PRECISEU

DELIVERABLE 4.3

Guidelines for Joint Interregional Projects (JIPs) on the use of health data

BSC, CLUST-ER HEALTH, BIOPRO & EATRIS

Deliverable's Due Date: 30/06/2026

Actual Submission Date: 12/03/2026

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 Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them.

Project information	
Project title	PeRsonalised medicine Empowerment Connecting Innovation ecoSystems across EUrope
Acronym	PRECISEU
Project URL	https://cordis.europa.eu/project/id/101161301
Grant Agreement n.	101161301
Call	HORIZON-EIE-2023-CONNECT-03
Call Topic	HORIZON-EIE-2023-CONNECT-03-01 - Implementing co-funded action plans for connected regional innovation valleys
Type of Action	HORIZON Programme Cofund Actions
Project start/end date	1 July 2024 – 30 June 2029
Project duration	5 years / 60 months
EU Project officer	Christina Nanou
Project coordinator	Montse Daban (Biocat)
Project manager	María Cejas (Biocat)

Table 1. Project information

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Deliverable information	
Deliverable n.	D4.3
Work package n.	4
Deliverable title	Guidelines for Joint Interregional Projects (JIPs) on the use of health data
Lead beneficiary	Barcelona Supercomputing Center (BSC)
Main authors	• Aída Moure, Laura Portell (BSC) • Charlotte Schlett, Romy Wentenschuh (BIOPRO) • Clémence Foltz (Clust-ER Health) • Emanuela Oldoni (EATRIS)
Contributors	• Cecilia Maini, Elisabetta Toschi, Stefano Carboni (ART-ER) • Liz Renzaglia (BIOVIA) • Apostolos Dimitriadis (FORTH) • Elina Drakvik (Sitra) • Jaakko Lähteenmäki (VTT) • Gökçe Banu Laleci Ertürkmen (SRDC) • Beatriz Barros (Sciensano) • Heikki Lehväslaiho (CSC) • Peija Haaramo (Findata) • Angelo Rossi Mori (CNR – IRPPS)
Reviewed by	María Cejas, Montse Daban (Biocat), Simona Giardina (BSC), Fatima Nazeefa (UiT)
E-mail Contact for queries	Aída Moure (aida.moure@bsc.es)Laura Portell (laura.portell@bsc.es)
Due date	30 June 2026
Document status	Final Report
Document version number	V2
Document type*	R
Dissemination level	Public
*Legend = **R** – Report // **DEM** – Demonstrator, pilot, prototype, plan designs // **DEC** – Websites, patents filing, press & media actions, videos, etc. // **DMP** – Data management plan // **OTHER** – Software, technical diagram, algorithms, models, etc.	

Table 2. Deliverable information

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Task 4.4 CONSORTIUM PARTNERS

	Name of the Entity	Acronym	Role	Country	ORG Type
1	BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACIÓN	**BSC-CNS**	B	Spain	
2	EATRIS ERIC	**EATRIS**	B	Netherlands	
3	BIOPRO BADEN-WUERTTEMBERG GMBH	**BIOPRO**	B	Germany	
4	CLUST ER INDUSTRIE DELLA SALUTE E DEL BENESSERE	**Clust-ER Health**	B	Italy	

***Legend** = Role in the Project: **C** – Coordinator // **B** – Beneficiary // **AP** – Associated Partner // Organization Type: **RTD** – Research and Technological Development// **PHI** – Public Health Institution // **Hosp –** Hospital // **SME** – Small and medium-sized enterprises. // **LC** – Large Company // **NGO** – Non-Governmental Org.

Table 3. Task 4.4 PRECISEU consortium

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

CONSORTIUM PARTNERS

	Name of the Entity	Acronym	Role	Country
1	Biocat LA FUNDACIO BIOREGIO DE CATALUNYA	**Biocat**	COO	ES
2	DEPARTAMENT DE SALUT - GENERALITAT DE CATALUNYA	**SALUT**	BEN	ES
3	BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION	**BSC-CNS**	BEN	ES
4	BIORN CLUSTER MANAGEMENT GMBH	**BIORN**	BEN	DE
5	BIOPRO BADEN-WUERTTEMBERG GMBH	**BIOPRO**	BEN	DE
6	AGENTIA PENTRU DEZVOLTARE REGIONALA NORD-EST	**NE RDA**	BEN	RO
7	ASOCIATIA DIGITAL INNOVATION ZONE ZONA DE INOVARE DIGITALA	**DIZNE**	BEN	RO
8	CLUSTERUL REGIONAL INOVATIV DE IMAGISTICA MOLECULARA SI STRUCTURALA NORD-EST (IMAGO-MOL)	**IMAGO-MOL**	BEN	RO
9	BIOTEHNOLOGICHEN I ZDRAVEN KLASTER	**HLSCB**	BEN	BG
10	STOLICHNA OBSHTINSKA AGENTSIA ZA PRIVATIZATSIA I INVESTITSII	**SIA**	BEN	BG
11	CLUST ER INDUSTRIE DELLA SALUTE E DEL BENESSERE	**CLUST-ER HEALTH**	BEN	IT
12	REGIONE EMILIA ROMAGNA	**RER**	BEN	IT
13	ART-ER-SOCIETA CONSORTILE PER AZIONI	**ART-ER**	BEN	IT
14	VLAAMSE GEWEST	**EWI**	BEN	BE
15	MEDVIA	**MEDVIA**	BEN	BE
16	VIESOJI ISTAIGA INOVACIJU AGENTURA	**IA LITHUANIA**	BEN	LT
17	BRG, BUSINESS REGION GOTEBORG AB	**BRG**	BEN	SE
18	EATRIS ERIC	**EATRIS**	BEN	NL
19	PLATAFORMA DE ORGANIZACIONES DE PACIENTES	**POP**	BEN	ES
20	AGENCIA PER A LA COMPETITIVITAT DE LA EMPRESA	**ACCIO**	BEN	ES
21	IDRYMA TECHNOLOGIAS KAI EREVNAS	**FORTH-ICS**	BEN	EL
22	REGION OF CRETE	**CRETE**	BEN	EL
23	SAHLGRENSKA SCIENCE PARK AB	**SSP**	BEN	SE
24	RIVNE INTERREGIONAL MEDICAL CLUSTER	**RIVNE**	BEN	UA
25	ASTRAZENECA FARMACEUTICA SPAIN S.A.	**ASTRA ZENECA**	BEN	ES
26	AGENCIA DE TRANSFORMACIÓN DIGITAL DE CASTILLA – LA MANCHA	**ATD**	BEN	ES
27	LAZIO INNOVA	**LI**	BEN	IT
28	NATIONAL INSTITUTE FOR BIOPROCESSING RESEARCH AND TRAINING	**NIBRT**	BEN	IE

Table 4. Consortium partners

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

WORK PACKAGES AND LEADERS

Work Packages Name		WP Leader
WP 1	Project Management and Coordination	Biocat
WP 2	Communication and Dissemination	NE RDA
WP 3	Interregional Collaboration and Partnership Bridging	IA Lithuania
WP 4	Use of Health Data	ART-ER
WP 5	Multistakeholder infrastructure to enable access to ATMP on large scale	BIO PRO
WP 6	Market and Patient Access	SSP
WP 7	Training and Cultural Change	HLSCB
WP 8	Adoption of PM innovations in the HealthCare System	SALUT
WP 9	Innovation Support Program	Biocat

Table 5. PRECISEU'S Work Packages and Leaders

How to quote this document
Aída Moure, Laura Portell/ BSC; Charlotte Schlett, Romy Wentenschuh/BIOPRO; Clémence Foltz/CLUST-ER Health; Emanuela Oldoni/EATRIS (2026). **PRECISEU D4.3 Guidelines for Joint Interregional Projects (JIPs) on the use of health data,** *Barcelona Supercomputing Center (BSC)*

Co-funded by the European Union
This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

TABLE OF CONTENTS

DISCLAIMER **2**

TASK 4.4 CONSORTIUM PARTNERS **5**

CONSORTIUM PARTNERS **6**

WORK PACKAGES AND LEADERS **7**

LIST OF ACRONYMS AND ABBREVIATIONS **9**

LIST OF TABLES **10**

LIST OF FIGURES **10**

EXECUTIVE SUMMARY **11**

1. INTRODUCTION **12**

1.1 The PRECISEU initiative and the Joint Interregional Projects (JIPs) *12*
1.2 Importance of the Joint Interregional Projects *13*
1.3 Objectives of the Guidelines *15*
1.4 Relevance/link with other EU initiatives *16*

2. GUIDELINES FOR THE JOINT INTERREGIONAL PROJECTS (JIPS) ON THE USE OF HEALTH DATA **18**

2.1 Data Collection *18*
2.2 Data Standardisation *22*
2.3 Data Publication *26*
2.4 Data Discovery *31*
2.5 Data Access *33*
2.6 Data Use *38*
2.8 Project Management & Governance *43*
2.9 Data Privacy and Security *45*

3. REFERENCES **50**

4. ANNEX: CASE EXAMPLES **51**

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition/Description
1+MG	1+ Million Genomes
AAI	Authentication and Authorisation Infrastructure
AI	Artificial Intelligence
ALS	Amyotrophic Lateral Sclerosis
API	Application Programming Interface
ART-ER	ART-ER (Italy) Società Consortile per Azioni
ATMP	Advanced therapy medicinal products
BBMRI-ERIC	European research infrastructure for biobanking
BIOCAT	Fundació BioRegió de Catalunya
BIOPRO	BIOPRO Baden-Württemberg GmbH
BSC	Barcelona Supercomputing Center
CDISC	Clinical Data Interchange Standards Consortium
CLUST-ER	Clust-ER Health (Italy)
CSV	Comma-separated values
DCA	Data Collaboration Agreement
DCAT	Data Catalog Vocabulary
DCAT-AP	Data Catalog Vocabulary — Application Profile
DICOM	Digital Imaging and Communications in Medicine
DMP	Data Management Plan
DOI	Digital Object Identifier
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DTA	Data Transfer Agreements
DUA	Data Use Agreement
DX.X	Deliverable X.X
EATRIS	European infrastructure for translational medicine
EGA	European Genome-phenome Archive
EGTC	European Grouping of Territorial Cooperation
EHDS	European Health Data Space
EHDS2	Secondary Use of Data (EHDS)
eHDSI	eHealth Digital Service Infrastructure
EHR	Electronic Health Record
EHRxF/EEHRxF	European Electronic Health Record Exchange Format
EISMEA	European Innovation Council and SMEs Executive Agency
ELIXIR	Research infrastructure for life-science data
EMA	European Medicines Agency
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
ENISA	European Union Agency for Cybersecurity
EOSC	European Open Science Cloud
ERN	European Reference Network
ETL	extract-Transform-Load processes
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FORTH	Foundation for Research and Technology – Hellas

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Abbreviation	Definition/Description
1+MG	1+ Million Genomes
GA4GH	Global Alliance for Genomics and Health
GDI	Genomic Data Infrastructure
GDPR	General Data Protection Regulation
GP	General Practitioner
GPAP	Genome-Phenome Analysis Platform
HDAB	Health Data Access Body/Bodies
HL7	Health Level 7
HTA	Health Technology Assessment
ICD-10	International Classification of Diseases, 10th Revision
IMI	Innovative Medicines Initiative
IPR	Intellectual Property Rights
JIPs	Joint Interregional Projects
KPI	Key Performance Indicator
LOINC	Logical Observation Identifiers Names and Codes
MIAME	Minimum Information About a Microarray Experiment (standard)
ML	Machine Learning
OMOP CDM	Observational Medical Outcomes Partnership Common Data Model
OMOP	Observational Medical Outcomes Partnership
PMO	Project Management Office
PRECISEU	PeRsonalised medicine Empowerment Connecting Innovation ecoSystems across EUrope
R&I	Research and Innovation
SCC	Standard Contractual Clause
SME	Small and medium-sized enterprises
SNOMED CT	Systematised Nomenclature of Medicine - Clinical Terms
SOP	Standard Operating Procedure
SPE	Secure Processing Environment
SRDC	Software Research, Development and Consultancy
TEHDAS	Joint Action Towards the European Health Data Space
TIA	Transfer Impact Assessment
UiT	Norges Arktiske Universitet
WP	Work Package

LIST OF TABLES

TABLE 1. PROJECT INFORMATION 3
TABLE 2. DELIVERABLE INFORMATION 4
TABLE 3. TASK 4.4 PRECISEU CONSORTIUM 5
TABLE 4. CONSORTIUM PARTNERS 6
TABLE 5. PRECISEU'S WORK PACKAGES AND LEADERS 7

LIST OF FIGURES

FIGURE 1. HEALTHDATA@EU DATA LIFE CYCLE (SOURCE: TEHDAS) 18

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

EXECUTIVE SUMMARY

This Deliverable (D4.3) defines the Guidelines for the Joint Interregional Projects (JIPs) on the Use of Health Data. JIPs are collaborative initiatives that will connect European regional innovation ecosystems to implement and scale personalised medicine solutions sharing best practices, health data and advanced therapies across regions. The guidelines contain different recommendations and steps to follow when undertaking a cross-regional project in terms of use and management of health data.

D4.3 is part of the PRECISEU project's Work Package 4 - *Use of Health Data*, which looks to assess health system readiness for secondary use of data, define data management needs and support cross-border interoperability. More specifically, Task 4.4 aims to define recommendations and suggestions for the use of health data along the development of JIPs on Personalised Medicine. This is reflected in the gathering of good practices related to health data (D4.3) and data management (D4.4)[1].

These guidelines are directly aligned with the PRECISEU Open Call requirements, offering applicants a practical and coherent framework to design Joint Interregional Projects that meet the Call's expectations on interoperability, data governance, FAIR principles, HealthData@EU-aligned architectures and responsible secondary use of health data for Personalised Medicine. Additionally, while these Guidelines are developed within PRECISEU to support the implementation of Joint Interregional Projects on the use of health data, they are equally intended as a transferable good-practice reference for any interregional or cross-border initiative in Europe aiming to advance data-driven Personalised Medicine. Their application can help regions, Member States and EU-funded projects converge towards shared standards, interoperable infrastructures and governance models that strengthen Europe's capacity to use health data safely, efficiently and at scale.

Although the European Health Data Space (EHDS) is not yet fully implemented, these Guidelines are aligned with its requirements. They follow the HealthData@EU and TEHDAS data lifecycle, providing recommendations and good practices for the collection, standardisation, publication, discovery, access, use and finalisation of data within the Joint Interregional Projects. In addition, the Guidelines address additional key elements of health data use and management, such as data security, ethics and project management. Finally, Guidelines also include cross-border personalised medicine use cases to highlight the added value of multi-country datasets and to illustrate why applying these Guidelines is important for such type of projects.

These guidelines provide a practical roadmap towards efficient, compliant and transparent health data use in multiregional settings. They offer step-by-step guidance across the full data life cycle, promote the responsible reuse of health data for personalised medicine and encourage cross-border interoperability through harmonised standards, tools and best practices. By reading and applying the content of the Guidelines, the JIPs will foster cross-border collaboration, data discovery, access and reuse in line with the FAIR principles, the EHDS and the Horizon Europe requirements. Altogether this will enhance research, innovation and policymaking across Europe.

[1] In accordance with the updated Description of Action and Internal Progress Report + Annual Work Planning, this Deliverable corresponds to the 'Guidelines for Joint Interregional Projects on Personalised Medicine' originally foreseen under WP4 Task 4.4 in the Grant Agreement (D4.3)

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

1. INTRODUCTION

1.1 The PRECISEU initiative and the Joint Interregional Projects (JIPs)

PRECISEU is a five-year EU-funded project (2024–2029) coordinated by Biocat in Catalonia (Spain) that brings together 28 partners from 15 regions in 11 EU countries and Ukraine.

