate the proposed functionality and identify areas for refinement.

Based on the feedback received, the draft requirements and specifications were further updated, and a prototype was developed. This prototype was subsequently demonstrated to the Working Group members and feedback was sought. Prototype screenshots with functionalities can be viewed in Appendix 5. Following the second round of feedback, the requirements and specifications described in the following sections were finalised. This iterative process ensured that the design of the nHDsC was evidence-based, aligned with EU-level initiatives, and informed by the practical experience and expertise of Working Group members. The following sections are structured to provide a comprehensive overview of these requirements and specifications. Section 2 details the business, functional and non-functional requirements, and Section 3 details the technical specifications that underpin the system design.

2. National Health Dataset Catalogue – Business Requirements

The National Health Dataset Catalogue (nHDsC) must meet the needs of a wide range of stakeholders including Data Holders, Data Users and policymakers. The business requirements outlined in this section reflect the overarching goals of enabling secure, standardised, and accessible metadata publication and dataset discovery.

2.1 Introduction

This section describes the specific business capabilities nHDsC must support — such as creating, publishing, updating, and searching for dataset metadata. These are followed by the functional features needed to support these capabilities, and a set of non-functional requirements to ensure usability, security, performance, and compliance with standards.

The business and functional requirements focus on four core capabilities:

- **Metadata input:** enabling Data Holders to submit dataset descriptions;
- **Metadata management:** metadata shall be stored, validated and maintained to ensure consistency with EHDS legal requirements, helping Data Holders to meet their obligations efficiently. This includes the ability of Data Holders to update and improve dataset descriptions and to document their annual audits;
- **Metadata output**: dataset descriptions shall be made available as HealthDCAT-AP via user-friendly search tools, facilitating discoverability by researchers and policy-makers;
- **Metadata access**: a user-centric interface will enable stakeholders to easily locate datasets through a nHDsC.

2.2 Requirements Traceability Matrix

The Requirements Traceability Matrix (RTM) provides a structured means of tracking the alignment between business requirements, their corresponding functional and non-functional requirements, and relevant specifications. This ensures transparency, completeness, and consistency throughout the system development lifecycle.

The RTM helps ensure that all business requirements are fully addressed through clearly defined functional and non-functional requirements. It also validates that specifications are rooted in genuine business needs and makes it easier to assess the

impact of any changes by clearly illustrating the relationships and dependencies between requirements. In addition, the RTM supports quality assurance, testing, and verification efforts by establishing clear traceability from initial business objectives to final system implementation.

The RTM table includes the following elements:

- Business Requirement (BR): A unique identifier and description of the high-level business objective.
- Linked Functional Requirements (FRs): Specific system capabilities needed to satisfy the business requirement.
- Linked Non-Functional Requirements (NFRs): Constraints or quality characteristics (e.g., usability, performance, accessibility) necessary for successful implementation.
- Specifications/Notes: Key design considerations or implementation details related to the requirement.

Table 1 Requirements Traceability Matrix (RTM) linking Business requirements with Functional and Non-functional requirements

Business Requirement (BR)	Linked Functional Requirements (FRs)	Linked Non-Functional Requirements (NFRs)	Specifications/Notes
BR 01 – Metadata Record Creation	FR 02, FR 03, FR 04, FR 05, FR 06, FR 07	NFR 04, NFR 05, NFR 06, NFR 07, NFR 08, NFR 09	Metadata must follow HealthDCAT-AP. Validation for EHDS compliance. Editable metadata before publication.
BR 02 – Metadata Record Publication and Update	FR 08, FR 09, FR 010, FR 011, FR 012, FR 013, FR 014, FR 015	NFR 010, NFR 011, NFR 012, NFR 013, NFR 014, NFR 015	Detailed metadata including EHDS fields. Versioning integrity and audit support.
BR 03 – Searching for Datasets	FR 016, FR 017, FR 018, FR 019, FR 020, FR 021	NFR 016, NFR 017, NFR 018, NFR 019, NFR 020, NFR 021	Filtering and comparison use HealthDCAT-AP fields. Saved searches for registered users. URL sharing supported.

General – Applies across all BRs	FR 01, FR 022	NFR 01, NFR 02, NFR 03, NFR 022, NFR 023, NFR 024	Language toggle and feedback option. All UI and content must be in English and Irish. UI compatibility across various browsers and accessibility on a wide range of devices. Additional factors that should also be taken into account during application development.

The business, functional, and non-functional requirements are detailed in the following sections.

2.3 Business Requirements

These requirements serve as a foundation for the development of functional and non-functional requirements and specifications.

BR 01 Metadata Record Creation

As part of this requirement, Data Holders must provide initial dataset descriptions using the HealthDCAT-AP application profile through the nHDsC. Dataset descriptions must include: Key metadata elements, including dataset purpose, structure, format, and access conditions, information on any legal or ethical constraints, including anonymisation or pseudonymisation levels.

The submitted dataset descriptions are validated to ensure compliance with the EHDS regulation. Any inaccurate metadata will be returned to the Data Holder for revision or revoked to ensure the reliability of published datasets.

Functional Requirements linked to Metadata Record Creation

1. Allow for User identification, authentication, and authorisation, including the option to enforce multi-factor authentication (MFA)
2. Create, Read, Update and Delete (CRUD) dataset descriptions using the HealthDCAT-AP application profile
3. Add data quality and utility labels
4. Enforce minimum elements for dataset descriptions

5. Submit dataset description
6. Review dataset description and data quality and utility labels.

BR 02 Metadata Record Publication and Update

As part of this requirement, the dataset descriptions must be published in nHDsC after validation.
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 the dataset metadata as necessary to ensure accuracy and relevance, with any changes being validated prior to publication. To enhance data governance, the system supports automated scheduling of periodic reviews and audits, sending timely notifications to stakeholders through email and in-portal to ensure compliance with relevant guidelines and regulations. Additionally, the nHDsC maintains detailed audit trails for all data quality management activities, providing a robust framework for monitoring actions taken on the metadata of datasets.

Functional Requirements linked to Metadata Record publication and update

1. Publish dataset descriptions using the HealthDCAT-AP application profile
2. Update or delete dataset metadata descriptions and maintain change history
3. Maintain Versioning of updated dataset metadata
4. Enable Rollback to Previous Version of Dataset Description when needed (with Admin Approval)
5. Automate scheduling of periodic reviews and audits
6. Enable export of dataset audit trail entries to CSV or Excel formats
7. Notify stakeholders about upcoming reviews
8. Maintain detailed audit trails for all dataset quality management activities.

BR 03 Searching for datasets

As part of this requirement, nHDsC should offer a comprehensive search functionality that enables Data Users to effectively browse and display entries from national datasets catalogues.

Data Users 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 HealthDCAT elements, allowing them to refine search results effectively. Data Users can save their search queries for easy access, receive notifications through email and in-portal about updates on their saved

datasets, and share these searches with others to enhance collaboration. The Data User can save the search only if they have an account created in the nHDsC portal, and the search can be shared with any user by copying the URL generated for specific search. Additionally, the systems allow Data Users 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 metadata.

Functional Requirements linked to Searching for datasets

1. Offer a search functionality for browsing and displaying metadata catalogue results
2. Enable selection and export of dataset metadata to CSV or Excel formats
3. Allow saving and sharing relevant searches
4. Notify the Applicant about updates on saved datasets
5. Provide filtering and sorting capabilities by DCAT and HealthDCAT-AP elements
6. Allow to select and compare datasets.

2.4 Functional Requirements and Features

The functional requirements mentioned in this section are specific descriptions of the functions a system must be able to perform to support business requirements. Functional requirements are essential in guiding the system's design, development, and testing processes, ensuring it delivers the necessary functionality. Additionally, the features mentioned in this section are individual high-level capabilities or attributes to achieve specific functional requirements. Each feature addresses a set of user needs or functional requirements, bridging the business requirements with concrete, user-facing functionalities within the system.

FR 01 Multilingual Support: English and Irish

i) Domain

User Interface/System Accessibility

ii) Requirement summary

The system must support both English and Irish languages across all user-facing content and application interfaces. Users must have the ability to select their preferred language at any time during their interaction with the application. (The specific content to be translated into Irish is yet to be confirmed).

iii) Actors

- Data Holders (Editor/Admin)
- System Admin
- Data Users

iv) Features

nHDsC must:

- Provide full bilingual support (English and Irish) for all UI components, labels, navigation, messages, metadata fields, and system-generated content.
- Allow users to switch between English and Irish using a clearly accessible language selector.
- Remember the user's language preference during their session or across visits, depending on login state.
- Ensure translated content is synchronised and equivalent in meaning and function.

v) Detailed Description

System functionalities include:

1. User Interface and Navigation Elements
   - All interface elements including menus, buttons, titles, tooltips, and error messages must be available in both English and Irish.
   - Static pages (e.g., homepage, about, help, contact) must be provided in both languages.
2. Dataset Metadata Fields and Labels
   - Field names, section headings, and controlled vocabularies visible in the application must be displayed in the selected language.
   - Where applicable, dataset descriptions added by Data Holders can support bilingual entry (e.g., title\en, title\ga).
3. Language Selection Mechanism
   - A language selector must be available on all public and authenticated pages, preferably placed in the application header.
   - The selector must offer options for:
     - English (EN)
     - Irish (GA).
4. Persistence of Language Choice

- For logged-in users, the system must store and remember language preferences between sessions.
- For anonymous users, language preference must persist throughout the session using browser cookies or session storage.

5. Compliance and Quality
   - All translations must be accurate, context-sensitive, and reviewed for compliance with public sector language accessibility policies in Ireland.
   - System messages, email and in-portal notifications, and alerts must be displayed in the user-selected language.
6. Fallback Behaviour
   - If a translation is temporarily unavailable, the system should display the English version by default and log the fallback for resolution.

vi) Systems Involved

- nHDsC

FR 02 Allow for User identification, authentication, and authorisation, including the option to enforce multi-factor authentication (MFA)

i) Domain

Datasets Access Applications and Requests

ii) Requirement Summary

The system must allow for Data Holder and user identification, authentication, and authorisation, including the option to enforce MFA, and support account lifecycle management including deactivation and account reassignment.

iii) Actors

- Data Holders (Editor/Admin)
- System Admin
- Data Users.

Note: All functionalities outlined in FR016 , FR017 and FR020 to FR022 shall remain accessible to users without logging into the application. Data Users shall only be required to log in to enable specific features such as saving search criteria (FR018), enabling user notifications (FR019), and adding dataset distributions to the basket (FR022).

iv) Features

nHDsC must:

- Allow log in
- Incorporate some identification, authentication and authorisation (IAA) mechanisms
- Support account deactivation/removal workflows with dataset ownership transfer
- Provide role and organisation-based access control.

v) Detailed Description

The systems must include processes for identification, authentication, and authorisation for Data Holders to provide electronic health data for secondary use (provide datasets).

System functionalities include:

1. Mechanisms for identification, authentication and authorisation:
   - nHDsC use identification and authentication (e.g. MFA).
   - nHDsC use Authorisation mechanisms to validate the user is related to the organisation obliged to create the dataset.
     - Establish a role-based access control (RBAC) system to assign permissions based on user roles (e.g., Data Holder Admin, Data Holder Editor, Data User and System Admin).
     - Implement organisation-specific access controls to ensure users are related to the relevant organisation.
     - Maintain an up-to-date user directory that maps users to their respective organisations, including roles and privileges.

2. Account Lifecycle Management:

User-Initiated Deactivation:

- Data Holders Admin/Data Holder Editor/Data User may request account deactivation via their account settings.
- System Admins can approve deactivation requests.
- While raising the deactivation request for Data Holder (Admin/Editor):
  - The requester must select a currently active Data Holder within the same organisation to whom their datasets will be reassigned.
  - The reassignment field is mandatory.

    - The system must validate that the selected transferee is eligible and active.
- While raising the deactivation request for Data User, they must fill in the reason for deactivation and submit the request.
- Upon approval:
    - The user's account is marked as "Deactivated".
    - All datasets are reassigned to the selected Data Holder (applicable for Data Holder deactivation).
    - Both the user and the transferee receive system and email notifications.

Data Holder Admin-Initiated Deactivation (On Behalf of Others):

- Data Holder Admins may submit deactivation requests on behalf of Data Holder Editors within their organisation.
- Requests must be approved by the System Admin.

System Admin-Initiated Removal:

- System Admins must be able to remove (terminate) user accounts, including inactive or unauthorised users
- Before removal:
    - System Admins may directly remove a user account.
    - The system must require reassignment of datasets before removal can be confirmed.
    - All actions are logged with the Actor (who performed the removal), Timestamp, Target user, Dataset transferee.

Dataset Reassignment Process:

- Dataset ownership transfer must update all metadata records and access permissions accordingly
- The transferee must receive a system notification and email confirming their new dataset responsibilities.
- The original user (if self-deactivating) and System Admin must receive a confirmation of deactivation and dataset reassignment.

3. Log-In Behaviour by Role:

- Data Holder Editors are limited to dataset creation and editing.
- Data Holder Admins have full control over dataset submission, user management within the organisation, and deactivation requests. One organisation can have multiple Data Holder Admins and Data Holder Editors.

- System Admins must authenticate with elevated privileges and have access to user management, audit logs, and approval workflows.
- Data Users must authenticate to search for datasets and manage their saved search profiles.

vi) Systems Involved

- nHDsC.

FR 03 Create, Read, Update and Delete (CRUD) dataset descriptions using the HealthDCAT-AP application profile

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must provide the functionality to Create, Read, Update, and Delete (CRUD) dataset descriptions using the HealthDCAT application profile

iii) Actors

- Data Holders Admin
- Data Holder Editor

iv) Features

nHDsC must:

- Provide text edition functionalities to complete the dataset description.
- Provide automated mechanisms to check quality and conformity of dataset descriptions (including validation tools to verify compliance with HealthDCAT before submission).

v) Detailed Description

Health data holders must submit the dataset descriptions in nHDsC for categories outlined in Article 51 of the EHDS. To support compliance with making specific data categories available for secondary use, the nHDsC provide the following functionalities.

System functionalities include:

1. Create New Dataset description: A Create a new dataset description option is clearly available on a landing page. There are two ways to do this:
   - Create a copy from a submitted dataset description either by selecting specific fields they want to copy or all the fields.
   - Create a new dataset description from scratch.
2. Guidance for data holders
   - A wizard tool guides data holders through a user-friendly process to determine their obligations under the EHDS Regulation.
     - 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.
3. 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-AP to ensure consistency and interoperability), 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 FR 05 Enforce minimum elements for dataset descriptions).
4. Dataset Description Form:
   - Includes mandatory, recommended and optional fields based on the type of dataset Access Rights (i.e Open Data, Protected Data and Sensitive Data) In HealthDCAT-AP, the sensitivity of a health dataset is defined by its relationship with the EHDS Regulation, categorising datasets as Open, Protected, or Sensitive. This classification determines the level of detail required in the metadata and the necessary access controls, with sensitive datasets (containing personal data) necessitating more extensive metadata, such as data dictionaries or samples, to ensure secure reuse of health data. The health dataset's sensitivity level dictates the mandatory, recommended, and optional metadata fields required for describing the dataset. A detailed breakdown of these fields is provided in Appendix 6.
   - 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.

5. 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 Editor has finalised with the dataset description, they can choose the Data Holder Approver name from dropdown and select the "Complete" option/feature that will trigger a process for submission (meaning it goes to the Data Holder Admin for submission).
When the Data Holder Admin has finalised with the dataset description, they can select "Submit" option/feature that will trigger a process for System Admin review.

vi) Systems Involved

- nHDsC

FR 04 Add data quality and utility labels

i) Domain

Datasets catalogue

ii) Requirement Summary

Systems must enable the addition of data quality and utility labels

iii) Actors

- Data Holder Admin
- Data Holder Editor

iv) Features

nHDsC must:

- Allow data holders to add data quality and utility labels to their dataset description
- Allow data holders to upload documentation supporting data quality and utility labels.

v) Detailed Description

In the nHDsC, an option to "Add data quality and utility labels" is available during the dataset description creation process, prior to submission for review. This option redirects users to a dedicated space for managing Data Quality and Utility Labels. These data quality labels refer to the characteristics or indicators that help assess and communicate the quality of the data contained in a dataset.

These quality labels will help to indicate:

- 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.

System functionalities include:

- Data Holder creates a dataset description in the nHDsC Portal.
- The Data Holder, while creating the dataset description, has an option to add data quality and utility labels.
- Once the Data Holder clicks on add data quality labels, the nHDsC Portal will present a form so that the Data Holder can fill-in the requested information regarding the quality labels.
- After adding the quality labels, the Data Holder will be able to save the form.
- Dataset description will have all the quality labels added by the Data Holder in the nHDsC Portal.

Note: The European Commission may establish the visual characteristics of the label in the future. The Data Quality and Utility Label (DQUL) will be created on a separate dedicated page (either within nHDsC or elsewhere), and data users will be able to import the DQUL information in a DCAT-AP compatible format.
As per the EHDS Regulation, labelling will be mandatory for datasets for which data collection was publicly funded (EU or national funding). This covers datasets for which the funding was specifically for the collection. For other datasets, providing a data quality and utility label is optional. For example, where public funding is used to set up a registry for research purposes, the label will be mandatory. Where a hospital receives public funding for providing treatment, the label will be optional, as the funding is for providing treatment, and the registration of data is incidental to this task.

nHDsC 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.

vi) Systems Involved

- nHDsC

FR 05 Enforce minimum elements for dataset descriptions

i) Domain

Datasets catalogue

ii) Requirement Summary

Systems must enforce minimum elements for dataset descriptions (format)

iii) Actors

- Data Holder Admin
- Data Holder Editor

iv) Features

nHDsC must:

- 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.

v) Detailed Description

Before submitting a dataset description, the nHDsC will include a format check procedure which will perform an assessment to ensure compliance with minimum common information elements and cross-border data standards.

System functionalities include in the nHDsC:

1. Automatic Control Checks:
   - Automated checks ensure metadata descriptions are valid and complete as well as HealthDCAT-AP compliance.
2. 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.

vi) Systems Involved

- nHDsC Portal

FR 06 Submit dataset description

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must allow to submit dataset description

iii) Actors

- System Admin
- Data Holder Admin

iv) Features

nHDsC Portal must:

- Provide a functionality to submit the dataset description.
- Provide a dashboard for data holders with the dataset descriptions that data holders have created.

v) Detailed Description

Once the dataset description is completed/filled in by Data Holder Editor, the nHDsC allow Data Holder Admin to submit the dataset description through a "Submit for Review" option/functionality, triggering the following process:

1. Submission Confirmation:
   - The nHDsC prompts a confirmation that the dataset description is submitted
2. Submission Notification:
   - Data Holder Editor and Admin receive a notification through email and in-portal that a dataset description and data quality and utility label have been submitted.
   - System admin receives a notification through email and in-portal that a dataset description and data quality and utility label has been created.
3. Data holder dashboard:
   - Each data holder has access to a dashboard displaying the status of their submitted dataset as "Pending Review".

vi) Systems Involved

- nHDsC

FR 07 Review dataset description and data quality and utility labels

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must allow to review dataset descriptions and data quality and utility labels

iii) Actors

- Data Holder Editor
- Data Holder Admin
- System Admin

iv) Features

nHDsC must:

- Allow to review metadata completeness, data quality, and utility labels added by data holders.

- Automatically notify the Data Holder Editor and Admin in case of the revocation of the data quality and utility label, including the justification.

v) Detailed Description

Once the data holders have provided a data quality and utility label, the nHDsC include a clear process that allows System admin to review dataset descriptions and data quality and utility labels (and possibly revoke them if not applicable).

1. New dataset description notification:
   - System admin receives a notification through email and in-portal that a dataset description and data quality and utility label should be reviewed.
2. Review:
   - System admin accesses the nHDsC following the notification.
   - System admin accesses a dashboard in the nHDsC with datasets descriptions for review (e.g. similar to a "Pending review"-inbox).
   - System admin opens the datasets descriptions (metadata completeness), data quality, and utility labels.
   - The nHDsC displays the dataset description and data quality and utility labels.
3. Revocation Option:
   - The portals provide a revocation option/feature/functionality for system admin to revoke descriptions and data quality and utility labels.
4. Justification Prompt:
   - Upon selecting the revocation option/feature/functionality, the portal prompts a free text field for providing a justification.
   - The system admin must enter a justification for revocation.
5. Revocation Process:
   - Once the justification is submitted, the system processes the revocation request.
   - The descriptions and data quality and utility labels is systematically marked as revoked, rendering it invalid.
   - The status of the dataset descriptions and data quality and utility labels in the database is updated to indicate it has been revoked.
6. Automatic Notification:
   - After revocation, the system automatically notifies the Data Holder Editor and Admin.
   - The notification through email and in-portal includes the justification for the revocation.

vi) Systems Involved

- nHDsC

FR 08 Publish dataset descriptions using the HealthDCAT-AP application profile

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must allow for the publication of dataset descriptions using the HealthDCAT application profile

iii) Actors

- System Admin
- Data Holder Editor
- Data Holder Admin

iv) Features

nHDsC must:

- Publish the contents of the applicable dataset descriptions.
- Provide an automated mechanism to associate a code with each dataset for identification and information purposes.

v) Detailed Description

After validation, the dataset descriptions are published in the nHDsC. This provides information about existing datasets to data users.

1. Publish:
   - If the System admin establishes the dataset description is complete and accurate, then they will click on the "Submit for publication" feature available.
2. Publish Confirmation:
   - The nHDsC prompts a confirmation that the dataset description is approved.
   - The nHDsC publish dataset descriptions after validation of the content and labels in the national dataset catalogue (publicly available and standardised machine-readable dataset catalogue).
3. Publish Notification:

- Data Holder Editor and Admin receive email and in-portal notification that a dataset description and data quality and utility label have been published.

4. Data holder dashboard:
   - Each Data Holder Admin and Editor has access to a dashboard displaying the status of their published dataset descriptions as "Published".
   - The dataset description is then removed from the "Pending review" dashboard from the nHDsC.
5. Automatic datasets identification processes:
   - The nHDsCs generate a unique version numbering for each new dataset description, following a code system that may include references to the country, Admin, data holder, and a sequential number for published datasets.
6. Public Availability:
   - The national datasets catalogue is made publicly available through an interface in the nHDsC (Note: datasets with trade secrets might not be displayed for public view).
7. 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.

vi) Systems Involved

- nHDsC

FR 09 Update or Delete dataset metadata descriptions and maintain change history

i) Domain

Datasets Catalogue.

ii) Requirement Summary

Systems must allow for updating dataset metadata descriptions and maintain change history.

iii) Actors

- Data Holder Editor
- Data Holder Admin
- System Admin

iv) Features

nHDsC must:

- Provide an interface to access their dataset descriptions
- Provide text-edition functionalities to update "published" datasets.

v) Detailed Description

If a dataset description requires updating or editing, Data Holder Admins and Editors can easily update it through the nHDsC.

System functionalities include:

1. 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.
2. Editing and Updating:
   - After identifying the dataset, the data holder selects an Edit function to modify metadata and content
   - This triggers the process described in requirements FR 06, FR 07 and FR 08 of this domain.
3. Delete metadata dataset
   - After identifying the dataset, the Data Holder Editor uses the Delete function to select the Data Holder Admin to whom the deletion approval request will be sent.
   - The Data Holder Admin reviews the request and submits it to the System Admin for final approval.
   - Once the System Admin reviews and approves the request, the dataset metadata is soft deleted.
   - Deleted dataset metadata remains visible to the Data Holder Admin and Editor on their dashboards, with the status marked as "Deleted", but not visible to the Data Users.
   - Admin-only functionality may allow full purging if required for compliance.
4. Versioning Process:
   - A new version of the dataset is automatically generated, leaving the previous version intact.

- 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.

vi) Systems Involved

- nHDsC

FR 010 Maintain Versioning of updated Dataset metadata

i) Domain

Datasets Catalogue.

ii) Requirement Summary

The system must maintain a comprehensive version history for all dataset descriptions, ensuring that updates are automatically versioned, traceable, and accessible for review.

iii) Actors

- Data Holder Editor
- Data Holder Admin
- System Admin

iv) Features

nHDsC must:

- Automatically create a new version upon each dataset description update
- Display a complete, structured version history for each dataset
- Provide comparison functionality between current and previous versions
- Support change traceability for quality assurance and audit, including an option to view the specific changes made in each version.

v) Detailed Description

The system must track all changes made to dataset descriptions by automatically generating a new version each time a description is updated. Versioning features must include:
Version History View

- Each dataset has an associated version history viewable in the dashboard

- The version history includes the data holder name who has made the update, version numbers, dates of creation, change summaries and status (e.g., current, deprecated, and showing which version is currently used).

Version Comparison

- Data holders can view metadata changes between current and previous versions using a side-by-side view, and the changes are highlighted.

Traceability and Integrity

- Every version must be assigned a unique ID.
- System must ensure consistent metadata relationships and validate integrity upon each update.
- All version updates are logged for audit purposes.

vi) Systems Involved

- nHDsC

FR 011 Enable Rollback to Previous Version of Dataset Description when needed (with Admin Approval)

i) Domain

Datasets Catalogue.

ii) Requirement Summary

The system must allow data holders to request rollback to a previous version of a dataset description, subject to System Admin approval. The rollback must result in a newly created version with a new identifier, clearly indicating which version it was copied from.

iii) Actors

- Data Holder Editor
- Data Holder Admin
- System Admin

iv) Features

nHDsC must:

- Allow selection of a previous version for rollback
- Require justification from the Data Holder Editor/Admin initiating the rollback
- If Data Holder Editor is initiating the roll back, it will request them to select the Data Holder Admin to whom the rollback approval request should be sent

- The Data Holder Admin reviews the request and forwards it to the System Admin for final approval
- Upon approval, create a new version with a unique identifier, displayed as a copy of the selected version
- Record and display rollback metadata in the version history
- Validate metadata integrity before rollback is finalised.

v) Detailed Description

Rollback functionality is essential for correcting errors or reverting to previously valid dataset descriptions while maintaining transparency, control, and auditability.

Initiation by Data Holders

- Data Holder Admins/Editors can request to roll back to a specific prior version using a rollback interface available in the version history view
- A justification field must be completed before submission.

Approval Workflow

- System Admins are notified of pending rollback requests
- System Admins can view version details, comparison views, and the justification before approving or rejecting
- Only upon Admin approval is the rollback executed.

Rollback Execution

- The rollback creates a new version with a new unique identifier
- In the version history UI, the new version will be displayed as: Version 8 (Copy of Version 5) — where Version 8 is the new ID, and Version 5 is the source of the rollback
- The version content will mirror the selected older version, but the rollback is treated as a forward-moving operation (i.e., no deletion or overwriting occurs).

Metadata Integrity and Validation

- Before finalising rollback, the system validates all linked metadata, vocabularies, and documentation references
- If inconsistencies or deprecated links are detected, the system issues warnings or blocks the rollback with explanatory messages.

Audit Logging

- The rollback process, including requester, approver, timestamp, original version, restored version, and rollback reason, is fully logged.
- Logs are maintained in a secure, read-only audit log for governance and traceability.

vi) Systems Involved

nHDsC

FR 012 Automate scheduling of periodic reviews and audit

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must support automated scheduling of periodic reviews and audits for data holders

iii) Actors

- System Admin
- Data Holder Editor
- Data Holder Admin

iv) Features

nHDsC must:

- 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 for the next periodic review and audit (annually since the last audit).

v) Detailed Description

The nHDsC clearly display the status of each dataset in their catalogues, indicating statuses such as "Published", “Draft”, “Deleted”, “Pending Review” and “Revoked”".

System functionalities include:

1. Status Display:
   - The system displays the current standing of each dataset, such as "Published", “Draft”, “Deleted”, “Pending Review” and “Revoked”.
2. Time-Tracking Mechanism:
   - Displays the date of the next scheduled review or audit of the dataset by the Data Holder Admin and Editor

- 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).

vi) Systems Involved

- nHDsC

FR 013 Notify stakeholders about upcoming reviews

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must notify data holders about upcoming audits and reviews

iii) Actors

- System Admin
- Data Holder Editor
- Data Holder Admin

iv) Features

nHDsC must:

- Email and in-portal notifications to the data holders about upcoming reviews, audit results and required actions

v) Detailed Description

The nHDsC should automatically notify Data Holder Editor and Data Holder Admin (i.e. the approver of dataset metadata) of any scheduled reviews or periodic audits.

System functionalities include:

1. Automatic Notifications:
   - The system automatically sends email and in-portal notifications to Data Holder Editor and Data Holder Admin (i.e. the approver of dataset metadata) about scheduled reviews or periodic audits.
   - The system automatically sends email and in-portal notifications to the System Admin about scheduled reviews or periodic audits and overdue reviews.
2. 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 within the system.

vi) Systems Involved

- nHDsC

FR 014 Maintain detailed audit trails for all dataset quality management activities

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must maintain detailed audit trails for all dataset quality management activities.

iii) Actors

- System Admin
- Data Holder Editor
- Data Holder Admin

iv) Features

nHDsC must:

- Maintain an audit trail of the whole lifecycle of dataset submitted

v) Detailed Description

nHDsC should capture and log metadata related to:

System functionalities include:

1. Datasets Submitted:
   - Captures and logs dataset ID, dataset name, submission date, data holder's name, and associated data quality and utility labels.
2. Datasets validated:
   - Captures and logs automated validation processes (including timestamp) and results of the validation (e.g. validated, not-validated, corrected, etc.).

3. Datasets Published:
   - Captures and logs dataset ID, dataset name, publication date, data holder's name, and the final data quality and utility labels applied.
4. 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").

vi) Systems Involved

- nHDsC

FR 015 Enable export of dataset audit trail entries to CSV or Excel formats

i) Domain

Datasets Catalogue

ii) Requirement Summary

System must allow Data Holders and System Admins to export selected or filtered dataset audit trail entries from the audit dashboard in CSV or Excel format.

iii) Actors

- System Admin
- Data Holder Admin
- Data Holder Editor

iv) Features

nHDsC must:

- Provide an option "Audit logs" to view the Audit dashboard
- Provide an export function on the audit dashboard interface
- Support exporting all entries, or only those currently visible via filters or selections
- Offer file format selection: CSV and Excel (XLSX)
- Ensure the exported data includes all visible columns: Dataset ID, Description, Status, Creation Date, Last Modified Date, Previous Review Date, and Next Review Date.
- Preserve applied filters (e.g., review status, next review date range) in the exported results.

v) Detailed Description

From the Audit Logs Dashboard, Data Holder (Editor/Admin) and System Admin must be able to:

- View audit records in a paginated and filterable table.
- Apply filters (e.g., status, review date) or manually select individual entries.
- Click an "Export" button to initiate file download:
  - System must prompt for file format (CSV or Excel).
  - Export must reflect the currently displayed or selected records.
- Downloaded file must:
  - Include column headers identical to those shown on-screen
  - Follow DD/MM/YYYY date format
  - Be UTF-8 encoded
  - Reflect sorting applied by the user (e.g., Next Review Date ascending).
- Additional System Behaviours:
- The Data Holder Editor must be able to view and download the audit logs of datasets they have created, updated, or deleted.
- The Data Holder Admin must be able to view and download the audit logs of all datasets created within their organisation.
- The System Admin must be able to view and download the audit logs of all datasets.
- If no results are available or selected, export button must be disabled or trigger a prompt: "No records selected or available for export."
- If the export process fails (e.g., timeout, file generation issue), a clear, non-technical message must appear with retry guidance.
- Export action must not interfere with ongoing dashboard interactions or data integrity.

vi) Systems Involved

nHDsC

FR 016 Offer a search functionality for browsing and displaying metadata catalogue results

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must offer a search functionality for browsing and displaying metadata catalogue results.

iii) Actors

- Data Holders (Admin and Editor)
- Data Users
- System Admin

iv) Features

nHDsC must:

- Provide a search function which enables browsing by data Users
- Provide automated mechanisms to correct and adjust elements of the search where necessary
- Provide automated mechanisms to detect and alert if no results are found from the search.

v) Detailed Description

Data Users must be able to discover existing datasets through a search engine available on nHDsC.

System functionalities include:

1. Accessing the Portal:
   - Open an internet browser to access the nHDsC.
2. Browsing the Catalogue:
   - Browse the entire metadata catalogue, displaying all entries.
3. Search Functionality:
   - Search using an open text field.
     - Propose commonly used search queries.
   - Create personalised search functions based on predefined filters.
4. Dataset Selection:
   - Display query results on the applicant interface.
   - Select a dataset and display its description.

Once the data user launches the search, the system functionalities include:

1. Matching and Results:
   - Match dataset descriptions to query results.

- Perform corrections, adjustments, and cancellations of the search.

2. 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.

vi) Systems Involved

- nHDsC

FR 017 Enable selection and export of dataset metadata to CSV or Excel formats

i) Domain

Datasets Catalogue

ii) Requirement Summary

System must allow data users to select one or more datasets from the metadata catalogue and export the corresponding metadata in CSV or Excel format.

iii) Actors

- Data Holders (Admin and Editor)
- Data Users
- System Admin

iv) Features

nHDsC must:

- Provide the ability for users to select one or multiple datasets from search results.
- Offer export options in commonly used formats such as CSV and Excel (XLSX)
- Ensure that exported metadata includes all core HealthDCAT-AP fields and relevant metadata extensions where applicable
- Provide the option to download the file directly through the browser.

v) Detailed Description

After performing a search and reviewing datasets in the metadata catalogue, data users must be able to:

- Select one or more datasets from the search results
- Click an "Export" action that allows selection of the desired format (CSV or Excel)
- Trigger the system to compile and generate a file that includes:
  Dataset title, description, publisher, coverage, update frequency, access conditions, DQUL labels, and any other standard metadata fields
- Automatically download the file to the applicant's device via the browser.

Additional System functionalities include:

- If no datasets are selected, the "Export" option is disabled or prompts the user to make a selection.
- If the export operation fails, the system must display a clear error message with recommended corrective action.
- The exported file should be UTF-8 encoded and contain headers in accordance with the HealthDCAT-AP schema.

vi) Systems Involved

- nHDsC.

FR 018 Allow saving and sharing relevant searches

i) Domain

Datasets Catalogue.

ii) Requirement Summary

Systems must allow saving and sharing relevant searches.

iii) Actors

- Data Users
- System Admin
- Data Holder (Editor and Admin).

iv) Features

nHDsC must:

- Allow to save search criteria and datasets descriptions by data users
- Provide a view of the search criteria and dataset descriptions saved in the Personal profile of the data user in the portals

- Generate a link to the search or descriptions in the portals to be copied-and-pasted and shared.

v) Detailed Description

The nHDsC allow data users to share relevant search results, criteria and dataset descriptions with colleagues or stakeholders, facilitating collaboration and reuse of search strategies.

System functionalities include:

1. Saving Search Criteria:
   - Applicants can save search criteria, including filters applied, keywords used, and search settings, through a "Save my search criteria" functionality.
     - A pop-up or menu allows applicants to name the search, add descriptions or notes, and choose save options such as receiving alerts for new or updated datasets matching their criteria.
   - Applicants can save dataset descriptions through a "Save this dataset description" functionality.
     - A pop-up or menu allows applicants to name the dataset description, add descriptions or notes, and choose save options such as receiving alerts for updates in the dataset description.
2. Managing Saved Searches:
   - Saved searches, dataset descriptions are stored in the applicant's profile.
     - Applicants can re-execute saved searches with a single click to retrieve updated results.
     - Applicants can adjust or refine saved searches by modifying filters or keywords and saving the new version.
     - Applicants can remove searches and dataset descriptions that are no longer needed.
3. Sharing Search Results:
   - Applicants can select relevant search results and dataset descriptions and share them using a "Share Search" or "Share dataset description" functionality on the results page
     - The system generates a shareable link
     - Applicants can send via email or other means to selected recipients.

vi) Systems Involved

- nHDsC

FR 019 Notify the user about updates on saved datasets

i) Domain

Datasets Catalogue

ii) Requirement Summary

Systems must notify the applicant about updates on saved datasets

iii) Actors

- Data User
- Data Holder (Editor and Admin)
- System Admin

iv) Features

nHDsC must:

- Provide the options to create alerts for updates on search results and datasets descriptions in a standardised, computer-readable format.

v) Detailed Description

The data user can chose to obtain alerts whenever new datasets matching the saved search criteria are added or when saved datasets descriptions are updated.

System functionalities include:

1. Alert Subscriptions:
   - The system allows the user to create alert subscription tied to the applicant's saved search and datasets descriptions.
     - The user must have option to select whether they want to be notified about new entries or updates
     - The applicants must be able to select to receive notifications (e.g. via email) if new entries or updates
     - Applicants must be able to opt to receive notifications directly within the portal.
2. Portal Notifications:
   - Upon logging into the Portal, a notification pop-up indicates updates relevant to their saved searches or datasets descriptions.
   - Users can also access updates from their profiles, which features a list of all saved search and datasets descriptions alerts and relevant system messages.

3. Notification Details:
   - Notifications include:
     - New dataset additions matching the search criteria
     - Updates to existing datasets, such as improvements in data quality, added metadata, or new data entries.

vi) Systems Involved

- nHDsC

FR 020 Provide filtering and sorting capabilities by DCAT and HealthDCAT-AP elements

i) Domain

Datasets Catalogue.

ii) Requirement Summary

Systems must provide filtering and sorting capabilities (including HealthDCAT AP elements)

iii) Actors

- Data User
- Data Holder (Editor and Admin)
- System Admin

iv) Features

nHDsC must:

- Provide filtering and sorting capabilities alongside the search function.

v) 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 characteristics, a wizard search tool, and queries using Health DCAT-AP elements. Data users can combine these filtering options to identify relevant datasets that meet their needs.

System functionalities include:

1. Basic Filtering:
   - Provides filters based on dataset characteristics, allowing selection according to criteria such as data holder, keyword, theme, scope, time, geographical coverage, provenance, access, format, data quality and utility label, and AI-suitable data.
2. Wizard Search Tool:
   - Features an applicant interface with a dialogue box that guides applicants through key questions and filters. The tool presents search results based on the answers provided and can be combined with basic filters for refined results.
3. Queries using Health DCAT-AP elements:
   - Features an 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.).

vi) Systems Involved

- nHDsC

FR 021 Allow the user to select and compare datasets

i) Domain

Datasets Catalogue.

ii) Requirement Summary

Systems must allow the user to select and compare datasets.

iii) Actors

- Data User
- Data Holder (Editor and Admin)
- System Admin

iv) Features

nHDsC must:

- Provide a functionality to select one or more datasets
- Display two datasets in the same screen.

v) Detailed Description

While completing a search process in the nHDsC, users can select one or multiple dataset descriptions for further analysis. To enhance this process, a "Compare" functionality is available, enabling applicants to view and evaluate differences between selected datasets.

