Building the Largest Federated Health Data Network in Europe: Advancing Collaboration and Real-World Evidence Generation

Clair Blacketer, Martijn J Schuemie, Maxim Moinat, Erica A Voss, Montse Camprubi, Peter R Rijnbeek, Patrick B Ryan

Submitted to: Journal of Medical Internet Research
on: March 18, 2025

Table of Contents

Building the Largest Federated Health Data Network in Europe: Advancing Collaboration and Real-World Evidence Generation

Clair Blacketer[1, 2, 3] MPH; Martijn J Schuemie[1, 3, 4] PhD, MS; Maxim Moinat[1, 2] MS; Erica A Voss[1, 3] PhD, MPH; Montse Camprubi[1, 5] MS; Peter R Rijnbeek[1, 2*] PhD, MS; Patrick B Ryan[1, 3, 6*] PhD, MS

[1] OHDSI Collaborators New York US
[2] Erasmus MC Rotterdam NL
[3] Johnson & Johnson (United States) Raritan US
[4] University of California, Los Angeles Los Angeles US
[5] Synapse (Spain) Madrid ES
[6] Columbia University New York US
[*]these authors contributed equally

Corresponding Author:
Clair Blacketer MPH

Johnson & Johnson (United States)
900 US Route 202
Raritan
US

Abstract

Background: The European Health Data & Evidence Network (EHDEN) is the largest federated health data network in Europe, designed to harmonize diverse real-world data (RWD) sources to the OMOP Common Data Model (CDM). By addressing challenges in data interoperability, governance, and quality assurance, EHDEN enables large-scale, high-quality real-world evidence (RWE) generation to support research and regulatory decision-making.

Objective: This paper provides an overview of the EHDEN network, examining its data harmonization efforts, quality control processes, and the range of data sources included in the network.

Methods: EHDEN recruited data partners across Europe through structured calls, providing financial and technical support for data harmonization. A network of certified small-to-medium enterprises (SMEs) assisted data partners in the extract, transform, and load (ETL) process. Data quality was systematically assessed using standardized tools, including the Data Quality Dashboard (DQD). Data sources were categorized by country, care setting, data capture method, and patient inclusion criteria.

Results: As of September 2024, EHDEN has harmonized 210 data sources from 30 countries, representing over 450 million individuals. The network encompasses diverse healthcare settings, including primary, secondary, and mixed care, and integrates various data capture methods such as electronic health records, claims, and registries. The standardized data have been used in large-scale observational studies, including research on medicine shortages and regulatory initiatives like DARWIN EU.

Conclusions: EHDEN has transformed the European RWD landscape, making standardized, high-quality data accessible for research and regulatory purposes. The transition to the EHDEN Foundation marks a new phase, emphasizing evidence generation and network sustainability to further enhance the utility and impact of real-world data across Europe.

(JMIR Preprints 18/03/2025:74119)
DOI: https://doi.org/10.2196/preprints.74119

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Building the Largest Federated Health Data Network in Europe: Advancing Collaboration and Real-World Evidence Generation

Clair Blacketer[1,2,5], Martijn J. Schuemie[1,5,6], Maxim Moinat[1,2], Erica A Voss[1,5], Montse Camprubi[1,3], Peter R. Rijnbeek[1,2]*, Patrick B. Ryan[1,4,5]*

1. OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA
2. Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, NL
3. SYNAPSE Research Management Partners, Madrid, ES
4. Department of Biomedical Informatics, Columbia University, New York, NY, USA
5. Johnson & Johnson, Raritan, NJ, USA
6. Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA

*co-last authors

Word Count: 2998
Keywords: Health Informatics, Epidemiology

Corresponding Author
Clair Blacketer
Observational Health Data Analytics
Johnson & Johnson
920 US Route 202
Raritan, NJ, 08869, USA
Telephone: +1 757-334-5788
Email: mblacke@its.jnj.com
ORCID ID: 0000-0003-2570-2124

Abstract

Background
The European Health Data & Evidence Network (EHDEN) is the largest federated health data network in Europe, designed to harmonize diverse real-world data (RWD) sources to the OMOP Common Data Model (CDM). By addressing challenges in data interoperability, governance, and quality assurance, EHDEN enables large-scale, high-quality real-world evidence (RWE) generation to support research and regulatory decision-making.

Objective
This paper provides an overview of the EHDEN network, examining its data harmonization efforts, quality control processes, and the range of data sources included in the network.

Methods
EHDEN recruited data partners across Europe through structured calls, providing financial and technical support for data harmonization. A network of certified small-to-medium enterprises (SMEs) assisted data partners in the extract, transform, and load (ETL) process. Data quality was systematically assessed using standardized tools, including the Data Quality Dashboard (DQD). Data sources were categorized by country, care setting, data capture method, and patient inclusion criteria.
Results
As of September 2024, EHDEN has harmonized 210 data sources from 30 countries, representing over 450 million individuals. The network encompasses diverse healthcare settings, including primary, secondary, and mixed care, and integrates various data capture methods such as electronic health records, claims, and registries. The standardized data have been used in large-scale observational studies, including research on medicine shortages and regulatory initiatives like DARWIN EU.
Conclusion
EHDEN has transformed the European RWD landscape, making standardized, high-quality data accessible for research and regulatory purposes. The transition to the EHDEN Foundation marks a new phase, emphasizing evidence generation and network sustainability to further enhance the utility and impact of real-world data across Europe.

Introduction

Real-world data (RWD) has become a cornerstone in healthcare research, especially in regulatory science, due to its ability to capture insights from diverse patient populations and clinical settings. Unlike data generated through traditional randomized controlled trials (RCTs), which often have stringent inclusion criteria, RWD reflects the everyday healthcare experiences of a broader patient base[1–5]. This breadth offers a richer context for understanding drug safety and effectiveness, guiding post-authorization safety monitoring, informing risk-benefit evaluations, and supporting regulatory decisions[6]. Regulators, industry, and academics alike rely on real-world evidence (RWE) derived from RWD to answer critical questions about healthcare interventions in clinical care settings that are more representative of routine practice[7–9].

Europe's healthcare landscape presents both challenges and opportunities for generating RWD[10]. Its diversity spans many different health systems, terminology systems, and data collection practices, with variability in healthcare delivery and data availability across countries. This heterogeneity complicates large-scale representative research but also offers a unique opportunity to study diverse populations[8,11–13]. However, capturing this potential requires overcoming technical, operational, and methodological barriers to ensure data harmonization and quality.

Federated data networks, like the European Health Data & Evidence Network (EHDEN), are well-suited for Europe's decentralized data landscape[14–19]. In a federated network, data remains within each local database and is analyzed in place, preserving the autonomy and governance policies of individual data holders. This approach allows for efficient multi-database studies across diverse populations without compromising patient privacy and data security. It facilitates compliance with Europe's stringent data protection standards, which is essential for building trust and enabling collaboration across borders.

EHDEN was established as an Innovative Medicines Initiative (IMI), now Innovative Health Initiative (IHI) public-private partnership in November 2018 to overcome the challenges and transform how health data is utilized in Europe[20,21]. The project built a federated data network that standardizes health data across participating sources, making data analysis more feasible and consistent. By harmonizing data and implementing quality assurance protocols, EHDEN enhances the usability and comparability of real-world data across Europe. This paper provides an overview of the EHDEN network, examining its data harmonization efforts, quality control processes, and the

range of data sources included in the network. Through this discussion, we aim to highlight the scope of RWD available across Europe and its potential for advancing healthcare research and regulatory decision-making.

Methods

Common Data Model

As the foundation for its network, EHDEN adopted the OMOP (Observational Medical Outcomes Partnership) Common Data Model (CDM)[22–24]. The OMOP CDM is widely recognized for its "structure + content" approach whereby the tables and fields (structure) as well as the vocabulary (content) are standardized, allowing for integration of data across multiple systems while maintaining data integrity. The model also supports a wide range of data types, including electronic health records (EHR), claims data, and patient registries.

The OMOP CDM is maintained by the Observational Health Data Science and Informatics (OHDSI) community, an open science effort that aims to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care[25]. The open-source nature of OHDSI allows for continuous community-driven improvements, making it adaptable to emerging healthcare needs[26–28].

