Important: Press F2 to save title in document property (for accessibility reasons)

xShare

Expanding the European EHRxF to share and effectively use health data within the EHDS

Public health Threats: state of the art, needs analysis and use cases- WP4 - EHTEL

Date: 17.02.2025

Project title: xShare - Expanding the European EHRxF to share and effectively use health data within the EHDS.

Grant Agreement: 101136734

Call identifier: HORIZON-HLTH-2023-IND-06-02

Dissemination level: Public

This project has received funding from the European Health and Digital Executive Agency (HADEA) under grant agreement no. 101136734.

Working paper description

Number and name of working paper:	Public health threats: state of the art, needs analysis and use cases-
Publishable summary:	This document summarises major current strategic and priority secondary use data collection initiatives and analyses them with the aim of bridging the gap between primary and secondary data use for better healthcare outcomes and policymaking. It documents important lessons learnt from "only once" initiatives and discusses the need of a better alignment between resources used in care, research, epidemiology and decision support.
Status:	Final
Version:	1.3
Last update:	17.02.2025
Deadline:	31.01.2025
Actual delivery:	18.02.2025
Lead beneficiary:	EHTEL
Contact:	Luc Nicolas luc.nicolas@ehtel.eu
Contributors:	Luc Nicolas (EHTEL), Eugenia Rinaldi (Charite), Sasha Milbeck (Sciensano) Nienke Schutte (Sciensano), Javier Samper- Zapater (UoV), Carola Schulz (EMP) Rebecca Baker (CDISC), Anderson Carmo (ITSCE): Reviewers
Editors:	Luc Nicolas (EHTEL), Eugenia Rinaldi (Charite), Sasha Milbeck (Sciensano)

Statement of originality

This working paper contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer

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

Table of Contents

List of abbreviations

Abbreviation	Term
AMR	antimicrobial resistance
AMS	antimicrobial stewardship
ATC	anatomical therapeutic chemical
ATHINA	Advanced Technology for Health INtelligence and Action IT system
CBRN	chemical, biological, radiological and nuclear substances
CDA	clinical document architecture
CDISC	Clinical Data Interchange Standards Consortium
CDM	common data model
CRG	clinical risk groups
DCAT-AP	Application profile for data portals in Europe
DRG	Diagnostic-Related Groups
EEA	European economic area
EEHRxF	European electronic health record exchange format
EHDS	European health data space
EHR	electronic health record
ETL	Extraction Transformation Loading
EU	European Union
ePI	electronic product information
FHIR	fast healthcare interoperability resources
GP	general practitioner
HDAB	Health Data Access Bodies
HEALTHDCAT-AP	Health Data Catalog Vocabulary- Application Profile
HERA	European Commission's Health Emergency Preparedness and Response Authority
ICD	international classification of diseases
LOINC	logical observation identifiers names and codes (Regenstrief Institute, Inc.)
MS	member state
NCD	non communicable disease
NCIt	National Cancer Institute thesaurus
OHDSI	observational health data sciences and informatics
OMOP	observational medical outcomes partnership
PHM	population health management
PHR	personal health record
PREMs	patient reported experience measures
PROMs	patient reported outcome measures
SDTM	study data tabulation model
SNOMED CT	SNOMED clinical terms

List of figures

Executive summary

The xShare project focuses on enhancing the European Electronic Health Record Exchange Format (EHRxF)[1] to effectively share and use health data within the European Health Data Space (EHDS). This working paper presents a state-of-the-art analysis, needs assessment, and use cases in public health, aiming to bridge the gap between primary and secondary data use for better healthcare outcomes and policymaking.

There is a significant disconnect between primary (clinical) and secondary (research and policy) data uses. This results in inefficiencies, duplicated data entry, and fragmented data models, posing challenges for healthcare professionals and innovation projects.

The EHDS regulation facilitates the standardisation and interoperability of health data. It emphasises the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, promoting better data alignment for primary and secondary uses.

This working paper surveys major European initiatives which are targeting priority public health topics including infectious diseases monitoring, cancer, resistance to antibiotics (AMR) and health system management focusing on communicable and non-communicable diseases (NCDs). The topics are analysed with the goals of improving the connection of current datasets with the EEHRxF (as specified by the EHDS), identifying priority use cases and considering adaptations of the formats which could deliver added value for public health.

A dedicated survey revealed significant variations among European countries in the standardisation of Electronic Health Record (EHR) data, which are used for datasets and support secondary use. The survey intended to update the existing repository of datasets in EU MS initiated by previous projects and identify use cases and datasets which require real-time or near real-time data collection. It revealed only a limited number of core datasets that feature continuous data updates and ensure real-time data exchange and availability. The observed difficulty to obtain information on use cases, datasets and processes which create a direct connection between primary and secondary use in EU Member States demonstrates that although critical for the successful implementation of the EHDS and for the development of use cases with a high added value, the issue is not yet sufficiently high in MS' agendas.

Several EU Member States have yet to fully commit to a coordinated effort to inventory and document all existing data sets. Under the EHDS, national or regional Health Data Access Bodies (HDABs) need to

[1] The EHDS Regulation uses the European Electronic Health Record Exchange Format (EEHRxF) as the means for ensuring the secure and interoperable exchange of electronic health data across the EU. The EEHRxF comprises a set of requirements and technical specifications supporting both structured and unstructured data types. The format's scope is based on six priority categories of personal electronic health data in the EHDS: patient summaries; electronic prescriptions; electronic dispensations; medical imaging studies and related imaging reports; medical test results, including laboratory and other diagnostic results and related reports; discharge reports. See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:PE_76_2024_INIT

provide a publicly available metadata catalogue, thereby addressing this gap. Adopting HealthDCAT-AP to describe datasets could further facilitate this process and serve as an important enabler for interoperability. As far as data models and terminologies are concerned, there are some encouraging signs such as the existing mapping between SNOMED-CT and MedDRA or ICD-10 but dealing with this complex issue requires both significant public investment both at national and EU levels such as official terminology servers and investment in both new human and technical resources at the level of healthcare organisations. While existing data models have all been built with the best of intentions considering the finality pursued, they usually failed to consider the data models used for other finalities. A progressive simplification and alignment is thus necessary to guarantee improved data integration.

While existing mapping between terminologies such as SNOMED-CT, MedDRA and ICD-10 encourage standardisation, their broader implementation requires substantial public investment at both national and EU levels. Support is needed to set up official terminology servers and improving human and technical resources within healthcare organisations. Moreover, existing data models although engineered with specific goals in mind, often overlook alignment with data models created for other purposes. Alignment of these models, such as the work performed by the BRIDG project, is necessary to improve data integration and interoperability.

The four “only once” best practice reports summarised in this document, with full versions available in the annexes, document initiatives that have moved beyond the proof-of-concept stage and are now fully implemented. These reports also outline the drivers required at technical, legal, financial and organisational levels to facilitate the FAIRification of data and the roles public and private actors can play in fostering cross-domain connections. Although those initiatives are not yet directly connected to the EEHRxF (as those still need to be officially described), the categories of data under consideration are very close to the main data fields of the 6 EEHRxF priority domains. All reports demonstrate that the buy-in from most of the data producers is necessary before considering a full-scale (legal) implementation. This buy-in is best acquired when the added value brought by the data workflows is multi-fold and different priority expectations can be met. The rapid development of AI might also significatively reduce the number of constraints imposed on the data providers and improve overall usability.

The working paper underscores the need for a cohesive approach to health data management, promoting interoperability and real-time data usage. It serves as a comprehensive guide to understanding the current landscape and future direction for the utilisation of health data in Europe and aligning with EHDS regulations to support public health and policymaking.

Thanks to those initial inputs, xShare WP4 is now well equipped to explore and validate public health use cases, enhance data governance, and foster collaborations across EU Member States to leverage health data for better public health outcomes.

1. Introduction

Primary and secondary use[2] of data have often been considered as two separate worlds with their own rules but also their own data models and standards. However, in many (but not all) instances, the original producer of data is the same: the clinician. There is an important paradox: while important resources are being used to structure, normalise and validate data for secondary use related to a specific finality, most of the information produced for primary use remains in free text and is loosely structured and very seldom coded. The data normalised in the context of secondary use are also rarely re-used to support continuity and quality of care. Clinicians often raise concerns that such data are either not fit for purpose or lack full reliability.

This disconnect between primary and secondary use of data raises significant efficiency concerns. Data often must be duplicated across multiple registration systems, which frequently use incompatible data models and terminologies. This creates a substantial administrative burden for healthcare professionals and organisations. Furthermore, data usage is often confined to one single purpose, limiting the development of secondary use. Each single innovation project, such as those testing new decision-support tools or knowledge generation and management processes, must often build a dedicated architecture from scratch to integrate, structure and normalise data. The "ad hoc" ontologies developed for these initiatives[3] are typically tailored to specific objectives, making their reuse challenging and also far from straightforward.

The COVID-19 pandemic has demonstrated the dire need for health systems to rapidly monitor and analyse data produced by clinical systems in order to improve decision making and responsiveness at all levels (local, regional, national, European and worldwide). Interoperability, security and governance are the three main pillars which can make this possible. The rapid uprise of AI also sheds a new light on the capacity to consolidate and generate knowledge for public health, when considering prediction scenarios and foresight for policymaking.

[2]According to the EHDS Regulation **primary use of electronic health data'** means the processing of personal electronic health data for the provision of health services to assess, maintain or restore the 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 while **'secondary use of electronic health data'** means the processing of electronic health data for purposes set out in Chapter IV of this Regulation. 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 the secondary use. See: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:PE_76_2024_INIT

[3] See for example: https://ceur-ws.org/Vol-3005/09paper.pdf

The EHDS regulation offers a unique opportunity to evolve towards a global "FAIRification[4]" of the categories of data critical for quality and continuity of care. The future capacity of clinical systems to generate data compliant with the European EEHRxF format provides clear references for all stakeholders of the health value chain. It provides for the first time a robust baseline for the creation of a dynamic and open ecosystem which should contribute to an improved alignment between data models and standards used for primary and secondary use.

xShare also hopes to bring a contribution for the preparation of implementing acts (related to article 80: "Minimum dataset specifications" which says: "*The Commission may, by means of implementing acts, determine the minimum specifications for datasets of high impact for the for secondary use of electronic health data, considering existing Union infrastructures, standards, guidelines and recommendations. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 98(2)*." With the phrase "datasets of high impact", we understand the data sets that can bring the most added value to society and/or the ones which can support decision-making and data discovery.

Article 51 of the EHDS Regulation[5] refers to seventeen categories of health data that should be made available for secondary use. Namely:

A) Electronic health data from EHRs
B) Data on factors impacting on health, including socio-economic, environmental and behavioural determinants of health
C) Aggregated data on healthcare needs, resources allocated to healthcare, the provision of and access to healthcare, healthcare expenditure and financing
D) Data on pathogens that impact human health
E) Healthcare-related administrative data, including on dispensations, reimbursement claims and reimbursements
F) Human genetic, epigenomic and genomic data
G) Other human molecular data such as proteomic, transcriptomic, metabolomic, lipidomic and other omic data
H) Personal electronic health data automatically generated through medical devices
I) Data from wellness applications
J) Data on professional status, and on the specialisation and institution of health professionals involved in the treatment of a natural person
K) Data from population-based health data registries such as public health registries
L) Data from medical registries and mortality registries
M) Data from clinical trials, clinical studies, clinical investigations and performance studies subject to Regulation (EU) No 536/2014, Regulation (EU) 2024/1938 of the European Parliament and of the Council, Regulation (EU) 2017/745 and Regulation (EU) 2017/746
N) Other health data from medical devices
O) Data from registries for medicinal products and medical devices

[4] FAIR stands for 'Findable, Accessible, Interoperable, and Reusable. See also: Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[5] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:PE_76_2024_INIT

P) Data from research cohorts, questionnaires and surveys related to health, after the first publication of the related results
Q) Health data from biobanks and associated databases

Although all data categories can be considered relevant, categories in bold (A), (B), (E), (F), (H), and (I) are those which are particularly relevant for xShare.

By giving every European citizen, the right to access their health data, the EHDS regulation also creates the conditions necessary to play a direct role in the monitoring, evaluation and development of public and population health policies. This is especially true when considering, for example, the scope of quaternary prevention. The power of the xShare yellow button lies in its ability to unlock data for an unlimited number of use cases, whether they are already documented or not. (An exemplative prospective use case is provided in D2.2: "EEHRxF in a nutshell).
This working paper aims to provide an in-depth exploration of the current state of play. While its key focus is health threats monitoring, it is important to clarify that this includes both communicable and non-communicable diseases. Non-communicable diseases (NCD), which are the most data-intensive, represent a major threat to the sustainability of health systems today. To enhance efficiency, both the data time delivery and volume of data at stake need to be considered. For communicable diseases, health threats monitoring requires that data to be provided as close to real-time as possible. In contrast, data related to NCD, although typically more substantial, can often be delivered with some delay without adversely affecting the quality of data analytics.

The EHDS regulation implementation timeline is now becoming more defined. Different timelines are set for specific parts of the Regulation. For example, a target for the patient summary and e-prescription/dispensation set for 2029. As shown in the figure below, the timelines for primary and secondary use of data are aligned. This approach will allow the necessary simplifications and alignments on one hand, while also allowing for the exploration and validation of the most promising public and population health use cases that can benefit from a compliant implementation of the EEHRxF specifications.

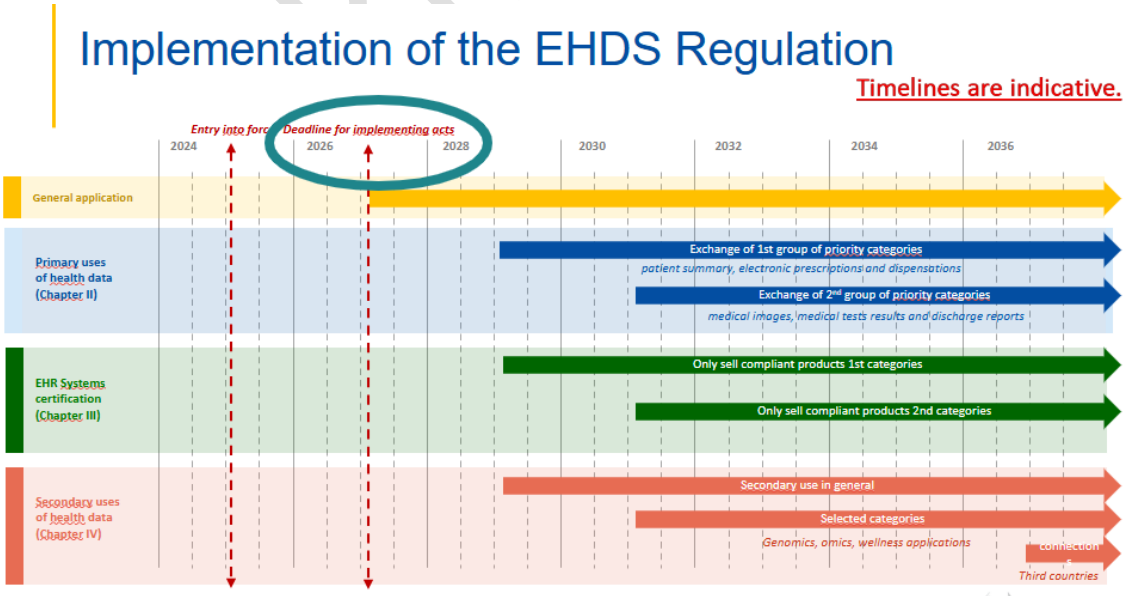


Figure 1: EHDS Implementation Timeline

This Working paper aims to provide a global state-of-play in EU MS, list existing use cases and associated data sets and list the key issues to be addressed by key players to harvest the expected results.

2. Methodology

To obtain a clear picture of the state of state-of-play in European Member States pertaining to the secondary use of data, we leveraged existing streams of information from past and ongoing projects and initiatives, complemented with gathering new information through a designated survey. Notably, the area that has been least explored is the potential connection between primary and secondary use ecosystems. In addition to interoperability issues, the factor of "time" is becoming increasingly important for supporting innovative use cases, such as real-time monitoring and clinical benchmarking.

2.1 Making maximal use of inputs from other projects

This Working paper provides a specific repository cataloguing past and current EU and national projects that could help us describe the current state of play. For each project, we have identified publicly available resources that can serve as potential inputs. In some cases, such as ongoing Joint Actions, due to the non-public or sensitive nature of the information, it has been agreed that the information will not be disclosed directly. Instead, it will be integrated into the broader analysis, with the priority use cases identified being included in the xShare public and population health business use cases repository.

2.2 Collecting direct information from stakeholders

Given the survey fatigue among EU projects, we created a new survey focusing solely on collecting information not yet available. The questions were carefully curated, and the survey was linked to existing sources of information. However, due to the novelty of our approach, we encountered a significant separation between the actors involved in primary and secondary data use. In many Member States, several of the questions raised had not yet been formally addressed. Despite diversifying our channels of diffusion, it remained sometimes challenging to reach individuals who could provide meaningful inputs. The limited response rate could also be interpreted as an indicator of the absence of relevant initiatives at this stage.

2.3 Identifying best practices

To identify countries where best practices could be found and documented, we employed both desk research and the extended network of xShare partners. Priority was given to initiatives in countries with a proven history of implementation, ensuring the collection of user feedback and consolidation of lessons learned. This approach facilitated the identification of the main initial drivers for successful implementation and the listing of main overarching objectives.

2.4 Link with other working papers

This working paper primarily focuses on describing the current situation regarding the inventory of consolidation actions to improve secondary use of data, as well as identifying alignment initiatives between primary and secondary use, including current use cases and best practices. In the context of WP2, an initial prospective use case has also been documented. Another working paper describes the xShare use cases to be tested with the dashboard, identify priority datasets that could be connected to the EEHRxF, and describe the most promising use cases, including the relevant datasets and their links with the EEHRxF. The requirements and use cases will be transferred to WP2 to feed the respective "x-bundles" and corresponding asset repositories. The four best practice reports documented in this working paper will also be utilised in WP7, serving as educational and training resources.

3. Inputs from selected papers

This section aims at summarising pertinent content originating from European documents of specific strategic importance, focusing on the priority public health topics listed in the Description of Action under table 3.1b.

3.1 Study on the barriers to effective development and implementation of national policies on antimicrobial resistance

The document "Study on the barriers to effective development and implementation of national policies on antimicrobial resistance"[6] published in 2023, coordinated by TETRA TECH and funded under the EU4Health programme highlights various strategies and core data collection requirements to improve data collection for antimicrobial resistance (AMR) management in clinical practice. AMR has been also chosen as one of the ***xShare yellow button prospective use case*** described in the context of WP2.

1. **Core Data and Variables**: Key data points essential for effective AMR management include:
    - **Antimicrobial Consumption**: Data on antibiotic use in **both hospital and community** settings, as well as specific consumption levels within healthcare facilities.
    - **Microbial Resistance Patterns**: Collection of data on resistance patterns ***for specific pathogens*** across different regions and healthcare facilities, supporting informed antibiotic choices.
    - **Infection Control Indicators**: Data on infection prevention and control (IPC) measures, such as hand hygiene and use of protective equipment, especially in healthcare-associated infections.
    - **Clinical Outcomes**: Information ***on treatment outcomes*** for patients with drug-resistant infections to assess the effectiveness of interventions.
    - **Diagnostic Data**: Use of diagnostics, including rapid point-of-care tests, helps refine antimicrobial prescriptions based on actual infection aetiology.
2. **Surveillance Systems**: The study suggests strengthening surveillance mechanisms, integrating data from both human and animal health sectors as part of the One Health approach. This also includes tracking AMR patterns and antimicrobial consumption trends across various healthcare settings.
3. **Data Collection Improvements**: Several barriers were identified, such as limited data coverage in specific settings (e.g., long-term care facilities), heterogeneous data collection standards across regions, and inadequate resources for surveillance. Addressing these gaps requires standardising data collection methods, enhancing interoperability between data systems, and expanding resources for training healthcare staff in data handling and AMR awareness.

[6] https://op.europa.eu/en/publication-detail/-/publication/8e6eeab8-4ba3-11ee-9220-01aa75ed71a1/language-en

Although not all above listed necessary data can be connected to the EEHRxF under development, one can immediately see the benefit of aligning core data specifications with those available in the EEHRxF.

These elements are crucial for a cohesive approach to AMR management, enabling the effective implementation of antimicrobial stewardship (AMS) programs and monitoring their impact on AMR trends.

The study identified **several EU countries with adapted data collection practices and AMR management systems** that serve as good practices:

1. **Denmark**: The **Danish Integrated Antimicrobial Resistance Monitoring and Research Programme (DANMAP)**[7] is a well-established model for One Health surveillance. It monitors antimicrobial use and resistance in humans, animals, and food sources, focusing on bacteria that are human and animal pathogens, zoonotic, and indicator types. Figure 2 provides a description of the DANAP organisation and how DNA sequencing might help better identify pathogenic bacteria. Part of the DANMAP data on veterinary use of antimicrobial agents derives from an IT monitoring programme called VetStat, which was initiated in 2000 by the Danish Government. VetStat collects data on prescribed medicine used in animals

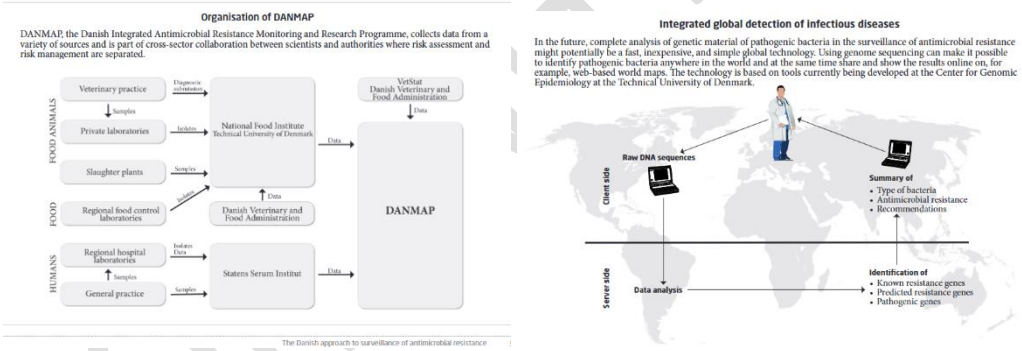


Figure 2: Danish Integrated AMR Programme

2. **Portugal**: The **PPCIRA-IQ quality index** is a tool that benchmarks hospital antimicrobial consumption, resistance, and infection control practices, incentivising compliance through performance-linked indicators. This index is also used to guide interventions and strategies for continuous improvement across healthcare institutions.

3. **France**: The **PROMISE**[8] **project** brings together multiple national networks and academic partners in a One Health framework, aiming to improve AMR data collection and interdisciplinary prevention efforts. When considering the current tasks allocated to the clinical research working group of this project, one can immediately perceive the added value of the EEHRxF and the xShare yellow button. The tasks of this project are outlined below:

[7] https://www.danmap.org/
[8] https://amr-promise.fr/

- To collect data from the different networks on antibiotic use and antibiotic resistance via a simple questionnaire
- To strengthen the conduct/development of clinical trials testing innovative drugs thanks to the epidemiological information from the networks
- To establish an innovative approach for conducting clinical trials on antibiotic resistance
- To improve the impact of antibiotic use and bacterial ecology outcomes related to antibiotic prescribing
- To highlight problems of clinical importance by linking epidemiology and clinical research
- To establish a dedicated clinical infectious disease research trial platform to continuously enrol patients in a study protocol targeting antibiotic resistance.

4. **Czechia**: The country has implemented innovative software to digitise veterinary records as part of its National Action Plan. This allows for streamlined data processing and sharing, enhancing AMR data analysis across veterinary sectors. Czechia also conducts annual evaluations of its National AMR Monitoring Programme, with a focus on AMR-resistant bacterial strains in the environment.

These countries have demonstrated commitment to advancing AMR surveillance and data collection, addressing key barriers, and integrating innovative tools to support AMR management. They can be considered good candidates to develop and validate the use cases and specifications in relationship with the EEHRxF.

3.2 Europe's Beating Cancer Plan and EU Mission on Cancer

Europe's Beating Cancer Plan[9] released in 2021 is a Communication from the European Commission to the European Parliament and the Council. It provides several recommendations to enhance data collection from clinical practice to support cancer knowledge and treatment.

The core data and variables to be collected include:

1. **Patient-Specific Data**:
   - Genetic and genomic information, enabling personalised cancer risk assessments and treatments.
   - Patient-reported outcomes and experiences to improve care and long-term management.
2. **Treatment and Outcomes Data**:
   - Longitudinal data on treatment protocols, including radiology, chemotherapy, and surgery.
   - Data on treatment effectiveness and survival rates, particularly for specific cancer types.
3. **Screening and Diagnosis Data**:
   - Information on cancer staging at diagnosis[10].

