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

Report on the findings of a readiness assessment of data holders in Ireland in relation to health data access body services

March 2026

HealthData@IE: Deliverable 8.12 – Report on the findings of a readiness assessment of data holders in Ireland in relation to health data access body services

Deliverable number and name	8.12 – Publication of readiness assessment
Work Package number	Work Package 8
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

Acknowledgments

The Health Information and Quality Authority (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 A for a list of the organisations represented, and the Working Group, see Appendix B for membership list, as well as the individuals within the organisations listed in Appendix C, who took part in the readiness assessment. Your input was valuable in informing the development of the readiness assessment and supporting materials.

Members of the Health Information and Standards Directorate, HIQA

This report was led by staff in HIQA's Health Information and Standards Directorate. The following individuals contributed to the management, technical writing or dissemination of this report:

Rachel Flynn, Barbara Foley, Linda Weir, Maria Ryan, Ruth Benson, Kevin O'Carroll, Shauna McCarthy, Stephen Mulryan and Mark Casey.

The following individuals contributed to the initial research which informed the background to this report:

Elaine Meehan, Suzanne Barror, Shane Kelly and Connie O' Regan.

The Health Information and Quality Authority (HIQA) is an independent statutory body with responsibility for setting standards for all aspects of health information and monitoring compliance against those standards, as set out in Section 8(1) of the Health Act 2007. Under the Act, HIQA is also charged with evaluating the quality of the information available on health and social care and making recommendations to the Minister and the Health Service Executive (HSE) in relation to improving the health information system.

Conflicts of Interest

None reported.

Table of Contents

Glossary of Terms

Term	Definition
Data access application management system (DAAMS)	A system to receive, track and process applications and to issue permits.
Data holder	Any natural or legal person, public authority, agency or other body in the healthcare or care sectors, including reimbursement services where necessary, as well as any natural or legal person developing products or services intended for the health, healthcare or care sectors; developing or manufacturing wellness applications; performing research in relation to the healthcare or care sectors; or acting as a mortality registry; as well as any union institution, body, office or agency that has either: • the right or obligation, in accordance with applicable union or national law and in its capacity as a controller or joint controller, to process personal electronic health data for the provision of healthcare or care; or for the purposes of public health, reimbursement, research, innovation, policy-making, official statistics; patient safety or for regulatory purposes; or • the ability to make available non-personal electronic health data through the control of the technical design of a product and related services, including by registering, providing, restricting access to or exchanging such data.
Data linkage	The matching of records from two or more discrete datasets. It can be either deterministic, which compares the matching variables across datasets to find exact matches, or probabilistic which estimates the probability that records match.
Data permit	An administrative decision issued to a data user by a HDAB or data holder to process certain electronic health data specified in the data permit for specific secondary use purposes, based on the conditions set out in Chapter IV of the European Health Data Space (EHDS) Regulation.
Data quality	The degree to which characteristics of electronic health data are suitable for secondary use.
Dataset	A structured collection of electronic health data.

Term	Definition
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.
Health data access body (HDAB)	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.
HealthData@EU	An infrastructure project connecting national contact points for secondary use of electronic health data and the central EU platform.
HealthData@IE	An EU-funded project, led by the Department of Health in collaboration with HIQA and the HRB, focusing on developing national infrastructures needed for data access, including data access infrastructure systems that have been identified by the EU Commission as being core 'Digital Business Capabilities' for HDABs.
Interoperability	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.
Metadata	Information that describes other data. It helps to explain what the data is, how it can be used and where to find it.
National health dataset catalogue	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.
Primary use of health data	The processing of personal electronic health data for the provision of health services to assess, maintain or restore the

Term	Definition
	state of health of the natural person to whom that data relates; including the prescription, dispensation and provision of medicinal products and medical devices, as well as for relevant social security, administrative or reimbursement services.
Pseudonymisation	The process of replacing identifiable characteristics of data with a pseudonym, a value which does not allow the data subject to be directly identified.
Secondary use of health data	The processing of health data for purposes such as research, innovation, training and policy-making. The data used may include personal electronic health data initially collected in the context of primary use, but also electronic health data collected for the purpose of secondary use.
Secure processing environment (SPE)	The physical or virtual environment and organisational means to ensure compliance with European Union law; such as Regulation (EU) 2016/679, in particular with regard to data subjects' rights, intellectual property rights, commercial and statistical confidentiality, integrity and accessibility as well as with applicable national law; and to allow the entity providing the secure processing environment to determine and supervise all data processing actions including the display, storage, download and export of data, and the calculation of derivative data through computational algorithms.

Abbreviations

Abbreviation	Definition
ACHI	Australian Classification of Health Interventions
AHIMA	American Health Information Management Association
ANP	Advanced Nursing Practitioners
API	Application Programming Interface
ASTM	American Society for Testing and Materials
ATC	Anatomical Therapeutic Chemical classification system
CDM	Common Data Model
CSO	Central Statistics Office
CSV	Comma Separated Values
DAAMS	Data Access Application Management System
DCAT-AP	Data Catalogue vocabulary Application Profile
DDI	Data Documentation Initiative
DICOM	Digital Imaging and Communications in Medicine
DSMP	Dataset Specification Management Process
ECDC	European Centre for Disease prevention and Control
EHDS	European Health Data Space
EHR	Electronic Healthcare Record
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
GP	General Practitioner
HADEA	European Health and Digital Executive Agency
HDAB	Health Data Access Body
Health DCAT AP	Health Data CATalogue Application Profile
HI	Health Information
HIQA	Health Information and Quality Authority
HL7	Health Level 7 standards

Abbreviation	Definition
HPVP	Health Performance Visualisation Platform
HRCDC	Irish Health Research Consent Declaration Committee
HRB	Health Research Board
HSE	Health Service Executive
IACR	International Association of Cancer Registries
ICD	International Classification of Diseases
ICD-9-CM	International Classification of Diseases, 9th revision, Clinical Modification
ICD-10-AM	International Classification of Diseases, 10th revision, Australian Modification
ICD-11	International Classification of Diseases, 11th revision
ICT	Information and Communications Technology
ICU	Intensive Care Unit
IEC	International Electrotechnical Commission
IHI	Individual Health Identifier
ISO	International Organisation for Standardization
ISSDA	Irish Social Science Data Archive
JRC-ENCR	Joint Research Centre for European Network of Cancer Registries
LOINC	Logical Observation Identifiers Names and Codes
MDR	Metadata Registry
NOCA	National Office of Clinical Audit
NTPF	National Treatment Purchase Fund
OMOP	Observational Medical Outcomes Partnership
PPSN	Personal Public Service Number
SDMX	Statistical Data and Metadata eXchange
SNOMED CT	Systematized NOmenclature of MEDicine Clinical Terms
SPE	Secure Processing Environment

Abbreviation	Definition
TEHDAS	Towards the European Health Data Space
UKIACR	United Kingdom and Ireland Association of Cancer Registries
UNODC	United Nations Office on Drugs and Crime
XML	Extensible Markup Language

Executive summary

The HealthData@IE project is a key national project led by the Department of Health, in collaboration with HIQA and the Health Research Board (HRB), and also involving the HSE and key stakeholders, funded through the EU4Health programme. It aims to support the establishment of Health Data Access Body (HDAB) services in Ireland, as required by the European Health Data Space (EHDS) Regulation. Once established, a HDAB will securely connect data users, such as researchers and policy-makers, with anonymised health datasets to support research and innovation, education and training, policy-making, health service management, and preparation of national statistics.

As part of the HealthData@IE project, a readiness assessment was conducted to examine data holders' levels of preparedness for the establishment of HDAB services to identify where gaps exist and what steps need to be taken to ensure data holders in Ireland can meet future obligations under the EHDS regulation. To conduct the assessment, three use cases were developed in the areas of influenza, diabetes, and colorectal cancer. The use cases served two purposes. First, to identify associated datasets and data holders to participate in the readiness assessment survey and focus groups. Second, to help explain and demonstrate certain stages related to the preparation of data for secondary use, by providing clear examples that data holders, who process data specifically relating to each use case, could examine when participating in the assessment. The use cases were selected to align with those included in a HealthData@EU pilot programme and the minimum categories of electronic data for secondary use, as outlined in the EHDS Regulation.

The readiness assessment involved 16 data holders who participated providing responses for 21 datasets:

- Ten of these data holders were under the governance of, or held a service level agreement, with the HSE
- Six of these data holders were outside the governance of the HSE

The majority of the data holders (nine) were responsible for national data collections, with five eHealth services, one national survey and one hospital electronic patient record also involved. For a full breakdown of data holders and datasets included, please see Appendix C.

Data Holder EHDS preparedness and information management maturity

This readiness assessment has identified that several steps need to be taken across Ireland's health and social care system to prepare for the full extent of the EHDS Regulation's requirements. It is important to note that the data holders relevant for these use cases included some of Ireland's more well-established health data collections. It is postulated that many of the challenges experienced by the data holders involved in this readiness assessment, will also be common for other data holders in Ireland, if not amplified for the smaller data holders. Yet, even when considering that some of the data holders involved have mature systems and structures, the readiness assessment identified mutual challenges preventing effective data sharing and linkage to support the use cases.

The readiness assessment identified that preparation is required at both the national level and data holders' level, to ensure Ireland is ready for the implementation of the EHDS Regulation and to enable the effective establishment of a HDAB service. In terms of timelines, the regulation will be implemented on a phased basis, with the majority of datasets expected to be made available through the HDAB by 2029. Implementing acts will be published by 2027 to provide detailed rules and specifications to support data holders and the system to prepare for the operationalisation of the EHDS regulation.

Furthermore, the implementation of the *National Standards for Information Management in Health and Social Care* which were mandated by the Minister for Health in 2024, will complement the regulation and support data holders in maintaining good data management practices to ensure data is of sufficient quality to be shared with a HDAB service.

Data linkage and anonymisation

Through the application of the use cases, the readiness assessment identified that significant work is required on structures and processes to enable the successful sharing and linking of data through the HDAB. In order for data holders to comply with the obligations outlined in the EHDS Regulation (including preparing and submitting data within three months of a request), effective infrastructure, governance structures and data sharing practices are essential. This includes implementing a unique identifier to facilitate linking datasets, using effective methods to protect personal information before sharing data, and building capabilities to meet legislative obligations at each stage of the data sharing process.

The findings of this study indicate that, of the data holders involved, only 57% and 48% are using the Individual Health Identifier (IHI) and Personal Public Service Number (PPSN) respectively, to uniquely identify individuals. The lack of a unique identifier across all data holders was viewed as one of the biggest challenges to the establishment and effective operation of a HDAB. Further guidance to support the implementation of a unique identifier at national level, including across hospitals, will be a critical enabler for data holders to solve many of their challenges in uniquely identifying individuals, pseudonymisation and data linkage. Another important enabler will be providing clarity on the interactions between national legislation, EU legislation and the development of clear guidance for data holders on what changes are required to comply with the legal framework within their area of responsibility. Improved clarity at national level will enable data holders to build on the legal expertise within and across their organisations.

Data quality

Access to high-quality data is of paramount importance for achieving the EHDS regulation goals, both for primary and secondary use of information. Good decision-making for better patient care can only be informed by having access to high-quality data at every level of the system. Furthermore, the effectiveness and value of a HDAB service in Ireland will require a focused effort on enhancing data quality throughout the data lifecycle.

All data holders of publicly-funded datasets will have to review how they assess and report data quality using the EHDS data quality and utility label. This will be a significant body of work for all data holders, even for those with data quality frameworks in place, as they move to reporting their quality in line with the EHDS data quality and utility label. The implementation of the label will require that data holders review their current processes for assessing and reporting on data quality, the aim of which is not only to increase transparency but also, to promote continued data quality improvement activities. The use of advanced data quality frameworks by data holders will be required to provide a clear structure to enhance data quality practices. These frameworks will need to set out a data quality strategy, governance structures, roles and responsibilities, scheduled assessments, internal audits, key performance indicators and quality improvement initiatives. Together, these efforts will assist data holders in preparing for adoption of the data quality and utility label and inclusion of their label results in the national data catalogue, along with the additional obligations under the EHDS Regulation to prepare and provide data to the HDAB once requested.

Increased governance and oversight which focus on enhancing data quality is also a key requirement under the *National Standards for Information Management in Health and Social Care*. Building the skills and capacity of staff at every level is critical to ensuring consistency and high-quality data. Conducting a training-needs analysis will support data holders in resource planning, as well as the identification of areas for capacity building. Furthermore, the adoption of the national data dictionary, once finalised, will further enhance data quality practices by improving the coherence and comparability of data.

Standards for data discovery, semantic interoperability and data exchange

A national health dataset catalogue is needed to facilitate data discovery, a data access application management system (DAAMS) to receive, track and process applications and to issue permits, and secure processing environments (SPEs) to ensure the secure processing of health data. Data holders will need to establish if they are required to make datasets available and discoverable for secondary use through the HDAB service.

Many data holders will need upskilling to move to describing relevant datasets, in a standard way, using the Health DCAT-AP to ensure data is discoverable. National guidance, dedicated resources and the implementation of a unified strategy are also required in order to standardise data sharing, in line with semantic interoperability and data exchange standards. This will ensure a consistent approach to making data readily available and fit-for-purpose when requested. Additionally, targeted training programmes need to be designed for clinical coders and data managers to improve source data quality and semantic mapping skills.

Next steps

The implementation of the EHDS regulation in Ireland brings with it welcomed developments for our national health information system, expediting the need to enhance digital health records and services, such as the establishment of a HDAB in Ireland. It requires a concerted effort across Europe to standardise practice in key areas, putting Ireland in a unique position to learn from the experiences of many experts across Europe at such an optimal time of planning and innovation. Positively, the aim of the HealthData@IE project is to support the development of a HDAB service. This will be the first time a national linkage authority will be established in Ireland for health data, delivering many positive developments for data holders and data users.

The project focuses on the establishment of national infrastructures needed for data access through a HDAB service, including:

- a DAAMS to receive, track and process applications and to issue permits
- a National Health Dataset Catalogue to facilitate data discovery
- a SPE to ensure the secure processing of health data.

Further, the project is delivering workstreams on data quality enhancement, seeking to improve the quality of health data made available by a HDAB service, as well as a structured programme of training and engagement with data users, data holders and the public on the EHDS.

There have been positive developments in European and national policy, such as the Digital Decade, EHDS Regulation, Sláintecare, Harnessing Digital (the Digital Ireland Framework) and the Health Information (HI) Bill, which will drive changes in the health care system in Ireland. When considering the current landscape in Ireland and the readiness assessment findings, this report highlights the need for further alignment to prepare for and implement these changes, identifying specific requirements across the system. Implementation of a comprehensive legal framework, including the enactment of the HI Bill and the EHDS Implementing Acts in 2027, will be a critical step in promoting the effective use of health data in Ireland.

Data holders need detailed rules and specifications to support changes in practices and to prepare for the operationalisation of the EHDS regulation and the establishment of a HDAB service. Governance of a HDAB, once established, is another significant factor that which will have implications for data holders and will need to be set out in national policy to ensure clarity in terms of roles and responsibilities.

To complement the implementation of the EHDS Regulation, there have also been positive strategic developments around health information. For example, the *Digital for Care: A Digital Health Framework for Ireland 2024-2030* and *Digital Health Strategic Implementation Roadmap*, both set out a plan for the integration of digital technologies in the Irish healthcare system, and aim to address the challenges for data holders in sharing data. The scope of this framework is mainly based on the primary use of data. Although enhancing digital records for primary use will positively affect the secondary use of information, a clear strategy to bridge these two uses is required. Furthermore, the current fragmented and siloed national health information sources require concerted efforts are needed across our health information system to achieve the technological and governance requirements, set

out by the EHDS Regulation. This need for broader strategic direction, specifically in the secondary use of information, is evident in reviewing the considerations for the establishment of a HDAB.

The HealthData@IE grant has been funded to address some of the specific gaps identified by data holders through the readiness assessment. This will build on progress made by organisations that have a remit in the area of health information. For example, HIQA currently has a remit to set standards for health information and monitor compliance against those standards, as well as to evaluate the quality of the health information available and make relevant recommendations to the Minister for Health and the HSE on these findings. The HealthData@IE project will also build on ongoing work to enhance data quality by developing various data quality guidance and training materials, hosting a peer-support network for data holders on the topic of data quality and developing a national interoperability framework. It will also establish a National Health Dataset Catalogue and the required infrastructure for processing applications and securely providing access to data. The outputs of this project will be made publicly available and ultimately aim to support the secondary use of health data in Ireland.

However, there are broader concerns identified by data holders, beyond the scope of the HealthData@IE project, which require further consideration, such as a national approach to guide the comprehensive implementation of a unique identifier. A clear roadmap to guide progress through the HealthData@IE project and beyond is required, identifying which changes should be addressed at a local level with data holders, and which changes should be addressed at a national level. This roadmap will be critical to building preparedness to the levels needed for the successful implementation of a HDAB service in Ireland.

1. Background

1.1 The European Health Data Space

The European data strategy, announced in February 2020, aims to build nine common data spaces to forge a single EU-wide data market. The first data space to be established is for health, the EHDS.[2] The 2022 Data Governance Act and 2024 Data Act establish its legal and regulatory backbone, while the Digital Decade programme sets targets for public service digitalisation and universal online medical record access by 2030.[3-5] The EHDS Regulation came into force in March 2025 and together with the Data Governance Act and the Data Act will function to promote secure and trustworthy use of data. A phased approach to the implementation of the

regulation has been adopted, with set timeframes for specific categories of data to be made available for secondary use between 2029 and 2031.

Once fully implemented, the EHDS Regulation will empower individuals to control their health data and will simplify cross-border exchange for care (primary use), and sharing of data for research, innovation, public health, policy-making and personalised medicine (secondary use). Pilot initiatives such as EHDS2[*] and the Joint Action Towards the European Health Data Space (TEDHAS1[†] and TEHDAS2)[‡] have prepared the principles, and are establishing recommendations, guidelines and technical specifications for the implementation of EHDS.[6-8] Building on the General Data Protection Regulation (GDPR), the Data Governance Act and the Data Act, EHDS also requires Member States to prepare their legal and operational frameworks. In Ireland, the HI Bill 2024 is part of planned legislative measures to give full effect to the EHDS Regulation.[9]

1.2 HealthData@IE 2023-2027

In respect of the secondary use of data, the EHDS Regulation places an obligation on Member States to establish the services of one or more HDAB. A HDAB service securely connects data users, such as researchers and policy-makers, with anonymised and pseudonymised health datasets to support research and innovation, education and training, policy-making, health service management and preparation of national statistics. The Department of Health, in collaboration with HIQA and the HRB, was awarded funding for the HealthData@IE project under the EU4Health programme to support the establishment of HDAB services in Ireland.[10]

Working with key stakeholders across the health system, the HealthData@IE project will focus on the development of national infrastructures needed for data access, including data access infrastructure systems that have been identified by the European Commission as possessing core digital business capabilities for HDABs. These include a national health dataset catalogue to facilitate data discovery, a DAAMS to receive, track and process applications and to issue permits, and SPEs to ensure the secure processing of health data. The HealthData@IE project will also deliver important programmes of work centred on data quality enhancement,

---

[*] EHDS2 Pilot Project: piloting connecting data platforms in a network infrastructure and developing services supporting the user journey for research projects using health data from various EU Member States.
[†] TEHDAS1: developing and promoting concepts for sharing of data in secondary use purposes for citizens' health, public health as well as health research and innovation in Europe.
[‡] TEHDAS2: developing guidelines and technical specifications for implementation of secondary use.

engagement and dissemination, as well as training and education for data users, data holders, HDAB staff and members of the public.

2. Introduction

The HealthData@IE project was funded to support the establishment of a HDAB. The project will focus on the development of national infrastructures needed for data access including a DAAMS to receive, track and process applications and to issue permits, a National Health Dataset Catalogue to facilitate data discovery, as well as a SPEs to ensure the secure processing of health data. Further, the project will deliver work streams on data quality enhancement to improve the quality of health data made available by HDABs, as well as a structured programme of training and engagement with data users, data holders and the public on the EHDS. An emphasis is placed on supporting data holders to prepare for the EHDS and the establishment of HDAB services in Ireland, ensuring they can meet their obligations relating to the secondary use of data. In this respect, one of the first steps is undertaking a readiness assessment.

2.1 Purpose of the readiness assessment

The overall **aim** of this readiness assessment is to obtain a point-in-time view of data holders' levels of preparedness for the establishment of HDAB services in Ireland, to identify where gaps exists and to determine what steps must be taken to ensure Irish data holders can meet future obligations, under the EHDS Regulation.[11]

The specific **objectives** of the readiness assessment are:

- To explore the feasibility of re-using and linking health and social care data from different sources for secondary use purposes in Ireland, by assessing if potential linkage variables are present across datasets.
- To identify whether there is capability and capacity among data holders to provide metadata and data in the necessary formats, to a future HDAB service.
- To determine the strengths and weaknesses of data holders' existing Information and Communications Technology (ICT) systems' interoperability and their ability to support discovery of their data, coding of data and the exchange of data with a future HDAB service.
- To raise awareness among data holders of what their obligations will be under the EHDS, and to explore perceived challenges and enablers to the implementation of the EHDS and the establishment of HDAB services in Ireland.
- To identify and prioritise data holders' data quality guidance and training needs.

- To identify the steps required and to outline a pathway to support data holders to make the required changes to ensure compliance with the EHDS Regulation.