Data Partner Calls

Any organization with access to a data source in Europe could apply to be included in the EHDEN network. In this context, a data source is defined as a distinct repository of healthcare-related data pertaining to a specific set of individuals. Except for the COVID-19 Rapid Collaboration Call, the 7 data partner calls executed between September 2019 and October 2022 were aligned to similar timelines for data partner identification, grant awarding, and initiation of data harmonization (figure 1). In each call, candidate partner organizations with access to one or more electronic healthcare databases applied to the EHDEN Harmonization Fund for a grant to implement or enhance their database (supplemental 2). Data Partners (DPs) were selected based on 3 criteria: data impact (size, coverage, quality), network impact (track record, uniqueness within network) and readiness (willingness to participate, governance) (supplemental 1), reviewed by a Data Source Prioritization Committee (DSPC). Each application was reviewed and scored by two reviewers and the top applicants per round were awarded a grant.

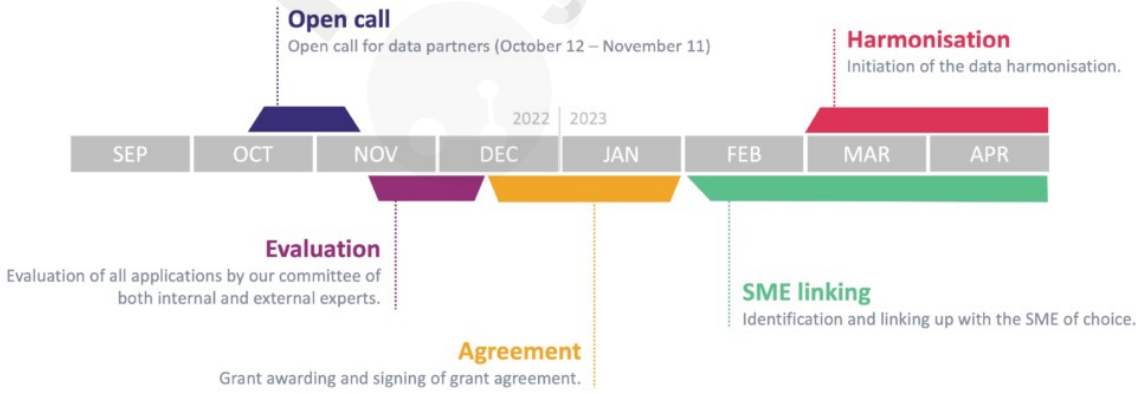

Figure 1: Data Call 7 timeline

Data Standardization

Once DPs were identified and grants awarded, each data source underwent standardization to the OMOP CDM. A crucial factor in EHDEN's long-term sustainability and success was the recruitment and training of local small-to-medium enterprises (SMEs). These SMEs were brought on board through separate calls from the EHDEN consortium and certified via the EHDEN Academy education program concluded by an on-site/online training. SMEs played a pivotal role in supporting data partners throughout the extract, transform, and load (ETL) process by providing guidance and expertise. In total 64 SMEs across 22 countries were certified by EHDEN to support DPs (see www.ehden.eu for more details about these SMEs).

The ETL process followed by the data partners and supported by the SMEs was largely uniform, as outlined by Voss et al., and involved four key steps: 1) summarizing the native data, 2) creating the ETL specification, 3) mapping source vocabulary codes, and 4) implementing the ETL[24]. This standardized approach ensured transparency in the followed procedure and adherence to the conventions in converting data sources to the CDM, while also allowing DPs to benefit from the SMEs' specialized knowledge.

Payments were structured based on output; to receive full funding DPs were required to meet 3 different milestones. The ETL specification document entitled data partners to 30%, ETL implementation and infrastructure released the next 40%, and the final 30% was received by the data partner after final inspection of the harmonized data (supplemental 3).

Data Quality

Each milestone was reviewed by an EHDEN consortium member who was part of the Milestone review committee. The ETL specification document required by milestone 1 was evaluated to ensure the mapping adhered to the OMOP CDM conventions and that the data partners/SMEs had a good understanding of the CDM and their own native data[23,29]. Milestone 2, the ETL implementation, had multiple review steps. The infrastructure was investigated to be sure the data partners were using a supported database platform[1]. The vocabulary mapping was evaluated to ensure most, if not all, source codes were included. The Data Quality Dashboard (DQD) was developed by EHDEN Work Package 5 to provide a standard structure for quality assurance[30]. It was utilized by data partners throughout the ETL process to continually improve the standardized data sources[31]. In milestone 3, final DQD results as well as the CDM Inspection Report were reviewed[32]. Once approved, the data partner then entered their information into the EHDEN portal, an online platform open to the public designed to catalog metadata on each data source [33].

Analyses

An individual data source was considered one entry in the EHDEN portal. Data sources were categorized based on country, person count, the levels of care represented (primary, secondary or mixed), why a person was included, and how data were captured. These categories were ascertained from the data source description and metadata provided to the EHDEN portal and verified with the data partners.

There are three reasons persons could be included in an EHDEN data source (person inclusion), as defined by: **population**, where a person enters the data source because they live in a certain geographical location or because they are registered with a practice or insurer; **encounter**, where a person enters the data source upon a visit to a healthcare provider, for any medical reason; or **disease**, where a person enters the data source when satisfying specific criteria, i.e. a person has a specific medical condition. The most restrictive reason was chosen as the classification for each data source.

[1] https://github.com/ohdsi/databaseconnector#features

We identified which types of data capture methods each source contained, it could be one or more of the following: electronic health record (EHR), a bill or adjudicated claim record for health services rendered (claim), measurements taken and results recorded (lab), a set of required information collected about participants in a registry (case report form), patient reported data (survey), documents analyzed by pulling structured data from unstructured data using a natural language processing algorithm (NLP), or death information from an official source or government entity (death certificate). If the data source did not provide this information they were categorized as unknown.

Results

As of September 1, 2024, there are 210 data sources in the EHDEN portal. Figure 2 shows all countries and the number of data sources available in each. The data sources span 30 countries with the largest representation from Italy, Great Britain, and Spain with 13.0%, 12.5% and 11.5% of the total data sources in the network, respectively. The mean number of persons per data source is 2,147,161 and the median number of persons is 457,664.