[9] https://health.ec.europa.eu/publications/europes-beating-cancer-plan_en

[10] This information is difficult to find because it is not always included in the patient's chart. To locate it, one must look for the original pathology report. This gap will be considered for D.4.2.

- Data from population-based cancer *screening programs* for breast, cervical, and colorectal cancers.

4. **Digital Health and Electronic Health Records (EHR)**:
   - Integration of patient health data into interoperable electronic health records to facilitate efficient sharing among healthcare providers.
5. **Environmental and Socioeconomic Data**:
   - Linkages between cancer incidence and environmental exposures or socioeconomic factors.

The plan emphasises the need for:

- Enhanced interoperability of health data systems across the EU.
- Investments in digital technologies like AI and high-performance computing to analyse large-scale health datasets.
- Development of guidelines for standardised data collection and ethical handling of patient information.

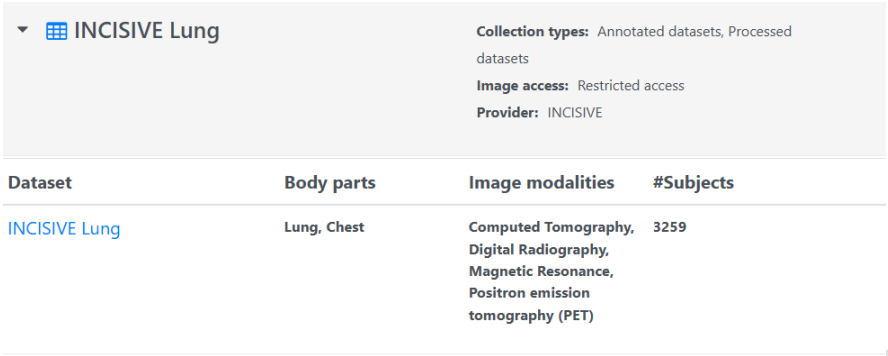
These initiatives are central to enabling personalised cancer treatments, improving early detection, and understanding cancer progression.

Initiatives and Tools for Data Collection at EU level:

- **European Cancer Information System (ECIS)**: Tracks the burden and trends of cancer and offers expanded features for detailed regional data analysis. The information available through ECIS[11] relies on available national datasets, but there is a notable time delay before historical data become available. The most recent update, covering data up to 2019, includes information for periods prior to 2014. Critical indicators such as incidence and mortality aligned with cancer types are typically obtained with a delay of 1-2 years. Additionally, there is a profound diversity between regional and national datasets, highlighting the varying challenges in data availability and consistency.
  If some of the core data elements were mapped to the EEHRxF, the delay could be significantly reduced, and the European Dashboard would be able to address a much broader number of questions. The three main semantic references for this are: ICD-10 (International Classification of diseases), ICD-O-3 (International Classification of Diseases for Oncology) and ICCC (International Classification of Childhood Cancer) (ICCC).

- **European Cancer Imaging Initiative[12] (EUCAIM)** Establishes a repository of anonymised cancer imaging data to support research and improve diagnostics. It is already ideally positioned to integrate the upcoming **European Health Data Space (**EHDS). This large database includes images as the one shown below, often annotated, which can be accessed according to diverse modalities.

[11] https://www.encr.eu/sites/default/files/Data_call/ECIS%20call%20for%20data%20protocol_20240226.pdf
[12] https://dashboard.eucaim.cancerimage.eu/



INCISIVE Lung

Collection types: Annotated datasets, Processed datasets
Image access: Restricted access
Provider: INCISIVE

Dataset	Body parts	Image modalities	#Subjects
INCISIVE Lung	Lung, Chest	Computed Tomography, Digital Radiography, Magnetic Resonance, Positron emission tomography (PET)	3259

The HL7 FHIR and OHDSI-OMOP data models have demonstrated their effectiveness in supporting research tasks within a federated setting, particularly for oncology clinical information and imaging data. Both data models have proven successful in facilitating ML-based oncological studies, clinical predictive modelling, federated learning medical applications and other ML-based analyses. While OMOP-CDM is more suitable as a common data model for storing all information used within EUCAIM, the initiative also supports the use of FHIR resources and messages. This approach supports the transfer of EHR data into the common data model, enhancing interoperability and data integration.
The upcoming EEHRxF imaging report, which will focus on the analysis of images, is an important enabler for establishing a stronger connection between clinical information and imaging data. This will enhance the integration of these data types.

The "Europe's Beating Cancer Plan" identifies several EU countries that have developed good practices in data collection for cancer treatment and knowledge management. Those countries could play a prominent role in developing advanced business use cases relying on EEHRxF standards.

1. **France**:
    - Implements comprehensive cancer data collection through national initiatives like **INCa (Institut National du Cancer)**, and supports research, prevention and patient care. It also promotes personalised medicine through integrating genomic testing into healthcare.
2. **Germany**:
    - Utilises a robust network of regional cancer registries, contributing to a unified national cancer database. This supports research and provides real-time insights into cancer trends and outcomes.
3. **Netherlands**:
    - Home to the **Netherlands Cancer Registry (NKR)**, which systematically collects and analyses cancer incidence, treatment, and survival data. The registry supports national cancer control planning.
4. **Sweden**:
    - Leverages its **Swedish Cancer Registry**, integrated with other national health registers. This enables advanced epidemiological studies and personalised care approaches based on real-world data.
5. **Denmark**:
    - Known for its **Danish Cancer Registry**, which is among the oldest nationwide cancer registries. It collects detailed data on cancer cases, supporting research and policy-making.

EU Mission on Cancer

The EU Mission on Cancer[13] supports the implementation of Europe's Beating Cancer Plan and all aforementioned initiatives. Among others, it aims to generate knowledge and further evidence in cancer understanding, prevention, diagnosis, treatment, and quality of life. It also aims to establish national cancer hubs in Member States and Associated Countries and provide a solid basis and scientific evidence for Europe's Beating Cancer Plan's implementation.

One flagship initiative of the EU Mission on Cancer is the **UNCAN.eu**[14] data exchange platform which will provide a better understanding of the development and progression of cancer.

All the structures and initiatives outlined in the EU cancer strategies would significantly benefit from the ability to use and receive data in the form of the EEHRxF. The extent to which this is necessary for their respective operations varies. For example, the National Cancer Data Nodes foreseen in the EU Mission on Cancer will need to adopt the necessary processes for technical interoperability with the EHDS, incorporating the EEHRxF. On the other hand, UNCAN.eu might only indirectly benefit from EEHRxF by hosting FAIRified data sets from data contributors who already utilise the format, such as Health Data Access Bodies and National Cancer Data Nodes.

To summarise, cancer is a compelling use case for a large-scale public health threat where research relies partly on the efficient transition of data from primary to secondary use contexts. The EEHRxF could have a significant impact by simplifying the data flow from real world to secondary use, provided it is consistently applied by the relevant actors, as outlined in the EHDS regulation.

3.3 Supporting Health system resources management and capacity monitoring

Although oriented towards identifying financial resources to improve health systems, the 2024 WHO Technical Support Instrument (TSI) report[15] highlights several approaches and core data requirements for improving data collection from clinical practice to support health system resource management and capacity monitoring.

The core data and variables to be collected include:

1. **Workforce Data**:
    - Information on workforce capacity, including numbers of healthcare professionals, roles, and regional distribution.
    - Training and skill levels of staff, particularly for new systems like digital infrastructure or triage systems.
2. **Facility Data**:
    - Infrastructure availability, such as the number of hospital beds, clinics, and specialised facilities.

[13] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe/eu-mission-cancer_en
[14] HORIZON-MISS-2024-CANCER-01-01
[15] https://www.eca.europa.eu/en/publications/SR-2024-18

   - Condition and utilisation rates of healthcare facilities.
3. **Service Delivery**:
   - Patient flow metrics, including average wait times and throughput for clinics and hospitals.
   - Data on accessibility of care, focusing on underserved populations and rural areas.
4. **Digital Health Infrastructure**:
   - Levels of adoption and interoperability of electronic health records (EHR) systems.
   - Usage metrics for telehealth and digital patient management tools.
5. **Financial Data**:
   - Budget allocations and expenditures for healthcare resources.
   - Cost-effectiveness of implemented reforms or interventions.
6. **Outcomes Data**:
   - Metrics on patient outcomes linked to healthcare resources, such as recovery rates and hospital readmissions.

The WHO Guidelines on Routine Health Information Systems (RHIS)[16] emphasises capturing consistent data on service utilisation, workforce capacity, and health outcomes also insists on those principles:

- Timeliness: Routine data provides real-time monitoring capabilities.
- Granularity: High detail at the subnational, facility, and community levels.
- Customisation: Flexibility to calculate specific indicators for population or facility needs.

This guideline and the aforementioned report also recommend among other things to:

- ***Leverage primary data from clinical records and secondary data from research and public health surveillance systems***
- **Use structured formats** for collecting patient demographics, healthcare resources, and service efficiency metrics
- Improve Interoperability of digital systems with common data standards to enable ***real-time monitoring*** of capacity and resource utilization
- Equip healthcare facilities with the necessary tools for ***automated data entry and reporting***
- Promote ***skill acquisition*** in data analytics to improve the interpretation and utilization of collected information
- Establish robust governance frameworks to oversee data collection and quality assurance.
- Implement regular audits and validation checks to ensure data reliability and completeness
- Develop dashboards and analytics platforms to visualise and analyse health system performance metrics
- Use predictive analytics to forecast resource needs and capacity constraints and resource allocation
- Encourage collaboration across EU Member States to share best practices and lessons learned

These strategies aim to create resilient health systems capable of dynamically allocating resources and responding effectively to changing demands.

[16] https://www.who.int/data/data-collection-tools/health-service-data/toolkit-for-routine-health-information-system-data/introduction

The **Belgian use case** analysed in this report is particularly relevant for xShare: Belgium wants to develop a digital dashboard for population health management (PHM) to integrate and analyse health data from diverse sources, enabling evidence-based decisions for integrated care. The main objectives are to promote multidisciplinary networks cantered around patients to enhance coordination between care providers but also to create training programs for healthcare professionals on PHM principles and data utilisation to foster localised healthcare improvements. The report highlights that a fragmented data landscape can hinder progress; therefore, the integration of databases will be essential for effective PHM. The report recommends a territorial approach where local population needs and regional collaborations drive implementation efforts.

Aside from Belgium, three other countries have implemented or used PHM approaches:

- **The Netherlands:** Since 2013, nine pilots using different strategies and approaches are implemented, which are using PHM as a cross-sectoral partnership to notably inform value-based payment models that aim at providing better care at a lower cost. All pilots acknowledge the importance of robust and comprehensive data to assess population health status, health costs and the performance in terms of quality of care but regret that data infrastructures for integrated use of PHM approaches are still lacking.
- **The United Kingdom**: Many NHS offices are using PHM approaches for integrated care purposes. During the COVID-19 pandemic, PHM was used to identify individuals who needed more support highlighting furthermore the link between poorer health outcomes with certain social determinants. The programme established aims at training multidisciplinary teams in health care to use data and analysis in their daily decision-making
- **France:** The health insurance provides the general public with a set of data on around fifty pathologies, chronic treatments and episodes of care and provides answers to the following questions: What is the number of patients treated for these different pathologies? How is the prevalence changing? How is the number distributed across France? What are the reimbursed expenses allocated to each of the pathologies identified?

Although it is clear that the EEHRxF will not provide all the data needed to feed the existing and forward-looking dashboards, it will be instrumental for providing data where time plays an important role. Furthermore, the capacity offered by the xShare yellow button will offer the opportunity to provide data produced by the patient such as PREMs and PROMs in direct connection with the core data of the EEHRxF. New use cases and data categories should be integrated on the format on demand.

3.4 Pharmaceutical Strategy for Europe

The Pharmaceutical Strategy for Europe[17] is a communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions released in 2020.

It emphasises that "We need to **break silos** so that various public authorities responsible for authorisation, health technology assessment, healthcare provision, health insurance and financing, work together. Increased cooperation in scientific advice and convergence on **key concepts, such as**

[17] https://health.ec.europa.eu/medicinal-products/pharmaceutical-strategy-europe_en

'unmet medical need', will facilitate the design of clinical trials, generation of evidence and assessment, ensuring that innovation matches the needs of patients and of the national health systems". While not explicitly stated, this clearly implies the necessity of establishing direct connections between regulatory, clinical and pharmacovigilance domains, encompassing the global value chain and extending beyond public bodies alone.

The core data and variables to be collected from clinical practice and patients to support the Pharmaceutical Strategy for Europe include:

1. Patient-Centered Data

- Access and Equity Metrics:
  - Geographic and demographic access to medicines.
  - Differences in patient access across Member States, focusing on vulnerable groups.
- Outcomes Data:
  - Long-term **treatment outcomes**, including recovery and survival rates.
  - Safety and efficacy of innovative medicines.

2. **Medicine and Therapy Data**

- Medicine Use and Availability:
  - **Usage rates of innovative and generic medicines**.
  - Monitoring the **availability** of essential and niche medicines.
  - **Data on medicine shortages**
- Affordability Metrics:
  - Costs of medicines and their impact on patient access and health system sustainability.
- Clinical Evidence:
  - Data from clinical trials and **real-world usage for** regulatory assessments[18].

However, other types of data such as **Supply Chain and Infrastructure Data, Research and Innovation Data, Environmental and Sustainability Data** need to be obtained from different sources.

To generate the necessary outcome data, the communication mentions the following key measures:

- Data Interoperability: Encourage harmonisation of data standards to ensure compatibility across national and EU-wide systems, leveraging artificial intelligence and big data analytics for evidence generation, drug discovery and decision-making:
  - Integration of electronic health records with health system metrics.
  - Access to federated and interoperable health data infrastructures.
- Developing standardised frameworks for health data reporting across the EU.
- Promoting real-world evidence generation to complement clinical trial data.
- Supporting federated access to genomic and health data for personalised medicine
- Revising pharmaceutical legislation to address technical requirements for digital transformation and interoperability in health data

[18] This supports connecting CDISC standards, OMOP standards, and FHIR standards. This approach allows clinical trial data to be combined with real-world data once the drug is approved and widely used. This will help with post-market drug safety surveillance.

- Implementing transparency mechanisms for pharmaceutical data, including electronic product information (ePI), to enhance patient and professional access across multilingual environments

From the above, one may conclude that the Medicinal Products ecosystem likely offers one of the most promising opportunities to connect primary and secondary use of data. Medicines constitute a core data element in 4 out of the 6 EEHRxF identified domains (e-prescription, e-dispensation, patient summary and hospital discharge report). Although initiatives are not yet fully connected (see DARWIN and UNICOM in next section), original interoperability efforts are currently being developed to enable this connection in the near future.

4. Inputs from relevant projects and initiatives

In this section we aim at identifying the projects and initiatives which can provide inputs, assets or use cases for xShare. Some of those projects have already come to an end while others are ongoing. We have thus divided this section in two parts: the first one referring to the former and the second one to the latter.

4.1 Inputs from previous initiatives and projects

4.1.1 Joint Action Towards the European Health Data Space- TEHDAS 1 (2021-2023)

The TEHDAS Joint Action[19] Health Data Space by developing principles for the cross-border secondary use of health data. Its Working paper 6.2[20] *Recommendations to enhance interoperability within HealthData@EU- a framework for semantic, technical and organisational interoperability* provides a few recommendations on semantic interoperability:

- While none of the existing taxonomies can cover all the health data types of interest in the HealthData@EU (the EHDS-infrastructure for secondary use of health data), data preparatory institutions (i.e. those acting as data holders) can reliably use them as a semantic layer when standardising their data. Among the taxonomies, SNOMED CT has proven to be the best equipped ontology for achieving semantic interoperability across controlled vocabularies and taxonomies related to medical concepts.
- As the current mapping of medical concepts from taxonomies and controlled vocabularies to SNOMED CT is not fully completed, it is recommended the European Commission fosters this effort and for Member States to progressively deploy SNOMED CT as an ontology of reference for medical concepts.
- There will still be a need to develop and share semantic maps beyond medical concepts, covering areas such as determinants of health (i.e. social, cultural, economic, environmental and genetic determinants). As these concepts are often instrumental to specific uses or research projects, it is recommended that data holders enrich their data collections with these maps and systematically share with other data holders within HealthData@EU.
- Preparing data for secondary use should go beyond concept mapping and involve the development of comprehensive data models accounting for individuals, place of residence, place of treatment, contacts with the system, treatments, and time. As for this purpose, it is recommended that the European Commission and Member States design and implement a dedicated development initiative. This should draw inspiration from the approach taken by the initiative promoting OMOP-CDM, which has successfully addressed principles of openness, transparency, technological neutrality and data portability and cooperation among public institutions.

Finally, the country profiles published by TEHDAS can help xShare in creating an overview of the current coding systems in use. TEHDAS carried out a mapping exercise among the participating member states and associated countries, to engage with national stakeholders and provide an

[19] ttps://tehdas.eu/tehdas1/
[20] https://tehdas.eu/app/uploads/2023/10/tehdas-recommendations-to-enhance-interoperability.pdf

overview of the state-of-play of the national health data management developments in relation to the European Health Data Space. The results are published in the shape of one-pagers[21].

4.1.2 HealthData@EU Pilot project (2022-2024)

The HealthData@EU[22] Pilot project aimed to build a pilot version of the European Health Data Space (EHDS) infrastructure for the secondary use of health data "HealthData@EU".

One work package focussed on standardising the descriptive metadata templates used to present and describe the available data collections in every node. The standard for this common descriptive metadata model is based on a health DCAT-AP[23] extension: HealthDCAT-AP[24]. This standardisation is crucial for the EHDS framework, enabling effective health data sharing across Europe. HealthDCAT-AP is specifically designed to describe health datasets and dataset access services, ensuring they are consistently represented and easily discoverable across various platforms. HealthDCAT-AP extends DCAT-AP by introducing two additional properties, hasCodingSystem and hasCodeValues with which coding systems in use (ex: ICD-10-CM, DGRs, SNOMED CT, ...) can be indicated. Using these properties, in the future EU dataset catalogue it will be fairly easy to get a quick overview of the coding systems in use for each EHDS Article 51 dataset.

In order to illustrate the feasibility and the potential of reusing data from several European countries, the project included use cases. Based on these, the HealthData@EU Pilot is working on building recommendations for the European Commission about data standards, legal requirements, costs and economic model needed to scale up the tested network. The five use cases are:

- Demonstrate the feasibility of using the EHDS to carry out infectious disease surveillance, focusing on antimicrobial resistance
- Foster a better understanding of the risk of thrombosis in COVID-19 patients
- Compare COVID-19 testing, vaccination and hospitalisation between the general population and vulnerable subpopulations
- Compare care pathways for cardiometabolic diseases in European countries and build prediction models, using artificial intelligence
- Mobilise and chain clinical and genomic data to enhance our understanding of colorectal cancer

The EEHRxF could contribute to enhancing the EHDS2 infrastructure by facilitating data transfer for the HealthData@EU use cases and beyond.

4.1.3 HealthyCloud (2020 – 2023)

The objective of the project HealthyCloud[25] was to generate a number of guidelines, recommendations and specifications that will enable distributed health research across Europe in the form of a Ready-

[21] https://tehdas.eu/tehdas1/packages/package-4-outreach-engagement-and-sustainability/tehdas-country-visits/
[22] https://ehds2pilot.eu/)
[23] https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe
[24] https://healthdcat-ap.github.io/
[25] https://healthycloud.eu/

to-implement Roadmap. This roadmap together with the feedback gathered from a broad range of stakeholders formed the basis to produce the final HealthyCloud Strategic Agenda for the European Health Research and Innovation Cloud (HRIC)[26].

One of the aimed services concerned metadata standards and data interoperability guidance. With regards to the interoperability, the Strategic Agenda lists the following:

"To use data from different collections and potentially link them at individual level, it is important to have them structured using the same standards (data models). Otherwise, data needs to be mapped to a common data model. Therefore, it is firstly important to **structure data at source**, at collection level and this support service would encourage the use of commonly recognisable standards at international level. Overall, this service aims to ensure the use of commonly recognised international metadata and data standards to facilitate findability and interoperability of datasets."

However, this comes with challenges, such as defining responsibilities of service provision and adoption and the possible risk of duplication of this service with existing initiatives (mandate provision). Although Research Infrastructures could be mobilised for such an endeavour, as they have the existing infrastructure and staff available, the experience, domain specificity and stable funding, are covering only part of the full health domain. Nonetheless, the Research Infrastructures listed in this section might aid the xShare efforts to reflect on strategies to identify and map existing data registries and to align data models and terminologies.

4.1.4 B1MG (2020-2023)

Genomic medicine holds significant potential to shed light on how genetic variation influences health, prevention strategies, and treatment responses. Launched in 2020, the Beyond 1 Million Genomes (B1MG) project aimed to establish a federated genomic data network across Europe, with Belgium participating as a pilot country. B1MG developed the Maturity Level Model (MLM), a framework for countries to self-assess the maturity of their genomic medicine practices. The MLM included 49 indicators across eight domains:

1. Governance and strategy
2. Investment and economic model
3. Ethics, legislation, and policy
4. Public awareness and acceptance
5. Workforce skills and organization
6. Clinical organization, infrastructure, and tools
7. Clinical genomics guidelines and infrastructure
8. Data management, standards, and infrastructure

The ongoing Genomic Data Infrastructure (GDI) project builds on the experiences of B1MG piloting countries and their MLM assessments.

As genomic data becomes increasingly significant in public health prevention strategies, it is essential for xShare to remain informed about the specific requirements and standards in this field.

[26] https://zenodo.org/records/8389643

4.1.5 GECCO dataset

The GECCO dataset[27] was published with its HL7 FHIR data model on Simplifier in May 2020 (https://doi.org/10.1186/s12911-020-01374-w). Developed as a German consensus dataset, it was designed to facilitate the consolidation and processing of study-related data for further use, serving as an important tool for German national and international projects such as NAPKON, CODEX, COMPASS, and ORCHESTRA. The GECCO dataset is based on ISARIC protocols.

The parameters captured using GECCO include etiological, diagnostic, and therapeutic data, as well as study-specific information such as certain inclusion criteria.

As part of the NAPKON project, an expanded dataset called GECCOplus was developed. Its purpose was to identify the common data elements across the three independently formed NAPKON cohorts (Cross-Sector Population, High-Resolution Cohort, and Population-Based Cohort).

Figure 3: the GECCO dataset [28]

The GECCO data set offers specifications for COVID-19- related information based on HL7 FHIR, SNOMED CT and LOINC. These could be re-used in xShare when defining specifications for use cases related to infectious diseases.

4.1.6 ORCHESTRA (2020-2024)

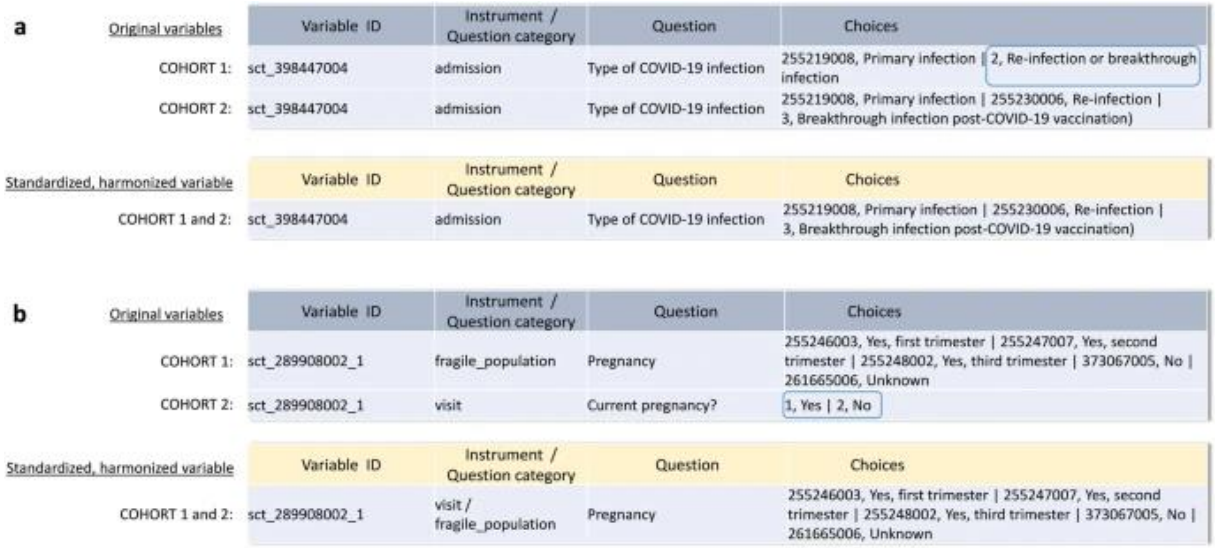
The project ORCHESTRA[29] was launched at the end of 2020 and supported the creation of a pan-European cohort of COVID-19 patients, involving 26 partners. However, differences in research protocols, languages, data structures, and definitions posed significant challenges to data integration.