System functionalities include:

1. Compare Functionality:
   - Applicants can select datasets and use a "Compare" functionality to display the dataset descriptions and quality and utility labels side-by-side on a single interface.
2. Highlighted Differences:
   - Discrepancies or variations between datasets are automatically highlighted in a prominent colour, ensuring immediate visibility.
3. Consistent Fields:
   - Fields where datasets share the same values or descriptions remain in standard text with no additional highlighting, improving readability.
4. Data Quality and Utility Label Comparison:
   - Special emphasis is placed on comparing data quality and utility labels, highlighting variations in data accuracy, completeness, relevance and so on (12 dimensions of Quality and Utility label along with overall rating).[20]
5. Comparison Summary Section:
   - A summary section at the top or bottom of the comparison webpage lists the key differences between the datasets, providing applicants with a quick, at-a-glance overview before diving into the detailed comparison.

vi) Systems Involved

- nHDsC

FR 022 Provide Feedback Functionality

i) Domain

Datasets Catalogue.

ii) Requirement Summary

System must allow users to provide feedback directly through the catalogue application.

iii) Actors

- Data User
- Data Holder Editor
- Data Holder Admin

iv) Features

nHDsC must:

- Provide a visible "Feedback" option accessible from the main interface
- Allow users to enter their Name and Email ID
- Allow users to select a Feedback Type from a predefined dropdown list (e.g., General Feedback, Technical Issue, Metadata Quality, Access Request, About revoked dataset)
- Provide a text box for users to enter their feedback content
- Submit the feedback securely to the system for review and follow-up.

v) Detailed Description

To encourage continuous improvement of the nHDsC and foster user engagement, a feedback functionality is integrated into the catalogue interface.

System functionalities include:

1. Feedback Access:
   - A "Feedback" button is prominently displayed in the navigation menu or footer across all relevant pages of the nHDsC.
2. Feedback Form Fields:
   - The system presents a structured form requiring users to provide:
     - Name (optional field)
     - Email ID (mandatory field for follow-up communication)

  - Feedback Type (dropdown with predefined categories)
  - Feedback Content (multi-line text box for detailed comments).
3. Validation and Submission:
   - Mandatory fields are validated before submission.
   - On successful submission, the user receives a confirmation message that their feedback has been received.
4. Storage and Accessibility:
   - Feedback entries are stored securely in the system database.
   - System administrators can access, review, and respond to feedback through an internal management interface.
5. User Confirmation and Acknowledgment:
   - Acknowledgment is sent to the email provided by the user, confirming receipt of feedback.

vi) Systems Involved

- nHDsC

FR 023 Add Dataset Distributions to Basket [†]

i) Domain

Datasets Catalogue.

ii) Requirement Summary

The system must allow Data Users to add selected Dataset Record Distributions to a Dataset Basket, review their selections, and proceed to request access through the Data Access Application Management System (DAAMS).

iii) Actors

- Data Users

iv) Features

nHDsC must:

- Provide a clearly visible "Add to Basket" option for each Dataset Record Distribution.

[†] This functionality will be added once details about requirements and specifications for DAAMS have been confirmed – to be developed in tandem with Work Package 5 during 2026.

- Allow users to select multiple Distributions across different Datasets and add them collectively to their Dataset Basket.
- Visually indicate the total number of Distributions currently added to the Basket via a numerical counter displayed beside the Basket icon in the main navigation bar.
- Group Distributions by their parent Dataset Record within the Basket view.
- Allow users to open the Dataset Basket to view all selected Distributions, confirm selections and Click "Proceed" to request access.
- Redirect users to the DAAMS portal upon clicking "Proceed," to complete the access request process.

v) Detailed Description

This functionality supports the streamlined selection and management of Dataset Record Distributions by Data Users prior to requesting access through DAAMS.

System functionalities include:

1. Add to Basket Interaction:
   - Each Dataset Distribution displayed within the catalogue must have an "Add to Basket" button.
   - When selected, the corresponding Distribution is added to the user's Dataset Basket.
   - The system must display a brief confirmation notification (e.g., "Distribution added to Basket").
2. Basket Overview and Grouping:
   - The Dataset Basket page must group all added Distributions under their respective parent Dataset Record.
   - Users can view all items currently in their Basket and remove individual entries if needed.
   - The total number of selected Distributions must be dynamically updated and displayed as a numerical badge beside the Basket icon.
3. Confirmation and Proceed Action:
   - Users can review their selected Distributions within the Basket.
   - By clicking "Proceed", users confirm their selections and are automatically redirected to the DAAMS portal for completing their access request.
   - During this process, the system must ensure session continuity and transfer relevant metadata to DAAMS through secure integration.
4. Error Handling and User Feedback:

- If the user attempts to proceed with an empty Basket, a message must display: "No Distributions selected. Please add at least one item to your Basket before proceeding."
- If the DAAMS service is temporarily unavailable, display: "Unable to connect to DAAMS. Please try again later."

vi) Systems Involved

- nHDsC

2.5 Non-Functional Requirements

This section outlines the quality attributes and operational constraints that the system must meet to ensure a consistent, reliable, accessible, and user-friendly experience. It includes specifications related to user interface and user experience (UI/UX), system performance, scalability, security, availability, error handling, and data integrity. These requirements support the effective implementation and sustainability of the functional requirements described above.

NFR 01 Cross-Browser Compatibility

The system must ensure full and consistent support across major modern web browsers, specifically Google Chrome, Microsoft Edge, Mozilla Firefox, and Apple Safari (latest two stable versions of each), to guarantee reliable functionality, accessibility, and user experience.

1. UI/UX and Accessibility:

- The user interface must render consistently across all supported browsers, including layout, fonts, spacing, colours, and interactive components.
- All interactive features (e.g., dropdowns, modal dialogs, accordions) must function as intended in each browser.
- Keyboard navigation, ARIA landmarks, screen reader compatibility (WCAG 2.1 Level AA), and contrast ratios must remain consistent across browsers.

2. Performance Requirements

- All dataset-related pages (e.g., dataset list, dataset description view, dataset submission forms) must load within three seconds on a standard broadband connection in each supported browser.

- UI responsiveness (e.g., tab switching, modal opening) must occur within 500 ms in all browsers.

3. Error Handling and Feedback

- Error states (e.g., invalid inputs, system errors, and timeouts) must present consistent messaging and styling across browsers.
- No client-side script or validation error should fail due to browser-specific rendering or execution.

4. Submission and Transaction Handling

- Form submission workflows (e.g., submitting or editing dataset metadata) must behave identically across all supported browsers. This includes field validations, auto-save functionality, progress indicators, and transactional integrity.
- All in-form actions must complete reliably without duplication or data loss, regardless of browser.

5. Audit Logging and Monitoring

- The system must capture browser-related metadata (user agent string, browser version) with each submission and major user interaction, enabling trend analysis and troubleshooting
- Monitoring tools must track and flag abnormal interaction patterns that could be browser-specific (e.g., unusually high error rates on a specific browser version).

NFR 02 Mobile Responsiveness

The system must support responsive web design principles to ensure optimal usability and accessibility across a range of mobile devices, including smartphones (≥360px width) and tablets (≥768px width). The layout, interactions, and performance must be adaptive to various screen sizes and input methods.

1. UI/UX and Accessibility:

- The user interface must dynamically adapt to various screen sizes and orientations (portrait/landscape) on mobile phones and tablets.
- Content must remain legible without zooming or horizontal scrolling.
- Navigation elements must be touch-friendly, with buttons and inputs sized appropriately (minimum 48x48px).

- Accessibility compliance (WCAG 2.1 Level AA) must extend to mobile interfaces, including screen reader compatibility, focus indicators, and sufficient colour contrast.

2. Performance Requirements

- On mobile devices connected via 4G or standard Wi-Fi, pages must load within five seconds and respond to user interactions (e.g., expanding sections, toggling views) within one second.
- Assets such as images, scripts, and stylesheets must be optimised for mobile bandwidth and device memory constraints.

3. Error Handling and Feedback

- Validation errors, warning messages, and success confirmations must be clearly displayed on mobile devices without being obscured by on-screen keyboards or overlapping UI elements.
- Toasts, modals, and banners must be mobile-optimised to ensure user awareness and ease of dismissal.

4. Submission and Transaction Handling

- Users must be able to complete dataset submissions, edits, and other transactional workflows on mobile devices without layout or input issues
- The system must handle intermittent connectivity gracefully by preserving form state (e.g., through local storage or progressive enhancement), and prevent duplicate or partial submissions.

5. Audit Logging and Monitoring

- The system must log interactions performed on mobile and tablet devices, including device type and screen resolution (where available).
- Monitoring tools should detect patterns indicating mobile-specific issues (e.g., form abandonments, high validation failure rates) to support proactive troubleshooting and UX optimisation.

NFR 03 Multilingual Support: English and Irish

1. UI/UX and Accessibility:

- A language selector must be available on all pages (public and authenticated), clearly placed (for example, in the header), allowing users to switch between English and Irish with a single interaction.
- The dataset description interface must display labels, field names, help text, system buttons, navigation menus, and controlled vocabularies in the selected language. Where bilingual entry is allowed (e.g., dataset title or description), the system must support input and display in both languages.
- The multilingual interface must comply with WCAG 2.1 AA, ensuring language attributes are properly applied and screen readers correctly identify and render language-specific content.

2. Performance Requirements

- Language toggling must complete with full interface update in under one second, including translation of dynamic elements.
- Upon user preference detection or language selection, the fully localised page must load in under three seconds under standard network conditions.
- Translated content must be efficiently cached to minimise redundant server calls, improving responsiveness, especially for users switching languages frequently.
- The system must maintain consistent language-switch performance for up to XXX (TBC- total number of users expected) concurrent users, ensuring multilingual support is unaffected by traffic spikes.

3. Error Handling and Feedback

- If a translation is missing, the system must:
  - Fall-back to English (default)
  - Log the event for tracking and resolution.
- If the system fails to change the language (e.g., due to backend failure or client error), the user must see a clear, non-technical message (e.g., "We were unable to switch the language. Please try again.").
- All validation messages, form errors, system alerts, and tooltips must be displayed in the currently selected language with appropriate tone and localisation.

4. Submission and Transaction Handling

- Anonymous Users: Language selection must persist for the session using cookies or session storage
- Logged-in Users: Language preference must be stored in the user profile and applied at login
- Language Selection During Dataset Submission - Data entry forms must:
  - Reflect the current language in labels and help text
  - Allow bilingual metadata entry where applicable (e.g., title in English and Irish).
- Context Retention - When switching languages during metadata submission or editing, users must remain in the same workflow step without data loss or interruption.

5. Audit Logging and Monitoring

- The system must log user-selected language settings for authenticated sessions, including:
  - User ID
  - Language selected
  - Timestamp
  - Access point (e.g., login page, dataset view).
- Any instance where the system falls back to English due to missing Irish content must be logged with:
  - Affected string identifier
  - Page/component
  - Timestamp
  - User language setting.

NFR 04 Provide Footer Functionality Across All Pages

1. UI/UX and Accessibility:

- The footer must be consistently displayed at the bottom of all nHDsC web pages, regardless of device or screen resolution.
- The footer must include the following clearly labelled links/options:
  - Give Us Feedback (opens feedback form)
  - About nHDsC (navigates to an informational page about the catalogue)
  - Privacy Policy (opens privacy statement in a new tab)
  - Accessibility (navigates to accessibility compliance page)

  - Logout (ends user session and redirects to login page).
- The footer must comply with WCAG 2.1 Level AA accessibility guidelines, ensuring full keyboard navigability, screen reader compatibility with descriptive link text and high-contrast text for readability.
- Footer links must be styled consistently with the overall nHDsC design system, ensuring visual alignment and responsive layout across desktop, tablet, and mobile.
- The footer must remain visible without overlapping content and adjust gracefully when content height is shorter than viewport height (sticky footer behaviour).

2. Performance Requirements

- The footer must load seamlessly with the main page content, without introducing additional latency (>100ms)
- All footer links must resolve navigation actions within two seconds under standard network conditions
- The footer must remain responsive across supported browsers and devices, maintaining usability for at least 1,000 concurrent sessions.

3. Error Handling and Feedback

- If a footer link fails to load its target page (e.g., Privacy Policy page unavailable), the system must display a contextual error message (e.g., "This page is temporarily unavailable. Please try again later.") and log the error for monitoring and support.
- If the Logout action fails, the system must display a clear warning ("Logout was unsuccessful. Please try again.") and ensure the user session is still protected (no partial logouts).
- If the Feedback link cannot load the feedback form, the system must provide an alternate message: "Feedback form could not be loaded. Please contact support at support email."

4. Submission and Transaction Handling

- Footer links to external content (e.g., Privacy Policy, Accessibility) must open in a new tab, ensuring that the main session is not interrupted.
- The Logout option must securely terminate the session, removing tokens/cookies from client storage and redirecting to the login page.

- The Give Us Feedback option must trigger the integrated feedback functionality described in FR 022.

5. Audit Logging and Monitoring

- The system must log user interactions with footer elements, including:
  - Timestamp of click
  - Link selected (Feedback, About, Privacy Policy, Accessibility, Logout)
  - User ID (if authenticated) or anonymous session ID.

NFR 05 Allow for User identification, authentication, and authorisation, including the option to enforce multi-factor authentication (MFA)

1. UI/UX – Accessibility and Layout

Accessibility:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA)
- Form fields must be clearly labelled, with inline validation messages (e.g., "Please enter a valid phone number")
- Required fields must be indicated with an asterisk `*`, and optional fields explicitly labelled.
- All screens must have consistent branding, navigation aids (e.g., Home > Login), and contact support links (e.g., Need help?).

Registration form UI/UX:

- The registration form should have below fields:
  - User Name: Accepts free text with max character limit (e.g., 50 characters)
  - Official Email ID: Must validate format using RegEx and disallow public domains (e.g., `gmail.com`, `yahoo.com`) if policy applies
  - Department/Organisation: Dropdown should be searchable and categorised where appropriate
  - User Role: Dropdown with common roles pre-listed; The roles should be Data Holder Admin, Data Holder Editor and Data User.
  - Phone Number: Must include country code; validate format and check for fake/test values

  - Password/Confirm Password: Must enforce password strength (minimum eight characters, upper/lowercase, number, symbol); inline tooltip for criteria
  - Provide password visibility toggle and strength meter
  - Password and confirm-password mismatch must be detected in real time.

- Submit Behaviour:
  - Upon clicking "Register", show a loading spinner and disable form.
  - Send confirmation email to user.
  - Show a submission confirmation message with estimated processing time.
  - Auto-forward the registration request to the system admin dashboard.

Admin Approval UX:

- System must receive an email and in-portal notification upon a new registration request
- nHDsC must show:
  - User details with all submitted fields
  - Approve/Reject buttons
  - Mandatory comment field when rejecting a request.
- On approval or rejection:
  - An email must be sent to the user indicating the decision
  - Rejection email must include the admin's comment.
- System admin UI must support filtering/sorting of pending and processed registration requests.

Login with OTP UI/UX

- Login Form
  - Two-step login process with email and password first
  - On correct credentials, user clicks Send OTP
  - Show real-time feedback: e.g., "OTP sent to your registered email.".

- OTP UI
  - OTP input must appear inline (not as a separate page)
  - Support six-digit numeric code input; auto-focus and tab through digits
  - Include "Resend OTP" button (active after 30 seconds)

  - OTP expiry should be shown visually (countdown timer)
  - On invalid OTP, display clear error: "Incorrect or expired OTP."

Forgot Password UX/UI:

- Trigger and Form
  - "Forgot Password?" link on login screen
  - Clicking opens form asking for registered email address
  - After submission, show message:
    "If this email is registered, you'll receive a password reset link shortly."
- Email must include:
  - Secure, unique, time-limited reset link (e.g., expires in 15 minutes)
  - Support contact details
  - Reset links are one-time-use and must use HTTPS
- Reset Password Page should have
  - New Password, Confirm Password and Reset Password button
  - New Password/Confirm Password must enforce password strength (minimum eight characters, upper/lowercase, number, symbol); inline tooltip for criteria
  - Password visibility toggle and strength meter
  - Password and confirm-password mismatch must be detected in real time.
- On submission:
  - Show message: "Your password has been successfully updated."
  - Provide direct link to Login page.
- If the reset link is invalid or expired, show error with option to resend.

Self-Deactivation:

- Must be initiated from account settings.
- Modal must enforce a warning about account deactivation and Dataset transferee selection (for Data Holder roles).
- Trigger reassignment and send notifications to transferee, deactivating user, and System Admin.

Admin/On-Behalf Deactivation:

- Data Holder Admins can raise deactivation requests for their organisation's Data Holder.
- Dataset transferee selection required (if applicable).

- Requests routed to System Admin for approval.
- Confirmation and logging required.

System Admin Removal:

- Must enforce reassignment before account removal.
- Confirmation UI and audit trail required.

2. Performance Requirements

- All identity-related actions (registration, login, OTP verification, password reset, deactivate/remove user) must complete within three seconds under normal network and system conditions.
- Admin approval interfaces must load user registration details and display actionable items within two seconds for 95% of use cases.
- OTP delivery must occur within 10 seconds of request submission.
- Password reset emails must be delivered within 30 seconds of request.
- Admin views (user approvals or deactivations) must load within two seconds in 95% of cases.

3. Error Handling and Feedback

User Feedback:

- Provide immediate and clear feedback for invalid form inputs, OTP errors, and expired/used reset tokens.
- Example messages:
  - "Incorrect or expired OTP."
  - "If this email is registered, you'll receive a reset link shortly."

Failed Login Attempts:

- Introduce rate limiting: after five failed attempts, delay login functionality by 60 seconds.
- Display non-technical error: "Too many attempts. Please try again later."

Downtime Handling:

- In the event of email/OTP delivery service failure, show fallback message and offer retry or support contact link.

Token Expiry:

- Expired/used reset tokens must result in a meaningful error and option to request a new link

4. Submission and Transaction Handling

All submissions must:

- Disable UI during processing
- Confirm success via on-screen message and email
- Automatically route requests (e.g., to System Admin dashboard).

Deactivation Requests:

- Require transferee selection for Data Holders.
- Validate transferee status.
- Block submission if missing or invalid.
- Notify all parties after approval.

Dataset Reassignment:

- Must update metadata and permissions.
- Notifications to transferee and deactivator (if applicable).
- Transactions (login, reset, approval) must be atomic: only complete on full success.

5. Audit Logging and Monitoring

- The system must log and monitor critical identity and access-related events.
  - Login events such as Success and failed login attempts and Lockout triggers after repeated failures.
  - OTP Events such as OTP generation, misuse attempts, and resend activity.
  - Password Reset Events such as Reset link creation, usage, and expiration.
  - Registration Flow such as Timestamp and user details for all submissions, approvals, and rejections.
  - Account Deactivation/Removal: Who initiated, when, dataset transfer log.
- All logs must include:
  - Timestamp
  - User ID or email
  - IP address
  - Action performed
  - Outcome (success/failure).

NFR 06 Create, Read, Update and Delete (CRUD) dataset descriptions using the HealthDCAT-AP application profile

1. Dataset Description UI/UX

Accessibility and Layout:

- The metadata record creation/editing interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- Wizard steps must be clear with a visual indicator of progress (e.g., Highlighting each steps, showing sequence of steps).
- Required fields must be marked with an asterisk `*`; optional fields must be labelled explicitly.
- Each field must include tooltips or contextual help for legal/technical terms (e.g., "Use controlled vocabulary for this field").
- Descriptions must be auto saved every 30 seconds or on field change
- Allow draft saving without validation and enable highlighting of incomplete sections with prompts.

Create Dataset Interaction Flow:

- When the data holder clicks on "Create Dataset". A modal/popup window must appear with the question:
  "Do you want to create a new dataset from existing metadata you previously added?"
- Two options must be presented:
  - Yes - Create from existing metadata:
    - A dropdown or list must allow the user to select from their previously created datasets.
    - Upon selection, the user must be able to choose which metadata fields to copy. An "All" option must be available to copy every field, or the user may select specific fields.
    - Selected fields (e.g., Title, Description, Tags, Licensing, Classification) must be auto-populated in the creation form.
    - Data Holders must be able to review and edit any pre-filled information.
  - No - Create a blank dataset:
    - If selected, the system must open an empty dataset creation form, requiring manual entry of all metadata fields.

- The wizard interface must support both paths and maintain accessibility, autosave, tooltips, and field guidance regardless of the creation method chosen.

Field details to create dataset:

- Text fields: Free-text with character limits and validation (e.g., 500 characters max).
- Field Types:
  - Controlled vocabularies: Drop-downs with search/filter.
  - Multi-choice: Checkboxes with logical grouping.
  - Date fields: Calendar picker enforcing date format (e.g., `DD-MM-YYYY`).
  - Sensitive data: Toggle to flag and explain justification for "Highly Sensitive" classification.
- The label entry section must be accessible via a dedicated tab or link on the metadata record creation page.
- Sections (e.g., "Assessment of Technical Quality", "Coverage", "Data Modifications") must be collapsible/expandable to reduce visual clutter.
- Each field must have tooltips or contextual help (e.g., For Keyword text tag the tooltips must mention "This explains the terms that may link this dataset to similar ones.").
- Required fields must be indicated with an asterisk `*`; recommended and optional fields must be clearly labelled.
- Field groupings must visually separate each Mandatory, Recommended and Optional fields using headers, spacing, or cards.
- Provide visual feedback for completed/incomplete sections (e.g., green checkmarks or red warnings).
- The interface must highlight incomplete or invalid sections upon attempting submission.

Field Input Types:

- Text fields: Free-text areas with max character limits (e.g., 500 characters), with autosave and inline validation.
- Dropdowns: Predefined options for controlled terms (e.g., "High", "Medium", "Low"), with search functionality.

- Justification fields: Optional comment boxes if needed to support specific label selections (e.g., why a dataset has "High Completeness").

Dataset metadata Summary UI:

- Users must be able to preview the dataset description summary/Overview before submission.
- Preview must mirror how users will see the information in the public-facing catalogue post-approval.

Dashboard UX:

- When data holders click on ”My Dashboard” they should be able to view a dashboard.
- The dashboard should display a pie or donut chart showing the number of dataset descriptions that are approved, pending review, rejected, and deleted.
- Dataset descriptions should be listed with ID, description, status, creation date, last modified date, and available actions (Edit / Delete).
- The Edit option must be disabled if the dataset metadata is in Pending Review status.
- The Delete option must be disabled if the dataset metadata is already in Deleted status.
- The ID/description should be a hyperlink, allowing the data holder to view the associated overview page by clicking on them.
- Include filters (e.g., Status, Dataset Type) and sort (e.g., alphabetical, date).
- Show a justification dialogue box that must be filled by data holders when deleting a dataset description. Also, display a confirmation modal before deletion.
- Pagination and Dataset Display Controls:
  - By default, 10 dataset descriptions must be displayed per page.
  - Users must be given the option to change the number of visible datasets using a dropdown menu, with options like 10, 25, 50, and 100.
  - The dashboard must display a line below the table showing:
    - The total number of dataset descriptions
    - The number currently visible on the page
    - Example: “Showing 1–10 of 42 entries”.

  - If the total number of dataset descriptions exceeds the current view setting, the system must provide pagination controls allowing users to navigate to:
    - Next page
    - Previous page
  - Specific page numbers (e.g., 1, 2, 3…).
  - All pagination and dropdown controls must be fully keyboard-navigable and screen reader compatible.

2. Performance Requirements:

- All CRUD actions (Create, Read, Update, Delete) must complete in under one second under normal load (100 concurrent users).
- Wizard step transitions and validations should occur within one to two seconds.
- The dashboard must load up to 100 entries in under two seconds.
- Auto-save should not disrupt typing or navigation; background process with no UI blocking.

3. Error Handling and Feedback:

Dataset description submission Form and Wizard:

- All validation errors must be shown inline next to fields (e.g., "This field is required").
- HealthDCAT validation failures must specify the exact field and reason for failure.
- On submission failure (e.g., server error), show fallback message: "Something went wrong. Please try again or contact support."
- Highlight unsaved changes when navigating away from the form.

Dataset metadata Uploads:

- Auto-Save and Recovery: Data must persist locally (via browser cache or local Storage) during unexpected connection loss.
- On reload after crash or disconnection, allow user to recover last unsaved draft.

4. Submission and Transaction Handling:

- Clicking on Submit button while submitting the dataset description should:
  - Show confirmation modal with summary of filled fields

  - Trigger full validation (both client-side and server-side)
  - Disable the form and show loading spinner during submission.
- All submissions must be atomic: either fully completed or rolled back on error.
- Confirmation message shown after successful submission: "Dataset description submitted to Admin for review."
- An automatic email and in-portal notifications must be generated and sent to the system admin, notifying them of the new dataset description and requesting review.
- A notification should also be displayed on nHDsC dashboard.

5. Audit Logging and Monitoring:

- Audit logs must capture:
  - All CRUD operations (who, when, what changed)
  - Dataset submissions and validation attempts
  - File uploads and deletions.
- System should log errors and validation exceptions with trace IDs for debugging.
- Health and usage metrics must be available for administrators (e.g., number of drafts, submissions per month).

NFR 07 Add data quality and utility labels

1. Data Quality and Utility Label UI/UX:

DQUL Interface Design:

- The DQUL form must comply with WCAG 2.1 Level AA accessibility standards, including full keyboard navigation, proper field labels, and screen reader compatibility.
- Data Holders must be able to:
  - Access the "Add Data Quality and Utility Labels" option while creating or editing a dataset description
  - Enter scores for all 12 dimensions of quality and utility labels
  - Upload supporting documentation (e.g., PDF, Word, or structured file formats) with clear instructions on file size and type limits.
  - Save progress in the DQUL form without submitting (draft mode)
  - Submit completed labels as part of the dataset description workflow.
- Each label field must include:
  - Field name and description

    - Input method (numeric scale, dropdown, or text field depending on the dimension definition)
    - Inline help text or tooltips to guide data holders.
- Completed DQUL information must appear in the dataset description under a dedicated "DQUL" tab.
- The DQUL tab must support both detailed view (all 12 dimensions with scores and explanations) and a summary view (aggregated scores or indicators). The interface must remain responsive and accessible across desktop, tablet, and mobile devices.

2. Performance Requirements

- The DQUL form must load and render in under two seconds under standard network conditions.
- Saving a draft of the DQUL form must complete within one second.
- Uploading supporting documentation up to 20MB must complete within 10 seconds, with progress indicators.
- Display of DQUL information on the dataset description page (summary + tab) must render within two seconds, even for datasets with complex labels.
- The system must handle at least 200 concurrent DQUL submissions without performance degradation.

3. Error Handling and Feedback:

- If a required dimension score is missing, the system must prevent submission and highlight the missing field and display a contextual error message: "All 12 quality and utility dimensions must be scored before submission."
- If an invalid score is entered (e.g., outside of defined range), the system must display: "Invalid entry — please select a valid score."
- If supporting documentation upload fails, the system must:
    - Allow retry.
    - Display a clear error message (e.g., "Upload failed. Please try again or use a supported file format.").
- If the DQUL service is temporarily unavailable, the system must save the dataset description without labels and provide a warning: "Data Quality and Utility Labels could not be saved at this time. Please try again later."

4. Submission and Transaction Handling:

- All DQUL data must be stored securely in the nHDsC database, linked to the associated dataset record.
- The system must allow the addition of one DQUL label per dataset.
- Draft and final versions of DQUL submissions must be timestamped and auditable.
- Supporting documents must be stored in a secure repository with reference links maintained in the dataset record.
- The system must export DQUL metadata in HealthDCAT-AP compatible format, ensuring future interoperability with EU-level catalogues

5. Audit Logging and Monitoring:

- The system must log the following for each DQUL event:
  - Dataset ID
  - Data Holder ID
  - Timestamp of submission or update
  - Dimension scores entered
  - Submission status (draft, final, failed).

Note: The current Add QUL Label functionality requires Data Holders to manually enter Quality and Utility label scores based on outputs from the QUANTUM tool. During the development of this report, this functionality was not yet available in the HealthData@EU Central Platform (Release five).[21] It is scheduled to be included in HealthData@EU Central Platform Release eight, planned for September 2026. Accordingly, this functionality shall be revised and updated to ensure alignment with the EU Central Platform releases.

NFR 08 Enforce minimum elements for dataset descriptions

1. Minimum Elements Validation UI/UX

Accessibility and Layout:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- Validation messages must appear inline next to relevant fields, using accessible colours and icons (e.g., red cross, exclamation mark).

- Before submitting the dataset, data holders must be able to access a dedicated "Summary" section that lists all missing or invalid fields in one place.
- Provide a tooltip or inline guidance for each mandatory field (e.g., "This field is required by EHDS Regulation – see metadata schema for reference.")
- The interface must visually differentiate between required, optional, and conditionally required fields (e.g., a legend or coloured tags).
- Navigation between failed sections must be supported via quick links or collapsible panels to improve error resolution flow.

2. Performance Requirements

- Automatic validation must trigger instantly when the user attempts to submit or when navigating between wizard steps, with results displayed in under one second.
- Pre-submission format checks must complete in under two seconds even with up to 100 metadata fields populated.
- Real-time field-level validation must not introduce noticeable delay in typing or navigating forms.
- The full metadata validation against HealthDCAT-AP rules must be executed client-side where possible to minimize load and latency.

3. Error Handling and Feedback

Inline Validation Errors:

- Errors must include:
  - Field label
  - Cause of failure (e.g., "Missing value for required field: `publisher`" or "Invalid URI format for `accessURL`")
  - Suggestion or link to help documentation (e.g., "Refer to HealthDCAT-AP Field Guide for correct format.")
- Prevent submission until all validation errors are resolved.
- Use persistent (non-dismissable) warnings for critical compliance issues until resolved.

Summary/Overview Panel:

- The overview page must display all dataset metadata fields entered by Data Holders.

- When a Data Holder clicks on any field in the overview page, the system must navigate to the corresponding field in edit mode. This enables users to correct data directly while reviewing the overview.
- An aggregated list of all errors must be displayed at the top of the page or in a side panel, with clickable links to navigate to the affected fields.
- If server-side validation fails after client-side validation passes (e.g., due to an invalid license URI), the system must display a clear error message: "Invalid licence URI entered. Please enter valid details or contact support."

4. Submission and Transaction Handling

- When a Data Holder Editor completes dataset creation, the system must provide the option "Complete and Send for Approval." The Editor must be able to select the Data Holder Admin to whom the approval request will be sent.
- Upon clicking "Complete and Send for Approval":
  - Perform full local validation (HealthDCAT-AP minimum + EHDS-specific checks).
  - If errors are found, halt submission and focus the user on the first failed field.
  - If validation passes:
    - Disable the form.
    - Display a loading spinner.
    - Submit metadata for review.
- The Data Holder Admin must be able to view approval requests in the "Pending Approval Requests" tab.
- The Data Holder Admin must review the dataset metadata, make updates if required, and have the ability to click "Submit for Review."
- Upon successful submission, display a success message:
  - "Dataset description passed compliance check and was submitted to Admin for review."
  - Provide options to navigate to "My Datasets" and "Create Datasets."

A confirmation email must be sent to both the Data Holder Admin and Editor, including a summary and timestamp of the submission.

5. Audit Logging and Monitoring

- Log all validation runs with:
  - Timestamp
  - User ID

- Dataset ID
- Pass/fail result
- List of fields that failed (if any).

NFR 09 Submit dataset description

1. Submission UI/UX

Confirmation UX:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- Before submitting the metadata, data user should be able to see the overview of the metadata details filled by them.
- "Submit for Review" button must be clearly visible and positioned at the bottom of the finalised dataset description form.
- After clicking "Submit for Review", a modal or inline message must confirm submission and display the next steps.

Visual Status Indicators:

- Dataset status (e.g., "Draft", "Pending Review", "Submitted", "Deleted", "Revoked", "Published") must be displayed using standardised badges or colour codes for quick recognition.

2. Performance Requirements

- Dataset submission confirmation must be displayed within two seconds of form submission under standard conditions
- Dashboard views (including filters and searches) must load within two seconds for up to 500 datasets per data holder.
- Email and in-portal notifications to both Data Holders and System Admins must be generated and dispatched within 10 seconds of submission.
- The change in the status of dataset metadata from "Draft" to "Pending to Review" should happen within two seconds of processing.

3. Error Handling and Feedback

- All mandatory fields must be validated client-side and server-side with inline messages (e.g., "Dataset title is required.")
- If submission fails (e.g., due to connectivity or server issues), the portal must display a clear, non-technical error message and preserve the user's input.

- Users must be allowed to retry failed submissions without data loss.
- If email delivery fails, fallback messages must be logged, and the system must attempt to resend (minimum three retries).

4. Submission and Transaction Handling

- Dataset description submissions must be processed as a single atomic transaction. If the confirmation, dashboard update, or notifications fail, the transaction must be rolled back.
- The system must display a message confirming successful submission and trigger email notifications.
- Submitted datasets must be automatically marked as "Pending Review" in the dashboard.
- The system must send notifications to:
  - The Data Holder Admin and Editor (confirmation)
  - The System Admin (alert of new submission and requesting for review)
  - Notifications must include dataset title, submission timestamp, and status.

5. Audit Logging and Monitoring

- The system must log all critical submission events, including:
  - Dataset submission attempts (successful and failed)
  - Dataset status changes (e.g., Draft → Pending Review)
  - Notification delivery logs for both recipient types
  - Dashboard access and filter actions (optional, for user activity insights).
- Each log must include:
  - Timestamp
  - Dataset ID
  - User ID and role (Data Holder, System Admin)
  - Action performed
  - Outcome (success/failure)
  - System errors (if any).

NFR 010 Review dataset description and data quality and utility labels

1. Review Workflow UI/UX

Accessibility and Layout:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- The review dashboard must visually separate dataset descriptions pending review, approved, and revoked, using accessible colour indicators and tags.
- Review actions (“Review”,"Submit for publication", "Revoke") must be clearly labelled and accessible via buttons.
- Each dataset entry in the dashboard must provide:
  - Metadata summary
  - Data quality and utility label summary
  - Expand/collapse sections for detailed review.
- Tooltips or inline help must be provided for review-related actions (e.g., “Click here to view label details”).
- For revoked items, the justification must be accessible as a tooltip or linked modal next to the item.

Admin Dashboard UX:

- When System Admin click on “My Dashboard”, they must be directed to a dashboard view displaying the status and details of dataset descriptions under their domain.
- The dashboard must be responsive, accessible, and optimised for both desktop and tablet devices.
- A pie or donut chart must be displayed prominently at the top of the dashboard, summarising the number of dataset descriptions by status:
  - Published
  - Pending review
  - Revoked
  - Deleted.
- Each segment of the chart should be interactive (e.g., clickable to filter the list view accordingly) and accessible with screen readers.
- The list of dataset descriptions must include the following columns:
  - Dataset ID (hyperlinked to the overview page)
  - Description (hyperlinked to the overview page)
  - Status (with colour-coded tags)

    - Creation Date
    - Last Modified Date.
- Available Actions (“Review”, ”Submit for publication”, ”Revoke”).
- The "Revoke" action must be disabled when a dataset description is in currently being updated by Data Holder.
- Upon clicking “Revoke”, a modal or overlay must prompt for justification with a minimum 20-character requirement.
- Include guidance text near the justification field (e.g., “Explain why the dataset or label does not meet standards”).
- Clicking on the ID or Description must open the dataset overview page in the same or a new tab.
- Filtering and Sorting:
    - Filter options, such as Status (Approved, Under Review, etc.) and Dataset Type.
    - Sorting capabilities on columns like Alphabetical Order, Creation Date, and Last Modified Date.
- Pagination and Dataset Display Controls:
    - By default, 10 dataset descriptions must be displayed per page.
    - Users must be given the option to change the number of visible datasets using a dropdown menu, with options like 10, 25, 50, and 100.
    - The dashboard must display a line below the table showing:
        - The total number of dataset descriptions
        - The number currently visible on the page
        - Example: “Showing 1–10 of 42 entries”.
    - If the total number of dataset descriptions exceeds the current view setting, the system must provide pagination controls allowing users to navigate to:
        - Next page
        - Previous page
        - Specific page numbers (e.g., 1, 2, 3…).
    - All pagination and dropdown controls must be fully keyboard-navigable and screen reader compatible.
- System confirmation messages must appear after successful review/revocation (e.g., “Revocation successful. Notification sent to data holder.”)
- Include breadcrumb navigation (e.g., Home > Review Submissions > Dataset Title) and back-to-dashboard link.

2. Performance Requirements

- The dataset review dashboard must load in <2 seconds for up to 1,000 pending datasets.
- Dataset metadata and labels must render in <1.5 seconds upon selection.
- Revocation requests must be processed and reflected in the UI within three seconds.
- Loading the reviewer dashboard must take no more than two seconds for up to 200 pending descriptions.
- Opening an individual dataset description and its associated DQUL must take under one second.
- "Submit for publication" and "Revoke" actions must complete and reflect status changes in under two seconds.
- Email and in-portal notifications must be sent to Data Holders within 10 seconds of revocation.

3. Error Handling and Feedback

- If a revocation fails, show a clear message:
  "Unable to perform the selected action. Please try again or contact support."
- If the catalogue or notification service is temporarily unavailable, show a fallback message and queue the revocation for retry.
- If no justification is entered during revocation, disable the submit button and show inline error:
  "Justification is required to revoke a dataset description or label."
- If automatic notifications fail, queue a retry.

4. Submission and Transaction Handling

- "Revoke" must not change dataset state unless justification is successfully submitted and processed
- Revocation must be processed as a single transaction including:
  - Status change to "Revoked"
  - Justification capture
  - Notification dispatch.
- If any of these steps fail, the system must roll back and prompt the user to retry.
- On successful revocation, display: "Dataset description and labels have been revoked and the data holder notified."

- The dataset metadata status should change to "Revoked"
- The revoked status must be immediately visible on both the admin review dashboard and the data holder's dashboard.

