



Building the European Virtual Human Twin

Call: Accelerating best use of technologies (DIGITAL-2021-DEPLOY-01)

Work program year: DIGITAL-2021-2022

Topic: ID DIGITAL-2021-DEPLOY-01-TWINS-HEALTH

Grant Agreement No: 101083771

Deliverable D6.2

Towards a Code of Conduct for the re-use and integration of VHTs

Due date of delivery: 30 November 2024

Actual submission date: 3 December 2024

Start of the project: 01 October 2022

End date: 31 December 2024



Reference

Name	EDITH_D3.1_Vision_roadmap_ouline_VPH_Mar2023
Lead beneficiary	Virtual Physiological Human institute (VPHi)
Author(s)	Lorenzo Cristofaro, Edwin Morley-Fletcher, Riccardo Traina Chiarini (Lynkeus), contributions from VITO (Frederic Jung, Simon Denil), HITS (Martin Golebiewski, Gerhard Mayer), UNIBO (Francesca Conte)
Dissemination level	Public
Type	Report
Official delivery date	30/11/2024
Date of validation by the WP Leader	30/11/2024
Date of validation by the Coordinator	03/12/2024
Signature of the Coordinator	03/12/2024

Version log

Date	Version	Involved	Comments
07/11/2024	v1.0	Lynkeus	First draft
22/11/2024	v2.0	HITS	Section 12
27/11/2024	v3.0	UNIBO	External review and Section 4.1.1
30/11/2024	v4.0	LYNKEUS	Completion and final review
03/12/2024	Final	VPHi	Final review & submission

Executive Summary

This deliverable D6.2 is designed as an **in-depth analysis of the essential elements to be considered when proceeding towards establishing a Code of conduct**, with the aim of serving as a compliance-enabler and an accountability tool under the General Data Protection Regulation, for the re-use and integration of Virtual Human Twins in healthcare.

D6.2 focuses on two primary objectives: (i) assisting all stakeholders composing the VHT ecosystem in **identifying the key legal and regulatory challenges** that need to be overcome for ensuring compliance with various applicable regulations, and (ii) **offering specific recommendations** with the aim of helping European policymakers to better identify and remove the obstacles, detected within the complex and multi-layered legal framework, that currently hinder a wider adoption of VHTs and Europe's global leadership in this field.

To this end, the following document first **analyzes the current EU general regulatory and policy scenario** in the AI and data-driven landscape, exploring the outcomes of some significant public initiatives carried out with the same scope of outlining the applicable barriers to scientific and technological innovation. Secondly, it sets out a comprehensive **state-of-the-art on data anonymization, data pseudonymization and Privacy-Enhancing Technologies**, both from a technical and a legal perspective, highlighting the key role that synthetic data may play in the future. Subsequently, all – already in force or forthcoming – **EU regulations with major impacts on the VHT ecosystem are comprehensively investigated**, including: the General Data Protection Regulation (providing a practical focus on some of EDITH's specific use cases); the European Health Data Space; the Artificial Intelligence Act; the Data Governance Act and the Data Act; the Clinical Trial Regulation; the Medical Device Regulation and the In Vitro Diagnostic medical devices Regulation.

Following this extensive analysis, **Intellectual Property Rights** management profiles are put under the magnifier lens, first focusing on copyright and database protection, trade secrets, and patentability, and finally also assessing innovative open-source approaches and licensing frameworks.

Section 11 evaluates the main **ethical implications and challenges** stemming from both the generation and the use of VHT, to ensure social license and wide acceptability under strict ethical frameworks based on internationally recognized principles and values.

Additionally, section 12 is dedicated to a summary of the main outcomes of the significant amount of work on **technical standards** which has been carried out during the EDITH CSA.

Finally, some brief **conclusions** are drawn both with regard to:

- (i) the **need to move towards a Code of Conduct** dedicated to the Virtual Human Twin, with the double aim of: setting out a framework of rules tailored to the specific needs and challenges of this groundbreaking technology; enhancing clarity and harmonization as to how regulations governing data, AI and data-driven digital models are to be implemented across Member States;
- (ii) the **steps and initiatives which need to be taken by EU policymakers** to address and mitigate the major barriers outlined in this document and to adequately benefit from the profound transformative impact on Europe's healthcare systems potentially implied by a growing uptake of VHTs, which could act in all Member States as growth-enhancing expenditure within the EU new economic governance framework.

Table of Contents

EXECUTIVE SUMMARY	3
TABLE OF CONTENTS	4
LIST OF FIGURES	5
ACRONYMS	5
1. INTRODUCTION: WHY MOVE TOWARDS A CODE OF CONDUCT	9
1.1 OBJECTIVES	11
1.2 TERMINOLOGY.....	11
2. SOME RECENT WORKS POINTING-OUT BARRIERS ALSO RELEVANT TO THE VHT	13
2.1 ‘THE FUTURE OF EUROPEAN COMPETITIVENESS’ REPORT	13
2.2 THE OUTCOMES OF TEHDAS	16
2.3 ‘ASSESSMENT OF THE EU MEMBER STATES’ RULES ON HEALTH DATA IN THE LIGHT OF GDPR’	19
3. THE KEY AND TRANSVERSAL ISSUE OF DATA ANONYMIZATION	23
3.1 THE CONCEPT OF PSEUDONYMIZATION (VS. ANONYMIZATION)	23
3.1.1 <i>Traditional and advanced pseudonymization techniques</i>	24
3.2 PRIVACY-ENHANCING TECHNOLOGIES	29
3.2.1 <i>Federated Learning</i>	31
3.2.1 <i>Differential Privacy</i>	32
3.3 SYNTHETIC DATA	34
3.3.1 <i>Technical scenario</i>	34
3.3.2 <i>Legal scenario</i>	37
3.4 ANONYMOUS DATA: STILL A BIG QUESTION MARK	39
3.4.1 <i>Judgement in case SRB v. EDPS (case T-557/20)</i>	42
4 GENERAL DATA PROTECTION REGULATION	44
4.1 THE FRAMEWORK FOR VHT.....	44
4.1.1 <i>GDPR and EDITH’s use cases</i>	50
4.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS.....	55
5 EUROPEAN HEALTH DATA SPACE	58
5.1 THE FRAMEWORK FOR VHT.....	58
5.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS.....	69
6 ARTIFICIAL INTELLIGENCE ACT	72
6.1 THE FRAMEWORK FOR VHT.....	72
6.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS.....	79
7 DATA ACT AND DATA GOVERNANCE ACT	85
7.1. THE FRAMEWORK FOR VHT	85
7.2. BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS	91
8 CLINICAL TRIALS REGULATION	92
8.1 THE FRAMEWORK FOR VHT.....	92
8.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS.....	97
9 MEDICAL DEVICE REGULATION / IN VITRO DIAGNOSTIC REGULATION	99
9.1 THE FRAMEWORK FOR VHT.....	99
9.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS.....	106
10 INTELLECTUAL PROPERTY	109
10.1 THE FRAMEWORK FOR VHT.....	109
10.2 CHALLENGES FOR IPR MANAGEMENT	115
10.2.1 <i>EHDS</i>	115
10.2.2 <i>Open source</i>	117
10.2.3 <i>Patentability of AI models and algorithms</i>	118