[27] https://art-decor.org/art-decor/decor-project--covid19f-

[28] Image source: https://www.bihealth.org/en/research/scientific-infrastructure/core-units/interoperability/home/translate-to-englisch-gecco)

[29] https://orchestra-cohort.eu/

The adoption of terminology standards such as SNOMED CT, LOINC, ICD, ATC, NCIt was crucial to maximizing the success of observational research. Variables were harmonised whenever possible for ORCHESTRA perspective studies and international codes assigned to the variables. For example, as shown in Fig 4, for variables enquiring the type of COVID infection or the pregnancy status, SNOMED codes were assigned both for the question and for the value set. Additionally, efforts were made to harmonise the value sets across the different cohorts to maximise precision and interoperability.



a

Original variables	Variable ID	Instrument / Question category	Question	Choices
COHORT 1:	sct_398447004	admission	Type of COVID-19 infection	255219008, Primary infection | 2, Re-infection or breakthrough infection
COHORT 2:	sct_398447004	admission	Type of COVID-19 infection	255219008, Primary infection | 255230006, Re-infection | 3, Breakthrough infection post-COVID-19 vaccination)

Standardized, harmonized variable	Variable ID	Instrument / Question category	Question	Choices
COHORT 1 and 2:	sct_398447004	admission	Type of COVID-19 infection	255219008, Primary infection | 255230006, Re-infection | 3, Breakthrough infection post-COVID-19 vaccination)

b

Original variables	Variable ID	Instrument / Question category	Question	Choices
COHORT 1:	sct_289908002_1	fragile_population	Pregnancy	255246003, Yes, first trimester | 255247007, Yes, second trimester | 255248002, Yes, third trimester | 373067005, No | 261665006, Unknown
COHORT 2:	sct_289908002_1	visit	Current pregnancy?	1, Yes | 2, No

Standardized, harmonized variable	Variable ID	Instrument / Question category	Question	Choices
COHORT 1 and 2:	sct_289908002_1	visit / fragile_population	Pregnancy	255246003, Yes, first trimester | 255247007, Yes, second trimester | 255248002, Yes, third trimester | 373067005, No | 261665006, Unknown

Figure 4: Harmonization of COVID-19 variables and identification of concepts with SNOMED codes[30]

Thanks to this effort, semantic interoperability across different studies was established.

The metadata for the ORCHESTRA datasets concerning the four main patient cohorts (general population, vulnerable population groups, healthcare workers, individuals with post-COVID conditions) are available on the ART DECOR platform (https://art-decor.org/art-decor/decor-project--orch-), along with their standardised terminology mappings.

ORCHESTRA COVID-19 data is accessible through the Data Portal (https://orchestra-cohort.eu/data-portal/). The portal offers open access for exploring the metadata used in ORCHESTRA and for obtaining information on bio banked samples. Additionally, it provides controlled access to pseudonymised data via Question-Oriented Data Export and Federated Data Analysis. Applications for data access can be submitted directly on the ORCHESTRA website at https://dataportal.orchestra-cohort.eu/data_access/access_form. Datasets The dataset and its terminology associations might be useful for providing specifications in use cases concerning infectious diseases

4.1.7 UNICOM (2020-2024)

UNICOM[31] was a major innovation and implementation project which ended in 2024. It aimed at supporting a wide and consistent implementation of the IDMP® suite of standards across the whole value chain which will allow a univocal identification of Medicinal Products. The project centered on conversion of key regulatory and clinical processes to use IDMP. These information value chains must

[30] Image source: https://www.nature.com/articles/s41746-022-00620-x

[31] https://unicom-project.eu/

be converted over their full length from data input to data repositories to data usage. The univocal identification of medicinal products relies on the standardised representation of substance, dose and strength. 19 countries were represented, including 26 national Drug and eHealth Agencies and all major Standard Development Organisations (SDOs)[32]. A wide IDMP implementation would guarantee interoperability by default all over Europe thanks to the compulsory use of official interconnected Medicinal Products Databases by all processes which make use of Medicinal Products. It also brings the promise of a very significant simplification by avoiding duplication of data across systems. The project achieved significant milestones, including the launch of an IDMP compliant registration platform (starting with variations) by the EMA to be used by industry, the completion of the EU substances database along with a collaborative platform to manage its evolution, the adaptation (or total renewal) of National Agencies Databases to ensure IDMP compliance, and the initial implementation within clinical domains. It has also provided a strong push to use the HL7 FHIR® standard to support exchange of data between the different actors. A first HL7 FHIR® implementation guide was also released. Finally, the first IDMP compliant implementations of cross-border substitution dispensation has been tested and the e-prescription and e-dispensation specifications of MyHealth@EU have been updated. The main results of this strategic project are summarised in this white paper.

Like xShare, UNICOM was one of the very few projects which considered the whole ecosystem and all the processes which make use of medicinal products. The IDMP suite of standards is complex and does not specify the terminologies and classifications which can be supported. Within the EU, the identifiers of substances are now a reality and are already mapped to major terminologies while a wide consensus exists to use EDQM and UCUM for dose forms. A final decision still needs to be made concerning strength although the issue is now well documented. Although a wide and full implementation of the IDMP® standards will still take time, the dynamic has now been launched with quick progress expected in several countries. The critical identifiers will need to become part of the EEHRxF specifications. Furthermore, the public availability of important assets such as the EU Substances Database already allows the creation of the basis of an integrated ecosystem and supports population health business cases such as the one documented under D.2.2- EEHRxF in a nutshell.

4.2 Connections with ongoing initiatives

4.2.1 PHIRI (2020 – current)

PHIRI[33], the Population Health Information Research Infrastructure is a European health research infrastructure that aims to facilitate and generate the best available evidence for research on health and well-being of populations. Already during its predecessor, the Joint Action for Health Information (InfAct)[34], it started with the developments of the Health Information Portal, www.healthinformationportal.eu[35]. The aim of the Health Information Portal is to provide access to

[32] See for the list: https://unicom-project.eu/partners?_sft_medicenter_gallery_category=sdos
[33] www.phiri.eu
[34] www.inf-act.eu)
[35] Tolonen H, Saso M, Unim B, Palmieri L, Schutte N, Peyroteo M, Lapão LV, Habl C, Bogaert P; Population Health Information Research Infrastructure and the Health Information Portal. European Health Information Portal: a one-stop shop for health information. Eur J Public Health. 2024 Jul 1;34(Supplement_1):i29-i34. doi: 10.1093/eurpub/ckad172. PMID: 38946446; PMCID: PMC11215314.

population health and healthcare data across Europe. The portal is the gateway for researchers and policy makers to make use of the services of the Research Infrastructure on Population Health Information and its coordinator, the Innovation in Health Information Systems unit at Sciensano. The Health Information Portal hosts a metadata catalogue that includes over 300 data sources in different categories: survey/interview data (e.g. health examination/interview surveys), administrative data (e.g. hospital discharge data), population data (e.g. causes of death), registry data (e.g. cancer registries), outpatient utilization data (e.g. morbidity data), social health insurance data, surveillance data of infectious diseases, biobank/sample/specimen data (e.g. biobanks), calculations (e.g. cancer statistics, life expectancy), hospital resources and healthcare administrative area resources (e.g. register of entities performing medicinal activities), hospital resources and healthcare resources (e.g. register of pharmacies), hospitalization statistics of the portals of the national health system (e.g. hospital waiting lists), customer record data (e.g. food consumption data), observational study data (e.g. cohort studies), multiple sources and other records. This rich catalogue serves as part of the input data of the analysis of identifying the core public health datasets in Europe.

PHIRI has developed a federated architecture and PHIRI analytical pipeline – that includes the development of a common data model – allowing the orchestration of the research question throughout a workflow that ensures legal, organisational, semantic and technological interoperability. Research questions coming from multiple European projects have been addressed using the PHIRI methodology for federated analysis which allows sensitive data to be mobilised to respond to multiple research queries in multiple sites, while preserving GDPR principles.

The uses cases focus[36] on:

- Vulnerable populations, inequalities and risk factors with direct or indirect impact on COVID-19 outcomes
- Delayed care in cancer patients
- Effects of the COVID-19 pandemic on maternal and newborn health
- COVID-19 related changes in population mental health
- COVID-19 vaccine(s) effectiveness in preventing SARS-CoV-2 infection
- Cross country population COVID-19 metrics

Standardization of data on the side of data collection (primary use) can speed up the development of the common data models, resulting in timely research evidence to underpin policy decisions. Where xShare stimulates the 'only once' practices, this could aid in reducing the time spent on data mobilisation and harmonisation. An example is long COVID, which is a 'hot topic' for quaternary prevention.

Finally, PHIRI has developed public health foresight activities[37] to provide guidance in identifying promising policy strategies and translating the information into knowledge. A comprehensive view of different health indictors is needed for high quality foresight activities to detect general health trends in a population. The EEHRxF specification will provide a promising opportunity to connect the clinical domain with real-world data.

[36] See https://www.healthinformationportal.eu/services/federated-demonstrators.
[37] https://www.phiri.eu/wp9

4.2.2 ELIXIR

ELIXIR[38] is a European research infrastructure for life science data, **integrating national bioinformatics resources** into a single platform. ELIXIR coordinates and develops life science resources across Europe so that **researchers can more easily find, analyse and share data, exchange expertise, and implement best practices**. ELIXIR is active in several large European Union grants, which support the work of the Platforms and Communities. ELIXIR aims specifically at:

- Developing ways that researchers across Europe can access, store, transfer and analyse large amounts of life science data. See the **Compute Platform**[39].
- Identifying key data resources across Europe and support the linkages between data and literature e.g. by making it easier to move from a scientific paper to the dataset that the paper was based on. See the **Data Platform**[40].
- Providing ways for researchers to find the best software to analyse their data. See the **Tools Platform.**[41]
- Establishing Europe-wide standards that can be used to describe life science data. This makes different data sets easier to compare and analyse. See the **Interoperability Platform**[42].
- Helping scientists and developers find the training they need and also provide that training. See the **Training Platform**[43].
- Developing communities, standards, databases and tools in selected life science domains (e.g. Marine Metagenomics, Human Data). See **Communities**[44].

ELIXIR has launched eight projects addressing the scientific and technological ambitions of its new Scientific Programme for 2024 to 2028. Three projects are content related and focus on cellular and molecular research; biodiversity, food security and pathogens; and Human data and translational research. The remaining five projects address ELIXIR's five technical priorities in **compute, tools, interoperability, data and training**. Each project connects different ELIXIR Nodes to achieve the technical aims of the ELIXIR's strategy in research data management and data sharing, reproducible analytics and infrastructure and federated service delivery.

[38] https://elixir-europe.org/platforms
[39] https://elixir-europe.org/platforms/compute
[40] https://www.elixir-europe.org/platforms/data
[41] https://elixir-europe.org/platforms/tools
[42] https://elixir-europe.org/platforms/interoperability
[43] https://elixir-europe.org/platforms/training
[44] https://elixir-europe.org/communities/

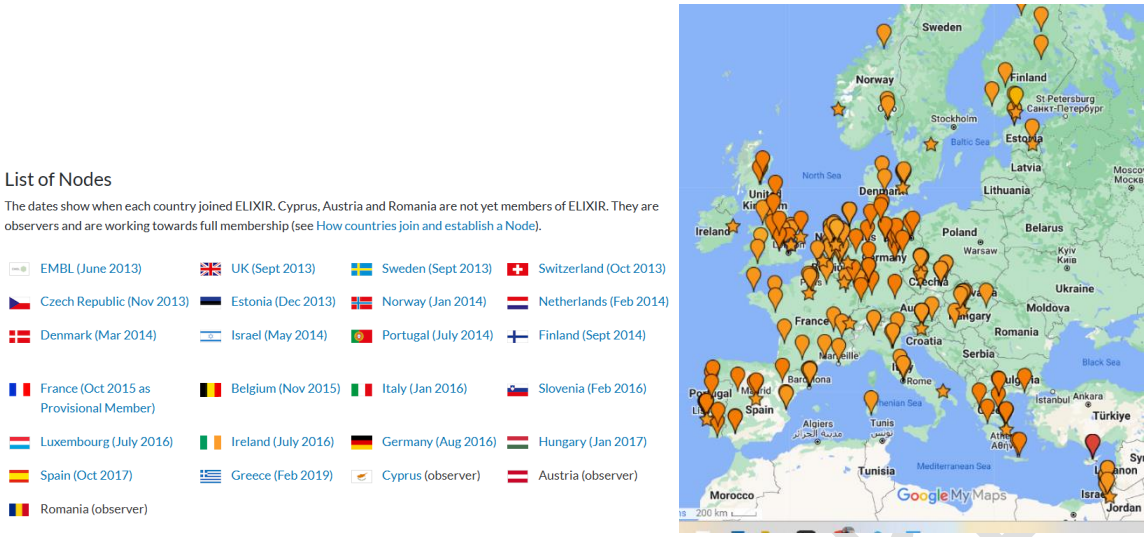


Figure 5: Nodes of the ELIXIR platform

Of particular interest for xShare is the work of the Interoperability Platform; its mission has evolved beyond technical and semantic interoperability, responding to changing needs and life sciences standards. It has already created an ecosystem uniting expertise and information and is transitioning to a **sustainable framework for enabling real-world data integration and reuse** by promoting and supporting interoperability, data management and FAIR principles. The strategy which centres on "the 3 Ps": Products, Processes and Practices. **Products entail mapping and coordinating FAIR resources, minimising duplication, and enhancing user experience**. Processes align products with user needs, ELIXIR Nodes, and global partners. Practices convert insights into actionable guidance, empowering users through the connection and use of diverse resources. This is evidently very much aligned with the xShare WP4 work.

The Research Data Management toolkit for Life Sciences[45] also includes a link towards tools and resources of 24 European countries. A FAIR cookbook has also been produced: it offers a combination of guidance, technical, hands-on, background and review types to cover the operation steps of FAIR data management, and **are classified according to the audience type**s, to serve all those involved in the data management life cycle.

[45] https://rdmkit.elixir-europe.org/

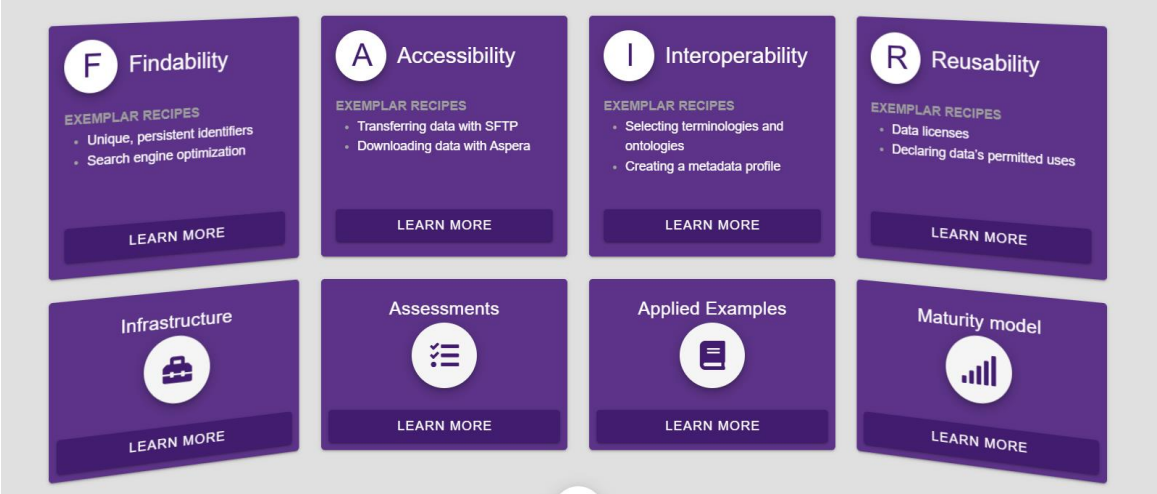


Figure 6: Services offered by the ELIXIR platform

4.2.3 BBMRI

BBMRI-ERIC[46] (Biobanking and BioMolecular Resources Research Infrastructure - European Research Infrastructure Consortium) is a pan-European research infrastructure connecting biobanks and researchers to facilitate access to biological samples and data for high-quality biomolecular and medical research. Established for an unlimited duration under the ERIC legal framework, it operates on a non-economic basis and is funded by annual membership fees from Member States.

One of its primary objectives is to assist researchers in locating ethically and legally compliant samples and data, ensuring their usability through enhanced visibility and adherence to high-quality standards.

In addition to aiding researchers, BBMRI-ERIC provides crucial support to biobanks by offering guidance on quality development, usability, and compliance with ethical and legal frameworks, including adherence to European regulations. This comprehensive approach ensures that biobanks operate effectively and meet the needs of the research community.

To fulfil these objectives, BBMRI-ERIC delivers a wide range of services tailored to various aspects of biobanking and research. These include IT tools, such as directories to connect biobanks with their users, and resources addressing ethical, legal, and social issues (ELSI), such as FAQs on GDPR compliance. The organization also provides consultancy to enhance quality systems in biobanks and supports biobanking development to strengthen their infrastructure and operations.

BBMRI-ERIC's activities are guided by a clear organisational framework. As part of its mandate, it establishes, operates, and develops a distributed infrastructure of biobanks and biomolecular resources. The organisation's work programme, determined by its Assembly of Members, is designed to address the shared challenges and requirements of the biomedical research community, fostering collaboration and innovation across Europe.

[46] https://www.bbmri-eric.eu/

Within BBMRI-ERIC, the Minimum Information About Biobank Data Sharing (MIABIS) initiative was established to standardise data elements used to describe biobanks, samples, and associated research data. Its primary goal is to enhance interoperability among biobanks, facilitating the sharing of valuable data and samples.

To enable sample and data exchange, MIABIS provides a list of attributes to describe Biobank, Collection, Research Resource, and Network.

Terminology is constantly being enriched by adding components describing dataset types and biobank capabilities, samples, sample donors, and events. This enhancement allows for more advanced querying capabilities. Current developments also include a new component focused on molecular pathology. The MIABIS components and their relationships are shown if Figure 7.

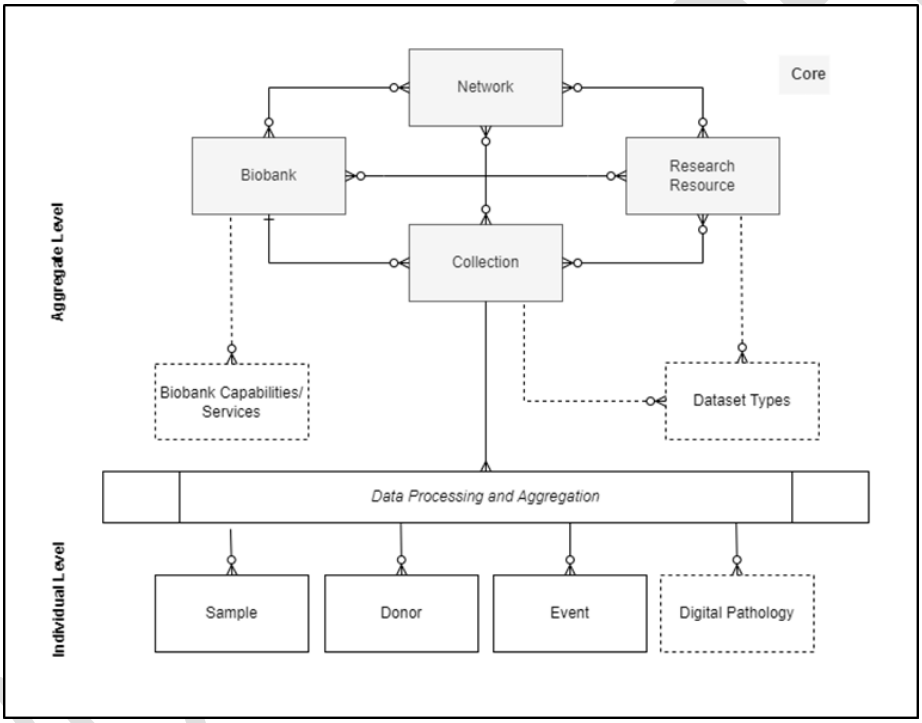


Figure 7: The MIABIS components and their relationships[47]

This project provides important insights concerning what information is needed to describe and exchange biological samples. A HL7 FHIR model is also available on Simplifier which might support the modelling of information for use cases related to infectious diseases.

4.2.4 DARWIN (2022- ongoing)

Real-World Data (RWD) are data relating **to patient health status and/or the delivery of health care routinely collected from a variety of sources,** while Real-World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD

[47] Image source: https://github.com/BBMRI-ERIC/miabis?tab=readme-ov-file

(thus excluding the clinical trials data). DARWIN EU delivers **real-world evidence** from across Europe on diseases, populations and the uses and performance of medicines. It will support non-interventional studies, including developing scientific protocols, integrating relevant data sources and interpreting and reporting study results which is a major interest for public and population health analysts.

The Darwin project was launched in 2022 and entered its operational mode in 2024. It aims at providing a **high-quality, validated real world data** on the uses, safety and efficacy of medicines. At this stage, DARWIN EU will routinely support the evaluation work of the European Medicines Agency (EMA)'s scientific committees and the national competent authorities. Organisations such as the European Centre for Disease Prevention and Control, Health technology assessment bodies and payers may make use of DARWIN EU in the longer term. Currently 20 data partners from 13 different countries including 23 data sources are participating.

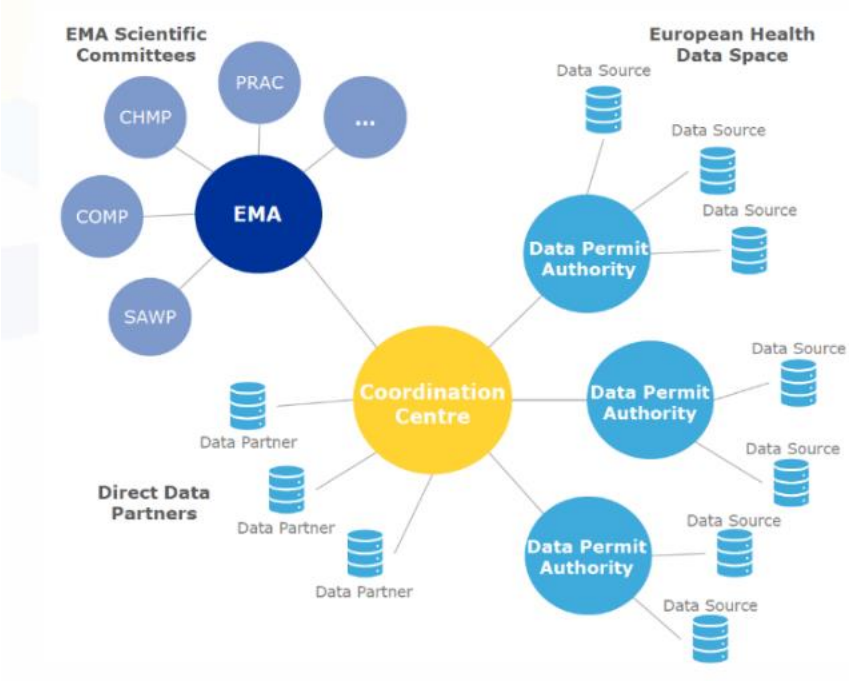


Figure 8: DARWIN ecosystem

The databases include both in and outpatient records originating from clinical practice and claims but also some linked data and biobank databases which also include lifestyle information. The following requirements are defined for candidate data sources:

- Data sources collecting health data routinely and **representative of the different types of real-world** data in terms of data elements, setting (primary & secondary care), population and origin (e.g. electronic health care records, claims)
- Data sources which collectively provide a broad geographical coverage
- Data sources containing patient-level data with a **unique patient identifier** linking all records relating to a given patient
- Medicines prescribed or dispensed identifiable with quantities (e.g. doses, package size) and dates allowing to calculate cumulative doses and duration of use and linked to individual but unidentifiable patients

- **Clinical events formally coded**, with accurate dates and linked to individual but unidentifiable patients
- Data already converted or planned to be converted into a common data model

It is also important to notice that at the minimum annual updates must be guaranteed with a **lag time of less than six months** in data availability.
Data needs to be converted into the Observational Medical Outcomes Partnership terminology (**OMOP**). There is a good alignment in Medications between OMOP and HL7 FHIR® except for medications administered as part of a procedure can be found in the OMOP data but not in HL7 FHIR. There is also good correspondence between OMOP and FHIR cross mapped by LOINC code for lab results. EMA has not reflected on how to integrate IDMP identifiers in DARWIN but this will certainly take place when IDMP implementation begins to be extended in the global value chain in Member States (See UNICOM).

The EEHRxF specifications will offer a unique opportunity to connect the clinical domain with both the clinical trials and real-world data models which have been built separately. In the context of medicinal products, the capacity to better connect all data models seems to be under reach.