PRECISEU is a Regional Innovation Valley project supported by Horizon Europe, a research and innovation funding programme which connects innovation ecosystems across Europe to drive advancements in personalised healthcare, focusing on the exchange and scalability of advanced health data practices and the implementation of deep-tech healthcare innovations. The consortium includes regional authorities, clusters, universities, research institutions and other key stakeholders from twelve European regions, joined by the Rivne region in Ukraine, aiming to foster cooperation for health innovation across the continent.

PRECISEU's overall objective is to increase the efficiency of the regional innovation ecosystems through strengthened collaboration and shared resources on strategic areas of regional strength and specialisation, enabling the development and implementation of innovative initiatives and facilitating the digital and sustainable transformation of European healthcare systems. The project has the ambition to:

- Accelerate the adoption of personalised medicine in Europe
- Connect innovation ecosystems across Europe
- Contribute to the transfer of practices and solutions among European regions
- Support the scaling-up of advanced technological innovations in the health sector, grounded in two pillars of personalised medicine: health data and advanced therapies
- Be a driving force, aligning the Personalised Medicine strategic agenda with the New European Innovation Agenda and the Smart Specialisation Strategy.

Among its other activities, PRECISEU funds specific Joint Interregional Projects (JIPs). These are collaborative initiatives that will connect European regional innovation ecosystems to implement and scale personalised medicine solutions sharing best practices, health data and advanced therapies across regions. JIPs will cover a broad spectrum of topics on the following domains:

- Use of health data for Research and Innovation (R&I) in compliance with the EHDS regulation
- Contribution to make Advanced Therapy Medicinal Products (ATMPs) sustainable and affordable

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- Build cross-regional value chains on Personalised Medicine to be undertaken by research and innovation entities from participating regions

- Contributions towards the PRECISEU project objectives.

Through the activation of JIPs, PRECISEU seeks to overcome fragmentation among European regions, bridge gaps between research and healthcare systems and stimulate a pan-European value chain that supports both innovation and patient-centred outcomes.

1.2 Importance of the Joint Interregional Projects

Healthcare is increasingly shifting towards personalised medicine, a paradigm in which prevention, diagnosis and treatment are tailored to the unique characteristics, genetic profiles, and real-world contexts of individual patients. This transformation is driven by rapid technological progress, expanding digital capabilities, the accumulation of vast amounts of health data, and the accelerating pace of genomic discoveries. Such personalised approaches promise to improve clinical outcomes, reduce adverse events, optimise healthcare resource allocation, and enhance preventative strategies. PRECISEU recognises that achieving this vision represents the future of healthcare, offering the potential to address patient needs with unprecedented precision and effectiveness.

However, realising this vision requires more than scientific potential. The successful development and implementation of personalised medicine is intrinsically linked to the scale and quality of the data available for analysis. It depends on the establishment of robust data infrastructures, harmonised methodologies, and clear and effective frameworks for collaboration across disciplines, institutions, regions, and national borders. Europe's healthcare environment remains highly fragmented, with data, expertise, and resources dispersed across numerous jurisdictions. Such fragmentation hinders the generation of sufficiently large, diverse, and interoperable datasets and limits the ability to draw meaningful conclusions from data that could otherwise drive innovation and improve patient care. To address this, PRECISEU fosters cross-border cooperation and promotes technical and ethical health data methodologies and standards, aiming to enable a healthcare ecosystem capable of delivering personalised solutions at scale. The Joint Interregional Projects - an Innovation Support Programme with funnelling and monitoring phases, supporting concrete interregional projects to be implemented by consortia with research and innovation entities from participating regions - aim to constitute trusted networks that support the secure sharing and responsible use of health data, enabling researchers, clinicians, and innovators to work with larger, more diverse datasets.

The benefits of JIPs are evident when considering the limitations of isolated regional health datasets. Whether they are derived from Electronic Health Records (EHRs), real-world data, biobanks, or genomic and proteomic repositories, health data that is held within a single region often lack the statistical power, diversity and representativeness that are necessary to support progress in personalised medicine. In many cases, they are too small or too homogenous to enable to detect subtle but clinically significant biomarkers, validate innovative therapies, or to ensure health equity in heterogeneous populations. This is particularly relevant in the case of rare diseases, where the number of patients is low and therefore meaningful research or clinical trials may not be conducted if considering only a single territory. Pooling data and expertise across borders generate added value by increasing the representativeness of datasets, integrating different

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

healthcare practices and population characteristics, and enabling an analysis of how treatments perform in different real-world contexts. They also make it possible to detect rare adverse events, understand long-term treatment outcomes, and validate results across varied populations, which are key requirements for evidence-based personalised medicine. European regions vary in their environmental exposures, genetic backgrounds, clinical guidelines, and models of care. Connecting these differences through shared data allows researchers to understand patient responses with greater nuance and to refine personalised interventions to be more equitable and more widely applicable. From an innovation perspective, by reducing data fragmentation and redundancy, JIPs will contribute towards lower research costs, create economies of scale, and open the way for a more attractive and integrated European market. Innovators benefit from clearer pathways to validate prototypes and therapies across multiple regions, while regulators and healthcare providers gain access to richer datasets that support faster and more informed decision-making. By enabling the responsible reuse of complementary datasets from different countries, JIPs will strengthen Europe's ability to produce high-quality evidence, accelerate therapeutic innovation, and deliver personalised healthcare solutions that benefit all citizens[2].

However, JIPs that involve cross-border health data use may encounter several difficulties:

Data governance. Although the General Data Protection Regulation (GDPR) provides a uniform legal base, its interpretation and application differ substantially between countries, regions and institutions. Besides that, the ethical requirements and internal data policies may also differ, and the EU Member States show different levels of readiness[3] to implement such frameworks as the European Health Data Space (EHDS). All these differences make it hard to establish shared procedures for data use (collecting, storing, accessing, reusing) at the cross-border level.

Technical barriers. Health data originate from different systems, formats, and standards, and the datasets may also vary in the completeness, coding practices, and quality aspects. Real-world data are very diverse, and they often have some missing, inconsistent, or incorrectly classified entries. On the other hand, unstructured data (such as clinical notes or imaging) make the situation even more complicated. The integration of these different sources into logical, interoperable datasets is a very demanding task in terms of harmonisation, curation, and standardisation.

Data quality. Different quality control procedures and variable data maturity may impact the validity and reproducibility of results, which is a serious issue when the results in question are to be used for clinical or regulatory decision-making. Moreover, ethical and privacy issues become more problematic in cross-border contexts: among other sensitive information, genetic and clinical data should be handled with the most reliable security measures, suitable consent mechanisms, and respect for patient rights. On top of that, public trust and cultural attitudes towards data sharing are different throughout Europe, which can influence participation and acceptance.

Organisational and administrative challenges. The coordination among institutions that have different governance structures, timelines, funding mechanisms, contractual rules, and professional cultures, requires

[2] Cascini, F. (2025). *Secondary Use of Electronic Health Data*. *SpringerBriefs in Public Health*. Cham: Springer Nature Switzerland.

[3] D4.1 PRECISEU readiness framework (Readiness/maturity assessment framework)

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

excellent project management and continuous communication. Also, the differences in digital infrastructure and technical capacity can lead to uneven contributions and benefits among the participating regions.

1.3 Objectives of the Guidelines

As mentioned before, the launch and implementation of JIPs present complex obstacles spanning technical, regulatory, ethical, and organisational domains, underscoring the need for guidelines and recommendations to facilitate efficient and compliant cross-border collaboration. The present guidelines for the use of health data in JIPs have therefore serve as a practical resource, a roadmap developed to transform the challenges of cross-border health data use into operational opportunities.

The Guidelines are informed by best practices from existing European initiatives, regulatory requirements, and the specific context of personalised medicine, where the complexity of data and the need for diverse patient populations create unique requirements. It also includes the views of European experts in data access and storage, data interoperability and secure processing domains. The interviewees comprised a Senior Lead at Sitra – Finnish Innovation Fund, a Scientific data management specialist at CSC - IT Center for Science, a Health Innovation Consultant from Sciensano and the Acting Head of Data Services at Findata in Finland. Also, a Senior Researcher at SRDC (Software Research, Development and Consultancy Ltd) in Ankara, a Principal Scientist by VTT Technical Research Center of Finland and a Senior Scientist at IRPPS-CNR in Italy. The interviews took place in different dates between May and October 2025 and explored real-world challenges, expectations, tools and use cases related to data interoperability, cross-border and cross-sector health data exchange, access, processing and storage. Insights from these conversations were instrumental in refining the Guidelines content and structure.

The first objective of the guidelines is to provide step-by-step guidance for launching JIPs that involve health data, highlighting critical aspects, and suggesting practical tools and available templates. The recommendations address the entire Data Life Cycle, recognising that effective data use requires considerations of data governance, technical infrastructure, ethical compliance, and scientific rigour from the early planning stage. By providing a comprehensive framework, the guidelines enable JIP teams to anticipate potential challenges, allocate resources appropriately, and establish foundations for successful cross-border cooperation.

A second ambition of the guidelines is to promote the responsible and effective use of both primary data, collected specifically for research purposes, and secondary data, originally collected for clinical care or administrative purposes but repurposed for research and innovation. As emphasised by the World Health Organisation, “data, and the knowledge derived from the use of that data, should be recognised as a global public good, and data-sharing and data reuse should be maximised in ways that are effective, ethical and equitable in order to improve public health”[4]. Reusing health data for secondary purposes offers a lot of possibilities to advance personalised medicine and improve the health of European citizens. Data obtained from electronic health records, registries, biobanks, clinical trials, genomics, proteomics, as well as citizen-generated health data through digital tools, provide a solid base to understand how treatments work, get the representation of diverse populations, study long-term effects, and identify rare events. If these data are responsibly reused and combined, they can provide evidence for clinical decision-making, regulatory

[4] https://www.who.int/publications/i/item/9789240044968

Co-funded by the European Union

This project has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement No 101161301

processes, and innovative therapies development. However, the realisation of this potential relies on resolving challenges related to data quality, privacy, ethics, and interoperability first. These guidelines cover the main considerations and direct JIPs to established, trusted tools and resources to help address the challenges of primary and secondary health data use across borders, ensuring secure, reliable and reproducible results for better treatments, efficient healthcare systems, and real benefits for European citizens.

In addition, cross-border collaboration introduces complexities that are absent in single-country projects. Real progress requires long-term collaboration and institutional alignment. However, different legislations, varying interpretations of regulations and technical incompatibilities can hinder cooperation or increase the risk of data or patient privacy breaches. Another goal of the present document is therefore to foster cross-border collaboration by suggesting common data formats, shared terminology, and harmonised procedures. By providing checklists, templates, reference materials and best practices for data management (pseudonymisation/anonymisation, secure data transfer and access, etc.) the guidelines aim to eliminate the need for project-specific solutions and enable trusted data sharing between JIP partners. By doing so, we aim to foster an environment of transparency and trust essential for successful cross-border collaborations such as the JIPs.

The final objective of these guidelines is to support the correct use of data, which involves technical competence, ethical responsibility, regulatory compliance and scientific rigour. To do so, the document provides clear, practical directions on how JIP partners should manage data throughout the entire Data Life Cycle. Central to this is the implementation of a robust Data Management Plan (DMP) to define the technical, ethical and procedural requirements for handling data. Beyond that, the guidelines also outline procedures to ensure data quality, including validation, cleaning, curation and versioning as well as data finalisation that involves archiving, sharing and maintaining long-term accessibility. This way, the guidelines enable JIP partners to handle data securely and responsibly while producing high-quality, trustworthy evidence that meets scientific, ethical and regulatory expectations, ultimately strengthening personalised medicine across the participating PRECISEU regions.

1.4 Relevance/link with other EU initiatives

The present guidelines on the use of health data in JIPs are closely linked to key European initiatives, some of which are part of the EU policy framework, that aim to support data-driven innovation and cross-border cooperation in healthcare.

- **European Health Data Space (EHDS):** the flagship EU initiative for enabling the secure and efficient use of health data across Member States for both primary (healthcare) and secondary (research, innovation, policy) purposes. By promoting the legal, technical and governance requirements of the still recent EHDS, the guidelines align with and anticipate its practical operationalisation for cross-border projects involving data sharing and use.

- **Towards the European Health Data Space (TEHDAS) Joint Action:** an EU-funded joint action that developed practical recommendations and governance models for the implementation of the EHDS.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

The document builds upon those relevant TEHDAS outcomes to ensure that data reuse within JIPs aligns with legal, technical and ethical standards.

- **European Open Science Cloud (EOSC):** an association that provides an open environment for the sharing, discovery and reuse of research data. The guidelines encourage JIPs to follow EOSC principles and use its tools and services e.g. for structuring data workflows in cross-border projects.

- **FAIR Guiding Principles for scientific data management and stewardship:** the guidelines advocate for the application of the FAIR (Findable, Accessible, Interoperable, Reusable) principles, which refer collectively to data, metadata and infrastructures. These principles are increasingly required in EU-funded projects and research infrastructures. By embedding FAIR requirements in the design and implementation of JIPs, the guidelines support enhanced data quality, transparency and interoperability, critical enablers for multi-regional personalised medicine research.

- **Data Analysis and Real-World Interrogation Network (DARWIN EU):** coordinated by the European Medicines Agency (EMA), DARWIN EU provides access to diverse data sources across EU countries, enabling the generation of real-world evidence to support regulatory decision-making for medicines. The guidelines promote both the contribution of data to and benefit from DARWIN EU infrastructure, ensuring that evidence generated through collaborative projects meets standards for regulatory use.

By adhering to these guidelines that are aligned with the above important European initiatives, JIPs will not only advance their specific personalised medicine projects but also actively contribute with valuable insights into the operational realities of implementing the EHDS, ensuring that data-driven innovation in health care is conducted legally, ethically, and efficiently across all participating regions. Conversely, as EHDS infrastructure and governance frameworks mature, they should provide essential support for JIPs, potentially streamlining data access procedures, standardising technical specifications, and clarifying legal frameworks.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

2. GUIDELINES FOR THE JOINT INTERREGIONAL PROJECTS (JIPS) ON THE USE OF HEALTH DATA

As Joint Interregional Projects are heavily aligned with the objectives and regulatory framework of the European Health Data Space (EHDS), the TEHDAS data life cycle defined within HealthData@EU is considered an essential pillar of these Guidelines. HealthData@EU constitutes the EU-level infrastructure for the secondary use of health data, enabling secure, interoperable, and trustworthy access to health data across Member States for purposes such as research, innovation, policymaking, and regulatory activities. It establishes a harmonised data life cycle encompassing all stages data go through, while ensuring compliance with data protection, privacy, and cybersecurity requirements. By following the data life cycle below, JIPs ensure legal certainty, cross-border interoperability, and alignment with EU standards for responsible health data reuse.