11	ETHICAL IMPLICATIONS	121
12	TECHNICAL STANDARDS	130
12.1	ASPECTS OF INTEREST AND RELEVANT IMPACTS/CURRENT USE:	130
12.2	CURRENT OPEN AND CHALLENGING POINTS	131
12.2.1	<i>Model credibility assessment standard</i>	131
12.2.2	<i>Best practices and clinical practice guidelines</i>	132
12.3	POLICY RECOMMENDATIONS	132
13	CONCLUSIONS	133

List of figures

FIGURE 1 - COUNTRIES THAT PARTICIPATED IN THE TEHDAS COUNTRY VISITS	17
FIGURE 2 - BARRIERS TO CROSS-BORDER SHARING OF HEALTH DATA FOR SECONDARY USE (TEHDAS)	18
FIGURE 3 - TEHDAS' PROPOSAL FOR ARCHITECTURE OF THE EHDS FOR THE SECONDARY USE OF HEALTH DATA	19
FIGURE 4 – LAWFULNESS GROUNDS FOR REUSING HEALTH DATA FOR MARKET APPROVAL OF MEDICINES AND DEVICES	21
FIGURE 5 - SPECIFIC NATIONAL LEGISLATION ADDRESSING THE REUSE OF HEALTH DATA FOR MONITORING OF MEDICAL DEVICE SAFETY AND/OR PHARMACOVIGILANCE.	21
FIGURE 6 - APPLICATION OF BASIC PSEUDONYMIZATION TECHNIQUES	25
FIGURE 7 – APPLICATION OF PSEUDONYMIZATION POLICIES	26
FIGURE 8 – ZERO-KNOWLEDGE PROOF FOR PSEUDONYMIZATION	27
FIGURE 9 - SIMPLIFIED APPLICATION OF SMPC	28
FIGURE 10 - FEDERATED MACHINE LEARNING ILLUSTRATION	31
FIGURE 11 - CENTRALISED AND DECENTRALIZED FL (FROM UK ICO'S GUIDANCE ON PETS)	32
FIGURE 12 – GLOBAL AND LOCAL DP IN UK ICO'S GUIDANCE ON PETS	34
FIGURE 13 - SOURCE: EHDS FACT SHEET	58
FIGURE 14 - DATA ACCESS PROCESS FOR SECONDARY USE	66
FIGURE 15 - HIGH-LEVEL COMPONENT VIEW OF A DT FOR PRECISION MEDICINE	92
FIGURE 16 – OVERVIEW OF THE ETHICAL ISSUES SURROUNDING DIGITAL TWINS IN HEALTH	121
FIGURE 17 – RISKS EMERGING FROM ETHICAL IMPLICATIONS OF DIGITAL TWINS IN HEALTH	122

Acronyms

Acronym	Full name
ABM	Agent-based model
ACP	Algorithm Change Protocol
AEs	Auto-Encoders
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act (Regulation (EU) 2024/1689)
AI HLEG	Artificial Intelligence High-Level Expert Group
API	Application Programming Interface
ASME	Americal Society of Mechanical Engineers
ASTM	Americal Society for Testing and Materials
Atrialmtk	Atrial Modelling Tool Kit
B2B	Business-to-Business
BBCT	Bologna Biomechanical Computed Tomography
BSC	Barcelona Supercomputer Centre

CDSS	Clinical Decision Support System
CE	Conformité Européenne
CJEU	Court of Justice of the European Union
COMBINE	Computational Modeling in Biology Network
CoP	Community of Practice
CoU	Context of Use
CSA	Coordination and Support Action
CT	Computed Tomography
CTIS	Clinical Trials Information System
CTR	Clinical Trial Regulation (Regulation (EU) 536/2014)
Data Act	Regulation (EU) 2023/2854
DGA	Data Governance Act (Regulation (EU) 2022/868)
DNA	Deoxyribonucleic Acid
DPIA	Data Protection Impact Assessment
DSM	Digital Single Market
DT	Digital Twin
DTH	Digital Twin in Healthcare
EC	European Commission
ECDC	European Centre for Disease Prevention and Control
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EEHRxF	European Electronic Health Record Exchange Format
EGC	European General Court
EGE	European Group on Ethics in Science and New Technologies
EHDS	European Health Data Space
EMA	European Medicine Agency
ENISA	European Union Agency for Cybersecurity
EPC	European Patent Convention
EPO	European Patent Office
EU	European Union
EUDAMED	European Database for Medical Devices
EUREC	European Network of Research Ethics Committees
FAIR	Findable, Accessible, Interoperable, Reproducible
FDA	Food and Drug Administration
FHIR	Fast Healthcare Interoperability Resources
FL	Federated Learning
FRAND	Fair, Reasonable, and Non-Discriminatory
FUTURE-AI	Fairness, Universality, Traceability, Usability, Robustness & Explainability in AI
GANS	Generative Adversarial Networks
GDPR	General Data Protection Regulation (Regulation (EU) 2016/679)
GMLP	Good Machine Learning Practice
GPL	Global Public License
GPU	Graphic Processing Unit
HDAB	Health Data Access Body
HE	Homomorphic encryption
HER	Electronic Health Record
HL7	Health Level 7
HMAC	Hash-based Message Authentication Code
HPC	High Performance Computing
HRAI	High-Risk AI
HTA	Health Technology Assessment
ICO	Information Commissioner's Office (UK)

ICT	Information and Communications Technology
ICU	Intensive Care Unit
ID	Identity
IEC	International Electrotechnical Commission
IoT	Internet of Things
IPR	Intellectual Property Rights
ISO	International Standards Organisation
IST	In Silico Trials
IT	Information Technology
IVDR	In Vitro Diagnostic Medical Devices Regulation (Regulation (EU) 2017/746)
MDCG	Medical Device Coordination Group
MDDT	Medical Device Development Tool
MDR	Medical Devices Regulation (Regulation (EU) 2017/745)
MDSW	Medical Device Software
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MS	Member State
MultiCellML	Multi-Cellular Modelling language
NIST	US National Institute of Standards and Technology
OECD	Organization for Economic Cooperation and Development
OEM	Original Equipment Manufacturer
OIDC	Open ID Connect
OpenVT	Open Virtual Tissues
OSS	Open Source Software
PCCP	Predetermined Change Control Plan
PEA	Privacy Enforcement Authorities
PerMedCoE	Personalised Medicine Centre of Excellence
PET	Privacy enhancing technology
PICAR	Population, Intervention, Comparison, Attributes of eligible Clinical Practice Guidelines, Recommendation
PICOR	Population, Intervention, Comparator, Outcome and Recommendation
PK	Public pseudonymization Key
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QMUL	Queen Mary University of London
QoI	Quantity of Interest
RDF	Resource Description Framework
RNG	Random Number Generator
SaMD	Software as a Medical Device
SDO	Standards Developing Organisation
SEP	Standard Essential Patents
SK	Secret (private) pseudonymization Key
SME	Small and Medium-sized Enterprises
SMPC	Secure Multi-Party Computation
SPS	SaMD Pre-Specifications
SRB	Single Resolution Board
TC	Technical Committee
TDM	Text and Data Mining
TPLC	Total Product Lifecycle
TS	Technical Specification
TSD	Directive on the Protection of Trade Secrets
TTP	Trusted Third Party