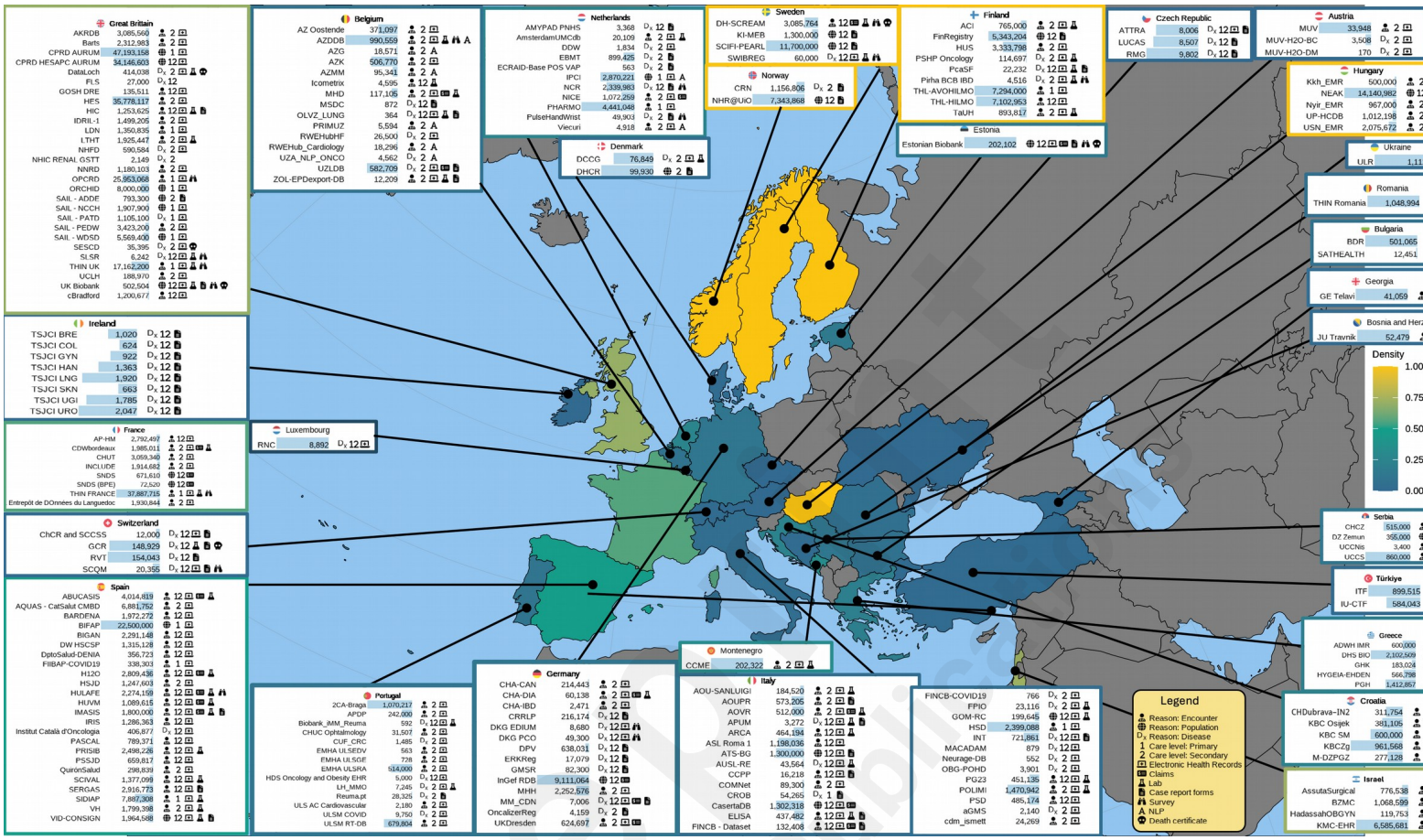

Figure 2: Map of all EHDEN data partners including patient count, care level, data capture method, and reason for person inclusion. The color of the countries represents the density of the data sources as compared to the total population

Table 1 provides the complete list of data sources and their attributes. One row in the table equates to one data source. The first column lists the data partner, which is the name of the institution or organization that is the custodian of the data source. The individual data sources are identified by an acronym which is also how they are identified in the EHDEN portal. Country of origin is represented by the two-digit country code. The number of persons, the person inclusion method, and care level are also provided. Each data capture category has its own column in the table. If a data source uses one of the capture methods, that box is colored green in the table.

Looking at care setting, 46.7% (98) of data sources represent data from the secondary setting only while 42.4% (89) represent data from mixed settings (primary and secondary). A comparatively smaller set of 11.0% (23) represent data only from the primary care setting (Table 2). Looking at the ways in which persons are included in the data sources, 55.7% (117) do so through healthcare encounters, 32.9% (69) through disease-specific data collection, and 11.4% (24) through population-based sources.

Figure 3 shows the number of data sources that receive information through each capture method and each combination of capture methods. EHR is the most common, with 74.7% (157) of data sources reporting at least one capture method as EHR. Over half of those data sources (85) report EHR as their only method for receiving data. Lab is the second most common way data sources capture information as it is reported in 29.5% (62) of data sources. Unlike EHR, lab data is more likely to be coupled with another data capture method as only one data source lists lab as the singular way they receive information.

Table 2: Stratification of person inclusion methods and care levels covered

Care Level	***Person Inclusion***			***Total, %***
	Disease	*Encounter*	*Population*	
Mixed	40	34	15	89 (42.4%)
Primary	2	14	7	23 (11%)
Secondary	27	69	2	98 (46.7%)
Total, %	69 (32.9%)	117 (55.7%)	24 (11.4%)	210

Table 1: Overview of 210 EHDEN data sources

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	ertificateDeath
Centro Clínico Académico – Braga, Associação (2CA-Braga)	2CA-Braga	PT	1070217	Encounter	Secondary	■						
INCLIVA	ABUCASIS	ES	4014819	Encounter	Mixed	■	■	■				
The wellbeing services county of Southwest Finland, VarHa	ACI	FI	765000	Encounter	Secondary	■		■				
Innovative Medical Research SA	ADWH IMR	GR	600000	Encounter	Primary	■						
Fondazione Casa Sollievo della Sofferenza	aGMS	IT	2140	Disease	Secondary	■						
Akrivia Health	AKRDB	GB	3085560	Encounter	Secondary	■						
Amsterdam UMC	AmsterdamUMCdb	NL	20109	Encounter	Secondary	■		■				
Stichting VUmc	AMYPAD PNHS	NL	3368	Disease	Mixed				■			
AZIENDA OSPEDALIERO UNIVERSITARIA SAN LUIGI GONZAGA	AOU-SANLUIGI	IT	184520	Encounter	Secondary	■		■				
University Hospital of Parma	AOUPR	IT	573205	Encounter	Secondary	■			■			
Azienda Ospedaliera Universitaria Integrata Verona	AOVR	IT	512000	Encounter	Secondary	■	■	■				
Assistance Publique - Hopitaux de Marseille	AP-HM	FR	2792497	Encounter	Mixed	■						
APDP	APDP	PT	242000	Encounter	Secondary	■						
Azienda Ospedaliero-Universitaria di Modena	APUM	IT	3272	Disease	Mixed	■		■	■			
Servei Català de la Salut	AQUAS - CatSalut CMBD	ES	6881752	Encounter	Secondary	■						
FONDAZIONE TOSCANA GABRIELE MONASTERIO PER LA RICERCA MEDICA E DI SANITA PUBBLICA (FTGM)	ARCA	IT	464194	Encounter	Mixed	■		■				
ASL Roma 1	ASL Roma 1	IT	1198036	Encounter	Mixed	■						
Assuta medical centers	AssutaSurgical	IL	776538	Encounter	Secondary	■				■		
ATS Bergamo	ATS-BG	IT	1300000	Population	Mixed	■			■			
Institute of Rheumatology	ATTRA	CZ	8006	Disease	Mixed	■			■			
Marco Massari (IRCSSE)	AUSL-RE	IT	43564	Disease	Mixed	■		■				
Az Oostende	AZ Oostende	BE	371097	Encounter	Secondary	■						
AZ Delta	AZDDB	BE	990559	Encounter	Secondary	■		■		■	■	
VZW AZ Groeninge	AZG	BE	18571	Encounter	Secondary	■					■	
AZ Klina	AZK	BE	506770	Encounter	Secondary	■						
AZ Maria Middelares	AZMM	BE	95341	Encounter	Secondary						■	
Servicio Navarro de Salud Osasunbidea (SNS-O)	BARDENA	ES	1972272	Encounter	Mixed	■						
Barts Health NHS Trust	Barts	GB	2312983	Encounter	Secondary	■						
National Scientific Programme "E-Health in Bulgaria"	BDR	BG	501065	Disease	Mixed	■			■			
Agencia Española de Medicamentos y Productos Sanitarios, AEMPS	BIFAP	ES	22500000	Population	Primary	■						