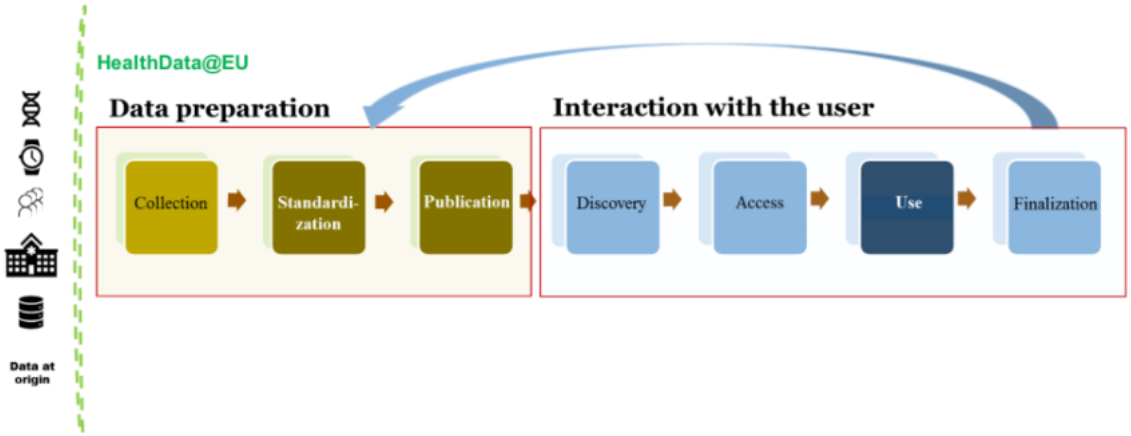

Figure 1. HealthData@EU data life cycle (source: TEHDAS)

2.1 Data Collection

Data collection is the first stage of the Data Life Cycle and, more specifically, the first step of the Data preparation phase. In the context of these guidelines, data collection is the process of obtaining and storing new information on specific variables of interest through various tools or techniques, such as running experiments, surveys, or observing behaviour in the field. It also includes gathering metadata, so that these new data is findable when published. Promoting a *data culture* and making preparation work visible are key to improving health data quality, particularly for precision medicine.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Frequently asked questions

- ***What is the purpose of your data collection?***

The purpose of data collection in your Joint Interregional Project is mainly defined by how the data will contribute to the project's specific goals and overall mission.

In addition to data, the collection of metadata is also essential. Metadata is highly structured data documentation which provides structured and searchable information so that a user can find existing data, evaluate its reusability and cite it. Collecting or generating metadata in a machine-readable or -actionable format makes your data more FAIR (Findable, Accessible, Interoperable and Reusable)[5].

In your project's data management plan (DMP) you will need to provide an overview of the data (and metadata) that will be collected and generated for the project during its entire lifetime. This would include datasets from multiple regions and disciplines relevant to precision medicine and innovation.

- ***Which type of data are you planning to collect?***

The type of data you plan to collect depends strongly on the nature, scope, and objectives of your project. Before any data collection begins, it is essential to clearly define the purpose of the data, ensuring it directly supports both your project goals and the PRECISEU objectives.

Since PRECISEU focuses on health data, it deals with the risk of discovering health-relevant information unrelated to the research objective. A protocol for Incidental Findings - such as the implemented in UK Biobank - should be considered to define how, and under what specific life-saving conditions, a re-identification process to inform a patient of a critical incidental finding would be triggered.

- ***How are you planning to collect the data?***

Data needs to be collected in a consistent and secure way to ensure it can be then accessed, analysed and compared across institutions and regions. However, collection of health data is organised differently across Member States, which makes difficult to compare data between regions and shared initiatives when made available[6].

In order to guarantee consistency in the data collected across the different regions and institutions, Standard Operation Procedures (SOPs) and harmonised protocols are necessary. Also, it is good practice to use shared templates and tools so all partners in your project collect data in the same format. This makes it easy to combine and compare data across borders efficiently within your project.

In addition, you must document in detail your data collection procedures under the project's DMP, to ensure consistency, transparency, and reproducibility.

[5] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship

[6] https://tehdas.eu/tehdas1/app/uploads/2022/05/tehdas-european-health-data-space-data-quality-framework-2022-05-18.pdf

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***Do you need any Data Collaboration Agreement (DCA)?***

A Data Collaboration Agreement (DCA) is a document that specifies how data will be shared amongst project partners. DCAs support cross-border projects by enabling partners to share research infrastructure, data, and expertise across countries, fostering harmonised standards and cooperation. They help bridge institutional and national boundaries, align projects with EU policy priorities, and enhance credibility and access to wider research networks.

This is especially relevant in cross-border collaborations, given that different regions may have different protocols for data sharing. A DCA describes the mechanisms in place (e.g., data sharing agreements, data access committees, joint governance), including how access will be managed and responsibilities distributed.

In case your project needs a DCA, this needs to be clearly specified under the project's DMP. The European Commission – via the Joint Research Centre – has established over 500 cooperation agreements across Europe and beyond.

- ***Who is going to collect the data?***

In order to account for quality and transparency of data collection, you will need to report in the project's DMP the people/institution responsible for collecting the data in each region. They will be responsible for:

  - Gathering data locally according to the project's DMP and using the shared templates and tools to ensure consistency across all partners and regions within the project.
  - Submitting updates according to the agreed schedule so the project team can process the data.

The project coordinator or data manager should provide guidance and support to make sure everyone follows the same procedures and maintains quality in data collection.

- ***Are there any legal and ethical requirements for collecting health data across regions?***

Collecting data across borders might has legal and ethical implications that need to be considered at the earliest stages of the project. Otherwise, you can end up with restrictions in data collection, penalties or delays in the project.

When collecting health data across borders, you must consider aspects such as:

  - Follow data protection laws (e.g., GDPR) to safeguard personal information.
  - Obtain ethical approvals from relevant committees before starting data collection.
  - Get informed consent from participants, explaining how their data will be used and stored.
  - Comply with local and regional rules on cross-border data transfers.
  - Data held within few regions may sometimes lack the diversity needed to ensure health equity. It is advisable to document the demographic and geographic representativeness of your multi-country datasets to ensure you avoid inadvertently discrimination against under-represented European regions.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- If you are developing on considering AI diagnostic tools in your project, you should be required to perform *an Ethical Impact Assessment* to prevent algorithmic bias, especially when combining heterogeneous data from different healthcare systems. More information can be found at the EU AI Act and in country- specific guidelines.

Check the following report from the European Commission for more information on legal and ethical requirements.

How do we ensure data quality at the collection stage?

Data quality refers to how well data fits the user's needs. This is, how efficiently data serves to the project's objectives and how much value it delivers to the different stakeholders.

TEHDAS provides some recommendations to ensure data quality, of which the following are particularly relevant at the collection stage:

- Adherence to TEHDAS framework quality dimensions. Data collected should adhere to the TEHDAS framework's quality dimensions, which include relevance, accuracy & reliability, coherence, coverage, completeness and timeliness. Ensuring these qualities during the data collection phase is crucial for the data's utility in subsequent stages.
- Make sure data meets the project needs (*relevance*).
- Verify that the collected data correctly and consistently reflects what was measured or reported (*accuracy and reliability*).
- In cross-border projects, data collection should be harmonised across all participating sites and regions to ensure that datasets are consistent, comparable, and interoperable, facilitating reliable analysis and secondary use (*coherence*).
- Check if the data collected is representative for your project. This means it adequately reflects the population, exposures and events in your project (*coverage*).
- Ensure that all necessary data elements are captured at the source and the missing data is minimal and accepted (*completeness*).
- Measure how up-to-date the information is collected (*timeliness*).
- Collect not only data but also metadata and related documentation (e.g. source, methods, etc.)
- Follow consent and privacy regulations from the point of collection (*compliance*).

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- Apply basic validations and controls during collection to prevent or minimise errors.

Other potential sources of information on data quality can be found on the QUANTUM project.

Relevant to DMP section(s):
• Data Summary • Increase data re-use (through clarifying licenses) • Ethics

2.2 Data Standardisation

Data quality, comparability and trust are critical to achieving meaningful secondary use. Data harmonisation focuses on ensuring that the meaning of data is consistent and shared, even when data originates from different systems, countries, or disciplines. This is achieved through the use of standards and controlled vocabularies. Shared standardised formats (e.g. OMOP CMD) and powerful computing environments support the path to interoperability.

Frequently asked questions

- ***How do we harmonise data from multiple regions?***

Harmonising data collected from multiple regions requires a coordinated approach that combines technical alignment, legal consistency, and active collaboration among partners to ensure data is contextually accurate and usable. The objective is to achieve interoperability while respecting national regulations.

The process begins by establishing a shared framework and a common understanding across all partners. Reaching early agreement on how data will be collected, described, and shared helps prevent inconsistencies later in the project.

In many different settings there is a tendency to work in silos and a lack of awareness in sharing data. To ensure interoperability, it is essential to adopt standardised structures and metadata. Shared data models and harmonised terminology help align datasets from different sources. To support this, PRECISEU has developed an Interoperability Framework[7] that enables partners to define and document their own approach, ensuring consistent and effective interoperability across partners in the JIP.

In this context, guidance from TEHDAS highlights the importance of metadata consistency, suggesting that the use of a DCAT profile specification (see in the following questions) should be considered to standardise data collection and reporting.

Finally, data harmonisation depends on sustained communication, governance, and capacity building. Regular reporting, transparent documentation, and mutual support enable partners in your project to align on standards and continuously improve data quality. Experienced partners can play a mentoring role in helping others strengthen their data management and compliance practices.

[7] D4.2 PRECISEU Interoperability Framework

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

How should metadata be structured to ensure interoperability?

Metadata should be structured using harmonised, standardised schemas so that datasets can be consistently described, discovered, and exchanged across systems, organisations, and countries. Well-structured metadata ensures that datasets are understandable and traceable for both humans and machines.

To achieve this, metadata should align with the FAIR principles, particularly interoperability and reusability. Each dataset should include a common set of descriptive elements such as title, creator, temporal coverage, version, access conditions, and provenance. These elements should use persistent identifiers and machine-readable formats.

Interoperability is best supported by adopting established metadata standards and profiles. Frameworks such as DCAT and its European application profile DCAT-AP provide a uniform structure for dataset descriptions across repositories.

Finally, metadata standards and rules should be documented in the Data Management Plan (DMP) and applied consistently by all project partners. This supports validation, integration, and long-term reuse of metadata across the full data lifecycle.

- ***How do semantic standards and common data models support data interoperability?***

Semantic interoperability requires that concepts are defined and used consistently. In practice, this means adopting recognised terminologies and ontologies such as SNOMED CT, LOINC, or ICD-10 in the health domain. These standards ensure that identical concepts are represented in the same way and that different terms are not mistakenly interpreted as different meanings.

Ontologies further define the relationships between concepts, allowing data from heterogeneous sources to be integrated and compared meaningfully. By mapping local terms to shared ontological concepts, organisations can align their data without changing internal systems.

When semantic standards are agreed upon and applied consistently across participating organisations, data can be interpreted unambiguously across borders and contexts. This semantic alignment is essential for reliable data integration, cross-country analysis, and reuse in collaborative and secondary-use scenarios.

In addition, common data models operationalise these semantic standards by embedding them into a harmonised data structure. The OMOP Common Data Model is a prominent example, as it defines both a standardised schema and the use of standard vocabularies to which local data are mapped. By transforming heterogeneous source data into the OMOP model, organisations achieve semantic and structural harmonisation while retaining their original data locally.

This approach enables federated and reproducible analyses, as the same analytical methods can be applied across multiple datasets without redefining mappings for each study. It also supports cross-border and cross-institutional collaboration by ensuring that data are interpreted consistently, regardless of their origin. For more information on semantic standards please take a look at D4.2 *PRECISEU Interoperability Framework*.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***What standards help make datasets easier to find and compare across borders?***

While increasingly mentioned and acknowledged in the EHDS context, there is a tendency to overlook the complexity of patient data. To make datasets easier to find and compare across borders, you should apply widely recognised data and metadata standards that support both interoperability and compliance with data protection principles. To find the correct standards to use for your project, PRECISEU has developed an Interoperability Framework[8] that enables partners to define and document their own approach, ensuring consistent and effective interoperability across partners in the JIP.

- ***What quality checks are recommended after standardisation?***

Once your datasets have been standardised, it is essential to carry out systematic quality checks to confirm that data compliance is maintained across all partners and systems. After standardisation, you should conduct the following key checks:

1. **Validation of data structure and format** - Verify that all datasets conform to the agreed standards and data models (such as OMOP or other Common Data Models). Ensure that data elements and metadata fields follow the same definitions, units, and value ranges defined in the project.

2. **Consistency and completeness checks** - Review datasets for missing values, duplicate records, or inconsistencies between sources. Automated scripts or validation tools can be used to confirm that all mandatory fields are populated and correctly formatted.

3. **Semantic alignment testing** - Assess whether controlled terminologies (e.g. SNOMED CT, ICD-10, LOINC) have been applied correctly and uniformly, ensuring that comparable data points have the same meaning across partners and countries.

4. **Verification of mappings and transformations** - Check that any data transformation, cleaning, or mapping performed during standardisation is documented and validated. Create a clear *data flow diagram* to trace these processes and verify that no unintended alterations occurred.

5. **Data protection and access control review** - Conduct post-standardisation checks to ensure that anonymisation or pseudonymisation has been applied correctly and that only authorised users have access to sensitive data.

6. **Quality reporting and audit trail** - Maintain a transparent record of all quality checks performed. Document these procedures as part of the *Data Management Plan (DMP)*, enabling traceability and accountability during audits or future reuse.

- ***How can we document the transformation or mapping process for transparency?***

To ensure transparency, the transformation and mapping process should be thoroughly documented in a way that allows every modification to be clearly traced, validated, and reproduced. This documentation not only supports scientific integrity but also strengthens trust among partners and demonstrates compliance with data governance standards. Also, AI-assisted semantic mapping and data quality validation should be considered.

[8] D4.2 PRECISEU Interoperability Framework

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Begin by creating a **comprehensive data flow description** that captures each stage of data processing - from the original source to the final integrated dataset. This should include diagrams or tables showing how raw variables are transformed, cleaned, harmonised, and mapped to the agreed data model. Each step must specify the purpose of the transformation, the methods used, the version of any scripts or software applied, and the individual or team responsible for execution and validation.

A detailed **data mapping file or codebook** should accompany this, listing all variable names, formats, and standard vocabularies applied. For example, if health concepts are mapped to controlled terminologies or common data models, those relationships should be explicitly recorded to ensure interpretability and consistency across partners in the JIP.

To maintain data integrity, you should apply version control to all transformation scripts, with changes tracked through a secure repository. Consistent **metadata annotation** supports transparency by describing both the transformation logic and any assumptions or exclusions used during data preparation.

Finally, the documentation should be integrated into the project's **Data Management Plan (DMP)**, ensuring that all partners can review, audit, and reuse the information as needed. Transparency in documenting transformations enables reproducibility, facilitates harmonisation in collaborative environments, and provides a solid foundation for regulatory and ethical accountability.

- ***Which data standards or common data models should be used?***

When selecting data standards or common data models, you should aim for recognised frameworks that enable interoperability, reuse, and comparability across projects and borders. Choosing well-established international standards ensures that data can be exchanged and understood consistently, regardless of the national or institutional systems involved.

For health and research data, **Common Data Models (CDMs)** such as **OMOP**, **CDISC**, or **FHIR** are widely recommended, as they provide a structured, consistent way to represent information. These models define standard data elements, relationships, and terminologies, allowing you to integrate data from diverse sources without losing meaning or context. Using controlled vocabularies and coding systems such as **SNOMED CT**, **LOINC**, or **ICD10** further enhances semantic harmonisation and makes datasets easier to analyse across studies. For more information on data standards, please see the D4.2 *PRECISEU Interoperability Framework.*

When standards differ between countries, you should adopt a **bridging approach** that prioritises compatibility while respecting local requirements. This can involve mapping national data elements to international standards, defining equivalence between terminologies, and documenting any deviations transparently. In cases where full alignment is not immediately feasible, you can use **federated or hybrid structures** that allow local datasets to remain under national standards while participating in cross-border analyses through standardised interfaces or queries.