5. Audit Logging and Monitoring

- The system must log all review and revocation-related activities, including:
  - Dataset view events (Dataset ID, Admin ID, Timestamp)
  - Revocation submissions (Dataset ID, Justification Text, Admin ID, Timestamp)
  - Notification dispatch logs (Success/Failure, Recipient Email, Timestamp)
  - Any system errors or retry attempts related to revocation.
- Logs must include:
  - Action type
  - Affected dataset ID
  - User ID and role
  - Timestamp
  - Status or outcome (success/failure)
  - Justification text (for revocations).

NFR 011 Publish dataset descriptions using the HealthDCAT-AP application profile

1. UI/UX – Accessibility and Layout

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).

Publish Action Interface

- The "Submit for publication" action must be clearly visible on the admin review interface, with confirmation modals prompting before execution.
- Include tooltips or info icons to describe what "publication" entails and its implications (e.g., public exposure, version control).

Metadata Display:

- Published metadata must be presented in a clean, readable layout with collapsible sections (e.g., Dataset Overview, Source, Scope, Data Type, Conditions of Access).
- Provide breadcrumb navigation and filters (e.g., by country, topic, data holder) for public catalogue viewers.

Version Code Display:

- The generated dataset identifier (e.g., `IE-HSE-0012-V01`) must be clearly displayed post-publication and explained via tooltip.

2. Performance Requirements

- Publication actions must complete in <3 seconds, including database update, code generation, and notification trigger.
- Public catalogue search and load must return results in <2 seconds for up to 10,000 published datasets.
- Code generation and versioning logic must scale to handle concurrent publication events with no duplication.

3. Error Handling and Feedback

- Ensure the dataset passes all mandatory field validations before enabling the publish button.
- If required fields are missing or incompatible with HealthDCAT-AP, disable the publish option and show error indicators.
- On publication failure, display a contextual error (e.g., "Dataset could not be published due to system error. Please try again.").
- If code generation fails, the system must retry up to three times and then flag for manual intervention.
- If an email and in-portal notification to the Data Holder fails, it must be logged and displayed as an alert to the admin.

4. Submission and Transaction Handling

- The following must occur as part of a single atomic transaction:
  - Validation confirmation
  - Generation and assignment of dataset identifier/version
  - Status change to "Published"
  - Removal from the "Pending Review" dashboard
  - Notification to Data Holder
  - Inclusion in the public catalogue.
- On success, display: "Dataset successfully published and added to the nHDsC."
- Dashboard must reflect new status instantly; "Pending Review" count must update in real-time.

5. Audit Logging and Monitoring

The system must maintain a robust audit trail for publication actions:

- Events Logged:
  - Dataset ID and Version Code
  - Publishing Admin (user ID, role)
  - Timestamp of publication
  - Validation status and source system
  - Generated dataset code
  - Notification delivery status.
- Logs must be accessible only to authorised roles (e.g., System Admin).

NFR 012 Update or delete dataset metadata descriptions and maintain change history

1. Update UI/UX – Accessibility and Layout:

Editing Interface Accessibility:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- In the "My Dashboard", data holders must see "View", "Edit", and "Delete" buttons adjacent to each dataset row.
- The "Delete" button must be clearly labelled and styled to indicate a destructive action.
- Edit options should be disabled for datasets marked as "Pending Review".
- Delete options should be disabled for datasets marked as "Deleted".

Editing Interface:

- Editing fields must:
  - Support plain text and markdown formatting
  - Provide contextual tooltips for fields based on HealthDCAT-AP metadata elements
  - Include inline validation messages and required field indicators.
- Version Timeline View:
  - Include an expandable "View History" panel showing version history in a collapsible timeline format.
  - Each version must display timestamp, editor, change summary, and a "View Details" link for full metadata.

Deleting Interface:

- Clicking "Delete" must trigger a modal confirmation prompt summarizing the dataset name and status.
- Modal must include: "Are you sure you want to delete this dataset description? This action cannot be undone."
- Version Timeline View:
  - "View History" must remain available for deleted datasets and it should be visible in the Data holder dashboard.

2. Performance Requirements

- Deletion operation (soft delete) must complete in ≤1.5 seconds from the time of user confirmation.
- Update the dashboard view within two seconds to reflect removal.
- System must maintain performance even with 1,000+ total datasets and 10+ deletions/updates occurring concurrently.
- Time to load the full version history for any dataset must be ≤1.5 seconds for up to 10 historical versions.
- Metadata updates must be saved and reflected in the system within three seconds of user submission.

3. Error Handling and Feedback

- If the form fails client-side or server-side validation, an error banner must clearly identify the issue (e.g., "Missing mandatory field: Description").
- On failed submission due to concurrency, show a message such as: "This dataset is currently being edited by another user."
- Partial updates/deletions must be rolled back to ensure consistency.
- If deletion/update fails due to permission issues or system error:
  - Display a clear error message: "Action failed – You do not have permission or the dataset is locked for review."
- Provide tooltips for disabled "Delete" and "Edit" buttons explaining why action is not available (e.g., "This dataset is currently under review").

4. Submission and Transaction Handling

Update Submission Flow:

- The system must prompt the Data Holder to enter a reason for changes before submission.

- When a Data Holder Editor updates dataset metadata, the system must allow them to select the Data Holder Admin to whom the update approval request will be sent before clicking "Complete".
- The Data Holder Admin must be able to review the approval request and click "Submit for Review."
- Upon clicking "Submit for Review," the system must display a confirmation modal summarizing the changes.
- During submission, the form must be disabled, and a spinner with status text ("Submitting changes…") must be shown.
- All changes must be committed as an atomic transaction: either all updates are saved or none.
- On success, the system must display a success message:
  - "Dataset description updated successfully and submitted to Admin for review."
  - Provide options to navigate to "My Datasets" and "Create Datasets."
- An automatic email must be sent to the Admin containing the dataset ID, version number, and the user-submitted change summary.
- A notification must be displayed in the Admin dashboard indicating that a dataset update is awaiting review.
- The system must generate a new version with an incremented identifier (e.g., `v1.2 → v1.3`).Versioning Rules:
  - Immutable: Past versions cannot be altered or deleted
  - Traceable: Each version must record who made the change, when, and what changed
  - Accessible: "View History" feature must display all previous versions in read-only mode.

Delete Submission Flow:

- The system must require explicit confirmation from the Data Holder in a modal before proceeding with deletion.
- When a Data Holder Editor requests deletion of dataset metadata, the system must request them to select the Data Holder Admin to whom the deletion request will be sent.
- The Data Holder Admin must be able to review and confirm the deletion request.
- During deletion, the system must disable other dashboard actions and display a spinner.
- On success, the system must display the message:

"Dataset description deleted successfully."

- Deletion must be handled as an atomic transaction: either all references and metadata are soft deleted, or the dataset remains unchanged.
- An automatic email/notification must be sent to the Admin containing the dataset ID, version number, and the user-submitted change summary or deletion justification. Datasets must be soft deleted by default (i.e., flagged as deleted, not purged), preserving:
  - Metadata
  - Version history
  - Audit logs.
- The deleted dataset metadata must remain visible to the Data Holder Editor, Data Holder Admin, and System Admin with the status "Deleted."
- Admin-only functionality may allow full purging if required for compliance (subject to access policies).

5. Audit Logging and Compliance

- Every update must be logged with:
  Dataset ID, version number, user ID, timestamp, IP address, and user-provided change summary.

NFR 013 Maintain Versioning of updated Dataset metadata

1. Dataset Description UI/UX:

Version History Accessibility:

- The dataset version history must be easily accessible by Data Holders and System Admins from the main dataset description dashboard with clear labels, indicators (e.g., current version badge), and consistent styling.

Structured Version Timeline:

- The UI must present versions in a reverse chronological timeline format with clear indicators for
  - Version number
  - Status (e.g., Current, Deprecated)
  - Created by
  - Creation date/time.

Version Differencing View:

- A visual comparison tool must be available to highlight changes between current and previous version, by clicking on 'Changes' button present along with each version timeline (e.g., inline highlights or side-by-side panels).

Accessibility Compliance:

- All interactive elements in the versioning UI (e.g., view, compare, filter, expand) must comply with WCAG 2.1 AA accessibility standards.

2. Performance Requirements

- Version Retrieval Speed – The system must retrieve and display version history for any dataset within two seconds under normal operating load.
- Scalable Performance – The system must efficiently handle datasets with up to 100 versions and scale to support 100,000 version records across the system without impacting user experience.
- Comparison Performance - When using the comparison (diff) tool, differences between versions must be displayed within three seconds for metadata descriptions up to 10,000 characters

3. Error Handling and Feedback

- Versioning Errors – If the system fails to generate a new version due to backend, validation, or connectivity errors, users must be shown a clear error message and advised on corrective actions (e.g., "Version could not be saved due to server timeout. Please try again.").
- UI Feedback – On successful version creation or update, the system must display real-time confirmation messages, including the new version number.
- Fallback Messaging – In case version history cannot be retrieved due to service disruption, the UI must display a fallback message and an option to retry.

4. Submission and Transaction Handling

- Atomic Versioning – The creation of a new version must be atomic, ensuring all metadata elements, audit logs, and version references are updated simultaneously or not at all in the case of failure.
- Conflict Prevention – The system must prevent concurrent edits to the same dataset description through optimistic locking or real-time session checks.
- Unique Version ID Generation – Every saved version must receive a system-generated unique identifier (e.g., UUID or sequential version number) that is immutable and referenced across UI and logs.

- Generate a new version with an incremented identifier (e.g., `v1.2 → v1.3`).
- Versioning Rules:
  - Immutable: Past versions cannot be altered or deleted
  - Traceable: Each version must record who made the change, when, and what changed
  - Accessible: "View History / Activity Stream" feature must display all previous versions.

5. Audit Logging and Monitoring

- Each version creation event must be logged with:
  - Version ID
  - Dataset ID
  - Timestamp
  - Actor ID (Data Holder / System Admin ID)
  - Summary of changes (when available).

NFR 014 Enable Rollback to Previous Version of Dataset Description when needed (with Admin Approval)

1. Dataset Description UI/UX:

- All UI components for rollback, including request, approval view, version preview, and rollback reason entry must comply with WCAG 2.1 AA accessibility standards.
- The rollback option must be available directly within the version history view of each dataset, clearly distinguished from edit/update actions.
- Before initiating rollback, Data Holder Admin/Editor must be able to preview the selected version in a side-by-side comparison with the current version and receive a clear, contextual confirmation prompt.
- Once rollback is completed and a new version is created, the version history must show the new version as, for example, "Version 8 (Copy of Version 5)". The original source version should be clickable from the new entry.
- When a Data Holder Editor requests a rollback, the system must require them to select the Data Holder Admin to whom the rollback approval request will be sent.
- The Data Holder Admin must be able to review the rollback request and submit it for System Admin approval.

- All rollback-related actions (submission, pending approval, rejection, completion) must provide real-time feedback using clear, non-technical language.
  For example, "Rollback request submitted to System Admin for approval."

2. Performance Requirements

- Submitting a rollback request must be processed and acknowledged by the system within two seconds under normal operating conditions.
- Admins must be able to load rollback requests, including version comparison and metadata, within three seconds.
- Once approved, the rollback action must result in the creation of a new version within five seconds, including validation and logging.
- The system must support up to 500 rollback requests per day across the platform without performance degradation.

3. Error Handling and Feedback

- If the system detects inconsistencies during rollback (e.g., broken references, schema mismatches), it must:
  - Halt the rollback
  - Display a specific error message
  - Offer corrective actions or direct users to contact support.
- Admins must receive error messages if approval submission fails due to network/server issues, and the rollback request status must reflect "Pending" until confirmed.
- Data holder must be notified via UI and/or email when:
  - Their rollback request is approved or rejected
  - The rollback is completed successfully and a new version is available.

4. Submission and Transaction Handling

- Rollback must follow a strict two-step flow:
  - Step 1: Data Holder submits rollback request with justification
  - Step 2: System Admin must approve it for execution.
- The rollback operation must be handled as an atomic transaction, ensuring that:
  - A new version is created only upon successful validation
  - Metadata dependencies are verified and updated consistently
  - If validation fails, no partial changes are saved.

- Each rollback operation must create a new version with a unique identifier, regardless of its similarity to a previous version.
- Display format must indicate the source version (e.g., Version 8 [Copy of Version 5]).
- The system must prevent rollback if a new version was created after the rollback request was submitted, prompting the requester to review recent changes first.

5. Audit Logging and Compliance

- All rollback-related actions must be logged with:
  - Data holder ID and timestamp
  - Target version selected for rollback
  - Admin approval/rejection timestamp and ID
  - Newly created version ID
  - Justification provided
  - Result status (Success/Failed with reason).
- Rollback approvals must include an immutable decision record by the System Admin, including any admin comments.

NFR 015 Automate scheduling of periodic reviews and audit + 2.7.10 Notify stakeholders about upcoming reviews + 2.7.11 Maintain detailed audit trails for all dataset quality management activities.

1. Accessibility and Layout

Audit Logs Dashboard Accessibility:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- The Audit Logs dashboard must be accessible from the main portal navigation for both Data Holders and System Admin.
- Visual indicators (e.g., red text or warning icons) must highlight datasets overdue for review.

Audit Logs Dashboard Layout:

- The audit dashboard must display audit data in a tabular view with the following columns:
  - Dataset ID (linked to dataset overview)
  - Description

  - Status
  - Creation Date
  - Last Modified Date
  - Previous Review Date
  - Next Review Date.
- Provide sorting by "Next Review Date" and filtering by status.
- Include tooltips that explain scheduling logic (e.g., "Review due annually from last audit date").
- All the date should be in DD/MM/YYYY format
- Tooltips must be provided for:
  - Review frequency logic (e.g., "Review due annually from last audit date")
  - Status indicators.
- Pagination and Display Controls:
  - Default view must show 10 entries per page, with a dropdown to select 10, 25, 50, or 100 items per page.
  - Below the table, display summary text like: "Showing 1–10 of 75 entries".
  - Provide pagination controls: Previous, Next, and Page Numbers.
  - All controls must be keyboard-navigable and screen-reader compatible.

Notification Center UX:

- The Notification Center must allow the Data Holder Editor, Data Holder Admin (who approved the dataset metadata), and System Admin to view both upcoming and past notifications related to audits, reviews, and required actions. Display key fields: Notification Type, Target Dataset, Date Sent, and Action Required.
- Provide links to relevant dataset entries or review actions.
- Notifications should be clearly grouped and timestamped.
- A visual indicator must appear for unread notifications.

2. Performance Requirements

- Dashboard Load Time: Must not exceed two seconds when loading up to 1,000 audit entries
- Review Date Recalculation: Following an audit, the system must recompute "Next Review Date" in **≤ 1 second**.

- Notification Dispatch: All queued notifications must be sent within five seconds of the trigger event.
- Log Persistence: System must record and store audit events in logs within one second of execution.

3. Error Handling and Feedback

- If the audit dashboard fails to load or retrieve data:
  A friendly, non-technical error message must be shown (e.g., "Unable to load audit logs at this time. Please try again.").
- If email and in-portal notifications fail to send, then display a warning icon in the UI and notify the system admin via internal alert.
- All failed operations (e.g., update audit record, schedule change) must show a clear error message to the user and log the failure event for review by administrators.
- Partial updates or invalid submissions must be rejected with full field-level validation errors displayed inline.

4. Submission and Transaction Handling

- When audits are completed and marked as reviewed, a confirmation modal must appear summarizing the action before submission.
- Upon confirmation:
  - The review must be submitted as a single atomic transaction
  - Form elements must be disabled and a loading spinner shown during submission.
- After successful submission:
  - A toast or modal must confirm "Audit successfully submitted. Next review date updated."
- The system must automatically calculate the Next Review Date as one year from the last completed audit. IT should be recalculated immediately as and reflected in the UI.
- Notifications triggered by submissions must appear in the Notification Center and should be sent by email (if enabled) to the relevant Data Holder or Admin.
- System must queue and retry failed notification deliveries every 15 minutes for up to five times.

5. Audit Logging and Monitoring

- For every audit and related action (submission, review, notification, update), the system must log:

  - Dataset ID
  - User ID (Reviewer/Auditor)
  - Action Timestamp
  - Audit Outcome
  - Comments or required actions
  - Client IP address.

NFR 016 Enable export of dataset audit trail entries to CSV and Excel

1. Accessibility and Layout

- The "Export" function must be clearly visible within the audit dashboard interface, positioned near the top-right of the audit table or within the filter bar.
- The export action must be keyboard-accessible and fully compatible with screen readers (WCAG 2.1 Level AA).
- The Data Holder Editor must be able to view and download the audit logs of dataset metadata they have created or updated.
- The Data Holder Admin must be able to view and download the audit logs of all dataset metadata created within their organisation.
- The System Admin must be able to view and download the audit logs of all datasets.
- A modal or dropdown must allow the user to choose between CSV and Excel (XLSX) formats using accessible radio buttons or toggles.
- If filters are applied to the audit log view, the UI must explicitly state: "Export will include only the currently filtered results."
- If no audit trail records are visible or selected, the export button must be: Disabled or Trigger a tooltip or modal: "No records available for export. Adjust filters or selection."
- Button labels, tooltips, and progress indicators must be translatable and support localisation (e.g., date formats in DD/MM/YYYY).
- Export file naming conventions must be human-readable and follow the format: `audit-log-export-[YYYYMMDD]-[HHMM].csv/xlsx`.

2. Performance Requirements

- Export file generation must complete within:
  - two seconds for up to 500 records
  - five seconds for up to 5,000 records.
- The system must support simultaneous export requests from at least 30 concurrent users without degraded response time or server impact.

- The generated file must not exceed 20 MB; if it does, the system must prompt the user to refine the filters or selection.
- For paginated audit dashboards, the export must include all matching records, not just those on the current page.

3. Error Handling and Feedback

- If the export fails due to backend or file system errors, users must see a message like: "Export failed due to a system error. Please try again later or contact support."
- If the user selects an invalid combination (e.g., requesting export without records), a non-intrusive inline message or modal must be shown.
- During export generation:
  - Display a loading spinner or progress indicator with the text: "Generating file… please wait."
  - If processing exceeds 10 seconds, show a timeout warning: "Export is taking longer than expected. You may wait or try again later."
- Errors must be logged internally with full diagnostic details and retry options must be made available to the user on failure.

4. Submission and Transaction Handling

- Export operations must be non-blocking and must not interfere with ongoing navigation, filtering, or audit log updates.
- Exports must reflect the user's current view, including all active filters, sort order, and selections.
- Each export operation must be atomic — a file is either fully generated and valid, or the system returns an error without any download.
- Export requests must not alter any state on the server or audit entries themselves.
- While processing an export, the system must disable the export button to prevent duplicate submissions.
- Export outputs must maintain data consistency, and values must match what is shown in the UI, including:
  - Date format (DD/MM/YYYY)
  - Review status labels
  - Dataset links (as text).

5. Audit Logging and Monitoring

- Each export event must be recorded in the system's audit logs, capturing:

- User ID and role (Data Holder or System Admin)
- Timestamp of export request
- Format selected (CSV or Excel)
- Number of records exported
- Active filters or parameters applied (e.g., status = "Overdue")
- Operation result (Success/Failure)
- IP address and session ID.

NFR 017 Offer a search functionality for browsing and displaying metadata catalogue results

1. UI/UX Accessibility and Layout

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- Provide both a Basic Search Bar and an expandable Advanced Search Pane
- Display results in a structured layout, showing Dataset Title, Description, Data Holder, Created date, file format and category.
- Basic search should allow free text input with autocomplete and suggestions.
- Advanced search panel must support field-level search using the following filters:
  - Title (partial or exact match)
  - Description
  - Publisher
  - Licence
  - Category
  - Format
  - Created After/Created Before (calendar picker)
  - Updated After/Updated Before (calendar picker).
- Advanced search filters must:
  - Be collapsible to conserve screen space
  - Retain filter selections on search result pages
  - Be screen reader compatible and keyboard-navigable.
- Provide a "Search Tips" tooltip icon near the search bar, displaying guidance on how to use:
  - Exact match: Use quotes, e.g., "health statistics"
  - AND operator: Combine terms, e.g., `male AND female`
  - OR operator: Alternative terms, e.g., `male OR female`
  - Minus sign or NOT operator: Exclude terms, e.g., `vaccines NOT COVID`

  - Wildcards and partial matches.

2. Performance Requirements

Search execution for normal search and advanced must return results:

- For up to 10,000 datasets within 500 milliseconds
- For large queries (with filters and sorting) within one second
- Autocomplete suggestions must appear within 200 milliseconds of user input
- Filtering or sorting search results must not take more than 500 milliseconds per interaction.

3. Error Handling and Feedback

- If a user submits an invalid query (e.g., empty input or unsupported characters), the system must:
  - Display a contextual message such as: "Please enter a valid search term."
  - Highlight the input field requiring correction.
- If no datasets match the user's query:
  - Display a user-friendly message (e.g., "No results found. Try adjusting your filters or search terms.")
  - Provide a "Search Tips" tooltip to guide users on refining their queries (e.g., use of `AND`, `OR`, exact match).
- If the search service fails to return results due to internal errors, timeouts, or service unavailability:
  - Display a message: "Search service is temporarily unavailable. Please try again later."
  - Offer an optional retry button on the UI.
- If a search query takes longer than the defined performance threshold (e.g., three seconds), show a loading indicator with a timeout fallback message: "Search is taking longer than expected. Please wait or try again shortly."

4. Submission and Transaction Handling

- The system must:
  - Allow combined use of filters and text queries
  - Use logical AND between selected filters unless specified otherwise.
  - Support range filters for date fields
  - Normalise terms (e.g., case-insensitive, Unicode-insensitive) before query execution.

- Filters must dynamically update the result count without requiring full page reloads
- The system must:
  - Automatically perform fuzzy matching and partial matching for misspelled or incomplete queries
  - Normalise terms (e.g., singular/plural handling, casing) for broader matching
  - Offer synonym-based matching using controlled vocabulary or metadata tagging.
- If no datasets match the query:
  - Alert the user within 500 milliseconds
  - Suggest related terms or allow retry with adjusted query.
- Store user-specific search history (if logged in) to:
  - Auto-suggest past queries
  - Recommend filters based on prior activity.
- If a search fails, the system must:
  - Retry backend execution up to three times
  - Space retries at 30-second intervals
  - Display retry option on UI if unresolved.

5. Audit Logging and Monitoring

- For each search interaction, the system must log:
  - Timestamp
  - Search query string (anonymised for unauthenticated users)
  - Filters applied
  - Number of results returned
  - Duration of query processing.
- Only authorised admin users may access audit logs via the monitoring system.

NFR 018 Enable selection and export of dataset metadata to CSV or Excel formats

1. UI/UX Accessibility and Layout

- The "Export" function must be clearly visible and accessible adjacent to the dataset search results.
- The export option must support both mouse and keyboard navigation, following WCAG 2.1 AA accessibility standards.

- The export format options (CSV, Excel) must be presented using accessible and descriptive labels within a dropdown or radio selection interface.
- Tooltips or brief explanatory text must be provided for the export feature to aid users unfamiliar with the function.
- The interface should notify users when no datasets are selected and guide them to make a selection to enable export.
- Icons and buttons associated with the export function must meet minimum contrast ratio requirements and be sized for easy interaction on both desktop and mobile devices.

2. Performance Requirements

- Export generation must complete within three seconds for up to 100 datasets selected.
- For larger exports (e.g., 500+ datasets), the system must process and provide feedback within five seconds, with a loading indicator or progress notification.
- System must support concurrent export requests from at least 50 users without performance degradation.
- Exported files must not exceed 20MB in size; if the export exceeds this limit, the user must be prompted to refine their selection.

3. Error Handling and Feedback

- If no dataset is selected, the Export button must download all dataset metadata currently visible on the dataset catalogue screen.
- If the system encounters an error during file generation (e.g., file corruption, format conversion failure), a user-friendly error message must be displayed: "Export failed. Please try again or contact support if the issue persists."
- Errors should not disrupt the rest of the interface; user must be able to retry without needing to refresh the page.
- The system must log all export-related errors with relevant metadata (timestamp, user ID, export format, number of records) for review.

4. Submission and Transaction Handling

- Export actions must be handled as non-blocking operations and should not require page reload.
- The system must validate that datasets selected for export are still available (i.e., not removed or restricted) before beginning the export process.

- Each export transaction must be treated as atomic — either a valid file is successfully generated and downloaded, or a clear error message is returned with no partial output.
- The export operation must not trigger server-side changes to dataset metadata or user state.

5. Audit Logging and Monitoring

- Every export operation must be logged with:
  - User ID or session ID
  - Timestamp
  - Dataset IDs included in the export
  - File format (CSV or Excel)
  - Operation status (success/failure).

NFR 019 Allow saving and sharing relevant searches

1. UI/UX Accessibility and layout

Accessibility and Usability:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- The "Save my search criteria and Descriptions" buttons must be clearly visible, accessible, and screen reader compatible (WCAG 2.1 AA).
- Tooltips or contextual help must explain saving and sharing functionalities.
- A modal or side panel must allow users to name saved items and add optional notes.
- All actions (save, share, delete) must be accessible via keyboard navigation.

Saved Searches Layout:

- Prompt the user to enter a custom name and optional description for the saved search.
- Show a “Save” and “Cancel” button.
- Auto-focus the name field on open.
- The personal profile section must include a dedicated “Saved Searches and Descriptions” tab or section.
- Each saved item must display the name, timestamp, number of results (for searches), and status (e.g., active/archived).

- The interface must allow re-execution, renaming, deletion, and editing of saved criteria.
- The system must provide the Data User with the option to enable or disable notifications/alerts for their saved searches.

Sharing UI:

- "Share" buttons must be present next to each saved item and search result
- The system must display a popup with a URL that can be copied.

2. Performance Requirements

- Generating and displaying saved items must not exceed:
  - 300 milliseconds for up to 50 saved searches/descriptions
  - 600 milliseconds for up to 500
  - Shareable links must be generated in under 500ms.
- Re-running a saved query must execute within the same limits as standard search (≤1 second)

3. Error Handling and Feedback

- If saving a search or dataset description fails (e.g., due to network issues, session expiry, or server error), the system must:
  - Display a clear, user-friendly error message (e.g., "Unable to save. Please check your connection or try again.")
  - Preserve any input data (e.g., search name, notes) to avoid data loss during retry.
- If generating a shareable link fails, the system must show a message indicating failure with suggested next steps (e.g., "Link could not be created. Please try again later.").
- If a saved search references outdated or deleted datasets show a warning in the result view: "Some datasets from this search may no longer be available."
- If generating a shareable link fails the display message "Link could not be created. Please try again later."
- If a shareable link is revoked and accessed the display "This link is no longer valid or has been disabled."

4. Submission and Transaction Handling

- Saved items must be persisted to the user's profile immediately upon save.
- Saved searches and dataset descriptions must be retained unless deleted by the user.

- Saved items must be user-specific and securely linked to the authenticated account.
- Sharable links must be:
  - Unique per saved item.
  - Valid unless manually revoked.
  - Usable by other users with equivalent search permissions (e.g., public datasets only).
- Re-executing a saved search must retrieve current, updated results without changing the original saved criteria.
- If datasets referenced in a saved description are deleted or archived, a warning message must appear upon access
- The data user who has created and saved the search should be able to edit the saved search
- The system must provide the Data User with the option to enable or disable notifications/alerts for their saved searches.

5. Audit Logging and Monitoring

- The system must log the following events:
  - Search Save Attempt (success or failure)
  - Search Share Attempt (success or failure)
  - Re-execution of a Saved Search
  - Deletion or Modification of Saved Searches
  - Access of Shared Link** (anonymous or authenticated).
- Each log entry must include:
  - Timestamp
  - User ID (if authenticated)
  - Action Type (save, share, delete, access)
  - Search Query Parameters
  - Result Count (if applicable)
    Error Code and Message (on failure).

NFR 020 Notify the applicant about updates on saved datasets

1. UI/UX Accessibility and layout

Accessibility:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).

- The system must provide the Data User with the option to enable or disable notifications/alerts for their saved searches. The notification must be sent to the users via email/in-portal only when it is enabled.
- Users must be able to enable/disable notification types from their profile settings (email, in-portal, both).
- Notifications must be accessible via a notification centre icon on the top menu of the portal interface.

Notification UI:

- Updates must appear in reverse chronological order, grouped by saved search or dataset title.
- Each email and in-portal notification must include clear labels (e.g., "New Dataset Matching Your Search", "Dataset Description Updated") with time-stamps.
- Pop-up alerts must follow accessibility best practices and support screen readers.
- Notifications must provide direct links to the updated dataset or search result
- Include a UI element to "Mark all as read" and manage past notifications (archive/delete).

2. Performance Requirements

- The system must be able to handle sending notifications to at least 10,000 applicants concurrently with no more than two seconds delay per user.
- Portal notification rendering time must not exceed one second for up to 50 unread alerts.
- Batch multiple alerts for the same user in a 15-minute window to minimise alert fatigue.
- Detect relevant updates and trigger notification logic within five minutes

3. Error Handling Requirements

- If notification delivery fails (e.g., invalid email address), the system must:
  - Retry three times within a 30-minute window.
  - Log the failure and display a non-intrusive warning in the applicant's profile area. (e.g., "We couldn't deliver one or more notifications. Please check your email settings").
- If a system error prevents generation of alerts, fallback notifications must be queued and sent once normal service resumes.

- Notification templates must be validated before dispatch to prevent formatting issues.

4. Submission and Transaction Handling

- Notification subscriptions must be:
  - Created automatically when users save a dataset or search (with opt-out available)
  - Persisted securely and tied to the authenticated user account
  - Revocable at any time via the profile settings.
- Notification transactions must:
  - Be atomic (generated and queued as a single unit)
  - Include rollback safeguards if link generation or user context lookup fails.
- Batching logic must:
  - Prevent spamming by combining alerts within a 15-minute window
  - Label batched alerts accordingly: e.g., "3 new datasets matched your saved search 'COVID Reports'".
- The system must support multiple active alert subscriptions per applicant.

5. Audit Logging and Monitoring

- The system must log:
  - When a notification is generated
  - Delivery status (e.g., sent, read, failed)
    Recipient user ID and dataset/search ID involved.

NFR 021 Provide filtering and sorting capabilities by DCAT and HealthDCAT elements

1. UI/UX Accessibility and layout

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- The system must provide a filter sidebar on search result pages with collapsible filter sections (e.g., Publisher, Format, Category, Quality Label).
- Filters must use checkboxes, dropdowns, range sliders (e.g., for dates or size), and searchable multi-select where applicable.
- The sorting dropdown must be located above the search results, offering clear labels (e.g., "Sort by: Name | Date Submitted | Size").

- A "Clear All Filters" and "Reset Sort Order" button must be visible and accessible.
- The wizard and DCAT-AP search interface must be visually distinguished from basic filters and provide step-by-step navigation with back and next options.

2. Performance Requirements

- The system must support filtering and sorting across 10,000+ datasets with no more than two seconds latency in high-load scenarios
- Filters must support lazy loading and pagination for large option sets (e.g., data holders list with 1,000+ entries).

3. Submission and Transaction Handling

- The system must update search results within two seconds of a filter or sort input
- All filters and sort options must be applied client-side for cached results, and server-side for large queries (over 100 records).
- Filtering and sorting must support combinatorial logic: AND/OR operations based on user selection.
- If a filter option yields no result, the interface must clearly indicate this and allow users to remove that specific filter without refreshing the entire page.
- Partial query failures (e.g., metadata fetch errors) must fallback gracefully and display a warning message without breaking the entire search flow.
- The system must support querying by mandatory and optional HealthDCAT-AP elements, including:
  `dct:title`, `dct:description`, `dct:publisher`, `dct:license`, `dcat:theme`, `dcat:mediaType`, `dct:created`, `dct:modified`, `dct:spatial`, `dct:provenance`, and custom extensions such as `dqv:qualityMeasurement`.
- Wizard interface must ensure semantic validation of user inputs to match DCAT/HealthDCAT structure.
- The wizard must support multi-step form interactions, save-and-resume capabilities, and tooltips explaining metadata concepts.

4. Error Handling and Feedback

- If metadata elements are missing or malformed in a dataset, filters using those fields must exclude the affected dataset from that filter result.
- If a filter request fails due to invalid response then selectively remove that filter and retain other active filters

- If a filter request fails (e.g., due to service downtime), display a contextual error message:
  "We couldn't apply this filter at the moment. Please try again later or contact support."
- Logs of filter and sort errors must be recorded with user ID, query parameters, and timestamp.

5. Audit Logging and Monitoring

- The system must log:
  - Each filter and sort operation
  - The dataset count returned per interaction
  - The timestamp and initiating user session ID.

NFR 022 Allow to select and compare datasets

1. UI/UX Accessibility and layout

Comparison Interface Design:

- The User interface must support full keyboard navigation and screen reader compatibility (Web Content Accessibility Guidelines (WCAG) 2.1 Level AA).
- Applicants must be able to:
  - Select datasets from the search results or their saved list via checkboxes
  - Use a single "Compare Selected" button to launch the comparison interface
  - Toggle between table view and summary-only view for better usability.
- The comparison page must display datasets side-by-side in a tabular format with clearly labelled fields aligned horizontally.
- Differences must be automatically highlighted using a distinguishable colour (e.g., yellow or red) with tooltips or markers indicating the nature of the difference.
- Common values between datasets should be displayed in a neutral format to improve scan-ability.
- The comparison view must support shareable URLs that retain selected dataset IDs and comparison mode.
- A summary section must appear at the bottom of the page, summarising key differences (e.g., data quality, coverage, access).

- The comparison interface must be responsive with collapsible sections for improved navigation on smaller screens.

2. Performance Requirements

- Dataset comparison must render results within two seconds for up to three datasets of standard metadata size (~100 fields per dataset).
- For bulkier metadata entries, the system must implement asynchronous rendering or loading indicators if processing takes longer than two seconds.
- The system must support the comparison of up to three datasets simultaneously without performance degradation or layout distortion.
- The application must maintain consistent UI responsiveness during rendering, scrolling, or toggling between views.

3. Error Handling Requirements

- If selected datasets cannot be retrieved, the system must:
  - Show a contextual error message (e.g., "One or more datasets could not be loaded for comparison. Please try again or contact support.")
  - Log the failed dataset IDs for system monitoring and debugging.
- If a user tries to compare unsupported dataset types, the system must block the action and display a warning: "Selected datasets are incompatible for comparison due to metadata format differences."
- If a comparison attempt exceeds allowed limits (e.g., more than three datasets), prompt: "You can compare up to three datasets at a time."

4. Submission and Transaction Handling

- The system must dynamically align metadata fields for side-by-side display based on standard metadata schemas (e.g., DCAT, HealthDCAT-AP).
- Optional and extended metadata fields must be displayed only when they exist in one or more selected datasets, with empty placeholders where absent.
- When shareable URLs are generated, the system must securely encode dataset IDs and view preferences using encrypted or tokenised parameters.
- Dataset comparison state must be stored securely in the client session or passed securely via the URL, with no sensitive data exposed.

5. Logging and Monitoring

- The system must log the following for every comparison event:
  - User ID (if authenticated)
  - Timestamp of initiation and duration of interaction
  - Dataset IDs selected for comparison

  - Error events (e.g., load failure, format incompatibility)
  - View type used (tabular or summary).
- Logs must support analytics on:
  - Frequently compared metadata fields
  - Most common errors
  - User engagement with DQUL metadata.

NFR 023 Provide Feedback Functionality

1. UI/UX Accessibility and layout

Feedback Interface Design:

- The feedback form must comply with Web Content Accessibility Guidelines (WCAG 2.1 Level AA), including full keyboard navigation and screen reader compatibility.
- Users must be able to:
  - Open the feedback form via a clearly labelled "Feedback" button available in the global navigation or footer
  - Enter their Name (optional) and Email ID (mandatory) into clearly marked input fields
  - Select the Feedback Type from a dropdown with predefined categories
  - Provide detailed comments in a multi-line text area with character counter (e.g., 1,000 characters max).
- Input fields must include descriptive labels, placeholders, and accessible error messages.
- The form must use responsive design, ensuring usability on desktop, tablet, and mobile devices.
- A confirmation message must be displayed on successful submission (e.g., "Thank you for your feedback. We will review and get back to you if required.")
- The acknowledgement email must be sent automatically to the address provided, confirming receipt.

2. Performance Requirements

- The feedback form must load and render within two seconds under normal network conditions.
- Feedback submissions must be processed and stored within one second after the user clicks "Submit".
- The system must scale to handle simultaneous submissions from at least 500 concurrent users without degradation.

- The acknowledgement email must be dispatched within 30 seconds of submission.

3. Error Handling Requirements

- If the email field is missing or invalid, the system must:
  - Display a clear error message ("Please provide a valid email address to submit feedback.")
  - Prevent form submission until corrected.
- If mandatory fields (Email, Feedback Type, Feedback Content) are empty, the system must:
  - Highlight missing fields
  - Display contextual error messages (e.g., "Feedback content cannot be empty.")
  - Prevent form submission until corrected.
- If the system fails to submit feedback due to server or connectivity issues, a message must be shown: "We could not submit your feedback due to a technical issue. Please try again later."
- Failed submissions must be logged with relevant error codes and user-provided input (excluding sensitive details).

4. Submission and Transaction Handling

- All feedback submissions must be transmitted securely using HTTPS/TLS encryption.
- User-provided email addresses must be validated and stored in compliance with GDPR and national data protection requirements.
- Feedback entries must be stored in the nHDsC database with metadata including:
  - Submission timestamp
  - Feedback type
  - User email (where provided)
  - Status (open, in review, closed).
- The acknowledgement email must be generated using secure templates without exposing system details.
- System administrators must have access to a secure interface for reviewing and managing feedback.

5. Logging and Monitoring

- The system must log the following for each feedback submission:
  - Timestamp of submission

- Feedback type selected
- Submission status (success, failure)
- User ID (if authenticated) or anonymous session ID.

NFR 024 Error events (validation failures, submission errors).Other Non-Functional requirements

The following cross-cutting non-functional requirements must be applied to the entire metadata catalogue application to ensure operational robustness, reliability, and maintainability:

1. Backup and Recovery
   - Daily Backups: The system must perform automated full backups of the application database and metadata repository daily during off-peak hours.
   - Retention Policy: Backups must be retained for a minimum of XXX days (TBC - Number of days back-up should be retained) in encrypted storage compliant with applicable data protection regulations (e.g., GDPR).
   - Integrity Checks: Backup integrity must be verified weekly using necessary mechanisms.
   - Restore Time Objective (RTO): In the event of data loss or corruption, the system must be able to restore the latest backup within XX hours (This normally depends on the data present in system, so TBC on Maximum timings).