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	ertificateDeath
Instituto Aragonés de Ciencias de la Salud (IACS)	BIGAN	ES	2291148	Encounter	Mixed	■						
Instituto de Medicina Molecular	Biobank_iMM_Reuma	PT	592	Disease	Mixed	■		■				
Bnai Zion Medical Research Foundation and Infrastructure Development Health Services	BZMC	IL	1068599	Encounter	Secondary	■		■		■		
Inspire-srl	CasertaDB	IT	1302318	Population	Mixed	■	■					
Connected Bradford	cBradford	GB	1200677	Encounter	Mixed	■						
Clinical Center of Montenegro	CCME	ME	202322	Encounter	Secondary	■		■				
Casa di Cura Privata del Policlinico (CCPP)	CCPP	IT	16218	Encounter	Mixed	■			■			
ISMETT	cdm_ismett	IT	24269	Encounter	Secondary	■						
Bordeaux University Hospital	CDWbordeaux	FR	1985011	Encounter	Secondary	■	■	■				
Charité - Universitätsmedizin	CHA-CAN	DE	214443	Encounter	Secondary	■						
Charité - Universitätsmedizin	CHA-DIA	DE	60138	Encounter	Secondary	■	■	■				
Charité - Universitätsmedizin	CHA-IBD	DE	2471	Encounter	Secondary	■						
Institute of Social and Preventive Medicine, University of Bern	ChCR and SCCSS	CH	12000	Disease	Mixed	■			■			
Clinical-hospital center Zvezdara	CHCZ	RS	515000	Encounter	Secondary	■		■				
Clinical Hospital Dubrava	CHDubrava–IN2	HR	311754	Encounter	Secondary	■						
Centro Hospitalar Universitário de Coimbra (CHUC)	CHUC Ophtalmology	PT	31507	Encounter	Secondary	■						
Centre Hospitalier Universitaire de Toulouse	CHUT	FR	3059340	Encounter	Secondary	■						
Modena Oncology Center - Azienda Ospedaliera Modena	COMNet	IT	89300	Encounter	Secondary	■						
Clinical Practice Research Datalink (CPRD)	CPRD AURUM	GB	47193158	Population	Primary	■						
Clinical Practice Research Datalink (CPRD)	CPRD HESAPC AURUM	GB	34146603	Population	Mixed	■						
The Norwegian Cancer Registry	CRN	NO	1156806	Disease	Secondary	■			■			
Basilicata Cancer Registry	CROB	IT	54265	Disease	Primary				■			
Krebsregister Rheinland-Pfalz	CRRLP	DE	216174	Disease	Mixed				■			
CUF	CUF_CRC	PT	1485	Disease	Secondary	■						
DataLoch	DataLoch	GB	414038	Disease	Secondary	■		■				■
Center for Surgical Science (CSS)	DCCG	DK	76849	Disease	Secondary	■		■				
Amsterdam UMC	DDW	NL	1834	Disease	Secondary	■						
Stockholm CREAtinine Measurements Project	DH-SCREAM	SE	3085764	Encounter	Mixed		■	■	■	■		■
University of Southern Denmark	DHCR	DK	99930	Population	Secondary				■			
DIGITAL HEALTH SOLUTIONS SA	DHS BIO	GR	2102509	Encounter	Primary	■		■				
German Cancer Society (DKG)	DKG EDIUM	DE	8680	Disease	Mixed	■				■		
German Cancer Society (DKG)	DKG PCO	DE	49300	Disease	Mixed	■				■		
Hospital de Denia	DptoSalud-DENIA	ES	356723	Encounter	Mixed	■						

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	ertificateDeath
University of Ulm, ZIBMT	DPV	DE	638031	Disease	Mixed				■			
Research Institute - Hospital de la Santa Creu i Sant Pau	DW HSCSP	ES	1315128	Encounter	Mixed	■		■				
Primary Healthcare Center Zemun	DZ Zemun	RS	355000	Population	Primary	■		■				
EBMT: The European Society for Blood and Marrow Transplantation	EBMT	NL	899425	Disease	Secondary				■			
European Clinical Research Alliance on Infectious Diseases (ECRAID)	ECRAID-Base POS VAP	NL	563	Disease	Secondary				■			
Centre Hospitalier Universitaire de Montpellier	eDOL Entrepôt de DOnnées du Languedoc	FR	1930844	Encounter	Secondary	■						
Fondazione IRCCS Policlinico San Matteo	ELISA	IT	437482	Encounter	Mixed	■		■	■			
EGAS MONIZ HEALTH ALLIANCE	EMHA ULSEDV	PT	563	Encounter	Secondary	■						
EGAS MONIZ HEALTH ALLIANCE	EMHA ULSGE	PT	728	Encounter	Secondary	■						
EGAS MONIZ HEALTH ALLIANCE	EMHA ULSRA	PT	514000	Encounter	Secondary	■						
European Rare Kidney Disease Registry (ERKReg)	ERKReg	DE	17079	Disease	Mixed				■			
University of Tartu	Estonian Biobank	EE	202102	Population	Mixed	■	■		■	■		■
FIIBAP	FIIBAP-COVID19	ES	338303	Encounter	Primary	■						
Fondazione IRCCS Istituto Neurologico Carlo Besta	FINCB - Dataset	IT	132408	Encounter	Mixed	■	■		■			
Fondazione IRCCS Istituto Neurologico Carlo Besta FINCB	FINCB-COVID19	IT	766	Disease	Secondary	■						
FinRegistry (Institute of Molecular Medicine Finland (FIMM), University of Helsinki)	FinRegistry	FI	5343204	Population	Mixed				■			
Queen Mary University of London	FLS	GB	27000	Disease	Mixed							
Fondazione Poliambulanza Istituto Ospedaliero	FPIO	IT	23116	Disease	Secondary	■		■				
Geneva Cancer Registry	GCR	CH	148929	Disease	Mixed			■	■			■
Telavi Regional Hospital	GE Telavi	GE	41059	Encounter	Secondary	■		■				
GENERAL HOSPITAL OF KAVALA	GHK	GR	183024	Encounter	Secondary	■						
MS Forschungs- und Projektentwicklungs-GmbH	GMSR	DE	82300	Disease	Mixed	■			■			
Grande Ospedale Metropolitano "Bianchi-Melacrino-Morelli"	GOM-RC	IT	199645	Population	Mixed	■		■				
GOSH	GOSH DRE	GB	135511	Encounter	Mixed	■						
Fundacion de Investigacion Biomedica del Hospital Universitario 12 de Octubre	H12O	ES	2809436	Encounter	Mixed	■	■	■				
Hadassah OBGYN	HadassahOBGYN	IL	119753	Encounter	Secondary	■						
Hospital Distrital de Santarém (HDS)	HDS Oncology and Obesity EHR	PT	5000	Disease	Mixed	■						
Harvey Walsh Ltd	HES	GB	35778117	Encounter	Secondary	■						
Health Informatics Centre (HIC)	HIC	GB	1253625	Encounter	Mixed	■		■	■			
SIMG	HSD	IT	2399088	Encounter	Primary	■						

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	ertificateDeath
Hospital Sant Joan de Déu	HSJD	ES	1247603	Encounter	Secondary	■						
Fundación para la Investigación del Hospital Universitario La Fe de la Comunidad Valenciana (HULAFE)	HULAFE	ES	2274159	Encounter	Mixed	■	■	■		■		
Hospital District of Helsinki and Uusimaa	HUS	FI	3333798	Encounter	Secondary	■						
Virgen Macarena University Hospital	HUVM	ES	1089615	Encounter	Mixed	■	■	■				
DIAGNOSTIC & THERAPEUTIC CENTER OF ATHENS “HYGEIA” SINGLE MEMBER SOCIETE ANONYME	HYGEIA-EHDEN	GR	566798	Encounter	Mixed	■		■				
Icometrix	Icometrix	BE	4595	Encounter	Mixed			■				
Lancashire and South Cumbria Integrated Care Board	IDRIL-1	GB	1499205	Encounter	Secondary	■						
Fundacio Institut d'Investigacions Mèdiques (FIMIM)	IMASIS	ES	1800000	Encounter	Mixed	■	■	■	■			
Lille University Hospital	INCLUDE	FR	1914682	Encounter	Secondary	■						
InGef - Institute for Applied Health Research Berlin GmbH	InGef RDB	DE	9111064	Population	Mixed		■					
Institut Català d'Oncologia	Institut Català d'Oncologia	ES	406877	Disease	Mixed	■						
Fondazione Istituto Nazionale dei Tumori	INT	IT	721861	Disease	Mixed	■			■			
NO GRANT	IPCI	NL	2870221	Population	Primary	■		■			■	
Consorci Corporació Sanitària Parc Taulí	IRIS	ES	1286363	Encounter	Mixed	■						
Istanbul University	ITF	TR	899515	Encounter	Mixed	■						
IUC Cerrahpaşa TIP Fakületesi	IU-CTF	TR	584043	Encounter	Mixed	■						
E-MEDIT D.O.O. & Hospital Travnik	JU Travnik	BA	52479	Encounter	Secondary	■		■				
IN2 d.o.o. & Clinical Hospital Center Osijek	KBC Osijek	HR	381105	Encounter	Mixed	■		■	■			
IGEA d.o.o. & University Hospital Center Sestre milosrdnice	KBC SM	HR	600000	Encounter	Primary	■		■	■			
Hierarchia & University Hospital Centre Zagreb	KBCZg	HR	961568	Encounter	Mixed	■						
MEB KI	KI-MEB	SE	1300000	Population	Mixed				■			
Bács-Kiskun Megyei Kórház a Szegedi Tudományegyetem Általános Orvostudományi Kar Oktató Kórháza	Kkh_EMR	HU	500000	Encounter	Secondary	■		■				
The Directorate of Government Medical Centers at the Israeli Ministry Of Health	KMC-EHR	IL	6585681	Encounter	Secondary	■		■		■		
Lambeth DataNet	LDN	GB	1350835	Encounter	Primary	■						
Hospital da Luz Learning Health	LH_MMO	PT	7245	Disease	Secondary	■	■	■				■
Leeds Teaching Hospitals	LTHT	GB	1925447	Encounter	Secondary	■		■				
OAKS Consulting s.r.o.	LUCAS	CZ	8507	Disease	Mixed				■			
MCS Grupa d.o.o. & Health Care Center of Primorje-Gorski Kotar County	M-DZPGZ	HR	277128	Encounter	Primary	■		■				
Azienda Ospedaliera SS Antonio e Biagio e Cesare Arrigo	MACADAM	IT	879	Disease	Mixed	■						