To manage these processes effectively, you should define and agree upon the chosen standards within the **Data Management Plan (DMP)** at the start of the project. This ensures that all partners follow the same conventions for terminology, metadata, and data exchange. Periodic reviews and quality checks should then confirm ongoing compliance with these standards and identify areas for improvement.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Ultimately, the key to resolving differences across countries lies in adopting **well-recognised international standards, ensuring transparency in mappings**, and establishing flexible data architectures that support both national autonomy and global interoperability.

Relevant to DMP section(s):
• Data Summary • Making Data Findable • Making Data Openly Accessible • Making (meta-)Data Interoperable • Increase Data Re-use

2.3 Data Publication

Data publication is the third and last step of the data preparation process in the Data Life Cycle, after data collection and harmonisation. Publishing your data means making it known and available to others. However, this process has relevant steps to be followed in order to make your data compliant with legal and ethical requirements as well as with the FAIR principles.

Frequently asked questions

- ***Which data can be published publicly or internally?***

Publishing research data is increasingly required by funders and publishers, but it can take different forms depending on the research outputs produced.

Under Horizon Europe, researchers are not obliged to publish their results in scientific publications. However, if they choose to publish a peer-reviewed scientific publication, it must be made open access. Open access means that the publication is freely available online so that anyone can read, use, and reuse it at no cost. When a scientific publication is produced, researchers are required by default to provide access to the data underpinning the publication, typically by depositing them in a trusted data repository and linking them to the article. Depending on the ethical, legal, or contractual constraints and the sensitivity of the data, researchers may choose either an open repository or a controlled-access repository. In all cases, metadata describing the datasets should be made openly available.

Independently of whether a scientific publication is produced, researchers may also be required or encouraged to share the data generated by their project via a data repository. This form of data sharing aims to ensure long-term preservation, discoverability, and reuse of research outputs and may include datasets not directly associated with a specific publication.

It is important to mention that some types of data cannot be made openly available due to ethical, legal, or contractual constraints, for example when they include personal, sensitive, or confidential information. In such cases, restricted or controlled access, anonymisation, or justified non-sharing may be applied in line with Horizon Europe requirements. We strongly recommend contacting the Legal Team or the Data Protection Officer (DPO) of your institution to obtain guidance on the appropriate management and sharing of such data.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

In addition, licensing is an important part of data publication. A licence defines what a user is allowed to do with a dataset. Even if your data is public – open access, licensing is relevant so that others know how to acknowledge your contribution. You can find more information on licensing in the questions below.

- ***What formats and repositories are recommended for publishing data?***

When publishing data, it is relevant to consider two broad areas: (1) what formats and metadata standards to use, and (2) which repositories to choose.

1. ***Formats and metadata standards***

a. Use **open, stable, non-proprietary formats** that are likely to remain usable in the long term (machine-readable, non-proprietary). For example: tabular data in CSV/TSV; structured metadata in JSON or XML; imaging data in standard DICOM (for clinical/medical imaging) or widely adopted community formats.

b. Provide **rich metadata** and adhere to **community minimal information standards**. It is critical that you document what the data are, how they were collected, processed, and any constraints on reuse (consent, privacy, etc.) Ensure that metadata allows data to be FAIR.

c. Document **provenance, access conditions, anonymisation/pseudonymisation**. Especially for human and other sensitive data, you must clearly state how data subject anonymisation/pseudonymisation was handled and any access restrictions. Ensure you attach persistent identifiers (e.g. DOIs) where possible so data can be cited and linked.

d. **Data organisation and packaging**: Data should be well-structured, with consistent folder/file naming, clear readme or documentation, and ideally packaged such that future users can understand it without needing to ask you.

2. ***Repositories***

a. A good repository **supports DOIs** and ensures **long-term accessibility**. It has clear policies on metadata, access, reuse and citation. It is also aligned with biomedical/clinical data relevant considerations (such as privacy, controlled access, consent).

b. Consider **discipline-specific repositories**: For example, genomic repositories, imaging repositories, clinical trial data archives. These often support domain-specific metadata, large data volumes, controlled access. Discipline-specific repositories also increase the FAIRness of the data.

c. Examples: Zenodo accepts a variety of research artefacts, issues DOIs for sensitive human biomedical data; repositories such as the European Genome-phenome Archive (EGA) which supports controlled access for human genomic/phenotypic data.

d. You can find more information on data publication and repositories on the RDMkit and the ELIXIR Core Data Resources.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***How do we include provenance, versioning and metadata in published datasets?***

Including provenance, versioning, and metadata in published biomedical datasets is essential to ensure that data are FAIR (Findable, Accessible, Interoperable, and Reusable). In practice, most of these requirements are already addressed when data are deposited in an appropriate, trusted repository.

Provenance refers to the lineage of your data, e.g. where it comes from, how it was generated, and how it has been processed or transformed. It is good practice to include a structured "provenance" section or file in your metadata file, describing:

- Source of the data (e.g., study name, project ID, instrument, sequencing platform)
- Data collection methods (e.g. protocols, instruments, software, parameters)
- Processing steps (e.g., normalisation, alignment, filtering)
- Responsible persons or institutions (e.g. roles, responsibilities)
- Links to related datasets, publications or workflows

Some of the widely used provenance standards are:

- W3C PROV model: A widely used standard for representing provenance in a machine-readable way.
- RO-Crate: A lightweight packaging standard that embeds provenance and metadata with data files (in JSON-LD). It is endorsed by ELIXIR and the EOSC community.
- ISA framework (Investigation-Study-Assay): Used in life sciences to record experimental provenance and metadata.

Versioning ensures users know *which* version of a dataset they are using, allowing reproducibility and transparency.

- Assign dataset version numbers. Use semantic versioning (e.g., v1.0, v1.1, v2.0) for major/minor updates.
- Maintain changelogs, i.e. keep a CHANGELOG.md or equivalent file summarising modifications (new samples, re-processing, corrections).
- Use repository versioning systems. Many European repositories (e.g., Zenodo, EGA, BioStudies) automatically mint new DOIs for new versions and link them to the original record. Version tracking can also be handled via Git/Git-LFS, especially for computational workflows or smaller datasets.
- Metadata versioning: Update dataset metadata files with a "version" and "date" field to ensure consistency between data and metadata.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Metadata describe the context of the dataset, e.g. what it is, who created it, how it was produced, and under what conditions it can be reused. Use standard schemas depending on your data type and include the core metadata fields such as title, authors, abstract, etc.

Some interesting tools:

- **FAIRsharing.org:** curated registry of data and metadata standards, databases and data policies with a primary focus on life sciences.
- **BioSamples (EMBL-EBI):** Metadata repository for sample descriptions linked to omics data.
- **DataCite Metadata Generator:** Helps create rich metadata records for Zenodo or institutional repositories.

Well-established disciplinary and generalist repositories guide researchers through the collection of standardised metadata and automatically capture key provenance information. They typically require information on how the data were generated and processed, link datasets to related publications or projects, and preserve this information in a machine-readable format.

Versioning is also largely handled at repository level: repositories assign persistent identifiers (e.g. DOIs or accession numbers), track dataset updates, and link different versions of the same dataset. This allows users to clearly identify which version of the data they are using and supports reproducibility over time.

To ensure good practice, researchers should:

- Select a trusted repository appropriate for their data type (e.g. domain-specific or certified generalist repositories)
- Complete all mandatory and recommended metadata fields provided by the repository
- Use repository-supported mechanisms for updating and versioning datasets
- Provide links between datasets, publications, and related resources.

By choosing the right repository and following its submission guidelines, researchers can ensure that provenance, versioning, and metadata are captured in a consistent, standardised, and FAIR-compliant manner, with minimal additional effort.

- ***Are you applying any type of licensing to your data?***

Applying a **license** to your biomedical data is an important part of making it *legally reusable* while ensuring proper attribution and protection of sensitive information. Licensing defines how others can use, share and build upon your data. Without a license, your dataset defaults to “all rights reserved” under copyright law — meaning others legally cannot reuse it, even for research or non-commercial purposes.

A clear, well-chosen license:

- Ensures compliance with funder or EU Open Science requirements (e.g., Horizon Europe)

Co-funded by the European Union

This project has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement No 101161301

- Enables reuse and citation
- Clarifies whether commercial use is allowed
- Protects you and your institution from misuse.

You can find more information on licensing and the types of licenses available here.

- ***What approvals or validation steps are required before publication?***

Understanding what approvals or validation steps are required before publishing biomedical data is critical, especially under European regulations (e.g., GDPR, Horizon Europe and national laws).

For any dataset involving *human participants* (including biological samples, imaging, genomic or clinical data) ethical and legal approvals are requested. They confirm that data collection, processing and sharing comply with ethical standards and that informed consent was properly obtained. Ensure your ethics approval covers data sharing and reuse, not just data collection. If not, re-consent or anonymise the data before sharing.

For any dataset containing *personal* or *pseudonymised* data from EU citizens GDPR Compliance is a must-have. Confirm with your Institution's Data Protection Officer (DPO) that all GDPR conditions are satisfied before publication.

Finally, ensure the quality of your data, perform an integrity check, and confirm that your Data Management Plan (DMP) is up to date and aligns with the dataset you are publishing. Validate file formats (e.g., DICOM for imaging, FASTQ/VCF for genomics, mzIdentML for proteomics). Check metadata completeness using repository-specific templates or FAIR assessment tools. Run basic QC scripts or pipelines to confirm reproducibility.

- ***How can I make results discoverable for other EU researchers or policymakers?***

Making your biomedical research results discoverable to other European researchers and policymakers is key to increasing impact, visibility and alignment with EU Open Science and FAIR data principles.

First, deposit both data and metadata in repositories that are:

- Compliant with FAIR principles
- Indexed by the European Open Science Cloud (EOSC)
- Assign persistent identifiers (DOIs)

Repositories such as Zenodo are integrated into OpenAIRE and EOSC, making your dataset automatically visible to other EU institutions and research infrastructures.

Second, provide rich metadata. Good metadata is what makes your data *findable*.

Third, clarify licensing and access. Clearly state whether data are open or controlled access and describe how others can request access.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Finally, link datasets to publications, projects and institutional repositories and promote them through OpenAIRE, EOSC Portal or Horizon Results Platform so that they reach EU researchers and policymakers.

You can find more information on data discoverability here.

- ***Which language should be used to describe the data in the project?***

Language remains an obstacle for health data sharing for EU cross border projects[9]. The recommended language for describing data in European research projects is English. Using English ensures that metadata and documentation are understandable and accessible across all EU member states and internationally. Most European repositories, research infrastructures, such as OpenAIRE and the EOSC index, and search metadata are primarily in English. Moreover, EU and Horizon Europe guidelines recommend English as the standard language for metadata, even if the dataset itself or certain variables are provided in another language.

Additional metadata fields or summaries may be included in national languages to support local users or participants; however, the core metadata, including README files and keywords, should always be provided in English.

Relevant to DMP section(s):
• Making data findable, including provisions for metadata • Increase data re-use (through clarifying licenses) • Data Security

2.4 Data Discovery

Data discovery is about knowing what data is available, where it is stored and how to access it. Effective data access begins with robust data discovery mechanisms and boosts collaboration, supports reuse of data and accelerates scientific progress in personalised medicine. Having access to clear metadata and documentation, knowing which public or institutional repositories to search, and working with standardised data formats that make it easier to understand and assess whether a dataset is fit for reuse are good practices for data discovery in your project, always considering legal, ethical and other potential requirements.

Frequently asked questions

- ***How can you find available datasets from different regions/stakeholders?***

To find available datasets from different regions or stakeholders - while waiting for March 2029, when the catalogue of datasets available for secondary use will also be provided by the HDABs. (art. 55, 57, 77 EHDS) -you can use European and global data discovery platforms that aggregate information from multiple repositories (trusted digital platform where data can be stored, preserved and shared for others to find and reuse):

[9] https://tehdas.eu/app/uploads/2023/09/tehdas-recommendations-for-best-practices-for-eu-crossborder-exchanges.pdf

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- OpenAIRE – integrates datasets, publications, and projects from thousands of European institutions and repositories
- EOSC Portal – the European Open Science Cloud catalogue, providing access to FAIR datasets, tools, and services across disciplines
- ELIXIR Core Data Resources – for life sciences data, linking national nodes and domain repositories (e.g., genomics, proteomics, imaging)
- BBMRI-ERIC Directory – to discover biobanks and biomolecular data resources across Europe
- European Data Portal – for open data from EU institutions, countries, and research infrastructures
- Google Dataset Search or DataCite Commons – for broader global dataset discovery
- Zenodo - free repository for sharing and citing research data.

Also, the data you need may come from primary sources (e.g. EHR, human genetic data, data from biobanks, wellness devices) or from data already collected for other purposes (i.e. you can reuse existing data) such as previous research studies or data registries. A data registry is essentially a centralised system for collecting, storing, and managing structured data about a specific group of people, objects, or events and it is a good source of data, often for research, healthcare, or policy purposes.

There are many EU data registries where data is collected for primary use (e.g. to promote a safe high-quality care across borders) and which is also available for secondary purposes. These registries have some specificities such as a cross-border nature, international coverage and open documentation that make them a good source of information for cross-border projects. TEHDAS provides a list of the cross-border registries with links to their available documentation.

- ***What metadata standards facilitate effective dataset discovery?***

Metadata standards provide a structured and consistent way to describe datasets, making them easier to find, interpret and reuse. In Europe, several well-established standards support FAIR and interoperable data management. Use metadata standards such as DataCite, Dublin Core, or ISA-Tab for general dataset description, and domain-specific standards (e.g., MIAME, DICOM, FHIR) for biomedical data types. These ensure your datasets are discoverable, interoperable, and reusable across European repositories and infrastructures such as OpenAIRE and EOSC. For more information, you can check the D4.2 *PRECISEU Interoperability Framework*.

- ***How do you know if the data are reliable and up to date?***

To ensure that datasets are reliable and up to date, always verify trusted, established sources, check metadata and versioning, and verify that data are curated, documented and recently updated.

- ***How do you identify and select relevant datasets for the project?***

To identify and select datasets suitable for your project:

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Define your requirements**: Clearly specify your research question, data type, format, scope, and quality criteria.
- **Search reliable sources:** Look for domain-relevant datasets in reputable repositories and ensure they follow FAIR principles (Findable, Accessible, Interoperable, Reusable)
- **Review metadata:** Check descriptions, keywords, and provenance to ensure the dataset fits your scope and context.
- **Assess quality and accessibility:** Prioritise well-documented, peer-reviewed datasets, with accessible and clear licensing conditions.
- **Check technical compatibility:** Ensure data formats and standards are interoperable with your tools and existing datasets.

Relevant to DMP section(s):
• Data Summary • Making Data Findable • Increase Data Re-use

2.5 Data Access

Data access is the secure and controlled use of health data for approved purposes, ensuring privacy, data protection and compliance with the law. Under the EHDS, data will be requested using data applications (or data requests) and if approved, a data permit is issued by the HDABs, as defined in the EHDS Regulation proposal. However, the process of accessing health data varies between countries so special considerations need to be taken into account when accessing data in cross-border settings, as analysed in the TEHDAS Joint Action.