2. Version Control
   - Source Code Versioning: All application code and configuration files must be managed using a version control system (e.g., Git), with a clear branching strategy (e.g., main, development, feature branches).
   - Audit Trail Integration: Each metadata version change must be linked to an audit log entry for traceability and compliance.

3. Hosting
   - Hosting arrangements for the nHDsC are yet to be confirmed

4. Expected System Load
   - Concurrent Users: The application must support up to XXX (TBC- total number of expected users) concurrent users during peak usage without degradation in performance.
   - Metadata Volume: The system must scale to accommodate over XXX (TBC – Total number of Datasets expected) metadata of datasets with average

metadata complexity (100–200 fields), supporting advanced filtering, sorting, and search operations with <2 seconds response time.

- Notification Volume: The system must handle dispatching notifications (in-portal and email) to 10,000+ users simultaneously with <2 seconds latency per user.

5. Disaster Recovery

- Disaster Recovery Plan (DRP): A comprehensive DRP must be documented, tested quarterly, and include procedures for system restoration in case of:
  - Complete data center failure
  - Infrastructure compromise (e.g., ransomware)
  - Application-level corruption.
- Recovery Point Objective (RPO): Maximum acceptable data loss should not exceed 24 hours (i.e., last successful backup).
- Failover: The system must support failover to a secondary site or cloud-based recovery environment to ensure continuity with no more than four hours of total downtime.
- Communication Protocol: In case of a major incident, stakeholders must be notified via email and in-portal within one hour, and regular updates must follow every two hours until service restoration.

3. National Health Dataset Catalogue – Technical Specifications

This section presents the Technical Specifications. It includes a summary of the system's scope, the epics and functional requirements and offers recommendations that guide the technical design. It also details key architecture considerations, authentication needs, and how the National Health Dataset Catalogue (nHDsc) is to be structured to ensure it aligns with HealthDCAT-AP. The focus remains on the design and operation of the dataset catalogue itself, rather than the data described within.

3.1 Introduction

The framework for this technical specification is defined by legal provisions that outline the requirements for national dataset catalogues, with a specific focus on health data under the EHDS regulation. While the EHDS provides the primary legal and operational foundation for health metadata catalogues, its implementation is supported by broader EU data governance frameworks, such as the Data Governance Act (DGA).[22] To meet the aforementioned legal requirements, it is necessary to establish a minimum set of technical and operational capabilities. These include:

- 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.

While the EHDS regulation outlines the mandatory outcomes, these technical specifications also propose recommended practices to enhance the efficiency, usability and adaptability of national dataset catalogues.
These recommendations aim to ensure:

- effective governance and maintenance;
- scalability to address future technological and legal developments;
- to make additional developments relevant to national priorities.

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

3.2 Scope

The objective of this technical specification is to provide a detailed breakdown of the system specifications required to implement functionalities within each of the business requirements mentioned in 2.3.

This chapter is organised into individual sections, each corresponding to a distinct domain. The specifications address:

- **Involved Systems and Core Features:** Details the systems involved and the features that fulfil the domain's requirements.
- **Proposed Services and Technologies:** Outlines the services requiring implementation, their objectives, recommended technologies and their contribution to the domain's functionality.
- **Architecture:** Presents key architectural implementation considerations for the domain.
- **Data Flow:** Demonstrates how the described services can interact to achieve essential domain processes.

Each section will provide insights into possible technology stacks and implementation considerations.

3.3 Epic and Functional Requirements for nHDsC

With functional requirements mentioned in 2.4 are seamlessly transitioned into the epic narratives and user stories, following the agile SAFe[3] approach. Each epic represents a larger, overarching theme within this project, while functional requirements break these epics into digestible, actionable domains.

This section describes four epics and their functional requirements which should be implemented in a nHDsC to meet all the requirements.

#	Domain	Epic Title	User Stories
E1	Authentication	Access Management and Security	1. Allow for User identification, authentication, and authorisation, including the option to enforce multi-factor authentication (MFA)

[3] https://scaledagileframework.com

E2	Dataset Catalogue	Metadata Record Creation	2. Create, Read, Update and Delete (CRUD) dataset descriptions using the HealthDCAT-AP application profile 3. Add data quality and utility labels 4. Enforce minimum elements for dataset descriptions 5. Submit dataset description
E3	Dataset Catalogue	Metadata Record Publication and Update	6. Review dataset description and data quality and utility labels 7. Publish dataset descriptions using the HealthDCAT-AP application profile 8. Update dataset metadata descriptions and maintain change history 9. Maintain Versioning of updated Dataset metadata 10. Enable Rollback to Previous Version of Dataset Description when needed (with Admin Approval) 11. Automate scheduling of periodic reviews and audits 12. Notify stakeholders about upcoming reviews 13. Enable export of dataset audit trail entries to CSV or Excel formats 14. Maintain detailed audit trails for all dataset quality management activities.
E4	Dataset Catalogue	Searching for a Dataset	15. Offer a search functionality for browsing and displaying metadata catalogue results 16. Enable selection and export of dataset metadata to CSV or Excel formats 17. Allow saving and sharing relevant searches 18. Notify the applicant about updates on saved datasets 19. Provide filtering and sorting capabilities by DCAT and HealthDCAT elements 20. Allow to select and compare datasets.

3.4 Architecture Considerations

The architecture employs a microservices design, where each microservice operates independently and interacts with other microservices either directly through API communication or asynchronously via a message broker.

The microservices architecture was selected for several key advantages:

- Independent deployment capabilities, allowing services to be updated or modified without affecting the operation of other services.
- The flexibility to utilise different technologies based on specific requirements. For example, systems requiring cost estimation calculations may benefit from Python's scientific libraries.

The system requires a user interface to facilitate interaction between Data users, Data Holders and System Admin. To avoid duplicating the user interface solution across each domain, the proposed solution will be described in the following section.

User Interface	
Objective	Provision of a user interface, a single access point, so that: ▪ Data Holders must be able to: – Create data description and add required quality and utility labels – Update and delete data set description – Submit the new or updated data descriptions for review and approval. ▪ Admin should be able to: – Review Data holders and Data Users access request and approve access request of trusted data holders – Review the data added by Data holders, validate whether it is as per EHDS – Approve and publish the data description. ▪ Data Users should be able to: – Search for required data – Save and share relevant search details – Filter and sort the data based on their requirement – Compare different datasets.
Proposed Solution	A component-based front-end framework such as **Vue, React** or **Angular**.

	By selecting a component-based front-end framework, the system will be scalable adding as many features as desired and can be modularised based on user roles.
Technology Selection Considerations	The front-end framework must be selected based on development team expertise. These frameworks can easily integrate with the backend microservices such as API Rest or GraphQL.

There is another common microservice, transversal for all microservice domains.

Notifications Microservice	
Objective	To provide notifications about status changes or further actions required, by email or within the user interface. The system must provide notifications to relevant users when key status changes occur: ▪ Data Holders should be notified when: – The status of their application access request changes – A dataset description they submitted is approved, published, or rejected – There are any scheduled reviews or periodic audits, audit results and required actions. ▪ Admin should be notified when: – A dataset description, including data quality and utility labels, is created and submitted for review. ▪ Data Users should be notified when: – New datasets matching their saved search criteria are added – Updates are made to datasets that fall within their saved search criteria.
Proposed Solution	AWS Simple Email Service (SES) or its own government-specific solutions ensuring secure communication protocols and compliant with GDPR, making it suitable for public sector use.
Features Covered	Notifications Communication: Sends notifications by email.

Technology Selection Considerations	The selection of AWS SES depends on the government, if the government already has a service acting for this purpose, then it will be re-used.
Integrations	This microservice integrates with a BPM for workflow orchestration, when it needed for any communication from the rest of microservices.

A reverse proxy serves as an essential cross-cutting component across all domains, enabling client communication with the various microservices. This architectural element requires careful consideration during implementation:

Routing Microservices (API Gateway/Reverse Proxy/Load Balancer)	
Objective	To manage incoming requests and ensure load distribution among microservices.
Proposed Solution	A **reverse proxy** component to accept the requests from a client and forward it into the different microservices and user interfaces, which can fulfil the request, returning the response to the client that performed the request. This selected component is the **httpd-proxy**, it is based on nginx and acts as a reverse proxy, responsible for serving the requests thought the front-end and back-end services.
Features Covered	▪ Api Management: Orchestrate API requests between different microservices.
Technology Selection Considerations	The selection of a reverse proxy will hide the backend infrastructure from client, reducing attack surface. It routes API requests consistently and supports efficient routing and microservices orchestration.

All microservices require comprehensive observability to ensure request traceability, error logging, action auditing and performance monitoring. This observability stack, serving as a critical cross-cutting component across all domains, must capture both microservice and infrastructure performance metrics:

Observability	
Objective	To monitor and audit distributed system operations, auditing the state of the system and ensuring compliance during the data application process. This microservice should include below system functionalities: ▪ Captures and logs dataset ID, dataset name, submission date, data holder's name, and associated data quality and utility labels. ▪ Captures and logs automated validation processes (including timestamp) and results of the validation (e.g., validated, not-validated, corrected, etc.). ▪ Captures and logs dataset ID, dataset name, publication date, data holder's name, and the final data quality and utility labels applied. ▪ 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").
Proposed Solution	▪ **Jaeger**: For tracking the trace of the requests between services. ▪ **Prometheus**: For monitoring metrics. ▪ **Loki**: For receiving logs and track the errors and map errors with a specific trace. ▪ **Grafana**: A user interface to centralize all these backend services.
Features Covered	▪ Observability: Manage traces, logs and metrics of the microservices. ▪ Audit Trail: Track logs for auditory purposes on actions performed by each microservice.
Technology Selection Considerations	Open Telemetry is the standard to be used for the monitoring and observability stack. The backend observability services specified are suggestions, but Jaeger can be replaced by Zipkin or Loki can be replaced by Tempo, depending on expertise of the DevOps team.

3.5 Authentication

This section details the specifications for User authentication, to all Data users and Data holders to login to nHDsC. It ensures secure login to application.

Authentication	
Epics Mapping	▪ E1 Access Management and Security
Involved Systems	▪ nHDsC
Description Core Features Covered	**nHDsC:** ▪ **User authentication, identification and authorisation** 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.

3.5.1 Proposed services

The architecture employs a microservices design pattern. For this domain, we propose specific solutions and technology recommendations.

Identity and Access Management(IAM)	
Objective	To ensure secure access though robust user authentication and authorisation, supporting Multi-Factor Authentication.
Proposed Solution	Leverage **Keycloak**, an open-source identity and access management solution, commonly used with **Piveau** for handling authentication and authorisation. Support for national identity services or **Electronic Identification, Authentication and Trust**

	Services (eIDAS) integration can be enabled via federated identity providers (e.g. SAML2, OIDC).
Features Covered	▪ Authentication and Authorisation: – Role-based access control (RBAC) – User and group management – Single Sign-On (SSO) – Federation with external identity providers – Support for MFA (e.g. TOTP, WebAuthn).
Technology Selection Considerations	▪ **Keycloak** is well-integrated into Piveau for IAM needs. ▪ Supports extensibility to connect with national or European identity frameworks. ▪ Evaluate the compliance of chosen identity provider with **eIDAS** Level of Assurance (LoA). ▪ Check for existing SAML or OIDC connectors with national eID services.
Integration	Supports integration with national identity frameworks or eIDAS Node for cross-border identity via standard protocols (OIDC/SAML). APIs and configuration options to support external trust frameworks and user federation.

3.5.2 Integrations

Piveau:
The portal integrates with **Keycloak**, the identity and access management component commonly used in the Piveau ecosystem. Keycloak supports user authentication, role-based access control, and federated identity with national and cross-border identity providers (e.g. eIDAS). This enables secure login for both Data Holders and Data users within the nHDsC.

3.5.3 High Level Data Flow

1. Data Holder/Data User accesses the nHDsC.
2. Data Holder/Data User is redirected to the integrated Keycloak service for authentication.
3. Keycloak authenticates the user via:

- Internal user store
- Federated identity provider (e.g., national eID or eIDAS Node via OIDC/SAML)

4. Upon successful authentication, Keycloak issues an access token and redirects back to the nHDsC with user roles.
5. User session is established, enabling access to permitted portal features (e.g., Metadata record creation).
6. Similarly, Data user accesses the portal and is authenticated via the same Keycloak mechanism.
7. Authorisation rules defined in Keycloak and or the application backend control access to data search, view, or download functionality.
8. Audit logs and session handling are managed centrally by Keycloak to ensure security compliance.

3.6 Dataset Catalogue

This section details the specifications for creating, modifying and publishing dataset metadata, whilst also defining specifications for dataset catalogue selection. It ensures accessibility and adherence to quality standards for dataset catalogue descriptors.

The proposed solution incorporates Piveau, an open-source platform for HealthDCAT-AP management, which enables interoperability across systems.

This domain encompasses three primary processes: metadata record creation, publication in national dataset catalogues, and dataset discovery. The latter includes search, filtering and selection capabilities for research purposes.

The solution prioritises security measures for both data holders creating metadata descriptors and data users discovering and selecting datasets.

Dataset Catalogue	
Epics Mapping	▪ E2 Metadata Record Creation ▪ E3 Metadata Record Publication and Update ▪ E4 Searching for a Dataset
Involved Systems	▪ nHDsC

Description Core Features Covered	**nHDsC:** ▪ **Create dataset Description** Allows health data holders to create detailed descriptions of their datasets, including source, scope, characteristics, and availability conditions, for publication in their national datasets catalogue. ▪ **Add Data Quality and Utility labels** Allows health data holders to fill in Data Quality and Utility labels for their datasets. ▪ **Update dataset description** Allows health data holders to update the detailed descriptions of their datasets, at least annually. ▪ **Submission and Publication of Data description and dataset** Allows data holders to submit their data description for review to Admin after creating/updating the data description. Allows admin to review/validate the added data, approve and publish that data. ▪ **Display dataset description** Allows reading dataset descriptions in the national datasets catalogue. ▪ **Search dataset catalogue** Includes the ability to find and retrieve metadata catalogues and particularly metadata records at the national level. ▪ **Assessment of dataset descriptions and labels** Receives from health data holders the description of the dataset and data quality and utility labels, allowing for review of their accuracy. ▪ **Provide decision on dataset descriptions and labels** Allows approval or revocation of the dataset description and labels. ▪ **Aggregate datasets** Oversees the aggregation, description, and publication of health metadata in the nHDsC.

The following architectural proposal presents a comprehensive solution that details the interactions between interconnected systems, designed to meet the technical specifications of the dataset catalogues domain.

3.6.1 Proposed services

When implementing the Piveau stack, its components (Piveau Hub, Piveau Metrics and Piveau Staging) are deployed independently within a containerised environment. This approach enables autonomous operation, whilst custom microservices interact with these components via their exposed APIs.

The microservice descriptions for this domain are as follows:

Dataset description Microservice	
Objective	To provide data holders with features for creating, reading and uploading dataset descriptors and data quality and utility labels.
Proposed Solution	Microservice core features must be implemented with **Spring Boot** or **.Net Core** communicating with the **Piveau Hub** component via Restful API exposed by this component. Piveau Hub uses **Elastic Search** database to persist metadata catalogues and our microservice relies on Piveau Hub to manage and index it.
Features Covered	Metadata Management: Manage Metadata entries, handling dataset descriptions.
Technology Selection Considerations	The selection of Spring Boot or Net Core depends on the development team expertise, both technologies are suitable for achieving the functionalities and will communicate with the Piveau Hub component. **Piveau Hub** acts as the main point for metadata management, handling dataset descriptions and indexing them for easy retrieval and search. **Elastic Search** is the database used by the Piveau Hub to index the dataset entries and allowing fast retrieval of the metadata records.
Integration	Depends on the Piveau Hub and Elastic Search because is a component used to delegate the dataset metadata description management.

Data Validation and Approval Microservice	
Objective	To temporarily persist metadata entries for validation before publication in the National datasets catalogue. It prepares, validates and processes datasets, ensuring these datasets meet quality and compliance standards before publishing them for discoverability.
Proposed Solution	**Piveau Staging** automates the assessment of the quality of the dataset based on configurable criteria, validating format, metadata completeness and other additional checks. After a successful validation, this component communicates with Piveau Hub for cataloguing. This component uses a **PostgreSQL** database.
Features Covered	Datasets Validation: Validates the quality of a dataset before publication.
Technology Selection Considerations	**Piveau Hub** – after submitting the descriptors of a dataset catalogue, **Piveau Hub** will communicate with the Piveau Staging for validating quality. If is a valid dataset, Piveau Staging will inform Piveau Hub that the dataset is ready for publication. **PostgreSQL** is the database used by the Piveau Staging.
Integration	Depends on the Piveau Hub and Piveau Staging for validating dataset catalogue before publication.

Metadata Record Publication Microservice	
Objective	To publish validated datasets to the national datasets catalogue
Proposed Solution	**Piveau Consus** is used as component that manages metadata publication **Piveau Hub** – after validating a dataset with Piveau Staging, **Piveau Hub** will communicate with the Piveau Consus to publish it into the national datasets catalogue. This component uses **Elastic Search** database.
Features Covered	Datasets Publication: Publishes a dataset in the national datasets catalogue.

Technology Selection Considerations	**Piveau Consus** manages dataset aggregation and facilitates synchronisation with the EU catalogue, ensuring that published datasets are consistent across national and at EU level. This component provides an API Rest to be used by other microservices. **Elastic Search** is the database used by the Piveau Consus.
Integration	Depends on the Piveau Hub, Piveau Staging and Piveau Consus for publishing dataset catalogue.

Data Standardisation Microservice	
Objective	To manage vocabulary, ontologies, ensuring that metadata descriptors are consistent.
Proposed Solution	**EC's Interoperability Test Bed SHACL Validator** is used as component that assesses vocabulary and terminologies for the descriptors, validating RDF data against SHACL shapes. Ensuring that only compliant datasets are stored and published. **Piveau Hub** – after validating a dataset with Piveau Staging, **Piveau Hub** will communicate with the EC's SHACL Validator for validation during dataset ingestion. Validation ensures metadata descriptors is compliant to legal and regulation.
Features Covered	Dataset Vocabulary Compliance: Validates datasets for legal and regulatory compliance.
Technology Selection Considerations	**SHACL Validator** ensures RDF datasets comply with SHACL shapes, critical for metadata interoperability, and DCAT-AP is based on RDF. This component provides an API Rest to be used by other microservices.
Integration	Piveau Hub is responsible for orchestrating dataset ingestion will call the REST API of the EC's SHACL Validator for validating dataset catalogue.

Search and Visualisation Microservice	
Objective	To discover and select in a user interface catalogues, by searching by specified keywords or advanced criteria fields, save and share the relevant search and compare the datasets.
Proposed Solution	**Piveau UI** is embedded within a frontend framework as standalone component, consuming the rest of Piveau components and Microservices. **Pivot-RP** (Research Professional), a platform often associated with Piveau, allows users to save their searches, enabling them to easily revisit those searches later and receive updates. **Piveau's pipeline** functionality allows users to define data processing workflows, including steps for cleaning, transforming, and potentially combining datasets. **Piveau's DQV** (Data Quality Validator) annotator helps assess the quality of individual datasets, including their structure, format, and potential interoperability issues.
Features Covered	▪ Datasets discoverability: Enables data users to search in the user interface the existing datasets. ▪ Enables data users to save and share relevant search. ▪ Enables data users to compare the datasets.
Technology Selection Considerations	**Piveau UI** has all the features for dataset discoverability, representation and selection and will be integrated with the frontend framework used in the rest of domains. **Piveau's pipeline** allows for some degree of comparison through the transformation and analysis of data. **Piveau's DQV** can be helpful in comparing datasets based on their quality and suitability for combined analysis.
Integration	Integrated with the frontend framework used in the portal.

These microservices will interact with the notification microservice, to notify Admin regarding the creation of dataset description and data quality and utility label by data holder, so they can review these details and make decision accordingly. A data holder will be notified about the status of the data added after the Admin's decision.

3.6.2 Integrations

The portal integrates with this Piveau components to support DCAT compliant metadata management and catalogue synchronisation.

The Piveau metrics software stack allows analysis of metadata with regards to the DCAT Application Profile for data portals in Europe (DCAT-AP) standard, which is based on the Data Catalogue Vocabulary (DCAT) developed by the W3C. Its specification for metadata records to meet the specific application needs of data portals in Europe while providing semantic interoperability with other applications on the basis of reuse of established controlled vocabularies (e.g. EuroVoc) and mappings to existing metadata vocabularies (e.g. Dublin Core, SDMX, INSPIRE metadata, etc.).

3.6.3 High Level Data Flow

1. Data Holder creates and submits dataset descriptors, including data quality and utility label with the Piveau Hub component.
2. After submission, Piveau Hub component invokes the Piveau Staging component to assess the descriptors of the metadata submitted.
3. Piveau Hub invokes EC's Interoperability Test Bed SHACL Validator and the RDF is assessed.
4. System Admin receive a notification that a dataset description and data quality and utility label have been created.
5. System Admin reviews the datasets descriptions (metadata completeness), data quality, and utility labels and submits for publication.
6. After this, the catalogue is ready for publication, Piveau Hub invokes Piveau Consus to publish it in the National Dataset Catalogues assigning a unique ID for the catalogue.
7. Data User authenticates in the nHDsC.
8. Data User search by using the Piveau UI component integrated in the portal and retrieves the desired catalogues.
9. Data User selects desired catalogues.

4. Glossary of Abbreviations

API	Application Programming Interface
AWS/SES	Amazon Web Services/Simple Email Service
BR	Business Requirement
CRUD	Create, Read, Update, Delete
CSV	Comma-Separated Values
DAAMS	Data Access Application Management System
DCAT	Data Catalogue vocabulary
DCAT-AP	Data Catalogue vocabulary Application Profile for data portals in Europe
DGA	Data Governance Act
DQUL	Data Quality and Utility Label
DRP	Disaster Recovery Plan
EC	European Commission
EHDS	European Health Data Space
EHR	Electronic Health Record
eID	Electronic Identification
eIDAS	Electronic Identification, Authentication and Trust Services
EN	English
EU	European Union
FR	Functional Requirement
GA	Irish (Gaelic)
GDPR	General Data Protection Regulation
HADEA	European Health and Digital Executive Agency
HDAB	Health Data Access Body
HealthDCAT-AP	Health-related extension of the DCAT application profile for sharing information about Catalogues containing Datasets and Data Services descriptions in Europe (DCAT-AP)
HIQA	Health Information Quality Authority
HRB	Health Research Board
HSE	Health Service Executive
IAA	Identification, Authentication and Authorisation
IAM	Identity and Access Management
ISO/IEC	International Organisation for Standardisation/ International Electrotechnical Commission
LoA	Level of Assurance

MFA	Multi-Factor Authentication
NFR	Non-Functional Requirement
nHDsC	National Health Dataset Catalogue
OIDC	OpenID Connect
OTP	One Time Password
Piveau's DQV	Piveau's Data Quality Validator
Pivot-RP	Pivot - Research Professional
RBAC	Role-Based Access Control
RDF	Resource Description Framework
RegEx	Regular Expression
RPO	Recovery Point Objective
RTM	Requirements Traceability Matrix
RTO	Restore Time Objective
SAML2	Security Assertion Markup Language
SDMX	Statistical Data and Metadata Exchange
SHACL	Shapes Constraint Language
SPE	Secure Processing Environment
SSO	Single Sign-On
TEHDAS	Towards a European Health Data Space
TOTP	Time-Based One-Time Password
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
UX	User Experience
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines
Wi-Fi	Wireless Fidelity (wireless internet)
XLSX	Excel Open XML Spreadsheet

5. Glossary of Terms

4G	Fourth generation of cellular network technology.
API (Application programming interface)	A mechanism that enable two software components to communicate with each other using a set of definitions and protocols.
Application	A software program designed to perform specific tasks, such as accessing and managing health data.
AWS/SES	Amazon Web Services are a range of scalable cloud-based computing services. Simple Email Service is the cloud-based email service they provide which can be integrated into any application.
Components	Modular parts of a software system that can function independently or together, such as data management, or back-office.
Data Access Application Management System (DAAMS)	A system to receive, track and process applications and to issue data permits.
Data Governance Act	The Data Governance Act was the first deliverable under the European strategy for data and became applicable in September 2023. The Data Governance Act regulates processes and structures that facilitate voluntary data sharing. The Act is a cross-sectoral instrument that aims to regulate the reuse of publicly held, protected data.(22)
Data Maturity	Data maturity represents an organisation's proficiency and sophistication in managing and leveraging its data assets. The data maturity level is an indication of how well an organisation understands, values, and utilises its data throughout various processes and functions.(23)
Data Permit	A data permit is an administrative decision issued to a data user by a health data access body or data holder to process the electronic health data specified in the data permit for the secondary use purposes specified in the data permit based on conditions laid down in EHDS regulation.(24)
Data Quality	Data quality refers to the degree to which characteristics of electronic health data are suitable for secondary use.(24)

Data Quality and Utility Label	A data quality and utility label is a graphic diagram, including a scale, describing the data quality and conditions of use of a dataset.[24]
Dataset	Dataset means a structured collection of electronic health data.[24]
Dublin Core	The Dublin Core Metadata Element Set is a set of fifteen "core" elements (properties) for describing resources.[25]
Electronic Health Record (EHR)	A system whereby the software, or a combination of the hardware and the software of that system, allows personal electronic health data that belong to the priority categories of personal electronic health data established under this Regulation to be stored, intermediated, exported, imported, converted, edited or viewed, and intended by the manufacturer to be used by healthcare providers when providing patient care or by patients when accessing their electronic health data.[3]
European Health Data Space (EHDS)	The EHDS is an initiative by the European Union that aims to create a unified framework for health data across EU Member States.[24]
Health Data Access Body (HDAB)	A HDAB is a service that allows data users, such as researchers and policy-makers, to apply for access to health datasets to support research and innovation, education and training, policy-making, health service management and preparing national statistics. Under the EHDS, each EU Member State will be required to establish one or more HDABs.[24]
Health Data Holder	A data holder is a person or organisation who has the right or obligation (through control of the technical design of a product and related services, and the ability to make data available), to register, provide, restrict access or exchange certain data.[24]
Health Data User	A natural or legal person who has lawful access to personal or non-personal electronic health data for secondary use.[24]
HealthData@EU	HealthData@EU is the infrastructure connecting national contact points for secondary use of electronic health data and the central EU platform.[24]
HealthData@EU Pilot Project	The HealthData@EU Pilot project is a two-year long European project that aims to build a pilot version of the EHDS infrastructure, HealthData@EU.[24]

HealthData@IE	The Department of Health has been awarded EU funding under the EU4Health programme to support the establishment of health data access body services in Ireland. The grant-funded project, known as HealthData@IE, is being delivered in collaboration between the Department of Health, the Health Information and Quality Authority (HIQA) and the Health Research Board (HRB) Department of Health.[9]
INSPIRE metadata	Metadata that conforms to the standards and rules of the Infrastructure for Spatial Information in the European Community directive.
Interoperability	Interoperability is the ability of organisations, as well as software applications or devices from the same manufacturer or different manufacturers, to interact towards mutually beneficial goals. It involves the exchange of information and knowledge without changing the content of the data.[24]
Metadata	Metadata is information that describes other data. It helps to explain what the data is, how it can be used and where to find it.[26]
Metadata catalogue	A centralised system that organises and manages metadata, making it easier to search, access, and manage data-related information.
National Health Dataset Catalogue (nHDsC)	A collection of dataset descriptions, which is arranged in a systematic manner and consists of a user-oriented public element, where information concerning individual dataset parameters is accessible by electronic means through an online portal.[27]
OIDC	OpenID Connect is an authentication protocol designed to simplify user identity verification.
Open Data	Open data is data that is openly accessible, exploitable, editable and shareable by anyone for any purpose.[28]
Piveau	The Piveau platform is a large-scale, open-source open data management system.
Psuedonymisation	The process of replacing and identifying characteristics of data with a pseudonym, a value which does not allow the data subject to be directly identified.[29]

RegEx	Regular Expression is a sequence of characters that forms a search pattern.
SDMX (Statistical Data and Metadata Exchange)	An initiative designed to manage and automate the process of data and metadata exchange. Aims to improve the quality of exchanges through standardisation, automation, validation, and data sharing.(30)
SSO/IAM (Single Sign-On/Identity and Access Management)	SSO allows access to multiple applications with one set of credentials. IAM manages user identities and access rights.
TEHDAS	The first TEHDAS joint action which ran between February 2021 and July 2023 helped EU Member States and the European Commission to develop and promote concepts for the secondary use of health data to benefit public health and health research and innovation in Europe. It provided elements to the European Commission's legislative proposal on the European Health Data Space.(31)
TEHDAS2	Builds on the work of the first TEHDAS joint action to produce concrete guidelines and technical specifications for the European Commission and Member States to ensure a harmonised implementation of the EHDS regulation.(32)

6. References

1. European Commission. European Health Data Space [Online]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en. Accessed on: 19 August 2024.

2. European Commission. Europe's Digital Decade: digital targets for 2030 [Online]. Available from: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en. Accessed on: 19 August 2024.

3. European Commission. 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 (Text with EEA relevance) Available from: https://eur-lex.europa.eu/eli/reg/2025/327/oj. Accessed on: 05 March 2025.

4. European Commission. Frequently Asked Questions on the European Health Data Space Available from: https://health.ec.europa.eu/document/download/4dd47ec2-71dd-49fc-b036-ad7c14f6ed68_en?filename=ehealth_ehds_qa_de.pdf. Accessed on: 25/09/2025.

5. Peolsson M. Building the EHDS: the importance of national metadata catalogues Available from: https://tehdas.eu/news/building-the-ehds-the-importance-of-national-metadata-catalogues-interview-with-michael-peolsson-swedish-ehealth-agency/. Accessed on: 26/09/2025.

6. Derycke P. HealthDCAT-AP Specification Available from: https://metadata.healthdataportal.eu/dev.py?N=66&O=2731&titre_page=HealthDCAT-AP%20Specification&titre_chap=Tools. Accessed on: 26/09/2025.

7. Derycke P. HealthDCAT-AP and health standards Available from: https://metadata.healthdataportal.eu/dev.py?N=64&O=2462&titre_chap=Scope%20of%20HealthDCAT-AP&titre_page=HealthDCAT-AP%20and%20health%20standards. Accessed on: 26/09/2025.

8. Health Information Bill Available from: https://www.gov.ie/pdf/?file=https://assets.gov.ie/299641/edc11db0-effc-4bd9-bdc9-98e8a233e657.pdf#page=null. Accessed on: 19 August 2024.

9. Department of Health. HealthData@IE– National Project Implementing EHDS [Online]. Available from: https://www.gov.ie/en/collection/87d0e-healthdataie-national-project-implementing-ehds/. Accessed on: 19 August 2024.

10. Health Information Quality Authority. National Standards for Information Management in Health and Social Care Available from: https://www.hiqa.ie/sites/default/files/2024-06/National-Standards-for-Information-Management-2024.pdf. Accessed on: 13 February 2025.

11. Tehdas. TEHDAS' proposals for clarifying EHDS governance [Online]. Available from: https://tehdas.eu/tehdas1/results/tehdas-proposals-for-clarifying-ehds-governance/. Accessed on: 03 September 2024.

12. European Union. HealthData@EU Central Platform Available from: https://acceptance.data.health.europa.eu/healthdata-central-platform/home?locale=en. Accessed on: 26/09/2025.

13. European Commission. HealthDataEU GitLab Available from: https://code.europa.eu/healthdataeu. Accessed on: 26/09/2025.

14. Tehdas. Second Joint Action Towards the European Health Data Space – TEHDAS2 [Online]. Available from: https://tehdas.eu/. Accessed on: 19 August 2024.

15. Health Information Quality Authority. Catalogue of National Data Collections Available from: https://www.hiqa.ie/reports-and-publications/health-information/catalogue-national-health-and-social-care-data. Accessed on: 01 October 2024.

16. Department of Public Expenditure. Public Service Data Catalogue Available from: https://datacatalogue.gov.ie/. Accessed on: 26/09/2025.

17. Department of Public Expenditure. Ireland's Open Data Portal Available from: https://data.gov.ie/. Accessed on: 26/09/2025.

18. Department of Public Expenditure. Open Data Technical Framework Available from: https://data.gov.ie/pages/opendatatechnicalframework. Accessed on: 26/09/2025.

19. Department of Public Expenditure. Guide on How to 'Add a Dataset' Available from: https://data.gov.ie/uploads/page_images/2018-08-08-163838.353849Add-a-Dataset-to-Portal-Guide---August-2018.pdf. Accessed on: 26/09/2025.

20. European Health Information Portal. Data Quality by Quantum [Online]. Available from: https://www.healthinformationportal.eu/data-services/data-quality. Accessed on: 07 November 2024.

21. European Commission: Directorate-General for H, Food S. HealthData@EU central platform – Open-source release 5 – User manual: Publications Office of

the European Union; 2025. Available from: https://op.europa.eu/en/publication-detail/-/publication/72c69188-98f9-11f0-97c8-01aa75ed71a1. Accessed on: 01/10/2025.

22. European Commission. European Data Governance Act [Online]. Available from: https://digital-strategy.ec.europa.eu/en/policies/data-governance-act. Accessed on: 19 August 2024.

23. Health Data Agency. How to Analyse your Organization's Data Maturity [Online]. Available from: https://academy.hda.belgium.be/course/view.php?id=4. Accessed on: 25 September 2024.

24. European Commission. Regulation of the European Parliament and of the Council on the European Heath Data Space Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0197. Accessed on: 08 August 2024.

25. Dublin Core. Dublin Core Metadata Element Set, Version 1.1: Reference Description Available from: https://www.dublincore.org/specifications/dublin-core/dces/. Accessed on: 02/10/2025.

26. Health Data Agency. An Introduction to Metadata [Online]. Available from: https://academy.hda.belgium.be/course/view.php?id=25. Accessed on: 25 September 2024.

27. European Commission. A European strategy for data [Online]. Available from: https://digital-strategy.ec.europa.eu/en/policies/strategy-data. Accessed on: 19 August 2024.

28. Open Knowledge. Open Definition - Defining Open in Open Data, Open Content and Open Knowledge [Online]. Available from: https://opendefinition.org/od/2.1/en/. Accessed on: 20 August 2024.

29. Data Protection Commission. Guidance note: Guidance on Anonymisation and Pseudonymisation Available from: https://www.dataprotection.ie/sites/default/files/uploads/2022-04/Anonymisation%20and%20Pseudonymisation%20-%20latest%20April%202022.pdf. Accessed on: 13 February 2024.

30. European Union. SDMX explained Available from: https://ec.europa.eu/eurostat/web/sdmx-infospace/sdmx-explained. Accessed on: 02/10/2025.

31. Tehdas. TEHDAS Project Available from: https://tehdas.eu/tehdas1/project/. Accessed on: 02/10/2025.

32. Tehdas. TEHDAS2 joint action to promote the secondary use of health data in the EU starts [Online]. Available from: https://tehdas.eu/news/tehdas2-joint-action-to-promote-the-secondary-use-of-health-data-in-the-eu-starts/. Accessed on: 06 September 2024.

33. World Health Organisation. WHO Family of International Classifications (FIC) Available from: https://www.who.int/standards/classifications. Accessed on: 26/09/2025.