The aim and objectives were achieved by developing three **use cases** (see Section 2.3 for more information on use cases). This readiness assessment will help to explore if adequate structures and processes are in place, to enable the secondary use of data across data holders, as well as identifying additional possible facilitators of overall system preparedness. The EHDS Regulation sets out a number of obligations which data holders will need to meet in order to achieve compliance.

Obligations include:

- **Identification of datasets** – review by data holders to establish if they are required to make datasets available for secondary use through the HDAB.
- **Making data discoverable** – information on relevant datasets will need to be published, and revised annually, through the national health dataset catalogue.
- **Describing data** – data holders are required to provide a description of the relevant datasets, through standardised reporting, using the metadata standard Health DCAT-AP.
- **Reporting data quality** – data holders will have to access and report the data quality using a data quality and utility label annually, through the national dataset catalogue.[12]
- **Submitting data** – once data is requested through the HDAB service, data holders will need to prepare and submit the requested data within three months.

Another obligation, which is beyond the function of the HDAB and aligns with Open Data Strategy, is the requirement for data holders to make non-personal data available through open databases.[13] In order to comply with these obligations, organisations will need to put effective governance structures and arrangements in place to meet the obligations, and to prepare the data for submission to a HDAB service and for sharing on open databases. This includes adopting a national unique identifier across datasets to enable linkage, aligning with the recommended use of standards to ensure interoperability, enabling exchange of data according to European guidelines and enhancing data quality to ensure data is fit for purpose for data users.

2.2 Purpose of the report

The purpose of this report is to present the key findings of the readiness assessment. Participants of this readiness assessment were selected based on the use cases and included some of Ireland's largest and well-established health data holders. The findings will help to inform implementation of the EHDS Regulation and policy development, preparing organisations for the establishment of a HDAB and the implementation of the EHDS. These findings will also inform many of the HealthData@IE deliverables for Ireland.

2.3 Methods and participants

Engagement with data holders in Ireland was key to conducting this readiness assessment. The methodological approach to this engagement involved an online survey and follow-up focus groups with 16 data holders.

As part of the readiness assessment, three use cases were developed which served two purposes. First, to identify associated datasets and data holders to participate in the readiness assessment survey and focus groups. Second, to help explain and demonstrate certain stages related to the preparation of data for secondary use by providing clear examples that data holders, who process data specifically relating to each use case, could examine when participating in the assessment.

The use cases were selected to align with those included in a HealthData@EU Pilot and the minimum categories of electronic data for secondary use, as outlined in the EHDS Regulation.[6] The topics of the Data Access Sharing Storage Linkage (DASSL) project case studies were also taken into consideration.[14] Steering Group§ and Working Group** members involved in the HealthData@IE project also provided input in the use case development process.

The use cases are outlined below, and a detailed description of the three use cases selected as part of the HealthData@IE project, can be found within the protocol.[11] Relevant definitions and considerations under the EHDS are set out in **Appendix D**.

---

§ The Steering Group, chaired by the Department of Health, supports the establishment of a health data access service in Ireland and acts as a steering committee for the development of digital capabilities under EU EU4H-2022-DGA-MS-IBA-4 — Direct grants to Member States: for setting up services by Health Data Access Bodies - Secondary use of health data.

** The Working Group, chaired by Health Information and Standards (HIS) Deputy Director, comprises of representatives from data holders, policy and government organisations. The group actively contribute to the completion of HealthData@IE data quality outputs through the provision of expert advice, providing feedback and supporting stakeholder engagement.

Detail on use cases

The three use cases for the readiness assessment are in the areas of: colorectal cancer, diabetes and influenza. Through the three use cases, 16 data holders were identified as being relevant and invited to take part initially in the survey and then subsequent focus groups, see **Appendix C**.

Colorectal cancer use case: to demonstrate the feasibility of linking clinical and genomic data to enhance our understanding of outcomes for people with colorectal cancer, including incidence, risk factors (for example, lifestyle, environment and genetic factors) and aetiology.

Diabetes use case: to demonstrate the feasibility of using available data to enhance our understanding of the provision of care for people with diabetes (type 1 and type 2), to compare care pathways, measure clinical outcomes, costs of care, and enable better planning of services and outcomes of care.

Influenza use case: to demonstrate the feasibility of using available data to carry out surveillance of influenza and explore rates of influenza testing, vaccination and hospitalisation in vulnerable groups of the population (for example, older adults).

The readiness assessment was conducted in four phases, as outlined below.

Phase 1	Phase 2	Phase 3	Phase 4
Design of survey	Conduct survey	Focus groups	Report on findings
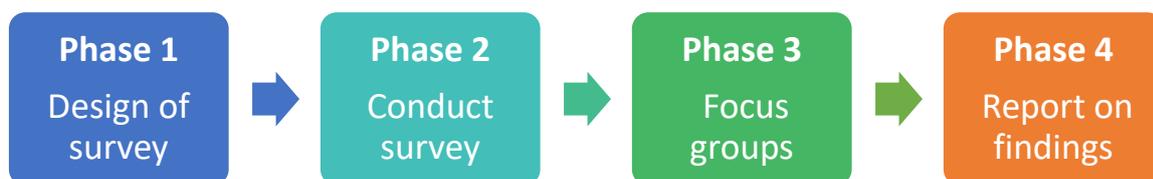
Phase 1 – Design of survey and supporting materials

The first phase of the readiness assessment included the design of the survey. The purpose of the survey was to gather information from data holders on a range of topics including general preparedness for the EHDS, use of identifier variables and de-identification techniques, use of standards for interoperability, current data quality practices, and data quality training and guidance needs.

The survey questions were developed based on requirements set out under the EHDS Regulation, findings of TEHDAS and the EDHS-2 pilot, and through exploration with the Working Group and Steering Group. **Table 1** sets out the overall composition of the survey. The survey included both quantitative and qualitative response options, with a total of 44 questions across six sections.

Table 1. Readiness Assessment Survey composition

Section	Scope of questions	Purpose
1	Details of the data holder and their dataset	To obtain an overview of the characteristics of the survey participants.
2	General EHDS preparedness and information management maturity	To assess overall preparedness for the implementation of the EHDS and the establishment of HDAB services in Ireland, as well as information management maturity to support such practices.
3	Data linkage and anonymisation	To ascertain if potential linkage variables are present across datasets, and assess current practices with regard to data linkage and anonymisation methods.
4	Data quality	To explore current data quality practices and levels of data quality expertise among staff.
5	Standards	To explore readiness to support and implement data discovery standards, standards for semantic interoperability and data exchange standards.
6	Data quality training and guidance	To assess preferences and priorities for future data quality training and guidance.

Phase 2 – Conducting the survey

The data holders (n=16) responsible for the selected datasets (n=21), identified through the use cases, were invited to partake in the survey. Data holders were asked to provide one survey response per dataset. As a result, some data holders submitted multiple survey responses relating to separate datasets; see **Appendix C** for a full list of the data holders who responded to the survey and the associated

dataset. Resources and information shared with participants to support their involvement in the readiness assessment, included a guidance for data holders, readiness assessment protocol and overviews of the three use cases.[11, 15]

To further support data holders in preparing to complete the readiness assessment, HIQA hosted an information session prior to the survey's launch in April 2025. Data holders were given six weeks to complete the online survey from 16 April to 28 May 2025.

Phase 3 – Focus groups

In June 2025, the Project Team hosted three focus groups with participants who responded to the survey. A total of 23 individuals, representing 16 data holders, took part in these focus groups. The aim of these focus groups was to explore in more detail the findings from the survey, specifically in the context of the use cases, addressing the challenges to secondary use of data in these contexts, and to further explore the perceived barriers and facilitators to the implementation of the EHDS and the establishment of HDAB services in Ireland.

Each focus group concentrated on one of the three use cases, exploring the themes identified from the survey including, EHDS preparedness; data linkage and anonymisation; data quality; and standards for data discovery, semantic interoperability and data exchange. These themes were used as a basis to undertake a framework analysis of the focus group data.[16] All participants that took part in the focus groups provided signed or verbal consent for audio recording of the focus groups.

Additionally, two focus groups were held in June 2025 with members of the Working Group. The purpose of these focus groups was to ascertain the capacity and skills requirements to support the establishment of HDAB services in Ireland. The findings of this engagement will support HIQA and the HealthData@IE project to improve awareness and understanding of the current challenges faced by data holders, identifying potential skills and capacity gaps.

Phase 4 – Reporting the findings

The online survey questions included a mix of closed-ended, binary (yes/no/don't know) questions, Likert-type scale questions and multiple-choice questions, of which responses were analysed using descriptive statistics. Responses to open-ended survey questions, as well as data collected during the focus groups and interviews, were analysed using a content analysis approach. The findings from the survey and focus groups are presented in this report.

3. Finding of the readiness assessment

This chapter provides an overview and summary of the feedback and responses received as part of the readiness assessment survey and focus groups. The survey achieved a 100% response rate as all invited participants completed the survey, including 21 responses; one for each identified dataset, (provided by 16 data holders).

3.1 General information relating to participants

Participants were asked to identify their data holder organisation, current role of the person completing the survey, the name of the dataset which their response pertains to and their contact email. A list of data holders and associated datasets can be viewed in **Appendix C**. Reporting of survey results is categorised either at an individual dataset level, or at the data holder level, whichever is more appropriate. In total, of the datasets included in the readiness assessment, 11 were under the governance of the HSE or had a service level agreement with the HSE, and 10 were outside the governance of the HSE. The majority of the data holders (nine) were responsible for national data collections, with five eHealth services, one national survey and one hospital electronic patient record also involved.

3.2 Feasibility of establishing data to support the use cases

The findings of the use case analysis indicate that data holders face similar challenges in preparing data for submission to the HDAB, some of which will be addressed through the establishment of the HDAB and the HealthData@IE project; while other data holders encounter challenges related to broader landscape developments. While individual data holders' readiness and practices varied widely, data holders identified similar challenges across use cases. When the data was analysed and compared by use case, the results showed no significant difference in the overall experience across the three use cases.

The addressing of readiness challenges currently lies with each individual data holder; however, to maximise efficiencies in sharing and re-use of data, a coordinated and national strategic approach will be required to structure and prioritise the implementation of secondary use of data, under the EHDS Regulation. Improving the value of data through the ability to effectively link data for the use cases is fundamentally linked to improving each contributing dataset whose data holders have independent practices, standards, technical infrastructure, and staff capabilities. As participants across use cases explained similar challenges and

facilitators, key themes were identified in their experiences. The findings will be reported in the following sections based on common themes, including:

- EHDS preparedness and information management maturity
- data linkage and anonymisation
- data quality
- standards for data discovery, semantic interoperability and data exchange
- training and guidance.

The following sections will explore these themes in greater detail, with a summary of significant findings at the end of each section.

3.3 EHDS preparedness and information management maturity across data holders

This section of the report outlines responses on how well data holders are prepared for the implementation of the EHDS Regulation and establishment of HDAB services in Ireland, in terms of putting in place the appropriate structures and processes to share their metadata and data for secondary use.

Furthermore, in line with the publication of HIQA's 2024 *National Standards for Information Management in Health and Social Care*, participants were asked questions to establish levels of information management maturity.[17] The relevant findings are reported below.

3.3.1 EHDS preparedness

When exploring the qualitative data, the overall findings indicate that many data holders are at an early stage of preparedness. In the qualitative survey responses and focus groups, participants reported reviewing existing governance groups or data management teams, evaluating overall systems readiness, examining improvements to metadata documentation and scoping organisational needs for additional resources. Others reported that preparations such as the development of new ICT systems (while not directly initiated in preparation for the EHDS Regulation), would help to meet specific requirements as set out under the regulation. Other participants were able to demonstrate various steps taken to prepare for the EHDS; some of these examples have not been implemented yet and participants indicated that significant work would be required to achieve compliance with the regulation.

The survey findings show that 11 out of 16 participants (70%) reported average to high levels of knowledge of the secondary use of data under EHDS and 87% (14 participants) reported having taken initial steps to prepare for the EHDS Regulation,

see **Figure 1**. A number of participants stated that their involvement in or participation with the HealthData@IE project was valuable in supporting their preparation for the upcoming regulatory changes. One other participant stated their involvement in another EU initiative also aligns with the changes required to meet their EHDS obligations. The data holders directly involved in EHDS or other EU initiatives appear to have a staff member responsible for ensuring the organisation takes steps to prepare for the EHDS. However, five participants reported that their organisation does not have a specific resource dedicated to this purpose.

"We are familiarising ourselves with the EHDS requirements, looking at what is needed and what we already have in place. We are at the 'getting our heads around it' phase."

"No significant steps have been taken to comply with the regulation but […] Information Governance Framework has been reviewed with a focus on data quality for interoperability."

Figure 1: Level of knowledge of secondary use of data

LEVEL OF KNOWLEDGE OF SECONDARY USE OF DATA (EHDS)

TOTAL ORGANISATIONS, 16

- 6.3% (n=1) — VERY LOW
- 25.0% (n=4) — LOW
- 37.5% (n=6) — AVERAGE
- 18.8% (n=3) — HIGH
- 12.5% (n=2) — VERY HIGH
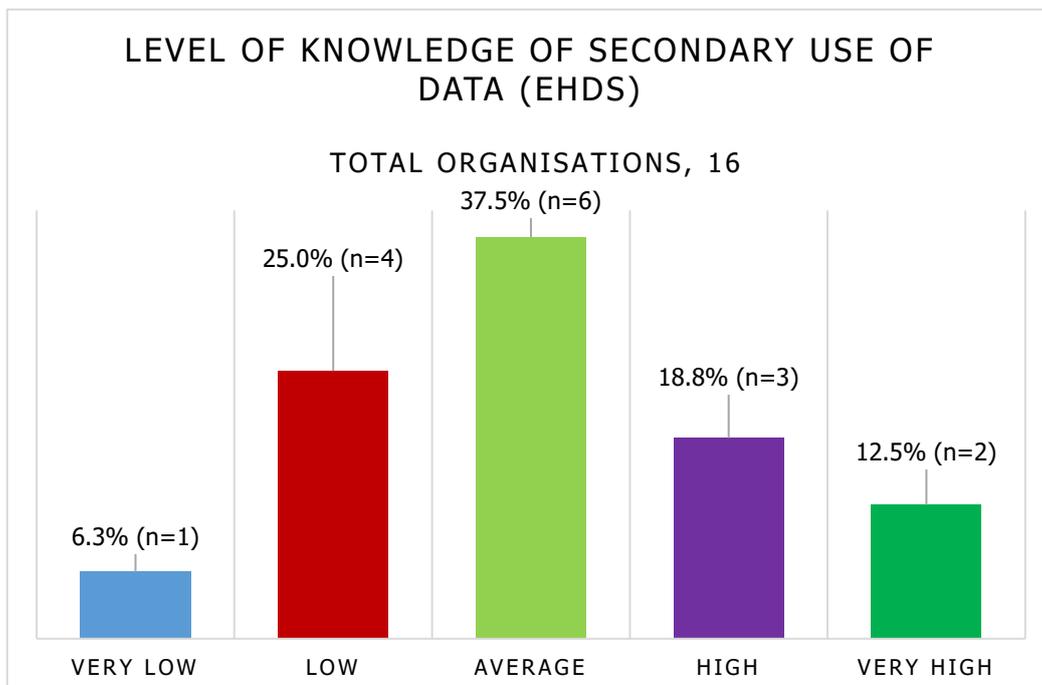
While participants broadly welcomed the vision of EHDS, some concerns were raised regarding current level of preparedness for the regulation due to uncertainty regarding their future legal remit. It was widely acknowledged that the absence of an enacted legal framework (including the HI Bill, and the implementing acts and delegated acts to give further effect to the EHDS regulation) is currently a primary challenge to effectively sharing and using health data in Ireland. The enactment of the HI Bill was seen by participants as a critical step to moving this agenda forward.

In particular, uncertainty was expressed about the roles of data holders and questions were raised regarding how current legal obligations, especially under the GDPR, would align with EHDS requirements and the HI Bill. In the absence of the specific detail regarding future requirements, there was general hesitancy to make significant changes. With the introduction of new regulations and legislation, data holders indicated a clear need for guidance and supports explaining these legal requirements and how they interact with one another for data holders to be able to progress preparedness for the EHDS.

In many cases, data sharing was reported to vary across organisations, which is often linked to governance and funding models as well as the associated skills and capacity. Participants called for better coordination across sectors and for universal support mechanisms to be provided for staff, in order to standardise and streamline practice. It was also acknowledged that to prepare for the implementation of a HDAB, clear and timely communication with the public and professionals is required, which staff need the appropriate guidance and training to achieve.

3.3.2 National Standards for Information Management in Health and Social Care

The vast majority of data holders (87%, 14 participants) reported average to high levels of knowledge of the *National Standards for Information Management in Health and Social Care*.[17] However, in the qualitative findings, participants indicated that they did not have any systematic or structured method of applying the standards in their organisation and that in some instances the standards were being applied by individual members of staff. Another participant noted that they have completed a systematic mapping exercise to assess where revisions and updates are required to internal policies and procedures to align with the standards. However, participants strongly emphasised the need to strategically improve skills and capacity in information management across the entire system. As the readiness assessment included some of the more well-established and mature data holders, there was also an acknowledgment of the benefit of data holders learning from each other through peer support networks.

"Technical skills are required. Health information management is a science in itself, but it's often an add-on to a role. You can't expect people to just be able to add as part of their role seamlessly. We need a recognition that there is more to it that just entering data."

"Organisations are usually risk averse, so they need guidance on how to share data safely and securely. A support network or peer network would be of huge benefit."

Participants were asked how they would overall rank the level of maturity of their organisation with regard to information management. The results show that 50% (eight participants) reported having either a managed or an optimised level of maturity, nearly 30% (five participants) reported having a defined level of maturity, while less than 20% (three participants) had either an initial or a repeatable level of maturity, see **Figure 2**. This finding further demonstrates the benefit of reviewing the practice of those organisations identifying as optimised, and managed to promote good practices and learning across the less-mature organisations.

Figure 2: Organisational maturity with regard to information management practices

Level of maturity on information management

Total organisations, 16

Maturity level	Percentage
Optimised	6.25% (n=1)
Managed	43.75% (n=7)
Defined	31.25% (n=5)
Repeatable	12.50% (n=2)
Initial	6.25% (n=1)
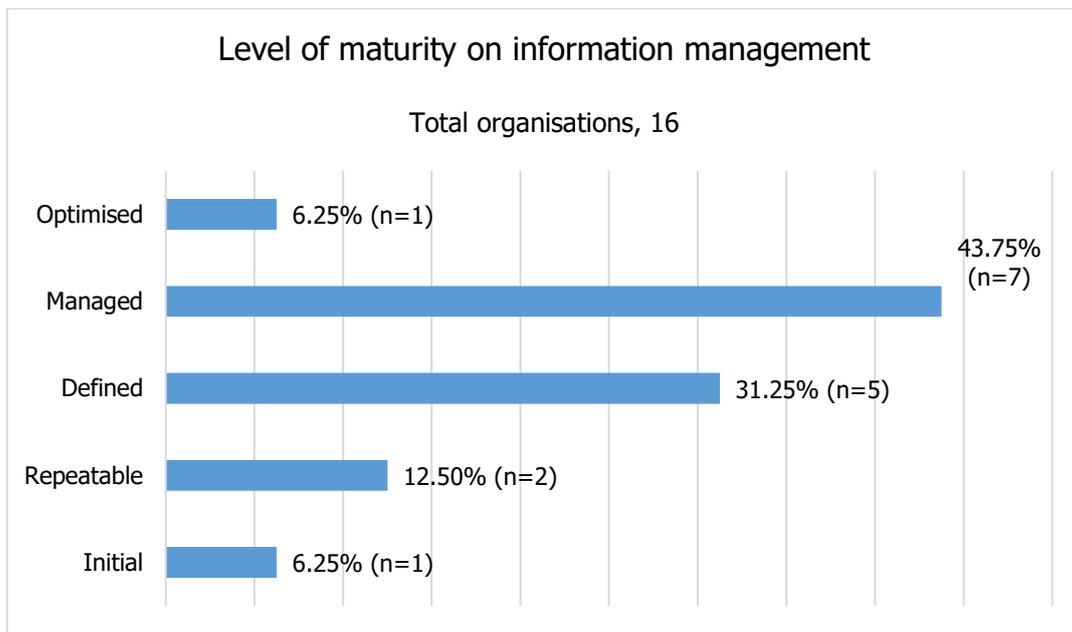
3.3.3 Significance of findings – Use cases, EHDS preparedness and information management maturity

The findings clearly demonstrate that the ability to study a health data use case is built on a foundation of individual data holders' practices. Currently, gaining a comprehensive understanding of the use cases is either not possible or very difficult due to the wide variation in data holder practices and systemic challenges.

Although participants of the survey reported initiating preparations for the EHDS Regulation, the qualitative findings from both survey responses and focus groups identified that significant steps need to be taken to prepare for the full depth of the regulation's requirements. The findings also show that data holders are awaiting policy direction and clear guidance before strategically planning, prioritising and implementing the significant changes needed for compliance with the regulation. While many participants reported progress in implementing *National Standards for Information Management in Health and Social Care*, the findings suggest that the degree of progress is linked to the maturity of the organisation's structure, systems, policies and procedures in this area. Enhancing compliance with these national standards will ultimately help with preparations for the establishment of the HDAB and implementation of the EHDS.