Frequently asked questions

- ***What steps are needed to request access to data in another region?***

Accessing health data across regions under EHDS typically involves the following steps aligned with the TEHDAS recommendations on data access and governance:

1. **Define the data need**: Clearly specify the purpose of the project, the datasets required, the type of data (e.g. EHRs, registries, genomics), and the intended analyses. Practical guidance on defining data scope and minimisation can be found in RDMkit – Sensitive data.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

3. **Identify the relevant datasets** you wish to use: Determine which organisations act as data controllers or holders for the required datasets (e.g. hospitals or data repositories). Depending on the country or region, access decisions may be handled by data access committees, regional health authorities, or other designated bodies that currently fulfil roles similar to the future Health Data Access Bodies (HDABs).

4. **Submit data access applications** to the relevant data holders or governance bodies in each country or region. Applications typically include project objectives, data requirements, and data protection measures. Procedures, formats, and timelines may vary significantly across jurisdictions in the absence of harmonised EHDS processes. In the future, EHDS aims to streamline this through an EU central platform and standardised application processes.

5. **Ethical and legal review**: Applications are assessed for compliance with GDPR, national legislation, and institutional requirements. Particular attention might be paid to data minimisation, risk mitigation, and safeguards for sensitive health data. For GDPR-specific considerations, see RDMkit – GDPR compliance page.

6. **Access authorisation and setup**: If approved, access is typically granted through existing secure data access mechanisms defined by the data provider, such as on-site access, remote secure environments, or trusted research platforms. While these may differ across regions, many reflect principles later formalised in EHDS. Technical models for such environments are described in TEHDAS guidance on Secure Processing Environments. Also, related projects are DARE-UK (architecture blueprint seen as a strong reference for Trusted Research Environment federation) or EOSC-ENTRUST (European Network of TREs).

- ***Who can request access to datasets?***

In the absence of a fully operational EHDS, access to health datasets may be requested by a range of organisations, subject to national and institutional rules. These typically include:

- Public or private research organisations
- Universities and academic institutions
- Healthcare providers and health system bodies
- SMEs and industry partners engaged in research, development, or innovation
- Public authorities and policy-making bodies

The general eligibility requirements are:

- Be legally established in the relevant country or region (or meet applicable conditions for international or cross-border access), including compliance with national implementing legislation of the EHDS and GDPR where applicable

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- Demonstrate a legitimate, clearly defined purpose such as research, innovation, public health, or policy support, with a documented lawful basis for data processing under GDPR (e.g., Article 6 and, where applicable, Article 9 for special categories of data), and adherence to purpose limitation)

- Put in place appropriate technical and organisational measures to ensure data protection, security, and compliance with GDPR, national legislation, and institutional policies. GDPR and data protection best practices documented by RDMkit – GDPR compliance page.

- Obtain approval from a competent Research Ethics Committee or Institutional Review Board. In addition, ensure provision of a Data Protection Impact Assessment (DPIA) where processing is likely to result in high risk.

- ***How can cross-border access be arranged while respecting local laws?***

Even though the European Health Data Space (EHDS) is not yet operational, projects can still plan for responsible cross-border health data access by following national, regional, and institutional frameworks, while aligning with emerging European guidance such as TEHDAS. They key principles for data access are:

- **Identify the relevant data holders and access bodies**: Determine who controls or manages the datasets in each country (e.g., hospitals, data repository, research infrastructures). These bodies will define the access requirements.

- **Understand local legal and regulatory requirements**: Comply with GDPR, national data protection laws and any institutional policies. GDPR, while designed to protect individuals' data, is interpreted differently across EU Member States and even among institutions within the same country[10]

- **Submit clear access requests**: Applications should specify the project objectives, datasets needed and intended analyses. Include details on how data protection, security, and compliance will be ensured. All this information should be included in the project's DMP.

- **Use secure data access and processing arrangement**s: Access is often provided through secure processing environments to ensure data protection while allowing multi-region collaboration.

- **Follow conditions for ongoing compliance**: Respect reporting obligations, output restrictions, and audit requirements set by the data holder. Ensure that all cross-border data use remains aligned with local laws and institutional policies.

When the EHDS is in place, cross-border access will be facilitated through the following mechanisms:

- **Single access point via Health Data Access Bodies** (HDABs): Data users apply to the HDAB in the country where the data is held. Once authorised, access is granted regardless of the applicant's country of establishment, enabling EU-wide use of the data.

[10] Gwen Gilderson (2025). PRECISEU D6.1 Cross Regional Report, Sahlgrenska Science Park

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **EU-wide recognition of permitted secondary uses**: EHDS defines a common list of allowed purposes (e.g. research, innovation, public health), reducing legal uncertainty and preventing divergent national interpretations from blocking cross-border projects.

- **Standardised access conditions and procedures**: EHDS introduces harmonised application requirements, access conditions, and safeguards, which significantly reduce fragmentation while ensuring consistency across Member States.

- **Respect for justified national constraints**: While EHDS enables cross-border access, Member States may apply additional safeguards in narrowly defined areas (such as particularly sensitive data types or ethical requirements), provided these are compatible with EHDS and EU law.

- **Built-in GDPR compliance**: EHDS operationalises GDPR principles—such as purpose limitation, data minimisation, and security—ensuring that cross-border access occurs within a strong data protection framework.

This means that data made available under EHDS will be accessible to eligible users across Europe, but access may be subject to specific conditions defined by the data-holding country, as assessed and enforced by its HDAB. PRECISEU D6.1 *Cross Regional Report* contains more information on health data accessibility.

- ***How is cross-border data securely transferred or accessed?***

Even without EHDS, cross-border access to health data for secondary use should minimise data transfers and maximise security. As a general rule, secure remote access is better than physical transfer of data. Cross-border access can be enabled in the following ways:

- **Access through Secure Processing Environments (SPEs):** Health data should be accessed and processed within an SPE when it is not fully anonymised or aggregated (e.g. pseudonymised data). These environments should promote user-friendly research user experience, interoperability, scalability and performance to handle large/complex and advanced analytics. Researchers and project partners from other EU Member States can access the data remotely, without the data leaving the country where it is held. Technical models for such environments are described in TEHDAS guidance on Secure Processing Environments.

- **Limited and controlled data transfer where authorised**: Where justified and explicitly authorised by the right ethical committee, certain data outputs or datasets may be transferred across borders. Such transfers are subject to strict conditions, including encryption, pseudonymisation or anonymisation, and compliance with GDPR requirements.

- **Federated and distributed access models**: Federated approaches in which analyses are performed locally on national datasets and only approved results or aggregated outputs are shared across borders. This allows joint interregional projects to obtain information without centralising sensitive health data. In this case, each participating node must operate within a Secure Processing Environment.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Strong technical and organisational safeguards**: Cross-border access is protected through authentication, role-based access controls, logging and monitoring, and output controls to prevent re-identification or unauthorised disclosure.

This means that most cross-border projects will access data remotely within secure environments, rather than receiving copies of datasets, ensuring a high level of protection while enabling EU-wide research and innovation.

- ***What agreements (e.g., Data Use Agreements) are required for access?***

Access typically requires one or more of the following:

- **Data Use Agreement (DUA) or Data Access Agreement**: Defines permitted uses, duration, security measures, and responsibilities
- **Consortium or Collaboration Agreements**: When multiple regions or partners are involved
- **Ethics approvals**: Required in some projects to be granted if clinical or sensitive data is used. This is especially important if there has to be data transfers across regions and no SPE can be used
- **Data Protection Impact Assessment (DPIA)**: Particularly when processing large-scale or sensitive health data.

In most EHDS use cases, a Data Transfer Agreement is *not required* because:

- Remote access through Secure Processing Environments (SPEs) is the preferred option
- Health data generally does not leave the country of origin
- No physical or logical "transfer" of raw data to the user occurs

However, a Data Transfer Agreement *may be required* in limited cases where:

- The HDAB explicitly authorises a cross-border transfer of data (e.g. specific datasets or outputs)
- Approved outputs are transferred to another environment
- Data is shared between jointly controlled environments

- ***Are there differences in access rules between countries/regions?***

Yes, differences can exist despite European regulations.

- **Harmonised** principles:
  - Core legal principles such as GDPR compliance, data protection safeguards, and clearly defined purposes for secondary use are generally consistent across EU countries.
  - National or institutional bodies (e.g., data access committees or ethics boards) enforce these principles.
- **Differences** remain due to:
  - National laws and ethical frameworks may impose additional requirements for sensitive data, such as genomics, rare diseases, or data on minors.
  - Procedures for access, including application forms, review processes, and approvals, differ between countries and institutions.
  - Technical infrastructure, secure processing environments, and rules on output or data transfer may also vary.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Implication** for projects:
  - While the legal basis is harmonised, operational procedures may differ by country
  - Projects should plan timelines and workflows accordingly, especially for multi-country studies

EHDS creates a shared legal framework, but access mechanisms and procedural details are still country-specific.

Relevant to DMP section(s):
● Making Data Openly Accessible

2.6 Data Use

The use of health data in this context refers to how health information is analysed and applied to support healthcare, research, innovation and personalised medicine across regions and countries. In this phase you will manipulate data in a trusted and secure way (e.g. using SPEs) to perform analyses using the information you have been granted access.

- ***How can you use/process cross-border data under EU principles?***

As already mentioned in the 2.5 *Data Access* section, once you got access to the health data, it should be processed within an SPE when it is not fully anonymised or aggregated (e.g. pseudonymised data). Researchers and project partners from other EU Member States can access the data remotely, without the data leaving the country where it is held. More information can be found in TEHDAS guidance on Secure Processing Environments.

In addition to the use of SPEs, there are two key techniques that enable the use of sensitive data for analysis: anonymisation and pseudonymisation.

Anonymisation is the process of irreversibly removing or altering personal identifiers from a dataset so that individuals can no longer be identified. Proper anonymisation places the data outside the scope of the GDPR, enabling safe reuse for research and other secondary purposes. Strong anonymisation often reduces the richness of the dataset, making it less useful for research or statistics. Therefore, there is a trade-off between privacy protection and analytical value. There are several anonymisation methods such as perturbation, generalisation, suppression, aggregation, encryption and tokenisation. More information on anonymisation methods can be found here.

Pseudonymisation is the process of transforming personal data so it cannot be directly linked to an individual without securely and separately stored additional information, often by replacing identifiers with codes or aliases. According to the European Guidelines to pseudonymisation there are two main ways of data pseudonymisation: *a) cryptographic algorithms* and *b) lookup tables*. In the former, the real identifier (e.g. a name or ID number) is run through a secure algorithm together with a secret key to produce a pseudonym. Without access to that secret key, it is extremely difficult to determine who the data belongs to. This is comparable to locking the identifier with a secret combination. It is an efficient way to pseudonymise data, but the algorithms and cryptographic keys must be regularly reviewed and updated as technology advances.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

In the lookup tables, each real identifier is replaced with a randomly generated pseudonym, and the controller keeps a separate, securely stored table that maps real identities to pseudonyms. Anyone without access to the table cannot re-identify people. This method is easily reversible for authorised users, but it requires securely storing and protecting the table, which itself is sensitive personal data. You can find more information on pseudonymisation techniques here.

Pseudonymised data remains personal data and is subject to GDPR, while for anonymised data GDPR does not apply. Please refer to TEHDAS documentation for further information, tools and practices on data (pseudo-)anonymisation and data minimisation.

For cross-border projects, federated approaches are important to consider, in which analyses are performed locally on national datasets and only approved results or aggregated outputs are shared across borders. This allows joint interregional projects to obtain information without centralising sensitive health data. In this case, each participating node must operate within a Secure Processing Environment.

- ***What types of data can be used (clinical, genomic, lifestyle, etc.)?***

Again, the data that you use/process/analyse should be meaningful for your specific project. A wide range of health-related data can be used, depending on the purpose and permissions you have. This includes clinical data from electronic health records, genomic or other molecular data for personalised medicine, lifestyle and behavioural data, imaging data, lab results and administrative or registry data. These different types of data can be combined to support the objectives of your project and should be detailed as much as possible in the project's Data Management Plan.

- ***Are there any restrictions that forbid data use (e.g. opt-out)?***

Under Art. 71(1) of the EHDS Regulation patients have the right to control who accesses their data. This can be done through opt-out mechanisms, in which patients and citizens can voluntarily express that they do not wish to share their data. This means that if a patient has refused consent or formally opted out, their data cannot be used for secondary purposes such as research or policy analyses. This right is generally unconditional and reversible. The EHDS opt-out right should not be confused with consent under the GDPR. Secondary use under EHDS is generally based on a legal access framework established by EU and national law, rather than on individual consent. However, national research legislation may still require consent in certain situations outside or in addition to the EHDS mechanism.

Regarding consent, there are some variations across countries. Some (e.g. Finland) allow broad consent for research, while others (e.g. Germany) require explicit permission for each use. Different regional consent practices could result in asymmetric consent, which represents an ethical issue. While the EHDS is not fully implemented, the potential impact on representativeness, bias and equity should be transparently assessed and documented.

Also, if you plan to use data in your project that is not fully anonymised or aggregated but you do not have a secure environment to process it, restrictions may apply.

It is recommended to check any potential restrictions for using data in your project with the data provider, ethics committee and relevant data protection or governance authorities; also, to review any data access agreements to understand what specific restrictions apply.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***What data processing tools are you planning to use for your project?***

You should use data processing tools that are appropriate and relevant for your study. Under the EHDS personal and sensitive data processing can only take place in secure processing environments that comply with the highest standards of privacy and cybersecurity and no personal data can be downloaded from these environments. If you have fully anonymised or aggregated data, this can be processed other tools such as statistical tools (e.g. R, Python), visualization tools (e.g. Microsoft®PowerBI, Tableau) or spreadsheets (e.g. Microsoft®Excel). Also, other tools which are not specifically for processing but are especially relevant in cross-border and collaborative projects, such as collaboration, control-version and reporting tools (e.g. GitHub, GitLab).

Relevant to DMP section(s):
• Data Summary • Increase Data Re-use • Data Security • Ethics

2.7 Data Finalisation

In the context of the data life cycle presented in these Guidelines, data finalisation is the last step, where decisions are made about what happens to data after it has been used. This means considering whether health data is securely kept for future treatment and research in the long-term, safely deleted or decommissioned if required, in line with legal and ethical rules; and learnings and further improvement documented.

Frequently asked questions

- ***How should datasets be managed for long-term preservation?***

Proper long-term preservation ensures that research outputs remain findable, accessible, interoperable and reusable (FAIR), supporting transparency, reproducibility and future research. For long-term preservation, datasets should be managed as follows in your project:

- Use **trusted repositorie**s. Deposit datasets in recognised institutional, national or domain-specific repositories that support long-term preservation and provide persistent identifiers (such as DOIs).
- Include **clear and sufficient documentation and metadata** so that others can understand and reuse the data in the future. This should cover:
    - how the data were created
    - what the data contain and their provenance
    - any limitations or conditions for reuse.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- Choose **sustainable file formats**. Store data in open, non-proprietary and widely used formats (e.g. CSV instead of XLSX, TXT instead of DOCX) to reduce the risk of software obsolescence

- Define **access and licensing conditions**. Clearly state access restrictions, ethical constraints and reuse licences. If data cannot be fully open, explain why and under what conditions access may be granted

- Ensure **data integrity and security**. Protect datasets against loss, corruption or unauthorised access through backups, checksums and appropriate access controls

- Plan **preservation in early stages of the project**. Long-term preservation requirements should be considered from the start of the project and documented in the project's Data Management Plan.

- Ensure that **datasets are stored in a way that complies with all relevant regional regulations**. For example, some data may need to remain in the country of origin or meet specific encryption requirement.