7. Appendices

Appendix 1. Steering Group Membership for HealthData@IE

Name	Organisation
Emer Doyle (Chair)	Health Information Policy, D/Health
Azul O'Flaherty	Health Information Policy, D/Health
Eoin Farrell	Health Information Policy, D/Health
Clodagh Thorne	Health Information Policy, D/Health
Lorna Gaffney-White	Health Information Policy, D/Health
Dr Barbara Foley	Health Information and Quality Authority
Rachel Flynn	Health Information and Quality Authority
Dr Kevin O'Carroll	Health Information and Quality Authority
Dr Teresa Maguire	Health Research Board
Bríd Burke	Health Research Consent Declaration Committee
Kathryn Kissane	Health Service Executive
Prof Richard Green	Health Service Executive
Loretto Grogan	Health Service Executive
Kerry Ryder	Health Service Executive
Denise Manton	National University of Maynooth, Ireland
Markus Helfert	National University of Maynooth, Ireland
Anthony Macken	Central Statistics Office
Dr Frances Burns	Northern Ireland Trusted Research Environment
Steve Dempsey	Irish Cancer Society (Patient Representative)
Kevin Devlin	National Research Ethics Committee/HRB PPI Reviewer (Patient Representative)
HIPU Team	Health Information Policy, D/Health
Elaine Murray	EIT Health Ireland and UK
Niall Sinnott	eHealth, D/Health
Eamonn Coyne	eHealth, D/Health
Ronan O'Kelly	eHealth, D/Health
Christopher Ryan	Research Services, D/Health
Laura Flannelly	Statistics and Analytics Unit. D/Health
Elizabeth McCrohan	Statistics and Analytics Unit, D/Health
John O'Neill	Research Policy and Innovation, D/Health
Róisín O'Neill	Research Policy and Innovation, D/Health

Appendix 2. Working Group for Work Package 6

Name	Organisation
Dr. Barbara Foley	Health Information and Quality Authority (HIQA)
Suzanne Barror	Health Information and Quality Authority (HIQA)
Fiona Dwane	National Cancer Registry of Ireland (NCRI)
Dr. Sarah Craig	Health Research Board (HRB)
Lorraine McNerney and Martin Troy (shared role)	Office of the Government Chief Information Officer (OoGCIO)
Loes Knaapen	The Irish platform for Patient Organisations, Science and Industry (IPPOSI)
Eamonn Coyne	D/Health
Loretto Grogan and Niall Halliday (shared role)	Health Service Executive (HSE)
Karen Kearns	Health Pricing Office (HPO)
Theresa Barry	Health Service Executive (HSE)
Sorcha Boyle	The Irish platform for Patient Organisations, Science and Industry (IPPOSI)
Ailish Kelly	D/Health
Helen Conroy	D/Health
Bríd Burke	Health Research Consent Declaration Committee (HRCDC)

Appendix 3. European Health Data Space - Categories of electronic data for secondary use [4]

	Categories	Examples for what is in scope	Examples for what is out of scope
Categories to be made discoverable by 26 March 2029	**Electronic Health Record (EHR) data**	EHRs contain a wide range of data about a patient's medical history, treatments, and outcomes generated by healthcare providers when providing treatment, such as diagnosis and problem list, medication lists and treatment plans.	EHR kept by a healthcare provider that qualifies as a micro-enterprise (unless that Member State extended the duty to make available data also to such entities.)
	Data on healthcare needs, resourcing, access, expenditure and financing	For example, resources allocated to healthcare covers data on the availability and distribution of healthcare resources, e.g. number of healthcare facilities (such as hospitals, clinics, nursing homes), number of healthcare professionals (e.g. doctors, nurses, GPs), availability of medical equipment and	Individual-level information on healthcare expenditure.

		technology. This is about aggregate-level non-personal data.	
	Pathogen genomic data	Collections of information on pathogens that can cause disease in humans, including bacterial, viral, fungal, parasitic or prion pathogens.	Pathogen data on pathogens only affecting animal health.
	Healthcare-related administrative data	Collections of information that is generated through the administration of healthcare services, typically used for billing, reimbursement, and healthcare management purposes.	Banking data such as account numbers related to reimbursement.
	Person-generated medical device data	Collection of health-related data that is generated by medical devices and kept by a health data holder.	Data stored locally on devices and not accessible from the outside.
	Wellness app generated data	Data from fitness tracker shared with health care providers or with the app developer.	Data stored locally on user device/in app without access by the developer or healthcare provider.
	Identification data on health professionals	For example, whether a treating physician referred to in an EHR is a general practitioner or a specialist (and if so, in what field).	Contact information of that health professional.
	Population-wide health data registries	Population-based registries are systematic collections of health-related data from a defined population. These registries are typically maintained by government agencies, health organisations, or research institutions to support public health decision-making, policy development, and healthcare planning.	
	Data from medical and mortality registries	Medical registries are systematic collections of data on patients with a specific disease, condition, or characteristic, such as transplantation registries containing collections of data on organ transplantation outcomes, including	

		patient characteristics, transplant procedures, complications, and graft survival rates. Mortality registries are a systematic collection of data on deaths, including information on cause, circumstances, and demographics, such as cause-of-death registries containing data on the underlying cause of death, including information on disease, injury, or condition.	
	Medical device-generated data	Data from pacemakers or other implanted medical devices held by manufacturer or healthcare provider.	Data stored locally on device without access by manufacturer or healthcare provider.
	Data from product and device registries	Collections of data on the use, safety, and effectiveness of medicinal products and medical devices, including two types of registries: Medicinal product registries: These registries collect data on medicinal products, including prescription and over-the-counter medications, vaccines, and biologics. Medical device registries: These registries collect data on medical devices, including implantable devices, diagnostic equipment, and software.	
	Biobanks and databases	Electronic health data kept by repositories of biological samples and associated health data, which are collected and stored for research purposes. These repositories can contain a wide range of biological samples as well as associated health data, such as medical records, lifestyle information, and environmental exposures. Examples: Population-based biobanks: collections of data on biological samples and health data from large populations, often for the purpose of	The biological samples themselves held by biobanks. EHDS secondary use is about the re-use of existing electronic health data, not about the generation of new data. EHDS secondary use cannot be used to request new analyses on biological samples (that would then generate new data).

		studying genetic and environmental factors that contribute to disease. Disease-specific biobanks: collections of data on biological samples and health data from individuals with specific diseases. Tissue banks: collections of data on human tissue samples for research purposes.	
Categories to be made discoverable by 26 March 2031	**Data impacting on health**	For example lifestyle factors analysis (smoking, alcohol consumption, surgeries, accidents...)	Detailed socioeconomic data collected outside healthcare settings, or purely environmental data not linked to health.
	Human genetic and genomic data	Human genetic data refers to the information contained in an individual's DNA, including their genes and chromosomes (e.g. genotyping data, genomic sequencing data, microarray data). Human epigenomic data refers to the information about the chemical modifications to an individual's DNA or histone proteins that can affect gene expression without altering the underlying DNA sequence (e.g. DNA methylation data, histone modification data). Human genomic data refers to the comprehensive information about an individual's genome, including their genetic and epigenetic data (e.g. whole-genome sequencing data, exome sequencing data, gene expression data).	
	Other human molecular data	Proteomic data from clinical research	Raw molecular data generated for non-health related purposes.
	Clinical trial data (after completion)	Data from completed clinical trials, investigations, and studies, subject to rules established in the legal acts governing them.	Data from ongoing trials, studies or investigations.

	Health research cohorts/questionnaires/surveys	Encompasses information collected from groups of individuals or populations to understand health-related phenomena, behaviours, or outcomes. These data are often used to identify risk factors, track trends, or evaluate the effectiveness of public health interventions. The requirement to declare such datasets applies after the first publication of result only.	

Appendix 4. HealthDCAT-AP and Health Standards [7]

Article 51 of the EHDS regulation expands the scope of health data to include additional domains that impact health, such as pathogen and environmental data. Many of these domains already rely on specific standards for describing, categorising, or modeling their data, often tailored to their unique purposes. Below is a brief, non-exhaustive overview of some of these standards and their focal areas:

Standards	Description
HL7 (Health Level Seven International)	This defines a set of international standards for the exchange, integration, sharing, and retrieval of electronic health information. HL7 standards provide a comprehensive framework for clinical and administrative data. Its primary scope is the exchange of individual clinical and administrative data elements (e.g. patient demographics, clinical observations). It is used to describe individual health records or transactions, not entire datasets. HL7's data models and messaging standards are HL7 V2, V3, and FHIR.
FHIR (Fast Healthcare Interoperability Resources)	This is a standard describing data formats and elements for exchanging electronic health records. Developed by HL7, it is designed to enable fast and efficient exchange of healthcare information. It uses modern web technologies and focuses on interoperability. FHIR focuses on specific elements like patients, observations, medications, and other clinical data points rather than on metadata for datasets as a whole.
OpenEHR	This is an open standard specification in health informatics that describes the management and storage, retrieval and exchange of health data in electronic health records (EHRs). Part of the OpenEHR

	framework, OpenEHR Archetypes are formal models or templates that define the structure, meaning, and relationships of health-related data in an interoperable and standardised way.
ICD (International Classification of Diseases)	This is a globally recognised standard, maintained by the World Health Organization (WHO), for coding diseases and health conditions. It provides standardising classification codes for diseases and health conditions. ICD codes and descriptions can be used to standardise the classification of health-related datasets in HealthDCAT-AP.
LOINC (Logical Observation Identifiers Names and Codes)	This is a universal standard for identifying health measurements, laboratory observations, and clinical data. LOINC codes can be used to describe lab tests, measurements, and other clinical observations in HealthDCAT-AP.
SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms)	This is a comprehensive clinical terminology that provides codes, terms, synonyms, and definitions used in clinical documentation and reporting such as diseases, clinical findings, and procedures. SNOMED CT can be utilised to describe clinical concepts and healthcare terms in HealthDCAT-AP.
The ISO/IEC 11179	This standard provides guidelines for metadata registries, including the registration and management of metadata for data elements. It offers a structured approach to define and manage metadata elements, which can be applied to health datasets.
OMOP (Observational Medical Outcomes Partnership)	This Common Data Model standardises the format and content of observational health datasets (i.e. clinical observations, treatments, and outcomes data).
CDISC (Clinical Data Interchange Standards Consortium)	This standards facilitate the exchange of clinical trial data and include models like CDASH (Clinical Data Acquisition Standards Harmonization) and SDTM (Study Data Tabulation Model).
SDMX (Statistical Data and Metadata Exchange)	This is an international initiative that aims to standardise the exchange of statistical data and metadata. It is used by statistical organisations to describe and exchange entire datasets and their metadata. It is the foundation for StatDCAT-AP.
DICOM (Digital Imaging and Communications in Medicine)	This is a standard for the handling, storing, printing, and transmitting information in medical imaging.
MeSH (Medical Subject Headings)	This is a comprehensive controlled vocabulary used by the National Library of Medicine (NLM) to index and organise biomedical and health-related information in databases like PubMed and MEDLINE. MeSH terms facilitate precise and consistent search and retrieval of scientific

	and medical information by categorising content into hierarchical topics and subtopics. This system includes descriptors, qualifiers, and supplementary concept records to cover various aspects of medical knowledge, ensuring that researchers, healthcare professionals, and librarians can find relevant information efficiently.
GSIM (Generic Statistical Information Model)	GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information used in the production of official statistics (information objects). The framework enables generic descriptions of the definition, management and use of data and metadata throughout the statistical production process.
WHO classifications	The WHO Family of International Classifications and Terminologies includes:[33] • the International Statistical Classification of Diseases and Related Health Problems (ICD), • the International Classification of Functioning, Disability and Health (ICF), • and the International Classification of Health Interventions (ICHI). These Reference Classifications serve as the global standards for health data, clinical documentation and statistical aggregation.

HealthDCAT-AP has been designed to serve as a comprehensive framework for describing diverse health data that utilise various data models, exchange services, formats, and vocabularies, including thesauri and ontologies. It provides an integrated solution for managing data relevant to Article 51 within health dataset catalogues. The vocabulary of HealthDCAT-AP must be robust enough to describe the standards associated with the dataset in a machine-actionable way.

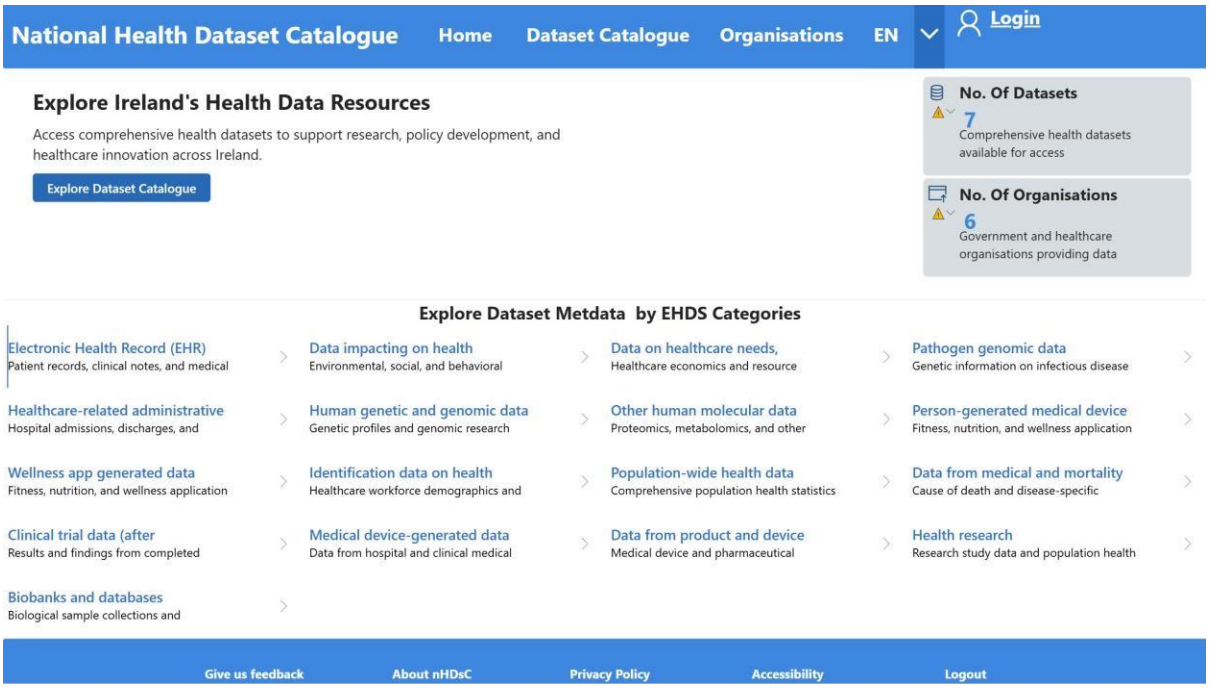
Examples of health Data Models, Services, Formats, and Ontologies for use in HealthDCAT-AP				
Data models	**Data services**	**Formats**	**Thesauri**	**Ontologies**
OMOP Common Data Model: A framework for transforming data from various sources (e.g. EHRs, claims data) into a common format. **OpenEHR:** Open standard for managing electronic health records (EHRs). **SDTM** (Study Data Tabulation Model) defines a standard structure for human clinical trial (study) data tabulations etc.	**FHIR API:** Standard APIs for exchanging healthcare data using RESTful principles. **OpenEHR API:** Facilitating the management and exchange of electronic health records. **DICOM Protocol:** Beyond being just a file format, DICOM also defines the communication protocol used to exchange medical images and associated information between medical devices, such as scanners, servers, workstations, and printers. This ensures that different systems and devices can communicate and interpret the data correctly.	**DICOM File Format:** The DICOM standard defines a file format for medical images, which includes both the image data (e.g. MRI, CT scans, X-rays) and associated metadata (e.g. patient information, imaging parameters). The file extension is typically .dcm. **FASTA** is a text-based format used for representing nucleotide sequences or peptide sequences (proteins) in bioinformatics. etc.	**ICD** is used globally for health management, epidemiology, and clinical purposes, providing codes for diseases, conditions, and procedures. **UMLS** integrates multiple health and biomedical vocabularies, providing a large compendium of healthcare-related terms and their relationships. **RxNorm:** A normalised naming system for generic and branded drugs **MESH** (Medical Subject Headings) etc.	**GO** (Gene Ontology) provides a controlled vocabulary to describe gene and gene product attributes across all species, focusing on biological processes, cellular components, and molecular functions. **OBO** (Open Biological and Biomedical Ontologies) **DOID** (Disease Ontology) etc.

Appendix 5. nHDsC – Prototype Screenshots with Functionalities

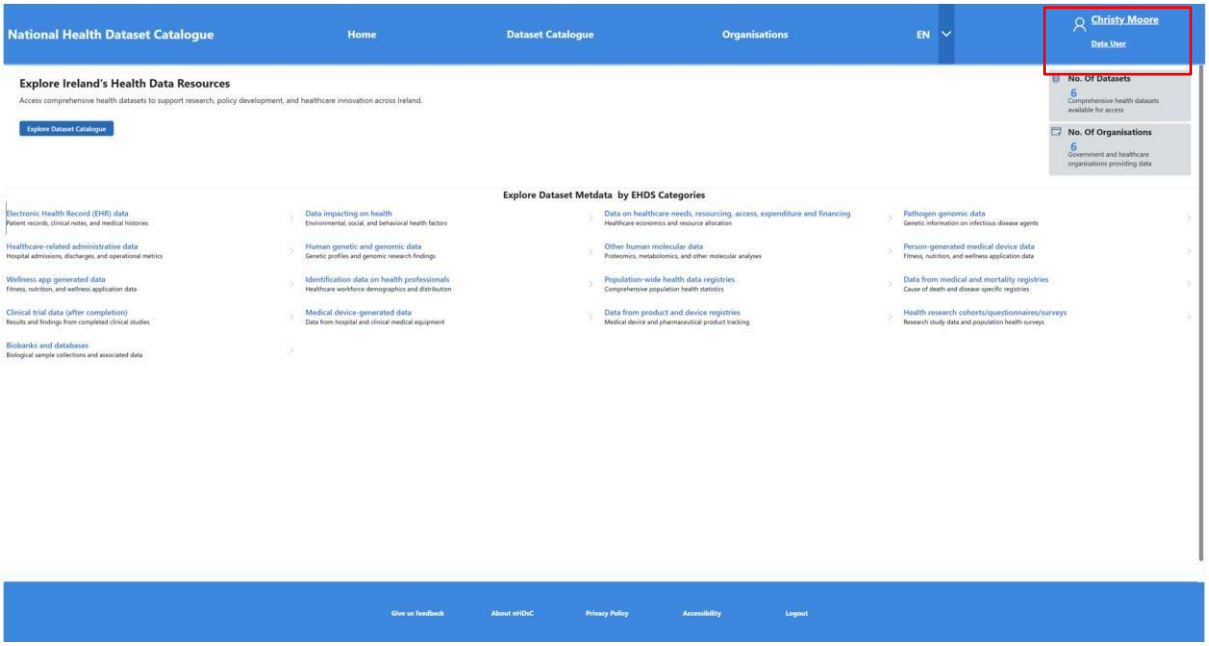
1.1 Functionalities for Data User

i. Home Page Overview

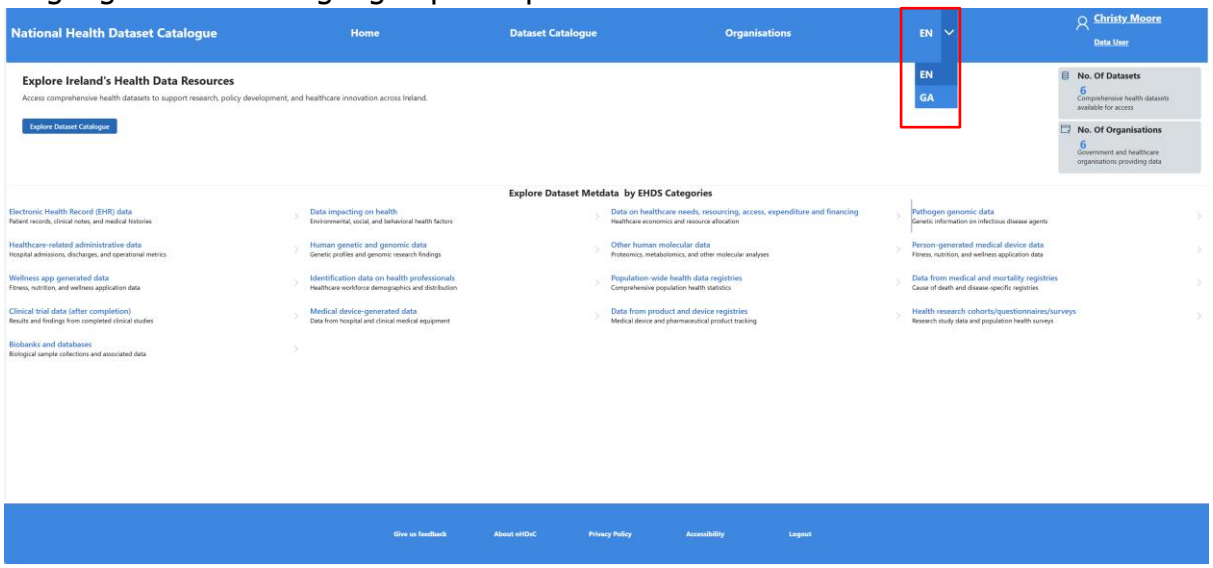
- When users access the application link, they land on the Home Page.



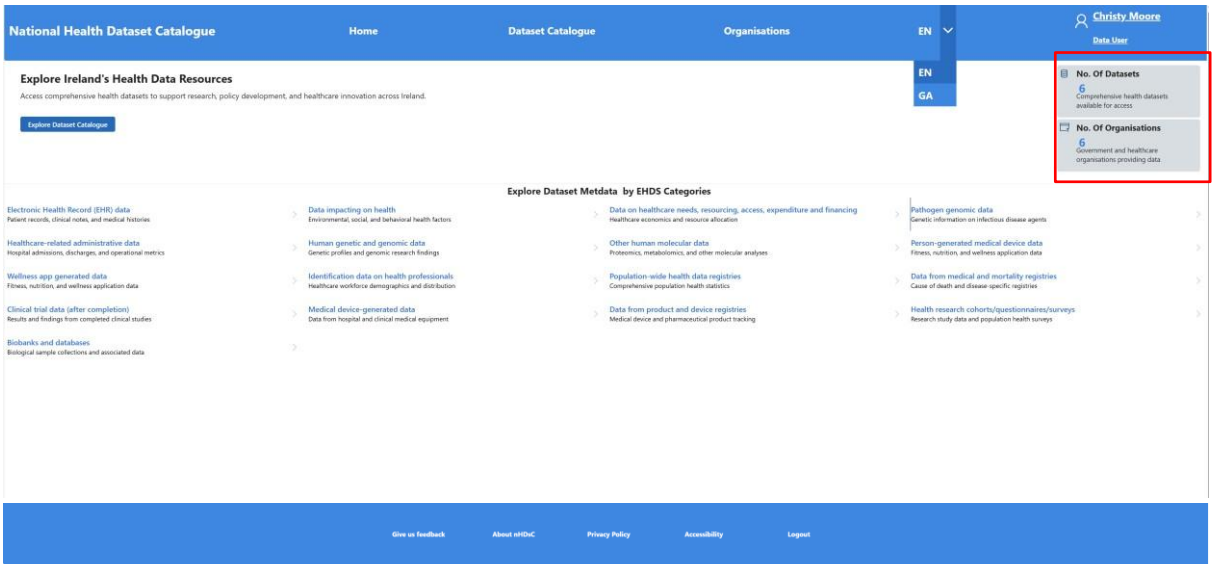
- Users can log in using the Login button. Once logged in, the Home Page reloads, displaying user details and role in the top-right corner.



- Language selection (English/Irish) is available. Users can select their preferred language from the language options provided.

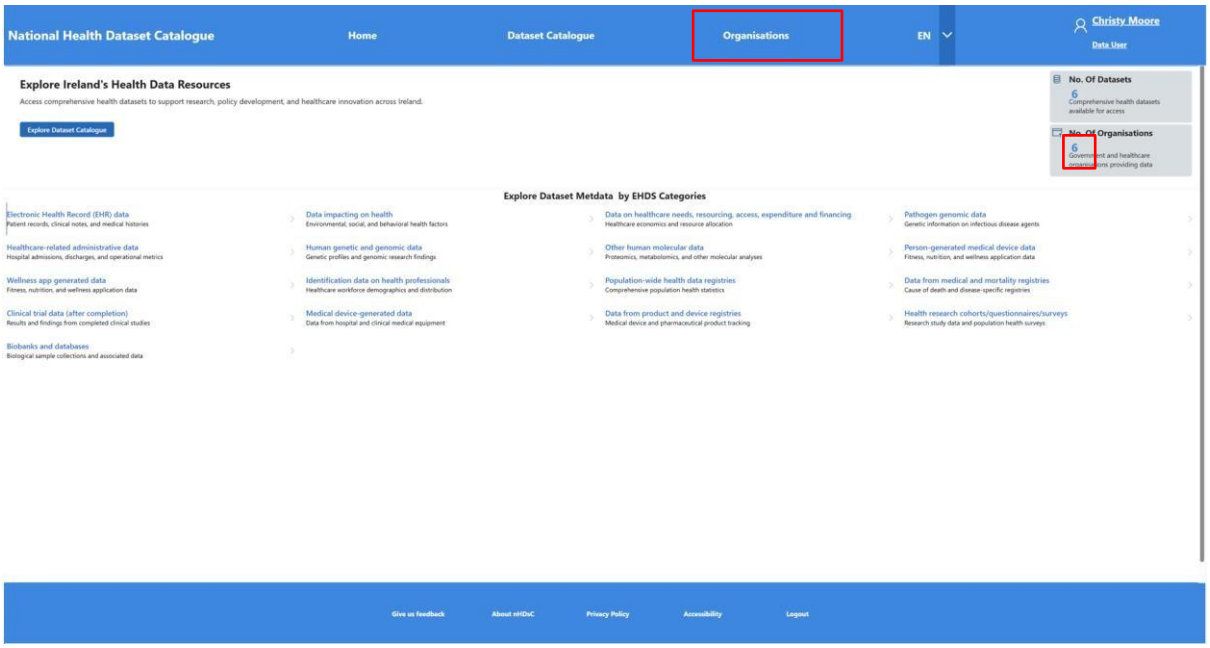


- The Home Page also displays key statistics, including:
  - The number of datasets currently discoverable through the nHDsC.
  - The number of organisations the Data Holder is associated with and enrolled in the nHDsC.

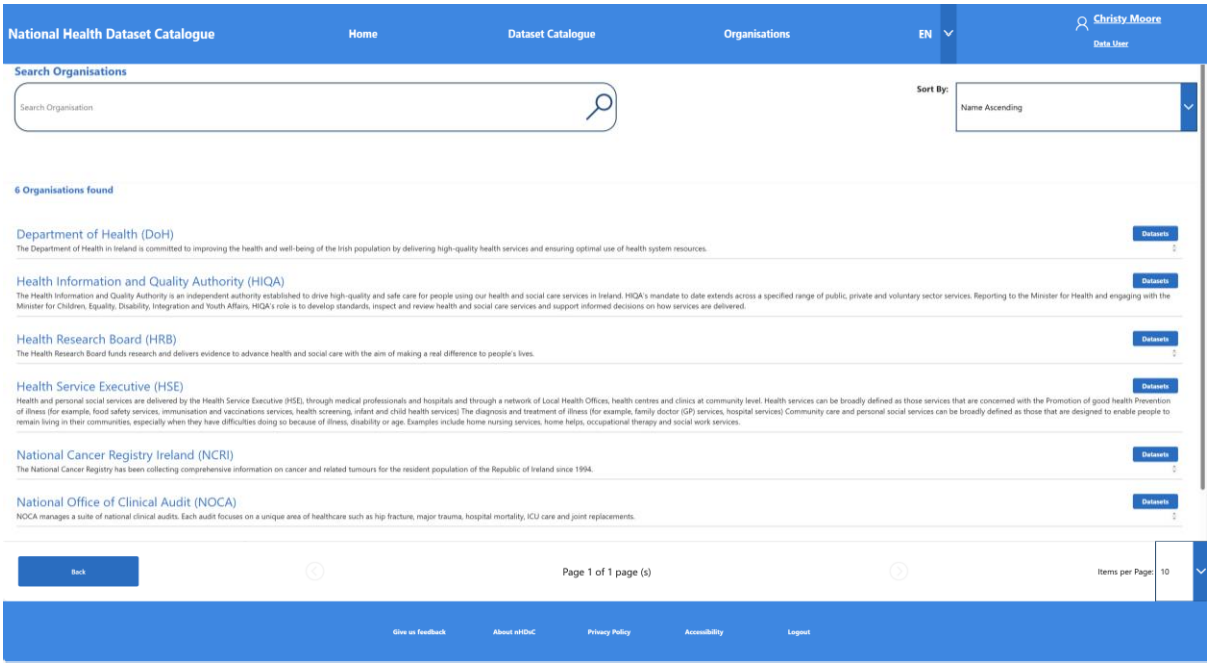


ii. Organisation List

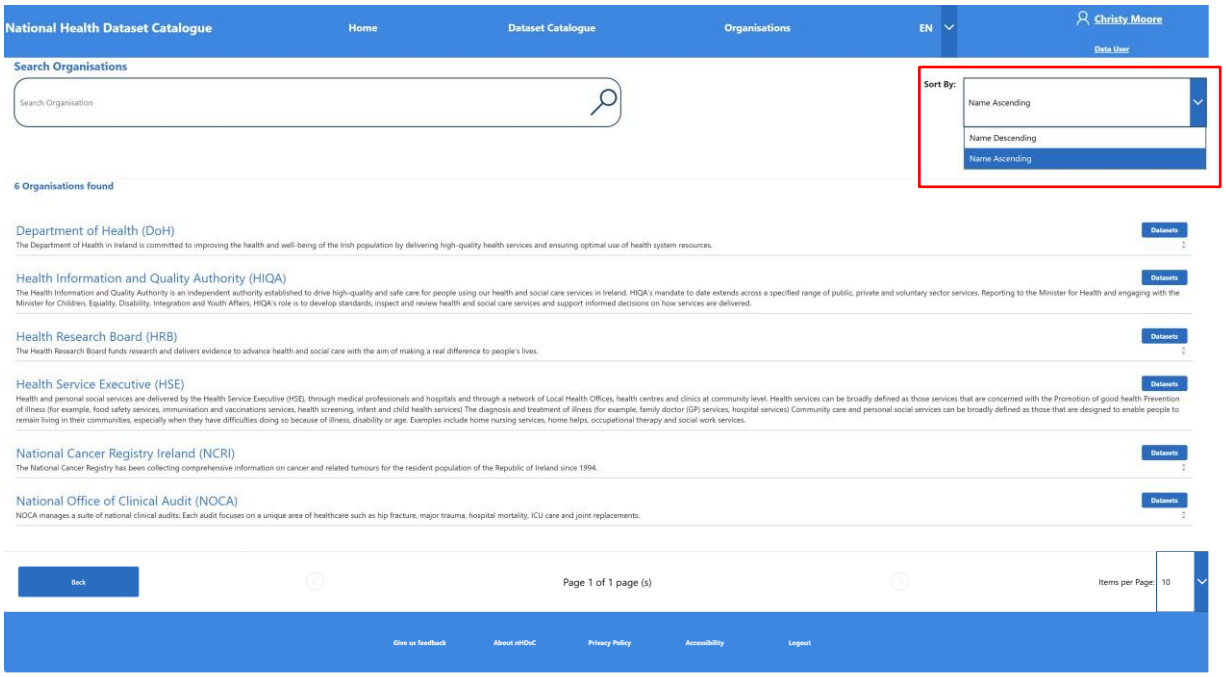
- Users can view a list of organisations by either:
  - Clicking the Organisation tab in the header, or
  - Clicking directly on the number of organisations shown on the Home Page.



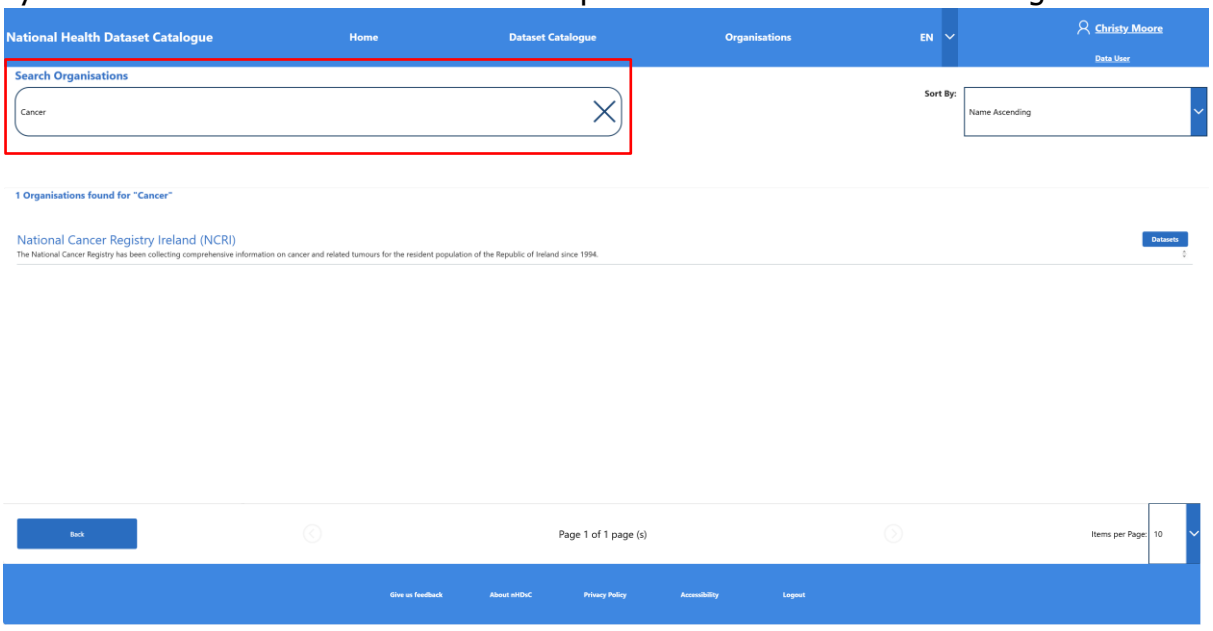
- Each organisation entry includes a title and description.



- Sorting functionality is available, allowing users to sort the list in ascending or descending order by organisation name.

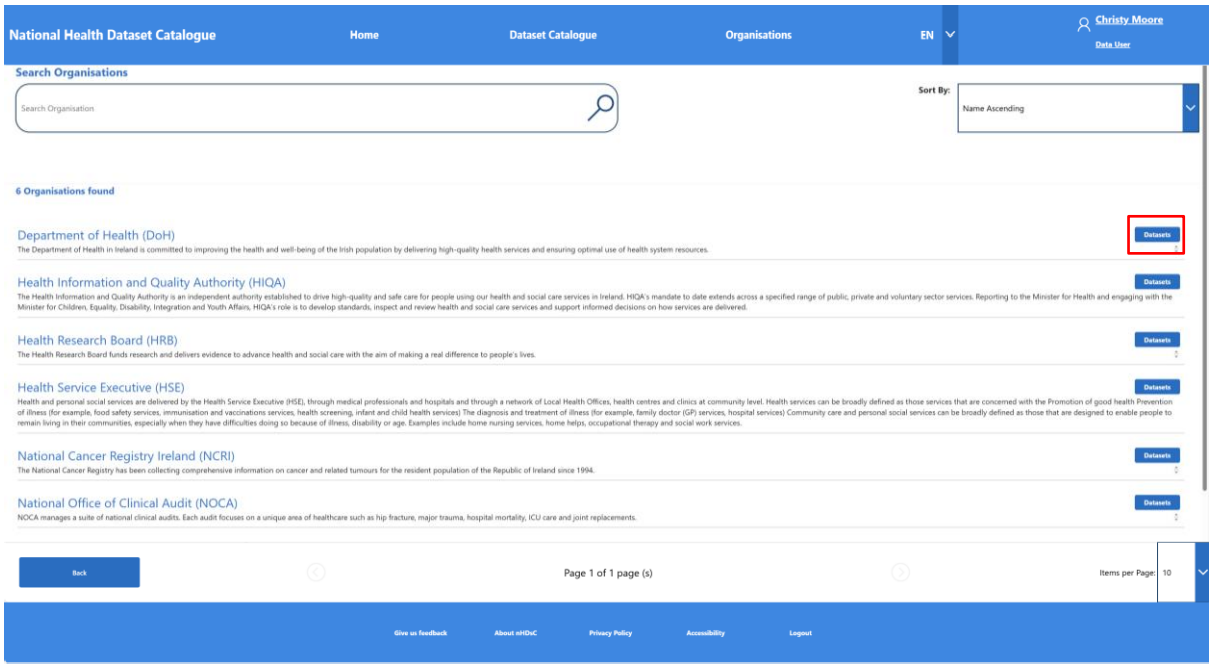


- A search box is also provided. By entering a keyword into the search box, the system searches both the title and description fields and returns matching results.

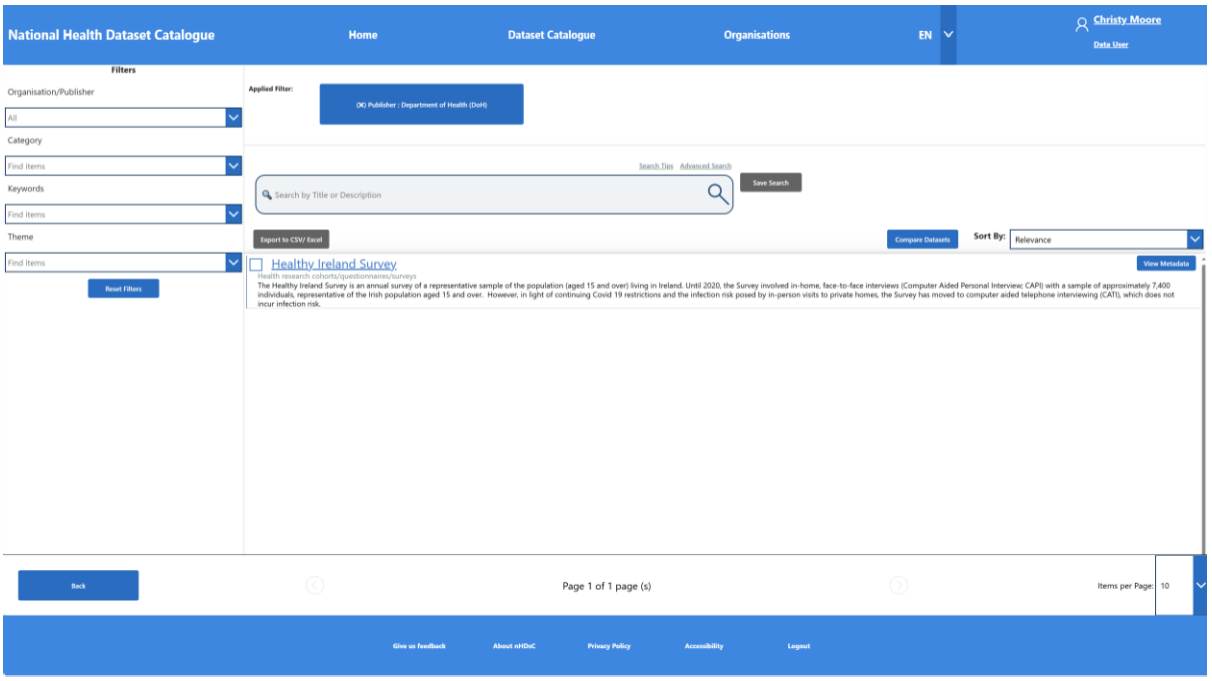


iii. Viewing Datasets by Organisation

- To view datasets published by a specific organisation, users click the Datasets button beside organisation name.