To put these findings in context, it is important to note that organisations which were included in the readiness assessment are more well-established data holders in Ireland, and several of these organisations are involved in the HealthData@IE project in various capacities. Therefore, these findings may be more positively biased towards the data holders that are more aware of the forthcoming requirements and hence may be more prepared, when compared to the wider population of data holders in Ireland. It is generally acknowledged that data holders will need significant supports, guidance and training to ensure they are ready for EHDS requirements and the effective establishment of a HDAB.

Improving the value of data, for example, by enabling the effective sharing and linkage of data for the use cases included in this study, will require clear strategic direction and supporting policy to decide how to address the depth of challenges faced by data holders. Although some of these issues will be addressed through the HealthData@IE project, there are broader concerns linked to the maturity of digital health data and a plan to bridge the link between primary and secondary data which require further attention at a national level. Identification of these through a clear roadmap for change is required, ascertaining which challenges can be addressed at a local level with data holders and which challenges should be addressed at a national level. This roadmap will be critical to building the level of preparedness required for implementation of HDAB services.

3.4 Data linkage and anonymisation

Improved usage of health data is a core goal of EHDS. By facilitating systematic access to health data, EHDS will enable researchers, policy-makers and healthcare managers to combine multiple datasets to reach stronger conclusions and drive greater understanding across the health system. The purpose of this section is to assess the data holders' readiness to facilitate effective data usage through the ability to combine datasets and anonymise data, making data available for secondary use under strict privacy and security safeguards.

3.4.1 Data Linkage Practices

Data linkage is the process of combining information from different datasets which relate to the same individual or individuals. It involves the secure identification of individuals across multiple datasets, then linking their data together to enhance the value of the data. For example, up to ten of the datasets surveyed could potentially be linked in the colorectal cancer use case. These consist of seven national data collections, one hospital EHR system and two eHealth services. These datasets are also governed by multiple organisations with five under the HSE, two under the governance of the HRB and the remaining three governed by other data holders, see **Appendix C** for more details.

In reviewing data linkage practices across the three use cases, it is clear that the health system and data holders in general are not sufficiently prepared to link data or have their data linked to other datasets via a HDAB, although some variation exists across data holders. Qualitative responses indicate that currently linking of datasets is either not taking place, or in instances where linkage does take place, it can be quite difficult to carry out and requires specific expertise. Many data holders highlighted that they approach data linkage with caution, or in some cases avoid it altogether due to uncertainty regarding legal remit or concerns around potential re-identification of data subjects. While some elements of linkage are improving, such as identification processes, the data linkage process faces ongoing challenges. The findings across use cases indicate that significant progress is required to improve the linking of data in Ireland.

Across the three use cases, data holders faced common challenges linking data. The findings indicate that variation in the use and collection of a unique identifier is a root cause to the challenges in linking data. Participants noted that a unique identifier is not consistently present in datasets and not always captured at the point of data collection, which leads to difficulty in linking data. In some instances,

participants explained how datasets must be individually reviewed and manually linked, involving complex and time-consuming processes.

Variation in practices were also highlighted in data holders' approaches to anonymisation, which is a process used to enable safe data sharing and to mitigate the risk of individual identification within a dataset. EHDS is aligned with personal identification protection requirements outlined under GDPR and expects the use of anonymisation methods. EHDS also introduces new obligations for data holders to supply certain datasets in anonymised formats to open data platforms. The research found that practices vary across data holders, with some lacking expertise in this area, undermining confidence in the preparation of data for linkage and sharing for secondary uses. This leads to a greater focus on governance controls and data usage agreements, with some reporting that this increases complexity throughout the data sharing process.

3.4.2 Identification for Data Linkage

Identifying individuals across datasets is the first step in the data linkage process. The survey identified that the most common variables used to identify individuals were date of birth (90%, 19 participants), sex (71%, 17 participants), first and last name (67%, 14 participants) and 'other' (71%, 15 participants), which covered a range of organisation-specific information, such as hospital health record numbers or new identifiers created by the data holders, see **Figure 3**.

Figure 3: Variables used to uniquely identify individuals

Variables used to uniquely identify individuals

Total Datasets, 21

Variable	Percentage
Date of Birth	90%
Sex	81%
Other	71%
First Name	67%
Last Name	67%
Individual Health Identifier (IHI)	57%
Address Line 1	57%
Gender	48%
Eircode	48%
Personal Public Service number (PPSN)	48%
Mother's Birth Family Name	14%
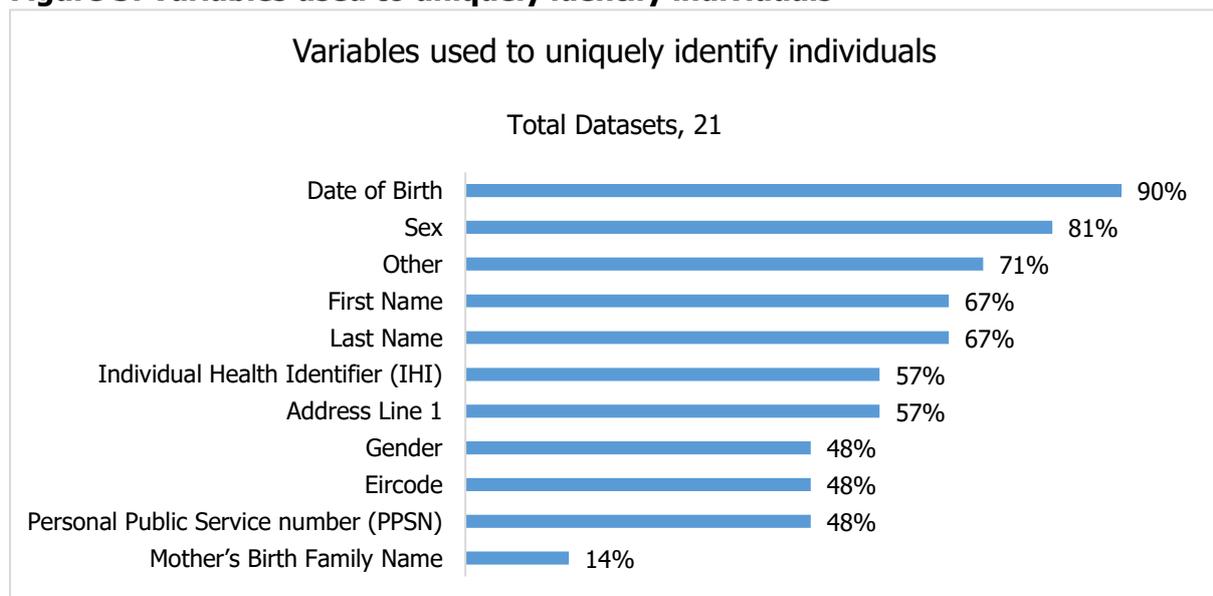
The findings indicate that different combinations of variables are currently used by data holders to anonymise data, resulting in significant challenges for data linkage. Furthermore, results of this readiness assessment show that 57% (12 participants) are using the IHI and 48% (10 participants) are currently using the PPS number. Adoption of a national unique identifier such as the IHI or PPSN as a common data linkage variable across datasets would be a key enabler for data sharing and linking; both important goals of EHDS.

Furthermore, while 57% (12 participants) report using the IHI, qualitative data from focus groups reveals that this process is more nuanced and involves multiple stages including, preparing the dataset for IHI; implementing IHI in data collection; and verifying that the collected IHIs match the individuals in the dataset. Several participants noted that they have prepared their datasets for inclusion of the IHI, but have not yet completed collection or verification of the IHI. One participant reported testing IHI verification with success rates of 98% or more, while other participants experienced difficulties with insufficient information often related to the IHI not being collected at source, preventing them from incorporating and using the IHI in their datasets.

For example, participants identified a significant variation in the implementation and use of the IHI at a hospital level. When an individual visits multiple hospitals as part of their care plan, each assigns a different local identifier or hospital number, thereby causing challenges with linking the patient's records across different settings. This creates a unique situation where patients' records cannot be linked across hospitals and other care settings in Ireland. The mandatory use of a unique identifier across all settings will resolve this. This has knock-on effects for use of data collected for both primary use and secondary use purposes.

The use of the PPSN as a unique identifier was reported by 48% (10 participants). One participant reported experiencing challenges using PPSN as an identifier due to duplication within the system, such as duplicate numbers or individuals with multiple numbers, leading to incorrect identification of individuals. Other participants stated that they lacked the legal basis to collect PPSN. The Health Information Bill introduces new provisions around the collection of the PPSN which aims to resolve this challenge.

Collecting PPSN in combination with IHI often led to strong IHI verification rates but relies on the presence of PPSN during data collection. Combined with concerns around legislative remit for collection and storage of unique identifiers, these experiences have led some data holders to delay using such unique identifiers. Participants also reported variation in practices in terms of the use of

identifiers, such as replacing data with organisation-specific or project-specific identifiers when sharing data. Notably, this practice may help to anonymise or pseudonymise data; however, it is not conducive to effective data linkage.

> *"Until we reach a point where we are happy with the data coverage and we also have data protection guidance about how we store and use these identifiers, we won't use them."*

While almost all participants could identify individuals within their data, the ability for data holders to handle an individual data subject's preference for secondary use is underdeveloped. Over 70% (15 participants) did not record an individual's preferences for data processing for secondary use. EHDS mandates the provision of a system to allow individuals to opt out of processing of their personal data for secondary use. While EHDS has provisions for public sector bodies for public interests within Member States to access health data for secondary use regardless of individual preference, the opt-out mechanism and preferences must still be used for all other types of data access requests, for example, health research projects. This may present challenges for data holders who previously did not need to handle individual preferences due to working exclusively in public interest since their data can now be requested by a much broader scope of data users. Additionally, EHDS will introduce new obligations for data holders to upload certain datasets to open data portals. Consequently, ensuring clarity for data holders on the impact of opt-out preferences will be essential.

> *"One of the key pieces is the availability of that information on the hospital system, whatever PAS system they are using. We are not separate, we are totally reliant on that, we can't collect that unless it's on those systems and we can pull it down. There is a whole piece about the quality, completeness and specificity around those source systems that we rely on."*

> *"Number one is that we need an individual health identifier and have it used within the clinical system. This issue is causing so many issues downstream."*

3.4.3 Performing Dataset Linking

Once individuals have been reliably identified within multiple datasets, their information from each dataset is linked together for further analysis. This is a fundamental activity in health data research and a critical component of EHDS's goal to improve data usage.

Results from the survey found that at least one third of participants (33%, seven participants) do not link their datasets with others. A further third (33%, seven participants) reported that datasets are routinely linked to other datasets, while the remaining 29% (six participants) link their datasets on an ad-hoc basis, see **Figure 4**. Combined, these results indicate that data sharing and combined usage is not taking place at the levels required for comprehensive information on each use case.

Figure 4: Is this dataset ever linked to other datasets?
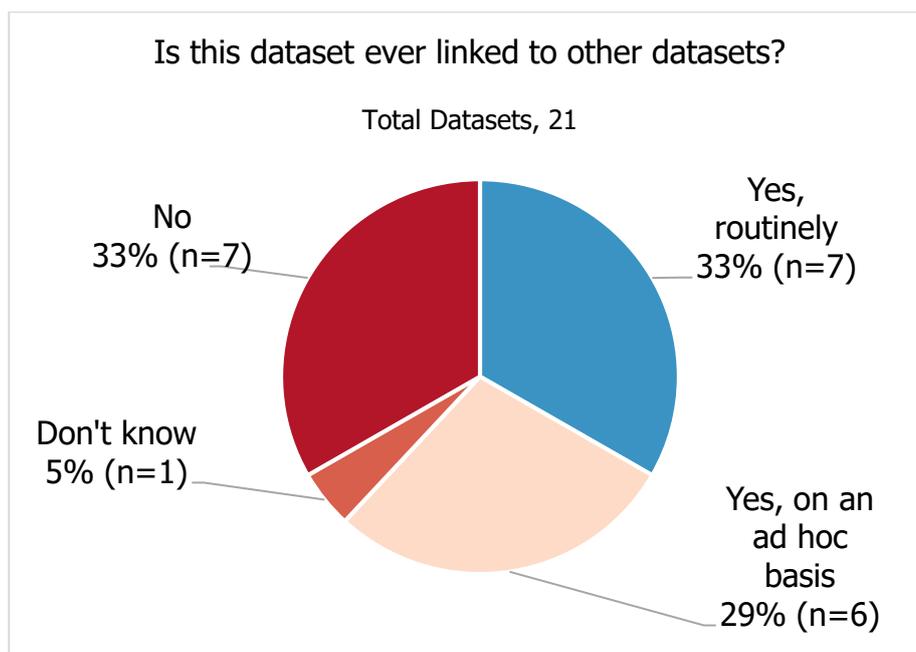
The first consideration is whether data holders have the legal basis to link data. There is variation in the bases upon which the data holders process data, some operate under consent, legal obligation and public task, each with their own parameters for what data can be collected, from who, what can be shared and how it can be used. Many participants reported not having the legal bases to link their datasets for the use cases. The establishment of a HDAB can provide a solution to this challenge, addressing existing legal limitations on data holders to link their datasets to other datasets.

The participants with successful linkage experiences highlighted challenges with the wider health data landscape, such as the timeliness, completeness and duplication of data. The data received is also reportedly difficult to use. Participants referenced receiving multiple file formats containing different types of data. One participant explained having to review data manually and subsequently attempting to link individual pieces of information. Another participant reported using internally-developed software to aid finding potential matches across personal identifier fields, while the final linkage stage is still a manual review process. This process is further complicated by the coherence and comparability of data, as well as significant variations in timeliness and availability of data. Received data is often un-coded, collected using different standards or using different definitions for the same variable. These challenges will be explored further in Section 3.5 and Section 3.6.

Data holders also expressed concern over the responsible re-use of data due to these challenges in shared understanding. In some cases, this has generated a reluctance around sharing data as data holders are concerned that the data will not be correctly interpreted, with the potential for wrong conclusions to be drawn from their data. These concerns stem from changes to data during any process or transformations performed by data users. Additional guidance based on ensuring consistent understanding is seen as essential by data holders, particularly on data provenance. This is imperative to ensure correct use and interpretation of data, once it is shared with a HDAB or on open data platforms.

"We have consent […] and we have only linked 24 records in the last 6 years. There is no centralised system to do it. It is a manual task unfortunately. Even though the information exists, there is no easy way to do it…. it's extremely resource burdensome."

"The challenge is that there are a lot of records across the system that are not digital. It will take a long time before they can be digitised and have an IHI applied to them."

"Even when we have the IHI, we manage multiple data collections, we may not be allowed to link them because of the agreed use for that data collection. We will still have to go through information governance to make sure that's ok."

"[On the topic of Colorectal Cancer] I am not aware of any digital sources of genomic data in the country at present available for secondary use."

3.4.4 Anonymisation for Data Sharing

EHDS requires that an individual's personal details are removed from datasets when sharing data, in line with GDPR. Anonymised data prevents the identification of individual data subjects during analysis. EHDS also allows for pseudonymised data, which means replacing any information that could be used to identify an individual with a value which does not allow the individual to be directly identified. A pseudonym is needed to link data. Both the IHI and PPSN are examples of pseudonymisation techniques.

Only 38% (8 participants) of participants reported using anonymisation techniques on their dataset. Pseudonymisation was more common, performed by 57% (12 participants). The most commonly-used approach combined procedural safeguards and data minimisation techniques, such as restrictions on the use of identifiable information, or removal of data records deemed more at-risk of identification.

Some participants reported having internal committees responsible for evaluating data requests to determine whether the datasets should be anonymised and shared. Participants often required ethics approval and data sharing agreements, with some requiring a consent declaration from the Irish Health Research Consent Declaration Committee (HRCDC) prior to releasing data. As different data holders process data on different legal bases, the interpretation of these legalities for linkage across multiple datasets involved in the use case, presented a challenge. Participants also noted that safeguarding the data to ensure re-identification was not always possible, especially when data linkage begins and further complicates the governance process. Participants also requested greater clarity on the distinction between anonymised and pseudonymised data, and the specific obligations that apply to each.

Many participants reported that anonymisation techniques are resource intensive as they are highly manual and time consuming. Data holders must ensure their anonymised data meets legislative requirements, governance requirements and the needs of the data user, all of which add to the workload. Staff responsible for preparing data often reported that they lack the sufficient legal understanding or the formal training in anonymisation techniques, to confidently make informed decisions. Some staff also reported lacking access to expert legal and technical guidance on performing anonymisation, making it difficult to improve their processes. This skills gap in anonymisation techniques and caution around re-identification at a later stage, sometimes led to participants adding terms to their data sharing agreements that prevent data linkage.

> "When we are providing data to an external user that there are conditions of use for the data set out. One of those conditions sets out a requirement not to try to link it with other sources in a way which could lead to a re-identification of the data set."

> "At national level, we struggle more with who can authorise the release of the data rather than the process of releasing it. This wider authorisation has become more of an issue over recent years."

This highlights a fundamental challenge for data holders who must balance data protection obligations under GDPR, with data sharing obligations under EHDS. Currently, the difficulties experienced in performing anonymisation leads to either an over-reliance on governance controls or in some cases, data holders not sharing their data at all. The establishment of the HDAB will provide solutions to this challenge when sharing data as part of a data request, but development is still required for data holders' new obligations under EHDS to anonymise and share certain datasets to open data platforms.

3.4.5 Significance of findings – Data Linkage and Anonymisation

The findings provide an outline of the current situation, identifying the challenges faced by data holders at the time of the assessment. Given that many of the datasets included in this readiness assessment are well established and the use cases examined are major healthcare topics, the findings highlight the difficulties in identifying individuals across datasets, connecting datasets and standardising anonymisation practices to protect personal information. Combined, these challenges currently prevent effective data linkage, in turn impacting on data users' ability to further our understanding of specific health topics. Significant work is needed for data holders to adequately prepare their data for sharing and linkage, which will be a contributing factor to their data submission and timeline obligations under the EHDS Regulation.

For the majority of participants, data linking and sharing were ad-hoc and difficult to undertake, or not done at all, as demonstrated in the use cases. The primary challenge is matching individuals across datasets, as identifying information collected varied widely and was often insufficient to match records successfully. While the IHI and PPSN both offer methods to improve this matter, both pose challenges and are not routinely used across data holders at present. Furthermore, both the IHI and PPSN require additional supporting information to be effective and more consistently applied throughout the healthcare landscape. Clear direction is needed to support the universal adoption of a unique identifier. This is vital for the effective functioning of a HDAB service.

These findings relate to a point-in-time, recognising that the establishment of a HDAB will address some of these challenges. As this will be the first time a linkage authority will be established in Ireland, the HDAB will bring many positive changes for data holders and data users. Findings emerging from the TEHDAS joint action outline multiple potential models of operation for a HDAB, so it will be left to individual member states to determine the model best suited to their system. Therefore, data holders are awaiting policy decisions, implementing acts and associated guidelines before they can confidently and effectively adapt their practices. A significant benefit to the implementation of the EHDS and the establishment of a HDAB will be the alignment of best practices in the area of linkage and anonymisation across Europe.

3.5 Data Quality

High-quality data is data which is fit for purpose; that is, it meets the needs of data users. Within the context of the EHDS, assurances with regard to the quality of available data are essential to facilitate the reliable secondary use of health data. Assessing the quality of health data is commonly undertaken through consideration of several dimensions including relevance; accuracy and reliability; timeliness and punctuality; coherence and comparability; accessibility and clarity. As set out in HIQA's *National Standards for Information Management in Health and Social Care*, organisations striving to ensure data quality should develop a data quality framework.[17] Essential elements of a data framework include a data quality strategy, a data quality assessment tool, data quality reports and a data quality improvement cycle.

3.5.1 Current data quality practices

While 85% of survey participants (18) reported having processes in place to improve and maintain the quality of their dataset, qualitative findings show a variance in the data quality processes applied by each data holder. With some deploying data quality checks at source, developing internal data quality rules, conducting data quality reviews and audits, or applying release guides which record changes to the data and ensure transparency and reproducibility. In a few instances, participants engaged in a formal and ongoing process of checking, validating, and cleaning of their data when issues are detected. Regardless, these practices are not applied systematically and at the times the processes occur on an ad-hoc basis, as issues are identified.

More than half of the survey participants reported having a data quality framework in place and some participants reported aligning to the *National Standards for Information Management Standards in Health and Social Care*.[17] However, in many cases this framework is internally focused and does not address the issue of interoperability with other systems, which is a key concern in preparing for the implementation of the EHDS Regulation. In addition, when examining findings from the survey on the components of a data quality framework in place, nine participants (50%) had implemented data quality improvement cycles, assessment methods and reports, while only five (23%) of these had a data quality strategy in place. This finding was supported by the qualitative responses, as some described elements of data quality framework or reactive responses to data quality issues. A strategy is essential to outline governance structures and oversight regarding data quality procedures, training and education and audit.

3.5.2 Data quality expertise

Participants responding to the survey largely reported building expertise in relation to data quality, as well as having high levels of confidence in their ability to communicate data quality issues to stakeholders. In the survey, all participants reported having some expertise, with some having specific roles dedicated to data quality with 38% (eight participants) reporting good expertise and 29% (six participants) reported having extensive expertise in this area.