- Check **what data needs to be preserved and for how long**. For example, those required by preserved by legal or ethical requirements or by the funder, publisher and/or institutional policies. Usually, preservation is considered for 5-10 years after the project ends.

You can find more information on data preservation in the ELIXIR RDMkit.

- ***When and how should data be deleted or decommissioned, if required?***

Controlled data deletion and decommissioning protect individuals' rights, ensures legal compliance and maintains trust, while proper documentation preserves accountability and transparency. When thinking about data deletion or decommissioning in your project, you should follow these principles:

- **Delete data only when justified**. Data may need to be deleted when required by data protection laws (e.g. GDPR), ethical obligations, consent agreements, contractual terms or institutional policies. If no such requirement exists, long-term preservation is generally preferred

- **Define retention periods in advance**. Retention periods and deletion conditions should be specified early, ideally in the project's Data Management Plan, to avoid ad-hoc or accidental data loss

- **Assess before deletion**. Before deleting data, verify whether:
  - the data have long-term scientific or societal value
  - anonymisation or restricted access could be used instead of deletion
  - related metadata should be retained for record-keeping or transparency.
- **Use secure deletion methods**. Data must be deleted using methods appropriate to the storage system (e.g. secure erasure for digital storage, certified destruction for physical media) to prevent unauthorised recovery
- **Document the process**. Record which data were deleted, when, why and by whom. Where applicable, keep metadata or summary information even after the data are removed

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Follow repository and institutional procedures**. When data are stored in repositories or institutional systems, follow their established decommissioning and withdrawal policies rather than deleting files manually

- **Deletion policies must respect regional legal requirements**. Some regions may mandate longer retention periods or require certified destruction methods. Check with if your institution, region or funder has any specific requirements

You can find more information on data deletion here.

- ***How do you document project learnings and improvements for future projects?***

Documenting learnings will help you and your team continuously improve data management practices, avoid repeating mistakes and build institutional knowledge over time.

Project learnings and improvements should be documented as follows:

- **Record lessons learned throughout the project**. Capture insights on what worked well, what did not, and what could be improved in areas such as data collection, documentation, tools, workflows and compliance. This should not be left only to the project end.

- **Use standard templates and formats**. Document learnings using agreed templates (e.g. "lessons learned" reports, final project summaries or data management reviews) to ensure consistency and comparability across projects and partners

- **Link learnings to concrete actions**. Where possible, describe how an issue was addressed or how a process was improved, so future projects can avoid the same issue or apply the same solution

- **Store documentation in shared, long-term locations**. Make lessons learned available in institutional repositories, project archives or knowledge bases that are accessible to relevant stakeholders

- **Include metadata and context**. Clearly indicate the project scope, timeframe, tools used and data types involved, so future users can understand the relevance of the learning

- **Review and reuse**. Encourage future projects to consult previous lessons learned during project planning and data management plan preparation

- Make sure documentation **captures any region-specific lessons or constraints**, so future projects in other regions can apply learnings without violating local rule.

Relevant to DMP section(s):
• Data Summary • Making Data Findable • Increase Data Re-use • Data Security

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

2.8 Project Management & Governance

Transparent and effective project management and governance are essential for a responsible use of health data across countries, regions and institutions. Clear roles, decision-making processes and coordination structures will help all partners in your project work together efficiently. When data is particularly sensitive, such as health data, good governance and project management ensure transparency, trust and ethical data use throughout the project. As a result, it is advisable to consider a specific governance structure in your project from the early stages and clearly depict it within the Data Management Plan.

Some stakeholders, such as SMEs and healthcare professionals, are usually underrepresented in governance discussions, mainly due to limited resources. Make sure all stakeholders in your project are well informed and represented. Project monitoring and reporting are handled through clear governance, standardised documentation, and continuous communication, ensuring that everybody in your project remains aligned, compliant, and result-oriented throughout the project lifecycle.

Frequently asked questions

- ***How do we handle data from countries / regions that are at different levels of maturity in terms of the EHDS?***

Data from countries or regions that are at different stages of readiness for the **European Health Data Space (EHDS)** should be managed through a **flexible and harmonised approach** that respects both inclusion and compliance:

- Build a **baseline framework that all members accept** (incident handling, logging, etc.)
- **Assess local readiness**: Identify each region's level of alignment with EHDS principles (governance, infrastructure, interoperability, and legal frameworks). More information can be found on D4.1 *PRECISEU readiness framework*
- **Apply common minimum standards**: Even where EHDS alignment is partial, use EU-level principles – such as GDPR compliance, FAIR data, and security-by-design – as a shared foundation
- **Use federated or hybrid models**: When full data integration is not feasible, allow local data to remain within national systems and enable access through federated architectures or secure queries
- **Ensure interoperability**: Adopt standardised formats, metadata, and APIs that facilitate future EHDS integration. More information can be found on D4.2 *PRECISEU Interoperability Framework*
- **Provide support and capacity building**: Work with less mature partners to strengthen their technical and legal capacity to meet EHDS requirements
- **Maintain transparent governance**: Use clear agreements and a common project's Data Management Plan to align responsibilities and safeguard trust.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***How do we handle legal and regulatory differences between countries/regions?***

Legal and regulatory differences across countries is a key challenge in international data collaboration. To manage these differences effectively, you should adopt a structured, transparent, and harmonised approach:

- **Map the legal landscape**: Identify early in the project which national or regional laws apply (e.g., data protection, ethics, health data access, secondary use)
- **Apply the highest common standard**: When in doubt, align all activities with the strictest applicable regulation — typically GDPR-level protection — to ensure overall compliance
- **Use standardised agreements**: Implement harmonised templates such as Data Transfer Agreements (DTA), Joint Controller Agreements, or Standard Contractual Clauses (SCC) for data exchange across jurisdictions
- **Rely on risk-based compliance**: For partners outside the EU or with different legal frameworks, conduct Data Protection Impact Assessments (DPIAs) and Transfer Impact Assessments (TIAs) to identify and mitigate risks
- **Ensure local legal input**: Engage local legal experts or data protection officers to validate compliance with national laws and ethics requirements
- **Document governance decisions**: Maintain a clear record of all legal bases, consents, and agreements to demonstrate accountability and simplify audits
- **Promote mutual understanding**: Organise training and knowledge exchange among partners to improve awareness of different legal environments.

In practice, the goal is not to eliminate differences, but to create a shared legal baseline and transparent governance, ensuring that all partners operate safely within their respective regulatory contexts while enabling responsible data sharing.

- ***How will project monitoring or reporting be managed?***

Project monitoring and reporting will be managed through a structured governance framework that ensures transparency, accountability, and compliance with both technical and ethical standards.

Key elements include:

- **Defined governance structure**: A central coordination team or project management office (PMO) oversees progress, ensures deliverables are met, and monitors adherence to timelines and budgets
- **Regular reporting cycles**: Partners submit periodic technical and financial reports summarising activities, achievements, risks, and deviations. These reports feed into consolidated updates to the funder or steering committee

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Use of standard project tools**: A shared project management platform (e.g., for milestones, data tracking, and document management) supports version control, transparency, and accessibility for all partners

- **Key performance indicators (KPIs)**: Quantitative and qualitative indicators measure progress against objectives - including data sharing milestones, compliance checks, and stakeholder engagement metrics

- **Risk management and mitigation**: Continuous identification and tracking of risks related to data, ethics, or regulatory compliance, with defined escalation procedures

- **Internal and external reviews**: Regular internal reviews (e.g., quarterly consortium meetings) and external audits or evaluations ensure quality and accountability.

- **Data and quality reporting alignment**: Data quality, integrity, and security are reported following FAIR and GDPR principles, integrated within the project's Data Management Plan (DMP).

Relevant to DMP section(s):
• Data Summary • Ethics

2.9 Data Privacy and Security

Data privacy and security are considered cornerstones in the use of health data. This type of data is highly sensitive and needs to be compliant to ethical and legal requirements, such as GDPR. For cross-border projects this requires ensuring lawful, transparent and purpose-limited data processing supported by robust technical and organisational security measures. This is essential not only for regulatory compliance but to enable safe data sharing across regions and maintain trust, fostering high-quality research and innovation in personalised medicine. While cross-border projects yield multiple benefits, security and citizen trust gaps remain ongoing challenges.

Frequently asked questions

- ***How is patient privacy ensured when sharing data across borders?***

Health data is considered as 'highly sensitive'. Privacy and security in health data refer to the protection of an individual's health data through strict access controls, consent management, anonymisation or pseudonymisation and secure processing environments. They ensure that data are used only for authorised purposes while maintaining confidentiality and integrity. This way, secondary use of data is in line with personal rights protection and promotes public trust. Different practices (e.g. data (pseudo-)anonymisation, federated analysis or federated learning), legal and governance requirements and tools (e.g. SPEs) take into consideration patient privacy when sharing data across borders.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Data privacy and security are especially relevant when sharing health data across borders. For secondary use, the EHDS enables secure and interoperable data exchange between Member States at European level via the HealthData@EU infrastructure. This includes not only safe measures for data access but also for data use, storage and analysis via Secure Processing Environments (SPEs). SPEs are a key element of the EHDS which allow for the secondary use of health data while upholding requirements for data protection, confidentiality, and information security. The EHDS states that personal data processing can only take place in these secure processing environments that comply with the highest standards of privacy and cybersecurity. Before transferring data to the SPE, the HDAB must implement encryption and access control mechanisms to prevent unauthorised access. However, there are still some concerns in how data originating from different countries or SPEs will be handled, as there is no clear framework for collaboration between SPEs or for ensuring interoperability across HDABs and EU Member states.

You can find here more information on how to use data securely using SPEs.

Data privacy incorporates the concept of patient's control over their data. The EHDS empowers individuals to access, control and share their electronic health data across borders for primary use, and to make it available for secondary use. In case patients do not wish their data to be reused for secondary purposes, opt-out mechanisms are provided.

More information about data privacy and security can be found here and here. Please also refer to the 'Data Security' section of the Data Management Plan.

- ***What level of (pseudo)anonymisation is appropriate before data are shared or processed for secondary use?***

Cross-border data sharing in health research requires careful consideration of (pseudo)anonymisation. Before data are shared or processed for secondary use, identifiers should be removed or replaced with pseudonyms to minimise the risk of re-identification. Different Member States or regions may interpret anonymisation and pseudonymisation requirements differently, meaning that a method considered sufficient in one country may fall short in another.

Given the sensitivity of personal electronic health data, it should be pseudonymised or anonymised as early as possible in the process of making data available for secondary use, ideally during the data preparation stage. However, these protections should be maintained throughout the whole Data Life Cycle. For example, when analysing and processing data using an SPE, results (including metadata) should be provided in an anonymous format.

A key aspect of pseudonymisation could be the ability to share, for the same subjects, data originating from different sources, not necessarily across borders. In such cases, the involvement of a trusted third party providing multiple data holders with a pure pseudonymisation service — without any knowledge of the content or context of the data being processed — and ensuring full compliance with GDPR privacy requirements, can be highly beneficial. Complementary to this technique, pseudonymised metadata should be managed by an additional data holder, so that health data managers cannot identify the assigned pseudonym through the combination of their data and metadata.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- **Apply data minimisation:** Before anything else, only the data necessary for the stated purpose should be shared (data minimisation). This concept is related to the GDPR and lies on the fact that only the minimum amount of personal health data that is adequate, relevant and limited to what is necessary for a specific purpose should be processed. Data minimisation must be respected to ensure data privacy. HDABs should assess whether all requested variables are required.

- **Prefer anonymisation when feasible:** If the research or secondary use purpose can be fulfilled without linking back to individuals (i.e. statistical or aggregate use) then data should be anonymised to the greatest extent possible.

- **Use pseudonymisation when necessary:** When fully anonymised data are not sufficient (for example, for longitudinal or individual-level analysis), pseudonymised data may be provided. In that case, the reidentification key (or "additional information") should not be held by you, but by the Health Data Access Body (HDAB). The deidentification process must be fully documented.

- The **level of (pseudo)anonymisation** should be **tailored to the project's context,** considering the identifiability risk, how data are linked, the number of subjects and the potential harm in case of reidentification. This corresponds to a risk-based mindset, rather than a one-size-fits-all rule.

When creating your Data Management plan, you must specify any information regarding data (pseudo)anonymisation measures for your project. More information on this topic can be found here and here.

- ***Are there any other alternatives to use data while preserving privacy?***

When data (pseudo)anonymisation is not possible, generalisation, suppression and randomisation of personal data can also be employed. Instead of replacing data with identifiers, the data itself is modified to reduce identifiability, often making re-identification impossible or much harder without significantly compromising data utility.

- **Generalisation** replaces specific values with broader categories (e.g., age ranges instead of exact birth dates).

- **Suppression** removes or hides high-risk data that could identify individuals (e.g., rare conditions).

- **Randomisation** introduces small, controlled variations in data to mask exact values while preserving trends.

Other techniques, such as federated analysis, in which only aggregated results are shared and the raw data never leaves its original location - are inherently privacy-preserving. They can also be combined with the methods above. This ensures compliance with different jurisdictions and reduces the risk of re-identification, even when only aggregated results are shared. In all cases, the principle of data minimisation must be respected.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

- ***What should be done in case of a security breach?***

A **personal data breach** is a security incident that consists in the accidental or unlawful destruction, loss, modification, unauthorised disclosure or access to personal data (transferred, stored or otherwise processed), as defined under the GDPR. Examples of data breaches could be (but are not limited to):

- Sending a file with patient data to the wrong email address
- Losing a laptop or USB stick containing health records
- A hacker accessing a database of user information
- Unauthorised internal access to personal data beyond assigned roles or permissions
- Malware attacks affecting the availability or integrity of health data

Breaches can trigger different legal obligations across countries, delay project timelines and affect trust between partners, making careful data handling and clear reporting essential.

In order to minimise effects from data security breaches:

- **Define clear roles**, establishing who is the *controller* and who is the *processor* of data in your project. This determines who must notify whom in case of a breach.

- As part of your governance arrangements, **design and agree a shared breach management plan** that includes detection, assessment, notification, communication and documentation. This should be also included in the Data Management Plan.

- Use **secure technical measures**, such as encrypting sensitive health data, ensuring and monitoring backups, using access controls.

- Make sure **everyone involved in the project can spot a potential breach and understands their responsibilities and the steps to fix it**.

- After any incident, **run a *postmortem***: what worked, what did not and improve your processes accordingly.

You can find more information on security breaches here.

- ***How can results be shared while protecting privacy and complying with GDPR and other EHDS requirements?***

Sharing data in personalised medicine is particularly relevant, as it enables prevention, diagnosis and treatments. At the same time, health data are highly sensitive and require strong safeguards to protect patients' privacy and to comply with GDPR – where health data are classified as a special category of personal data - and the European Health Data Space (EHDS), which involves additional requirements for research, innovation, policy-making and clinical care. These challenges are amplified when conducting projects across different populations, regions or countries.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

1. **Share results, not individual data**. Results should be disseminated in (pseudo)-anonymised, aggregated or model-based forms. Individual patients should never be identifiable from published outputs. Particular care should be taken in personalised medicine, where rare conditions or highly specific profiles may increase re-identification risks. Check GDPR art. 5(1)(c) and 5(1)(b) and this TEHDAS guidance for more information

2. **Integrate data protection measures from the very beginning in your project**, adopting a privacy-by-design approach. Use pseudonymisation throughout data processing, keeping re-identification keys strictly separated and protected. Apply privacy-preserving analytics, such as federated learning or distributed analysis, especially for cross-border projects where data remain within national or institutional boundaries. Finally, a good practice is to assess, audit and document re-identification risks – especially when combining datasets from different countries. Check GDPR art. 25 and 32 and this TEHDAS guidance for more information

3. **Use secure, harmonised infrastructures for cross-border collaboration**. Cross-border sharing of results should take place through trusted and secure platforms, rather than uncontrolled data exchanges. This involves the use of trusted and EHDS-compliant research environments and access by trusted HDABs. Implement role-based access, logging and audit trails; and use common data formats and standards to ensure consistency across countries. Check GDPR art. 32 and 44-49 and this TEHDAS guidance for more information.