4.2.5 The second Joint Action Towards the European Health Data Space (TEHDAS2)

TEHDAS2[48] aims to produce concrete guidelines and technical specifications for the European Commission and member states to ensure a harmonised implementation of the European Health Data Space (EHDS) regulation for secondary use. Member states can use TEHDAS2 results to support their implementation efforts at the national level, while at the same time, the work supports the European Commission, for instance, in the drafting of the implementing acts defined in the EHDS regulation.

Part of the work in TEHDAS – work package 5: data discovery – is dedicated to data discovery. In this work package, guidelines for data holders to fulfil their duties towards data description are being drafted. This work will both fine tune and validate the HealthDCAT-AP (see section on the HealthData@EU Pilot project) and will provide full text definitions of the EHDS Article 51 *Minimum categories of electronic health data for secondary use*. This work package will also develop guidelines for common policies for Health Data Access Bodies (HDABs) on limitations in relation to the purposes that are enlisted for secondary use according to EHDS and technical specifications on the national metadata catalogue.

4.2.6 Health data access bodies – Community of practice

Set up in January 2024, the Community of Practice (HDABs-CoP) brings together all recipients of the Direct Grant for Member State authorities aimed to support the implementation of the EHDS for secondary use by setting up the Health Data Access Bodies in those member states. The European Commission supports the HDABs-CoP by co-chairing the General Assembly, serving as its secretariat as well as solution provider for the HealthData@EU infrastructure central services. Its aims are to support the establishment of HDABs; strengthen cooperation inside and outside the HDABs-CoP; establish a platform for sharing information and aligning strategies; sharing of technical knowledge and expertise; streamline solutions and promote the 3Rs principle – Reduce, Reuse and Recycle; to increase capacity

[48] http://www.tehdas.eu

building; foster harmonisation, standardisation and interoperability; and discuss strategies for disseminating results.

The Community of Practice hosts six different subgroups, focussing on different aspects of EHDS2 and its stakeholder involvement.[49] For xShare, especially subgroup **SG2 on Data Quality and Utility and Health Datasets and Metadata Catalogues**, as well as SG6 Stakeholders Fora are relevant. The former ties in with the EEHRxF's position of bridging primary and secondary use of health data. In contact with SG6, xShare will strive to position itself as a relevant stakeholder to be consulted on the EEHRxF and its associated use cases, as well as standards implementation. SG5 provides an overall picture of the implementation status of each MS. In its last public report, SG5- Deployment and Operations- identified several challenges, including the ability to assess MS needs for operations, lack of participation of MS for a harmonised implementation, and a common EHDS interpretation. Both SGs provide xShare with a regular state-of-play on the implementation of the EHDS for secondary use.

4.2.7 PREVENTNCD Joint Action

Joint Action Prevent Non-Communicable Diseases (JA PreventNCD)[50] is a European project designed to support strategies that aim to reduce the burden of cancer and NCDs. A primary goal of the project is to address the social determinants of health that can predispose individuals to NCDs and build a comprehensive European infrastructure for monitoring factors relating to cancer and other NCDs. To achieve these goals, the project prioritises harmonising methodologies and data sources, alongside improving systems for data collection and analysis.

This commitment to enhancing interoperability and the efficiency of data systems aligns closely with xShare's objectives. Strong collaboration between the projects can improve the quality of NCD and cancer registries, ensuring that data collection processes are effective and allow for data reuse. Collaboration has been initiated between xShare WP4 and PreventNCD WP 8: monitoring. A specific use case of relevance to xShare is PreventNCD's pilot-activity of gathering data with the intention of identifying people at increased risk of cancer using machine learning. WP8 is also exploring synergies between primary and secondary use of data, similarly to xShare.

4.2.8 UNITED4 SURVEILLANCE Joint Action

UNITED4 SURVEILLANCE (U4S)[51] is a joint action involving over 40 partners dedicated to improving Europe's preparedness for potential new infectious outbreaks. Its focus is on integrating existing and emerging data sources to enhance infectious disease surveillance, prevention, and control across the EU/EEA. By strengthening surveillance capabilities within Europe and beyond, U4S aims to contribute to global health security. The primary goal of U4S is to advance national infectious disease surveillance systems by enhancing the integration, interoperability, and digitalisation of data sources. Key areas of focus include outbreak detection, laboratory-based reporting, hospital surveillance, and One Health surveillance. Considering the growing global threat of infectious disease and antimicrobial resistance, U4S will offer an important insight into public health needs.

[49] https://health.ec.europa.eu/document/download/4ef405ce-49f1-41b2-8c48-593ec70573be_en?filename=ehealth_20240125_mi_en.pdf
[50] https://preventncd.eu/
[51] https://united4surveillance.eu/

U4S aims to identify needs, particularly in digitising and integrating data, to support early outbreak detection, improve surveillance systems, and implement a One Health approach. One Health aims to enhance the detection of (re)emerging pathogens with zoonotic potential and support comprehensive public health risk assessments.

U4S shares common objectives with xShare WP4, such as conducting a needs and gaps analysis in public health, promoting interoperability across European countries, and identifying best practices. Contact has been established with the purpose to establish a collaboration between the two initiatives that can enable them to benefit from each other's findings.

The use cases identified by U4S are:

1) Improve algorithms for outbreak detection and pandemic preparedness by enhancing real-time surveillance for better-coordinated responses. Strengthening national surveillance systems will bolster Europe's overall surveillance capabilities

2) Build a foundation for timely, comparable, and representative surveillance of severe infections leading to hospitalization in each Member State

3) Provide recommendations and training for implementing One Health surveillance

In this context, xShare can learn from the needs assessment performed by U4S for the three use cases and contribute by providing technical specifications to advance these efforts.

4.2.9 QUANTUM

While the second Joint Action Towards the European Health Data Space provides the European Commission with input for most of the implementing acts and delegated acts, there is a separate CSA on data quality: QUANTUM[52]. This project aims to create a common label system for Europe that guarantees the quality and utility of datasets for scientific and health innovation purposes. It will do so in 3 steps: first it will reach consensus on which elements describe data sets' quality and utility and the level of data holders' maturity. A thorough literature study was followed by an EU wide Delphi and organised experts' groups to reach consensus (January 2024-September 2024). Currently the project is designing, developing and testing the label system with data holders; a small-scale pilot has just concluded, and another larger pilot will start early next year (step 2). And finally, the QUANTUM label system results are translated into sustainable recommendations for the European Commission.

The data quality label will be part of the metadata record of a dataset. The standard that will be set by the European Commission to describe metadata will be the HealthDCAT-AP, a DCAT application profile that is extended to accommodate health data. The HealthDCAT-AP contains the property 'Quality Certificate', which is defined as annotation that associates a resource (especially, a dataset or a distribution) to another resource (for example, a document) that certifies the resource's quality according to a set of quality assessment rules. It is here where the QUANTUM-data quality label will be linked to the metadata record. Data holders should update (or at least review) their metadata record annually, providing them with an opportunity to also update their quality label (if data quality has been improved in the meantime).

[52] https://quantumproject.eu/

Ensuring that high-quality health data is registered at the source—during the care provision process—is crucial for reliable decision-making, research, and patient outcomes. By standardising data collection at the source, such as through structured formats and interoperable systems, healthcare providers can reduce errors, enhance consistency, and improve the usability of data across systems. This approach not only supports care delivery but also enables seamless data sharing for public health and research purposes

4.2.10 JPIAMR

JPIAMR is an international collaborative platform engaging 29 nations and the European Commission to curb antimicrobial resistance (AMR).

The Joint Programming Initiative on Antimicrobial Resistance (JPIAMR)[53] facilitates international collaboration by coordinating national investments and funding for research aimed at tackling antimicrobial resistance (AMR) worldwide. Its activities span a variety of initiatives tied to the six priority areas outlined in its Strategic Research and Innovation Agenda (SRIA), focusing on identifying and promoting solutions to limit the spread of resistant bacteria. This work is carried out in partnership with 29 member nations and is supported by the European Commission. JPIAMR is one of the few platforms financing research that applies a One Health approach, addressing human, animal, and environmental health collectively. JPIAMR´s objectives include bridging the gap between research and policy by collecting and sharing information and developing global AMR research strategies and programs through alignment of national and international research programs.

4.2.11 One Health approach

One Health is a collaborative, multisectoral, and interdisciplinary approach that aims to achieve optimal health outcomes by acknowledging and addressing the interconnectedness of humans, animals, plants, and the environments they share. The One Health approach is extremely important in preventing infectious disease outbreaks and reducing antimicrobial resistance.

[53] https://www.jpiamr.eu/

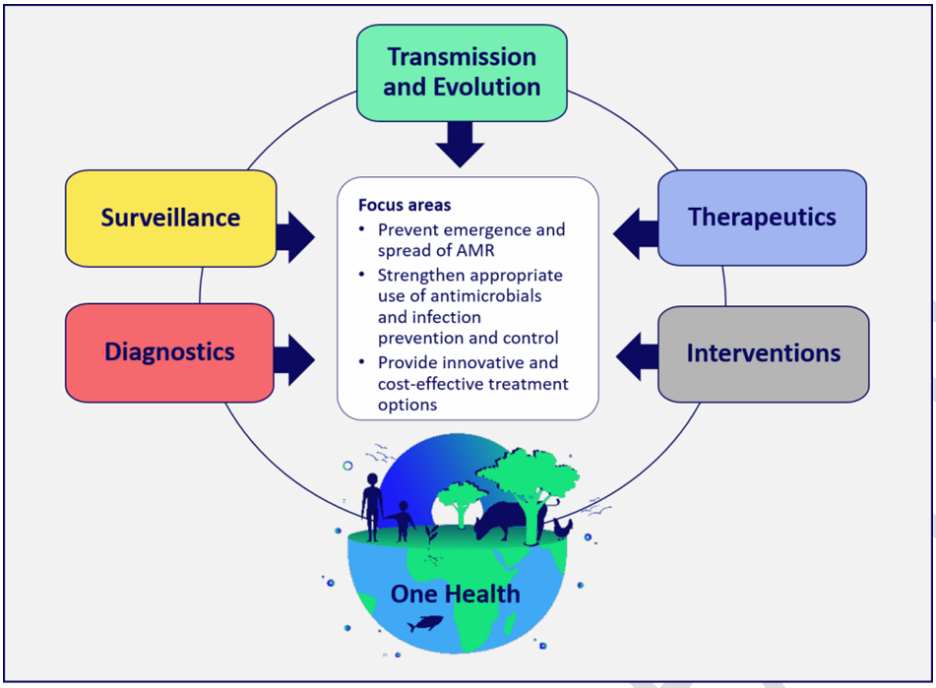


Figure 9: The One Health approach [54]

Building on the foundation laid by JPIAMR, the European Partnership on One Health Antimicrobial Resistance (EUP OHAMR) seeks to enhance coordination and alignment of AMR activities and funding across countries and with the European Commission. It also aims to foster coherence at the national level among various services and ministries responsible for different facets of AMR, such as human health, agriculture, environment, industry, and finance.

EUP OHAMR will address knowledge gaps in AMR, including those highlighted in the European One Health Action Plan against AMR and the EU Strategic Approach to Pharmaceuticals in the Environment. It will also provide support for regulatory science and offer evidence-based insights to inform policymaking. Antimicrobial resistance is a major global threat in public health and xShare efforts to create specifications in this field should be aligned with the requirements identified within this partnership.

4.2.12 PancareSurPass (2021-2025)

PanCareSurPass[55] is a EU funded Research and Innovation action aiming to study the scale-up and implementation of the digital Survivorship Passport to improve people-centred care for childhood cancer survivors. It is the result of a succession of several projects which aim at providing a life-long support to young cancer survivors.

The main reason of inclusion of this project in this review lies in the double objective pursued: conduct a multi-country implementation study of SurPass v2.0 for people-centred care with >5-year CCS and to assess implementation in terms of people-centred care (e.g. activation, empowerment, satisfaction, increase in knowledge), feasibility and cost effectiveness while conducting a parallel observational

[54] https://www.jpiamr.eu/activities/one-health-amr/
[55] https://www.pancaresurpass.eu/

qualitative study using SurPass v2.0 for delivering people-centred survivorship care to <5-year CCS and >5-year CCS (but with limited treatment data available).

The project has created an HL7 FHIR® implementation guide[56] that includes, among other things, the Structured Data Capture specification. This guide provides an infrastructure to standardise the capture and expanded use of patient-level data collected within an EHR. It features two components designed to support more sophisticated questionnaire/form use cases, such as those needed for research, oncology, pathology, and other clinical domains. Additionally, it facilitates the pre-population and auto-population of EHR data into forms and questionnaires for uses outside direct clinical care, including patient safety, adverse event reporting, and public health reporting. This is clearly a use case which needs to be considered, especially because of its cross-border dimension.

4.2.13 Medical Informatics Initiative (2018-current)

The MII is a national German initiative launched by the German Federal Ministry of Education and Research (BMBF) which started in 2018 with the aim to enhance the utility and significance of data derived from healthcare and research. The MII offers a framework that translates research findings into tangible benefits for patients. It includes 37 universities organized into four consortia (DIFUTURE, HiGHmed, MIRACUM, SMITH). Each Consortium focuses on specific concrete use cases for which strategies for shared data use and exchange need to be developed.

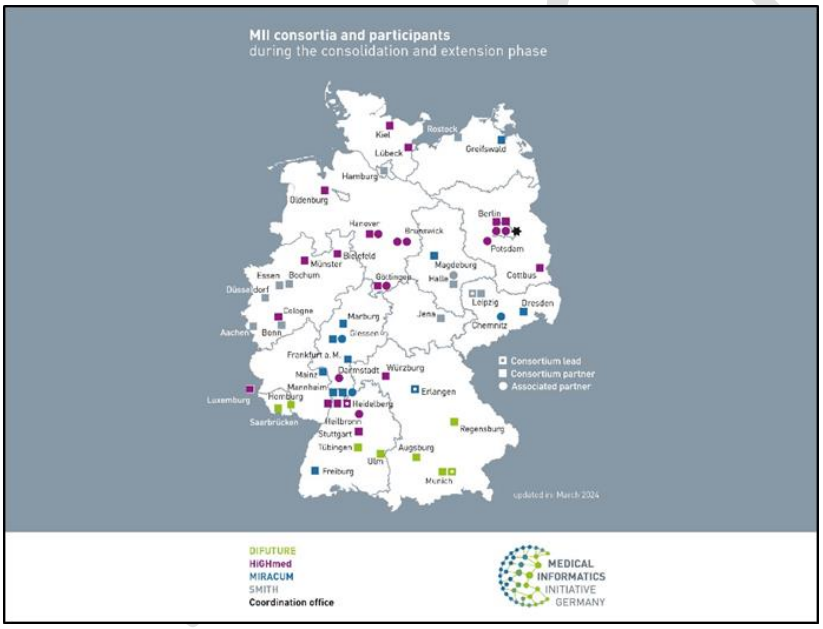


Figure 10: Consortia and participants of the Medical Informatics Initiative [57]

Data Integration centers (DICs)

DICs are hospital-based facilities where healthcare and research data are gathered. They develop the IT protocols necessary for processing medical data from various hospital information systems and ensure it is provided in a standardised format. DICs oversee data access related to projects, along with managing patient consent and pseudonymisation procedures. They also assist in ensuring the privacy-

[56] https://build.fhir.org/ig/hl7-eu/pcsp/

[57] https://www.medizininformatik-initiative.de/en/about-initiative

compliant use of patient data and handle data usage requests for the hospital. Additionally, DICs contribute to the design and execution of research projects, ensuring that data is used in a legally, ethically, and technically optimised manner.

The Core data set

Within the MII, a task force on core data sets is focused on defining a minimum set of patient-related information that all hospitals should be able to exchange. Patient-related information has been divided into several modules, organised into core and extension modules. The core modules are more general, while the extension modules are more specific to particular disciplines. For each module, a set of data elements has been identified, for which format and terminologies need to be provided based on HL7 FHIR R4, SNOMED CT, and LOINC specifications.

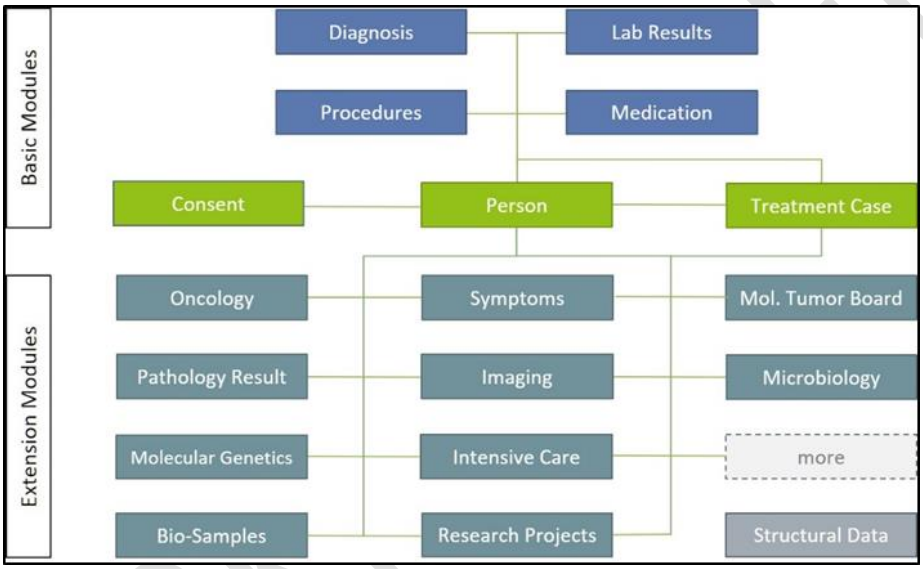


Figure 11: the MII Core Data Set[58]

First version data models were released for all core modules and for some extension modules. Subsequent updates integrate new requirements and possible standard evolution.

4.2.14 Secondary use of data: the German Portal for Medical Research Data (FDPG)

The FDPG provides researchers with the ability to centrally request data and biospecimens from MII sites (DICs) that is compliant to the MII Core Data set specifications. Requests can be made for data based on the MII core data set, which covers a broad range of datasets across various indications. During the testing phase, only scientists from MII sites were allowed to submit feasibility inquiries and data access requests. Starting in May 2023, all researchers can request access to patient data and biospecimens for medical research and submit feasibility inquiries. A positive ethics approval from the researcher's institution for the project is required when submitting an application.

This initiative not only provides an example of HL7 FHIR-based secondary use of data but also offers freely available data models and value sets that can be re-used within the xShare use cases.

[58] Image source: https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set

4.2.15 EU-HIP

For intelligence gathering and threat assessment, DG Health Emergency Preparedness and Response Authority (HERA) needs support from Member States and associated countries, and a comprehensive state-of-the-art IT system generating actionable insights for decision-making is crucial. The upcoming HERA IT platform for intelligence gathering (ATHINA) will only be operational if Member States have strong national IT systems that are interoperable with HERA's IT system and other relevant systems. The EU-HIP project supports participating countries to enhance and improve national IT systems in an efficient and coordinated manner, with the objective to obtain the needed interoperability with HERA's IT platform. To this end, EU-HIP aims to strengthen and align existing IT systems for the assessment of health threats and for intelligence gathering in medical countermeasures at national level. The project is currently conducting health information system assessment for early warning and response. These mappings could aid in xShare's task to identify missing data needed for public authorities (and future health data access bodies) to efficiently monitor and prevent disease outbreaks, health services availability and use, or other high priority issues.

5. WP4 survey

5.1 Objective

One of the goals of WP4 is to develop a thorough and up-to-date overview of public and population health datasets collected across European countries. This includes assessing their potential alignment with the European Electronic Health Record Exchange Format (EEHRxF), as outlined in the European Health Data Space (EHDS) regulation.

Such an overview is important for identifying the priorities and gaps in the public health landscape and understanding the mechanisms currently used to enhance the quality and efficiency of data collection. Additionally, promoting greater technical and semantic interoperability between data collection systems requires an in-depth understanding of the registries and infrastructures used to gather data continuously and "only once" with the intention of reuse. Collecting data "only once" aims to eliminate fragmentation and duplication across multiple systems, thus improving data quality and streamlining processes.

5.2 Methodological approach

5.2.1 Survey Architecture

To gather information from European Member States (MS) on public and population health datasets, as well as their involvement in initiatives aimed at standardising data and enhancing system interoperability, a detailed web survey was created[59].

The survey was developed using LimeSurvey, a platform commonly used by Sciensano. The survey was made available for all European Member States, and welcomed contributions from Albania, Bosnia, Iceland, Norway, Serbia, Switzerland, and the United Kingdom. The survey consisted of questions organised into five sections:

- current public health data collection efforts
- cross-sectoral data collection efforts (e.g. linking human health data with information from environmental, animal, traffic, energy, and mobility sectors)
- intended future public health data collection strategies
- only once strategy and good practices (e.g. use of interoperability standards and data exchange formats)
- data visualisation and dissemination strategies (e.g. use of public health dashboards and AI)

5.2.2 Rationale for Country-Specific Questions

For the first section, current public health data collection efforts, surveys were tailored per country with previously catalogued national datasets listed for selection. This tailored design not only facilitated the identification of relevant datasets but also promoted alignment with data collection

[59] An example of the survey form in pdf format applied to one country (Ireland) can be found here: https://xshare-project.eu/wp-content/uploads/2025/01/xshare-WP4-survey-PDF.pdf

strategies from previous European projects, such as the Population Health Information Research Infrastructure (PHIRI)[60], in which Sciensano was a partner.

PHIRI successfully created a comprehensive catalogue of European public health datasets, accessible online via the Health Information Portal[61]. In accordance with xShare's objective to identify critical public health datasets, survey participants were asked whether they worked on datasets previously catalogued on the Health Information Portal. This approach reinforced the "only once" principle by avoiding redundant data collection and only collecting information that was not previously known. Such information included the standards and formats used during data collection, and the delay between data collection and availability. This ensures researchers can effectively understand and leverage existing resources for interoperability and further studies.

An example of identified datasets and the additional information required can be seen below.

			Terminology Standards					Data Exchange Formats			
	Delay	Length of delay	SNOMED	LOINC	ICD-10/ICD-11	ATC	others	HL7 CDA	HL7 FHIR	OMOP CDM	others
Belgian Cancer Registry	Rea	1-2	☐	☐	☐	☐	☐	☐	☐	☐	☐
Belgian Healthcare statistics	Sele	Sele	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 12: Additional information required for existing datasets (Survey)

If survey respondents worked on different datasets that have not been yet catalogued on the Health Information Portal, respondents were asked to complete a metadata record for the dataset. The data was collected according to the HealthDCAT-AP[62], a new standard meant to support the interoperability of health data for secondary use in the EHDS.

A printable version of the survey, specific for Denmark, can be found here.

5.2.3 Question Development

Expanding upon the collected public health datasets and related health information, the survey also aimed to gather insights into future data collection initiatives and strategies to enhance efficiency through the "only once" data collection principle.

A primary focus was understanding which terminology standards (e.g., SNOMED CT, LOINC, ICD-10/ICD-11) and data formats (e.g., HL7 FHIR, HL7 CDA, OMOP CDM, CDISC SDTM) are currently used and expected for future use. Additionally, the survey anticipated to determine whether countries have plans in place to standardise within the Electronic Health Record (EHR) industry, integrate EHR data into registries, or develop applications that rely on standardised and secure ETL (Extract, Transform, Load) processes.

Furthermore, the survey explored each country's use of public health dashboards, and the potential adoption of AI-driven tools to enhance their functionality. By collecting this information, the survey

[60] https://www.phiri.eu/
[61] https://www.healthinformationportal.eu/
[62] https://healthdcat-ap.github.io/

foresees to identify opportunities for improved interoperability, data visualisation, and how future xShare-supported dashboards can fill information gaps.

5.2.4 Survey Distribution

The finalised survey was distributed through numerous channels to maximize reach. First, all project partners were encouraged to promote the survey via their social media platforms, including Twitter and LinkedIn, and it was also shared through xShare's official communication channels. Also, targeted emails were sent to key stakeholders including public health institutes, industry representatives and federal health data authorities.

To further extend its visibility, the survey was showcased in several newsletters, including the xShare newsletter, the U4S and the International Association of National Public Health Institutes (IANPHI) network. The objective was not only to reach a broad range of stakeholders but also to encourage respondents to share the survey within their own communities, increasing its dissemination and engagement.

5.3 Results

5.3.1 General Information and Response Rate

The survey was launched in June 2024 and initially, was meant to remain open until September 30, 2024. However, at the time of the survey's end date, the only country that provided a detailed response was Germany. Thus, the decision was made to extend the survey until October 31, 2024, and prioritise extending the reach of the survey. The additional month welcomed significantly more contributions and by the survey's end date, 18 out of 34 countries (53%) responded to the survey.

5.3.2 Response Distribution

A downfall of the survey was the wide variation in participation across states. Certain countries provided substantive responses from actors involved in local, regional, and national public health initiatives (Germany, Netherlands, Denmark), while other countries only had participation from actors working on specific local projects (e.g. Spain).