Figure 5: Data quality expertise and ability to manage data quality

Current Levels of data quality expertise and ability to manage data quality

Total Datasets, 21

	No expertise	Little expertise	Some expertise	Good Expertise	Extensive expertise	Don't know
	0.00% (n=0)	0.00% (n=0)	28.57% (n=6)	38.10% (n=8)	28.57% (n=6)	4.76% (n=1)
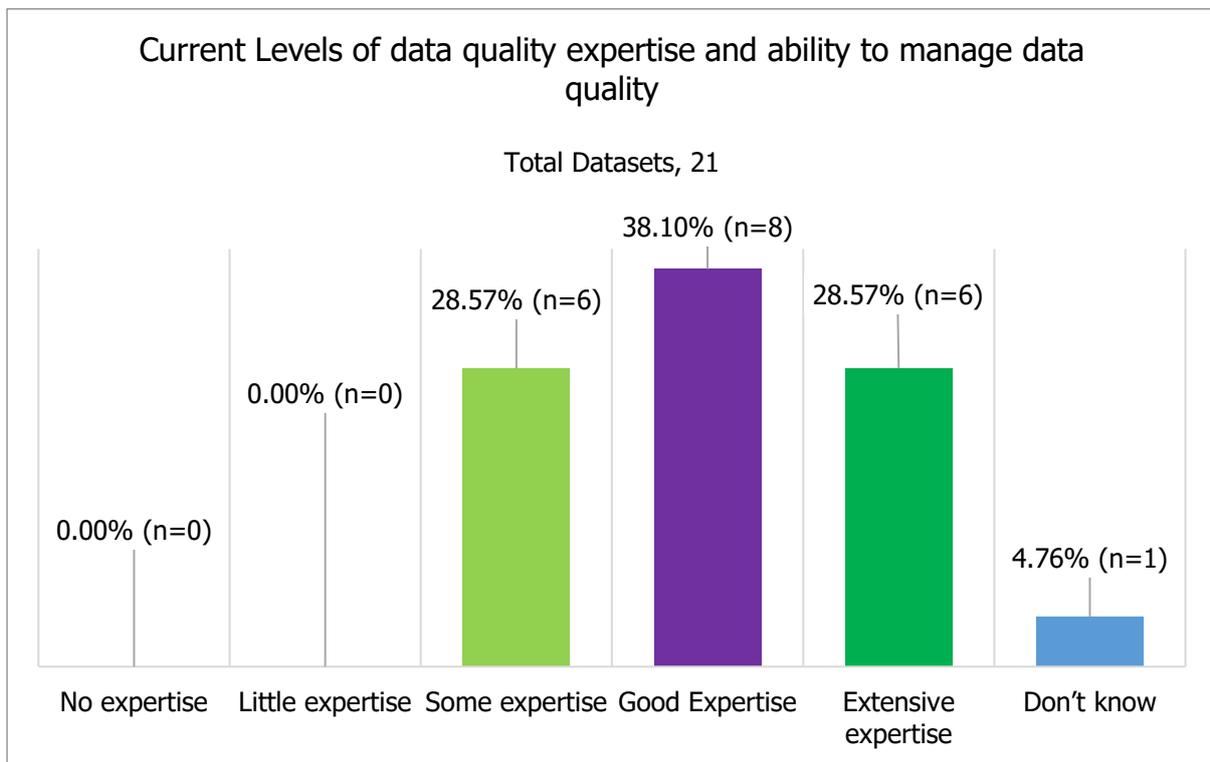
Participants reported that a range of formal and informal training sessions were provided for staff within their organisations, often covering general data quality training, as well as trainings aiming to strengthen staff capacity on internal systems, best practice in handling and describing data. Participants reported formalised training schedules with mandatory courses in place for staff to complete ongoing education and induction courses on data quality. Participants also highlighted peer support workshops facilitated within their organisations aimed at supporting staff on a range of data quality topics. This range of supports reflect the data holders' size and capacity to develop and host structured training programmes, often linked to the funding model and governance structures. Smaller organisations do not have the same capacity and expertise to carry out similar programmes of work. However,

discussions also identified that data quality is not always given the same precedence as training in areas such as GDPR or cyber security, due to current legal and reputational implications in these areas.

3.5.3 Enablers and challenges to data quality

Analysis of responses from participants identified several enablers and facilitators to supporting and maintaining data quality. Enhancement of data quality is supported through the adoption of a data quality framework, data dictionaries, analysis of key performance indicators, as well as the use of automated tools and software which promote data quality efforts. In some cases, participants described how they used the findings from data quality checks and improvements, when communicating with data users, promoting transparency and effective use of data.

Participants reported variance in the digital and data literacy of staff within the workforce, which was evident in inconsistent approaches to data entry, data documentation and coding, all of which are challenging to the progression of data quality. Additionally, participants discussed a high dependency on staff members with knowledge and expertise in this area, with little contingency if staff members leave. Therefore, some participants reported high staff turnover as a significant issue through loss of knowledge and disruption to an organisation's ability to demonstrate continuity and evolution of practices in this area.

Another key challenge to improving and strengthening data quality practices centred on the lack of a unique identifier and specifically the delays in adopting a national unique identifier across datasets and within hospitals' administration systems. Low adoption or use of unique identifiers, as detailed in Section 3.4, impedes the ability to create high-quality data and the effective use of data. Without this, datasets cannot be easily linked, and datasets suffer from completeness and duplication issues which lead to ongoing and avoidable cleaning challenges.

"That strategic identity management piece should come forward as part of the EHR and Shared Care record. It will improve things for us all."

"We have a data dictionary where you can see what type of quality of data is in each variable. If its 'green', it indicates that it's fairly good quality data. If its 'amber' there may be some issues in the dataset."

3.5.4 Data quality dimensions

Participants raised challenges in ensuring that data is **accurate** and **reliable**. This impacts the ways in which data can be used, and the confidence data users have in the data which directly informs decisions around provisions, planning, policy development and research. Issues were highlighted in relation to the consistency and verification of data upon the point of entry, as well as the duplication of data due to hospitals gathering data based on episodes of care as opposed to an individual's entire patient journey. Challenges prevail in the collection of data in clinical services; gaps exist in processes and systems to check for missing and accurate data as the data moves through its lifecycle. Data gathered in clinical settings is susceptible to data quality issues, largely because clinical data is gathered for the primary purpose of providing direct care, often in busy settings, lacking the same quality checks that curated datasets for secondary uses are subjected to. In addition, it was noted that the secondary use of clinical data will become more significant with the evolution of the national electronic health record. Further clarity is required regarding best practice for the enhancement of data quality pertaining to data collected in clinical settings.

Dimensions of **coherence** and **comparability** were discussed as being particularly important to the establishment of a HDAB service, requiring exchange of data both nationally and within Member States. However, national guidance and coordination are essential to ensuring consistency of the data collected across organisations and over time. In particular, participants of the focus groups identified enablers including a national approach to implementing a data dictionary, further decisions and guidance on coding standards and enhancing mapping capabilities as key to improving these aspects of data quality.

Participants also expressed concern relating to the **timeliness** of data, reporting practical issues with manually entering data into systems, and problematically having to enter the same data onto different disparate systems. This leads to significant delays in accessing 'real-time' data, often with significant backlogs of data entry into national systems. This is a broader national health information issue but nevertheless impacts on the effective use of data and the optimal functioning of a HDAB service.

"*There is no point having data if it is not accurate and relevant to the patient or the service provided. You need to be able to get data out of the system. Accuracy and reliability is fundamental. If you are getting data out and it means nothing, what is the point of it?*'"

"*They have their own standards of coding which may not align with some of our coding…we have some legacy coding, local coding from the past. We have to work across all the systems to align the codes.*"

3.5.5 Roles and responsibility for data quality

Findings from the readiness assessment identified that data quality tasks are often performed by individuals without recognition of a formal data quality role or responsibility. In these scenarios, the task is informally added to other core duties and consequently is not always prioritised. This was particularly evident in organisations without a data quality strategy or framework. In these cases, practices are not formalised, and dedicated resourcing is often not assigned to data quality.

Findings from participant feedback also highlight the need to emphasise the importance of data quality, at all stages of the data lifecycle. To facilitate this, participants highlighted that training should be strategically managed to address general training across all health and social care professionals through emphasising the importance of data quality, as well as specialised training for those with specific data quality roles.

In addition to the formalisation of data quality roles, it was suggested that a national approach to data quality was required to standardise practice and to create clearer structures to support data quality improvement. This could include the adoption of national frameworks and improved integration of data quality processes into routine operations. While the need for training was reiterated, it was emphasised that broader organisational recognition, as well as policy alignment on the value of data quality work, would be essential to sustained improvement. Participants recognised that they are currently in a changing landscape and that they will have to align current data quality practices with the changing requirements to comply with the EHDS Regulation. Training and guidance on data quality enhancement are aspects of the HealthData@IE project, which will address some of the concerns identified by data holders.

"It is a clear barrier if people have different levels of understanding of the impact of poor-quality data on the overall chain…We don't have a uniform way of understanding the same concepts of the same values throughout the different stages."

"There is a need to revise the current HIQA data quality framework to align with the [EHDS] data quality [dimensions], will be useful in order to highlight what to focus on for good data quality."

"Data quality is a skillset in itself; however, it's often added on as an extra item to a person's role and because of this there could be a lack of understanding of what is required. There is a need for more investment into skills on data quality. There is a need to upskill workforce on data quality. There is a gap in skills currently and people have busy day jobs already. There is a need for investment in people at the coalface of care to improve data quality from initial stage of collection."

3.5.6 Significance of findings – Data quality

The success of a HDAB depends on the quality of datasets, therefore the EHDS regulation outlines an obligation for data holders to report on the quality and utility of datasets on an annual basis. Across the survey and focus groups, data quality was a common concern amongst participants of the readiness assessment. Participants noted the implications for legal compliance, public trust, and the reliability of any secondary use of data. Data holders in general are concerned about the quality of data and the required preparations in terms of assessing and reporting data quality through the common data quality and utility label. Staff capacity and capability in terms of roles and responsibility for data quality, as well as training and education programmes for staff varied significantly among data holders. Larger organisations with specific data quality roles appeared to be more prepared to provide support and resources to relevant staff on enhancing data quality practices.

For data holders, the application of the data quality and utility label, as well as moving to new data quality dimensions, will require significant capability and capacity to ensure organisational alignment with EHDS. To complete this work, data holders will need to improve their data quality skills to meet the requirements of the EHDS Regulation. The findings also highlighted that many data quality practices focus on internal processes, rather than the coherence with other data and integration with other systems. In addition, findings indicate that most data holders lacked a data quality strategy to outline governance structures and oversight regarding data quality procedures, training and education, as well as ongoing monitoring and improvements of data quality. This oversight plays a vital role in implementing a data quality framework as well as providing senior management with information to help monitor and prioritise data quality. All data holders, regardless of their size, will need to establish strategies with specific goals in order to respond to the legal requirements, as set out by the EHDS Regulation. Smaller data holders without specific data quality roles or expertise in this area may need additional support, to meet the obligations of assessing and reporting on the label. In relation to specific dimensions, a number of key findings were identified which need further consideration in terms of system and data holders' preparedness. Firstly, ensuring the accuracy and reliability of data at all stages of the data lifecycle, especially at the point of data collection. This is particularly relevant to data gathered for operational purposes, such as within EHRs, as this data often lacks the quality and rigour necessary for secondary use. In this instance, a clear link needs to be established for data collected for primary and secondary use, and guidance is needed on processes to enhance the quality of data collected for primary use in clinical settings, but which will also be used for secondary purposes.

Moreover, the importance of data quality needs to be part of the education and training for those working at each key stage of the data life cycle, including health and social care professionals. Regarding coherence and comparability in preparation for the EHDS Regulation, it is important that data holders become nationally aligned to a data dictionary and agreed coding standards, and that mapping capabilities are enhanced to ensure comparability of datasets. Significant progress is required to prepare the system and data holders in this regard. Additionally, regarding timeliness, manual data entry and issues with system integration have caused delays to this real-time access. These issues will impact on data users' ability to access health data, and the effective use of data in Ireland.

Furthermore, participants also highlighted the need to report data provenance in a standardised way across data holders, to ensure the correct interpretation of data and the appropriate use of data for secondary purposes. In particular, making the history of the data origins and any transformations the data underwent available as part of the standard report. Consistent documentation of data provenance is critical for building trust and transparency in health data re-use. In addition, further efforts to engage with participants around the 'data discovery' process will need to be undertaken to improve the accuracy of data requests and utilising a 'data label' to convey limitation of a dataset's quality.

The EHDS Regulation's capacity-building framework (Article 82) will support the development of guidelines and training modules to promote best practices in data quality.[5] The HealthData@IE project will also deliver outputs on data quality enhancement to improve the quality of health data, including the development of data quality guidance and associated training modules, an information management maturity assessment tool and a compliance assessment framework. The project will also develop a national interoperability framework to support data holders in aligning with best practice and in preparation for submitting data to a HDAB. The HealthData@IE project will include the development of a peer support network for data holders on the topic of data quality. Together, the recommendations from the QUANTUM project[*] and the outputs from the HealthData@IE project will help to address some of the identified gaps and supports required in this area.

---

[*]QUANTUM aims to develop recommendations on a common label system that can be used across Europe to report on the quality and utility of datasets so that data users can identify high-quality data for different purposes. The EHDS Regulation will enforce the application of the label for datasets made available through a HDAB.

3.6 Standards for data discovery, semantic interoperability and data exchange

When HDAB services are established in Ireland, data holders will be required to use the Health DCAT-AP specification to submit metadata describing their health-related datasets for inclusion in a national health dataset catalogue. Once the HDAB service has been established, data holders will also be required to provide data to the service in response to granting of a permit following a data access request. This section aims to assess the data holders' readiness around existing practices around data discoverability, the ability to publish to a national metadata catalogue and to provide coded data to HDAB through the use of data exchange standards.

Overall, the insights gained during this assessment indicate a range of structural, legal, and operational challenges that must be addressed to support health data sharing in Ireland. Technical solutions and an enhanced legislative framework will be essential. Although organisations already have some expertise in describing metadata and exchanging data with other systems, many practices still need to be aligned with the EHDS regulation. Participants repeatedly emphasised that successful implementation would hinge on organisational readiness and the support of the workforce. Participants expressed a willingness to engage with the HDAB but noted that sustained progress would require consistent policy direction, adequate resources, and cross-sectoral collaboration.

3.6.1 Data Discovery

Participants were surveyed on their current use of descriptive metadata standards and their plan for the use of Health DCAT-AP, the descriptive metadata standard preferred by TEHDAS and EHDS. Surveys results showed that 43% (nine participants) reported that a specific descriptive metadata standard is used, with 38% (eight participants) currently having a plan for adoption of Health DCAT-AP, see **Figure 6**.

Figure 6: Plans to use Health DCAT-AP

Do you have plans to use Health DCAT-AP to describe the dataset?

Total datasets, 21

Don't know, 14.3% (n=3)

Yes, 38.1% (n=8)

No, 47.6% (n=10)
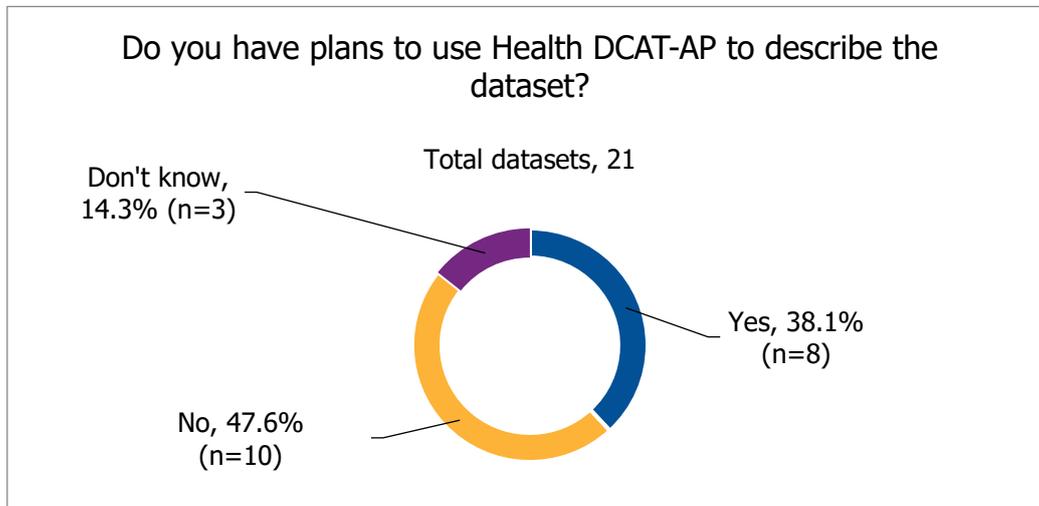
Of the 43% (nine participants) who currently use a specific descriptive metadata standard to describe a dataset, the participants pointed to the use of a custom data dictionary, custom templates, custom data validation specifications, custom data reporting specifications, the Data Documentation Initiative (DDI) standard and the American Health Information Management Association (AHIMA) guidelines for Data Dictionaries.

A follow-up question requesting participants to provide further details was asked of the eight participants currently planning to use Health DCAT-AP. In total, four data holders submitted nine responses to the follow-up question. As some details were identical across datasets within organisations, only four unique responses remained. Organisations reported being at different stages of planning:

▪ one participant stated that they are exploring the area to see the requirements
▪ one is reviewing the requirements to see how best to apply DCAT-AP to their dataset
▪ one is developing a new procedure which will include guidance for the use of DCAT-AP
▪ one participant noted that resources and expert supports are required to facilitate this work.

Participants of the focus groups highlighted that they were familiar with frameworks such as Health DCAT-AP, a data dictionary and Data Quality Framework but these frameworks were not commonly implemented. This was attributed to both limited awareness and a lack of capacity within organisations to operationalise these

standards, tools and frameworks. Metadata was frequently reported as incomplete or absent, which undermines efforts to discover and re-use datasets effectively.

"Health DCAT is something we will all have to get on top of very quickly. We will need knowledge on how to manage this. The characterisation of data. We need help on how to do this properly."

"The way it's coded or recorded can differ so much that it's hard to align across organisations."

The selection of national metadata standards was identified as a critical next step. However, participants reported that 81% (17 participants) presently share aggregated data externally and 52% (11 participants) share patient level data while 71% (15 participants) reported sharing descriptive metadata with other systems.

Figure 7: Methods facilitation discovery of datasets by data users

Are details about your dataset discoverable by data users by any of the following?

Total datasets, 21

Method	Percentage
Other	52.4% (n=11)
Browsing Catalogue	33.3% (n=7)
Viewing Metadata	42.9% (n=9)
Search engine	47.6% (n=10)
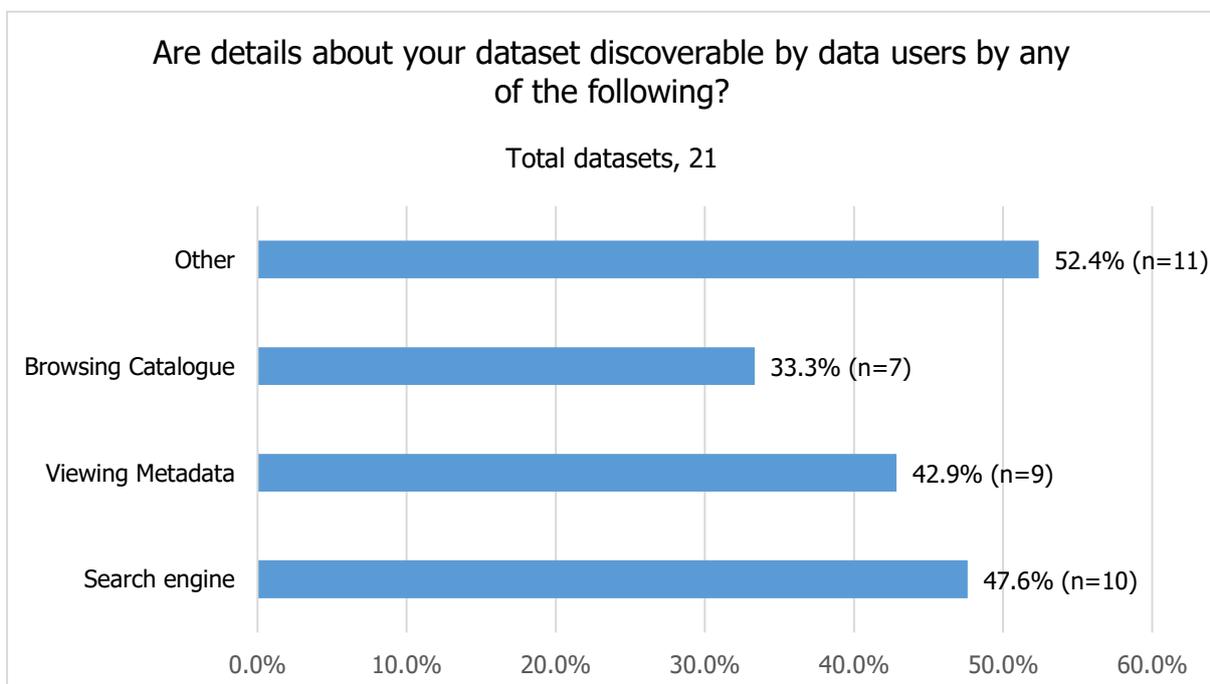
Nearly half (47.6%, 10 participants) reported that the dataset is discoverable by using a search engine, 43% (nine participants) reported that it is discoverable by viewing the metadata, while one-third (33%, seven participants) reported that it is discoverable by browsing the catalogue.