4. Make sure your **data is shared and reused only for clearly defined and legitimate purposes**. Document the intended uses in your project's Data Management Plan and data sharing agreements; assess whether the reuse of results is compatible with the original purpose of the data and explicitly forbid uses that could lead to misuse (e.g. discrimination, insurance decision-making, etc.) Check GDPR art. 5(1)(b), 6 and 9 and this TEHDAS guidance for more information.

5. Ensure **your project clearly states how data contributes to personalised medicine and cross-border collaboration**, providing transparent information on data use and sharing. Publish summaries of privacy safeguard and governance and ethical reviews. This enhances transparency and social trust, key under the EHDS. Also, involve the right experts (DPOs, clinical experts, ethical boards, etc.) and conduct DPIAs for high-risk projects (e.g. those involving AI). Check GDPR art. 12, 14, 24 and 35 and this TEHDAS guidance for more information.

Relevant to DMP section(s):
• Data Security • Ethics

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

3. REFERENCES

- Cascini, F. (2025). *Secondary Use of Electronic Health Data*. *SpringerBriefs in Public Health*. Cham: Springer Nature Switzerland *[accessed 2025]* doi: https://doi.org/10.1007/978-3-031-88497-9
- European Commission. European Health Data Space Regulation (EHDS) [Internet]. Brussels: European Commission [accessed 2026]. Available from: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en
- European Data Protection Board. Data breaches [Internet]. EDPB; [accessed 2026]. Available from: https://www.edpb.europa.eu/sme-data-protection-guide/data-breaches_en
- European Data Protection Board. Guidelines 01/2025 on Pseudonymisation [Internet]. EDPB; 2025 [accessed 2026]. Available from: https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en
- European Data Protection Board. Secure personal data [Internet]. EDPB; [accessed 2026]. Available from: https://www.edpb.europa.eu/sme-data-protection-guide/secure-personal-data_en
- European Innovative Health Initiative & European Federation of Pharmaceutical Industries and Associations. Data Sharing Playbook 2024 [Internet]. [accessed 2026]. Available from: https://www.ihi.europa.eu/sites/default/files/uploads/Documents/ProjectResources/IMI_IHI_DataSharingPlayBook_2024.pdf
- European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU) [Internet]. EMA; [accessed 2026]. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/real-world-evidence/data-analysis-real-world-interrogation-network-darwin-eu
- European Partnership for Personalised Medicine (EP PerMed). Guidelines for data reusability. Version 2.1 [Internet]. 2025 [accessed 2026]. Available from: https://www.eppermed.eu/wp-content/uploads/EPPerMed_Guidelines-for-data-reusability_v2-1.pdf
- European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union: L 119/1–88, 04 May 2016 [accessed 2026]. Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- European Union. HealthData@EU – Data acceptance portal [Internet]. European Union; [accessed 2026]. Available from: https://acceptance.data.health.europa.eu/
- GO FAIR. Principles for scientific data management and stewardship guiding FAIR data practices. FAIR Principles: Findable, Accessible, Interoperable, Reusable [Internet]. GO FAIR; [accessed 2026]. Available from: https://www.go-fair.org/fair-principles/
- PRECISEU 4.2 Interoperability Framework
- PRECISEU 4.4 Data Management Template for Joint Interregional Projects
- PRECISEU D4.1 PRECISEU Readiness Framework (Readiness/maturity assessment framework) [accessed 2026]. Available from: https://preciseu.eu/wp-content/uploads/2025/10/D4.1-Asessing-readiness-to-EHDS.pdf
- QUANTUM – The health data quality label [Internet]. QUANTUM project; [accessed 2026]. Available from: https://quantumproject.eu/
- Svingel LS, Jensen CE, Kjeldsen GF, Pedersen MH, Kalra D, Christiansen CF, Vad KH. Shaping the future EHDS: recommendations for implementation of Health Data Access Bodies in the HealthData@EU infrastructure for secondary use of electronic health data [Internet]. European Journal of Public Health. 2025 Sep;35(Suppl 3):iii32–iii38 [accessed 2026]. Available from: https://doi.org/10.1093/eurpub/ckaf033
- TEHDAS .Joint action and project information on the secondary use of health data and the European Health Data Space. Towards the European Health Data Space [Internet]. TEHDAS; [accessed 2026]. Available from: https://tehdas.eu/
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, et al. The FAIR Guiding Principles for scientific data management and stewardship [Internet]. Sci Data. 2016;3:160018 [accessed 2026]. Available from: https://doi.org/10.1038/sdata.2016.18
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, et al. The FAIR Guiding Principles for scientific data management and stewardship [Internet]. Scientific Data. 2016;3:160018 [accessed 2026]. Available from: https://doi.org/10.1038/sdata.2016.18
- World Health Organization. Sharing and reuse of health-related data for research purposes: WHO policy and implementation guidance [Internet]. Geneva: World Health Organization; 2022 [accessed 2026]. Available from: https://www.who.int/publications/i/item/9789240044968

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

4. ANNEX: CASE EXAMPLES

Below you can find some case examples of cross-border scenarios and initiatives that demonstrate the value of health data sharing across regions and countries, as well as showcase the various approaches to address the fundamental tension between maximising data use and protecting patient privacy and data sovereignty.

***Case example 1: FLUTE - Federated* Learning and mUlti-party computation Techniques for prostatE cancer**

Cross-border relevance	FLUTE demonstrates privacy-preserving cross-border health data use by integrating health data hubs from three different European countries (Italy, Belgium, and Spain) to develop AI solutions for prostate cancer diagnosis.
Description	FLUTE develops novel methods for privacy-preserving cross-border federated learning in healthcare, advancing secure multi-party computation for AI model development. The platform enables solution developers to create AI tools that allow healthcare professionals to analyse their patient data and compare it with data from other hospitals and research centres across borders while maintaining strict GDPR compliance. The project focuses on developing a federated AI toolset for diagnosing clinically significant prostate cancer using multiparametric MRI imaging data, with multinational clinical validation across participating sites.
Added value of cross-border datasets	Accessing diverse patient cohorts from multiple countries improves the accuracy and robustness of AI prediction models for aggressive prostate cancer, and reduces bias linked to single-country healthcare practices. The integration of data from different healthcare systems and populations enhances model generalisability while avoiding unnecessary biopsies and reducing healthcare costs.
Interoperability testing potential	**Technical**: Contribution to HL7 FHIR standard development, privacy-enhancing technologies (secure multi-party computation), generation and utilisation of synthetic data **Semantic**: Common data models for medical imaging (MRI) data **Organisational**: Development of GDPR-compliant cross-border federated learning guidelines **Legal**: Novel frameworks for GDPR-compliant cross-border data utilisation, privacy metric tools for certification
Type of data use	Secondary use (research and development of AI diagnostic tools)
Website	http://fluteproject.eu/

Case example 2: IDERHA - Integration of heterogeneous Data and Evidence towards Regulatory and HTA Acceptance

Cross-border relevance	IDERHA builds one of the first pan-European health data spaces aligned with European Health Data Space (EHDS) requirements, creating a federated data infrastructure connecting multiple European institutions.
Description	IDERHA creates a scalable platform for seamless integration and linkage of diverse real-world data across European borders to support healthcare professionals, patients and researchers in improving lung cancer patient outcomes. The project establishes an open, disease-agnostic federated data space enabling connectivity, access, use and reuse of digital health data from multiple countries. Using lung cancer as a pilot disease, the platform analyses

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

	data along the entire patient journey, from early screening of at-risk citizens to remote monitoring of late-stage patients, employing AI and machine learning for improved early detection and personalised care.
Added value of cross-border datasets	Integration of heterogeneous data types (electronic health records, imaging, genomics, patient-generated health data) from multiple European institutions enables development of more comprehensive AI/ML algorithms for risk profiling, malignancy prediction, diagnosis and prognosis. Cross-border data access facilitates better understanding of disease patterns across diverse populations and healthcare systems.
Interoperability testing potential	**Technical**: HL7 FHIR, DICOM, OMOP Common Data Model, Eclipse Data Space connectors, DCAT-AP metadata standards **Semantic**: Standardised terminologies and ontologies, FAIR data principles **Organisational**: Health Data Access Bodies coordination, federated machine learning infrastructure **Legal**: Alignment with EHDS2 requirements, data sovereignty principles, consensus policy recommendations for regulatory and HTA decision-making
Type of data use	Both primary use (enabling personalised care and remote patient monitoring) and secondary use (research, innovation, policy-making, and regulatory decision-making for HTA acceptance)
Website	http://iderha.org/

Case example 3: SHAIPED - Supporting Health Data Access Bodies to establish AI pathways enabling Deployment of AI as medical device tools

Cross-border relevance	SHAIPED creates a European consortium of 30 partners from 11 EU member states to establish AI pathways for medical device development, testing and deployment across borders, directly implementing the European Health Data Space framework.
Description	SHAIPED addresses the challenge of accessing real-world health data across European borders for training, testing and validating AI-based medical devices. The project facilitates cross-border collaboration between Health Data Access Bodies (HDABs) by creating frameworks and agreements for data sharing that ensure developers have necessary datasets to train and validate AI tools. Three clinical use cases demonstrate the platform: chronic kidney disease management, metastasis detection for lung cancer and breast cancer, and heart failure prevention in patients with cardiac implantable devices.
Added value of cross-border datasets	Access to diverse, high-quality health data from multiple European countries enables AI models to adapt to varied patient cohorts, improving clinical effectiveness at scale and accelerating time-to-market for AI medical devices while ensuring they meet EU standards across different populations.
Interoperability testing potential	**Technical**: Standardised integration of AI medical devices into European health information systems, compliance with European Health Data Space infrastructure **Organisational**: Optimised collaboration between Health Data Access Bodies across Europe **Legal**: Leverages EHDS Regulation and AI Act for improved patient outcomes, compliance with highest data protection standards
Type of data use	Secondary use (development, testing, and validation of AI-based medical devices for research, innovation and improved healthcare delivery)
Website	http://shaiped.eu/

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Case example 4: TRUMPET - TRUstworthy Multi-site Privacy Enhancing Technologies

Cross-border relevance	TRUMPET develops a privacy-enhanced federated AI platform enabling analysis of siloed, multi-site, cross-domain, cross-border European datasets, with partners from Belgium, France, Italy, Spain and the United Kingdom.
Description	TRUMPET addresses privacy concerns in federated learning by developing novel privacy enhancement methods and delivering a highly scalable federated AI service platform. The platform enables researchers to conduct AI-powered studies on European cancer datasets that were previously inaccessible due to privacy concerns, while ensuring GDPR compliance with enhanced privacy guarantees. The project develops *armoured federated learning* that addresses inference attacks and implements secure multi-party computation to prevent re-identification of data subjects through parameter updates in AI models.
Added value of cross-border datasets	Enables researchers and policymakers to extract AI-driven insights from cross-border, cross-organisation cancer data while protecting patient privacy and anonymity. The platform allows healthcare professionals to analyse their patient data and compare it with data from other European hospitals and research centres without compromising data protection.
Interoperability testing potential	**Technical**: Development of novel privacy metrics and automated measurement tools, secure multi-party computation methods, privacy-enhancing technologies integrated into federated learning platform **Organisational**: Dual-sided platform addressing needs of both data providers and data users **Legal**: Novel metrics for GDPR compliance certification in federated learning implementations, contribution to EU cybersecurity policy making, collaboration with ENISA and European Cybersecurity Competence Centre
Type of data use	Secondary use (research and AI model development for cancer diagnosis and treatment, with clinical validation in European cancer hospitals)
Website	http://trumpetproject.eu/

Case example 5: IASIS - Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients

Cross-border relevance	IASIS brings together experts from five countries working in medicine, genomics, neuroscience and AI to integrate heterogeneous big data from different sources across Europe.
Description	IASIS developed an AI platform for integrating and analysing heterogeneous data from disparate sources including genomics, electronic health records, medical imaging and scientific bibliography. The platform was designed to support clinicians in developing more personalised diagnosis and treatments by combining patient history, memory test results, genomic data and imaging data. The project focused on two major diseases where personalised therapy is essential: lung cancer and Alzheimer's disease, creating a global knowledge base that converts clinical notes into usable data and integrates them with genomic data, related bibliography and imaging.
Added value of cross-border datasets	Integration of diverse data types from multiple European institutions enabled development of AI models for early detection of Alzheimer's disease through analysis of spoken language, identification of treatment effectiveness indicators and discovery of toxicity biomarkers for lung cancer. Cross-border data integration allowed for better prediction of disease prognosis and more personalised treatment recommendations.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Interoperability testing potential	**Technical**: Common representation schema for heterogeneous data sources, semantic integration using knowledge graphs, advanced analytics for pattern discovery **Semantic**: IASIS knowledge graph integrating multiple data modalities and sources **Organisational**: Multi-country collaboration between medical, genomics, neuroscience and AI research institutions
Type of data use	Both primary use (supporting clinical decision-making for individual patients) and secondary use (research for disease prediction, public health planning, and development of precision medicine tools). The project produced actionable information for authorities for planning public health activities and policies.
Website	https://project-iasis.eu/

Case example 6: XShare - Expanding the European EHRxF to share and effectively use health data within the EHDS

Cross-border relevance	XShare is a Horizon Europe project with 40 partners from 13 EU and EFTA member states, creating a sustainable European EHRxF Standards and Policy Hub to enable seamless cross-border health data sharing.
Description	XShare develops the *Yellow Button* application allowing EU citizens to seamlessly share their health data with healthcare providers across borders using their mobile devices. The project implements the European Electronic Health Record Exchange Format (EEHRxF) across three domains: healthcare portability for continuity of care, population health and cross-border health threats, and clinical research. A key innovation is IPS+ (International Patient Summary extended for research), which extends the standard patient summary to support secondary data uses. The Yellow Button empowers citizens to exercise GDPR data portability rights, enabling them to share structured, coded health data in EEHRxF format with a single click.
Added value of cross-border datasets	Enables patients to carry health data across European borders, improving continuity of care for mobile populations. For research and public health, the standardised EEHRxF format enables aggregation and analysis of health data from multiple countries while maintaining semantic interoperability. The IPS+ extension bridges primary use (clinical care) with secondary use (research and public health), creating unified data infrastructure serving both purposes.
Interoperability testing potential	**Technical**: HL7 FHIR-based European Electronic Health Record Exchange Format, Yellow Button mobile application, integration with national EHR systems across 13 countries **Semantic**: Standardised terminologies (SNOMED International, LOINC), mapping between healthcare and research standards (HL7 FHIR, CDISC), IPS+ specification for secondary use **Organisational**: European EHRxF Standards and Policy Hub coordinating six Standards Developing Organisations, collaboration with national eHealth centres, three Open Calls for EHDS standards adoption **Legal**: Implementation of GDPR data portability rights, patient-controlled data sharing with consent management, alignment with EHDS regulation
Type of data use	Primary use (continuity of care across borders, emergency care, treatment coordination) and secondary use (clinical research, population health monitoring, cross-border health threat surveillance, public health policymaking).
Website	https://xshare-project.eu/