UDI	Unique Device Identification
UDI-DI	Unique Device Identification – Device Identifier
UDI-PI	Unique Device Identification – Production Identifier
UNIBO	University of Bologna
UP	Unitary Patent
UPC	Unified Patent Court
VAEs	Variational Auto-Encoders
VHT	Virtual Human Twin
VITO	Flemish Institute for Technological Research
VV-40	Verification and Validation 40
VVUQ	Verification, Validation and Uncertainty Quantification
WIPO	World Intellectual Property Organization
WP	Work Package
WP29	Article 29 Working Party
ZKP	Zero-Knowledge Proof

1. Introduction: why move towards a code of conduct

Virtual Human Twins sit exactly at the intersection of all major data, technology and health regulations.

In all likelihood, there is no other subject to date which is more crosscutting at a regulatory level and suitable, at the same time, to serve as a litmus test of the foresight and overall coherence of the European legislation on health care and medical research.

Digital Twins in health need large volumes of data, either personal (General Data Protection Regulation) or non-personal (Data Act), imply in almost all cases the secondary processing of health information (European Health data Space and Data Governance Act), require data exchange infrastructures and rules both at B2B and B2C level (Data Act), entail the use of artificial intelligence models and systems (Artificial Intelligence Act), may qualify as software-as-a-medical-device (Medical Device Regulation), can be used in the context or for the purpose of clinical trials (Clinical Trial Regulation), may undergo stringent cybersecurity requirements, when used by certain operators in the health sector (NIS2 – Network and Information Security Directive), pose new challenges in terms of intellectual property rights management and enforcement (IPR Directive and Open Data Directive). And many more.

In the context of such a complex and evolving legal landscape, in as far as they can be achievable **codes of conduct** would stand as a **crucial and practical accountability tool**.

In all cases when law provisions need to be applied, and therefore tested, in a specific operational context, with the need to take into account sectoral limits and peculiarities, codes of conduct would enable operators to weave a mesh capable of connecting the ‘general’ with the ‘particular’, tailoring the obligations provided for by applicable laws on the basis of the needs of a specific industry or business.

After all, one of the most significant challenges for legislators, both at the EU and national level, is the need to identify rules that are sufficiently general to adapt to as many concrete cases as possible, while at the same time ensuring that the necessary level of abstractness does not hinder the concrete application of the rules which are subsequently laid down. Inevitably, **this challenge often leads to a misalignment, which can sometimes be slight but in other cases wider, between the set regulatory framework and everyday reality**, all the more in fleeting environments such as scientific, medical and technological research.

Indeed, more than any other market phenomenon, **new technologies require flexible and adaptable regulatory instruments, that ought to be easily modelled based on legal and operational needs that by definition vary over time**, so as to be able to take into account the novelties, the risks and the specificities that emerge, on a case-by-case basis, from the daily experience of all stakeholders and business operators.

The rationale behind this fundamental compliance-enhancement tool is well outlined in several pieces of legislation, which are worth mentioning:

- Recital 98 of the GDPR, which encourages associations and organizations representing market players to set up codes of conduct, in order to *“facilitate the effective application of this Regulation, taking account of the specific characteristics of the processing carried out in certain sectors and the specific needs of micro, small and medium enterprises. In particular, such codes of conduct could calibrate the obligations of controllers and processors, taking into account the risk likely to result from the processing for the rights and freedoms of natural persons”*¹;
- Recital 165 of the AI Act, which states that *“Providers of AI systems that are not high-risk should be encouraged to create codes of conduct, including related governance mechanisms,*

¹ Recital 77 of the GDPR specifies that *“Guidance on the implementation of appropriate measures and on the demonstration of compliance by the controller or the processor, especially as regards the identification of the risk related to the processing, their assessment in terms of origin, nature, likelihood and severity, and the identification of best practices to mitigate the risk”*, could be provided in particular by means of codes of conduct.

intended to foster the voluntary application of some or all of the mandatory requirements applicable to high-risk AI systems, adapted in light of the intended purpose of the systems and the lower risk involved and taking into account the available technical solutions and industry best practices (...)”;

- Recital 32 of the Data Governance Act, which in turn specifies that *“In order to increase trust in (...) data intermediation services, in particular related to the use of data and compliance with the conditions imposed by data subjects and data holders, it is necessary to create a Union-level regulatory framework which establishes highly harmonised requirements related to the trustworthy provision of such data intermediation services, and which is implemented by the competent authorities. That framework will contribute to ensuring that data subjects and data holders, as well as data users, have better control over access to and use of their data, in accordance with Union law. The Commission could also encourage and facilitate the development of codes of conduct at Union level, involving relevant stakeholders, in particular on interoperability”*;
- Recital 79 of the Data Act, which echoes Regulation (EU) 2018/1807,² by encouraging *“providers of data processing services to develop and effectively implement self-regulatory codes of conduct covering best practices for, inter alia, facilitating the switching of providers of data processing services and the porting of data”*.

The leitmotiv underlying all the above considerations, as well as most of the regulations with major impacts on VHT, are data. Which means that benefits stemming from industry-specific or sectoral codes of conduct – as provided for by the above data-centred laws – would become even broader and deeper.

Furthermore, codes of conduct would offer the most effective solution to the extremely harmful regulatory fragmentation that often characterizes those areas where the European Legislator has allowed Member States to adopt additional rules at national level, such as in connection with primary and secondary processing of health data, as well as with scientific and medical research. In these cases, **a well pondered set of self-imposed rules can act as a glue between the various locally scattered regulatory fragments.**

The rules of conduct would apply on a voluntary basis in order to strengthen both the degree of accountability of the adhering parties and the trust afforded by society, in general, and more specifically by patients, users and consumers. However, the **development of such rules of conduct** would certainly require a **very high level of regulatory maturity, in terms of specialized knowledge and subject-awareness by competent Supervisory Authorities and the many public bodies which are already or will be operating in accordance with the various applicable or forthcoming legislations** (e.g., the European Data Protection Board in the GDPR, the AI Office and the European Artificial Intelligence Board in the AI Act, the European Data Innovation Board in the Data Governance Act (DGA) and the Data Act, the European Health Data Space Board in the EHDS, the European Medicine Agency, the Medical Device Coordination Group, and others) **which is currently far from being achieved with regards to the VHT.**

For these reasons, this document is conceived as a work aimed at providing an overall analysis of the elements to be taken into account for moving towards a code of conduct for the re-use and integration of DTs, focusing on two main objectives:

- help stakeholders and especially VHT developers and users, to identify the main legal challenges that they face, in terms of compliance with several applicable regulations, in relation to DTs;
- provide specific recommendations for supporting European policymakers in identifying what, to date, are the main factors that, in the complex and multi-layered applicable legal landscape,

² Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.

slow down or sometimes hinder the uptake of the VHT and Europe’s global leadership in this sector.