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	CertificateDeath
Medaman	MHD	BE	117105	Encounter	Secondary	■	■	■				
Hanover Medical School	MHH	DE	2252576	Encounter	Secondary	■						
CancerDataNet GmbH	MM_CDN	DE	7006	Disease	Mixed	■	■		■			
University MS Center	MSDC	BE	872	Disease	Mixed				■			
Medical University of Vienna	MUV	AT	33948	Encounter	Secondary	■						
Medical University of Vienna	MUV-H2O-BC	AT	3508	Disease	Secondary	■						
Medical University of Vienna	MUV-H2O-DM	AT	170	Disease	Secondary	■						
IKNL	NCR	NL	2339983	Disease	Mixed			■	■	■		
National Institute of Health Insurance Fund Management Hungary	NEAK	HU	14140982	Population	Mixed	■						
AO Card. G. Panico - Center for Neurodegenerative Diseases and Aging Brain	Neurage-DB	IT	552	Disease	Secondary	■						
Queen Mary University of London	NHFD	GB	590584	Disease	Secondary	■						
King's College London	NHIC RENAL GSTT	GB	2149	Disease	Secondary			■				
University of Oslo	NHR@UiO	NO	7343868	Population	Mixed				■			
National Intensive Care Evaluation foundation	NICE	NL	1072259	Encounter	Secondary	■	■					
UK National Neonatal Research Database	NNRD	GB	1180103	Encounter	Secondary	■						
Szabolcs-Szatmár-Bereg Megyei Kórházak és Egyetemi Oktatókórház	Nyir_EMR	HU	967000	Encounter	Secondary	■		■				
Bambino Gesù Children's Hospital	OBG-POHD	IT	3901	Disease	Secondary	■						
Onze-Lieve-Vrouwziekenhuis Aalst-Asse-Ninove	OLVZ_LUNG	BE	364	Disease	Mixed	■		■	■			
GermanOncology	OncalizerReg	DE	4159	Disease	Secondary				■			
Optimum Patient Care Limited	OPCRD	GB	25953068	Encounter	Primary	■				■		
Royal College of General Practitioners (RCGP)	ORCHID	GB	8000000	Population	Primary	■						
Rioja Salud	PASCAL	ES	789371	Encounter	Mixed	■						
University of Turku (Prostate Cancer Registry of South West Finland)	PcaSF	FI	22232	Disease	Mixed	■		■	■			
ASST Papa Giovanni XXIII	PG23	IT	451135	Encounter	Mixed	■		■				
Papageorgiou General Hospital	PGH	GR	1412857	Encounter	Secondary	■						
STIZON	PHARMO	NL	4441048	Encounter	Primary	■						
BCB Medical Ltd	Pirha BCB IBD	FI	4516	Disease	Secondary	■		■		■		
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico	POLIMI	IT	1470942	Encounter	Secondary	■		■				
UZ Brussel	PRIMUZ	BE	5594	Encounter	Secondary			■			■	
Fundació Institut d´Investigació Sanitària Illes Balears	PRISIB	ES	2498226	Encounter	Mixed	■		■				
IRCCS Policlinico San Donato	PSD	IT	485174	Encounter	Mixed	■		■				
Finnish Clinical Biobank Tampere	PSHP Oncology	FI	114697	Disease	Secondary	■		■				

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	CertificateDeath
Parc Sanitari Sant Joan de Déu	PSSJD	ES	659817	Encounter	Mixed	■						
Harm Slijper	PulseHandWrist	NL	49903	Disease	Secondary				■	■		
Quironsalud	QuirónSalud	ES	298839	Encounter	Secondary	■						
Registo Portugues de Doentes Reumaticos	Reuma.pt	PT	28325	Disease	Secondary				■			
Czech Myeloma Group	RMG	CZ	9802	Disease	Mixed				■			
Registre National du Cancer du Luxembourg	RNC	LU	8892	Disease	Mixed	■						
Vaud Cancer Registry	RVT	CH	154043	Disease	Mixed				■			
LynxCare	RWEHub_Cardiology	BE	18296	Encounter	Secondary						■	
LynxCare	RWEHubHF	BE	26500	Disease	Secondary	■						
SAIL Databank	SAIL - ADDE	GB	793300	Population	Secondary				■			
SAIL Databank	SAIL - NCCH	GB	1907900	Population	Primary	■						
SAIL Databank	SAIL - PATD	GB	1105100	Disease	Primary	■						
SAIL Databank	SAIL - PEDW	GB	3423200	Encounter	Secondary	■						
SAIL Databank	SAIL - WDSD	GB	5569400	Population	Primary	■		■				■
SAT Health	SATHEALTH	BG	12451	Encounter	Secondary		■					
Gothenburg University	SCIFI-PEARL	SE	11700000	Population	Mixed			■	■			
Servicio Cántabro de Salud and IDIVAL	SCIVAL	ES	1377099	Encounter	Mixed	■		■				
HUG and SCQM	SCQM	CH	20355	Disease	Mixed	■			■	■		
Consellería de Sanidade	SERGAS	ES	2916773	Encounter	Mixed	■			■			
University of Edinburgh	SESCD	GB	35395	Disease	Secondary	■						■
SIDIAP - The Information System for Reseach in Primary Care	SIDIAP	ES	7887308	Encounter	Primary	■		■				
King's College London	SLSR	GB	6242	Disease	Mixed	■		■		■		
Health Data Hub	SNDS	FR	671610*	Population	Mixed		■					
Bordeaux PharmacoEpi	SNDS (BPE)	FR	72520*	Population	Mixed		■					
SWIBREG	SWIBREG	SE	60000	Disease	Mixed	■		■		■		
Pirkanmaa Hospital District	TaUH	FI	893817	Encounter	Secondary	■		■				
CEGEDIM HEALTH DATA	THIN FRANCE	FR	37887715	Encounter	Primary	■		■		■		
CEGEDIM HEALTH DATA	THIN Romania	RO	1048994	Encounter	Primary	■				■		
CEGEDIM HEALTH DATA	THIN UK	GB	17162200	Encounter	Primary	■		■		■		
Finnish Institute of Health and Welfare	THL-AVOHILMO	FI	7294000	Encounter	Primary	■						
Finnish Institute for Health and Welfare (THL)	THL-HILMO	FI	7102953	Encounter	Mixed	■						
Trinity St James's Cancer Institute	TSJCI BRE	IE	1020	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI COL	IE	624	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI GYN	IE	922	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI HAN	IE	1363	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI LNG	IE	1920	Disease	Mixed				■			