The majority of responses came from national public health institutes, researchers affiliated with national registries, and universities. For some countries, numerous responses were recorded per public health institute (e.g. Sciensano in Belgium, Robert Koch Institute in Germany, Statens Serum Institut in Denmark, National Institute of Public Health in Norway).

Overall, the survey results were significantly less comprehensive than expected, with many countries either not responding or being underrepresented. This limited the ability to draw definitive conclusions about the overall public health data landscape across Europe. The countries represented and the organizations that contributed feedback can be seen in the graphs:

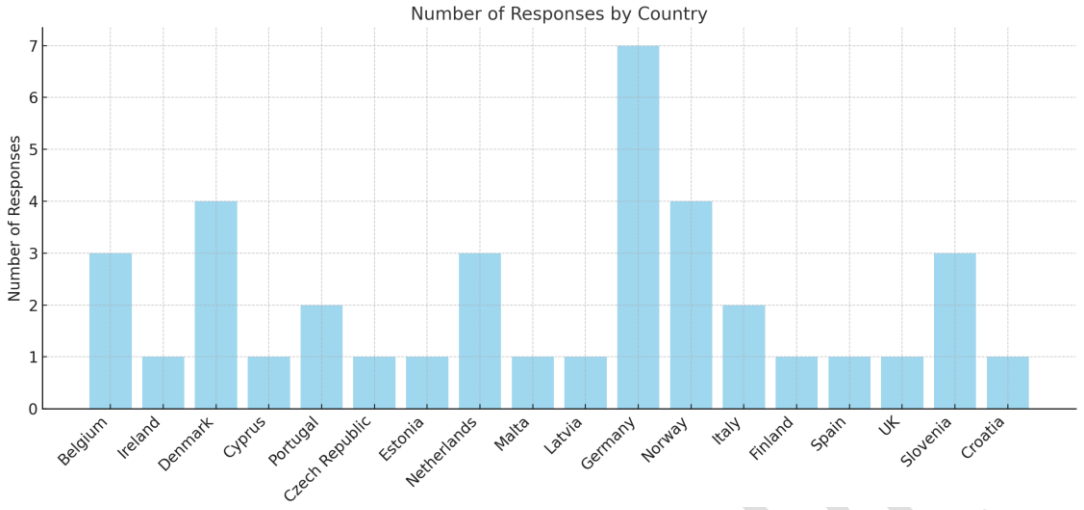


Figure 13: Number of responses obtained by country (Survey)

Country	Organization(s)	Number of Responses
Belgium	Sciensano	3
Ireland	Health Information and Quality Authority (HIQA)	1
Denmark	Danish Health Data Authority, Statens Serum Institut (SSI), Medcom	4
Cyprus	National eHealth Authority	1
Portugal	Regional Secretariat of Health and Civil Protection, Serviços Partilhados do Ministério da Saúde (SPMS)	2
Czech Republic	National Institute of Public Health	1
Estonia	Republic of Estonia Health Board	1
Netherlands	National Institute for Public Health and the Environment (RIVM), Netherlands Comprehensive Cancer Organisation (IKNL), Health-RI	3
Malta	Infectious Disease Control and Prevention Unit at the Ministry of Health	1
Latvia	Centre for Disease Prevention and Control	1
Germany	Robert Koch Institute, Otto-von-Guericke University, FDW, Bundesinstitut für Arzneimittel und Medizinprodukte	7
Norway	National Institute of Public Health	4
Italy	Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Fondazione Policlinico Universitario Agostino Gemelli (IRCCS)	2
Finland	Finnish Institute for Health and Welfare (THL)	1
Spain	Foundation for the Promotion of Health and Biomedical Research of the Valencian Community (Fisabio)	1
UK	Scottish Government	1
Slovenia	Institute of Oncology Slovenia, Slovenian Institute of Public Health (NIJZ)	3
Croatia	Croatian Institute of Public Health	1

Figure 14: Organisations participating in the survey

5.3.3 Key trends

Identified Datasets: The survey identified various types of datasets, which can be described as follows:

- Disease Registries: Includes registries for cancer, diabetes, rare diseases, infectious diseases, and antimicrobial resistance. These are common across many countries, supporting research and surveillance.

- Epidemiological and Public Health Data: Covers COVID-19 databases, national surveillance systems for healthcare-associated infections, and vaccination data.

- Administrative and Health System Data: Includes hospital billing, healthcare statistics, and registries of healthcare providers.

Although many datasets have been identified, the survey revealed only a limited number that feature continuous data updates and ensure real-time data exchange and availability. The following notable datasets were highlighted:

Denmark: The Danish Microbiology Database (MiBa) is a nationwide, automatically updated database of microbiological test results. Established in 2010, MiBa allows healthcare professionals across Denmark to access patient test results regardless of where the samples were taken. Maintained by the Statens Serum Institut, the database relies on automatic data extraction from electronic health records (EHRs). It uses national and local codes for terminology standards and MedCom's national XML standard for data exchange.

Norway: Norwegian stakeholders identified the Norwegian Immunisation Registry, and the Norwegian Emergency Preparedness Register for COVID-19 as examples of datasets with continuous data exchange and real-time availability. These datasets employ SNOMED CT and ICD-10/ICD-11 for terminology standards and use HL7 CDA and HL7 FHIR for data exchange.

Croatia: Croatian stakeholders highlighted the Croatian Immunisation Registry as a best practice for ensuring that data is collected once and made available in real time for multiple purposes. This dataset uses ATC and OMOP CDM terminology standards. Another example is the Croatian Medical Birth Registry, which is updated hourly and relies on ICD-10/ICD-11 and OMOP CDM standards.

Slovenia: Several Slovenian datasets were identified for their real-time availability and continuous updates, including the Slovenian Hospital Discharge Database, the Register of Vaccinated Persons and Adverse Reactions, the Slovenian Communicable Disease Registry, and the Primary and Secondary Care Data Outpatient Database. However, information on terminology standards and data exchange formats was not provided.

Malta: The National Hospital Information System Database collects hospital activity data from state and private hospitals in Malta. This dataset is continuously updated, with real-time data availability, and uses ICD-9/ICD-10 terminology standards.

Unfortunately, respondents from other countries did not provide details about core public health datasets with continuous real-time data collection. Further research is needed to identify additional datasets that can serve as best practices across Europe, particularly those that align with the "only once" strategy, ensuring data can be utilised for various purposes efficiently.

Standards and Formats: The survey collected information on the current and prospective use of health data standards and data formats across European countries. Respondents provided information about the specific terminology standards implemented, such as SNOMED CT, LOINC, and ICD-10/ICD-11, as well as the formats used for data, including HL7 FHIR, HL7 CDA, and OMOP CDM. Usage of these standards and formats are important for ensuring semantic and technical interoperability and supporting cross-border health initiatives.

The following graphs show the distribution of standards and data formats reported by the responding countries. While many countries have adopted international standards and formats, there remains reliance on local and regional codes.

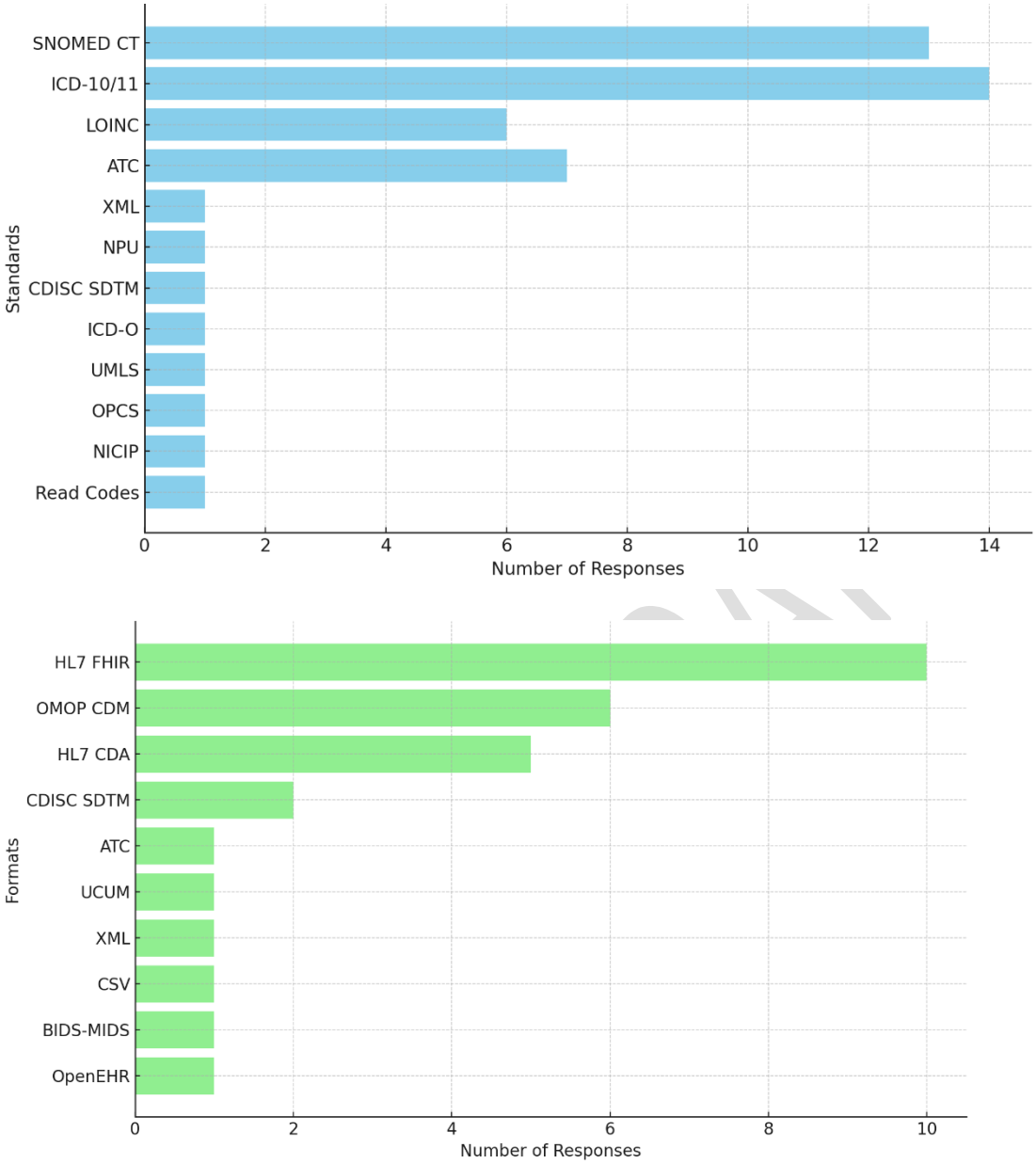


Figure 15: Adoption of standards - Survey response

Metadata Catalogues: Almost half (8, 48%) of the responding countries reported having established metadata catalogues, which are needed for organizing and managing health data for secondary use. Two countries (Germany and the Netherlands) mentioned the transition towards the HealthDCAT-AP application for categorising metadata and adhering to EHDS guidelines.

Dashboards and AI Integration: Many countries (16, 89%) indicated prior use of public health dashboards to visualize and monitor health data. Specifically, dashboards are used for surveillance of infectious diseases and have been established following the COVID-19 pandemic. A lesser number of states (12, 67%) indicated interest in incorporating AI capabilities to enhance analytics and decision-making. The concrete plans of expanding AI usage are unknown.

Standardisation in EHR Systems: Many countries (14, 78%) have initiated efforts to improve data standardisation, primarily through Electronic Health Records (EHR) and using EHR data to feed registers. Others are aiming to standardise but are still relying on less structured local approaches (Cyprus, Latvia). Half (9, 50%) of responding countries also reported on already implementing secure Extract, Transform, and Load (ETL) processes to extract and validate data from EHR, laboratory, and imaging systems for use in national registers and other applications.

Data Linkage: Cross-sectoral data linkage (linking health data with other sectors) is frequently mentioned, suggesting a trend towards multi sectoral health surveillance. A few countries (7, 39%) mentioned that they are prioritizing the linkage of human health data with environmental and animal data, specifically in the context of pandemic preparedness and infectious diseases. Cross-sectoral projects that were reported included European EXHAUSTION and Next Generation BiOactiVe NAnocoatings (NOVA) projects. Data linkage with the mobility sector was mentioned by respondents in 3 countries (17%). The UK reported additional data linkage with traffic accidents, wastewater surveillance, and the energy sector. Slovenia also reported linkage with education data for epidemic surveillance in adolescents.

5.4 Conclusion and Mitigation

There is a clear trend toward improving health data interoperability through the adoption of global standards (SNOMED, ICD-10/ICD-11, HL7 FHIR, etc.) and improving the interoperability and reuse of EHR data. Cross-sectoral data linkage and focus on secondary data use for surveillance and public health research are also common themes. However, due to the limited response rate to the survey, it is difficult to gain a full understanding of the scope of public health data collection and standardisation across Europe. To address this gap, detailed country profiles have been developed for specific countries that can be considered best practices. The countries highlighted later in this working paper are Denmark, Finland, Belgium and Spain.

6 Standardisation Initiatives in the Public/Population Health Domains

6.1 Inventory of Public and Population Health data Sets in MS

6.1.1 Current state of affairs in the EU

Accurate and timely health information is needed not only by healthcare providers to deliver precise and equitable patient care but also for researchers and policymakers to drive evidence-informed decision-making and foster innovative research. Traditionally, health information—at both patient and population levels—has been gathered and shared by a wide range of stakeholders across different countries, often stored in numerous websites and portals. This fragmented approach has made it difficult to obtain a comprehensive and user-friendly overview of health datasets and registers across Europe. To address this challenge, two European Joint Actions—the Joint Action on Health Information (InfAct, 2018–2021) and the Population Health Information Research Infrastructure (PHIRI, 2020–2023)—prioritised the creation of the Health Information Portal, a centralised one-stop shop for health information in Europe.

The Health Information Portal (HIP)[63] serves as a centralised gateway to population health and healthcare data across Europe. It is organised into several sections, providing access to catalogues of data sources, national and European projects, research infrastructures, capacity-building initiatives, and COVID-19-related resources. These resources are designed to support researchers with timely and reliable health information.

One key section of the portal focuses on data sources, offering an extensive collection of nationally organised datasets and registers along with their metadata. Currently, the catalogue features over 300 data sources categorised into areas such as survey and interview data, administrative data, infectious disease surveillance, and more. Authorised stakeholders can contribute datasets, ensuring the data on the portal remains accurate, updated, and well-maintained. An example of registered datasets on the HIP for Belgium can be seen below.

[63] https://www.healthinformationportal.eu/services/find-data

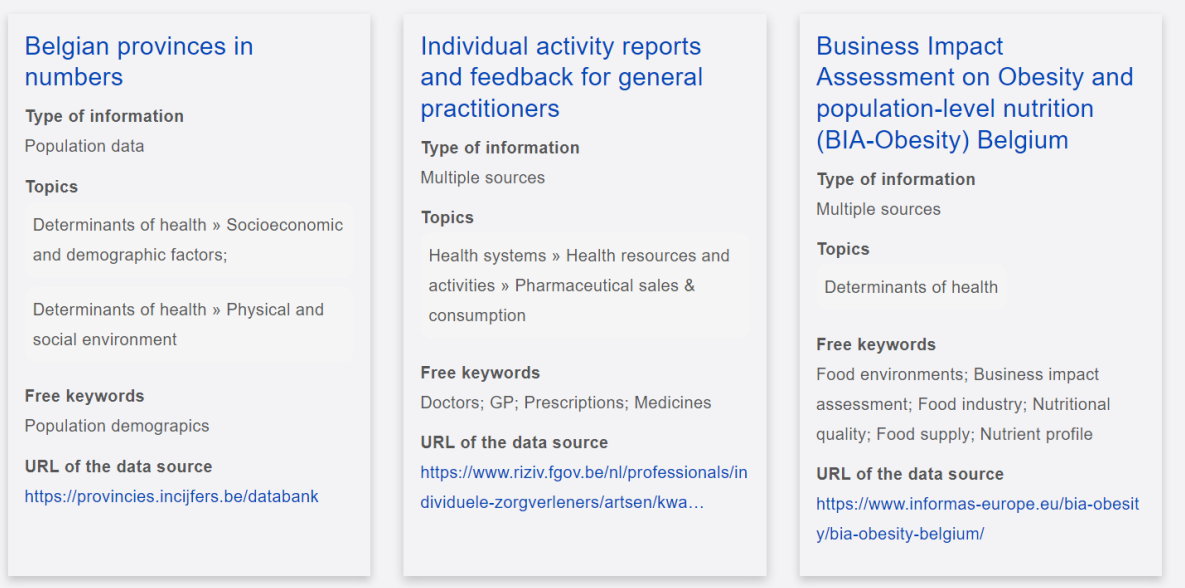


Figure 16: Examples of EU PH datasets available on the HIP portal

6.1.2 Towards a common metadata standard for (Population) Health Data sets

The EHDS Regulation establishes clear rules and processes for data availability, usage conditions, and the repurposing of data originally collected for other purposes, such as research, innovation, and policymaking. The EHDS is a cornerstone of the broader EU Data Strategy, which aims to create a unified European data space—a true single market for data. The EU Data Strategy outlines the development of Common European Data Spaces across strategic domains such as health, agriculture, manufacturing, energy, mobility, finance, public administration, skills, the European Open Science Cloud, and the Green Deal. Key standards for interoperability and security must remain consistent across all EU Data Spaces to ensure efficiency and alignment.

Article 60 of the EHDS regulation specifies the duties of health data holders with regards to data description:

> *3. The health data holder shall communicate to the health data access body a description of the dataset it holds in accordance with Article 77. The health data holder shall, at a minimum on an annual basis, check that its dataset description in the national dataset catalogue is accurate and up to date.*

Furthermore, **Article 77** reads that these new entities, health data access bodies, should provide national metadata catalogues:

> *1. Health data access bodies shall, through a publicly available and standardised machine-readable dataset catalogue, provide a description in the form of metadata of the available datasets and their characteristics. The description of each dataset shall include information concerning the source, scope, main characteristics, and nature of the electronic health data in the dataset and the conditions for making those data available.*

Finally, **Article 77** provides information on the elements that should be provided to describe a dataset:

4. By ... [two years from the date of entry into force of this Regulation], the Commission shall, by means of implementing acts, set out the minimum elements health data holders are to provide for datasets and the characteristics of those elements.

These articles pave the way for datasets to surface, i.e. making them findable in the European health Data Space through **metadata catalogues, enhancing data discovery.**

The DCAT Application Profile (DCAT-AP) serves as a key specification for metadata records, enabling semantic interoperability in the EHDS and other European data spaces. The HealthData@EU pilot[64] developed a first version of HealthDCAT-AP, a health-specific extension of DCAT-AP. This model enhances the discovery and accessibility of health data, including electronic health records, while prioritizing privacy, security, and responsible data sharing. The Joint Action TEHDAS2 will finetune and validate the HealthDCAT-AP in the broader health community. The draft version of HealthDCAT-AP is available on GitHub.

6.2 Current use of Standards in the Public Health domain

6.2.1 Alignment and mapping of standards used in clinical, research and public health domains.

International standards have been built to serve as exemplary models within a specific ecosystem but often lack real consideration for the overarching ecosystem. Consequently, those standards usually lack alignment. With the implementation of the EHDS, a better alignment between standards used in clinical practice, clinical trials, observational studies and public health statistics and benchmarking needs to become a priority. This is an absolute condition to improve data quality, reduce inefficiencies, reduce time lags and support innovative use cases deemed to create real value.

While the WHO classifications have traditionally been most widely used in public health, and sometimes for clinical purposes that fall outside its intended mandate, SNOMED CT is increasingly being recognized as the reference terminology for clinical practice[65]. Many European Member States have invested significant resources in creating validated translated subsets of SNOMED CT. Yet, even in countries where translated sub-sets are available, the actual use of SNOMED-CT remains limited partially due to its comprehensiveness (usability) and conditions of use (patent). The rapid uptake of AI might however significatively improve the usability of SNOMED-CT, moving towards an AI-supported coordination of concepts rather than the manual selection of a pre-coordinated concept.

CDISC-SDTM[66] is the undisputed reference when dealing with clinical trials while OMOP has recently become extremely popular when dealing with observational studies and real-world evidence. The CDISC controlled terminology is curated by the NIH National Cancer Institute (NCI) Enterprise Vocabulary Services (EVS)[67]. The typical dictionaries associated with CDISC-SDTM are MedDRA[68],

[64] http://www.ehds2pilot.eu
[65] 22 EU countries are today members of SNOMED international.
[66] https://www.cdisc.org/standards/foundational/sdtm
[67] https://www.cdisc.org/standards/terminology/controlled-terminology
[68] MedDRA is used to standardize the Medical History, Clinical Events, Adverse Events (Medical/Health Conditions- the domain is dependent on the timing of the event and the protocol specification). In FHIR this is the Condition Resource, In OMOP this is the condition occurrence..

WHODrug, and CDISC own Controlled Terminology. MedDRA and SNOMED-CT have also recently been mapped. Both LOINC and SNOMED-CT dictionaries may be used when applicable to the data. CDISC includes other standards related to the clinical research process, including for non-clinical data, clinical study planning, and clinical data collection. Further, data exchange standards are supported by CDISC such as dataset metadata submitted using the Define-XML standard.

At its core, OMOP CDM is designed to harmonize disparate healthcare data sources. OHDSI (Observational Health Data Sciences and Informatics) which maintains OMOP claims to support more than 100 semantic resources although a number of them have been abandoned or have been retired. OMOP philosophy is that each organisation or user builds its own Common Data Model (CDM) while tools are provided for assistance with Exchange, Transform, and Lead (ETL) processes to transform data. However, the use of OMOP CDM can require significant effort to transform the data. Organisations need to invest in training and technology to effectively implement and use the standard effectively. Finally, OMOP CDM has limited flexibility outside of its standardised structure. Adaptations and/or extensions to the model are not supported by OHDSI, therefore studies that require data elements not currently in the model would introduce complexity. Let's also mention the ATHENA platform[69] which u-is a result of the OHDSI Clinical Trial Working group where mapping between SDTM and OMOP is being performed.

A HL7 FHIR resource defines a set of data elements, the relationship between the data elements, and constraints on the data. There are currently 157 resources each defining a health data concept (e.g., Patient, Condition, Observation...). Resources can be further constrained to form HL7 FHIR profiles tailored to meet specific use cases. All resources have metadata and a human-readable component, and each resource of the same type is formatted in the same way. Resources or data elements each have a tag that acts as a unique identifier, like the URL of a webpage. HL7 FHIR is optimized for Real Word Data and hence for EHR and claims data. HL7 FHIR also leverages web standards such as JSON, HTTP, AToM, OAuth and others. HL7 FHIR however lacks many of the concepts necessary to represent clinical research data. HL7 FHIR relies typically on dictionaries such as SNOMED, LOINC, ICD, EDQM or RxNorm. The innovation of HL7 FHIR relies in the granular organization of information which makes data exchange particularly efficient.

It is also important to consider the patient-centric approach followed by openEHR[70]. It is based on clinically-led clinical data models (archetypes) designed and managed by the International Clinical Knowledge Manager (CKM), nd the modelling community consists of nearly 3000 clinicians from 90+

https://ohdsi.github.io/CommonDataModel/cdm54.html#condition_occurrence
https://hl7.org/fhir/uv/ips/StructureDefinition-Condition-uv-ips.html

[69] OMOP provides big data and provide signals while SDTM is designed to test hypothesis and is very controlled and rigorous. The CDASH case report forms (which are present in RedCap) can be leveraged for clinical trials. This would help in a pandemic because the domain specific outlines are present.

[70]

https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html#_ontological_separation

countries. Over the past 15 years, they have created over 800 clinical data models[71] which are published freely. OpenEHR claims that unlike mainstream healthcare IT, all clinical models are built by healthcare domain experts. OpenEHR provides a framework for developing information models that mimic data structures in external systems and international standards, to effect both import and export. International and local terminologies are used via bindings in the models. It aims to provide interoperable and vendor-independent data repositories. However, several shortcomings exist including poor compatibility with medical data other than EHR, difficulties in defining prototypes, steep learning curve and the lack of mature development tools and guidelines. Both openEHR and OMOP CDMs are ***designed for exporting and reusing data from a distributed clinical database***. The former is suitable for collecting data from distributed EHR systems and building medical big data warehouses, while the latter is a better model for sharing data in some decentralised medical database.

Finally, it is important to mention the Dutch Initiative of Clinical Building Blocks: A specific initiative "**data registration at the source**"[72] supports organisations which want to better exploit and reuse their data. It relies on the use of *Healthcare Information Building Blocks* (ZIBs) when building or designing the EHR and the specification of a basic data set (BgZ)[73]. A ZIB Compliance framework[74] has been designed to guarantee consistency but unfortunately these models have not always been implemented consistently and the results are therefore limited.