1.1 OBJECTIVES

Taking inspiration from the overly ambitious goal of attaining a code of conduct approved under Art. 40 of the GDPR, this deliverable D6.2 basically aims to the more limited objective of identifying best practices and envisaging solutions to help stakeholders overcome, in a homogeneous and consistent manner across all Member States, the main legal or regulatory hurdles that currently arise in connection with the VHT, with a view to unleashing its full potential by removing barriers that still make the use of a technology capable to revolutionize entire industries so infrequent and uncommon.

In more detail, this document:

- as a first step, goes through some important works that have recently already identified many of the major limitations and obstacles to the full development of the VHT;
- secondly, provides a general picture of the current scenario, both legal and technical, regarding anonymous and pseudonymous data and Privacy-Enhancing Technologies, including synthetic data;
- subsequently identifies, among the many, the EU laws with major impacts and implications for the VHT, providing a brief description of their scope and goals;
- then describes, for each of these regulations, eventual gaps, or loopholes, or misalignments or inconsistencies with other legislations, when applied or associated with the VHT;
- finally offers, for each regulatory obstacle or barrier identified, a proposal for a solution, or enhancement, or step forward, addressed to the competent authorities.

Europe has never before been so prolific in terms of digital, technology and data strategy regulations. But the more articulated the set of rules to navigate, the more challenging it becomes to find the thread that guarantees overall consistency of the applicable laws and that all relevant ‘pieces’ appropriately fit in the regulatory puzzle which must be made up, on a case-by-case basis, for each specific horizontal technological application.

1.2 TERMINOLOGY

‘Virtual Human Twin’ and ‘Digital Twin in Healthcare’

As in the Roadmap, the term ‘Virtual Human Twin’ is used in this document to indicate the concept of combining resources towards the development of a fully integrated Digital Twin for personalised health and care. In addition, it is also used as the brand name of the overarching EU initiative on VHT, as well as the infrastructure to be developed. In the context of this initiative, Digital Twins represent smaller entities focusing on specific applications, as they are virtual representations of a physical object, process or system across its life-cycle. It uses data and other sources to enable learning, reasoning, and recalibrating (either dynamically or through human-in-the-loop decision making) for monitoring, diagnostics and prognostics.

For the sake of clarity, please note that the following terminology is used in this document:

- **Roadmap** refers to the final output of EDITH CSA project. A first version of this roadmap has been made available (D3.2)³;
- **Virtual Human Twin** is defined as a systematic, ever-growing digital and quantitative representation of the actionable knowledge available on human pathophysiology. The European VHT infrastructure will enable the pooling of resources and assets to develop digital twins in healthcare and assess their credibility. It entails the development of a federated public infrastructure and the collection of appropriate resources (data, models, algorithms, computing power, storage etc.), driven by the engagement of a collaborative ecosystem. The term VHT is only used to indicate the broader initiative or its infrastructure;
- **Digital Twin** are examples of the Virtual Human Twin concept, focusing on specific applications, as they are virtual representations of a physical object, process or system across its life-cycle. It uses data and other sources to enable learning, reasoning, and recalibrating (either dynamically or through human-in-the-loop decision making) for monitoring, diagnostics and prognostics.
- **Infrastructure** refers to the trinity of software Catalogue-Repository-Platform;
- **European VHT infrastructure** is a federated public infrastructure and the collection of appropriate resources (data, models, algorithms, computing power, storage etc.). It encompasses a catalogue, repository and simulation platform. The implementation of the Advanced VHT simulation platform is not part of the EDITH CSA, but the subject of a separate program, starting with the already ongoing EC procurement process for a Platform for Advanced VHT Models.

³ <https://zenodo.org/records/8200955>

2. Some recent works pointing-out barriers also relevant to the VHT

The European VHT Initiative is a flagship action of the EU Commission “to foster and accelerate the development of integrated, validated digital representations of the human body. These Digital Twins hold substantial potential for medical research and healthcare delivery, contributing to a deeper understanding of human physiology, pathology, and disease aetiology, as well as facilitating personalised, patient-centric medicine”⁴.

This Initiative includes:

- the envisaged European VHT platform, enabling the pooling of resources and assets to further develop the science and technology required for building advanced VHT-based solutions in health and care;
- the VHT Manifesto, a ‘Statement of intent on development, evidence, and adoption in healthcare systems’, with the aim of promoting collaboration in the current VHT ecosystem, by bringing together the various stakeholders to demonstrate their support for the European VHT Initiative, while also facilitating the context for further collaboration⁵.

Bearing in mind the goals of this initiative, the analysis of the results of some works that were carried out in the recent period and which already highlighted which legal and regulatory barriers are most likely to hinder the use of data-driven or AI based technologies, helps to understand both strengths and weaknesses of the current legislation from the perspective of the VHT.

2.1 ‘THE FUTURE OF EUROPEAN COMPETITIVENESS’ REPORT

On 9 September 2024, the long-awaited Mario Draghi’s report ‘The future of European competitiveness’ was published⁶.

The document sets out some remarks that deserve to be reported:

1. *“Companies in Europe face three main hindrances from the rising weight of regulation. First, they need to comply with the accumulation of or frequent changes to EU legislation over time, translating into overlap and inconsistencies. For example, a Business Europe gap analysis of 13 pieces of EU law flagged duplication across 169 requirements, including differences (29%) and outright inconsistencies (11%). Second, EU companies face an extra burden due to national transposition, for instance as Member States “gold plate” of EU legislation or implement laws with divergent requirements and standards from one country to another. (...) GDPR in particular has been implemented with a large degree of fragmentation which undermines the EU’s digital goals. Third, EU regulation imposes a proportionally higher burden on SMEs and small mid-caps than on larger companies, yet the EU lacks a framework to assess these costs (...)”.*
2. *““Gold-plating’ by Member States of the GDPR and a lack of consistency in its enforcement adds to EU companies’ administrative burden. The GDPR, which entered into force in 2016 and is directly applicable in all Member States, aims to offer a harmonised EU approach to privacy enforcement. However, it gives Member States the possibility to define privacy rules in 15 areas, leading to fragmentation and legal uncertainty stemming from the widespread use of specification clauses, ‘gold-plating’ and inconsistent enforcement by national Data Protection Authorities (DPAs), and the fact that some Member States have several DPAs doing so (e.g.,*

⁴ <https://digital-strategy.ec.europa.eu/en/policies/virtual-human-twins>

⁵ <https://www.virtualhumantwins.eu/>

⁶ https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

16 in Germany). This could hinder cross-border entrepreneurship and innovation, including the development and deployment of new technologies and cybersecurity solutions (...).”