More than half of participants (52%, 11 participants) who selected 'other' indicated that their organisations have many different ways in which their datasets are discoverable. The most frequently reported method was by publishing details of the datasets on their organisation's website (29%, six participants) with the listing of the dataset on HIQA's catalogue of national data collections[††] the second most common (24%, 5 participants). Following on from that, there was a variety of methods, including the Central Statistics Office (CSO) Statbank, published academic papers, reports and aggregated data shared with the EU and United Nations Office on Drugs and Crime (UNODC) and annual reports.

Those who reported that their organisation shares descriptive metadata with other systems were asked to provide information on the methods that are utilised to share the descriptive metadata. The results are presented in **Figure 8**.

Figure 8: Methods used to share descriptive metadata with other systems

Methods used to share descriptive metadata with other systems

Total datasets, 21

- HDAB service harvests data via an API (pull): 0.0% (n=0)
- Uploading data via an application programming interface (API) (push): 14.3% (n=3)
- Manual submission through form-based user interface (UI): 9.5% (n=2)
- Manual submission through file uploads (for example, CSV): 19.0% (n=4)
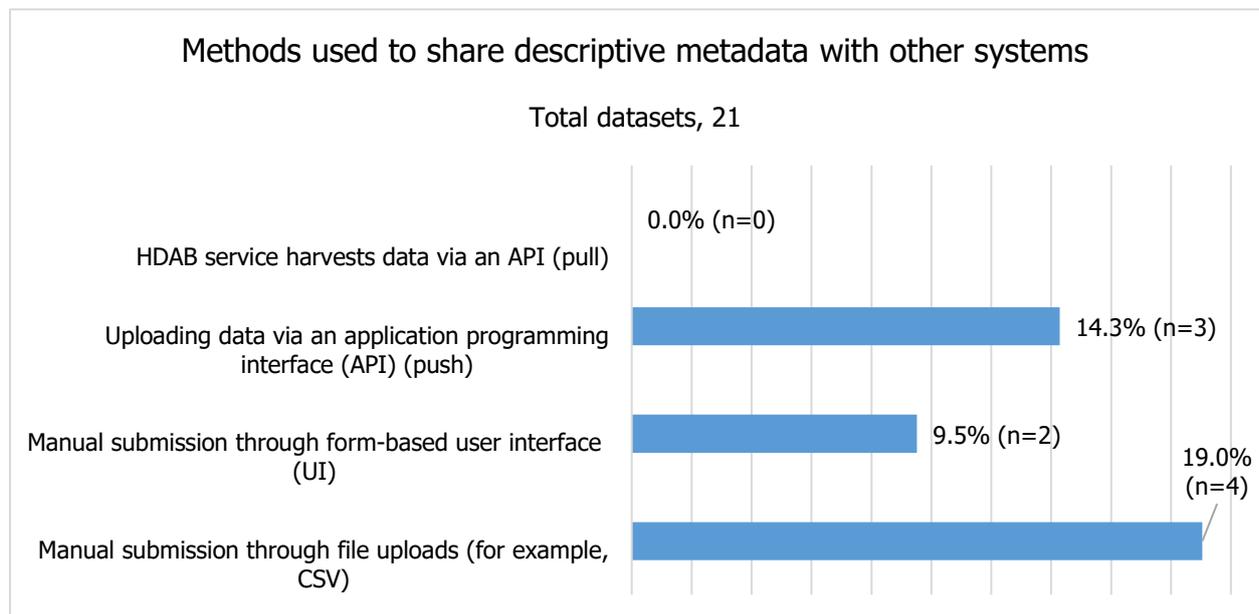
[††] National data collections are repositories of routinely collected health and social care data, including administrative sources, censuses and national patient registries in Ireland.

Methods of sharing metadata include manual submission either through file uploads (19%, four participants), through form-based user interfaces (10%, two participants) and through use of APIs (14%, three participants).

Participants were given the opportunity to provide further detail relating to the metadata they currently share with other systems, with five responses to this question. The methods used by data holders when sharing their metadata included HL7, extraction to Health Performance Visualisation Platform (HPVP), National Treatment Purchase Fund (NTPF), Department of Health and Advanced Nursing Practitioners (ANP), CSO Statbank, CSV files sent to a HSE-shared site, and metadata shared with EU and UNODC through a user interface.

From the focus groups, participants identified that preparations are underway to support better management and documentation of their metadata. Participants noted that organisations are awaiting national initiatives to support data discovery and the capacity for data discovery will be enhanced by the development of a National Dataset Catalogue in which the Health DCAT AP will be used.

> "*We have been looking at the management of our metadata that's describing our data and our data handling processes, and we are looking at improving that documentation.*"

> "*I think one of the biggest barriers is that there is no repository for us in Ireland to deposit that metadata.*"

Participants highlighted the benefit and the need for a national data dictionary in promoting consistent practice and to bring consistency across datasets. However, one participant noted that this should extend to data collected in all settings – public, private and voluntary. For example, one participant noted:

> "*I know there is work ongoing with a HSE Data Dictionary and I hope that will become a National Health Data Dictionary. A lot of our care is provided by non-statutory groups, such as voluntary and private hospitals.*"

One participant noted that the HSE has established a Dataset Specification Management Process (DSMP) to ensure a consistent approach to dataset standardisation, and they are also developing a data dictionary based on the International Organisation for Standardization (ISO)/ International Electrotechnical Commission (IEC) 11179 Metadata Registry (MDR) standard. Participants also reported using Logical Observation Identifiers Names and Codes (LOINC) and Systematized NOmenclature of MEDicine Clinical Terms (SNOMED).

3.6.2 Semantic Interoperability and Data Exchange

Figure 9: Sharing aggregated data with other systems

Does your organisation currently share aggregated data with other systems?

Total datasets, 21

No, 19.0% (n=4)

Yes, 81.0% (n=17)
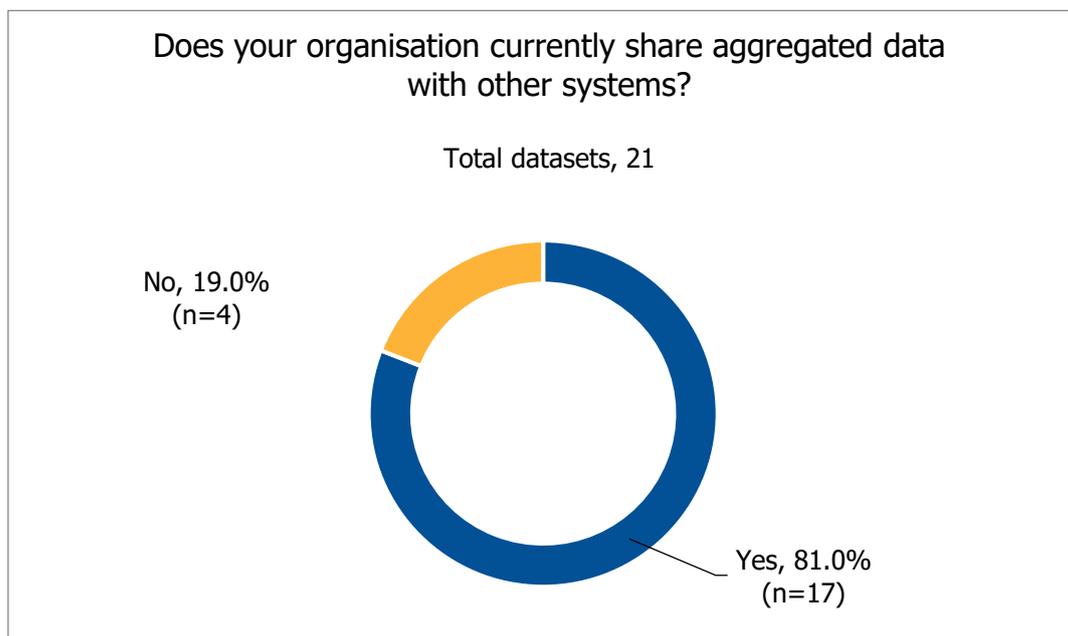
More than 80% (17 participants) reported that their organisation currently shares aggregated data with other systems.

The survey elicited 16 responses when participants were asked to provide information on the data they currently share with other systems, and the systems and methods they utilise to share data. Of these, 10 responses indicated that aggregated data is shared with a third-party system or organisation via user interfaces and secure online platforms, while six participants indicated that aggregated data is shared via reports or other formats through their organisation's website. Examples of the methods or systems used for data sharing include:

- European Centre for Disease Prevention and Control (ECDC)
- Open Data (data.gov.ie)
- Irish Social Science Data Archive (ISSDA)

Figure 10: Sharing individual-level data with other systems

Does your organisation currently share individual-level data with other systems?

Total datasets, 21

No, 47.6%
(n=10)

Yes, 52.4%
(n=11)
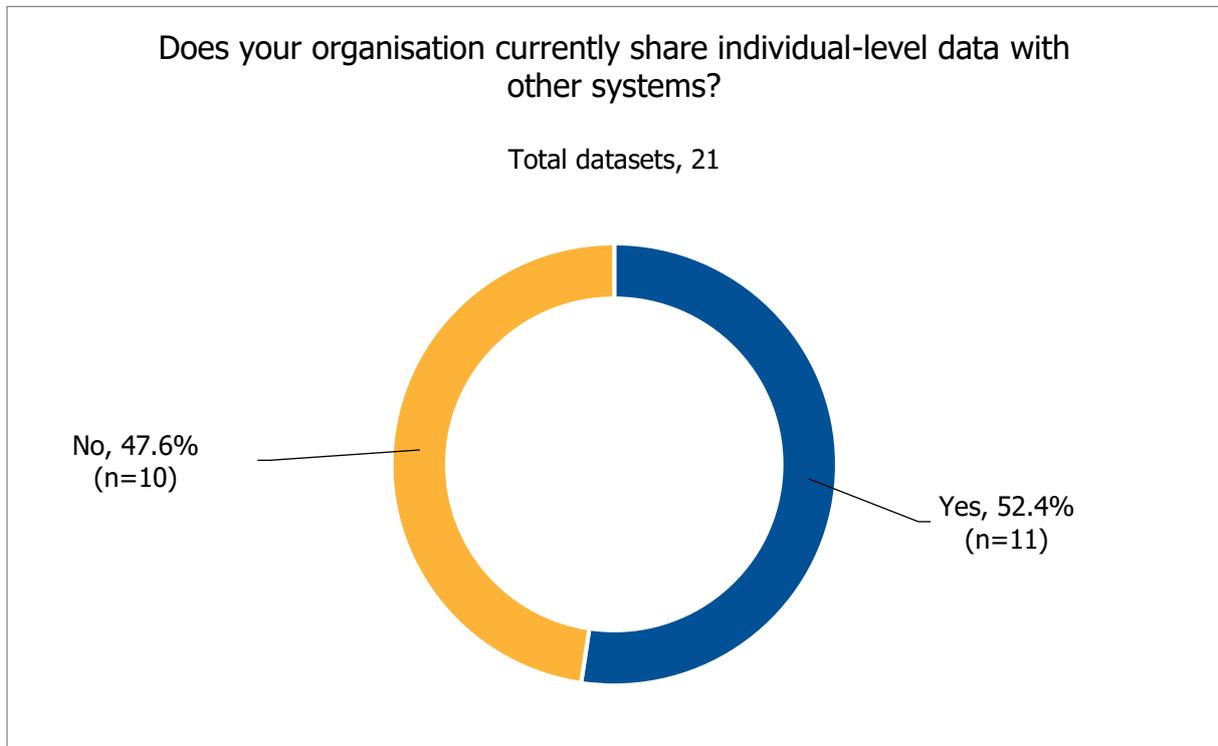
More than half (52%, 11 participants) reported that they share individual-level data with other systems. Those who responded yes to the above were asked to provide information on the methods used to utilise the sharing of individual-level data, the results are presented in the graph below.

Figure 11: Methods used to share individual-level data with other systems

Methods used to share individual level data with other systems

Total datasets, 21

Method	Percentage
HDAB service harvests data via an API (pull)	0.0% (n=0)
Uploading data via an application programming interface (API) (push)	23.8% (n=5)
Manual submission through form-based user interface (UI)	14.3% (n=3)
Manual submission through file uploads (for example, CSV)	38.1% (n=8)
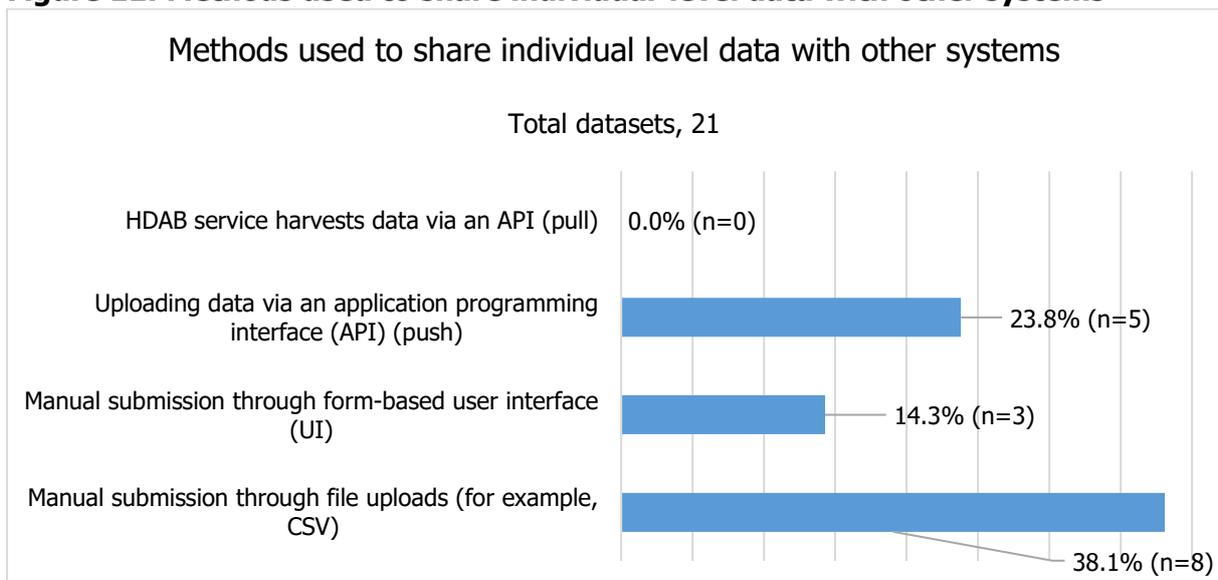
From the 11 responses gathered, 16 methods were identified. Of those that responded, eight data holders were found to share individual-level health data by means of a manual submission through file uploads; three data holders share individual-level health data by manually submitting it through a form-based user interface; while five organisations were found to share individual-level health data by uploading data using an application programming interface (API).

Participants who do share individual-level data with other systems were asked to provide qualitative responses on the data they share with other systems. Participants from two data holders provided further information. One response indicated the use of HL7 to share individual-level data with other systems. The other response referenced a sequence of different data which is shared, such as pharmacy dispensing, some vaccination data, certain oncological data, outpatient parenteral treatment data, medical card GP consultations, special services data and multiple sclerosis and anti-viral therapies data.

Participants were asked to indicate their data holder's level of preparedness for providing data to a HDAB, the results of which are presented in the **Figure 12** below.

Figure 12: Level of preparedness for providing data to a HDAB service
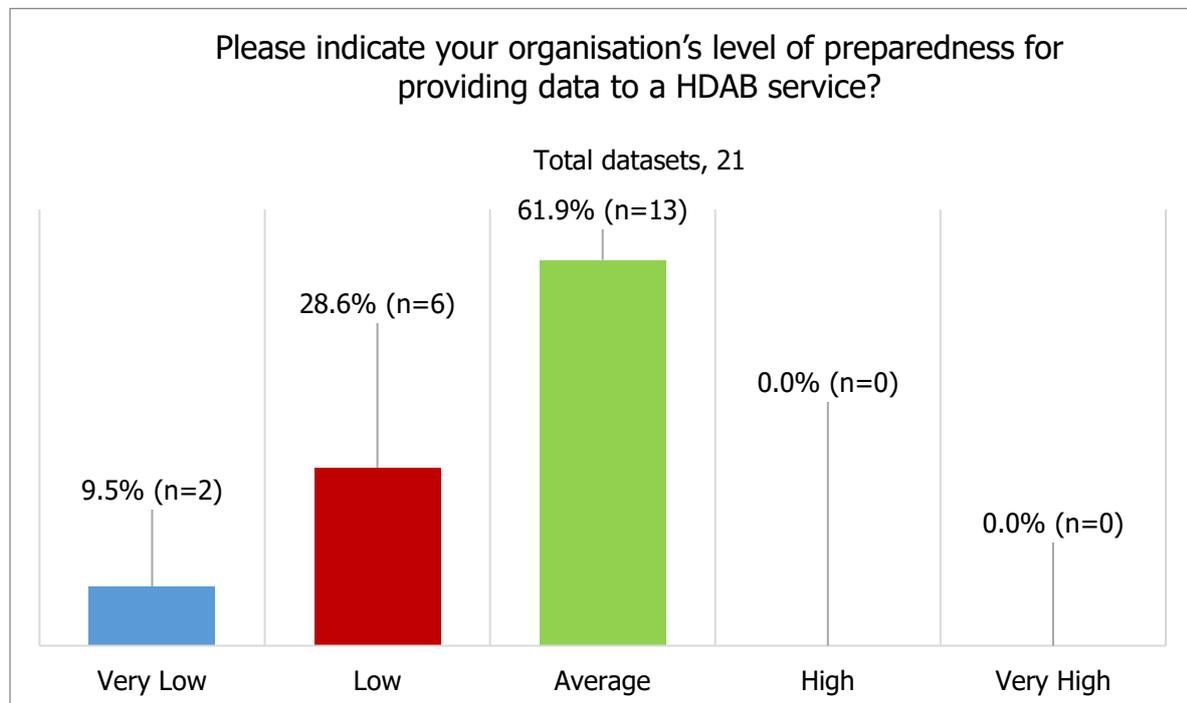
Survey responses showed that 62% (13 participants) reported that their organisation has an average level of preparedness for providing data to a HDAB service, while 38% (eight participants) reported having a low to very low level of preparedness.

Challenges identified for data holders in releasing data to the HDAB included the additional resources required to prepare data for release to the HDAB, specifically in situations where there is a misalignment of the coding systems used, which would require a mapping of data to a different standard, with the complexity of that mapping potentially varying between different datasets. Other challenges include the lack of completeness in data fields, the need to effectively communicate dataset completeness with data users, and to provide support to data users to ensure appropriate data interpretation. This aligned with reported data that less than one-quarter of participants (24%, 5 participants) have a plan for providing data to the HDAB service using the standards for interoperability identified at the European level, see **Figure 13**.

Participants were generally in agreement that the responsibility for data preparation ahead of release to a HDAB and the process of mapping their data to other coding systems, should sit with the data holder or that data holders should be central to any decisions made about the release of data, while also acknowledging the additional resources this would require.

When asked about their experiences of coding data prior to sharing it externally, a number of participants reflected on the requirements to engage with the data users as early as possible to ensure that the correct dataset can be curated to meet their needs. This process can prove to be resource intensive as the user may not be aware of the full extent of the dataset or of any issues present within the dataset.

Figure 13: Plans to provide interoperable data to the HDAB service

Does your organisation have a plan in place for providing data to the HDAB service using the standards for interoperability identified at the European level?

Total datasets, 21

Don't know, 28.6% (n=6)

Yes, 23.8% (n=5)

No, 47.6% (n=10)
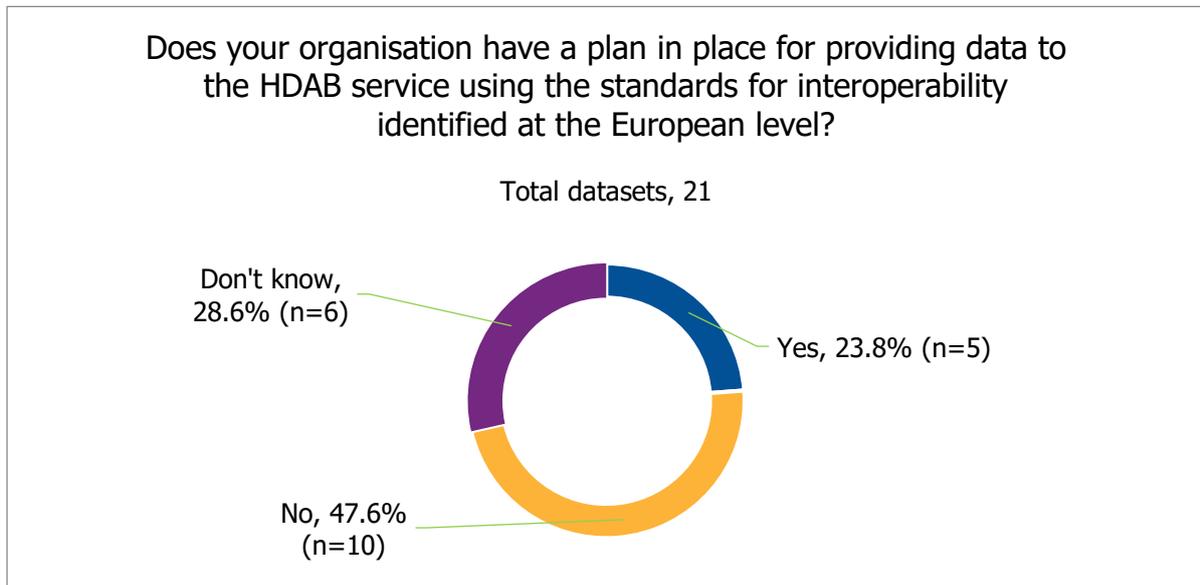
Participants were asked to indicate the level of maturity of the dataset regarding the implementation of standards for semantic interoperability and interoperable communication, see **Figure 14**.