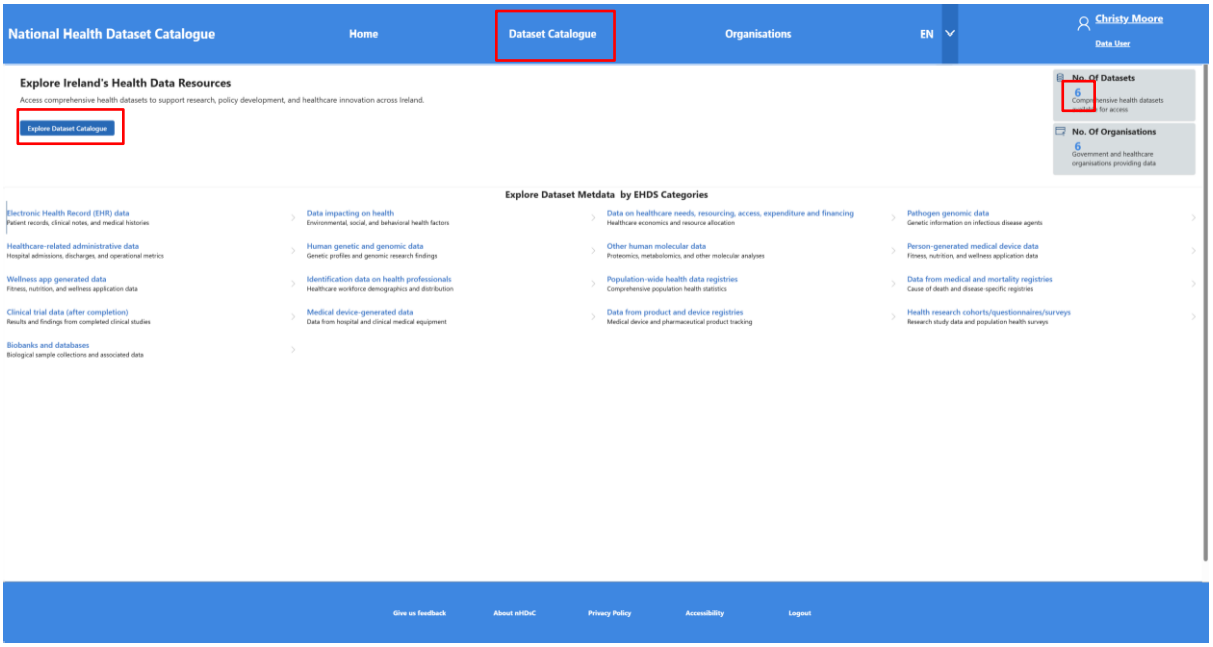
- This action opens the Dataset Catalogue Page, automatically filtered by the selected publisher/organisation.



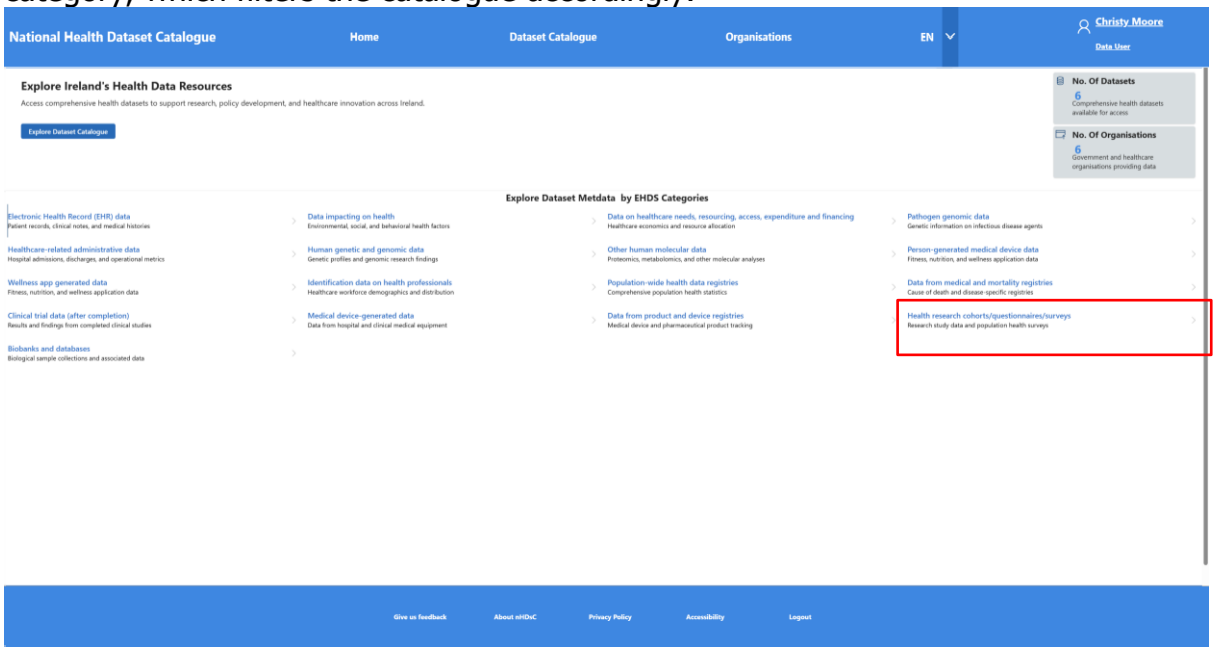
iv. Navigating to the Dataset Catalogue

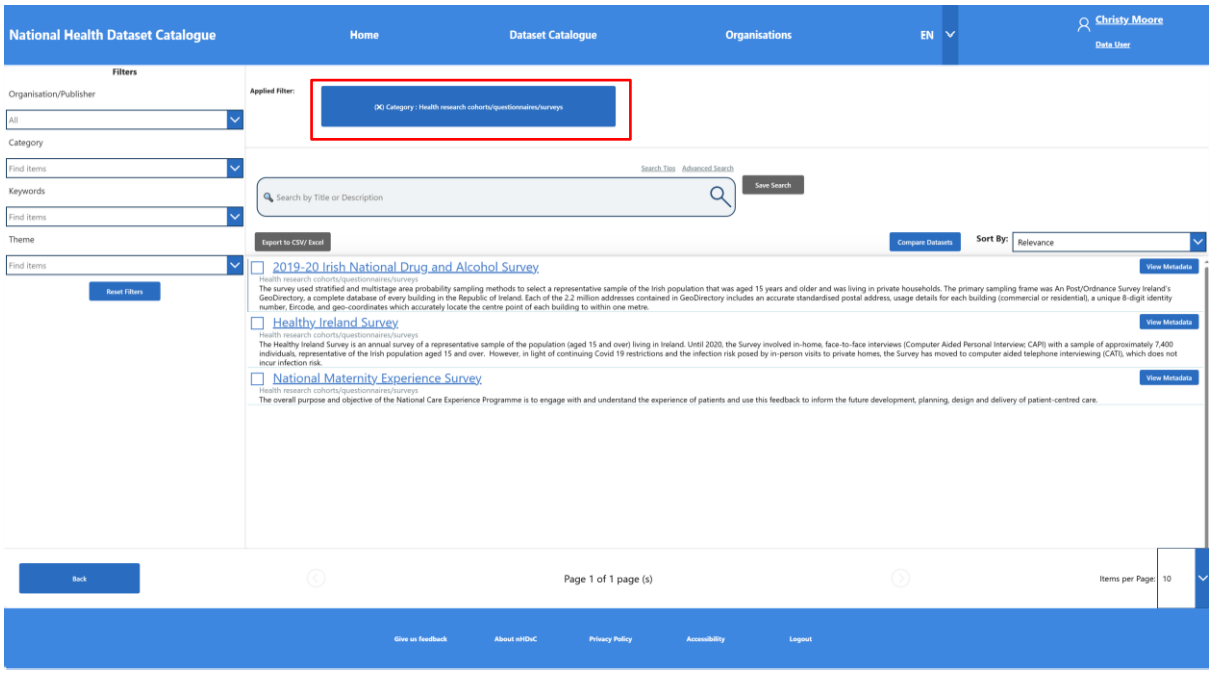
- Users can also access the Dataset Catalogue Page directly by:

- Selecting Dataset Catalogue from the header,
- Clicking Explore Dataset button, or
- Clicking the dataset count shown on the Home Page.



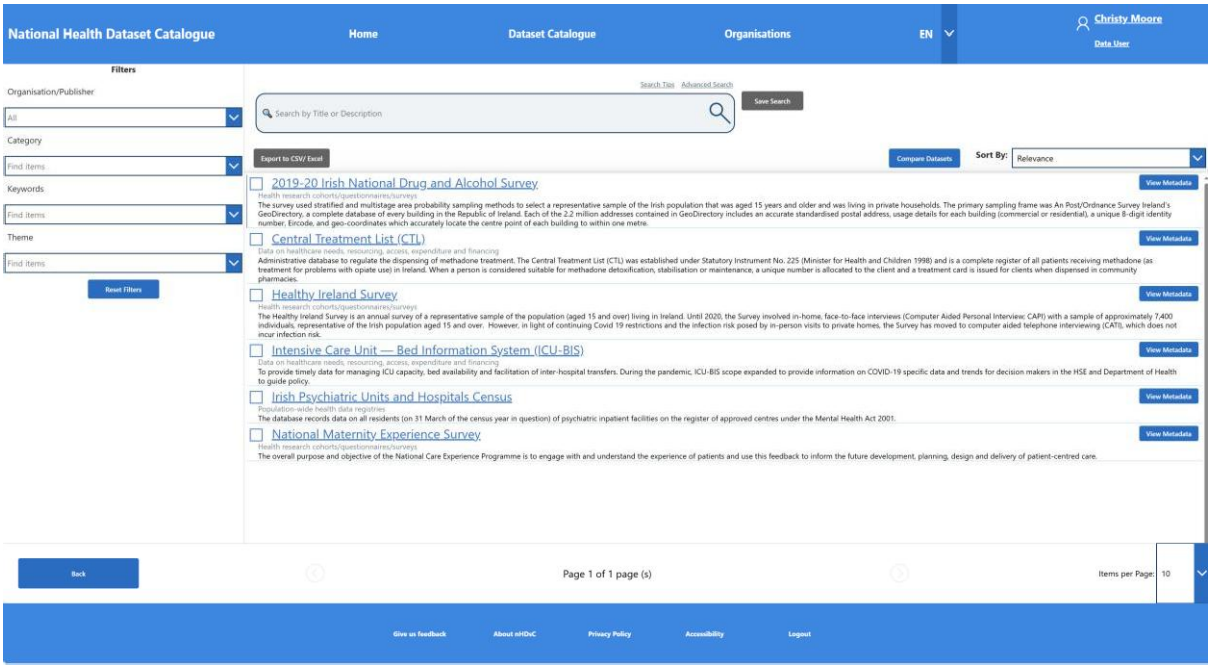
- Viewing Datasets by EHDS Categories
  Users may also browse datasets grouped by EHDS categories by selecting a category, which filters the catalogue accordingly.





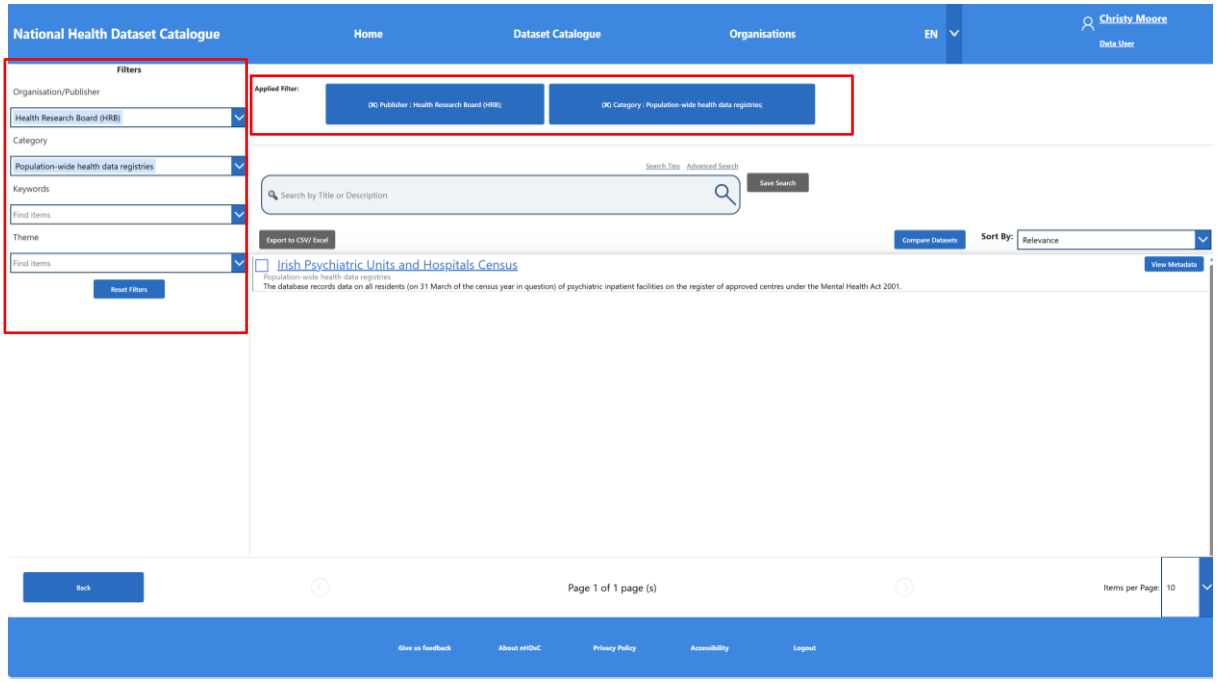
v. Dataset Catalogue Page

- Overview
  Displays a list of discoverable datasets with title, category, and description.

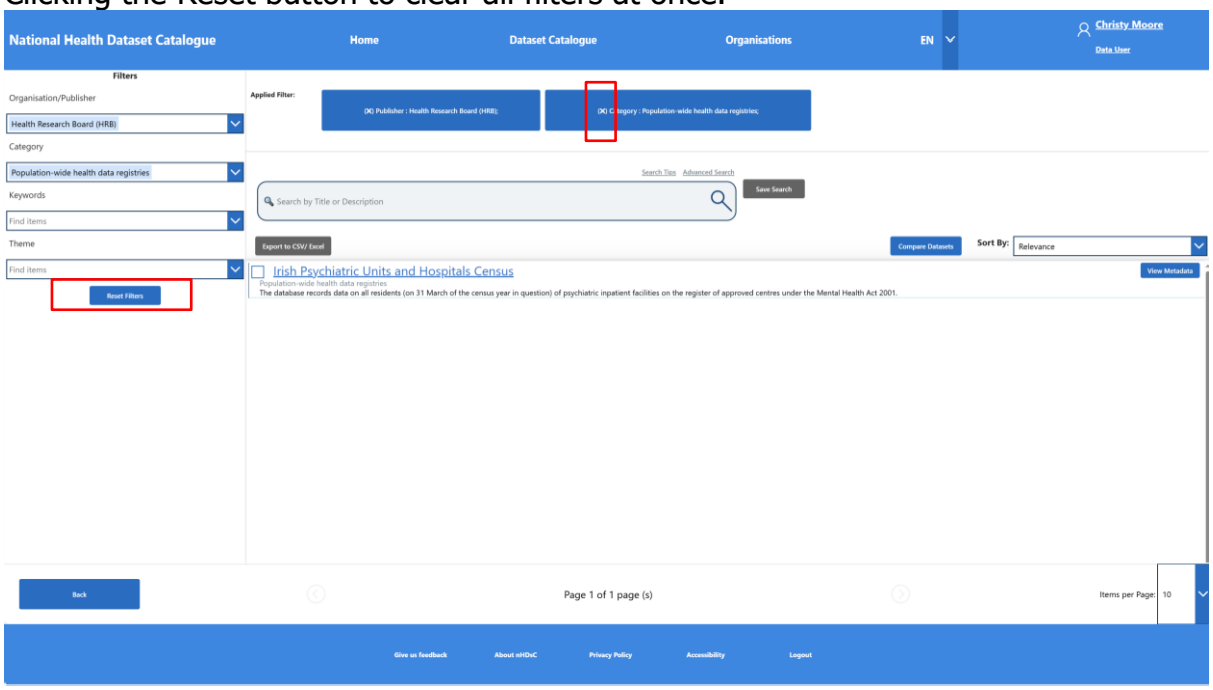


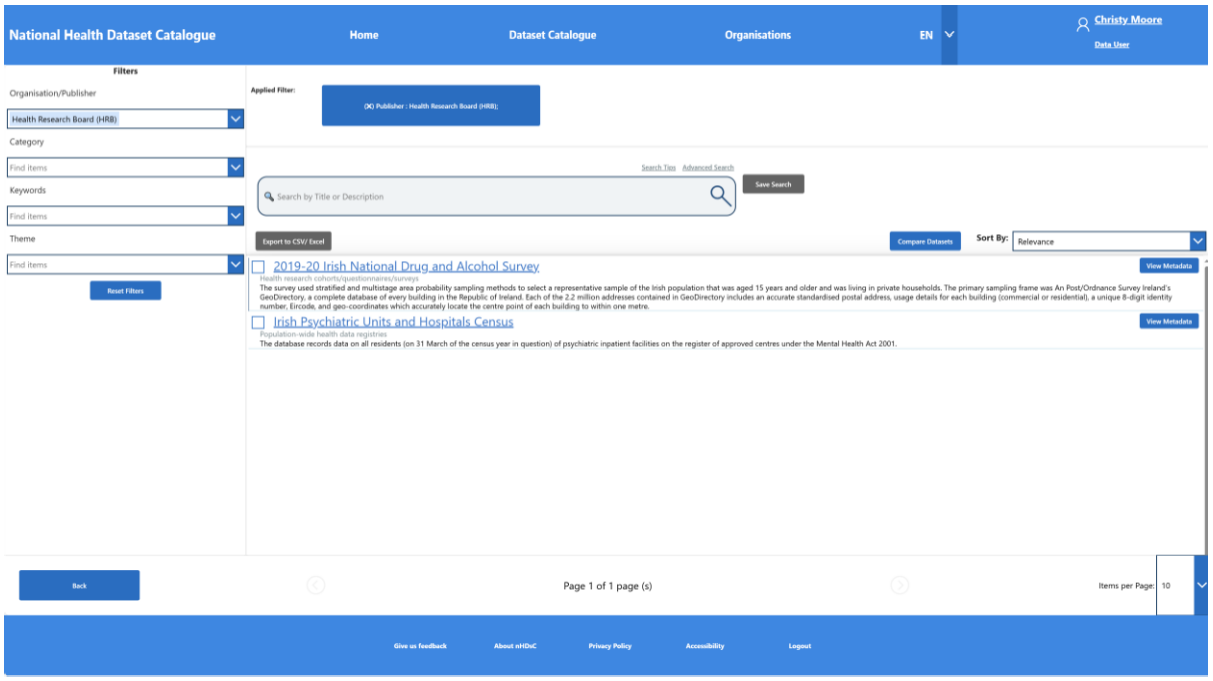
- Filtering and Search
  Users have multiple filter options to refine the dataset list.

Note: Additional filter options and fields will be introduced in the final implementation. For the prototype, only a limited set of fields has been included.

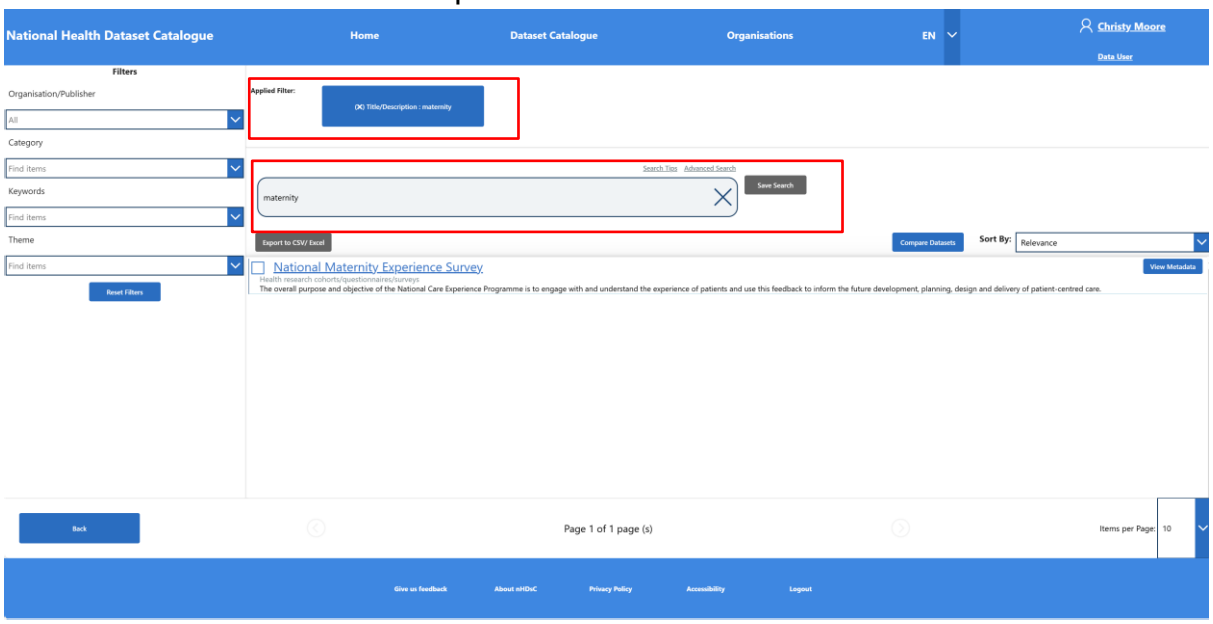


- Filters can be removed either by:
  Clearing them individually from the applied filters section, or
  Clicking the Reset button to clear all filters at once.

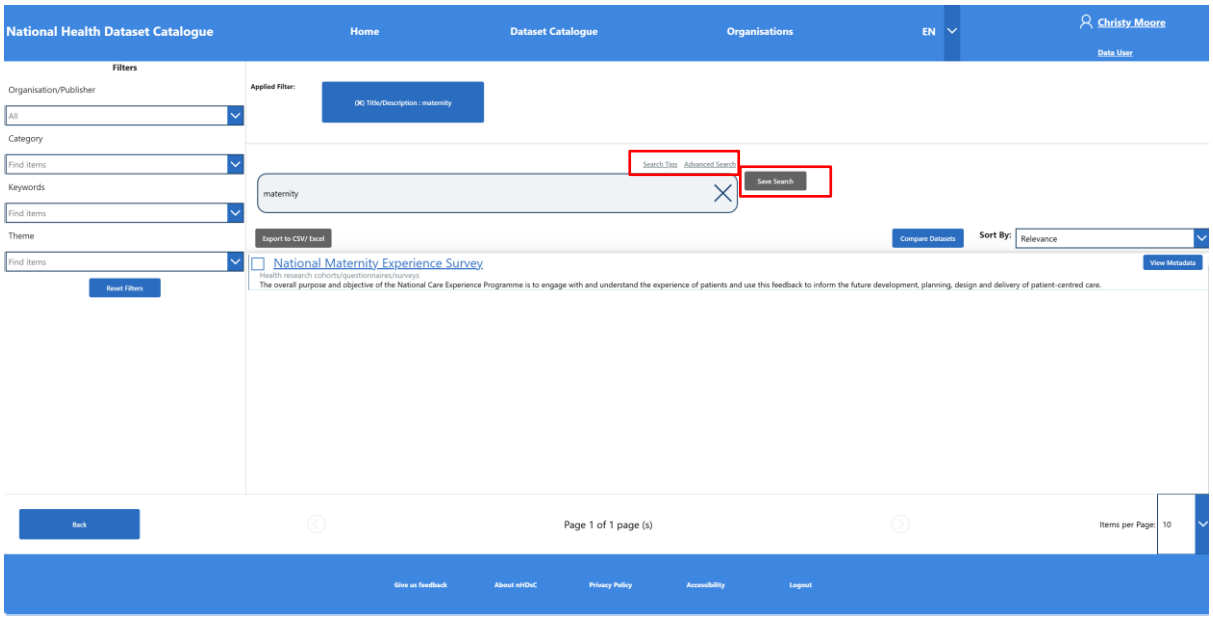




- A search box is provided, allowing users to search datasets by entering keywords from the dataset's title or description.

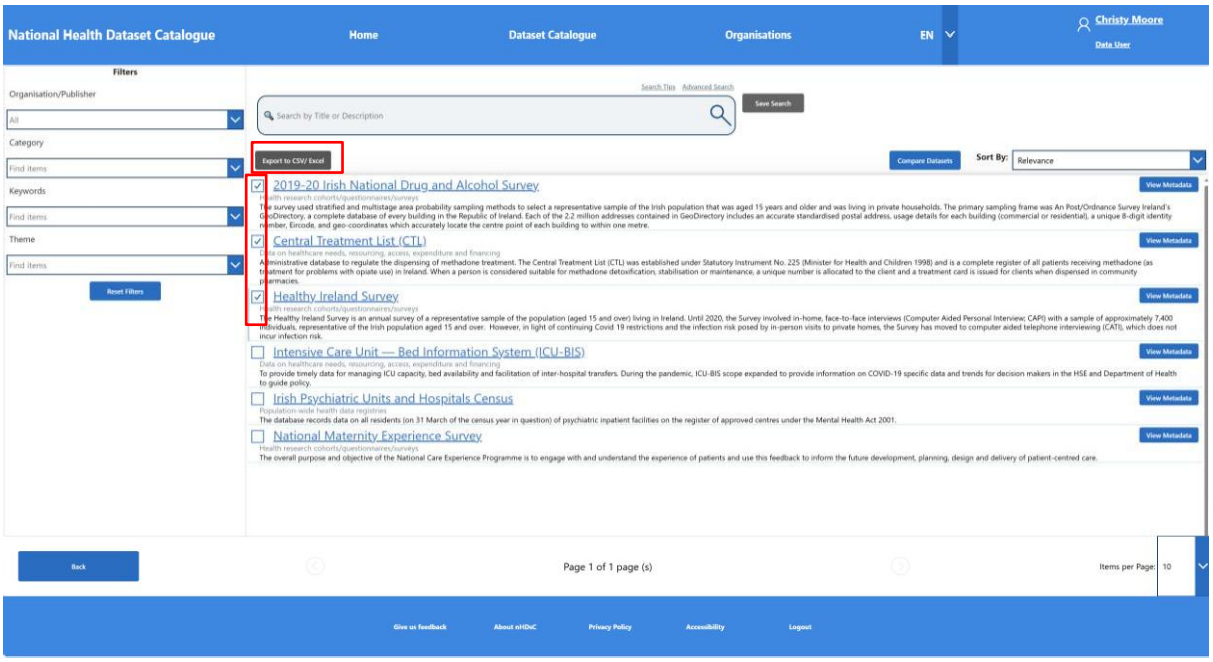


- Additional search options:
  Additional options include:
  Search tips to help refine searches,
  Advanced search for more detailed search, and
  Save search functionality. The saved searches can be accessed anytime upon login.

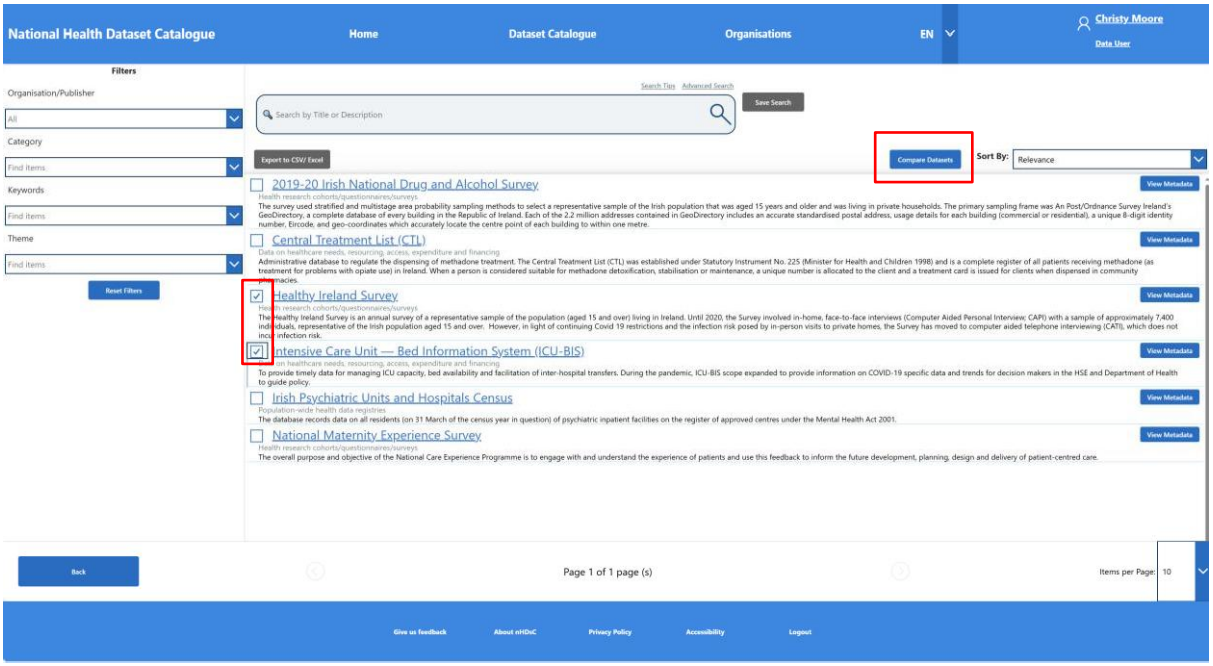


Users may enable/disable notifications for saved searches. When enabled, notifications are sent for updates or newly added matching datasets.

- Export Options
  Export options allow downloading dataset metadata in CSV or Excel format, either for all datasets visible on the screen or for specifically selected datasets.

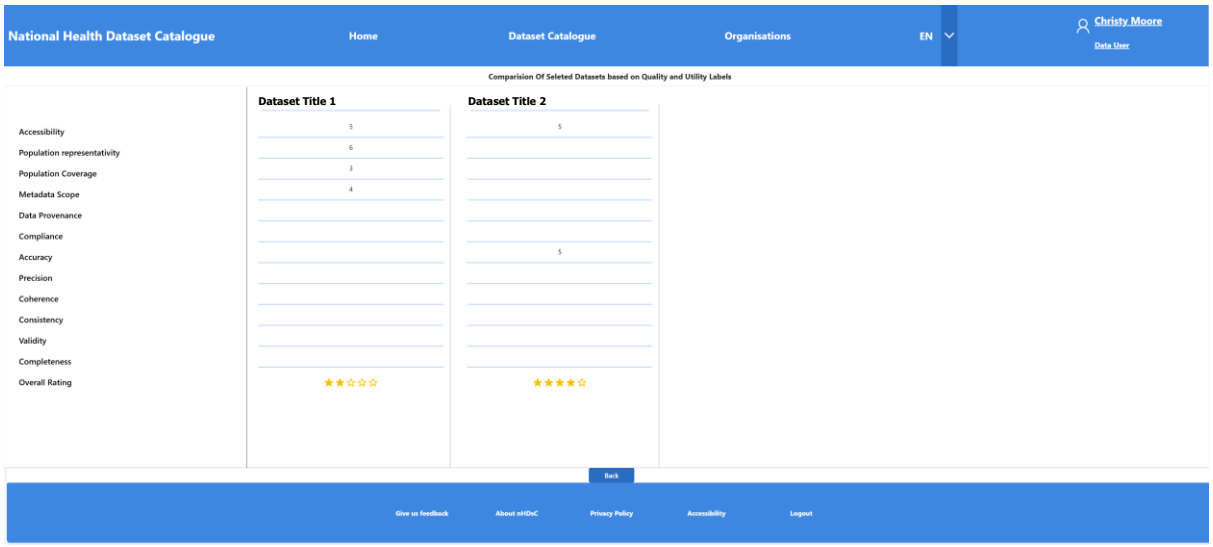


- Dataset Comparison
  Users may compare multiple datasets by selecting them and clicking Compare Datasets.



The comparison highlights differences based on the Quality and Utility Labels provided by the Data Holder

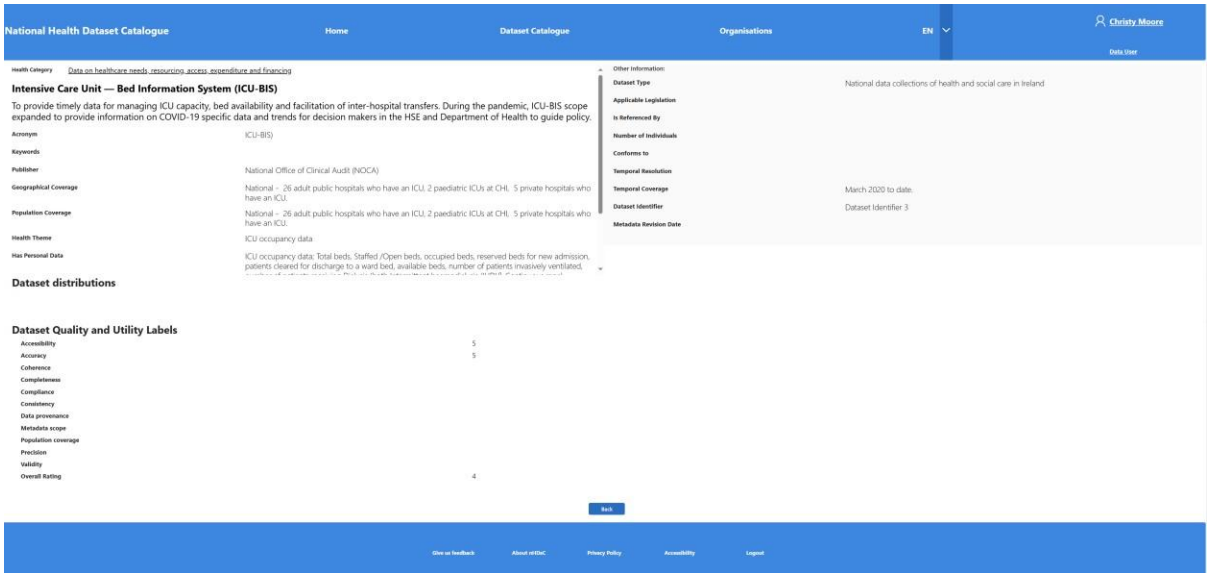
Note: In addition to Quality and Utility Labels, other metadata fields will also be displayed. For the prototype, only QUL has been included. The fields used for comparison will be finalised and confirmed during the pilot phase.



vi. Viewing Dataset Metadata

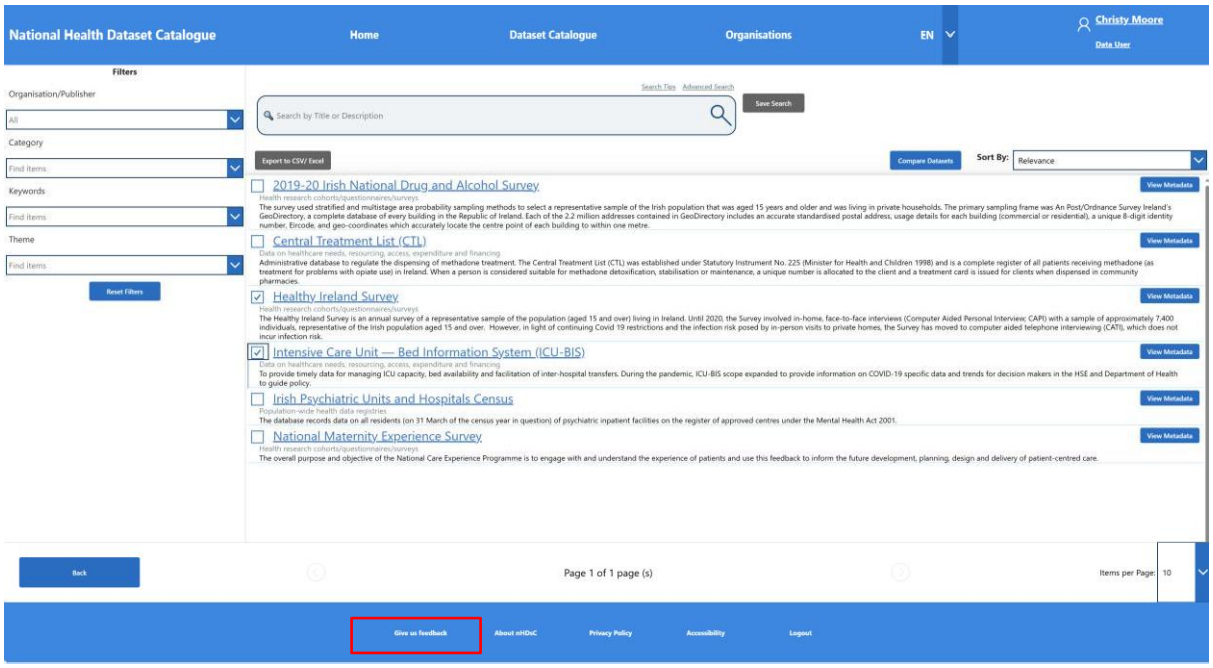
To view detailed metadata, users can click on View Metadata for any dataset. This displays metadata structured according to HealthDCAT-AP, including: Dataset details, Associated distributions, and Quality and Utility Labels.

Note: Additional HealthDCAT-AP fields will be visible in the live system (prototype displays a limited set).