Data Partner	Data Source Acronym	Country	Person Count	Person Inclusion	Care Level	EHR	Claim	Lab	rmCase Report	Survey	NLP	ertificateDeath
Trinity St James's Cancer Institute	TSJCI SKN	IE	663	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI UGI	IE	1785	Disease	Mixed				■			
Trinity St James's Cancer Institute	TSJCI URO	IE	2047	Disease	Mixed				■			
Clinical centre of Nis	UCCNis	RS	3400	Encounter	Secondary	■		■				
Clinical Center of Serbia	UCCS	RS	860000	Encounter	Secondary	■		■	■			
University College London Hospitals	UCLH	GB	188970	Encounter	Secondary	■						
University College London (UCL) (UK Biobank)	UK Biobank	GB	502504	Population	Mixed	■		■	■	■		■
University Medicine Dresden	UKDresden	DE	624697	Encounter	Secondary	■	■					
National Cancer Institute	ULR	UA	1112	Disease	Mixed				■			
ULS AC Cardiovascular	ULS AC Cardiovascular	PT	2180	Encounter	Secondary	■						
ULSM	ULSM COVID	PT	9750	Disease	Secondary	■						
Unidade Local de Saúde de Matosinhos	ULSM RT-DB	PT	679804	Encounter	Secondary	■						
University of Pécs	UP-HCDB	HU	1012198	Encounter	Secondary	■	■			■		■
Semmelweis University	USN_EMR	HU	2075672	Encounter	Secondary	■						
University Hospital Antwerp	UZA_NLP_ONCO	BE	4562	Disease	Secondary						■	
Universitaire Ziekenhuizen KU Leuven	UZLDB	BE	582709	Disease	Secondary	■	■		■			
Vall d'Hebrón Hospital Campus	VH	ES	1799398	Encounter	Secondary	■		■				
FISABIO-HSRU	VID-CONSIGN	ES	1964588	Population	Mixed	■		■	■			
VieCuri Medisch Centrum	Viecuri	NL	4918	Encounter	Secondary	■					■	
Ziekenhuis Oost-Limburg	ZOL-EPDexport-DB	BE	12209	Encounter	Secondary	■		■	■			

*This is a subset of the full data source.

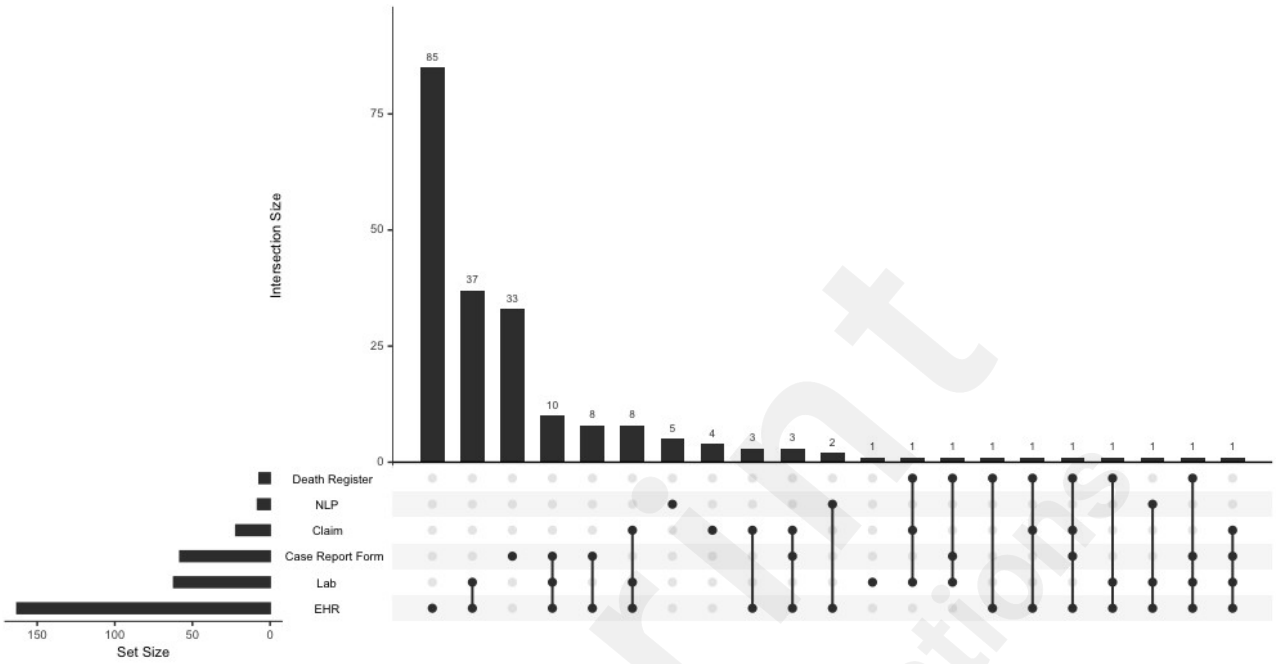

Figure 3**:** Intersection of data capture methods and number of data sources that use each method(s)

Discussion

The varied healthcare data across Europe, as demonstrated by the summary of 210 data sources in EHDEN from 30 countries, underscores the critical need to generate evidence from more than one data source to comprehensively represent the healthcare needs or experiences of the entire European population. Across the person inclusion and care levels represented in the network, the data sources are well distributed, emphasizing how healthcare systems, populations, and data capture methods can differ substantially. While 74.7% of the data sources report EHR as at least one of their data capture methods, only 40.4% report EHR as their only data capture method. The other 34.3% report some combination of EHR, Lab, Case Report Form, Claim, NLP, and Death Register data, showcasing the tremendous heterogeneity of data available in Europe.

Prior Initiatives

Prior initiatives like EU-ADR and IMI-EMIF laid the groundwork for EHDEN, with learnings from those projects directly impacting this project[15,18,34,35]. EU-ADR demonstrated the feasibility of building a federated data network for large-scale drug safety monitoring in Europe using common data analysis files. IMI-EMIF made the first transition from using Common Input Files like those in EU-ADR to the OMOP CDM but it was not scalable due to the lack of funds and need for trained SMEs, both problems which EHDEN addressed.

Sustainability and Success of EHDEN

The sustainability of the EHDEN initiative has been achieved through a combination of mechanisms that foster shared leadership, collaboration, and long-term value creation. One key factor has been the stimulation and enablement of both national and European collaborations. The establishment of OHDSI National Nodes has provided a platform for data partners within

individual countries to collaborate, share best practices, and enhance data quality[36]. These nodes facilitate national-level harmonization while ensuring compliance with local regulations and coding systems, thereby strengthening the network's integrity. Beyond this, EHDEN's adoption in multiple European projects has further expanded its influence, including its pivotal role in enabling large-scale initiatives such as the Data Analysis and Real World Interrogation Network (DARWIN EU®). This has also influenced how the European Federation of Pharmaceutical Industries and Associations (EFPIA) is standardizing its data, demonstrating EHDEN's transformative impact across sectors.

EHDEN has also delivered significant economic value by creating local ecosystems that support SMEs and data partners. Through the Harmonisation Fund, EHDEN has injected resources into the European healthcare data landscape, with the return on investment yielding a multiplier effect. By recruiting and training SMEs through the EHDEN Academy, the initiative has built local expertise to support data partners throughout the ETL process, ensuring decentralized and sustainable support for the network.

One of the goals of EHDEN has been to standardize health data—akin to utilities like electricity or the internet—essential and accessible to a rapidly growing number of stakeholders across Europe. Now that EHDEN has transitioned from a project under IHI to the non-profit EHDEN Foundation, the focus has shifted to sustain, expand, and improve the network while leveraging the harmonized data for evidence generation. This next phase aims to generate meaningful real-world evidence (RWE) for research and regulatory purposes. A recent report by The European Commission on the future of European competitiveness highlights EHDEN's foundational role in shaping the future of the European Health Data Space, further solidifying its legacy as a critical driver of innovation and collaboration in European healthcare[37].