There is no single approach which can meet all requirements, and a clear trend has emerged favouring hybrid approaches. This means that all major models are supported, with data submitted in the model most suitable for the data producer. However, this approach demands greater effort in terms of alignment, as each domain must also consider the requirements of other domains. The work performed by xShare WP5 related to the IPS+R (Patient Summary adapted for research) is very exemplary in this respect.

The list below is the result of the TEHDAS1 work related to the identification of International Standards relevant for secondary use of data. However, it is not complete due to the lack of inclusion of practices including nursing and dentistry that have also developed specific international terminologies. Many Member States are also still using proprietary (national) coding systems for a number of data sets, which can be found on the national resources page on the PHIRI portal. The TEHDAS1 list presented in figure 17 gives a first overview of the 24 semantic resources to be considered a priority.

[71] https://github.com/openEHR/gdl-guideline-models/blob/master/archetypes/openEHR-EHR-CLUSTER.diabetic_retinopathy_classification.v1.3.adl See also: https://specifications.openehr.org/releases/RM/latest/ehr_extract.html

[72] https://www.registratieaandebron.nl/over-het-programma: Support has however stopped in 2022.

[73] https://informatiestandaarden.nictiz.nl/wiki/BgZ:V2.0_Beta_2_BgZ_2017_Technical_IG

[74] https://www.registratieaandebron.nl/files/Raamwerk-zib-compliance.pdf

Acronym	Typology	Utility	Domain	Information (URL)
DCAT-AP2	Meta-data standard	Discoverability	Public reporting of data catalogues	https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en
INSPIRE	Meta-data standard	Discoverability	Data geo-allocated	https://inspire-geoportal.ec.europa.eu/#
FairSharing	Meta-standard	Discoverability	Data sources, meta-data standards, taxonomies of any health domain	https://fairsharing.org/biodbcore/?q=health+record
BBMRI/MIABIS	Meta-data standard	Discoverability	Bio samples	https://github.com/BBMRI-ERIC/miabis
Beacon	Metadata standard	Discoverability	Genomics, clinic data	https://beacon-project.io
Bioimage archive	Meta-data standard	Discoverability	Bio-Image	https://www.ebi.ac.uk/bioimage-archive/
CEDAR	Meta-data standards	Discoverability	Biomedical experiments	https://metadatacenter.org/
ECRIN	Meta-data standard	Discoverability	Randomized clinical trials	https://ecrin.org/tools/clinical-research-metadata-repository
CESSDA	Meta-data standards	Discoverability	Social data collections	https://datacatalogue.cessda.eu/
PHIRI	Meta-data standard	Discoverability	Population health data collections	https://www.healthinformationportal.eu
OMOP	Layer of interoperability	Common data model	EHR, claims data	https://ohdsi.org/omop/; https://athena.ohdsi.org/search-terms/terms?domain=Condition&sort=vocabulary_id&order=asc
SDTM | CDISC	Layer of interoperability	Common data model	EHR, claims data	https://www.cdisc.org/
ORDO	Layer of interoperability	Taxonomy	Rare-diseases	http://www.orphadata.org/cgi-bin/index.php#ontologies ; http://www.orphadata.org/cgi-bin/index.php
SNOMED-CT	Ontology	Data provenance	HER, bio sample	https://www.snomed.org/
LOINC	Ontology	Data provenance	Lab data	https://loinc.org/; LOINC is now part of SNOMED-CT
ISO23494	Meta-data standard	Data provenance	Bio samples	In development here: https://www.iso.org/standard/80715.html
ISO 8000 110	Meta-data standard	Conformance messaging	Any data master file	https://www.iso.org/standard/78501.html
HL7 FHIR	Meta-data standard	Conformance messaging	EHR	https://www.hl7.org/fhir/; See also https://art-decor.org/mediawiki/index.php/Main_Page
DICOM	Meta-standard	Conformance messaging	Images	https://www.dicomstandard.org/
SPOR	Meta-data standard	Conformance messaging	Medicinal products (Substances, Products, Organization, Referentials)	https://www.ema.europa.eu/en/human-regulatory/research-development/data-medicines-iso-idmp-standards/substance-product-organisation-referential-spor-master-data
EDQM Standard Terms	Meta-data standard	Conformance messaging	Medicinal products (Dose forms, Routes of administration, Packaging, Units of presentation, etc.)	https://standardterms.edqm.eu/
ICF	Ontology	Common data model	Functioning	https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health
EMDN	Meta-data standard	Conformance messaging	Medical devices	https://ec.europa.eu/health/sites/default/files/md_sector/docs/md_emdn_eudamed_nomenclature_en.pdf

Figure 17: List of standards recorded by TEHDAS1

A few Member States have created a dedicated infrastructure to facilitate alignment between semantic resources used by the different domains, and most importantly, to allow organisations and companies to map legacy proprietary systems with international standards. Specific examples include initiatives launched in **France** and in **the Netherlands** which both claim to be HL7 FHIR compatible. Snowstorm is an open-source terminology server developed and maintained by SNOMED International. Snowstorm offers a standard HL7 FHIR API that implements the HL7 FHIR Terminology Module specification. This provides access to the terminology using HL7 FHIR Operations and returning content structured as HL7 FHIR Resources. The existence of a European Terminology Server which would provide users with an instant access to updated European and international semantic assets (and their mappings when available) would be an important enabler to accelerate both alignment and deployment. National Terminology servers could then mainly focus on the specific needs related to translations and national legacies.

Aside from the availability of tools and semantic resources, an adapted methodology can also help. The project InteropEHRate has proposed via a dedicated whitepaper[75] an **end-to-end data integration methodology**, supported by a suite of graphical tools, while simultaneously pushing the boundaries of precision and automation. Pending stricter alignment between domains, this methodology provides

[75] https://www.interopehrate.eu/blog/2022/05/17/white-paper-towards-interoperable-health-data/

an interesting path to reduce inefficiencies and create better links between domains at the level of major clinical organisations such as hospitals.

In order to map, transform, and translate health data, the system needs to acquire an in-depth understanding of the data. For this to happen, the system needs to be fed with knowledge about local data and practices by a human data manager. The data manager (1) sets up the formal knowledge relevant to the interoperability problem to be solved; (2) defines the transformations and mappings on terms, codes, data attributes and values; and (3) curates and maintains these transformations and mappings over time. The data manager typically uses and needs material help in the form of (a) example data; (b) specifications of the interoperable target data representations; and (c) existing crosswalks.

Finally, with the upcoming EHDS and the duty for health data holders to describe their data using HealthDCAT-AP, it will be possible to have a better overview of the standards and coding systems in use, as the application profile contains the 2 properties *Coding system* [healthdcatap:hasCodingSystem] and *Code values* [healthdcatap:hasCodeValues]. *Coding system* indicates the coding system used within a dataset, providing crucial information for its reuse and discoverability. For example, if a dataset uses ICD-10 for disease classification, this property allows data users to search for datasets with the same coding system. As a machine-actionable property, it also facilitates automated processes, making dataset discovery more efficient. The *Code Values property* complements the *Coding System* property by detailing the specific codes used within a dataset. It improves the discoverability of datasets by enabling searches for specific diseases or conditions using coding systems like the ICD-10 example provided. Like the *Coding System* property, this is machine-actionable, allowing for automated searches and processes to enhance dataset accessibility. As metadata records of health data sets become more and more available, this gives the unique opportunity to map the use of standards and coding systems in Europe.

6.2.2 Connecting national and European Union Infrastructures

Currently, several connections between national and European (Union) infrastructures are already in place. Examples include:

ECDC: TESSY[76]: In accordance with the ECDC (European Center for Disease Prevention Control) and founding regulation (Regulation (EC) 851/2004) the EU Member States have to provide ECDC *in a timely manner with the available scientific and technical data relevant to its mission*. The European Surveillance System (TESSy) is provided by ECDC in order to collect, analyse and disseminate surveillance data on infectious diseases in Europe. All EU Member States and EEA countries report data on communicable diseases to the system.

Health statistics by Eurostat[77]**:** The EU regulation 1338/2008, ensures that health statistics provide adequate information for all EU members to monitor actions in the field of public health and health and safety at work. This regulation lists 5 domains:

- Health status and health determinants: statistics based on the European health interview survey

[76] https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy
[77] https://ec.europa.eu/eurostat/web/health/overview

- Health care (health care expenditure and financing)
- Causes of death
- Accidents at work
- Occupational diseases: experimental statistics

In addition, data on selected health issues are based on the legislation of the 2 social surveys used, namely the legislation on EU statistics on income and living conditions and the legislation on the labour force survey ad-hoc modules on accidents at work and other work-related health problems.

JRC's EU Wastewater Observatory for Public Health[78]: The EU Wastewater Observatory for Public Health aims to collect and share wastewater surveillance data and transform them into information to support health decision making. One sub-project is called 'DEEP': The Digital European Exchange Platform that connects and collects data from national, regional or local wastewater surveillance activities. Partners include Health or Environmental Services, private sector bodies, Research and Academia, and NGOs.

[upcoming]: ATHINA platform by HERA: The new Health Emergency Preparedness and Response Authority (HERA) was set up to strengthen Europe's ability to prevent, detect, and rapidly respond to serious cross-border health emergencies. According to HERA, high priority health threats that require coordination of measures at the EU level to bolster our response capacities include 1) pathogens with high pandemic potential; 2) chemical, biological, radiological, and nuclear threats (CBRN); and 3) threats resulting from antimicrobial resistance (AMR). In order to support this, HERA is developing the IT system ATHINA (Advanced Technology for Health INtelligence and Action IT system). The national IT systems for these data should be compatible with this system. As this will only be operational if Member States have strong national IT systems that are interoperable with HERA's IT system and other relevant systems, the EU-HIP project supports participating countries to enhance and improve national IT systems in an efficient and coordinated manner (see also Section 4.2.17).

HERA developed a list of three high priority health threats that require coordination of measures at the EU level to bolster our response capacities: Pathogens with high pandemic potential; Chemical, biological, radiological, and nuclear threats (CBRN) and Threats resulting from antimicrobial resistance (AMR).

[78] https://wastewater-observatory.jrc.ec.europa.eu/

7 Only once best practice country reports

Registries are characterised by diverse methodologies in technology (e.g., paper, fax, e-mail, web applications, batch transfer, web services) and data structures (holding the content), and often have repeated requests for the same data from data providers.

The origin of this heterogeneity in the methodology of research registries – most of which have existed for decades – was strongly determined by what was technically and organisationally feasible at the launch of each registry with the data collector and the participating data providers.

In addition to the diversity of data collections, there is a great diversity of data providers. Starting with the type of care providers. These include large (e.g. hospitals, laboratories) and medium-sized (e.g. mental health centres) institutions, group practices (including multidisciplinary) or individual care providers (e.g. general practitioners, paramedics). Within each of these types of healthcare providers, there is also a great diversity in terms of the accessibility of medical information in the IT systems used: from almost maximum to partial availability of the data sought in operational IT systems, or even the absence of an IT system at the data provider.

The consequences of this diversity for providers and collectors of this data include reduced efficiency when recording and processing information (a lot of manual data cleaning, re-encoding and mapping), real risks to privacy, and the dispersion of IT and human resources for the same tasks. In addition, this context is not motivating for the (highly qualified) employees who are associated with these data collections (both at data providers and researchers). Finally, this situation results in a major financial effort (direct and indirect) for data providers, researchers and their clients.

This working paper focuses on 4 official initiatives in Denmark, Belgium, Finland and Valencia Region (Spain) which attempted to navigate this issue and identify a number of lessons learnt.

This working paper only reproduces the summary of the comprehensive reports which can be found in the annexes. The section below outlines comprehensive summaries of benefits and lessons learnt.

7.1 Denmark

This “only once” best practice is one of the very first European MS Initiatives to connect primary and secondary use of data. Although it started as a public driven initiative, the private sector has quickly understood the strategic importance of the issue and developed its own initiative. The report focuses mainly on the public initiative called SENTINEL but also mentions the parallel initiative- the Health Hub- which has been developed as a collaborative effort by the different GPs EHRs software’s operating in Denmark.

7.1.1 Origin and objective

The SENTINEL System is a digital solution aimed at improving **data monitoring, operational efficiency**, and **decision-making** processes. The system automates tasks that were previously manual, enabling more **accuracy** and **timely reporting**. The main initial motto has been the capacity to provide healthcare providers with the benchmarking deemed useful for their practice and build collaboratively the knowledge which may improve quality of care. With the rising demand for data by different authorities, the initiatives are increasingly connecting the EHR with external Databases and third-party applications with adapted privacy modalities.

SENTINEL addresses specific challenges such as:

- **Data fragmentation**: Difficulty in consolidating and interpreting information.
- **Operational delays**: Manual methods caused inefficiencies.
- **Limited analytics capabilities**: The lack of real-time data limited decision-making.
- **Data duplication and data quality validation**

Sentinel offers a **centralised platform** for integrating data from the different specialties' practices, streamlining workflows, and enabling **better data insights**. It currently mainly supports the following mainly quality related finalities:

- Data quality check and support to diagnostic codes
- Individual and practice benchmarking
- Reporting to national approved clinical quality databases
- Compulsory reporting (diagnosis codes) to the regions
- Quality improvement in priority treatments (e.g. diabetes)

The data consumers of the system are:

- The Healthcare providers themselves
- The 28 national registers and the 85 quality registers
- The Regions

The main common structured data that are currently captured are:

- Diagnostic codes
- Laboratory results/values
- Medication information
- Patient ID
- Practice identification information

Other types of data captured depend on each specific approved project. For the Coding of diagnostics, ICD-10 is used.

From January 2024 on, the provision of diagnostic codes to the Regions and the Danish Health Data Agency has become legally compulsory.

Sentinel automatically copies all structured data from the HCP journal: it collects data from the patient record and returns it to the data producer in a structured and personally identifiable form. Copied data are stored in a local Sentinel database that is part of the HCP journal system. Data are then also kept separate for each clinic in the sundhed.dk database in the same way as an account in an ordinary bank. Depending on requirements of each quality project, information that is not already structured in the EHR may be needed. This could, for example, be about assessing how a clinical guideline has been implemented. As shown in figure 18, the data to be provided and which were not present in the EHR are related to questions and variables agreed upon by each healthcare speciality. In such cases, a pop up is displayed if, for example, the HCP enters a specific diagnostic code. The pop-up can function both as a reminder and decision support in the patient treatment. Each healthcare provider remains free to select the data they agree to share.

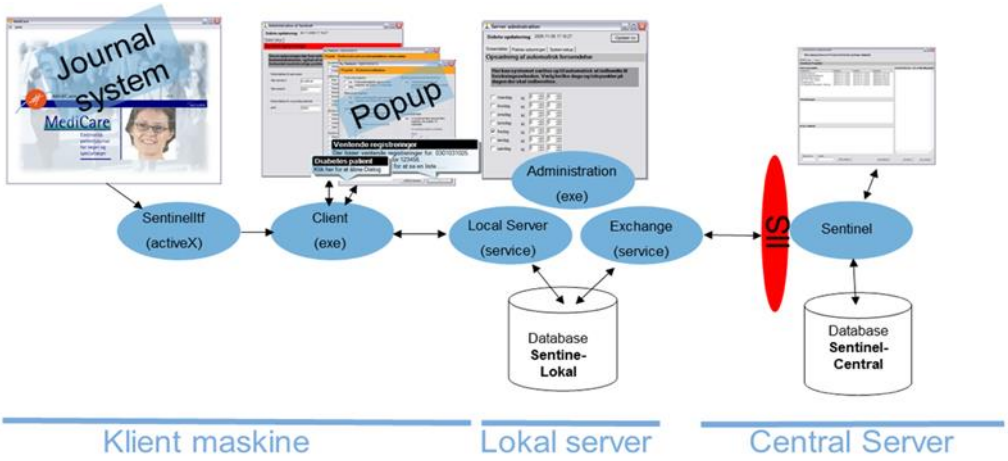


Figure 18: SENTINEL (DK) simplified workflow

7.1.2 Key learnings

The overall implementation highlights the importance of addressing both technical and human factors during digital transformations.

The success of the Danish initiative lies in its capacity to make sure that:

- Efforts required are carefully calibrated focusing only on essential data and limiting initially data compulsory coding to diagnosis.
- There is an immediate added value for the data producers, sometimes for reasons not immediately identified such as the need to be able to better objectify.
- Specific objectives and requirements have been discussed with each specialty which promoted the system in its own network.
- Each HCP remained free to decide the level of involvement.
- Legal requirements came only when the solution has been largely adopted and is technically mature.

Users of SENTINEL also emphasised efficiency gains and impact on quality of care and minimisation of human errors for specific pathologies. With a high patient turnover (around 40 patients per day), reducing administrative burdens, such as data entry, is critical for private practitioners. Data is crucial for engaging policymakers and securing resources for private practices, which handle a significant portion of healthcare in Denmark. Tools like pop-ups for mandatory fields (e.g., diabetes) are acceptable only if they align with the doctor's workflow and provide clear value. Practitioners prioritise minimising interactions with systems that distract from clinical tasks. The system also enables tracking of disease progression (e.g., diabetes and retinopathy), supporting personalised care plans and contributes to the reduction of visit frequencies. The SENTINEL's "plug and play" nature and its integration with existing EHRs reduced the need for extensive training or system interface management.

The main identified barriers to the use of the system are:

- While representative user associations were involved in initial development, broader input from specialists was limited, leading to gaps in addressing specific needs.

- Although ICD-10 coding is mandatory, the lack of immediate utilisation of coded data by authorities reduced the perceived value for private practitioners.
- Data portability is not perceived as a possible benefit for users who remain usually faithful to the solutions they have been using.
- Not all practitioners prioritise contributing to health system discussions or research due to their workload.
- While the system supports decision-making by comparing clinical patterns and practices, its potential is still underutilised due to insufficient detail and engagement.

The interview highlighted the balance between reducing administrative burdens for practitioners while leveraging their participation in data-driven healthcare improvements. Clear incentives, streamlined processes, and integration into workflows are crucial for broader acceptance and sustained engagement. Increasing awareness of the system's added value, backed by statistics (e.g., diabetes screening success), can boost acceptance and usage.

SENTINEL and Health Hub

The Health Hub, a successful competing initiative launched by the GPs software's Industry is also highlighting important lessons learnt: The creation of a dedicated company to manage interactions with third parties and external databases came as an answer to address privacy, transparency, and usability concerns with the first SENTINEL developments. Independent governance of the new company allowed focus amidst competing development priorities. Strategically this segment of industry also understood that this was to become a strategic element and that remaining owners of the solution would provide considerable business advantages. It is however a very original initiative as competing companies have agreed to create a dedicated company to deal with all EHR externalities. The company also benefited from public subsidy to make initial developments: This was justified by authorities as the GPs were not originally the priority target of SENTINEL but were nevertheless important for a number of metrics. A close partnership has been established with MEDCOM and the GP association (PLO) to design a unified approach. The industry considered that it was also best able to address GPs' demand for feedback and research data on specific diseases.

Main motivators include allowing GPs to showcase their productivity and adherence to care pathways, portfolio analysis and providing visual tools to compare patient demographics and regional benchmarks. These have been the primary immediate drivers.

The Health Hub is now trying to attract other segments than the GPs. The Health Hub believes that the agility of the system developed, and the governance of the company will allow it to adapt swiftly to new requirements, including those which include cross-border aspects. All GPs EHR software adopted a common data model: They are using ICD-10 but also increasingly other standards, addressing legacy systems while planning for alignment with global standards like FHIR (See summary of current connectors in annex).

Key messages:

- Requirements for data reuse need to be correctly calibrated and related to the expected return on investment for the data producers themselves.

- Denmark SENTINEL is an excellent example of an initiative co-created with the users with a good mix of drivers before considering legal enforcement. Other complementary co-created initiatives such as CAPRI and INCEPT (see annex) have been launched to support iterative research.
- Industry can play a key role in providing adapted solutions for connecting primary and secondary use of data provided that their competitive advantage can be protected.
- Although an eHealth European champion, Denmark has been slow to engage in discussions around interoperability challenges in Europe and beyond. National legacies remain numerous and pervasive. Both Public Competence Centre and Industry seem to have understood the necessity to make swift progress in this regard. The EEHRxF specifications are evidently considered with a lot of attention and expectation.

7.2 Belgium

Belgium's public health data domain, particularly around interoperability in health records and laboratory data, reflects a fragmented but ambitious history shaped by regional and national policies, as well as European initiatives. Due to the country's distinct regional federated structure, hospitals and laboratories in different regions in Belgium (Flanders, Wallonia, and Brussels) developed their own systems and standards without standardised national data formats or protocols. The lack of interoperability between regional systems made it challenging to create cohesive national standards and efficient systems to collect data that can be processed and reused more than once.

A better alignment between primary and secondary use of data has been discussed quite early in the Belgian Digital Health journey. In 2012, the first Digital Health roadmap was created through a co-creation process inclusive of all stakeholders of the value chain. This process led to the description of 20 SMART [79]actions. Action 18 was dedicated to the inventory and consolidation of all datasets in the context of secondary use. Multiple new data collection initiatives had been launched, often without coordination between them, and as consequence, resulted in a new burden for the healthcare organisations and professionals. Hence, the very rationale of this action point was to rely as much as possible on an "only once" strategy where data for secondary use would be captured as much as possible from existing EHRs. The action point mentioned the following priorities:

- The creation of a detailed inventory of all existing registries and databases by Sciensano. This inventory had to provide a precise description of the data fields collected, their status (optional or not), their format, the codifications (proprietary or not) used, the type of user and institutions concerned as well as any other essential contextual information.
- The establishment of governance rules for the creation of any new register which would require the consideration of data available in EHR systems as apriority.
- The systematic adoption of standardised system-to-system communication protocols and the gradual phasing out of dedicated data collection platforms tied to specific datasets in the mid-term

During the first revision of the action plan in 2015, it was specified that the datasets in question are not limited to those managed by Sciensano but should also encompass datasets created to support epidemiology, vigilance and quality of care. The revised plan also provides significantly more details on

[79] Specific, Measurable, Achievable, Relevant, and Time-Bound

the 13 sub-actions required to achieve its goals. The Healthdata.be platform is mentioned for the first time, with the HD4DP connector identified as a key channel for implementing the “only once” strategy. Additionally, new types of data, such as Next Generation Sequencing (NGS), are considered.

Among other initiatives, the extended use of a Belgian adaptation of the Dutch Clinical Building Blocks, referred to as Care Sets, was planned, along with the adoption of SNOMED CT concepts for all values supported by these building blocks. However, these efforts have not yet resulted in practical implementation, though the rollout of Care Sets is expected to occur soon. Very significant progress has however been achieved to create a structured and governed connection between primary and secondary use.

Today, Belgium has placed a strong focus on improving coordination and interoperability across its regional healthcare systems and supporting centralised and standardised data registries and warehouses. A central component of this effort is healthdata.be, a collaborative initiative between Sciensano and the National Institute for Health and Disability Insurance. To support these goals, healthdata.be developed the Health Data for Data Providers (HD4DP) application. This report examines how hospitals and laboratories are interacting with this application, shedding light on the challenges they face and the successes they have achieved.

7.2.1 Background on Health Data for Data Providers (HD4DP) application

Health Data for Data Providers (HD4DP (currently version 2.0)) is a data collection application used by healthcare providers in hospitals and laboratories for the recording of health and healthcare data. Installed locally at hospital sites and laboratories, it enables the secure and standardized submission of medical data. The software is maintained by the healthdata.be platform within Sciensano.

7.2.1.1 Technical Architecture and Installation

The HD4DP 2.0 Local is an electronic data capture (EDC) system: a computerized system designed for the collection of clinical data in electronic format for use in research supporting human public health policy. The application is meant to replace traditional paper-based data collection methodology and streamline data collection, analysis and reporting.

HD4DP 2.0 local application contains the following components: NextGen Connect, Form.io, HD Connect (LOCAL Proxy), and Local data warehouse. All components work together to translate standards, support more efficient API development, store data securely in localised data warehouses, and ensure that data is in the correct formats for reporting and extraction. Additional technical details of the HD4DP application can be found here.

The software is provided without cost and installed remotely on the infrastructure of the healthcare organization by healthdata.be. Healthcare organizations are provided the system requirements for installation of HD4DP 2.0. The application is then maintained without cost remotely on the infrastructure of the healthcare organization by healthdata.be. The infrastructure on which the application HD4DP v2 Local is installed, should be maintained by the healthcare organization.

The technical components that must be present at the data provider to participate in the data collections of the healthdata.be platform are:

- primary operational systems of the data provider (e.g. electronic medical record, laboratory information management system,...)
- HD4DP software of the healthdata.be platform
- an encryption module
- sending client for the eHealthbox,
- eHealthbox

Except for the HD4DP software and the encryption module, the other components are mostly already present in hospitals and laboratories. HD4DP software is usually installed by the healthdata.be team in hospitals and labs.