3. “(...) **access to health data is one of the preconditions for the development of AI in the pharma industry but is constrained by fragmentation.** In particular, although GDPR contains options to use patient data for health research, take up has been uneven across Member States, preventing the industry from tapping into a wealth of available electronic data”.
4. “(...) while the ambitions of the EU’s GDPR and AI Act are commendable, their complexity and risk of **overlaps and inconsistencies can undermine developments in the field of AI by EU industry actors.** The differences among Member States in the implementation and enforcement of the GDPR (...), as well as overlaps and areas of potential inconsistency with the provisions of the AI Act create the risk of European companies being excluded from early AI innovations because of uncertainty of regulatory frameworks as well as higher burdens for EU researchers and innovators to develop homegrown AI. As in global AI competition ‘winner takes most’ dynamics are already prevailing, the EU faces now an unavoidable trade-off between stronger ex ante regulatory safeguards for fundamental rights and product safety, and more regulatory light-handed rules to promote EU investment and innovation, e.g., through sandboxing, without lowering consumer standards. This calls for **developing simplified rules and enforcing harmonised implementation of the GDPR in the Member States, while removing regulatory overlaps with the AI Act** (...). This would ensure that EU companies are not penalised in the development and adoption of frontier AI (...) While it is early to fully gauge the impact of these landmarks regulations, their implementation must avoid producing administrative and compliance burdens and legal uncertainties as the GDPR’s and must be enforced within shorter timeframes and more stringent processes for compliance provisions”.
5. “Harmonise national ‘AI Sandbox regimes’ across all Member States to enable experimentation and the **development of innovative AI applications in the selected industrial sectors and ensure harmonised and simplified implementation of the GDPR.** Regular assessments should be carried out of potential regulatory hindrances deriving from EU or national legislation, with feedback from research centres to regulators and the EU. On this basis, it is recommended to introduce regular and fast review process of the main AI-related regulations (e.g., every three years), as technological developments can make regulations rapidly obsolete in this sector. In this context, develop simplified rules, particularly for SMEs, and enforce harmonised implementation of the GDPR in the Member States, while removing regulatory overlaps with the AI Act”.
6. “In the absence of EU hyperscale companies, developing AI verticals requires strong coordination between multiple actors, including AI developers, Research and Technology Organisations (RTOs), and industrial players. For instance, discovering whether an innovative product can be developed by a factory using its **AI-powered digital twin** requires the replication of the factory, its robots, processes and the overlay of an AI algorithm. In the absence of clear coordination at an early stage, the product would not be developed, leading to a market failure. EU-wide collaboration and coordination among Member States on AI verticals would enable EU players to reach the required scale in terms of data, investment and market share, potentially enabling them to compete with US hyper-scalers”.
7. “The complex emergence of a **European Health Data Space (EHDS).** There is significant untapped potential to leverage health data in the EU, as demonstrated by the considerable possibilities to access and link datasets in healthcare relative to the US. Currently, the GDPR allows the processing of health data for the provision of health or social care, public health and scientific purposes based on EU or national law. Data can be processed without explicit consent provided that suitable and specific measures are put in place to safeguard the rights and freedoms of data subjects. Some Member States already benefit from these possibilities under their own national law. However, **the uptake of these options by Member States has been uneven and has resulted in the ineffective secondary use of health data.** To overcome this challenge, the Commission has proposed a regulation to enable a European Health Data

Space (EHDS) by building on possibilities offered by the GDPR for a specific EU law with particular safeguards. In spring 2024, the European Parliament and the Council reached a political agreement on the proposed regulation. The proposal aims to develop a European framework inspired by the actions taken by several Member States that have adopted similar national legislation for the secondary use of health data”.

8. "Ensure the **optimal implementation of the EHDS Regulation** by supporting the accessing and sharing of electronic health records and capacity building for national health data access bodies. The regulation is expected to start to apply two years after its entry into force with a staggered application thereafter and a first partial evaluation after eight years. To optimise its implementation, it is key to **make short-term resources available for the introduction of EU requirements and standards in electronic health records at the national level**. This is important notably to enable the cross-border provision of healthcare and patient rights to access their health data in a structured interoperable format (...) National health data access bodies have a pivotal role as they are tasked to decide on data access applications. Their proper functioning will be crucial for the overall implementation of the EHDS Regulation. The clarification and cross-country coordination of opt-out mechanisms will need to be ensured".
9. "**Leverage existing health data for regulatory, policy and clinical decision-making by stepping up the standardization of pre-existing ‘legacy’ health data**. In the run-up to the full application of the EHDS Regulation, it will be necessary to continue and increase efforts to standardise existing data sources to a common data model building on the work initiated by the European Health Data Evidence Network (EHDEN), set to end by October 2024. The initiative can be set up as a new public-private partnership, aiming to work in full alignment (forward compatibility) with the EHDS. Through this work, standardised health data will be leveraged to generate evidence for regulatory, policy and clinical decision-making"⁷.
10. "The use of ‘real-world evidence’ may help streamline the process of patient recruitment and data collection for pricing and reimbursement. An example of how real-world data can be applied at the EU level is the Data Analysis and Real-World Interrogation Network (DARWIN EU). DARWIN EU was established in 2022 by the EMA and the European Medicines Regulatory Network as a coordination centre to provide timely and reliable evidence from real-world healthcare databases across the EU on the use, safety and effectiveness of medicines (...) **Leverage the DARWIN EU network to generate evidence for innovation in medicine development and for policy and clinical decision-making supported by the use of AI**. Existing expertise and experience need to be geared towards generating ‘real-world’ evidence by running non-interventional studies drawing on the existing data source catalogue to expand activities building on additional data sources in Member States made available by the EHDS. AI has huge potential to accelerate the management and analysis of health data for this purpose”.
11. "Guidance is gradually disseminated until 2027 by the EMA and national medicine agencies, under their AI work programme. Importantly, it will need to maximise the possibilities offered by the forthcoming EHDS Regulation and the recent AI Act. This should cover the analysis of ‘raw’ clinical data transmitted to the EMA by the industry as planned under current proposals, as well as data collected for pharmacovigilance purposes. Opening up the secondary use of health data for research purposes has particular potential to anchor R&D activities within the EU. Guidance can also build on the experience gained through the DARWIN EU® network”

Again, the common thread – and underlying issue – linking all the above considerations, made with reference to the European competitiveness level, is represented by data, or rather, by their scarcity resulting – despite the huge volumes of information generated every single minute in any industry – from the existing limitations on data sharing and reuse, especially between different Member States.