The success of EHDEN in harmonizing data to the OMOP CDM has led to significant advances in methods research and evidence generation[38–41]. Many of the data partners involved in EHDEN have used their standardized data to conduct analyses across a broad spectrum of use cases. For instance, several studies have been conducted to describe the natural history of diseases, the safety and effectiveness of treatments, and healthcare utilization patterns across diverse populations[42–44]. EHDEN's standardized data has also enabled and improved the development of predictive models, allowing for personalized predictions of treatment outcomes and disease progression[45,46]. The harmonization of data has facilitated large-scale population-level studies, which are crucial for understanding trends in public health and informing healthcare policy decisions[47,48]. These examples of evidence generation illustrate the broad applicability of the data in the network, which serves both academic researchers and regulatory agencies.

One notable demonstration of EHDEN's success is the network's use in providing timely information on medicines under surveillance due to shortages in multiple European countries. More than fifty Data Partners contributed data to a study titled "Incidence, Prevalence, and Characterization of Medicines with Suggested Drug Shortages in Europe[49]." This study represents the largest observational database study conducted across Europe, both in terms of the number of databases involved and its geographic scope. The findings will support European efforts to monitor the use of critical medicines, contributing to the global fight against medicine shortages.

The EHDEN data network has also affected other European collaboratives around RWD. IMI projects like PIONEER, BigData@Heart, EU-PEARL, HARMONY, also use the OMOP CDM and have partly continued the mapping work done in EHDEN[50–53]. In the EMA commissioned Data Analysis and Real World Interrogation Network (DARWIN EU®) initiative, among the 20

DPs onboarded in the first two years, 16 are also EHDEN DPs[54].

Future Directions

Building on the progress achieved through the EHDEN network, several key areas offer opportunities for future development. One priority is fostering sustained engagement with data partners. Continuous collaboration will be essential to ensure that data partners remain active contributors to the network by regularly updating and improving their data contributions. Strategies to incentivize engagement, provide ongoing support, and ensure mutual value will be vital for the network's long-term success, particularly as efforts shift toward more robust evidence generation and ongoing enhancements in data quality.

Expanding the network's reach and optimizing its databases for specific research use cases are also key areas for growth. With 210 data sources currently included, there is a significant opportunity to onboard additional data partners and expand the network's coverage across Europe. Future studies will also help identify gaps where further data optimization is required, such as refining mappings or addressing specific quality issues to ensure that the evidence generated is robust, reproducible, and generalizable.

Finally, the newly established EHDEN Foundation will play a critical role in these efforts. By securing funding and fostering collaborations, the Foundation can drive the onboarding of new data partners, address emerging research questions, and ensure that EHDEN continues to adapt to the evolving healthcare landscape. These directions will position the network to remain a cornerstone for real-world evidence generation in Europe, supporting both research and regulatory innovation.

Conclusion

The results of this study demonstrate that the identification, harmonization, and standardization of data sources through EHDEN have contributed significantly to understanding the diverse RWD landscape and advancement of evidence generation across Europe. These efforts are not only transforming healthcare research but are also influencing broader regulatory initiatives, such as DARWIN EU®, which builds on the foundational work of EHDEN to leverage real-world data for regulatory decision-making. Now that the initiative has transitioned to the EHDEN Foundation, there is an exciting opportunity to focus even more on generating high-quality evidence, further solidifying the role of real-world data in improving healthcare and informing policy decisions across Europe.

Abbreviations

EHDEN	European Health Data & Evidence Network
RWD	Real-world data
OMOP CDM	Observational Medical Outcomes Partnership Common Data Model
RWE	Real-world evidence
SME	Small-to-medium enterprise
ETL	Extract, transform, and load
DQD	Data Quality Dashboard
RCT	Randomized controlled trial
IMI	Innovative Medicines Initiative
IHI	Innovative Health Initiative
DPs	Data Partners

EHR	Electronic health records
DSPC	Data Source Prioritization Committee
DARWIN EU®	Data Analysis and Real World Interrogation Network
EFPIA	European Federation of Pharmaceutical Industries and Associations

Declarations

Author Contributions

All authors (CB, MJS, MM, EAV, MC, PRR, PBR) were involved in data collection. CB, MJS, PBR, MM, PRR were involved in the study design, analysis, and interpretation of results. CB, MJS, PBR, PRR contributed to writing and all authors revised and approved the final draft.

Funding

This project has received support from the European Health Data and Evidence Network (EHDEN) project. EHDEN received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Competing Interests

CB, MJS, EAV, PBR, are employees of Johnson & Johnson and hold stock and stock options. PRR works for a department that receives/received unconditional research grants from Amgen, Chiesi, Johnson and Johnson, UCB Biopharma, the European Medicines Agency and the Innovative Medicines Initiative.

Patient Consent

Not required

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Data Availability

All data supporting this work can be found on the EHDEN portal at https://portal.ehden.eu/summary.

Acknowledgements

The authors would like to express their deepest gratitude to all EHDEN Data Partners who contributed their time, effort, and expertise in harmonizing their data to the OMOP Common Data Model. Their commitment to data quality and transparency has been fundamental to the success of this initiative.
We extend our appreciation to the Small-to-Medium Enterprises (SMEs) that played a pivotal role in supporting Data Partners through the extract, transform, and load (ETL) process. Their expertise in data mapping and technical implementation has been invaluable in ensuring high-quality and standardized data across the network.

We would also like to acknowledge the contributions of the EHDEN Work Packages and their leads, whose dedication to data harmonization, quality assessment, infrastructure development, training, and sustainability efforts have made this project possible. Their leadership and vision have shaped EHDEN into a robust and scalable federated data network.
Finally, we extend our thanks to all members of the EHDEN Consortium who have worked tirelessly to build and sustain this network. This includes researchers, data scientists, software engineers, governance and regulatory experts, and the broader Observational Health Data Sciences and Informatics (OHDSI) community, whose open-science collaboration and innovation continue to drive the success of EHDEN.
We recognize that the continued success of EHDEN is the result of collective contributions from numerous stakeholders across Europe, and we sincerely appreciate the efforts of everyone involved.

References:

1. U.S. Food and Drug Administration. *Framework for FDA's Real-World Evidence Program*.; 2018.

2. Heads of Medicines Agency, European Medicines Agency. HMA-EMA Joint Big Data Taskforce Summary Report. Published online February 13, 2019:48.

3. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology. Accessed September 6, 2024. https://encepp.europa.eu/encepp-toolkit/methodological-guide_en

4. Jonker C, Bakker E, Kurz X, Plueschke K. Contribution of patient registries to regulatory decision making on rare diseases medicinal products in Europe. *Frontiers in Pharmacology*. Published online 2022. doi:10.3389/FPHAR.2022.924648

5. Eskola SM, Leufkens H, Bate A, Bruin ML, Gardarsdottir H. The Role of Real-World Data and Evidence in Oncology Medicines Approved in EU in 2018-2019. *Journal of Cancer Policy*. Published online 2023. doi:10.1016/J.JCPO.2023.100424

6. Bolislis WR, Fay M, Kühler TC. Use of Real-world Data for New Drug Applications and Line Extensions. *Clinical Therapeutics*. Published online 2020. doi:10.1016/J.CLINTHERA.2020.03.006

7. European Medicines Agency. Guidance on good pharmacovigilance practices (GVP). Module VIII. Post-authorisation safety studies. Accessed August 6, 2024. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf

8. Santoro A, Genov G, Spooner A, Raine J, Arlett P. Promoting and Protecting Public Health: How the European Union Pharmacovigilance System Works. *Drug Saf*. 2017;40(10):855-869. doi:10.1007/s40264-017-0572-8

9. Goedecke T, Morales DR, Pacurariu A, Kurz X. Measuring the impact of medicines

regulatory interventions - Systematic review and methodological considerations. *Br J Clin Pharmacol*. 2018;84(3):419-433. doi:10.1111/bcp.13469

10. OECD. *Health at a Glance: Europe 2022: State of Health in the EU Cycle*. Organisation for Economic Co-operation and Development; 2022. Accessed August 22, 2024. https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-europe-2022_507433b0-en

11. Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. *Clinical Pharmacology & Therapeutics*. 2019;106(1):36-39. doi:10.1002/cpt.1426