7.2.1.2 Data Extraction using HD4DP

Data can be submitted to the application via: Manual entry, CSV upload, or through an API integration. Then once data is in the HD4DP 2.0 application, data is split into two streams:

1. Patient Data: Transmitted through eHealth, where the patient's ID is pseudonymized. Healthdata.be only receives a technical identifier to ensure patient anonymity.

2. Medical Data: Sent via SFTP and tagged with the technical identifier. This allows the pseudonymised patient ID and medical data to be later combined for research purposes.

Once collected, the data is stored in the healthdata.be data warehouse and made available for research purposes to approved clients such as Sciensano researchers, scientific organizations, pharmaceutical companies, and universities. Clients seeking access to healthdata.be data must submit a formal request form and undergo an internal assessment to ensure the project aligns with healthdata.be's scope and procedures. This structured process ensures transparency, compliance, and alignment with institutional priorities.

7.2.2 Laboratory use of HD4DP application in the context of the Epilabo Project

7.2.2.1 Background and objectives of Epilabo project

Epilabo is a project aiming at collecting laboratory test results from a network of sentinel microbiology laboratories in order to monitor the epidemiology of various infectious diseases in Belgium. The project has been operating since 1983 through Sciensano and relies on the voluntary participation of laboratories across Belgium in weekly submitting diagnostic data for 40 + pathogens. Epilabo aims to improve public health through monitoring epidemiological trends and detecting potential outbreaks, assessing the impact of prevention programs, predicting the potential spread of infectious diseases, and reporting data to international health authorities. Epilabo is a part of the European project EU interoperability with HERA's IT platform (EU-HIP) that develops new IT systems as well as strengthens and enhances existing national IT systems for improving health threat assessment in European countries.

7.2.2.2 COVID-19 pandemic as an enabler for HD4DP use

During the pandemic, laboratories transmitted daily nominative COVID-19 data to healthdata.be, where the data was standardised and forwarded to regional health authorities, enabling them to trace

positive cases back to individual patients. The data was then also pseudonymized and made accessible in the healthdata.be data warehouse and within Sciensano's Epidemiology and Infectious Diseases service. This facilitated the secondary use of the data for research, reporting, and surveillance purposes.

The enhanced data collection process implemented during the pandemic was supported by Health Data for Data Providers (HD4DP 2.0). Laboratories were able to provide System-to-System (S2S) transfer of JSON files via an API, ensuring real-time, standardised data exchange that were efficiently distributed to relevant stakeholders. The application uses the Belgian Preparedness Architecture for Infectious Diseases (be.Prepared), as seen in Figure 19, which is designed in response to the need for one central data retrieval for exchanging personal data concerning public health.

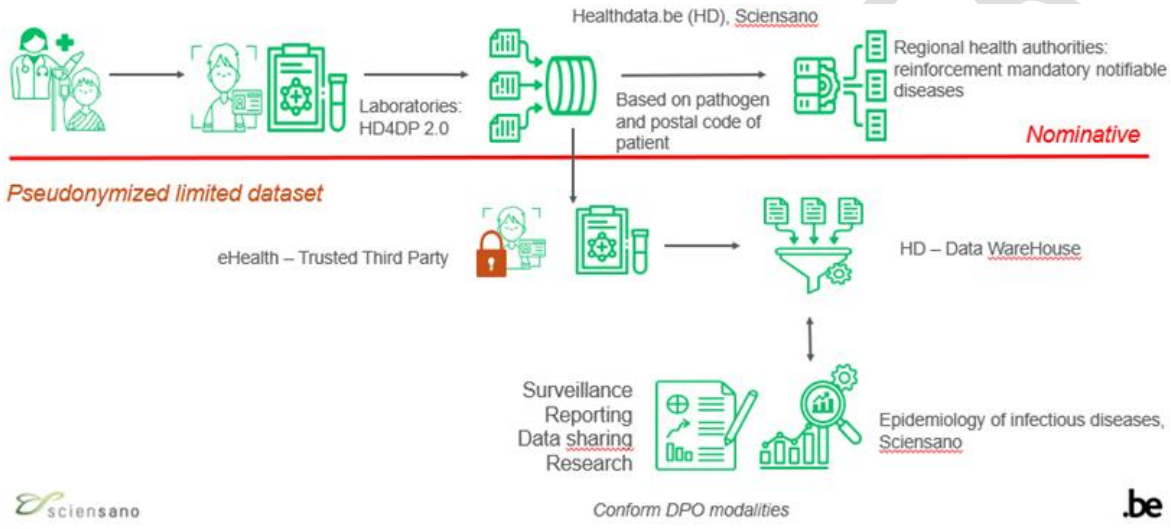


Figure 19: EPILABO (Belgium) data workflow

Currently, Epilabo has only used this application in the scope of the COVID-19 pandemic, but is gradually hoping to use it in the future for other pathogens in the scope of initiative Epilabo 2.

7.2.2.3 Epilabo 2: Improving data collection and standardisation

The goal of Epilabo 2 is to build on experience obtained during the COVID-19 pandemic for a sustainable Epilabo data collection. The project aims to:

- implement a system where data is extracted weekly from laboratory systems and sent to healthdata.be
- ensure a consistent weekly update in healthdata.be's data warehouse
- standardise laboratory data, and encourage laboratories to adopt LOINC codes
- facilitate the mapping of data to international SNOMED codes, enhancing interoperability and usability for research and analysis
- collaborate with Belgian authorities on initiatives that incentivise the use of LOINC and HL7 FHIR standards to improve interoperability across healthcare systems.

7.2.2.4 Further Integration with Healthdata.be and HD4DP

HD4DP 2.0 has yet to be fully implemented for Epilabo. During the COVID-19 pandemic, laboratories were mandated to use the infrastructure and were financially incentivized through payments tied to the number of tests reported. However, in the current context, the rollout for Epilabo requires laboratories to adapt their systems to extract data in the appropriate format and integrate with HD4DP 2.0. This process is both time-intensive and lacks the financial incentives that facilitated rapid adoption during the pandemic. However, pilot projects with two experienced laboratories will commence in January 2025 to test the feasibility of using HD4DP for Epilabo. These pilots will provide valuable insights into the practical challenges and time requirements for full implementation.

7.2.3 Clinician Feedback on HD4DP application

Interviews were conducted with clinicians and medical information officers at Belgian hospitals. All were asked to share their experience using the HD4DP 2.0 application and offer insight on suggestions for improved interoperability between data for primary use (treatment of patients) and secondary use (research, surveillance, and policymaking for Belgian authorities). The key points are summarised below:

Clinicians highlighted the purpose of HD4DP and its intended benefits for avoiding re-entry of data but also admitted that the application itself has heightened administrative stress. The key reasons for this are as follows:

1. *Data entry into the HD4DP application often takes as long as medical procedures.*

Clinicians feel that filling out the application diverts time from patient care, and they are burdened by mandatory compliance without adequate tools for implementation. A reason for this is that some hospitals prioritise care production over administrative tasks such as filling out the application. This can create a conflicting paradigm for clinicians who are required to fill out HD4DP to avoid payment delays but also are constrained by hospital mandates and the need for quick production of care. Only recently has the responsibility of data input and coding for registries become a requirement for clinicians rather than designated data providers at hospitals. Thus, there is a natural resistance to change and a need to reorganise responsibilities between the different actors of hospital information systems.

2. *Clinicians argue that inefficiency arises from re-encoding data that has already been documented elsewhere, and the application requires submission of data that clinicians do not routinely collect for primary use and treatment of their patients.*

In many hospitals, other parallel data collection systems are still in place and some registries have not yet been integrated within the HD4DP application. Additionally, some registries ask for data not commonly collected by EHRs, which makes it difficult for data providers to successfully fill in the registry. Given this, in order for hospitals and clinicians to maximise the use of the HD4DP application, it is vital for data producers to align data models and terminologies and have specific guidelines on what information is needed and how to properly encode it.

3. *Improper integration between HD4DP and electronic health record (EHR) systems forces manual data re-entry for clinicians.*

Clinicians argue for better integration of EHR systems in the HD4DP application, and a streamlined process to extract and transfer data between systems using different terminologies. A key area of concern was that misalignment of data models, terminology standards, and exchange formats between clinicians' databases, EHR software, and HD4DP leads to lack of interoperability and harmonisation of data.

7.2.4 Guidance for Improvement

While at its core, HD4DP is meant to facilitate the only once strategy and bridge the gap between primary and secondary use of data, some improvements are needed to support clinicians and data providers.

1. *Single input, multiple output*

Clinicians emphasize the importance of a unified data entry system that supports both primary and secondary data use cases. This can be facilitated by greater adoption of international standards. Many hospitals are already using SNOMED CT, LOINC, ICD-10 terminology standards, OMOP queries, and HL7 FHIR for data exchange. Universal adoption and push for international standards can support interoperability and ease data production for care providers. Ideally all data needs to be entered at the first point of contact and then seamlessly pulled into the notes across the system.

2. *Clear guidelines and support*

Greater collaboration between care providers, EHR developers, and federal authorities can streamline data collection and support both primary and secondary use cases. Initiatives such as the Belgian Terminology Center and the Community Support for Clinical terminologies (CSCT) aim to support healthcare professionals in standardisation and understanding the benefits of using international terminology such as SNOMED CT.

7.2.5 Key learnings

HD4DP exemplifies Belgium's effort to standardise and streamline the process of acquiring data from EHRs through a unified application. However, clinician feedback highlights the urgent need for greater harmonisation across systems. The fragmentation within the Belgian healthcare system could be mitigated by adopting international standards and data exchange formats, such as those proposed by the EEHRxF. With the support of Belgian authorities, and the current evidence of hospitals pushing for the use of international standards such as SNOMED CT and OMOP CDM, Belgium stands to significantly benefit from the implementation of the EEHRxF. Changes are essential to ensure that data collection systems are interoperable, prioritise patient care, and alleviate the burden on data providers.

7.2.6 Conclusion

Belgium's progress towards standardising health data systems offers both challenges and achievements. The country continues to address fragmentation by implementing national platforms and aligning with international terminology standards and data exchange formats. Efforts to centralise data collection systems, ensure semantic interoperability, and minimise data redundancy are key to improving data quality and usability.

The HD4DP application can be presented as an example of Belgium's dedication to bridging the gap between primary and secondary use. However, based on the feedback from hospitals and laboratories working with the application, it is evident that greater priority needs to be placed on using international standards and simplifying data entry for users.

7.3 Finland

7.3.1 Background and objectives

Over the past 20 years, Finnish legislation has actively supported the digitalisation of health data through various laws. Notably, in 2007, the framework for a centralised patient data archive was established, and in 2019, the Secondary Use Act was introduced to enable the effective and secure processing of personal social and health data for secondary purposes, in alignment with the EHDS legislation concerning providing a consistent, trustworthy, and efficient system for reusing health data for research, innovation, policy-making, and regulatory activities. In EHDS, citizens have a right to opt-out specifically from secondary use, in an easy and reversible way. However, for certain important public interests and under strict safeguards, including transparency requirements, data may still be used. Under Finnish law, there is no explicit right to opt out. Individuals can exercise their right to object under GDPR Article 21 by submitting their objection to Findata, the data permit authority, providing a personal justification. There are circumstances where an applicant for a data permit could override such objections. In cases where objections are overridden, individuals who opted out must be informed and given the opportunity to appeal the decision.

It is important to note that Findata is not the original controller of health and social data. Consequently, submitting an objection to Findata regarding the use of personal data does not prevent the data from being disclosed for secondary use by other controllers specified in Section 6 of the Act on the Secondary Use of Health and Social Data (Secondary Use Act).

Finland lacks a centralised system that would allow individuals to universally and conclusively prohibit the secondary use of their data in a manner binding on all parties.

It should also be emphasised that Finland's national data protection law, which complements the GDPR, requires data controllers to assess on a case-by-case basis whether it is necessary to restrict the right to object or other rights under Article 89(2) of the GDPR for a specific research project.

Healthcare services in Finland are provided by both public and private organisations. Public services are mainly financed by the state and private services by customer fees. All data collected by public health care and social welfare service providers must be connected to the national data archives. Also, private health and social service providers using an information system for the purpose of processing of client and patient data need to share their data with the central archive. The national roadmap includes moving to FHIR, adopting international terminologies and complying with the EHDS requirements.

7.3.1 Centralisation of data

7.3.2.1 The Kanta services

Since 2007 Kanta is the nationwide information system for managing patient, client and welfare data. Its use is mandatory for private health and social service providers that process client and patient data through an information system.

A key component of Kanta is the Patient Data Repository, where patient data is securely stored and easily accessed by healthcare professionals.

When a healthcare provider uses an electronic patient information system, it must be integrated with the Patient Data Repository. Kanta publishes the specifications for each individual service outlining the requirements that systems must meet to obtain certification and be authorised to send data to Kanta. A testing environment is provided by Kanta for this purpose. This centralisation allows healthcare data from both public and private providers to be collected in one place, ensuring up-to-date and easily accessible information. Patient information is entered into the provider's system and automatically stored in Kanta. Patients can access their health information via the MyKanta portal or app, and healthcare professionals' benefit from continuity of care, even when patients change providers. The archiving of imaging documents is also a feature of the Patient Data Repository, where care documents such as requests, study records, and reports are stored within the Kanta Services. These imaging care documents are connected to the images saved in the Imaging Data Repository, enabling healthcare professionals to access both the imaging results and their associated care documents in one centralised location.

In addition to patient records, and diagnostic images, Kanta includes prescriptions, and pharmaceutical product data. Recent updates in 2024 added appointment information and wellbeing data to Kanta, further expanding its scope. Discharge reports are available in Kanta as part of general health record entries, but not as separate documents in a fixed structured format.

- All the data from healthcare providers are sent to Kanta, and the information includes:
    - records of treatment given at a health centre or hospital
    - records of dental care
    - records of laboratory tests
    - records of imaging examinations
    - records of vaccinations
    - Prescription and dispensation data
    - Living wills and organ donation testaments
    - Wellbeing data and measurement results
    - Social services client data
    - Medical certificates and statements
    - Consents and denials of consent
    - Appointments

Data is automatically accessible in MyKanta as soon as healthcare providers upload it to the Kanta services.

7.3.2.2 Kanta PHR

Wellbeing data and measurement results can be entered and edited directly by citizens and constitute the Kanta Personal Health Record (PHR) national data repository. Applications can integrate with the Kanta PHR, which provides interfaces and standardised national data content based on the HL7 FHIR standard. The FHIR Rest API interfaces are detailed in the Capability Statement description.

Applications that include a back-end service operating in a trusted environment can be integrated with the Kanta PHR. The back-end service is responsible for managing communication with the Kanta PHR. Currently, applications installed directly by end users that do not connect to a back-end service are not supported.

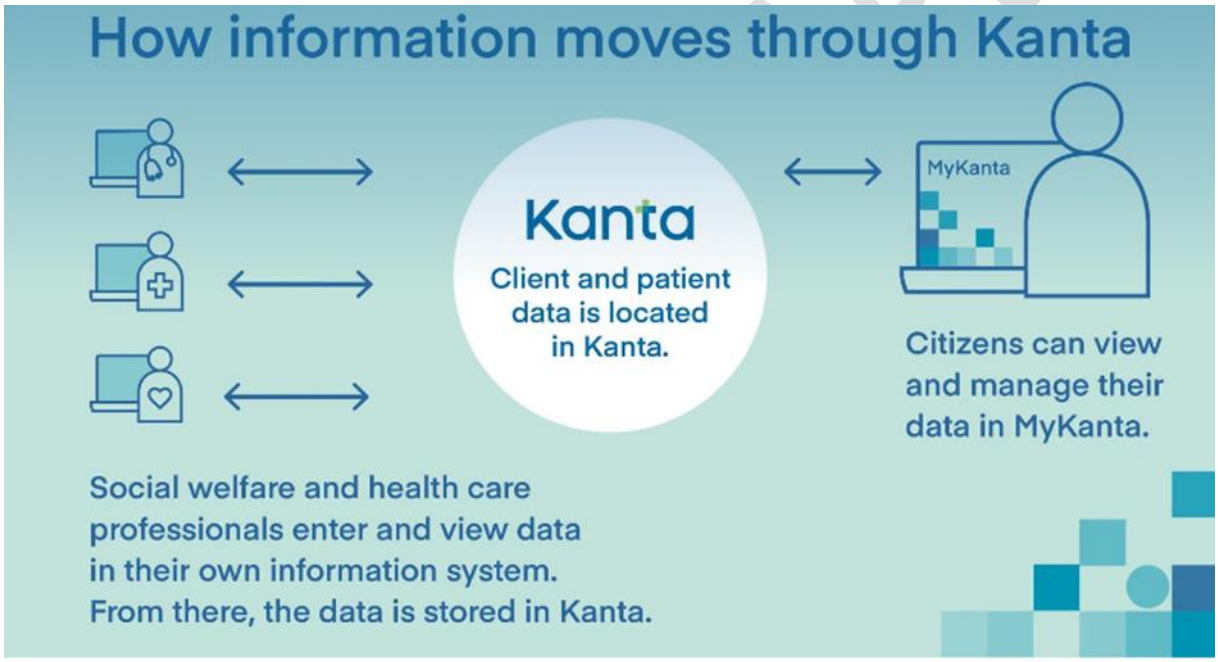


Figure 20: The architecture of the Kanta services (Finland)[80]

7.3.2.3 Connection models

Public-sector social welfare and health care service providers plus pharmacists join the Kanta Services via a direct connection. Private-sector social welfare and health care providers can choose a joint connection model where a main joining party submits an application to join Kanta and administrative burdens are shared.

In certain cases, private-sector social welfare service providers can access the Client Data Repository for Social Welfare Services through a parallel connection model. This is allowed if a public-sector service organiser, such as a wellbeing services county, grants the provider access to its client

[80] https://www.kanta.fi/en/data-in-kanta

information system. This way, private providers do not need to apply to join Kanta Services if all stored client data belongs to the wellbeing services county's client register.

Each service enabler utilising Kanta Services operates through at least one Kanta access point. An access point serves as a communication hub, connecting the service enabler's information system to the Kanta Services.

The entity responsible for implementing a Kanta access point is known as a Kanta provider. Service enablers rely on Kanta providers to establish connections between their information systems and the Kanta Services.

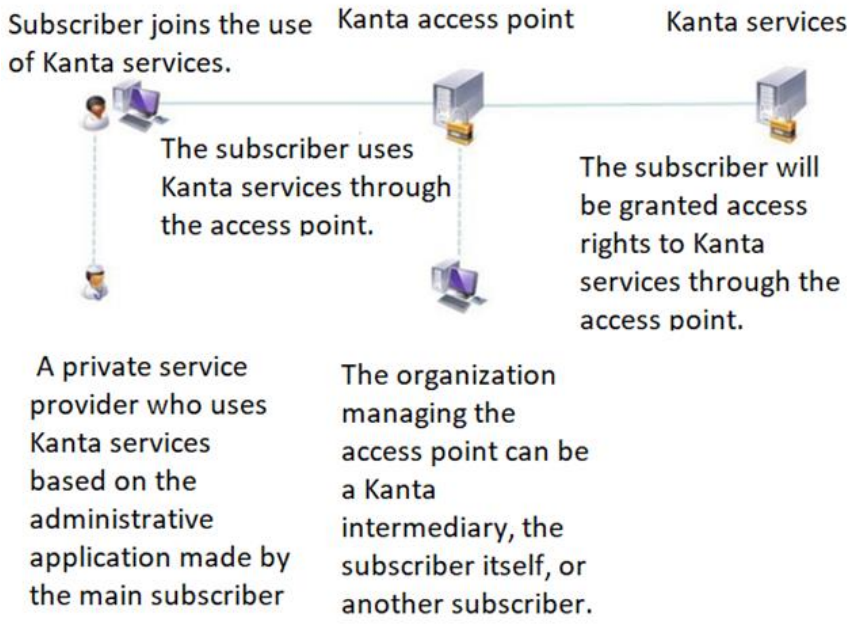


Figure 21: KANTA Access point connection[81]

Kanta providers are listed in the Provider Register maintained by the Finnish Institute for Health and Welfare. This register includes intermediaries authorised to function as Kanta providers when connecting to the Kanta Services, excluding pharmacies or health and social service organizations.

7.3.2.4 Data security

Using the Kanta Services requires healthcare and social welfare information systems to be integrated with Kanta. All social and health care providers, pharmacies, and data intermediary organisations that electronically process client and patient data must comply with Kanta's data security requirements.

Information systems, wellbeing applications, intermediary services, and other systems connecting to Kanta must undergo a data security assessment conducted by an approved information security inspection body, as stipulated by the Client Data Act.

[81] source: https://www.kanta.fi/documents/d/guest/kanta-tekniset-liittymismallit-3_16

A data security certificate, valid for up to three years, is issued upon successful completion of the assessment. The costs of the data security assessment are borne by the respective manufacturers or providers of the information system, wellbeing application, or technical intermediary service.

7.3.2.5 Software Developers

System providers are responsible for classifying their information systems correctly in accordance with the THL (National Institute for Health and Welfare) Regulation 4/2024, as the classification determines how key requirements are verified. Compliance with interoperability, data security, data protection, and functionality standards are essential, and compatibility with Kanta Services is ensured through joint testing.

Providers must notify Valvira to register their systems meeting the Client Data Act requirements, and deployment is contingent on inclusion in Valvira's information system database. Additionally, systems connected to Kanta Services must have an approved data security assessment, which the provider coordinates with an inspection body.

If multiple parties are involved, they must mutually agree on certification responsibilities, including for system entities and subsystems.

7.3.3 Services with the EHDS Regulation

Currently, Finland offers two European electronic health services:

- Cross-Border Prescription: Allows the use of Finnish prescriptions abroad and foreign prescriptions in Finland.
- Patient Summary: Enables the transfer of patient information between European countries.

The specifications for these services are available in Finnish on their respective subpages:

- Specifications for Cross-Border Prescription[82]
- Specifications for Patient Summary[83]

With the EHDS, new European health services will be gradually introduced in Finland. These upcoming services will enable the exchange of additional types of health data between European countries, including:

- Laboratory results and reports, expected around 2028.
- Medical images and related reports, expected around 2030.
- Patient discharge summaries, expected around 2032.

7.3.4 Access to data for secondary use and “only once” strategy

The Act on the Secondary Use of Health and Social Data created the framework for a centralised system, where Findata acts as data permit authority. Data from Kanta are sent to Findata which is responsible for granting permits when data is needed for secondary use.

[82] https://www.kanta.fi/jarjestelmakehittajat/rajat-ylittava-resepti
[83] https://www.kanta.fi/jarjestelmakehittajat/potilastietojen-yhteenvedon-maarittelyt

This system includes secure user environments and interfaces for data provision, ensuring strong privacy protection and safe data usage.

The primary responsibility of Findata is to provide guidance and issue permits for the secondary use of social and health data, ensuring secure processing and handling.

The issuance of permits is based on fees, which are determined by a decree from the Ministry of Social Affairs and Health on charges of work carried out by the health and social data permit authority, Findata. In addition, under the Secondary Use Act, controllers can use their own data for knowledge management without permission from Findata. In Finland the government agency THL is the main body responsible for public health. THL can access data from Kanta directly without going through Findata.

7.3.5 Standards

The standard format for the Patient repository and ePrescriptions is HL7 Version3 CDA. HL7 FHIR has been implemented in the latest developments of Kanta: the personal health record, medication list, social services disclosures, appointments and new implementation guides.

Finland has a national terminology server[84] with glossaries and vocabularies for the health services which is not based on HL7 FHIR. International codes such as ICD are used mainly for diagnoses. LOINC is not used to exchange laboratory information and proprietary coding are still used instead.

Although Finland is a member of SNOMED, the adoption of SNOMED CT is presently limited to the domain of Pathology and Problem list, it is however increasingly being adopted in other domains.

The standards ICD-10 & ICF, LOINC, ATC, ICPC-2, NCSP, MeSH, UCUM, CCC and various ISO classifications are used for nursing classification.

7.3.6 Challenges

Despite the advanced digital service infrastructure and comprehensive health data legislation, certain challenges persist and require careful consideration.

- The increased number of mandatory data fields in the information system required by Kanta coupled with the selection of appropriate codes from the structured information format weighs on the administrative burden of healthcare providers.
- Information systems often lack usability and sufficient analysis and search capabilities to fully leverage structured data, which diminishes physicians' motivation in providing the information.
- While a significant amount of data is being collected, the focus should now shift to making effective use of it.
- System developers need to face the economic burden of adapting to the common (evolving) specifications and of obtaining conformance certifications

[84] https://koodistopalvelu.kanta.fi/codeserver/pages/classification-list-page.xhtml?clearUserCachedLists=true

- Accessing data for research through Findata has become increasingly costly and subject to longer waiting times
- In Finland, general practitioners are assigned randomly, which often hinders continuity of care. Integrating information systems with patient lists could help doctors maintain their own list of patients, thereby improving continuity.