As it will be shown in detail in this D6.2, crucial and urgent steps forward are needed in this respect.

⁷ Further information on EHDEN may be found on <https://www.ehden.eu/>.

2.2 THE OUTCOMES OF TEHDAS

TEHDAS, the joint action ‘Towards the European Health Data Space’, helped EU Member States and the EU Commission to develop and promote concepts for the secondary use of health data to benefit public health and health research and innovation in Europe⁸.

The project was focused on:

- engaging other European projects and policymakers in a dialogue about the EHDS;
- ensuring sustainability of the secondary use of health data in the EU;
- developing a governance model for cross-border cooperation in the secondary use of health data between European countries;
- promoting the reliability and compatibility of and access to health data for secondary use;
- clarifying the role of individuals in the secondary use of health data and including them in dialogue about the use of health data for research and policymaking.

As a result of extensive cross-border works and analyses, TEHDAS produced several recommendations to help the EU Commission and the Member States to overcome the barriers to the secondary use of health data and it also provided many valuable feedback that were taken into account during the EHDS legislative process, fostering a pan-European dialogue on the matter.

Some of the main barriers identified by TEHDAS (1) are set out below:

1. in order to learn from others and engage all stakeholders in a dialogue about the EHDS, TEHDAS organized Project Forums, Policy Forums and fact-finding country visits. All the 12 countries involved (see *Figure 1* below) agreed on the benefits of the EHDS, including the facilitation of research studies that require access to health-related datasets and the use of common European templates and guidelines in the process of reusing such data. At the same time, however, they all also acknowledged that there are still many practical and legal challenges to overcome, such as the fact that **there are diverse health data management and governance systems in different countries, as well as divergent national interpretations of what data anonymization is.**

⁸ The TEHDAS project, started on 1 February 2021 and ended on 31 July 2023, involved partners from 21 Member States and four other European countries. The project was coordinated by the Finnish Innovation Fund Sitra. TEHDAS was funded by the Health Programme of the European Union and by the European countries involved. Sitra has been appointed to lead a new joint action (‘TEHDAS 2’), under EU4Health Programme, that will start in 2024 with the participation of 29 European countries.

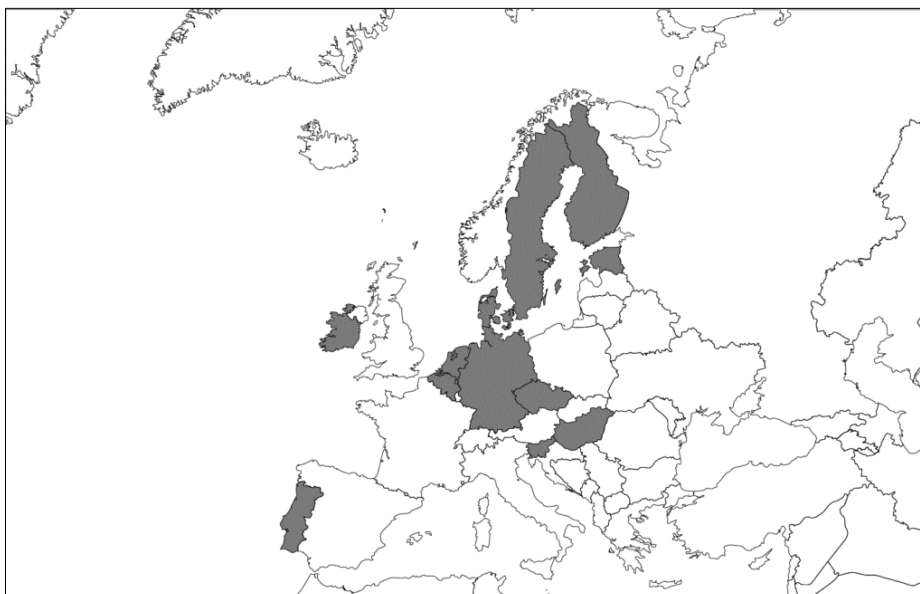


Figure 1 - Countries that participated in the TEHDAS country visits

2. **A lack of standardization of datasets and variations in legal interpretations of EU data protection law** are examples of the most common hurdles that make secondary processing of health data and transnational studies very complex to achieve and increase the costs of research and compliance. The main barriers that slow down, complicate, condition and sometimes prevent data re-use have been described by TEHDAS in the table below. Clearly, almost all of them are of a legal nature, or in any case related to limitations arising from regulations in force⁹:

Barrier description	Theme
There are differences in governance and health data systems in Europe.	Infrastructure Legal
There is no common European interpretation of what constitutes 'sufficient anonymisation' to transform personal data to non-personal data.	Legal
There is no common European interpretation of what constitutes 'pseudonymisation'.	Legal

⁹ Crucial considerations and recommendations are made by TEHDAS in a number of reports, with special regard to data sharing and data barriers on <https://tehdas.eu/tehdas1/results/tehdas-identifies-barriers-to-data-sharing/> and <https://tehdas.eu/tehdas1/results/tehdas-suggests-options-to-overcome-data-barriers/>.

There is no common European interpretation of what is, and what is not, 'secondary use' of data.	Legal
European countries have national legislation/rules concerning health and research data in addition to the GDPR.	Legal
European countries have the ability to set their own derogations under the GDPR. This lack of harmonisation may create additional barriers.	Legal
European countries have different preferences as to the choice of legal basis for processing under the GDPR. This impedes the cross-border collaboration and data sharing.	Legal
Health data is considered sensitive data, meaning for example special category data under GDPR, and is treated differently from other types of data when it comes to health data ethics, management and use.	Data
No standardised data sharing agreements exist for products developed by private sector providers using public sector health data to (a) facilitate safe data sharing and (b) protect taxpayers' investment.	Trust and Transparency
Across Europe, different taxonomy and ontology codes are used to label the same health condition, making comparisons between data sets a challenge.	Data
Poor data management procedures reduce the ability to reuse data.	Data

Figure 2 - Barriers to cross-border sharing of health data for secondary use (TEHDAS)

As said above, **reliance on anonymization is currently risky due to diverging interpretations by competent Supervisory Authorities as to the degree of irreversibility that individual de-identification must achieve so that data can be deemed anonymous and not pseudonymous.** Among others, also BBMRI-ERIC specifically reported this criticality¹⁰.