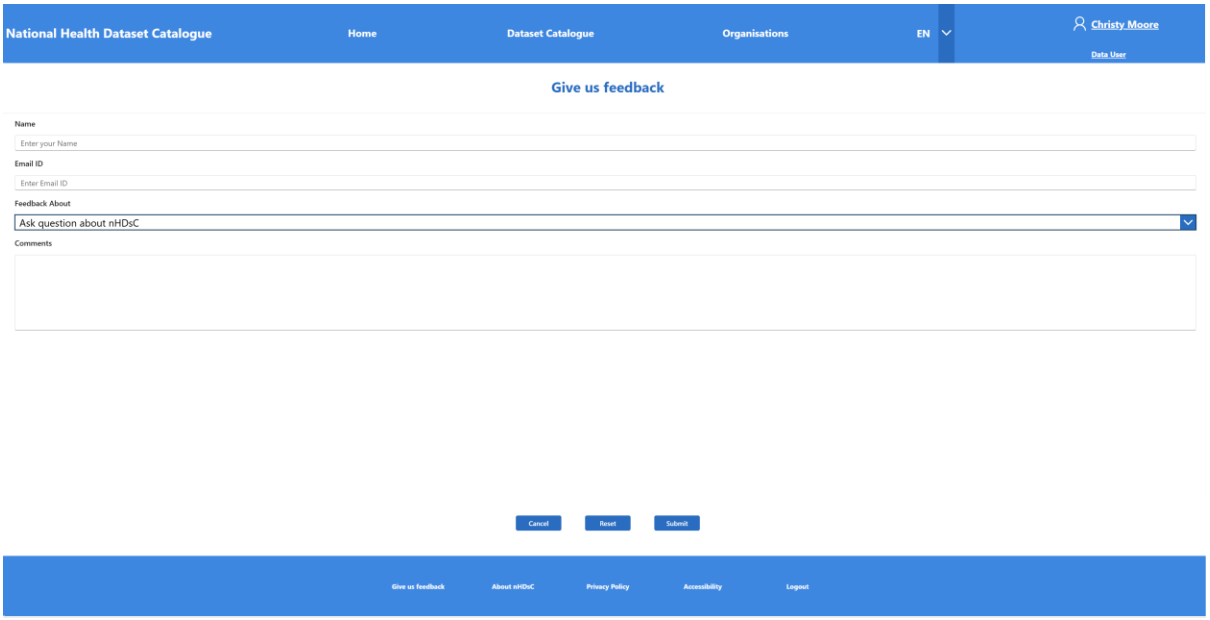


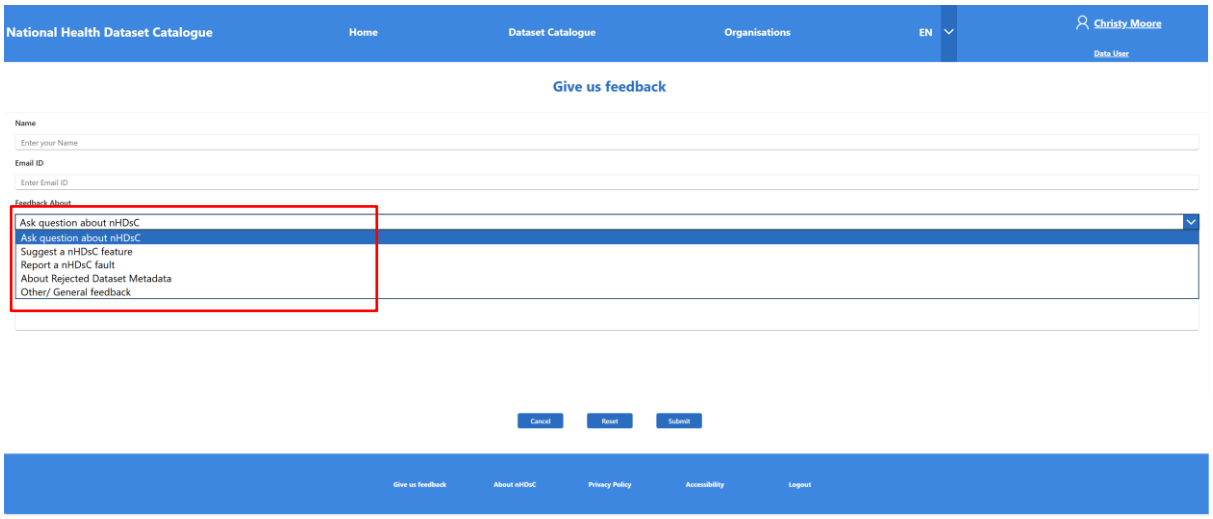
vii. Feedback Form

- Users and Data Holders can provide feedback, suggestions, or queries regarding the catalogue through the Give Feedback option located in the footer.



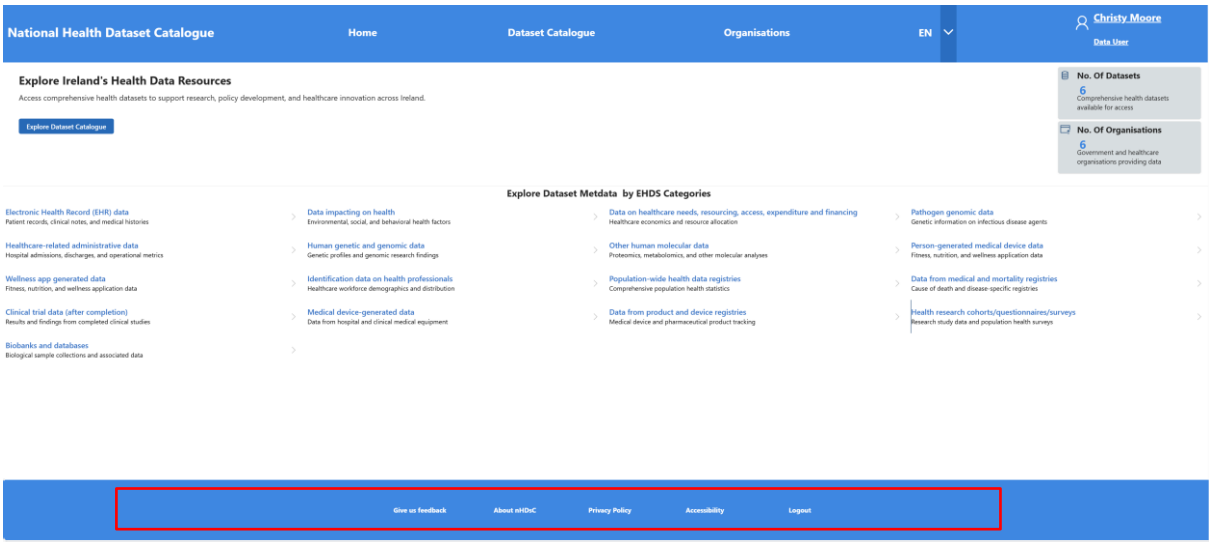
- Selecting this option opens a feedback form, where the user must:
  - Enter their name and email ID
  - Select the reason for contacting the system administrator, and Provide their comments.
  - Once submitted, the system admin will respond via email.





viii. Footer Options

The application footer includes: About nHDsC, Privacy Policy, Accessibility and Logout.

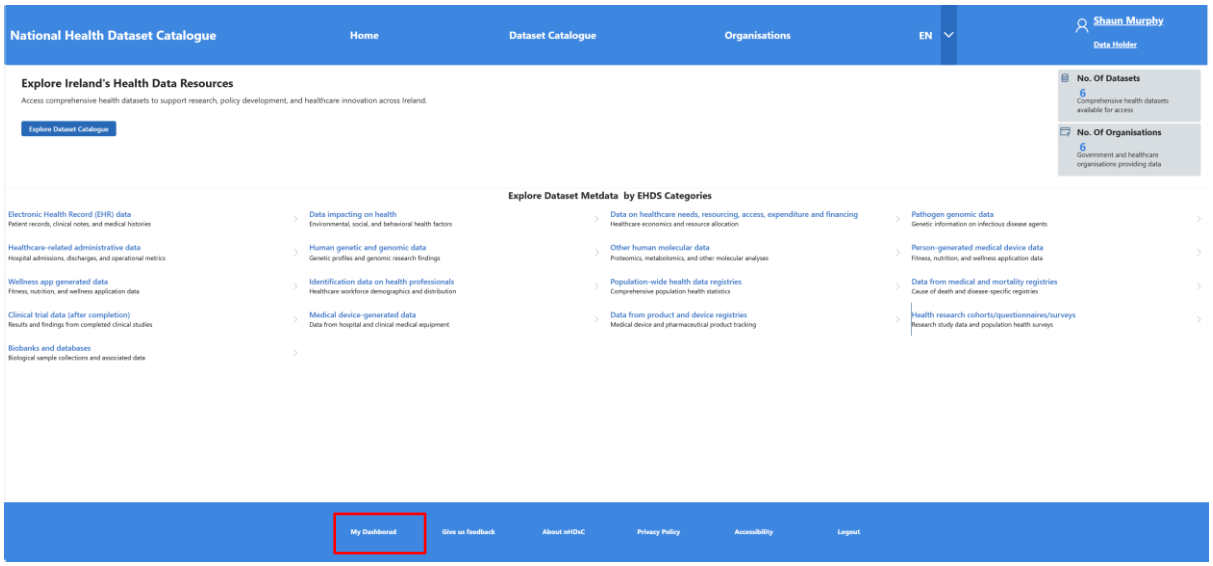


1.2 Functionalities for Data Holder

- Overview

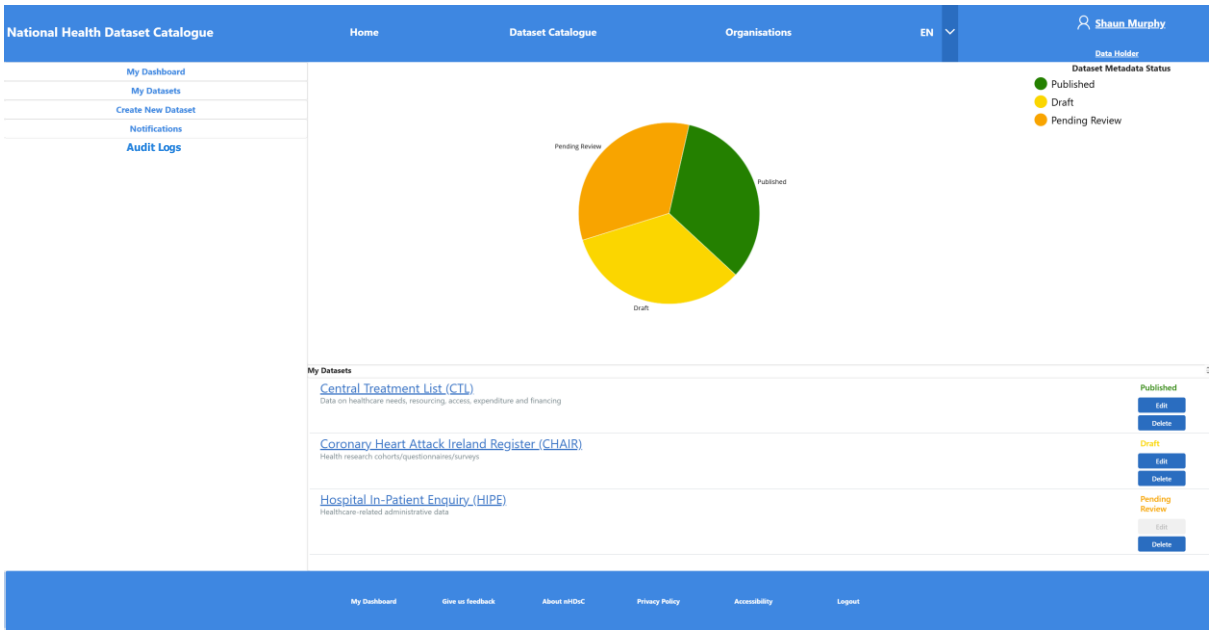
When a Data Holder logs in, they have all Data User functionalities.

In addition, they have access to the My Dashboard option, available in the footer.



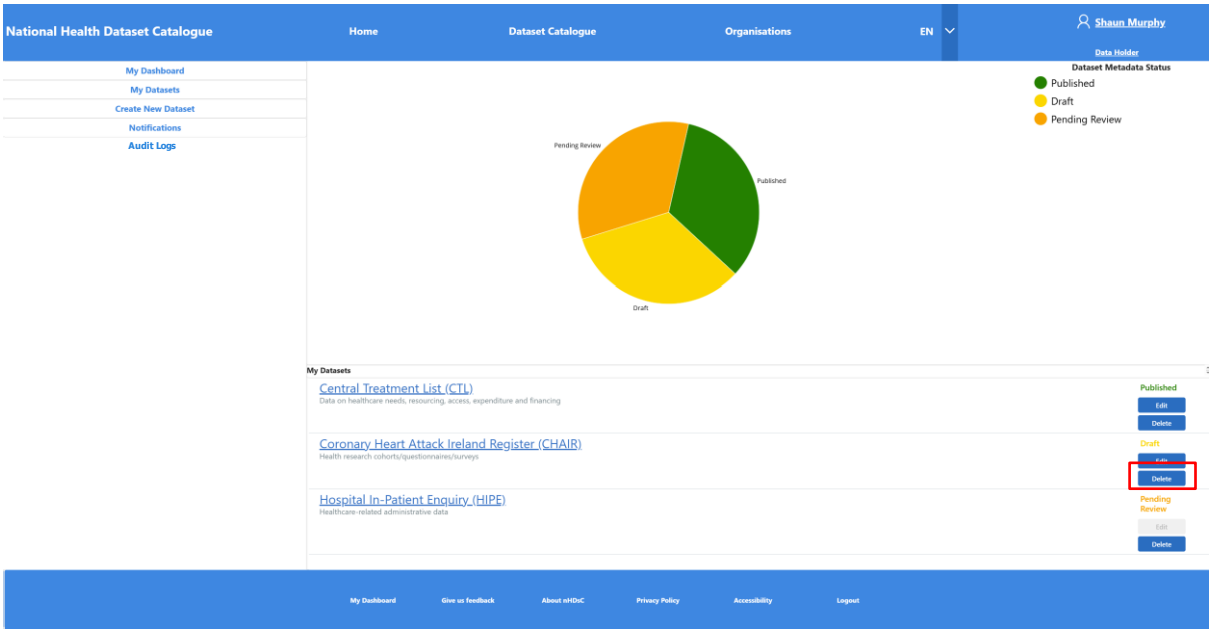
i. My Dashboard

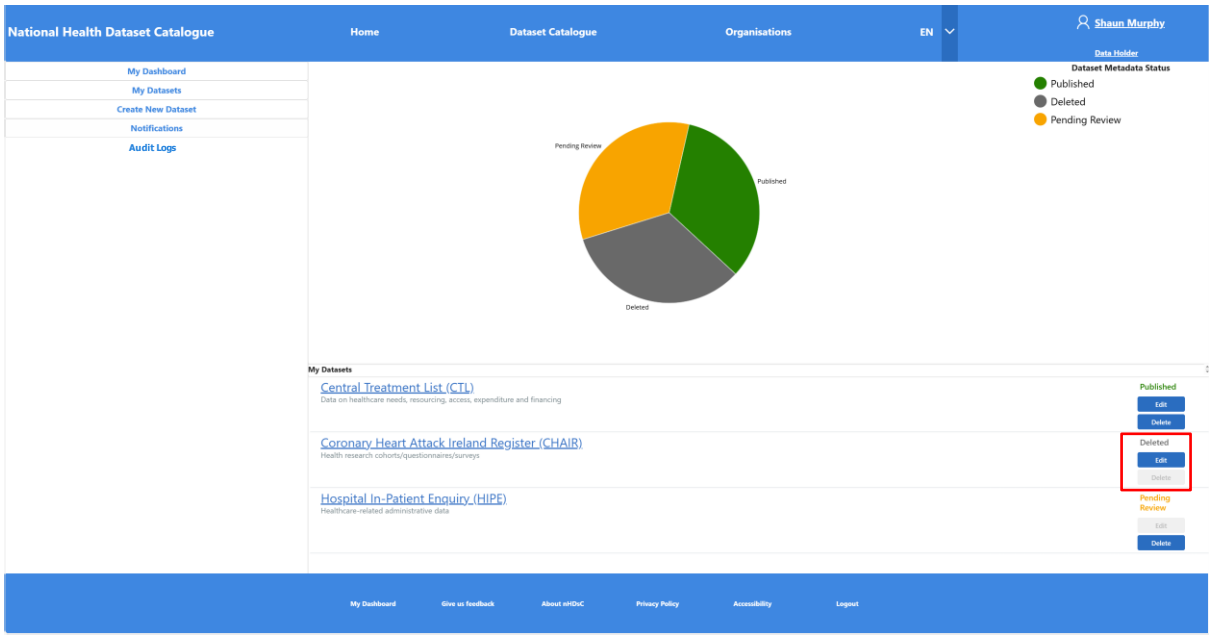
- Clicking on My Dashboard opens a pie chart summarising statuses of datasets created by the Data Holder.
- The possible statuses include:
    - Draft – dataset metadata is created and saved but not yet submitted.
    - Pending Review – dataset metadata is submitted to the System Admin for review.
    - Published – dataset metadata is reviewed and approved by the System Admin.
    - Rejected – dataset metadata is reviewed but rejected by the System Admin. A rejection reason is mandatory. Rejected datasets can be queried by data holders via the Feedback Form (by selecting the option About Rejected Dataset Metadata).
    - Deleted – dataset metadata that was created but later deleted by the Data Holder. Deletion of dataset metadata requires approval from the System Admin.



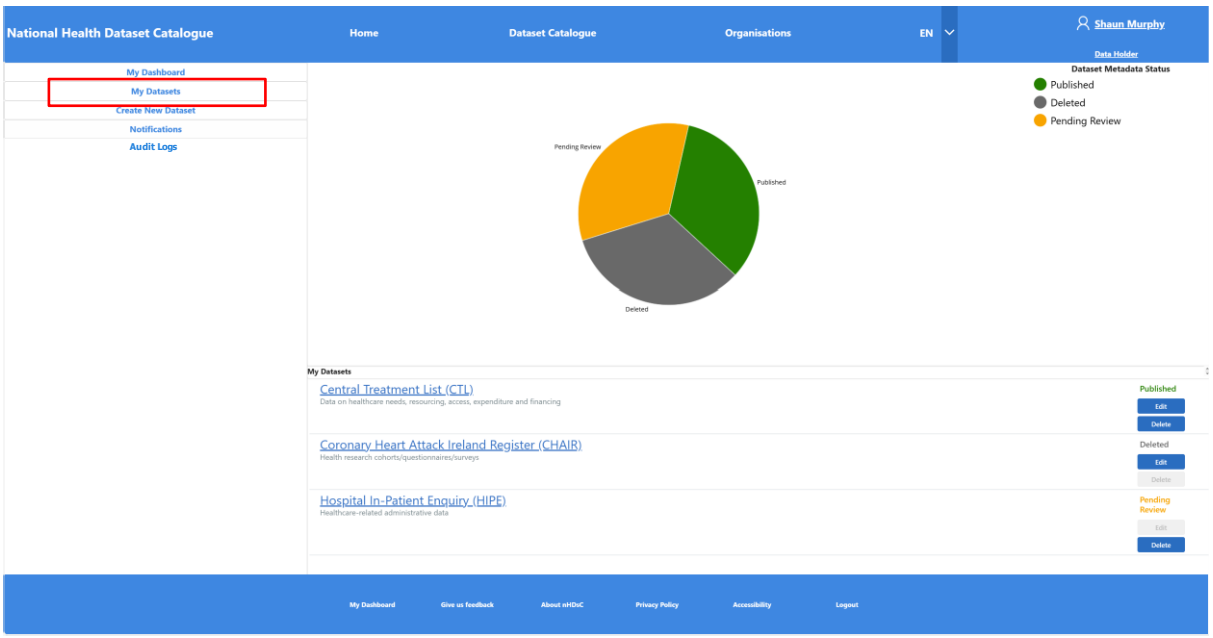
ii. Dataset Management

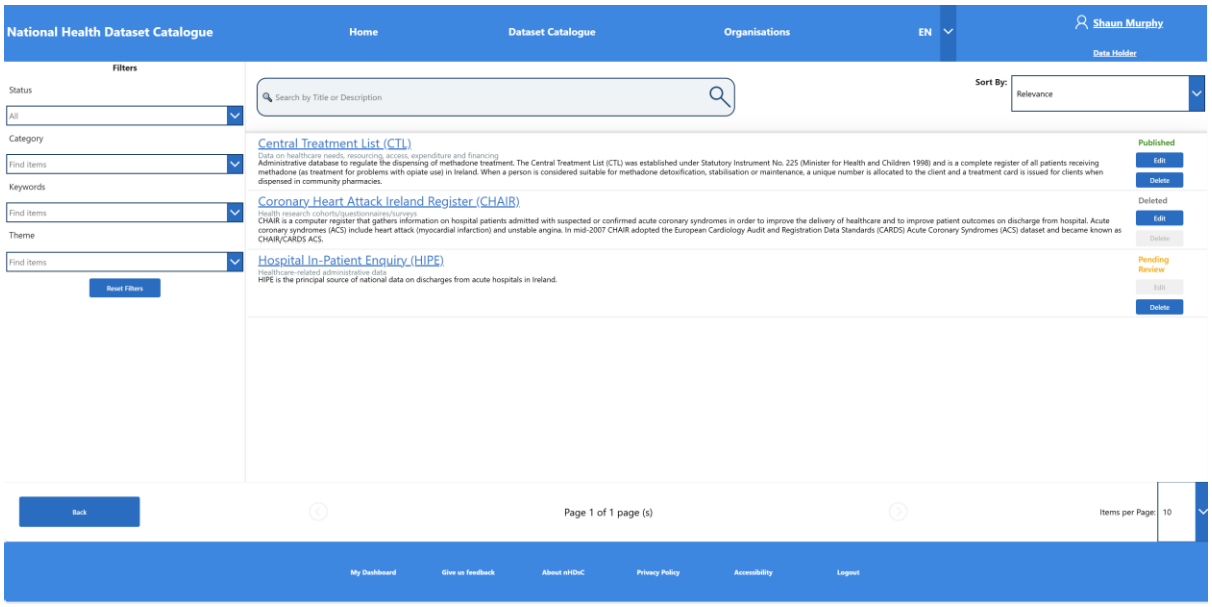
- Data Holders can edit or delete dataset metadata using the corresponding edit/delete options.
  While deleting the dataset it's mandatory to provide the reason for deletion.



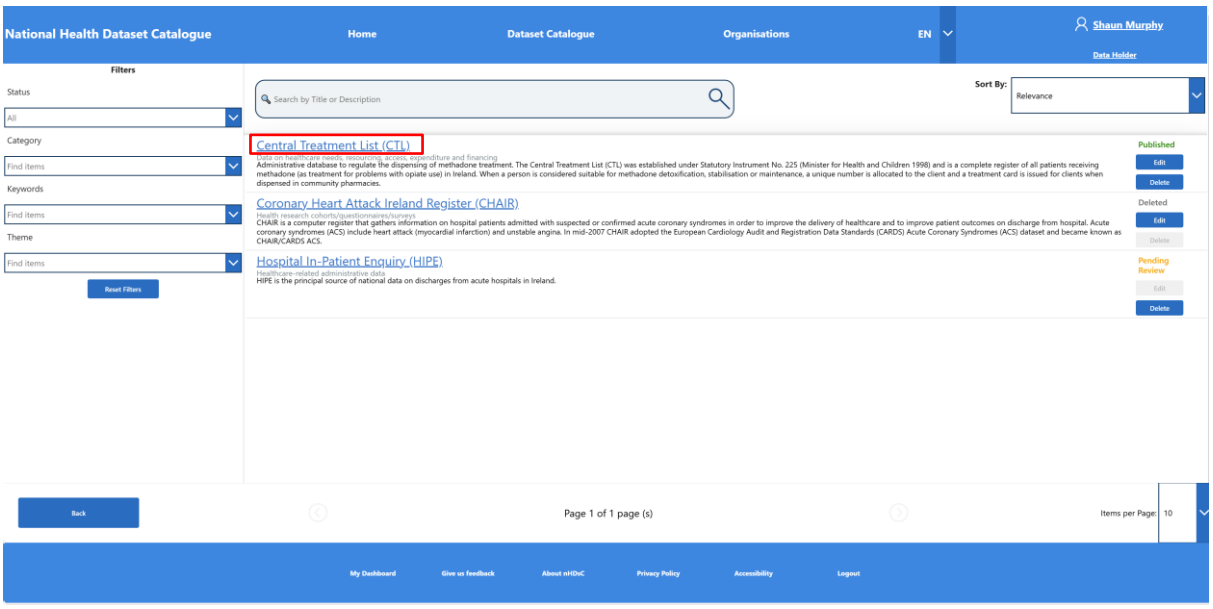


- They can view a list of datasets they have created by selecting My Datasets. The My Datasets page includes filtering, sorting, and search functionalities.

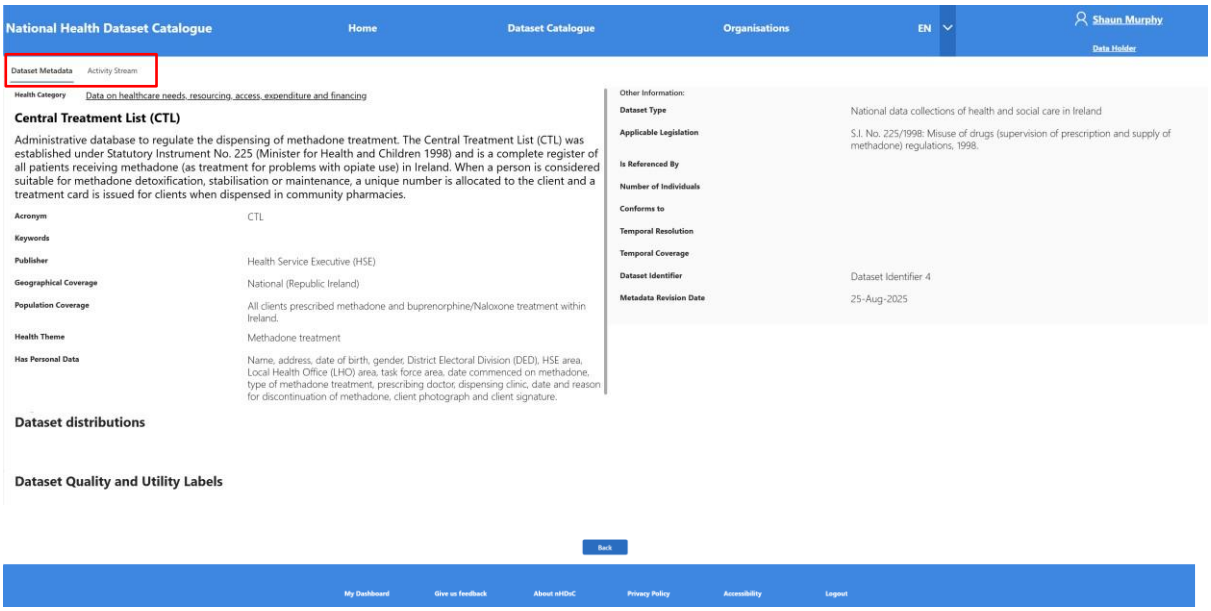




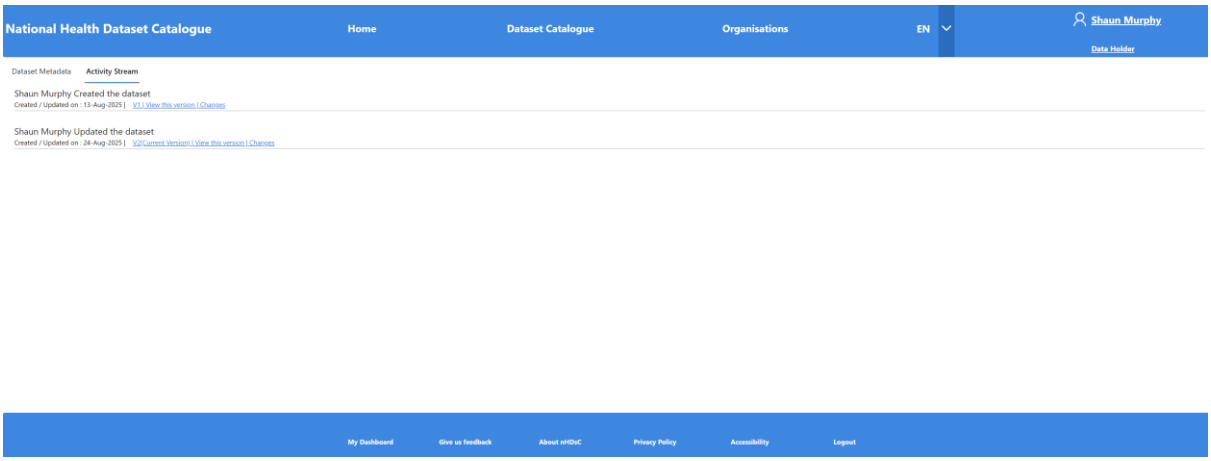
- When data holder clicks on View dataset, they will be able to view dataset metadata and Activity stream tab.



- Each dataset entry includes access to:
  - Metadata view – dataset description, distributions, and DQULs
  - Activity stream – dataset-related activities.



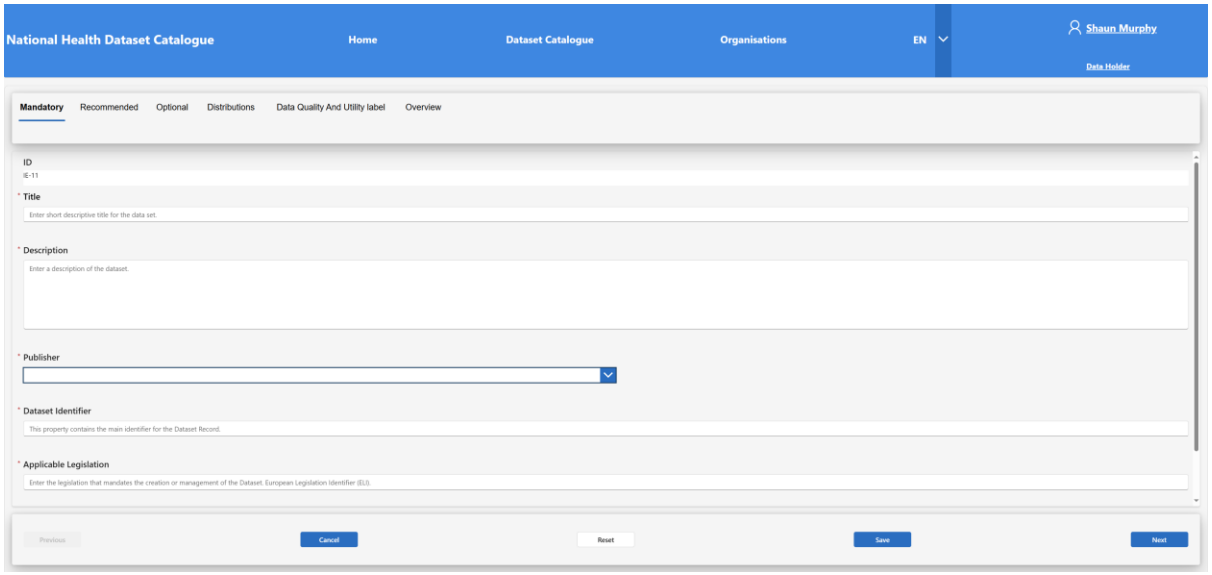
When data holder clicks on activity stream, it will show the activities related to dataset, along with version ID (can be used to view version), and an option to view changes (which highlights fields that were modified).



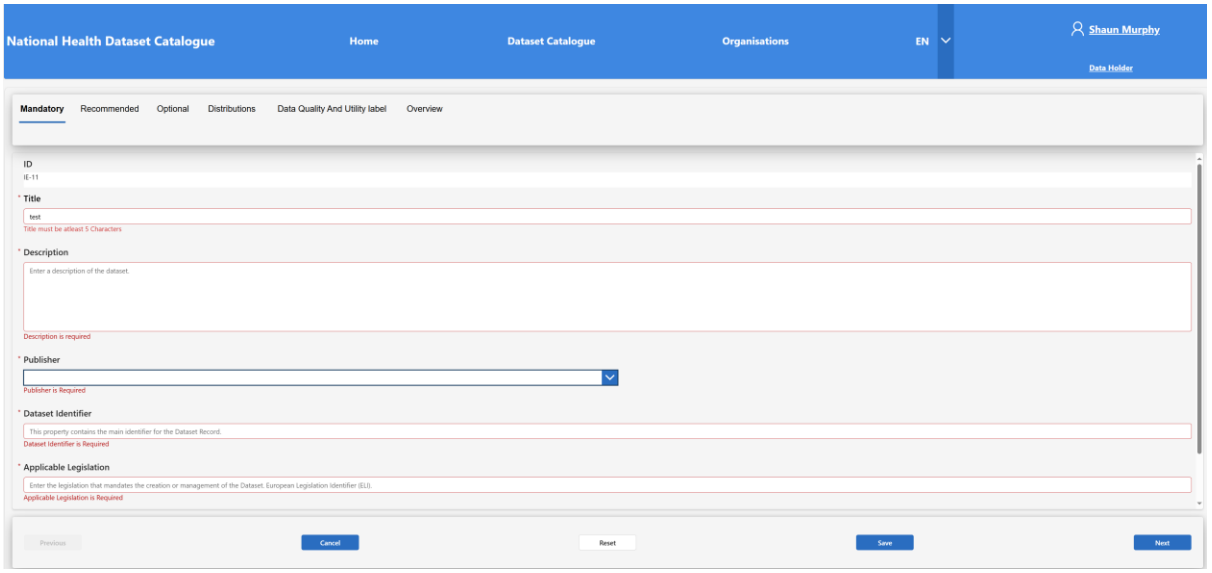
iii. Creating a Dataset

- Clicking Create Dataset opens the dataset creation flow (prototype supports creating from scratch only; cloning from existing metadata will be available in pilot).
- Metadata fields follow the HealthDCAT-AP standard, structured as:
  - Mandatory
  - Recommended
  - Optional
- Each field will include a placeholder describing the information to be entered. Required fields will be marked with an asterisk (*), while optional fields will be

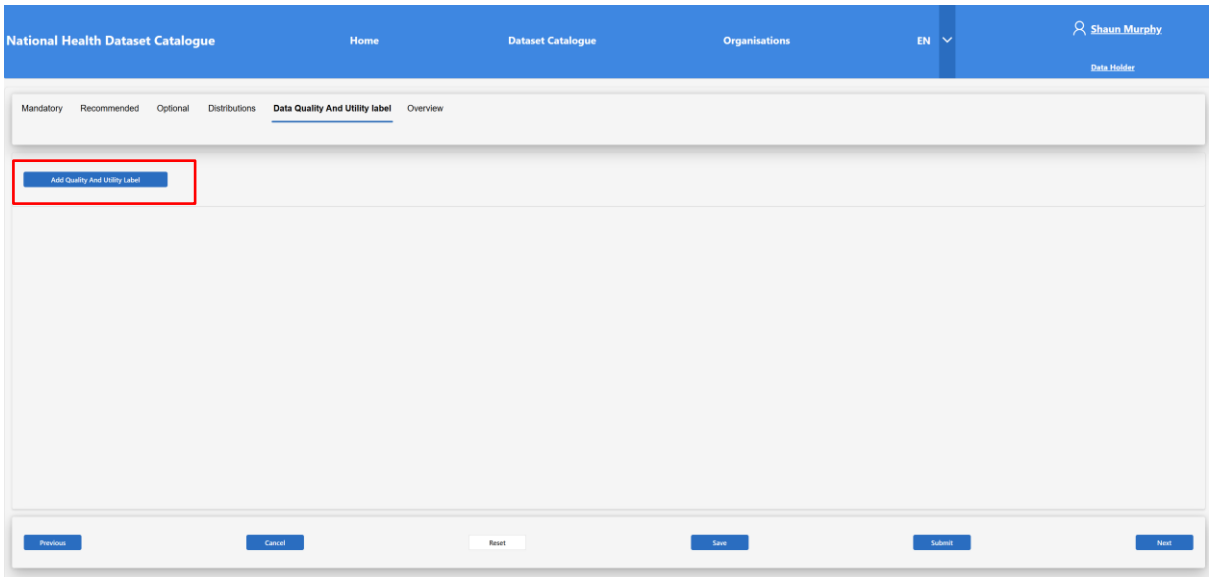
explicitly labelled. (The mandatory, optional and recommended fields will differ for Open, Protected and Sensitive data).

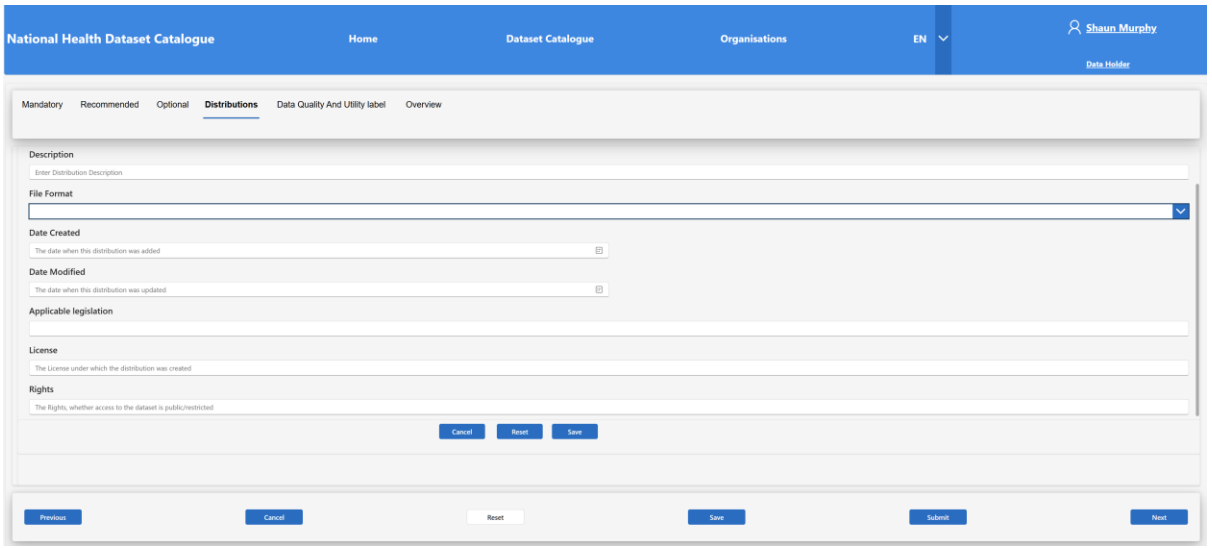


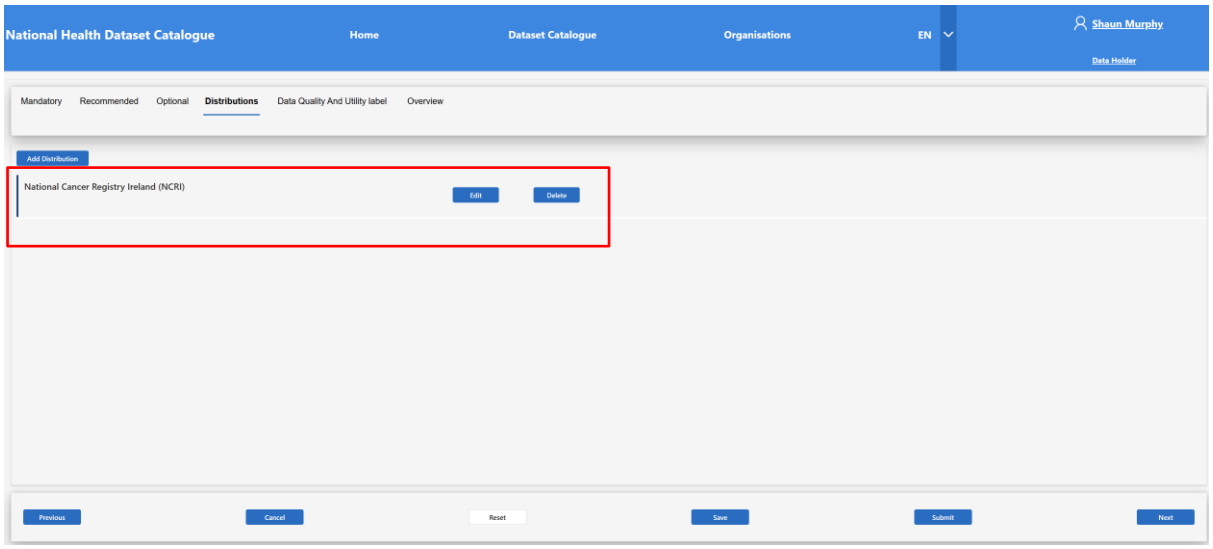
- Error messages highlight missing or incorrect information.