7.3.7 Future developments

Findata is actively working to improve its service and functionalities. To this purpose, it has launched a four-year project, FinHITS, co-funded by the European Union that aims to strengthen Finnish health data ICT for secondary use and to enable Finland's seamless integration into the EHDS.

The Kanta Services are also undergoing a modernisation process shifting from document-based HL7 v3 CDA structures to an information model-based framework based on HL7 FHIR. The transition will occur gradually, but the newest contents are already compliant with the FHIR standard: booking healthcare appointments, managing notifications of social welfare disclosures, and sharing wellbeing data stored in MyKanta with healthcare and social welfare professionals.

There is also a plan to adopt a CSIRO (Commonwealth Scientific and Industrial Research Organisation) HL7 FHIR-based terminology server. It is probable that proprietary terms will continue to be used together with international terminologies.

Additionally, efforts are underway to integrate patient lists into information systems, thereby enhancing continuity of care.

7.3.8 Key learnings

- Finland's healthcare system is constitutionally mandated and is based on the social welfare and healthcare services offered by the 21 counties.
- Finland boasts comprehensive, high-quality information resources and advanced digital service infrastructure, anchored by the national centralised Kanta services system.
- Legislation was fundamental for Finland to achieve harmonisation and centralisation of health data.
- Data from Kanta are automatically available for secondary use via the centralised national entity Findata.
- Finland has a national terminology server which includes proprietary terms as well as international ones.
- Information systems developers must comply with the necessary specifications to connect their applications to the Kanta services.
- Despite the advanced legislation and underlying infrastructure, there is significant criticism from physicians and researchers regarding the difficulty of using the collected data.
- Criticism also concerns the difficulty of expressing thoughts within predefined structured data codes.

- Information systems often lack usability and sufficient analysis and search capabilities to fully leverage structured data, which diminishes physicians' motivation in providing the information.
- Efforts to improve Kanta and Findata services as well as to switch to HL7 FHIR and international terminologies are in place.

7.4 Valencian Community (Spain)

Abucasis is an integrative project initiated by the Generalitat Valenciana's Health Department, aiming to establish a unified electronic health record (EHR) for patients receiving ambulatory care across the Valencian Community. It is a centralised system that connects primary care centres, hospitals, and specialised units, allowing comprehensive management of patients' health information. The system enables healthcare professionals to access patient histories from any outpatient care point, including health centers and hospital outpatient consultations.

7.4.1 Origin and Objective

The Abucasis electronic health record (EHR) system has significantly transformed healthcare delivery in the Valencian Community, Spain. Implemented in 2006, it serves as the primary EHR for both primary and specialised outpatient care, achieving 96% population coverage by 2009[85]

Developed by Indra, Abucasis unifies ambulatory medical records across the region, encompassing electronic records for 98% of Valencia's population. It supports all primary care centers and outpatient services, benefiting over 15,000 daily users[86].

The Abucasis Project has undergone and continues to undergo continuous modifications in its configuration, but in general terms it was created to respond to a series of objectives that we detail below (https://healthgroup.es/medico-en-valencia-abucasis/) :

1. Establishment of a single Clinical History for each patient treated on an outpatient basis in the Valencian Community.
2. Access to the Clinical History from any healthcare point in the outpatient healthcare network (Health Centres and Outpatient Clinics of Hospitals and Specialty Centres).
3. Integration with other existing information systems (especially relevant for this working paper):
    - SIP (Population Information System)
    - GAIA (Health Care Management)
    - RVN (Vaccination Registry)
    - IRIS/HIGIA (Hospital HIS)
    - Future Systems (Metabolic Disorders, Food Hygiene, etc.)

[85] https://pmc.ncbi.nlm.nih.gov/articles/PMC7394961/
[86] https://www.indracompany.com/en/noticia/indra-puts-autonomous-region-valencia-fore-front-digital-health?

7.4.2 Key issues from a technical and end-user perspective.

Architecture and Technology

- Web-based architecture for interoperability between different modules
- RESTful web services for communication with the central database
- Standard technologies: HL7 for information exchange (HL7 v2.6 + HL7 FHIR, CDA), ICD-10 for diagnostic coding, nursing diagnoses coded (NANDA) , nursing interventions coded (NIC), etc.
- Java-based application layer with Oracle database backend

Abucasis is a component of the Valencia Health System Integrated Database (VID), a comprehensive set of public, population-wide electronic databases (see Figure 22). VID includes hospital discharge records, emergency care discharge records, birth registries, and more, supporting extensive healthcare data analysis and research. **(Catálogos EMA).**

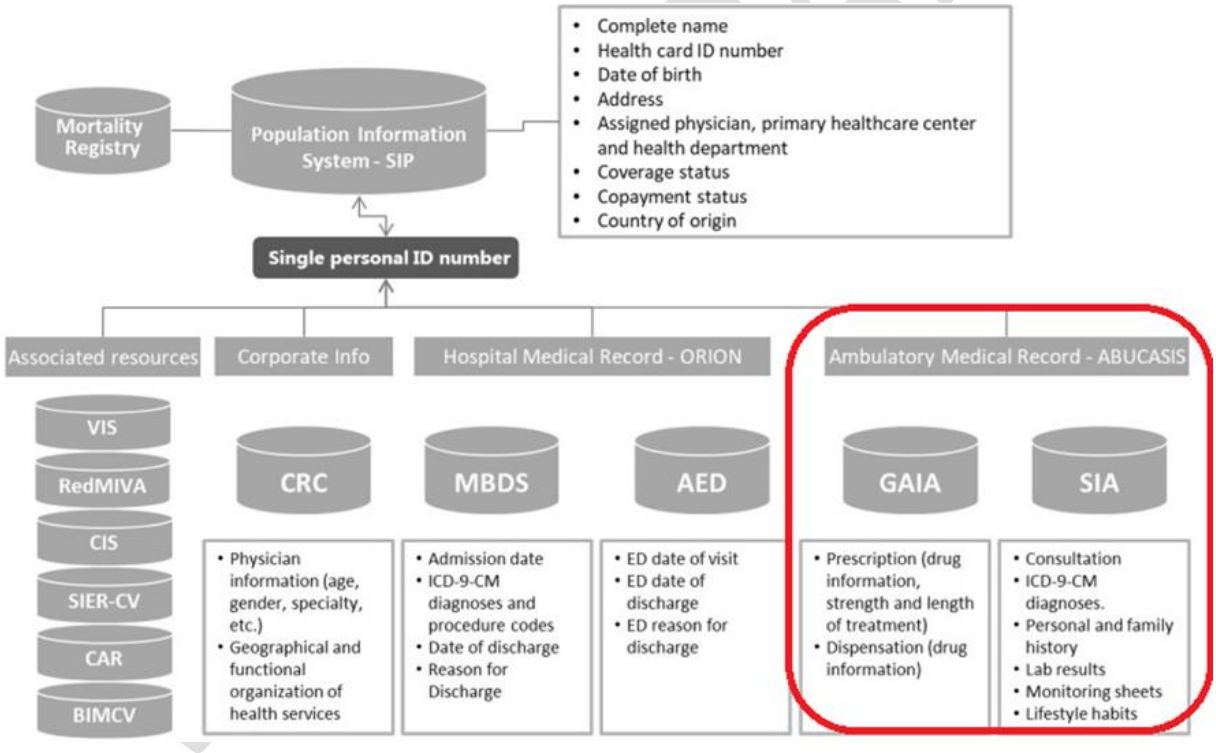


Figure 22/ The Valencia Health System Integrated Database (VID).[87]

It is important to highlight that VID is an active member in the DARWIN EU® project. This participation demonstrates the VID's commitment to leveraging real-world data for secondary use in research and regulatory decision-making at the European level.

[87] Source: García-Sempere A, Orrico-Sánchez A, Muñoz-Quiles C, Hurtado I, Peiró S, Sanfélix-Gimeno G, Diez-Domingo J. Data Resource Profile: The Valencia Health System Integrated Database (VID). Int J Epidemiol. 2020 Jun 1;49(3):740-741e. doi: 10.1093/ije/dyz266. PMID: 31977043; PMCID: PMC7394961.87

Abucasis Key Features:

- Centralised patient records
- Real-time access to medical information
- Appointment management
- Electronic prescriptions
- Integration with laboratories and pharmacies

Data Processing and Security

ABUCASIS processes clinical histories, diagnostic codes, and prescription data, while ensuring data security and compliance with healthcare regulations. The system is designed to comply with GDPR and allows integration with other healthcare systems across Spain.

The implementation of Abucasis II has led to the digitalisation of a significant portion of the tasks performed by medical and healthcare personnel. To ensure both the legal validity of the documents involved and the security of electronic transactions, the system incorporates advanced measures.

It integrates the use of digital certificates issued by the Certification Authority of the Valencian Community, alongside advanced electronic signatures at every stage of medical actions. These recognised certificates are provided on secure devices, such as cryptographic cards (Abucasis Card), ensuring compliance with the required security standards and safeguarding sensitive patient information.

This robust security framework not only reinforces trust in the system but also aligns with national and EU regulations for digital healthcare platforms (https://healthgroup.es/medico-en-valencia-abucasis/).

Key constraints to the users

ABUCASIS II imposes several key constraints on its users to ensure data quality, standardisation, and compliance:

1. Semantic standardisation: Users must adhere to standardised medical codes and terminologies such as ICD-10 for diagnostics, Clinical Risk Groups (CRG), etc.
2. Structured data entry: The system requires input of structured data, including diagnostic codes, treatment plans, medication history, and demographic details.
3. Access controls: Different user roles (e.g., healthcare providers, administrative staff) have specific access levels to protect patient privacy and ensure data security.
4. Mandatory fields: Certain data fields are required to ensure complete and accurate patient records, for example, the SIP (Population Information System).
5. Compliance with regulations: Users must follow protocols that ensure compliance with data protection regulations like GDPR.

6. Interoperability standards: The system uses HL7 FHIR standards, which must be adhered to for data exchange and integration with other healthcare systems.
7. Training requirements: Users are expected to undergo training to use the system effectively, although the frequency and quality of this training have been reported as insufficient.
8. Data entry protocols: Healthcare professionals are required to follow specific protocols for data entry to maintain consistency and quality across the system.

These constraints aim to ensure data accuracy, patient privacy, and system efficiency, but they also present challenges for users, particularly when changes are implemented without adequate notice or training.

The integration of Abucasis II with other existing information systems (Figure 23), such as the Population Information System (SIP), GAIA (Healthcare Provision Management), and the Vaccine Registry, further emphasises the importance of adhering to these constraints to maintain a cohesive and efficient healthcare information infrastructure. In addition, the medical history of any user is associated with their SIP number, which guarantees a unique and univocal history throughout the healthcare field of the Valencian Community.

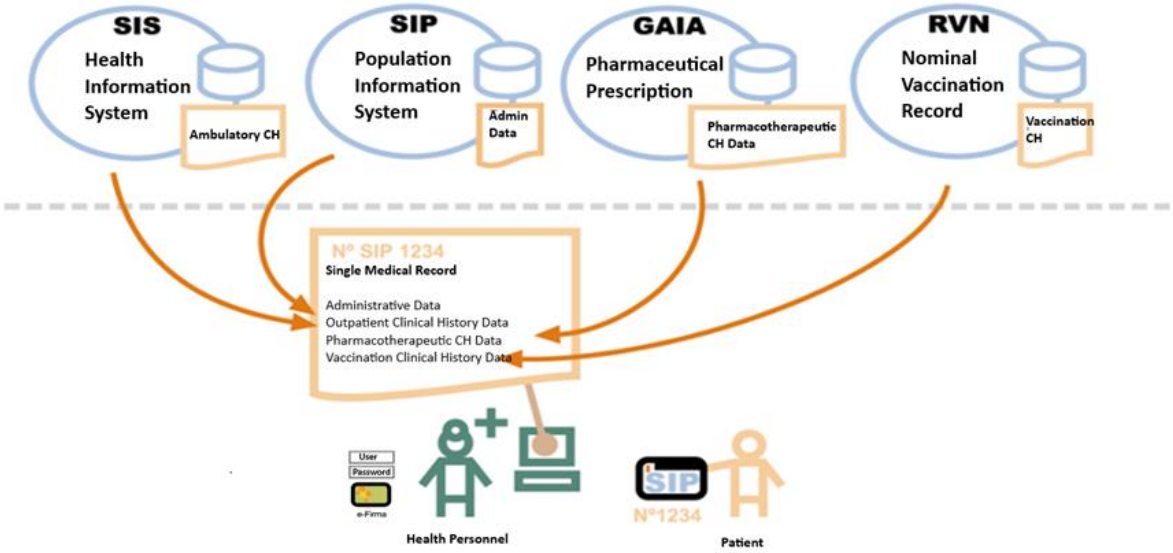


Figure 23: : Integration of Abucasis [88]

Healthcare personnel can seamlessly access various corporate systems from a single platform, ensuring transparent navigation between applications. This access is facilitated using their credentials.

[88] *Source: 13:30 El proyecto Abucasis II en la Generalitat Valenciana*

Data Sharing Requirements

The sharing of data within Abucasis appears to be largely compulsory, given its role as a centralised EHR system for the Valencian Community. However, the level of obligation may vary depending on the type of data and their intended use.

Compulsory Data Sharing:

1. Clinical Histories: Healthcare providers are required to input and share patient clinical histories within the system
2. Prescription Data: Medication and prescription information must be entered into the system to maintain a comprehensive patient record
3. Lab Results and Imaging Data: These are required to be shared within the system to ensure a complete patient record.

Potentially Optional Data Sharing:

1. Administrative Information: While billing and appointment scheduling data are captured in the system, it's not clear if sharing this information is mandatory for all healthcare providers.
2. Electronic Communications: Discharge summaries and referral letters are included in the system, but the level of obligation for sharing these may depend on specific regional policies.

Data Use and Access

1. Healthcare Providers: Access to patient data during treatment for diagnosis and care decisions is compulsory.
2. Regional Health Authorities: Access to anonymised data for benchmarking and quality control is compulsory.
3. National Health Reporting: Sharing of anonymised data on diagnostic codes, treatments, and prescriptions with national healthcare authorities for reporting and research purposes is mandatory.
4. Patient Access: Patients have the right to access their own medical data, including diagnosis, treatment plans, and clinical histories.
5. Researchers: Access to the data may be requested by any researcher (providing the corresponding documentation required) from the Valencia Health System Data Commission .

Databases and registers

The VID, in which Abucasis is integrated, includes sociodemographic and administrative information (sex, age, nationality, etc.) and healthcare information such as diagnoses, procedures, lab data, pharmaceutical prescriptions and dispensations, hospitalisations, mortality, healthcare utilisation and public health data. It also includes a set of specific associated databases with population-wide information on significant care areas such as cancer, rare disease, vaccines and imaging data.

Therefore, a large set of information is linked and accessible through a Single Personal Identification Number in the Valencia region.

End-User perceptions

The interview (annex) highlighted that end-users, particularly healthcare professionals, have generally received the system positively. The implementation has led to significant improvements in organisation, data storage, and accessibility of patient information.

The benefits Perceived have been:

- Better global view of patient history.
- Improved access to patient data.
- Enhanced use of patient information.
- Streamlined decision-making process.

And the challenges Identified:

- Insufficient training and information about system updates.
- Limited time given for system adoption.
- Lack of involvement of healthcare personnel in system improvements.

Related to the User Experience and Support, the system is used daily, with healthcare professionals spending 7 to 24 hours (during on-call duty) interacting with it. While it initially slowed down workflows, users recognize long-term benefits in patient care.

Initial training was provided during the system's implementation. However, ongoing training and support have been limited, with users reporting a lack of information about system changes and updates.

Some suggestions for improvement were:

- Involve medical and nursing staff in system advisory roles
- Provide more notice and training for system changes
- Improve the effectiveness and efficiency of the system based on user feedback

In conclusion, while Abucasis has significantly improved healthcare data management and patient care in the Valencian Community, there is room for improvement in user training, support, and involvement in system development. The level of current constraints on the data provider is relatively low, leading to only a limited number of FIAR data available for secondary use. The local authorities expressed their interest in aligning with EU EEHRxF formats.

7.4 Summary of current use cases and registries possibly connected with EEHRxF

As evidenced in the previous sections, there are different ways to look at the public health use cases which can benefit from the implementation of the EEHRxF. Likewise, the EEHRxF can take inspiration from other initiatives, to help guide its implementation.

With the upcoming EHDS, most countries have prioritised the creation of a dedicated infrastructure for the secondary use of data under the responsibility of a dedicated Health Data Access Body. However, only a limited number of MS have already identified strategies to better integrate data flows in the primary and secondary use domains.

All finalities are meant to converge, but based on the history of "only once" best practices documented in this DEV, the following original finalities have been identified:

- Creating an integrated big data infrastructure BOTH for primary and secondary use (Finland, Valencia)
- Reducing administrative burden for data producers (Belgium)
- Providing added value to end-users (Denmark)

In the latter two cases, the approach prioritised achieving the most efficient "Return on Investment", minimising user effort while maximising benefits across the entire value chain. Use cases were often designed to build upon each other creating a virtuous cycle where drivers are progressively aligned.

Another approach is to examine use cases based on specific required functionalities, following the strategy of WP5 -European EEHRxF in Clinical Research- with the IPS+R, which focuses on patient discovery. This approach is also pertinent for Public and Population health as exemplified in D.2.2.: EEHRxF in a nutshell.

Aside from the focus on finalities and functionalities, the third approach is to focus on priority thematic areas and their associated data sets. In this context, the issue of time is particularly relevant. Public health use cases that require "near real time" data availability of data should be prioritised. While the COVID-19 crisis served as an important catalyst, use cases related to new public health patterns of prevention and prediction are still emerging.

Analysis of past and ongoing EU initiatives helped in the identification of additional use cases where the EEHRxF and xShare could have a role in improving public health outcomes. In general, the EEHRxF helps with a transition from real world data to secondary uses. This primarily pertains to enhancing public health surveillance through better EHR data interoperability. Prominent use cases here are antimicrobial resistance, and registries for non-communicable diseases including cancer. This is extremely relevant since both relate to global health threats, which also constitute the focus of global health policies.

The working paper also outlined the potential of the EEHRxF to support a flow of higher quality interoperable real-world data – primarily related to EHR data (pharmaceutical strategy), supporting data discoverability, and linking between clinical trial data and real-world models. Lastly, EEHRxF could also help facilitate EHR data use for tasks like healthcare provider capacity monitoring.

The EU initiatives also illustrate areas which can inspire xShare and populate its business use case (BUC) registry. The project could leverage country-level reports, such as those published in TEHDAS1, to gain Member State level insights for its BUCs including on coding systems. It should build upon existing mappings of public health data registries, terminologies and data models such as those for infectious diseases. Additionally, BUCs should stay informed about future developments in specific data types, including genomic data, antimicrobial data (especially in the context of the One Health approach), medication identification, and bio sample data. BUCs can also draw valuable lessons from established research data flows within other public health research infrastructures, as well as from identification of emerging data needs.

Evidently, xShare can benefit, advance, and learn from ongoing EU-level efforts - both in relation to specific public health topics and from a more a process-oriented perspective

7 Next Steps

The information contained in this working paper provides a global overview of the dynamic present in the health data secondary use ecosystem and describes real life implementations to better connect primary and secondary use. It also identifies the public health domains that would benefit most from closer alignment with specifications implemented in systems used by clinicians, particularly those that collect data in real or near-real time. It also shows that integrating the primary and secondary use ecosystems in terms of organisation and technical aspects is essential for health threats monitoring as demonstrated by the ATHINA platform created in the HERA-supported EU-HIP project.

The upcoming working paper (D.4.2) will focus on key datasets which are currently considered a priority by different European Initiatives and show the need for a clear organisational and technical connection between primary and secondary use of health data. It will also document in detail the use cases proposed by the three xShare WP4 pilots (Charite, Sciensano and Monasterio). It will also consider more prospective use cases such as the one proposed by the University of Valencia on traffic accidents and the one described in D.2.2 which reflects the potential use of the xShare Yellow Button in the context of Public Health. The working paper will be structured in such a way that that it can provide a direct contribution to the X-Bundle registry.

WP4 will also collaborate with WP5 in the context of the IPS+R specifications as the use cases supported by IPS+R, such as patient discovery, are similarly relevant for population health.

The contacts established with different relevant ongoing initiatives including the Joint Actions mentioned under section 4 will also be sustained with the objective to share knowledge, define use cases priorities and collect further inputs.

8 Conclusions

Until recently, the collection of data for public health finalities such as benchmarking, knowledge management, health system management, policy evaluation and health threats monitoring was performed in closed silos. Often ad hoc data collection and data mobilisation processes were employed, without adhering to internationally set standards. Furthermore, there was hardly any consolidated and structured inventory of the existing datasets. The use of the collected data was often restrained to a single finality and to the same few data consumers. The multiplication of those data collection processes led to a paradox: while structured and coded data were made available to PH registers, data produced on the point of care by clinicians were still largely in free text. In instances where semantic standards were used, they often did not align with the specific needs of the task. The proliferation of data collection requests further led to significant frustration among clinicians, who felt that valuable time and resources were being diverted from patient care without clear added value. The Covid-19 crisis served as a wake-up call, convincing Member States of the urgent need for more direct and efficient access to data. At the same time, the rapid advancements of AI offer significant new opportunities to support the FAIRification of data, advanced analytics and evidence-based decision making. With the adoption of the EHDS regulation, progress is being made, including the establishment of dedicated health data agencies, which provide a strong push in the right direction. However, the availability and accessibility of FAIR, interoperable health data also require a more seamless connection between primary and secondary use of data. This will allow data produced at the point of care to be reused for secondary purposes with minimal effort

The work outlined in this working paper leads us to the following preliminary conclusions:

- Primary and secondary use are still considered by many Member States as two different ecosystems and reflection on how to better connect them is still in its early stages.
- A significant number of national and European projects and initiatives have been launched to help with the FAIRification of data for secondary use in priority domains of infectious diseases, cancer and AMR. Other initiatives have also been developed to standardise and share specific data types such as genomics, images or biomarkers, across one or multiple domains.
- Some of the minimum categories of health data in article 51 of the EHDS regulation that should be made available for secondary use are clearly connected to primary use: this is particularly the case for EHR data (globally) and data produced by the patient (or a patient facing app) related to life habits, wellbeing, environment, PREMs and PROMs, and potentially social domains-related data. The EEHRxF and the xShare yellow button provide a strategic opportunity to connect primary data produced by both clinicians and citizens to priority public health use cases. The objective to have a wide adoption of the HealthDCAT-AP standard to describe datasets is an important first step in aligning standards and coding systems.
- Real time (or nearly real time) data collection for secondary use is currently limited to infectious diseases monitoring (i.e. ECDC's TESSY). With the support of HERA, a European health threats data monitoring is in the making.
- The alignment between data models and semantic standards remains in its early stages. While some progress has been made, there is currently no immediate miracle solution. Main enablers today include the availability of accessible tools, such as national and European terminology servers that offer resource mapping the adoption of interfaces with robust

machine learning capabilities and the establishment of highly skilled and diverse health data teams within organisations.

- The four countries "only once" best practice reports provided valuable insights. While holistic initiatives such as the one implemented in Finland are very promising, as they establish the legal foundation for centralising data and services, they currently face challenges in addressing practical aspects such as user feedback on usability, clinician motivation, and the resources required for data reuse. An initiative aimed at consolidating datasets for secondary use, such as the one developed in Belgium, promised to reduce administrative burden and improve data quality by leveraging existing data in EHR systems. However, implementing such a system without fully consolidating all relevant registers and significantly improving EHR systems may limit the expected benefits and lead to frustration. Nonetheless, these efforts are expected to yield positive results in the medium term. Countries that had established voluntary sentinel systems for infectious diseases are now able to consider achieving full universal coverage. Meanwhile, initiatives that began with a more limited bottom-up approach, such as in Denmark, focused initially on use cases with direct added value for users and minimal constraints before expanding to include broader public health outcomes. However, Denmark has faced challenges due to the historically slow adoption of international standards. In this context, the role played by industry, particularly through the development of a collaborative connector, is also noteworthy.

  From this analysis of best practice use cases, it can be concluded that the buy-in of data producers is best acquired when the added value brought by the data workflows is multi-fold and different priority expectations can be met.

  For countries which already have in place direct connection processes between EHR systems and PH Datasets, the availability of the EEHRxF specifications is expected to bring greater fluidity, improved efficiency and support a significant number of new use cases.

Annexes

To limit the size of this working paper, we are providing hereunder the links to the annexes.

Denmark full only once best practice report

Best Practice WP4 report SENTINEL DK final.pdf

Belgium full only once best practice report

Best Practice Report Belgium.pdf

Finland full only once best practice report

Best Practice WP4 report Finland.pdf

Valencia region (Spain) only once best practice report

Best Practice WP4 report Valencia.pdf