Figure 14: Level of maturity - standards for semantic interoperability and interoperable communication

Level of maturity of dataset in the implementation of standards for semantic interoperability and interoperable communication

Total datasets, 21

38.1% (n=8)

28.6% (n=6)

19.0% (n=4)

14.3% (n=3)

0.0% (n=0)

Initial – little coordination and practices remain ad hoc

Repeatable – basic practices are in place, but variation exists between people and teams

Defined – standardised practices are documented and in place but there is still room for improvement

Managed – practices are established and controlled, and regularly monitored

Optimised – practices are continuously monitored to enable optimisation and improvement

The largest share of datasets (38.1%, eight participants) are at the 'managed' stage, indicating solid, repeatable processes but not full optimisation. The smallest share of datasets (14%, three participants) are at a 'defined' level of maturity, while over two-thirds of datasets (47.6%, 10 participants) are at the 'initial' or 'repeatable' stages, highlighting significant room to formalise standards and practices.

Through focus groups, participants were asked about their experience coding data within their own datasets. While variances in the coding systems emerged, general difficulties were discussed around clinical data and coders being reliant on the quality of the information gathered at source. Participants discussed the complexities of coding from paper medical records, as coders require training to input, interpret and translate the data into the coding system they are using. In these cases, participants discussed how audits are required to monitor the consistency across coders, sites and patient groups. Participants also highlighted the automation of coding and provision of training amongst coding staff as being potential facilitators in the coding process. This will be particularly relevant going forward in the context of data being more readily available when requested.

Participants reported usage of standardised coding systems, such as SNOMED in order to maintain consistent value sets. When asked about the supports in place, participants across the three focus groups discussed their active involvement with DSMP managed through the SNOMED CT National Release Centre.

Participants were asked a multi-choice question with five response options regarding semantic interoperability standards used in the dataset. The three standards identified as candidate standards by TEHDAS were included. International Classification of Diseases ICD-10-AM was included, as it is used within health information systems in Ireland. Additionally, participants were provided a free-text option as 'other' to provide further information.

The two most common semantic interoperability standards in place are ICD-10-AM (44%, eight participants) and SNOMED CT (33%, six participants).

Figure 15: Semantic interoperability standards implemented in the dataset
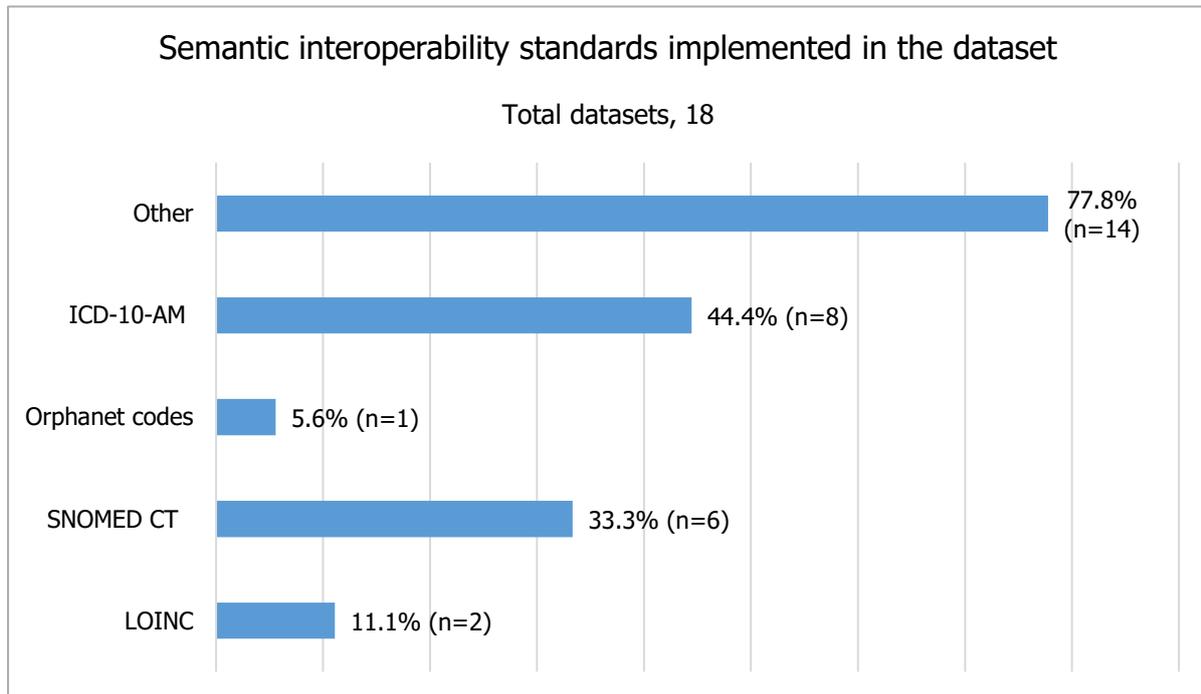
All 14 of the participants who selected 'other' from the initial question provided further information on coding and terminology standards used in their systems. A number of different variations of the ICD standard were identified by participants along with other coding and terminology standards. The standards identified included:

- ICD-9-CM, ICD-10 (4) and ICD 11 (x2),
- Australian Classification of Health Interventions (ACHI) (2)
- Health Level 7 (HL7) (x2) and Digital Imaging and Communications in Medicine (DICOM,)
- ICD-O-Vx,
- Local, Anatomical Therapeutic Chemical (ATC), Intensive Care Unit (ICU) critical care and European Centre for Disease prevention and Control (ECDC) metadata standards.

Participants were requested to indicate data exchange standards used in their datasets, with four options provided, see **Figure 16**. The most common data exchange standard in place is HL7 version 2.x (53%, nine participants).

Figure 16: Standards for data exchange implemented in the dataset

Standards for data exchange implemented in the dataset

Total datasets, 17

Category	Value
Other	58.8% (n=10)
DICOM	11.8% (n=2)
HL7 FHIR	5.9% (n=1)
HL7 version 2.x	52.9% (n=9)
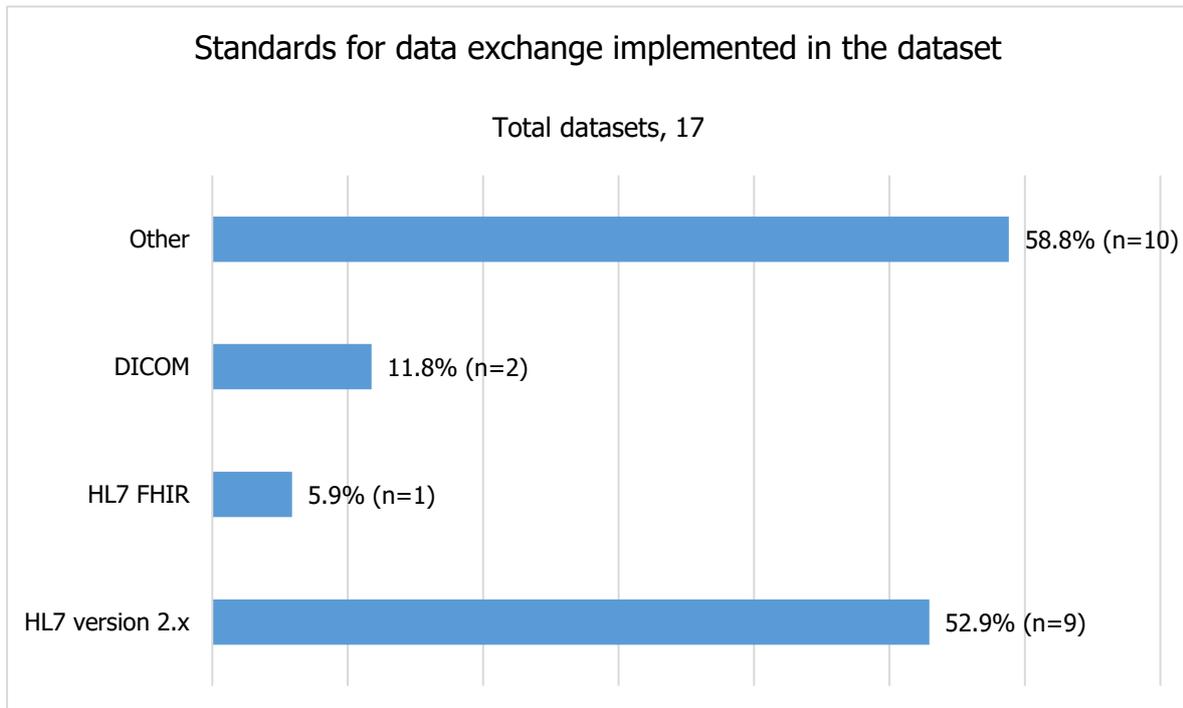
When excluding 'other' as an option, the most common standard for data exchange implemented in datasets is HL7 version 2.x (53%, nine participants).

Of the 10 participants who selected 'other' from the initial question, nine provided further details. Three organisations hope to move towards implementing HL7 in the future. Other organisations who provided details gave a range of answers, which include:

- Extensible Markup Language (XML)
- Statistical Data and Metadata eXchange (SDMX)
- American Society for Testing and Materials (ASTM)
- ICD-10
- In-house policies.

3.6.3 Significance of Findings - Standards for data discovery, semantic interoperability and data exchange

Current data sharing practices of the data holders who share aggregated level data is commonly conducted through third party systems or, less regularly, through reports or other formats on their organisations' websites. Individual level data is generally shared using manual submission through file uploads, form-based user interfaces and the use of APIs. Participants have a low to average level of preparedness for providing data to a HDAB. Additionally, the results of this readiness assessment show that a large majority of participants do not have a strategic plan in place to provide data to a HDAB using standards for interoperability. Data holders will require national guidance prior to the creation of such a strategic plan. This finding highlights an immature landscape for data sharing, where methods are varied. Support will be required to improve and standardise the approach to data sharing, along with forming a strategic plan to provide both metadata and national data to a HDAB.

Through the various stages of engagement, data holders demonstrated good practices in the implementation of sematic interoperability standards, as well as the use of available resources to improve skills and capacity. However, inconsistent metadata practices were identified. Data holders are awaiting further guidance to support their preparations for the implementation of a HDAB services in Ireland. Furthermore, data holders are currently waiting on the establishment of national infrastructure to guide this preparation.

The implementation of Health DCAT-AP and national metadata registry will play a central role in enabling data discovery by providing a common metadata framework for describing, cataloguing and discovering datasets across data holders. Implementation of Health DCAT-AP will ensure that data users can easily identify relevant datasets and understand their structure, content and access conditions. Through the publication of these catalogues in a central resource, data users will have a single place of reference when developing their requests for data access. For data holders, Health DCAT-AP provides a consistent approach for documenting their datasets.

Regarding the capability and capacity among data holders to provide metadata and data in the necessary formats to a future HDAB service, there were positive aspects, as well as areas of further opportunity. Almost half of data holders use a specific metadata standard. However, almost half do not currently plan to use the Health DCAT-AP standard and more than two-thirds do not currently share descriptive metadata with other systems.

Although data holders employ multiple semantic interoperability standards across their systems, they may be required to submit data to the HDAB using a limited selection of those standards, necessitating targeted mapping or upskilling. Findings indicate that SNOMED CT, LOINC, Orphanet codes and numerous variations of both ICD 10 and ICD 11 codes are among the semantic interoperability standards currently in use. In addition, the assessment findings indicate that HL7 version 2.x is currently the main data exchange standard used among data holders. Participants discussed their approaches to coding, however the recurring concern was that the level of data quality at source was insufficient, with subsequent implications for coding. Areas for support may include increasing digital skills capabilities across the healthcare system, particularly by building the capacity of staff involved in handling and preparing data through ensuring that outputs are consistent, structured and that the value of the work is understood.

Supports for data holders in the coding process currently exist, including the DSMP managed by the SNOMED National Release Centre, which was referenced by many participants. On the maintenance of value sets, organisations demonstrated good practice when describing their attempts to standardise data as much as possible using internal data dictionaries, libraries, databases, catalogues, use of industry standards and by putting governance structures in place to control the approval of the use of new fields and or data types. On the last two topics covered, the participants highlighted key concerns as the need for standardisation between the HDAB and the data holders across coding, data catalogues and processes. On mapping of data, data holders felt they were best placed to map the data but also were concerned around the resourcing requirements for such a role.

The Healthdata@EU use case findings identified that harmonising validation procedures and data models across EU nations is essential for consistent, interoperable datasets and affects the quality of the study. Though many challenges exist in the standardisation of data and provision of data to a HDAB, the development and implementation of a national interoperability framework for secondary use of health data through the HealthData@IE project, will deliver many benefits for data users, data holders and health data access services.

For data users, a national interoperability framework will provide more consistent, high quality and semantic-aligned datasets, consequently enabling linkage and analysis across multiple data sources.

For data holders, the national interoperability framework will offer both standards and guidance on structuring data and the processes for provision of data to the health data access services. A HDAB service will benefit from improved semantic and technical interoperability which will ultimately facilitate smoother and more efficient data flows.

3.7 Training and guidance

As part of the HealthData@IE project, a range of guidance and training materials will be developed for data holders. The area of data quality was specifically addressed in the survey to identify needs and requirements. However, development of training and guidance was also a common thread across all themes, as a result this section will present additional trainings and guidance, as identified by participants, which will be required to effectively prepare for the establishment of the HDAB.

3.7.1 Data Quality

Generally, the readiness assessment identified that there is a significant need to develop the capabilities and skills of staff and data holders to prepare the system and data holders for the establishment of a HDAB. There is a need to expand the training and educational supports available to data holders and other key stakeholders who are involved in key stages of the data life cycle. There was widespread acknowledgement of the need for guidance and training across all topic areas included in the survey (see **Figure 17**). For example, 95% (19 participants) reported a need for guidance and training on data quality standards and data quality improvement strategies, and 75% (15 participants) reported a need for guidance and training on data maturity assessment.

Figure 17: Data quality guidance and training needs

Data quality guidance and training needs

Total datasets, 20

Topic	Value
Standards for data exchange	85% (n=17)
Standards for semantic interoperability	80% (n=16)
Data quality improvement strategies	95% (n=19)
Data governance strategy and policy development	90% (n=18)
Data quality metrics	85% (n=17)
Data quality audit and reporting	90% (n=18)
Effective data quality tools and technologies	80% (n=16)
Data maturity assessment	75% (n=15)
Data quality standards	95% (n=19)
Metadata management and documentation	80% (n=16)
Applying a data quality label to datasets	80% (n=16)
Key data quality criteria and indicators	85% (n=17)

The participants also identified a number of data quality tools and resources, which they found useful. Current data quality resources that participants listed as being beneficial to enhance data quality include the following:

- Guidance on a data quality framework for health and social care
- National Standards for Information Management in Health and Social Care
- European Network of Cancer Registry (ENCR) recommendations and guidance
- International Association of Cancer Registries (IACR) standards and guidance
- Joint Research Centre for European Network of Cancer Registries tools and resources to support quality checks and adherence
- UKIACR (United Kingdom and Ireland Association of Cancer Registries) meetings
- Logical Observation Identifiers Names and Codes (LOINC) coding sessions (Delivered by the HSE Clinical Terminology Architecture Lead)
- Internal procedural documentation supporting data quality.

3.7.2 Data discovery, semantic interoperability and data exchange

Generally, the readiness assessment identified that there is a significant need to develop the capabilities and skills of staff and data holders in key areas relating to data discovery, semantic interoperability and data exchange to prepare the system and data holders for the establishment of a HDAB. Data holders requested that training and guidance are required at a national level to enhance the capabilities of data holders in respect of data sharing, metadata practices, anonymisation techniques, as well as mapping between semantic interoperability standards.

Specifically in the area of metadata, data holders indicated a need for guidance on how to implement Health DCAT-AP. Additionally, the findings highlighted the need to provide targeted training for staff involved in handling and preparing data across the broader healthcare system. Such training will improve digital skills, make source-captured data more consistent and structured, and ensure everyone understands the value of this work.

3.7.3 Other training and guidance

Throughout the focus groups, data holders reported that they require guidance to support their understanding of the changing legal framework in this area. Data holders highlighted several training and educational materials which currently support their enhancement of data quality including HIQA's Guidance on a data quality framework for health and social care and the associated modules. Additional

topics or skills, identified through the qualitative responses that the participants would like to see included in future training and guidance included the following:

- data quality management best practice
- data quality and utility label requirements
- introduction to Health DCAT-AP
- preparing data for linkage including anonymisation techniques
- mapping ICD-10 and ICD-11
- mapping SNOMED and ICD codes
- data provenance documentation
- submission process to HDAB and timelines
- Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Largely, participants listed in-person workshops, live webinars and self-paced online learning courses as their preferred method of training.

3.7.4 Significance of findings – Training and guidance

During discussion on system and dataset preparedness for the EHDS, participants of the readiness assessment regularly raised the need for training and education for data holders. This would suggest a need for coordinated and organised training and education programme to enhance skills across these subject areas, at a national level. However, a more in-depth training-needs analysis is required to inform the development of such a programme.

Some of the training and guidance requirements will be addressed as part of the HealthData@IE project. As part of the project, HIQA will deliver outputs on data quality enhancement for data holders. These outputs have been outlined in section 3.5.6 and include updating the guidance on a data quality framework for health and social care, as well as the associated online training modules to support data holders in improving the quality of their data. In conjunction with this, the project will also develop an information management maturity assessment tool and a compliance assessment framework, both of which will aim to support data holders in enhancing their data quality practices. Furthermore, the project has also committed to a programme of work focusing on the development of online training materials specifically for data holders and for more general audiences, all to be informed by a training-needs analysis. Outputs from the QUANTUM and TEHDAS projects will also be adapted and adopted for an Irish context, as part of the HealthData@IE project.

Participants also identified that some data holders are emerging as leaders in specific areas. Many participants identified that they would benefit from a peer-learning network wherein knowledge sharing can be facilitated on a peer-to-peer level, as well encouraging learning and discussion on relevant topics, as they arise. A peer support network is being developed as part of the HealthData@IE project.

Beyond the HealthData@IE project, further plans will need to be in place to support the effective implementation of a HDAB. Clear guidance and support on the application of legislation relating to data sharing is also required. Data holders will also individually have to address development of capability and capacity at an organisational level requiring skills analysis and training requirements in the short, medium and long term.

4. Conclusion

The implementation of the EHDS Regulation in Ireland brings with it many welcome developments for our national health information system. First, it expedites the need to enhance electronic health records and services such as the establishment of a HDAB, both of which have been identified as essential developments to align with best international practice. Second, it will encompass a concerted effort across Europe to standardise practice in key areas by establishing recommendations, guidelines and technical specifications for the effective functioning of a collective health data space. This puts Ireland in a unique position to learn from the experiences of many of our European counterparts, at such an optimal time of planning and innovation.

For example, it has been documented in many published reports over the past number of years that while Ireland has valuable health information sources, they have not developed in a strategic and coordinated fashion, and therefore remain fragmented and siloed.[18] Despite progress made in relation to digital health, there is still a reliance on paper records in many settings. As the EHDS regulation is primarily designed to regulate the use and exchange of electronic health data across the EU, focused effort is required to achieve the plans set out in *Digital for Care: A Digital Health Framework for Ireland 2024-2030*, which has set out six principles for digital care, and the associated *Digital Health Strategic Implementation Roadmap*.[19, 20] Together, the framework and implementation roadmap set out a plan for the integration of digital technologies in Ireland's healthcare system, advancing key programmes of work such as the National Shared Care Record Programme, the Electronic Health Record (EHR) and the HSE App, as well as other initiatives. The Framework and Roadmap highlight challenges and enablers for the sharing of data. Key among these challenges is the standardisation of data quality practices, specifically the coding across systems and alignment with national and EU law to ensure compliance with privacy, security and consent. It also must be noted that scope of this framework is mainly on the primary use of data. Although enhancing digital records for primary use will ultimately affect the secondary use of information positively, a clear focus on bridging these two uses is also required. This need is growing increasingly acute in reviewing the considerations for the establishment of a HDAB.

Positively, the aim of the HealthData@IE project is to support the development of a HDAB service. This will involve the establishment of a national linkage service in Ireland for health data, in turn bringing many positive changes for data holders and data users. The project focuses on the development of national infrastructures

needed for data access through a HDAB service including a DAAMS to receive, track and process applications and to issue permits, a National Health Dataset Catalogue to facilitate data discovery, and a SPE to ensure the secure processing of health data. Further, the project is delivering work streams on data quality enhancement, seeking to improve the quality of health data made available by a HDAB service, as well as a structured programme of training and engagement with data users, data holders and the public on the EHDS.