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Case example 7: GA4GH Beacon

Cross-border relevance	The Beacon protocol, developed by the Global Alliance for Genomics and Health with major support from ELIXIR (European bioinformatics infrastructure) enables federated discovery and sharing of genomic data across international borders through a standardised API.
Description	Beacon provides a simple yet powerful mechanism for discovering whether specific genomic variants exist in datasets worldwide without compromising privacy or requiring data sharing. Organisations worldwide can "light" Beacons to make their genomic datasets discoverable while maintaining full control over their data. The Beacon Network connects multiple international Beacon instances, allowing researchers to query aspects such as "Who has observed this allele?" across distributed databases. Version 2 of the Beacon protocol extends functionality to include phenotypic and clinical data queries, enabling case-level and cohort-level requests, particularly beneficial for rare disease genetics and cancer research.
Added value of cross-border datasets	Enables researchers to discover relevant datasets across international boundaries without data sharing, facilitating unprecedented collaboration while maintaining data sovereignty. Particularly valuable for rare disease research where aggregating information from multiple countries increases statistical power and validation opportunities. Helps avoid unnecessary complicated access requests by allowing rapid identification of datasets containing sequences of interest.
Interoperability testing potential	**Technical**: RESTful API specification, standardised query and response protocol, integration with VCF files and Phenopackets, handover mechanisms to other data exchange formats **Semantic**: Common data model for genomic and clinical information, support for semantic codes (CURIEs) for phenotypes and disease codes **Organisational**: Federated network model allowing autonomous local databases, tiered access levels (open, registered, controlled) **Legal**: GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data, compliance with varying international data protection regulations
Type of data use	Primarily secondary use (research and clinical variant interpretation) with potential handover to primary use cases. Supports discovery for research, clinical diagnostics and variant interpretation across borders while enabling data custodians to control access based on patient consents and institutional policies.
Website/links	https://genomebeacons.org/ https://pmc.ncbi.nlm.nih.gov/articles/PMC9322265/

Case example 8: RD-Connect Genome-Phenome Analysis Platform

Cross-border relevance	RD-Connect establishes a federated rare disease data infrastructure connecting genomic and phenotypic data repositories across multiple European countries.
Description	RD-Connect creates a Genome-Phenome Analysis Platform (GPAP) that federates rare disease data resources across Europe, enabling researchers and clinicians to access and analyse genomic and phenotypic data from multiple countries. The platform connects European rare disease registries, biobanks and genomic databases, facilitating collaborative research on rare diseases where patient populations are geographically dispersed and data is sparse. By aggregating data from multiple countries, the platform addresses the challenge of small patient cohorts inherent in rare disease research.

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Added value of cross-border datasets	Rare diseases affect small numbers of patients in any single country, making cross-border data aggregation essential for achieving statistical power in research studies. Federated access to genomic and phenotypic data from multiple European rare disease resources enables more comprehensive genotype-phenotype correlations, improved diagnostic yield and better understanding of disease mechanisms and natural history.
Interoperability testing potential	**Technical**: Federated data architecture enabling distributed queries across multiple repositories **Semantic**: Standardised phenotype ontologies (Human Phenotype Ontology), genomic data standards **Organisational**: Connection of European rare disease registries, European Reference Networks (ERNs) integration **Legal**: Compliance with varying national and European data protection frameworks while enabling research access
Type of data use	Both primary use (clinical diagnostics and patient care through improved variant interpretation) and secondary use (rare disease research, natural history studies and development of therapeutic approaches)
Website/links	https://platform.rd-connect.eu/ https://pubmed.ncbi.nlm.nih.gov/35178824/

Case example 9: The MinE project

Cross-border relevance	Project MinE represents an international whole-genome sequencing consortium for Amyotrophic Lateral Sclerosis (ALS) research, involving participants from multiple countries coordinating cross-border genomic data sharing.
Description	Project MinE conducts large-scale whole-genome sequencing of ALS patients and controls across multiple countries to identify genetic factors contributing to ALS. The consortium coordinates international sample collection, sequencing and data analysis, creating one of the largest ALS genomic datasets through cross-border partnership. Researchers from participating countries contribute samples and expertise while accessing the aggregated dataset for analysis, demonstrating successful international coordination in rare neurological disease genomics.
Added value of cross-border datasets	ALS is a relatively rare disease, making international collaboration essential to achieve the sample sizes needed for genomic discovery. Cross-border data aggregation enables identification of rare genetic variants and modifiers that would be undetectable in single-country studies, improving understanding of ALS genetic architecture and potential therapeutic targets across diverse populations.
Interoperability testing potential	**Technical**: Standardised whole-genome sequencing protocols, common bioinformatics pipelines for variant calling and quality control **Semantic**: Standardised phenotype data collection, common variant annotation frameworks **Organisational**: International consortium governance model, data sharing agreements **Legal**: Multi-country ethical approvals and consent frameworks for genetic research
Type of data use	Secondary use (genomic research to identify genetic risk factors, understand disease mechanisms and inform potential therapeutic development)
Website/links	https://projectmine.com/ https://pubmed.ncbi.nlm.nih.gov/29955173/ https://pubmed.ncbi.nlm.nih.gov/31280677/

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Case example 10: EHDEN - European Health Data & Evidence Network

Cross-border relevance	EHDEN creates a large-scale federated network of real-world data sources standardised to a common data model across European countries, enabling cross-border observational research.
Description	EHDEN establishes a federated network of healthcare databases across Europe, standardising diverse national and regional healthcare data sources to the OMOP Common Data Model. The network enables researchers to conduct observational studies across multiple European countries without data leaving source institutions. Through standardised data transformation and federated analysis tools, EHDEN allows execution of common analysis protocols across participating sites, generating real-world evidence for regulatory decision-making, health technology assessment and clinical research.
Added value of cross-border datasets	Federated access to real-world data from diverse European healthcare systems enables larger sample sizes, increased statistical power and better generalisability of findings across different healthcare contexts, patient populations, and treatment patterns. Cross-border harmonisation through common data models facilitates rapid evidence generation for regulatory and HTA purposes.
Interoperability testing potential	**Technical**: OMOP Common Data Model implementation, federated analysis tools, standardised extract-transform-load (ETL) processes **Semantic**: Standardised vocabularies and terminologies (SNOMED CT, LOINC, RxNorm), common outcome definitions **Organisational**: Network of data partners across Europe, coordinated governance structure **Legal**: Federated approach maintaining data sovereignty, compliance with GDPR and national regulations
Type of data use	Secondary use (observational research, health technology assessment, regulatory decision support, evidence generation for clinical guidelines and policy-making)
Website/links	https://www.ehden.eu/ https://pmc.ncbi.nlm.nih.gov/articles/PMC12331365/

Case example 11: European Reference Networks (ERNs) Registries and FAIRification efforts

Cross-border relevance	ERN registries implement FAIRification efforts to enable cross-border rare disease research by making registry data Findable, Accessible, Interoperable and Reusable across European reference centres.
Description	European Reference Networks bring together healthcare providers across Europe for rare and complex diseases. ERN registries collect standardised patient data from reference centres in multiple countries, implementing FAIR principles to enhance data quality, interoperability and research utility. FAIRification efforts focus on standardising data collection, implementing common data elements, using standardised ontologies and terminologies, and developing technical infrastructure for federated queries and analyses. This enables collaborative research and quality improvement initiatives across ERN centres while maintaining data at source institutions.
Added value of cross-border datasets	For rare diseases managed through ERNs, cross-border data integration is essential because individual centres may only see small numbers of patients. Aggregating data from multiple ERN centres across Europe enables natural

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

	history studies, quality of care benchmarking, identification of best practices, and sufficient sample sizes for research on ultra-rare conditions.
Interoperability testing potential	**Technical**: Implementation of common data elements, standardised data entry systems, federated query capabilities **Semantic**: Use of standardised ontologies (Human Phenotype Ontology, Orphanet Rare Disease Ontology), common data dictionaries **Organisational**: Coordination across ERN centres in multiple countries, integration with national rare disease registries **Legal**: Compliance with GDPR while enabling cross-border clinical collaboration and research
Type of data use	Both primary use (supporting clinical care coordination across ERN centres, enabling virtual consultations and collaborative patient management) and secondary use (rare disease research, registry studies, quality improvement and policy development)
Website	https://link.springer.com/article/10.1186/s13023-022-02558-5

Case example 12: BigData@Heart project

Cross-border relevance	BigData@Heart, an IMI (Innovative Medicines Initiative) project, creates a data-driven translational research platform integrating cardiovascular data from multiple European institutions and cohorts.
Description	BigData@Heart develops a platform for cardiovascular precision medicine by integrating diverse data sources including clinical trials, electronic health records, registries, biobanks and imaging repositories from multiple European countries. The project creates a comprehensive cardiovascular data ecosystem that enables researchers to access harmonised data for biomarker discovery, risk prediction model development and personalised treatment approaches. The platform facilitates collaboration between academic institutions, healthcare providers and industry partners across Europe.
Added value of cross-border datasets	Cardiovascular disease manifests differently across populations with varying genetic backgrounds, environmental exposures and healthcare systems. Cross-border data integration enables development of more robust and generalisable predictive models, identification of population-specific risk factors and validation of biomarkers across diverse European populations, ultimately supporting more effective precision medicine approaches.
Interoperability testing potential	**Technical**: Data integration platform supporting multiple data types (omics, imaging, clinical), standardised data pipelines **Semantic**: Harmonised phenotype definitions, standardised cardiovascular outcomes, common terminologies **Organisational**: Public-private partnership model (IMI), collaboration between academia, healthcare providers and industry **Legal**: Data sharing agreements across institutions and countries, GDPR-compliant infrastructure
Type of data use	Secondary use (translational research for biomarker discovery, predictive model development, validation studies and evidence generation to support personalised cardiovascular medicine)
Website	https://www.ihi.europa.eu/projects-results/project-factsheets/bigdataheart

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Case example 13: European Genomic Data Infrastructure (GDI) project

Cross-border relevance	The GDI Starter Kit enables cross-border access to genomic and phenotypic data across Europe, building on the 1+ Million Genomes (1+MG) initiative to create a federated European genomic data infrastructure.
Description	The European Genomic Data Infrastructure develops a technical and organisational framework enabling secure cross-border access to genomic and phenotypic data from participating European countries. The GDI Starter Kit provides participating nations with tools, standards and best practices for establishing national genomic data infrastructures that can interoperate within a federated European ecosystem. The infrastructure supports both discovery of available datasets through Beacon protocols and access to detailed genomic information for authorised researchers, enabling personalised medicine applications including cancer genomics, rare disease diagnostics and pharmacogenomics.
Added value of cross-border datasets	Enables researchers and clinicians to access genomic data from across Europe, increasing sample sizes for rare variant discovery, enhancing statistical power for genome-wide association studies and enabling validation of findings across diverse European populations. Cross-border genomic data sharing is particularly valuable for rare diseases and cancer precision medicine where larger cohorts improve diagnostic yield and treatment selection.
Interoperability testing potential	**Technical**: Implementation of GA4GH standards (Beacon, Phenopackets), federated authentication and authorisation infrastructure (AAI), secure data transfer protocols **Semantic**: Standardised phenotype ontologies, common genomic data formats (VCF, CRAM), harmonised variant annotation **Organisational**: Coordination across 1+MG signatory countries, alignment with national genomic medicine initiatives **Legal**: Framework for cross-border data access under GDPR, ethical approval coordination, compliance with national regulations
Type of data use	Secondary use (genomics, rare diseases, infectious diseases and cancer research)
Website	https://gdi.onemilliongenomes.eu/

Case example 14: HealthData@EU Pilot

Cross-border relevance	The HealthData@EU Pilot tested the first version of the European Health Data Space infrastructure by connecting national data platforms from 12 European countries, with cancer genomics as a key use case demonstrating cross-border secondary use of health data.
Description	The HealthData@EU Pilot project validated the feasibility of cross-border health data exchange for secondary use by implementing a pilot infrastructure connecting national health data access points. The cancer genomics use case demonstrated how researchers can discover and access cancer genomic data across European borders through standardised metadata discovery, common data access request processes and federated analysis capabilities. The pilot tested technical architecture, developed the Health DCAT-AP metadata standard, assessed data quality and transfer protocols, and provided practical guidance for implementing the European Health Data Space.
Added value of cross-border datasets	Cancer genomics research benefits significantly from cross-border data access, as larger sample sizes enable identification of rare genomic alterations, better stratification of patient subgroups and validation of predictive biomarkers across

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

	diverse populations. The federated approach allows researchers to access relevant cancer genomic data from multiple countries while respecting data sovereignty and governance requirements of participating nations.
Interoperability testing potential	**Technical**: HealthData@EU infrastructure architecture, federated metadata discovery system, common data access request processes, privacy-enhancing technologies for secure data transfer **Semantic**: Standardised metadata framework for health datasets, Health DCAT-AP metadata standard, common data quality metrics, harmonised cancer genomics terminologies **Organisational**: Connection of national health data access bodies, coordinated governance across 12 countries, alignment with EHDS regulation requirements **Legal**: Implementation of EHDS principles for secondary use, GDPR compliance framework, cross-border data transfer safeguards
Type of data use	Secondary use (cancer genomics research, precision oncology research, development of predictive models for cancer treatment response, and policy-making for cancer care improvement)
Website/links	https://www.health-data-hub.fr/page/healthdataeu-pilot https://pmc.ncbi.nlm.nih.gov/articles/PMC12420906/

Case example 15: Cross-Border Hospital of Cerdanya (France-Spain)

Cross-border relevance	The Hospital of Cerdanya is Europe's first fully cross-border hospital, established through a European Grouping of Territorial Cooperation (EGTC) and jointly managed by French and Catalan health authorities, serving populations on both sides of the Franco-Spanish border.
Description	The Cross-border Hospital of Cerdanya serves approximately 30,000 permanent residents in a mountainous region, providing emergency services, acute care, obstetrics, and specialised mountain sports medicine to both French and Spanish citizens without administrative barriers. The hospital operates with a binational governance structure, multilingual staff, and integrated care protocols combining elements from both healthcare systems. It covers 53 municipalities across the border, with facilities including 68 beds, operating rooms and advanced imaging equipment. The hospital demonstrates practical implementation of cross-border health data exchange for clinical care, with patient records integrated into both French and Spanish national health information systems.
Added value of cross-border datasets	Eliminates geographical and administrative barriers to healthcare access in an isolated region, ensuring proximity-based care regardless of national borders. The integration creates operational efficiencies and enables sharing of specialised equipment and expertise that neither region could sustain independently.
Interoperability testing potential	**Technical**: Integration of French and Spanish health information systems, bilateral health data exchange, common medical protocols **Semantic**: Multilingual medical terminology (French, Spanish, Catalan), harmonised clinical protocols, unified patient identification **Organisational**: EGTC governance with joint decision-making, binational staffing model, coordinated referral pathways **Legal**: Resolution of cross-border legal issues (births registration, repatriation, employment, insurance reimbursement), compliance with both French and Spanish healthcare regulations and data protection requirements

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301

Type of data use	Primary use (direct patient care for emergency services, acute care, obstetrics, and specialised medical services). Patient health data is accessible and interoperable across the French-Spanish border to ensure continuity of care, with real-time cross-border health data exchange infrastructure aligned with eHealth Digital Service Infrastructure (eHDSI) and EHDS requirements.
Website	https://ec.europa.eu/regional_policy/en/projects/france/a-cross-border-hospital-in-cerdanya-improves-healthcare-for-french-and-spanish-citizens

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101161301