3. **The definition of EU-wide data quality standards is lagging, to the detriment of health research and innovation, where accurate big data are necessary**, e.g., to develop a new drug, to find the causes of a disease, to understand which patients might benefit the most from a treatment, or to assess which prevention strategy is better to reduce mortality in the population. Likewise, policymakers and regulators need quality data, for example to safely introduce new treatments, to ensure that the most vulnerable patients receive appropriate care, to assess gaps in the timely response to patients with critical acute care needs, or to tailor treatments to those patients that would benefit the most. Legal provisions must be laid down to indicate the quality of a dataset, or the mechanism for informing data users about the value to them of a specific data collection¹¹.
4. Ensuring secure and private access to health data involves the interaction of various computational systems, including hospital and primary care information services, systems used by health authorities or ethics committees to authorize data access, and secure environments for processing the sensitive data. Therefore, **the current lack of homogeneous data governance and data security standards must be remedied through appropriate legal rules and technical guidelines**, precisely allocating obligations and relevant responsibilities. This is needed in order to avoid that the necessary interactions between services and systems under the EHDS unintentionally expose patients' personal information and to ensure that all processing activities take place exclusively within a secure environment. To this end, TEHDAS also put forward a specific architecture – shown in Figure 2 – based on a network of nodes, operated by the Member States and supported by core services provided by European Commission, acting as service providers for

¹⁰ BBMRI stressed that “The existing regulatory framework seems insufficient to deliver on the promises of the EHDS. Health data governance remains fragmented at national and regional level, hindering any effort to scale up research and healthcare solutions. Most importantly, it is necessary to protect and promote the use of health data, defining clear pan-European rules to overcome the existing gaps in practice” (<https://www.bbmri-eric.eu/wp-content/uploads/statement-on-european-health-data-space.pdf>).

¹¹ More specific and detailed indications in this respect can be found in several TEHDAS reports (such as <https://tehdas.eu/tehdas1/results/tehdas-proposals-for-data-quality-and-utility-in-ehds/> and <https://tehdas.eu/tehdas1/results/tehdas-proposals-for-data-quality-and-utility-in-ehds/>).

data users. These nodes connect the data holders and the data permit authorities (*i.e.*, the national ‘Health data Access Bodies’ under the EHDS) that will decide on the granting of access to the data, as well as the secure processing environments where the data will later be analysed by the data user.

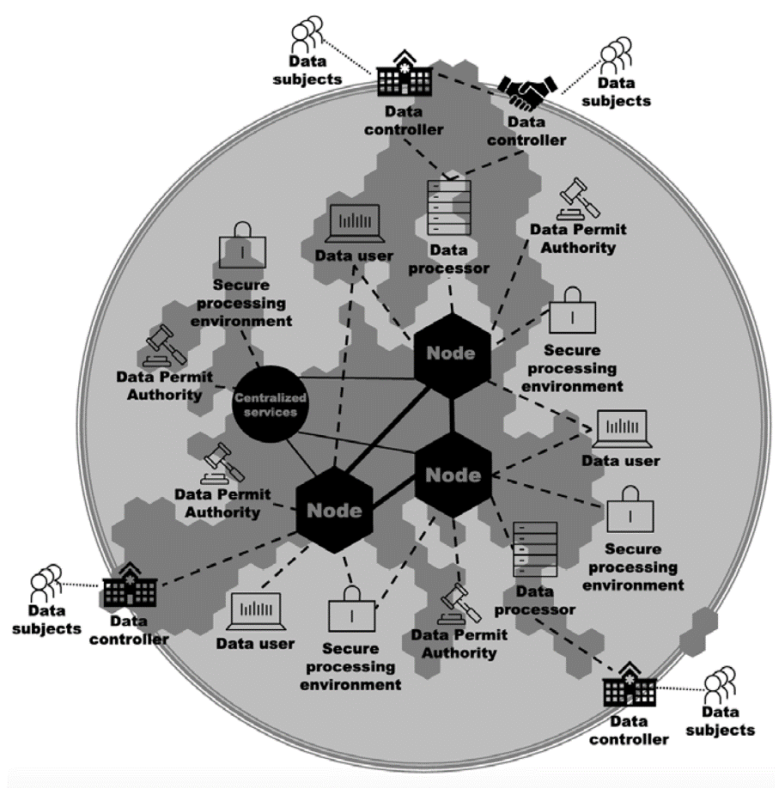


Figure 3 - TEHDAS' proposal for architecture of the EHDS for the secondary use of health data

5. People's perspective on the secondary use of health data is embedded in their perception of the intrinsic value and sensitivity of this special category of data. The trustworthiness of decision-making processes on the secondary use of health data stands as an important factor, especially when they include a plurality of views and actors: the relationship depends strongly on what the health data will be used for and who will benefit from it. **Citizens also perceive health data as power: although they tend to support the secondary use of their health data, this was dependent on the conditions and purposes of processing.** To balance this power, on the one hand, the opportunity must be given to the individuals to make meaningful and active choices as the use and reuse of their data, and on the other hand, data should be processed in a way that makes it as difficult as possible to re-identify the data subjects, guaranteeing the utmost transparency about who accesses their data and for what specific admitted purpose. This urges policymakers and stakeholders from the ecosystem of the secondary use of health data to **integrate the voice of citizens and patients in the definition and process of the secondary use of health data**, to ensure its benefits are fully and safely exploited while ensuring the minimization of relevant risks. Taking these steps helps build public trust, which in turn will foster positive attitudes and social license towards the secondary use of health data.

2.3 'ASSESSMENT OF THE EU MEMBER STATES' RULES ON HEALTH DATA IN THE LIGHT OF GDPR'

During the first half of 2020, the EU Commission conducted a study based on a series of workshops with the participation of representatives from Member States' Ministries of Health, Data Protection

Authorities, stakeholders and independent experts across the EU, with the objective to examine and present the national rules that govern the processing of health data in light of the GDPR. As a result, a valuable and detailed ‘*Assessment of the EU Member States’ rules on health data in the light of GDPR*’ was published¹².

The study was aimed at highlighting differences and identifying elements that might affect the cross-border exchange of health data in the EU, as well as at examining the potential for EU level action to support health data use and re-use.

The EU Commission was indeed already well-aware that the achievement of several strategic goals at a European level is conditional on the availability, and consequently the sharing, of as many data as possible, *e.g.*, to support clinical decision making, to allow for healthcare system planning, supervision and improvement, to empower patients to engage actively in their healthcare and wellness management and, last but surely not least, to foster research and innovation in the field.

In this sense, health data are necessary:

- ✓ for public purposes, including planning, management, administration and improvement of health and care systems; prevention or control of communicable diseases; protection against serious threats to health and ensuring high standards of quality and safety of healthcare and of medical products and medical devices. This implies the secondary processing of health data that were collected initially in the context of providing care, but which can later be reused by public entities such as national health systems, public health bodies (*e.g.*, universities or public health laboratories) and by regulators, such as medicines agencies;
- ✓ for scientific research purposes, as pursued by both public and private sector organizations (third parties, not being the original data controllers), including the pharmaceutical and medical technology industries, research centres and organizations, AI developer, insurance providers and many more.