The findings of the readiness assessment demonstrate areas of good practices that currently exist across data holders in Ireland, including progress of implementation of the National Standards for Information Management in Health and Social Care, application of data quality frameworks and the use of automated tools and software to promote data quality efforts, as well as implementation of semantic interoperability standards. However the assessment also identified that significant steps need to be taken across Ireland's health and social care system to prepare for the full depth of the EHDS Regulation's requirements and the establishment of a HDAB service. A summary of these include:

- Data holders are awaiting further policy direction and clear guidance to support strategic planning, prioritisation and implementation of the changes needed for compliance with the regulation.
- Clear direction is also needed to support the universal adoption of a unique identifier. This is vital for the effective functioning of a HDAB service. A national system to uniquely identify individuals in the health system has not been implemented across all data holders, limiting the potential to link data for the varied purposes. Many data holders have developed their own approaches to uniquely identifying individuals, which further fragments the system and reduces the ability to link data.
- Significant skill gaps in anonymisation techniques exist and concerns regarding the potential for re-identification post-linkage are evident. However the development of practices to support the safe sharing of data will depend on the operating model chosen for a HDAB in Ireland. Data holders will need clear guidelines to support changes to practices, in line with associated policy decisions.
- As the success of the HDAB service rests on data quality, data holders require strong data strategies with specific goals to continuously monitor and improve data quality. Data holders also need to respond to the legal requirements for data quality assessment and reporting, as set out by the EHDS Regulation. Guidance and training to support data holders in aligning with the dimensions set out by EHDS data quality and utility label will be essential to help them prepare for this requirement. Furthermore, specific support will be required

for smaller data holders with no specific data quality role who may struggle to meet the obligations of the assessing and reporting label without internal expertise in this area.

- Data holders also require a strategic plan to align with recommendations emerging from the European level, with respect to both metadata and data standards. Data holders demonstrate good practices in the implementation of sematic interoperability standards, however inconsistent practices in the use of interoperability and metadata standards has been reported. Data holders are awaiting further guidance regarding the use and reporting of metadata, as well as the development of a National Interoperability Framework to support submission of data to a HDAB, in order to guide practices in line with best practice across Europe.

- A coordinated and organised training and education programme to enhance skills across data holders is essential. This should be developed through an in-depth training-needs analysis to inform the development of such a programme.

The readiness assessment highlights that improving the value of data by enabling effective data sharing and linkage will require clear strategic direction and supporting policy to decide how to address the challenges faced by data holders. The HealthData@IE project has been funded to address some of the specific gaps identified by data holders through this readiness assessment. For example, the project will develop various data quality guidance and training materials, host a peer support network for data holders on the topic of data quality, develop a national interoperability framework, a National Health Dataset Catalogue and the required infrastructure for processing applications and securely providing access to data. The outputs of this project will be made publicly available and are designed to support the secondary use of health data in Ireland. However, there are broader concerns, beyond the scope of the HealthData@IE project which require further consideration at a national level, including a plan to bridge the link between primary and secondary data. Development and implementation of a clear roadmap for change in response to these identified issues is required, with identification of challenges that can be addressed at a local level, with data holders and challenges that require action at a national level. This roadmap will be critical to building preparedness for HDAB implementation.

In addition to the above findings and linked to the need to consider the strategic approach to the secondary use of information, there are also challenges in the way national data holders are organised in Ireland. Ireland does not have an integrated approach to national sources of data compared to other European countries. National data collections tend to be developed in silos, with a single purpose and

have their own infrastructure and governance structures. This additional challenge calls for a clear strategic direction and supporting policy to address the wider health information landscape in Ireland. This will be critical to building preparedness to EHDS at both a national and at a data holder level. While major national improvements are in progress for primary use of data, the readiness assessment has also found a clear need for a coordinated and strategic approach to the implementation of the secondary use of data under the EHDS. This will be critical for, not only compliance with the regulation, but also in supporting the goal of improving healthcare through enhancing data access for important uses like planning and management of services, policy-making, public health and research.

Appendix A – Steering Group Membership for HealthData@IE

Name	Organisation
Emer Doyle (Chair)	Health Information Policy, Department of Health
Dr Barbara Foley	Health Information and Quality Authority
Dr Kevin O'Carroll	Health Information and Quality Authority
Rachel Flynn	Health Information and Quality Authority
Bríd Burke	Health Research Consent Declaration Committee
Prof Markus Helfert	National University of Ireland, Maynooth
Denise Manton	National University of Ireland, Maynooth
Kerry Ryder	HSE
Kathryn Kissane	HSE
Richard Greene	HSE
Loretto Grogan	HSE
Ana Terres	HSE
Dr Teresa Maguire	Health Research Board
Anthony Macken	Central Statistics Office
Azul O'Flaherty	Health Information Policy, Department of Health
Clodagh Thorne	Health Information Policy, Department of Health
Eoin Farrell	Health Information Policy, Department of Health
Brendan Brady	Health Information Policy, Department of Health
Cáit Ní Chorcora	Health Information Policy, Department of Health
Caitriona Wray	Health Information Policy, Department of Health
Helen Conroy	Health Information Policy, Department of Health
Ailish Kelly	Health Information Policy, Department of Health
Laura Flannelly	Statistics and Analytics, Department of Health
Elizabeth McCrohan	Statistics and Analytics, Department of Health
Eamonn Coyne	eHealth and Health Information Systems, Department of Health
Ronan O'Kelly	eHealth and Health Information Systems, Department of Health
Christopher Ryan	Research Services Unit, Department of Health
Róisín O'Neill	Research Policy and Innovation Unit

Appendix B – Working Group for Work Package 8

Name	Organisation
Barbara Foley (Chair)	Health Information and Quality Authority
Maria Ryan	Health Information and Quality Authority
Kevin O'Carroll	Health Information and Quality Authority
Helen Conroy	Health Information Policy, Department of Health
Brendan Brady	Health Information Policy, Department of Health
Anthony Macken	Central Statistics Office
Sarah Craig	Health Research Board
Jean Kelly	National Cancer Registry of Ireland
Cliona O'Donovan	National Office of Clinical Audit
Lorraine McNerney	Office of the Government Chief Information Officer
Frank Moriarty	Royal College of Surgeons in Ireland
Loretto Grogan	Office of the Chief Clinical Officer, HSE
Niall Halliday	Office of the Chief Clinical Officer, HSE
Patricia Wilson	Office of the Chief Information Officer, HSE
Louise Cullen	Health Protection Surveillance Centre
Gabriela Avram	HSE National Patient Forum and IPPOSI Patient Rep
Leonard Browne	University of Limerick

Appendix C – Survey and focus group participants

Participating data holders n=16   Datasets included n= 21

Data Holder	Dataset	Governance	Description	Use Cases
1. **National Screening Service**	BowelScreen	HSE	National Data Collection	Colorectal Cancer
	Diabetic Retina Screen	HSE	National Data Collection	Diabetes
2. **Healthcare Pricing Office**	Hospital Inpatient Enquiry	HSE	National Data Collection	Influenza, Diabetes, Colorectal Cancer
3. **HSE - National Health Schemes Data**	Primary Care Reimbursement Service	HSE	National Data Collection	Influenza, Diabetes, Colorectal Cancer
4. **Chronic Disease Management Programme**	Chronic Disease Management Programme	HSE	eHealth service	Diabetes
5. **Health Protection Surveillance Centre**	Computerised Infectious Disease Reporting system	HSE	National Data Collection	Influenza
6. **HSE – Office of the Chief Clinical Officer**	COVAX: National COVID-19 Immunisation System	HSE	eHealth service	Influenza
7. **HSE - MedLIS**	MedLIS – National Laboratory Information System	HSE	eHealth service	Diabetes
8. **HSE - National Integrated Medical Imaging System**	National Integrated Medical Imaging System	HSE	eHealth service	Colorectal Cancer

Data Holder	Dataset	Governance	Description	Use Cases
9. HSE - National Cancer Control Programme	National Cancer Information System	HSE	eHealth service	Colorectal Cancer
10. St James's Hospital	Electronic Patient Record	St James's Hospital‡‡	Hospital Electronic Health Record	Colorectal Cancer
11. HRB	National Drug Treatment Reporting System	HRB	National Data Collection	Colorectal Cancer
	National Drug-Related Deaths Index	HRB	National Data Collection	Colorectal Cancer
	National Psychiatric Inpatient Reporting System§§	HRB	National Data Collection	Not related to a specific use case, see footnote§§
	National Ability Support Services§§	HRB	National Data Collection	Not related to a specific use case, see footnote§§**Error! Bookmark not defined.**
12. NOCA	Irish National ICU Audit	NOCA	National Data Collection	Influenza, Diabetes
	National Paediatric Mortality Register	NOCA	National Data Collection	Influenza, Diabetes
13. National Cancer Registry of Ireland	National Cancer Registry of Ireland	National Cancer Registry of Ireland	National Data Collection	Colorectal Cancer

---

‡‡ As a Section 38 Agency, St James's Hospital has a Service Level Agreement in place with the HSE.

§§ These datasets were not included in the Readiness Assessment protocol as they were self-identified as additional datasets, which were reviewed by the project team and deemed suitable for inclusion in the analysis and report.

Data Holder	Dataset	Governance	Description	Use Cases
14. Children's Health Ireland at Tallaght University Hospital and Trinity College Dublin	Irish Childhood Diabetes National Register	Children's Health Ireland and Trinity College Dublin	National Data Collection	Diabetes
15. CSO	Vital Statistics	CSO	National Data Collection	Influenza, Diabetes, Colorectal Cancer
16. Trinity College Dublin	The Irish Longitudinal Study on Ageing	Trinity College Dublin	National Survey	Influenza

Appendix D – Use case definitions and EHDS requirements

Definition	EHDS Requirements
Data quality is the degree to which the elements of electronic health data are suitable for their intended primary use and secondary use.[2]	'Data quality management and quality assurance are integral pieces within the data governance of an organisation. A data quality assurance framework should be transversal to all data management processes including monitoring, detection and resolution of incidences, and data enrichment procedures. Data quality management should be applied throughout the full data life cycle focusing on: • data collection, curation, storage and staging • data integration with relevant sources and systems • data description and metadata management (meaning use of meta-data standards) • data quality assessment, data profiling and remediation • data modelling, transformation, operationalisation and servicing.'[21]
Data linkage is the matching of records from two or more discrete datasets. It can be either deterministic which compares the matching variables across datasets to find exact matches, or probabilistic which estimates the probability that records match.[14]	'Data holders should implement data management procedures to allow datasets linkage and IDs persistence. In the case of sensitive data, those individual IDs should be pseudonymised and persisted across datasets and overtime.'[21]
Data interoperability is the ability of organisations, as well as of software applications or devices from the same manufacturer or different manufacturers, to interact through the processes they support, involving the exchange of information and knowledge, without changing the content of the data, between those organisations, software applications or devices.[2]	'Once data quality management and quality assurance procedures ensure a certain level of quality, data holders may need to implement a semantic and syntactic interoperability layer across datasets. • Syntactic as how data is structurally persisted within a dataset (literal name, standard abbreviation and encoding) • Semantic as consistency in data meaning across datasets (clear description, operational definition rules, mapping across controlled vocabularies and standards).'[21]

Appendix E – Survey graphs

EHDS preparedness and information management maturity

EHDS ORGANISATIONAL KNOWLEDGE

TOTAL ORGANISATIONS, 16

- VERY LOW: 6.3% (n=1)
- LOW: 25.0% (n=4)
- AVERAGE: 37.5% (n=6)
- HIGH: 18.8% (n=3)
- VERY HIGH: 12.5% (n=2)

EHDS organisational preparedness

Total organisations, 16

- Very Low: 6.3% (n=1)
- Low: 25.0% (n=4)
- Average: 56.3% (n=9)
- High: 12.5% (n=2)
- Very High: 0.0% (n=0)
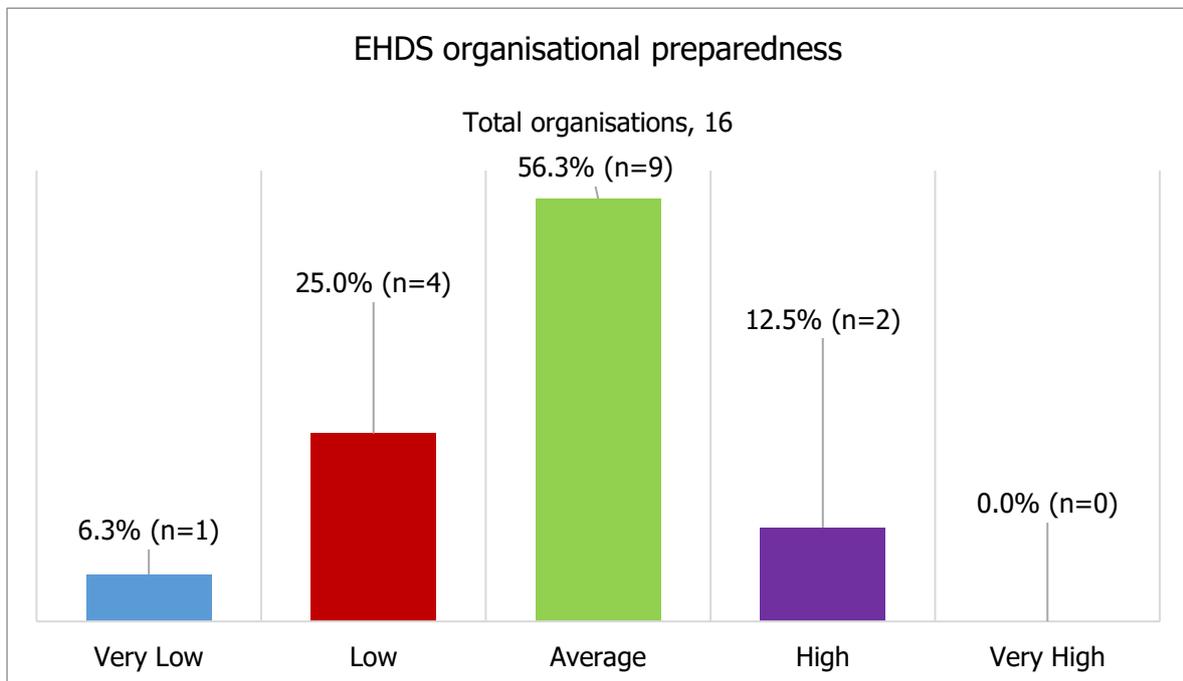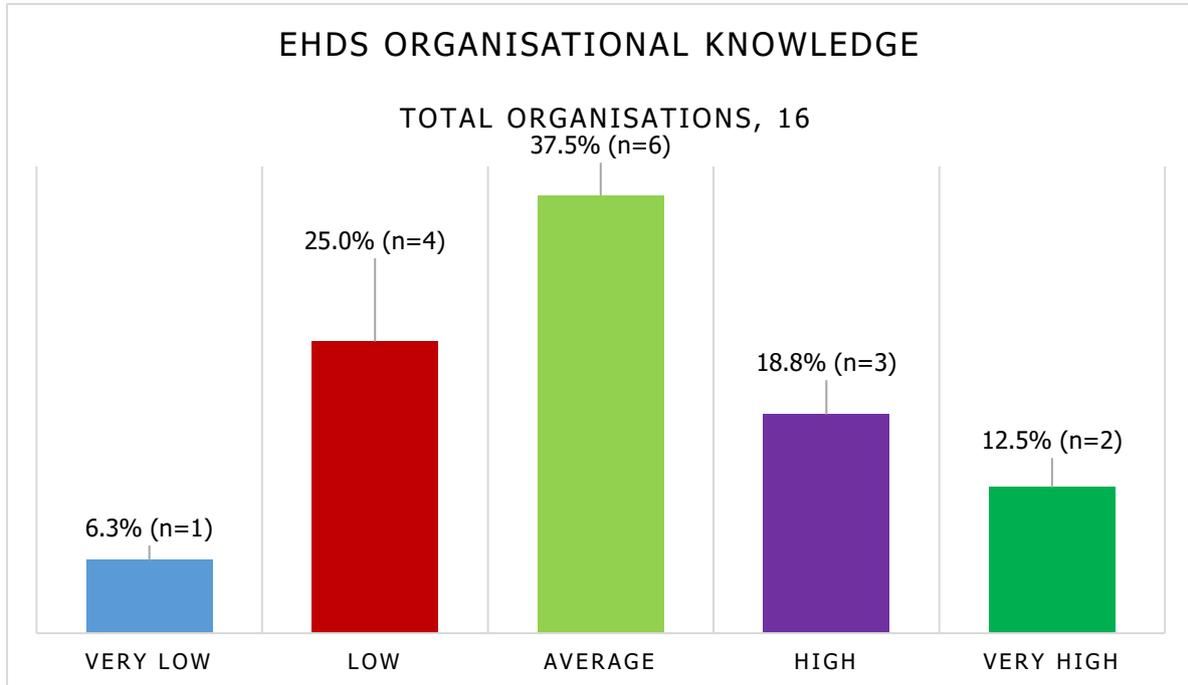
Is the organisation taking steps to prepare for EHDS?

Total Organisations, 16

No, 12.50%
(n=2)

Yes, 87.50%
(n=14)
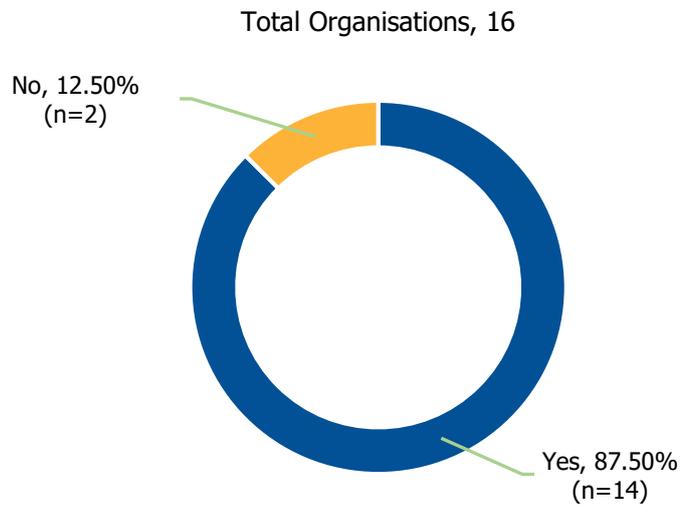
Is there an individual responsible for ensuring the organisation takes steps to prepare for the EHDS?

Total Organisations, 16

No, 31.25%
(n=5)

Yes, 68.75%
(n=11)

Level of knowledge of *National Standards for Information Management Standards in Health and Social Care*

Total organisations, 16

50.00% (n=8)

37.50% (n=6)

0.00% (n=0)

12.50% (n=2)

0.00% (n=0)

Very Low	Low	Average	High	Very High

Have the *National Standards for Information Management Standards in Health and Social Care* been implemented?

Total Organisations, 16

no   0.00% (n=0)

Don't know   25.00%(n=4)

Yes, partially implemented   56.25% (n=9)

Yes, fully implemented   18.75% (n=3)

Level of maturity on information management

Total organisations, 16

Level	Percentage
Optimised	6.25% (n=1)
Managed	43.75% (n=7)
Defined	31.25% (n=5)
Repeatable	12.50% (n=2)
Initial	6.25% (n=1)

Data linkage and anonymisation

Is there a method to uniquely identify individuals in the dataset?

Total Datasets, 21

No, 4.76% (n=1)

Yes, 95.24% (n=20)

Variables used to uniquely identify individuals

Total Datasets, 21

Variable	Percentage
Mother's Birth Family Name	14% (n=3)
Personal Public Service number (PPSN)	48% (n=10)
Eircode	48% (n=10)
Gender	48% (n=10)
Address Line 1	57% (n=12)
Individual Health Identifier (IHI)	57% (n=12)
Last Name	67% (n=14)
First Name	67% (n=14)
Other	71% (n=15)
Sex	81% (n=17)
Date of Birth	90% (n=19)

Is the dataset ever linked to other datasets?

Total Datasets, 21

Response	Percentage
No	33.33% (n=7)
Don't know	4.76% (n=1)
Yes, on an ad hoc basis	28.57% (n=6)
Yes, routinely	33.33% (n=7)
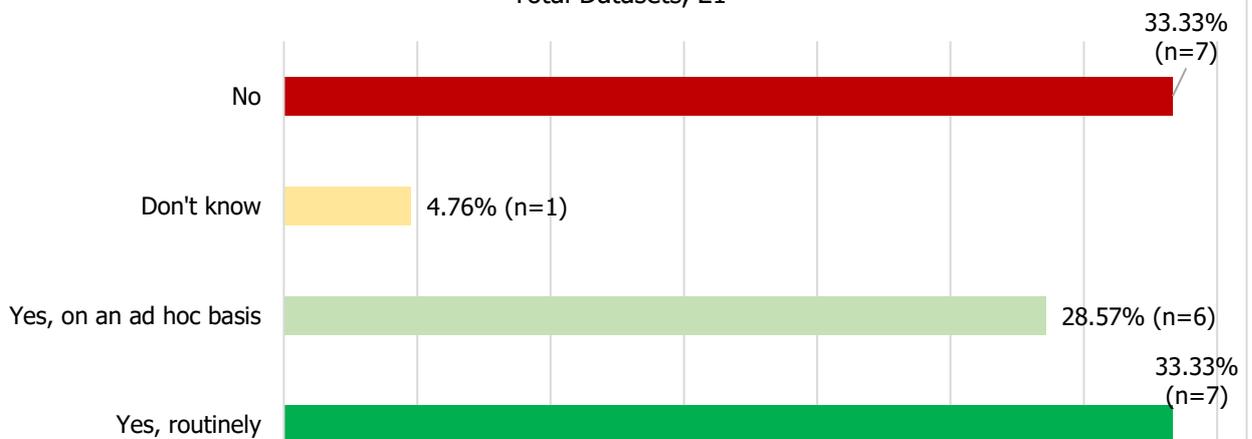
Is there a mechanism for recording an individuals "opt out" preference?