12. Bakker E, Plueschke K, Jonker C, Kurz X, Starokozhko V, Mol P. Contribution of Real-World Evidence in European Medicines Agency's Regulatory Decision Making. *Clinical pharmacology and therapy*. Published online 2022. doi:10.1002/CPT.2766

13. Flynn RWV, Plueschke K, Quinten C, et al. Marketing Authorization Applications Made to the European Medicines Agency in 2018–2019: What was the Contribution of Real-World Evidence? Published online 2021. doi:10.1002/CPT.2461

14. Gini R, Schuemie M, Brown J, et al. Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. *eGEMs*. Published online 2016. Accessed September 6, 2024. https://dash.harvard.edu/handle/1/26318777

15. Coloma PM, Schuemie MJ, Trifirò G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiology and Drug Safety*. 2011;20(1):1-11. doi:10.1002/pds.2053

16. Trifirò G, Coloma P, Rijnbeek P, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *Journal of internal medicine*. Published online 2014. doi:10.1111/JOIM.12159

17. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010;153(9):600-606. doi:10.7326/0003-4819-153-9-201011020-00010

18. Reynolds RF, Kurz X, de Groot MCH, et al. The IMI PROTECT project: purpose, organizational structure, and procedures. *Pharmacoepidemiol Drug Saf*. 2016;25 Suppl 1:5-10. doi:10.1002/pds.3933

19. EFPIA Recommendations on a Connected Data System in Europe. Accessed September 6, 2024. https://www.efpia.eu/news-events/the-efpia-view/efpia-news/efpia-recommendations-on-a-connected-data-system-in-europe/

20. European Health Data Evidence Network. ehden.eu. Accessed March 12, 2021. https://www.ehden.eu/

21. Innovative Health Initiative | IHI Innovative Health Initiative. September 11, 2024. Accessed September 12, 2024. http://www.ihi.europa.eu/

22. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553-564. doi:10.1093/jamia/ocu023

23. Observational Health Data Sciences and Informatics. OMOP Common Data Model (CDM). https://ohdsi.github.io/CommonDataModel/

24. Voss EA, Blacketer C, van Sandijk S, et al. European Health Data & Evidence Network-learnings from building out a standardized international health data network. *J Am Med Inform Assoc*. 2023;31(1):209-219. doi:10.1093/jamia/ocad214

25. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-578.

26. Schuemie M, Reps J, Black A, et al. Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research. *Stud Health Technol Inform*. 2024;310:966-970. doi:10.3233/SHTI231108

27. The Book of OHDSI. https://ohdsi.github.io/TheBookOfOhdsi/

28. Reich C, Ostropolets A, Ryan P, et al. OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *Journal of the American Medical Informatics Association*. 2024;31(3):583-590. doi:10.1093/jamia/ocad247

29. THEMIS | THEMIS Convention Repository for the OMOP CDM. Accessed September 13, 2024. https://ohdsi.github.io/Themis/

30. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *Journal of the American Medical Informatics Association : JAMIA*. Published online July 27, 2021. doi:10.1093/jamia/ocab132

31. Blacketer C, Voss EA, DeFalco F, et al. Using the Data Quality Dashboard to Improve the EHDEN Network. *Applied Sciences*. 2021;11(24):11920. doi:10.3390/app112411920

32. EHDEN/CdmInspection: R Package to support quality control inspection of an OMOP-CDM instance. Accessed September 13, 2024. https://github.com/EHDEN/CdmInspection

33. EHDEN Portal. Accessed September 13, 2024. https://portal.ehden.eu/wordpress/index.php/landing-page-2/

34. Klungel O, Kurz X, Groot MC, et al. Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project. *Pharmacoepidemiology and drug safety*. Published online 2016. doi:10.1002/PDS.3968

35. Lovestone S. The European medical information framework: A novel ecosystem for sharing healthcare data across Europe. *Learn Health Syst*. 2019;4(2):e10214. doi:10.1002/lrh2.10214

36. Observational Health Data Sciences and Informatics. National Nodes. Accessed September 27, 2024. https://www.ohdsi-europe.org/index.php/national-nodes

37. Mario Draghi. The Future of European Competitiveness. Published online September 9, 2024. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en#paragraph_47059

38. Yang C, Fridgeirsson EA, Kors JA, Reps JM, Rijnbeek PR. Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. *Journal of Big Data*. 2024;11(1):7. doi:10.1186/s40537-023-00857-7

39. Fridgeirsson EA, Williams R, Rijnbeek P, Suchard MA, Reps JM. Comparing penalization methods for linear models on large observational health data. *J Am Med Inform Assoc*. 2024;31(7):1514-1521. doi:10.1093/jamia/ocae109

40. Williams RD, den Otter S, Reps JM, Rijnbeek PR. The DELPHI Library: Improving Model Validation, Transparency and Dissemination Through a Centralised Library of Prediction Models. *Stud Health Technol Inform*. 2023;302:139-140. doi:10.3233/SHTI230085

41. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. 2021;113:103655. doi:10.1016/j.jbi.2020.103655

42. Morales DR, Ostropolets A, Lai L, et al. Characteristics and outcomes of COVID-19 patients with and without asthma from the United States, South Korea, and Europe. *Journal of Asthma*. Published online February 11, 2022. doi:10.1080/02770903.2021.2025392

43. Li X, Ostropolets A, Makadia R, et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. *BMJ*. 2021;373:n1435. doi:10.1136/bmj.n1435

44. Voss EA, Shoaibi A, Yin Hui Lai L, et al. Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study. *EClinicalMedicine*. 2023;58:101932. doi:10.1016/j.eclinm.2023.101932

45. Williams RD, Reps JM, Rijnbeek PR, Ryan PB, Prieto-Alhambra D. 90-Day all-cause mortality can be predicted following a total knee replacement: an international, network study to develop and validate a prediction model. *Knee Surgery, Sports Traumatology, Arthroscopy*. Published online December 6, 2021. doi:10.1007/s00167-021-06799-y

46. Reps JM, Williams RD, You SC, et al. Feasibility and evaluation of a large-scale external validation approach for patientlevel prediction in an international data network: validation of

models predicting stroke in female patients newly diagnosed with atrial fibrillation. Published online May 6, 2020. doi:10.1186/s12874-020-00991-3

47. Lane JCE, Weaver J, Kostka K, et al. Risk of depression, suicide and psychosis with hydroxychloroquine treatment for rheumatoid arthritis: a multinational network cohort study. *Rheumatology*. 2020;(00):1-13. doi:10.1093/rheumatology/keaa771

48. Burn E, Weaver J, Morales D, et al. Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study. Published online November 7, 2019. doi:10.1016/S2665-9913(19)30075-X

49. Pineda-Moncusí M, Rekkas A, Pérez ÁM, et al. Trends of drug use with suggested shortages and their alternatives across 41 real world data sources and 18 countries in Europe and North America. Published online August 29, 2024:2024.08.28.24312695. doi:10.1101/2024.08.28.24312695

50. PIONEER - European Network of Excellence for Big Data in Prostate Cancer. PIONEER - European Network of Excellence for Big Data in Prostate Cancer. Accessed November 7, 2024. https://prostate-pioneer.eu/

51. BigData@Heart > Home. Accessed November 7, 2024. https://www.bigdata-heart.eu/

52. HARMONY is part of the IMI “Big Data for Better Outcomes” program. Accessed November 7, 2024. https://www.harmony-alliance.eu/en/harmony-is-part-of-the-imi-big-data-for-better-outcomes-program

53. EU-PEARL - Innovative Patient Centric Clinical Trial Platforms. EU-PEARL. Accessed November 7, 2024. https://eu-pearl.eu/

54. European Medicines Agency. Darwin EU. Accessed September 26, 2024. https://www.darwin-eu.org/

Supplementary Files

Multimedia Appendixes

EHDEN Data Partner call description.
URL: http://asset.jmir.pub/assets/a4e81cd0b4ac8ba646c1f5806e03e87f.pdf

EHDEN sub-grant agreement model.
URL: http://asset.jmir.pub/assets/1858852801b8e6103877c69c32c2ea52.pdf

EHDEN framework for quality benchmarking.
URL: http://asset.jmir.pub/assets/3addc50e3742251db9665dc22163d554.pdf