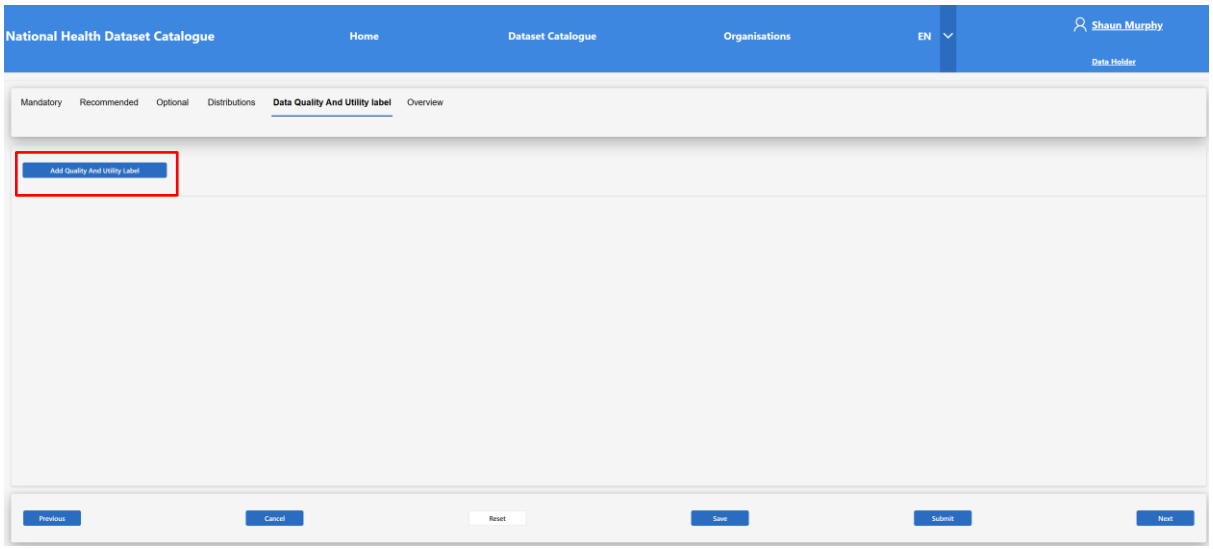
- Multiple distributions can be added to a single dataset.

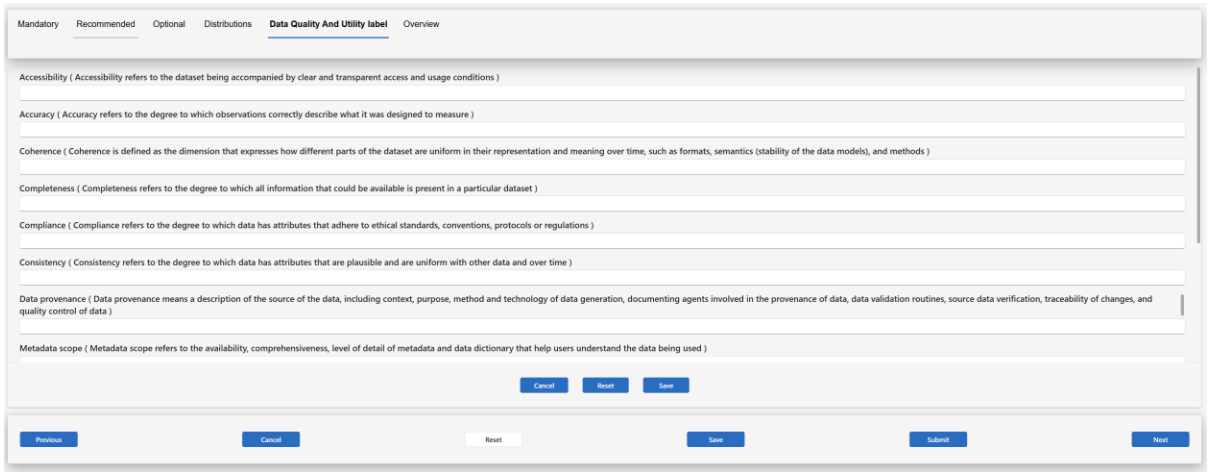

National Health Dataset Catalogue
Home
Dataset Catalogue
Organisations
EN
Shaun Murphy
Data Holder
Mandatory
Recommended
Optional
Distributions
Data Quality And Utility label
Overview
Add Quality And Utility Label
Previous
Cancel
Reset
Save
Submit
Next

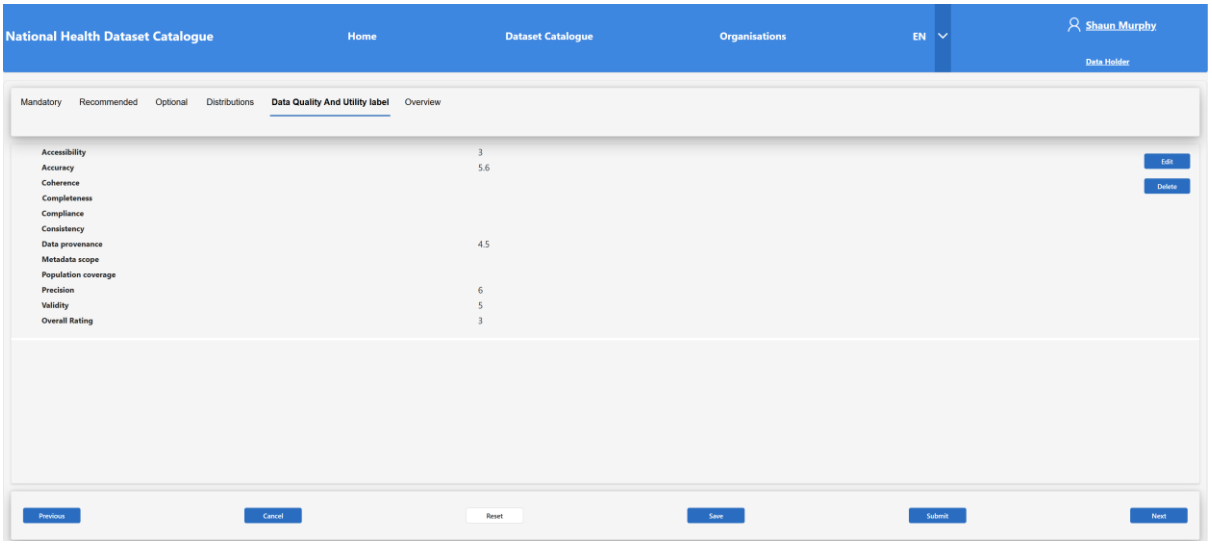

National Health Dataset Catalogue
Home
Dataset Catalogue
Organisations
EN
Shaun Murphy
Data Holder
Mandatory
Recommended
Optional
Distributions
Data Quality And Utility label
Overview
Description
Enter Distribution Description
File Format
Date Created
The date when this distribution was added
Date Modified
The date when this distribution was updated
Applicable legislation
License
The License under which the distribution was created
Rights
The Rights, whether access to the dataset is public/restricted
Cancel
Reset
Save
Previous
Cancel
Reset
Save
Submit
Next



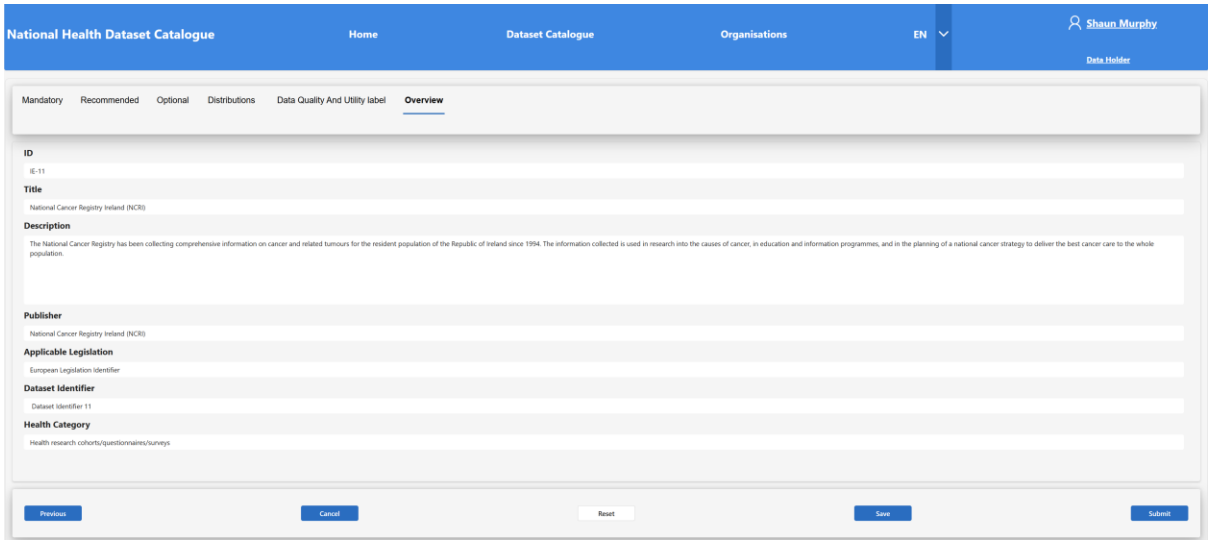
- One data Quality and Utility Labels (DQULs) can be added per dataset, based on QUANTUM 12 dimensions. (Currently, scores and the overall rating must be entered manually)



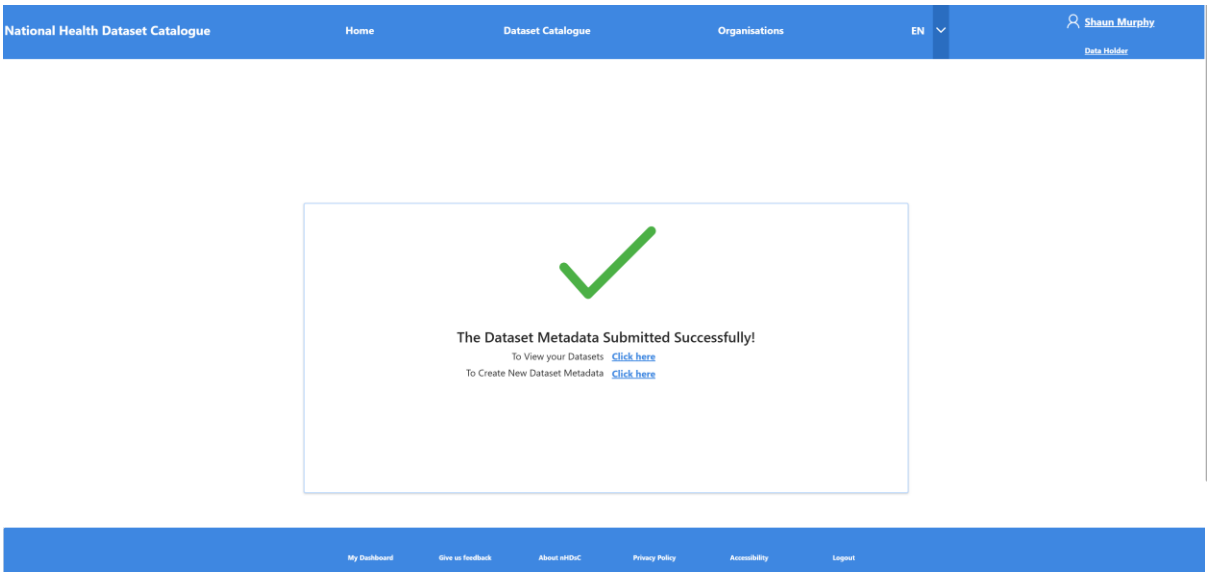




- Before submission, an Overview Page summarises all the details entered for the dataset.

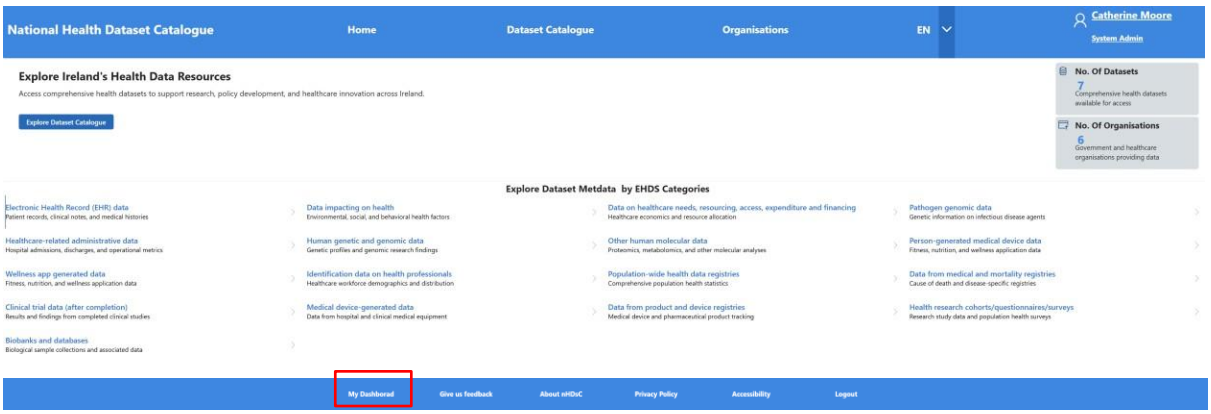


- Once submitted, status changes to Pending Review until approved (status changes to Published) or rejected (status changes to Rejected) by the System Admin.



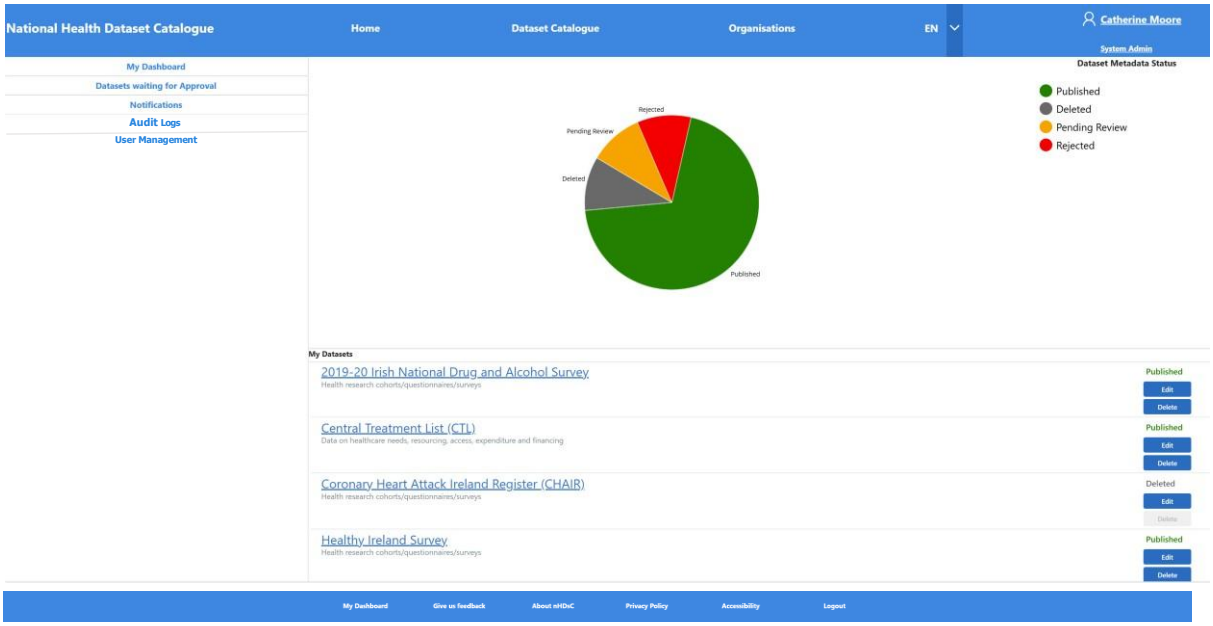
1.3 Functionalities for System Admin

- Overview
  System Admins have all Data User functionalities, plus full system-level oversight and controls available via My Dashboard.



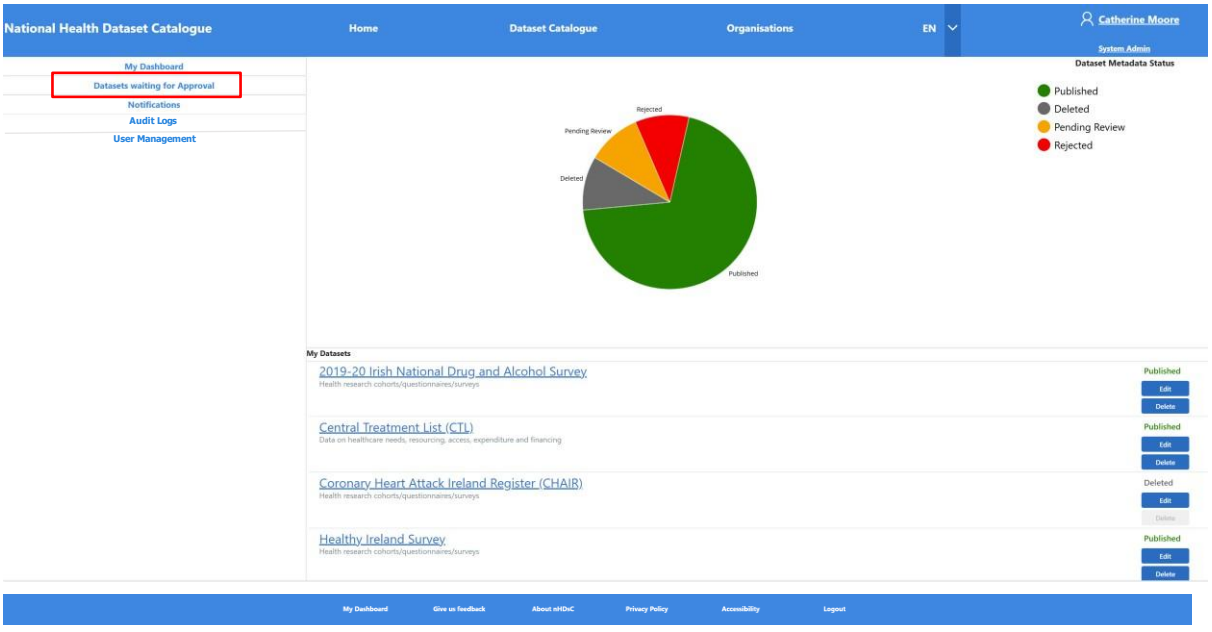
i. My Dashboard

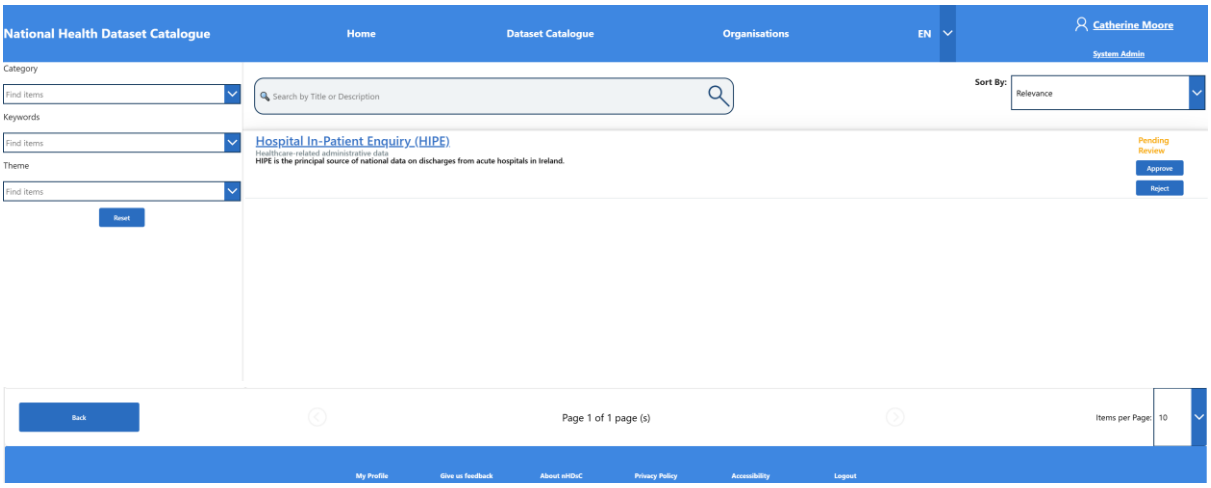
- Displays graphs of dataset statuses across the nHDsC. System Admins can view all datasets except draft.



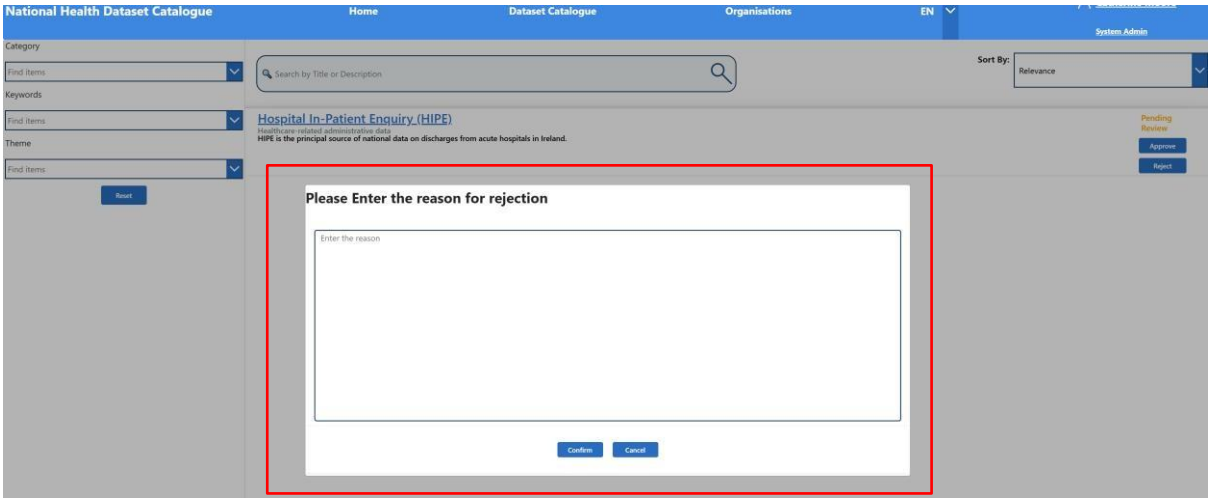
ii. Dataset Management

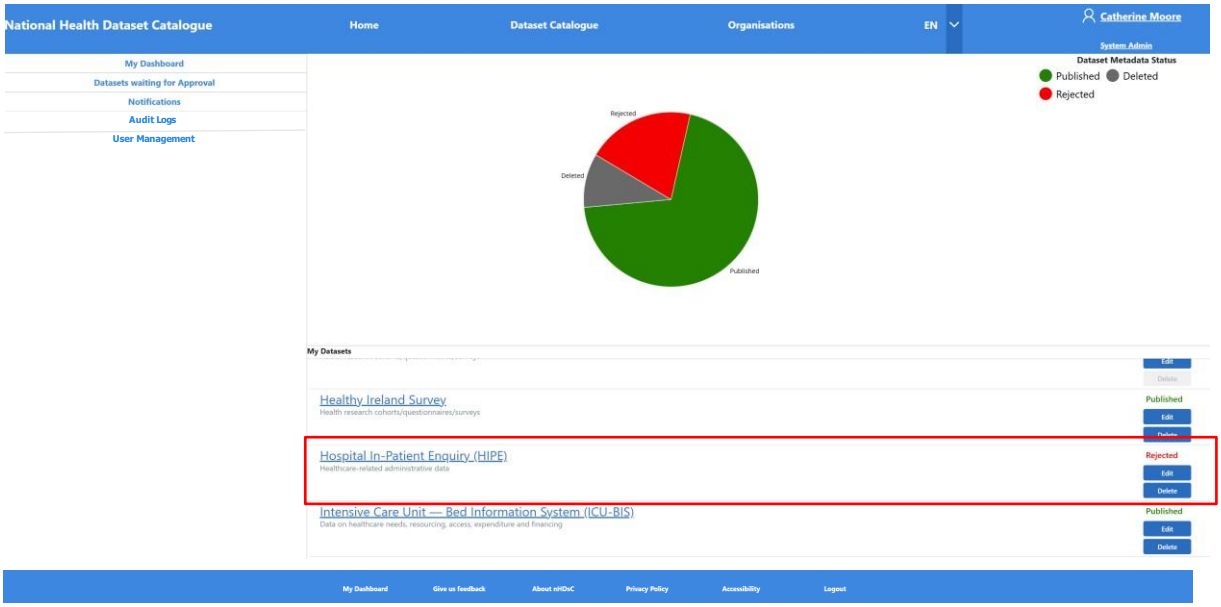
- Datasets Waiting for Approval shows a list of datasets pending review.





- Admins can:
  - Approve → dataset status changes to Published
  - Reject → mandatory to enter reason; dataset status changes to Rejected.
- Notifications are sent to Data Holders upon approval or rejection





Note: Additional Planned Functionalities (Pilot Phase)

The following features are not implemented in the prototype but will be included in the pilot phase:

- Account management requests: Data Holder Editors/Admins and Data Users can request account deactivation
- If deactivating a Data Holder account, the requester must provide the name of the new Data Holder for dataset ownership transfer.
- System Admins can manage user account additions and deletions.
- Data Holders and System Admins will have access to audit logs, enabling them to view upcoming dataset audits and download logs.

Appendix 6. National Dataset Catalogue for health data (nHDsC) - Prototype (EN) - Content of the metadata catalogue

Note: This is based on HealthDCAT-AP Unofficial Draft 22 December 2023 and EU Central platform Release 5

Class: Dataset

Name of the Property	RDF Example	Definition	Public Property Type (HealthDCAT-AP) (Mandatory/ Recommended/	Restricted Property Type (HealthDCAT-AP) (Mandatory/ Recommended/	Non-Public Property Type (HealthDCAT-AP)

HealthDCAT-AP			Optional)	Optional)	(Mandatory/ Recommended/Optional)
Title	dct:title: rdfs:Literal [1..*]	A name given to the Dataset.	Mandatory	Mandatory	Mandatory
Description	dct:description: rdfs:Literal [1..*]	A free-text account of the Dataset.	Mandatory	Mandatory	Mandatory
Theme	dcat:theme (dct:subject): skos:Concept [1..*]	A category of the Dataset.	Mandatory	Mandatory	Mandatory
Identifier	dct:identifier: rdfs:Literal: xsd:anyURI [1..*]	The main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue.	Mandatory	Mandatory	Mandatory
Health data access body	healthdcatap:hdab foaf:Agent [1..1]	Health Data Access Body supporting access to data in the Member State.	Mandatory	Mandatory	Mandatory
Applicable legislation	dcatap:applicableLegislation eli:LegalResource [1..*]	The legislation that mandates the creation or management of the Dataset.	Mandatory	Mandatory	Mandatory
Access rights	dct:accessRights: dct:RightsStatement [1..1]	Information that indicates whether the Dataset is publicly accessible, has access restrictions or is not public.	Mandatory	Mandatory	Mandatory
Health Category	healthdcatap:healthCategory: (dct:subject) skos:Concept [1..*]	The health category to which this dataset belongs as described in the Commission Regulation on the European Health Data Space laying down a list of categories of electronic data for secondary use, Art.51.	Mandatory	Mandatory	Mandatory
Dataset distribution	dcat:distribution: dcat:Distribution [1..*]	An available Distribution for the Dataset.	Mandatory	Mandatory	Mandatory
Purpose	dpv:hasPurpose dpv:Purpose [0..*]	A free text statement of the purpose of the processing of data or personal data.	Recommended	Recommended	Recommended
Keyword	dcat:keyword: rdfs:Literal [0..*]	A keyword or tag describing the Dataset.	Recommended	Recommended	Mandatory
Population Coverage	healthdcatap:populationCoverage rdfs:Literal [0..*]	A definition of the population within the dataset	Recommended	Recommended	Recommended

Number of Unique Individuals	healthdcatap:numberofUniqueIndividuals rdfs:nonNegativInteger [0..1]	Number of records for unique individuals.	Recommended	Recommended	Recommended
Number of Records	healthdcatap:numberofRecords rdfs:nonNegativInteger [0..1]	Size of the dataset in terms of the number of records.	Recommended	Recommended	Recommended
Minimum Typical Age	healthdcatap:minTypicalAge rdfs:nonNegativInteger [0..1]	Minimum typical age of the population within the dataset	Recommended	Recommended	Recommended
Maximum Typical Age	healthdcatap:maxTypicalAge rdfs:nonNegativInteger [0..1]	Maximum typical age of the population within the dataset	Recommended	Recommended	Recommended
Personal Data	dpv:hasPersonalData dpv:PersonalData [0..*]	Key elements that represent an individual in the dataset.	Recommended	Recommended	Recommended
Legal Basis	dpv:hasLegalBasis dpv:LegalBasis [0..*]	The legal basis used to justify processing of personal data	Recommended	Recommended	Recommended
Retention Period	healthdcatap:retentionPeriod dct:PeriodOfTime [0..1]	A temporal period which the dataset is available for secondary use.	Recommended	Recommended	Recommended
Contact Point	dcat:contactPoint: vcard:Kind [1..*]	Contact information that can be used for sending comments about the Dataset.	Recommended	Recommended	Mandatory
Publisher	dct:publisher: foaf:Agent [0..1]	An entity (organisation) responsible for making the Dataset available.	Recommended	Recommended	Recommended
Coding System	healthdcatap:hasCodingSystem dct:Standard [0..*]	Coding systems in use (ex: ICD-10-CM, DGRs, SNOMED=CT, …)	Recommended	Recommended	Recommended
Health Theme	healthdcatap:healthTheme: (dct:subject) skos:Concept [1..*]	A category of the Dataset or tag describing the Dataset.	Recommended	Recommended	Mandatory
Code Values	healthdcatap:hasCode Values: skos:Concept [0..*]	Health classifications, or other classification-like systems and their codes that can be associated with the dataset.	Recommended	Recommended	Recommended
Quality Annotation	dqv:hasQualityAnnotation dqv:QualityCertificate [0..*]	A statement related to quality of the Dataset, including rating, quality certificate, feedback that can be associated to the dataset.	Recommended	Recommended	Recommended
Sample	adms:sample: dcat:Distribution [0..*]	A sample distribution of the dataset.	Recommended	Recommended	Recommended

Analytics	healthdcatap:analytics : dcat:Distribution [0..*]	An analytics distribution of the dataset.	Recommended	Recommended	Recommended
Provenance	dct:provenance: dct:ProvenanceStatement [0..*]	A statement about the lineage of a Dataset.	Optional	Optional	Mandatory
Creator	dct:creator: foaf:Agent [0..*]	An entity responsible for producing the dataset.	Optional	Optional	Optional
Was generated by	prov:wasGeneratedBy: prov:Activity [0..*]	An activity that generated, or provides the business context for, the creation of the dataset.	Optional	Optional	Optional
Language	dct:language: dct:LinguisticSystem [0..*]	A language of the Dataset.	Optional	Optional	Recommended
Geographical Coverage	dct:spatial: dct:Location [0..*]	A geographic region that is covered by the Dataset.	Optional	Optional	Mandatory
Temporal Coverage	dct:temporal: dct:PeriodOfTime [0..*]	A temporal period that the Dataset covers.	Optional	Optional	Recommended
Temporal Resolution	dcat:temporalResolution rdfs:Literal: xsd:duration [0..1]	The minimum time period resolvable in the dataset.	Optional	Optional	Recommended
Spatial Resolution	dcat:spatialResolutionInMeters: rdfs:Literal: xsd:decimal [0..*]	The minimum spatial separation resolvable in a dataset, measured in meters.	Optional	Optional	Optional
Frequency	dct:accrualPeriodicity: dct:Frequency [0..1]	The frequency at which the Dataset is updated.	Optional	Optional	Recommended
Qualified Attribution	prov:qualifiedAttribution prov:attribution [0..*]	An Agent having some form of responsibility for the resource.	Optional	Optional	Optional
Other Identifier	adms:identifier: adms:Identifier [0..*]	A secondary identifier of the Dataset, such as MAST/ADS17, DataCite18, DOI19, EZID20 or W3ID21.	Optional	Optional	Optional
Conforms to	dct:conformsTo: dct:Standard [0..*]	An implementing rule or other specification.	Optional	Optional	Recommended
Related Resource	dct:relation: rdfs:Resource [0..*]	A related resource.	Optional	Optional	Recommended
Is Referenced by	dct:isReferencedBy: rdfs:Resource [0..*]	A related resource, such as a publication, that references, cites, or otherwise points to the dataset.	Optional	Optional	Recommended
Landing Page	dcat:landingPage: foaf:Document [0..*]	A web page that provides access to the Dataset, its	Optional	Optional	Recommended

		Distributions and/or additional information.			
Documentation	foaf:page: foaf:Document [0..*]	A page or document about this Dataset.	Optional	Optional	Recommended
Version	owl:versionInfo: rdfs:Literal [0..1]	The version indicator (name or identifier) of a resource.	Optional	Optional	Optional
Version notes	adms:versionNotes: rdfs:Literal [0..*]	A description of the differences between this version and a previous version of the Dataset.	Optional	Optional	Optional
Release date	dct:issued: rdfs:Literal: xsd:date [0..1]	The date of formal issuance (e.g. publication) of the Dataset.	Optional	Optional	Optional
Modification date	dct:modified: rdfs:Literal: xsd:date [0..1]	The most recent date on which the Dataset was changed or modified.	Optional	Optional	Optional
Type	dct:type: skos:Concept [1..1]	A type of the Dataset.	Optional	Optional	Mandatory

Class: Distribution / Sample / Analytics

Name of the Property HealthDCAT-AP	**RDF Example**	**Definition**	**Public** Property Type (HealthDCAT-AP) (Mandatory/ Recommended/ Optional)	**Restricted** Property Type (HealthDCAT-AP) (Mandatory/ Recommended/ Optional)	**Non-Public** Property Type (HealthDCAT-AP) (Mandatory/ Recommended/Optional)
Access URL	dcat:accessURL; rdfs:Resource [1..*]	A URL of the resource that gives access to a distribution of the dataset (e.g. landing page, feed, SPARQL endpoint).	Mandatory	Mandatory	Mandatory
Applicable Legislation	dcatap:applicableLegislation eli:LegalResource [1..*]	The legislation that mandates the creation or management of the distribution.	Mandatory	Mandatory	Mandatory
Description	dcterms:description; rdfs:Literal [0..*]	A free-text account of the distribution.	Recommended	Recommended	Recommended
Licence	dcterms:licence; dcterms:LicenceDocument [0..1]	A legal document under which the distribution is made available.	Recommended	Recommended	Recommended

Format	dcterms:format; dcterms:MediaTypeOrExtent [0..1]	The file format of the distribution.	Recommended	Recommended	Recommended
Title	dcterms:title; rdfs:Literal [0..*]	A name given to the distribution.	Optional	Optional	Optional
Language	dct:language: dct:LinguisticSystem [0..*]	A language used in the distribution.	Optional	Optional	Optional
Status	adms:status; skos:Concept [0..1]	The status of the distribution in the context of maturity lifecycle.	Optional	Optional	Optional
Temporal Resolution	dcat:temporalResolution; xsd:duration [0..*]	Minimum time period resolvable in the dataset distribution.	Optional	Optional	Optional
Spatial Resolution	dcat:spatialResolutionInMeters; xsd:decimal [0..*]	The minimum spatial separation resolvable in a dataset distribution, measured in meters.	Optional	Optional	Optional
Linked Schemas	dcterms:conformsTo; dcterms:Standard [0..*]	An established schema to which the described distribution conforms.	Optional	Optional	Optional
Documentation	foaf:page: foaf:Document [0..*]	A page or document about this distribution.	Optional	Optional	Optional
Download URL	dcat:downloadURL; rdfs:Resource [0..*]	A URL that is a direct link to a downloadable file in a given format.	Optional	Optional	Optional
Rights	dcterms:rights; dcterms:RightsStatement [0..1]	A statement that specifies rights associated with the distribution.	Optional	Optional	Optional
Access Service	dcat:accessService; dcat:DataService [0..*]	A data service that gives access to the distribution of the dataset.	Optional	Optional	Optional
Has Policy	odrl:hasPolicy; odrl:Policy [0..1]	The policy expressing the rights associated with the distribution if using the ODRL vocabulary.	Optional	Optional	Optional
Availability	dcat:availability; skos:Concept [0..1]	An indication how long it is planned to keep the distribution of the dataset available.	Optional	Optional	Optional

Release Date	dcterms:issued; rdfs:Literal; xsd:date [0..1]	The date of formal issuance (e.g., publication) of the distribution.	Optional	Optional	Optional
Modification Date	dcterms:modified; rdfs:Literal; xsd:date [0..1]	The most recent date on which the distribution was changed or modified.	Optional	Optional	Optional
Media Type	dcat:mediaType; dcterms:MediaType [0..1]	The media type of the distribution as defined in the official register of media types managed by IANA.	Optional	Optional	Optional
Packaging Format	dcat:packageFormat; dcterms:MediaType [0..1]	The format of the file in which one or more data files are grouped together, e.g. to enable a set of related files to be downloaded together.	Optional	Optional	Optional
Compression Format	dcat:compressFormat; dcterms:MediaType [0..1]	The format of the file in which the data is contained in a compressed form, e.g. to reduce the size of the downloadable file.	Optional	Optional	Optional
Byte Size	dcat:byteSize; rdfs:Literal; xsd:nonNegativeInteger [0..1]	The size of a distribution in bytes.	Optional	Optional	Optional
Checksum	spdx:checksum; spdx:Checksum [0..1]	A mechanism that can be used to verify that the contents of a distribution have not changed.	Optional	Optional	Optional

Published by the Health Information and Quality Authority (HIQA).

For further information, please contact:
Health Information and Quality Authority
George's Court
George's Lane
Smithfield
Dublin 7
D07 E98Y

+353 (0)1 814 7400
info@hiqa.ie
www.hiqa.ie