Total Datasets, 21

Don't know, 4.76% (n=1)

Yes, 23.81% (n=5)

No, 71.43% (n=15)

Are anonymisation techniques used on this dataset?

Total Datasets, 21

Yes, 38.10% (n=8)

No, 61.90% (n=13)

Are pseudonymisation techniques used on this dataset?

Total Datasets, 21

Don't know, 4.76% (n=1)

No, 38.10% (n=8)

Yes, 57.14% (n=12)

Are any other anonymisation techniques used on this dataset?

Total Datasets, 21

Don't know, 14.3% (n=3)

No, 42.9% (n=9)

Yes, 42.9% (n=9)

Data quality

Are there processes to improve and maintain data quality in this dataset?

Total datasets, 21

Don't know, 4.76% (n=1)

No, 9.52% (n=2)

Yes, 85.71% (n=18)

Is there a data quality framework in place for this dataset?

Total datasets, 21

Don't know, 4.76% (n=1)

No, 38.10% (n=8)

Yes, 57.14% (n=12)

Components of a data quality framework in place for this dataset

Total datasets, 21

Component	Percentage
Data quality improvement cycle	42.86% (n=9)
Data quality reports	42.86% (n=9)
Data quality assessment methods	47.62% (n=10)
Data quality strategy	23.81% (n=5)

Which components are included in your data quality strategy?

Total Datasets, 21

Component	Percentage
Standards for data quality	23.81% (n=5)
Data quality audits	23.81% (n=5)
Training, education and development programme for data quality	23.81% (n=5)
Data quality policies and procedures	23.81% (n=5)
Governance, leadership and management structures in relation to data quality	23.81% (n=5)

Overall maturity of data quality practices for dataset

Total datasets, 20

Optimised	25% (n=5)
Managed	30% (n=6)
Defined	20% (n=4)
Repeatable	25% (n=5)
Initial	0% (n=0)

Effectiveness of data quality policies & procedures in meeting organisations needs for producing fit-for-purpose data?

Total Organisations, 16

Not effective at all	Slightly effective	Moderately effective	Very effective	Extremely effective
0.00% (n=0)	6.25% (n=1)	56.25% (n=9)	31.25% (n=5)	6.25% (n=1)

Is data quality training and education available?

Total organisations, 17

Don't know, 5.88% (n=1)

No, 11.76% (n=2)

Yes, 82.35% (n=14)

Current Levels of data quality expertise and ability to manage data quality

Total Datasets, 21

- No expertise: 0.00% (n=0)
- Little expertise: 0.00% (n=0)
- Some expertise: 28.57% (n=6)
- Good Expertise: 38.10% (n=8)
- Extensive expertise: 28.57% (n=6)
- Don't know: 4.76% (n=1)

Perception of the current levels of expertise among staff with data quality responsibilities

Total datasets, 20

Legend: ■ No expertise ■ Little expertise ■ Some expertise ■ Good expertise ■ Expert expertise

Category	No expertise	Little expertise	Some expertise	Good expertise	Expert expertise
Metadata management and documentation		5%	35%	35%	25%
Using relevant data quality standards		5%	35%	50%	10%
Data cleaning techniques			30%	45%	25%
Data quality auditing and reporting		10%	45%	30%	15%
Data integration and consolidation methods		10%	35%	40%	15%
Developing and applying data quality metrics		10%	40%	40%	10%
Using data quality improvement software	10%	10%	35%	40%	5%
Communicating data quality issues to stakeholders		5%	30%	25%	40%

Standards for data discoverability, semantic interoperability and interoperable communication

Is a specific descriptive metadata standard used to describe the dataset?

Total datasets, 21

Don't know, 28.6% (n=6)

Yes, 42.9% (n=9)

No, 28.6% (n=6)

Has the general DCAT-AP specification been used to describe the dataset?

Total datasets, 21

No, 100.0% (n=21)

Do you have plans to use Health DCAT-AP to describe the dataset?

Don't know, 14.3% (n=3)

Total datasets, 21

Yes, 38.1% (n=8)

No, 47.6% (n=10)

Are details about your dataset discoverable by data users by any of the following?

Total datasets, 21

Other — 52.4% (n=11)

Browsing Catalogue — 33.3% (n=7)

Viewing Metadata — 42.9% (n=9)

Search engine — 47.6% (n=10)

Does your organisation currently share descriptive metadata with other systems?

Total datasets, 21

No, 28.6% (n=6)

Yes, 71.4% (n=15)

Methods used to share descriptive metadata with other systems

Total datasets, 21

Method	Value
HDAB service harvests data via an API (pull)	0.0% (n=0)
Uploading data via an application programming interface (API) (push)	14.3% (n=3)
Manual submission through form-based user interface (UI)	9.5% (n=2)
Manual submission through file uploads (for example, CSV)	19.0% (n=4)

Does your organisation currently share aggregated data with other systems?

Total datasets, 21

No, 19.0% (n=4)

Yes, 81.0% (n=17)

Does your organisation currently share individual-level data with other systems?

Total datasets, 21

No, 47.6% (n=10)

Yes, 52.4% (n=11)

Methods used to share individual level data with other systems

Total datasets, 21

Method	Percentage
HDAB service harvests data via an API (pull)	0.0% (n=0)
Uploading data via an application programming interface (API) (push)	23.8% (n=5)
Manual submission through form-based user interface (UI)	14.3% (n=3)
Manual submission through file uploads (for example, CSV)	38.1% (n=8)

Does your organisation have a plan in place for providing data to the HDAB service using the standards for interoperability identified at the European level?

Total datasets, 21

Don't know, 28.6% (n=6)

Yes, 23.8% (n=5)

No, 47.6% (n=10)

Please indicate your organisation's level of preparedness for providing data to a HDAB service?

Total datasets, 21

61.9% (n=13)

28.6% (n=6)

0.0% (n=0)

9.5% (n=2)

0.0% (n=0)

| Very Low | Low | Average | High | Very High |

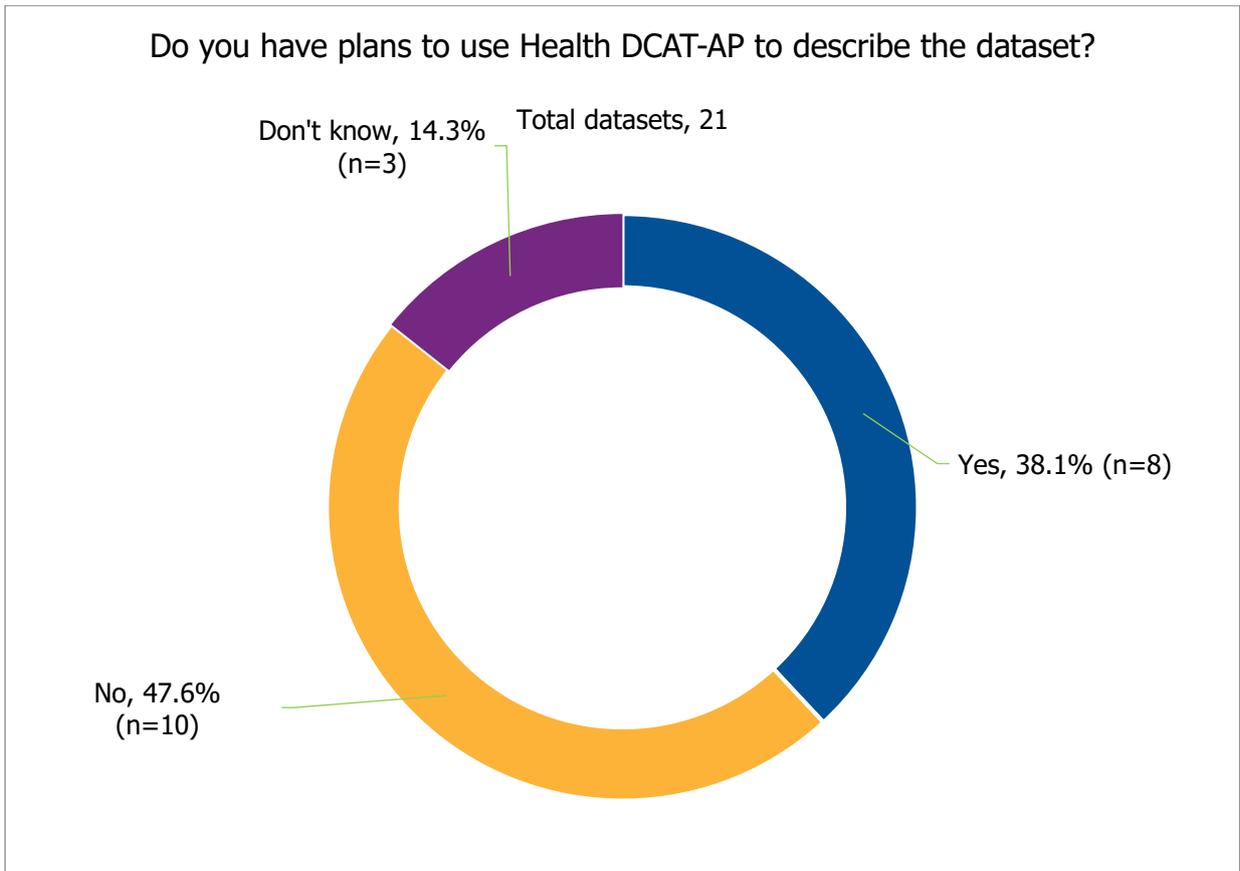## Level of maturity of dataset in the implementation of standards for semantic interoperability and interoperable communication

Total datasets, 21

38.1% (n=8)

28.6% (n=6)

19.0% (n=4)

14.3% (n=3)

0.0% (n=0)

| Initial – little coordination and practices remain ad hoc | Repeatable – basic practices are in place, but variation exists between people and teams | Defined – standardised practices are documented and in place but there is still room for improvement | Managed – practices are established and controlled, and regularly monitored | Optimised – practices are continuously monitored to enable optimisation and improvement |

Semantic interoperability standards implemented in the dataset

Total datasets, 18

Standard	Percentage
Other	77.8% (n=14)
ICD-10-AM	44.4% (n=8)
Orphanet codes	5.6% (n=1)
SNOMED CT	33.3% (n=6)
LOINC	11.1% (n=2)

Chart Standards for data exchange implemented in the dataset

Total datasets, 17

Standard	Percentage
Other	58.8% (n=10)
DICOM	11.8% (n=2)
HL7 FHIR	5.9% (n=1)
HL7 version 2.x	52.9% (n=9)

Preferences and priorities for future training and guidance

Data quality guidance and training needs

Total Data Holders, 20

Category	Value
Standards for data exchange	85% (n=17)
Standards for semantic interoperability	80% (n=16)
Data quality improvement strategies	95% (n=19)
Data governance strategy and policy development	90% (n=18)
Data quality metrics	85% (n=17)
Data quality audit and reporting	90% (n=18)
Effective data quality tools and technologies	80% (n=16)
Data maturity assessment	75% (n=15)
Data quality standards	95% (n=19)
Metadata management and documentation	80% (n=16)
Applying a data quality label to datasets	80% (n=16)
Key data quality criteria and indicators	85% (n=17)

Most suitable format of data quality education - First to third preferences

Total datasets, 21

Legend:
- Other
- Peer-to-peer learning opportunities
- Checklists
- Short informative videos
- Reading materials, including process guides, guidance documents
- Self-paced online courses
- Live webinars
- In-person workshops

First Preference
- 4.8%
- 9.5%
- 9.5%
- 19.0%
- 19.0%
- 23.8%
- 19.0%
- 52.4%

Second Preference
- 9.5%
- 4.8%
- 4.8%
- 9.5%
- 52.4%
- 4.8%

Third Preference
- 4.8%
- 4.8%
- 9.5%
- 19.0%
- 33.3%
- 4.8%
- 14.3%
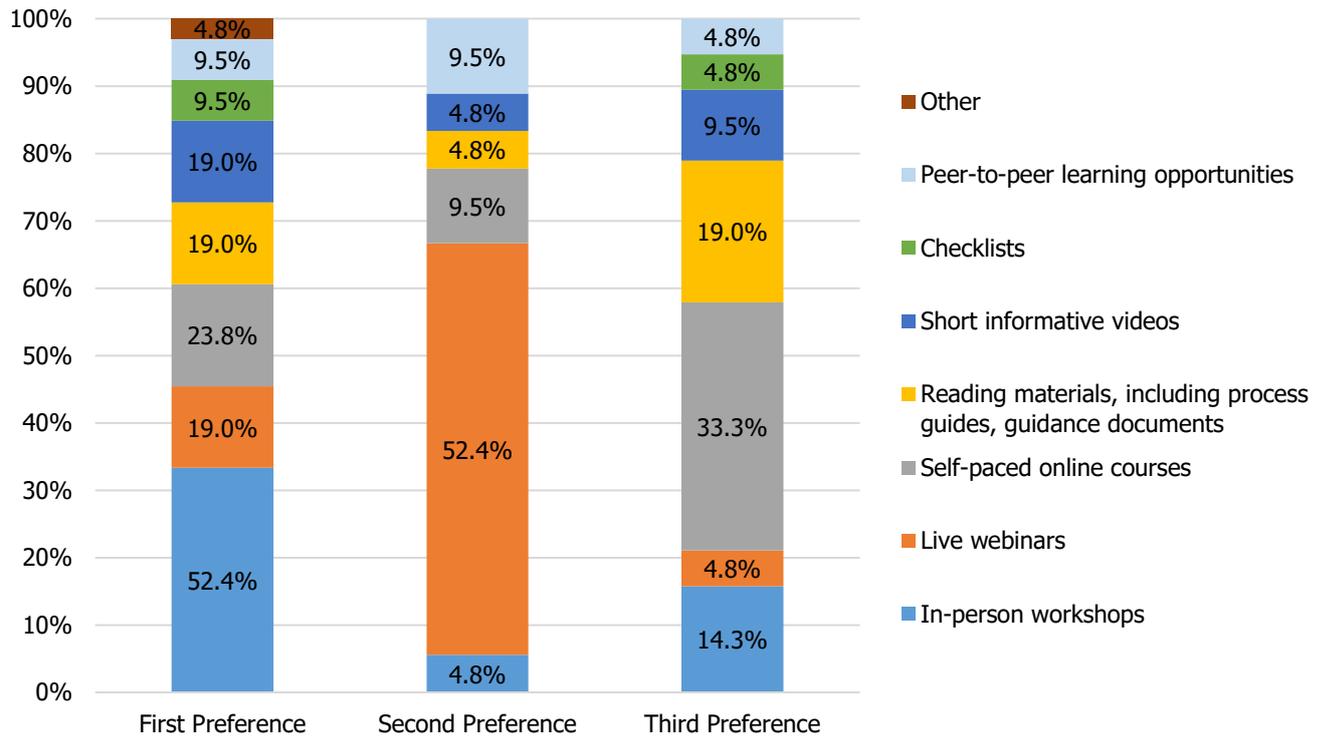
References

1.   TEHDAS. *Identification of relevant standards and data models for semantic harmonization.* Finland: 2022. Available from: h[ttps://tehdas.eu/tehdas1/app/uploads/2022/02/tehdas-identification-of-relevant-standards-and-data-models-for-semantic-harmonization-2022-02-03.pdf.](https://tehdas.eu/tehdas1/app/uploads/2022/02/tehdas-identification-of-relevant-standards-and-data-models-for-semantic-harmonization-2022-02-03.pdf) Accessed on: 30 August 2024.

2.   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), Brussels: EU Commission; 2025. Available from: h[ttps://eur-lex.europa.eu/eli/reg/2025/327/oj.](https://eur-lex.europa.eu/eli/reg/2025/327/oj) Accessed on: 05 March 2025.

3.   European C. *European Data Governance Act* [Online]. Available from: h[ttps://digital-strategy.ec.europa.eu/en/policies/data-governance-act.](https://digital-strategy.ec.europa.eu/en/policies/data-governance-act) Accessed on: 19 August 2024.

4.   European C. *Data Act* [Online]. Available from: h[ttps://digital-strategy.ec.europa.eu/en/policies/data-act.](https://digital-strategy.ec.europa.eu/en/policies/data-act) Accessed on: 19 August 2024.

5.   European C. *Europe's Digital Decade: digital targets for 2030* [Online]. Available from: h[ttps://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.](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.

6.   Pilot EHE. *Health Data@EU Pilot: Piloting an infrastructure for the secondary use of health data* [Online]. Available from: h[ttps://ehds2pilot.eu/.](https://ehds2pilot.eu/) Accessed on: 19 August 2024.

7.   TEHDAS. *Joint Action Towards the European Health Data Space – TEHDAS1* Available from: h[ttps://tehdas.eu/tehdas1/project/.](https://tehdas.eu/tehdas1/project/) Accessed on: 13 October 2025.

8.   TEHDAS. *Second Joint Action Towards the European Health Data Space – TEHDAS2* [Online]. Available from: h[ttps://tehdas.eu/.](https://tehdas.eu/) Accessed on: 19 August 2024.

9.   Health Information Bill, Dublin: The Stationary Office; 2024. Available from: h[ttps://www.gov.ie/pdf/?file=https://assets.gov.ie/299641/edc11db0-effc-4bd9-bdc9-98e8a233e657.pdf#page=null.](https://www.gov.ie/pdf/?file=https://assets.gov.ie/299641/edc11db0-effc-4bd9-bdc9-98e8a233e657.pdf#page=null) Accessed on: 19 August 2024.

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

11.   Health Information and Quality Authority. *Preparing for the establishment of health data access body services in Ireland under the European Health Data Space Regulation: Readiness Assessment Protocol* 2025. Available from: h[ttps://www.hiqa.ie/sites/default/files/2025-04/Readiness-assessment-protocol.pdf.](https://www.hiqa.ie/sites/default/files/2025-04/Readiness-assessment-protocol.pdf) Accessed on: 19 August 2025.

12.   Quantum. *Quantum - the health data quality label* [Online]. Available from: h[ttps://quantumproject.eu/.](https://quantumproject.eu/) Accessed on: 20 August 2024.

13.   Department of Public Expenditure and Reform. *Open Data Strategy 2017-2022.* 2017. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/h[ttps://data.gov.ie/sites/default/files/files/Final%20Strategy%20online%20version(1).pdf.](https://data.gov.ie/sites/default/files/files/Final%20Strategy%20online%20version(1).pdf) Accessed on: 09 September 2025.

14.   Fennelly O, Moriarty F, Corrigan D, Grogan L, Wong SS. *Proof of concept: Technical prototype for Data Access, Storage, Sharing and Linkage (DASSL) to support research and innovation in Ireland*. Dublin: Health Research Board; 2022. Available from. Accessed.

15.   Health Information and Quality Authority. *Preparing for the establishment of a health data access body services in Ireland: Guidance for data holders.* 2025. Available from: h[ttps://www.hiqa.ie/sites/default/files/2025-04/Health-Information-Readiness-Assessment-Guidance.pdf.](https://www.hiqa.ie/sites/default/files/2025-04/Health-Information-Readiness-Assessment-Guidance.pdf) Accessed on: 13 October 2025.

16.   Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*. 2013; 13(1):117. Available from: h[ttps://doi.org/10.1186/1471-2288-13-117.](https://doi.org/10.1186/1471-2288-13-117) Accessed.

17.   Health Information and Quality Authority. *National Standards for Information Management in Health and Social Care.* Dublin: 2024. Available from: h[ttps://www.hiqa.ie/sites/default/files/2024-06/National-Standards-for-Information-Management-2024.pdf.](https://www.hiqa.ie/sites/default/files/2024-06/National-Standards-for-Information-Management-2024.pdf) Accessed on: 13 February 2025.

18.   Health Information and Quality Authority. *The need to reform Ireland's national health infomraiton system to support the delivery of health and social care services.* Available from: h[ttps://www.hiqa.ie/sites/default/files/2021-10/The-need-for-reform-of-the-health-information-system.pdf.](https://www.hiqa.ie/sites/default/files/2021-10/The-need-for-reform-of-the-health-information-system.pdf) Accessed on.

19.   Department of Health. *Digital for Care: A Digital Health Framework for Ireland 2024-2030* [Online]. Available from: h[ttps://assets.gov.ie/293780/5c6e1632-10ed-4bdc-8a98-51954a8da2d0.pdf.](https://assets.gov.ie/293780/5c6e1632-10ed-4bdc-8a98-51954a8da2d0.pdf) Accessed on: 13 February 2025.

20.   Department of Health. *Digital Health Strategic Implementation Roadmap.* 2024. Available from: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://assets.publications.hse.ie/media/file_based_publications/Digital_Health_Strategic_Implementation_Roadmap.pdf. Accessed on: 26 September 2025.

21.   TEHDAS. *Recommendations on a Data Quality Framework for the European Health Data Space for secondary use.* Finland: 2023. Available from: https://tehdas.eu/tehdas1/results/tehdas-proposals-for-data-quality-and-utility-in-ehds/. Accessed on: 19 August 2024.