In almost all of these cases, the conditions applicable to the secondary processing (and very often also to the primary use) of personal data, all the more when qualifying as health data, vary a lot all across the EU, with different levels of limitations and requirements established by Member States.

By way of example, focusing on the reuse of health data, and so the possibility to develop specific DTs for market approval of medicines and devices by competent agencies, the EMA, HTAs and notified bodies, the distinction must be made between *ex ante* approval on the one hand and post marketing surveillance or pharmacovigilance on the other.

In both cases, however, the legal grounds that may permit the reuse of the health data, which were originally collected in connection with the provision of health and care to patients, differ at a national level, as clearly shown here below.

¹² https://health.ec.europa.eu/publications/assessment-eu-member-states-rules-health-data-light-gdpr_en

Legal basis for market approval	Total MS	
6(1)(c) Legal obligation + 9(2)(i) public interest in the area of public health	7	BG, CZ, DK, IE, HR, IT, FI
6(1)(c) legal obligation + 9(2)(h) health or social care	3	BG, DK, HR
6(1)(f) legitimate interest + 9(2)(h) health or social care	0	
6(1)(e) public interest + 9(2)(h) health or social care	3	BG, DK, HR
6(1)(e) public interest + 9(2)(i) public interest in the field of public health	4	BG, DK, FR, HR
Other combination	3	EE, FR, MT
No specific legislation	17	BE, DE, EL, ES, CY, LV, LT, LU, HU, NL, AT, PL, PT, RO, SI, SK, SE, [UK]

Figure 4 – Lawfulness grounds for reusing health data for market approval of medicines and devices¹². MS: Member State.

Similarly, with reference to the second area described above, although pharmacovigilance is based on the Directive 2010/84/EU, as it does not state that personal data may be processed (or rather, re-used) for this purpose, in all cases where a Member State national legislation (allowing authorities or even pharmaceutical companies to process personal data to attain the relevant pharmacovigilance objectives) does not exist, health data could not be leveraged, unless anonymized or in presence of the individuals' express consent for this activity. Figure 5 clearly shows this fragmented scenario.

Legal basis medical device safety and pharmacovigilance	Total MS	
6(1)(c) Legal obligation + 9(2)(i) public interest in the area of public health	15	CZ, DK, DE, IE, EL, ES, FR, HR, IT, LV, LU, MT, AT, SI, FI
6(1)(c) legal obligation + 9(2)(h) healthcare	7	DK, DE, HR, LV, MT, AT, SI
6(1)(e) public interest + 9(2)(h) healthcare	5	DK, EE, EL, HR, LV
6(1)(e) public interest + 9(2)(i) public interest in the field of public health	6	CZ, DK, EE, EL, HR, LV
6(1)(f) legitimate interest + 9(2)(h) healthcare	0	
Other combination	7	DE, EL, ES, FR, AT, PL, SE
No specific legislation	9	BE, BG, CY, LT, HU, NL, PT, RO, SK, [UK]

Figure 5 - Specific national legislation addressing the reuse of health data for monitoring of medical device safety and/or pharmacovigilance¹². MS: Member State.

But more importantly with a view to the progress and reinforcement of the EU-wide VHT ecosystem, it must be noted that data-driven scientific research is hampered by local barriers and often still siloed at the national level, mainly due to the following reasons identified by TEHDAS:

- Member States have not exploited the margin of manoeuvre granted to them by the GDPR (Art. 9, as it will be better detailed in the dedicated section of this document) to adopt further conditions for the processing of health data “*in a homogenous way, resulting in a complex and fragmented landscape for researchers to navigate. Consequently, differences between Member States in the way the GDPR is implemented and interpreted in the area of scientific research*”

*has made data exchange between Member State and EU bodies for research purposes difficult and in some cases highly technical*¹³;

- *“Divergence arises between Member States as to what are considered “tools” likely to be used to identify individuals (...) in practice some Member State authorities work on the basis that full anonymity can never be achieved for health-related data while still keeping the data useful for research, others believe anonymity within the meaning of GDPR can be achieved (...) Similar differences in interpretation also exist with respect to pseudonymization (...) As a result, a number of misconceptions have arisen as to its meaning*¹⁴.

As evident, the findings of the three studies briefly discussed above, though centred on different aspects, largely align in identifying the main regulatory barriers to innovation and research, particularly when it comes to the processing of personal data and the development and use of artificial intelligence models or systems. These criticalities will be better explored in specific sections of this document.

¹³ Ibid. (Par. 5.1.2). The study also highlights that *“It is evident there is a variance of safeguards and lawful basis leading to confusion and technical difficulty when conducting inter-jurisdictional research”* (p. 73).

¹⁴ Ibid. (Par. 5.2.2). Regarding anonymization, helpful indications are provided in the joint paper prepared by the European Data Protection Supervisor and the Spanish Data Protection Authority (Agencia Española de Protección de Datos) titled *‘10 misunderstandings related to anonymization’* (https://www.edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en).

3. The key and transversal issue of data anonymization

While closely related to the GDPR (which will be discussed later), the legal question of when and under what conditions data can be considered anonymous permeates all the other regulations under review in this document and often plays a crucial role in determining whether they, or at least a part of their provisions, must apply or not.

Given the complexity and prominence of this issue, we believe it is important to address it separately and as a preliminary matter, in order to facilitate a better understanding of some of the main arguments that will be later explored concerning other areas relevant to the VHT.

3.1 THE CONCEPT OF PSEUDONYMIZATION (VS. ANONYMIZATION)

Even though pseudonymization represents a crucial tool to ensure privacy-by-design and reinforce accountability, the complexity of its practical interpretation, with particular regard to its distinction from anonymization, made the application of pseudonymization quite tricky – especially in some contexts, such as healthcare – and therefore much less frequent than expected.

While no exact definition of ‘anonymization’ or ‘anonymous data’ is provided for by the GDPR, ‘pseudonymization’ is described as “*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*” (Art. 4(5) GDPR).

In brief, pseudonymization aims at protecting personal data by hiding the identities of individuals in a dataset, *e.g.*, by replacing one or more personal identifiers with the so-called pseudonyms, and appropriately protecting the link between such pseudonyms and the initial identifiers.¹⁵

At a very basic level, pseudonymization starts with a single input (the original data) and ends with two outputs (the pseudonymized dataset and the additional information). Together, these can reconstruct the original data. However, with respect to the individuals concerned (the data subject), each output has meaning only in combination with the other.

Therefore, pseudonymization refers to techniques that replace, remove or transform information that identifies individuals, and keep that information separate and secure.

It has a prominent role in the GDPR both as a security measure (Art. 32) and as a tool to achieve privacy-by-design (Art. 25) and data minimisation (Art. 5.1(c))¹⁶: this technique can go beyond hiding real

¹⁵ The report on ‘Pseudonymization techniques and best practices’ by the European Union Agency for Cybersecurity (ENISA), dated November 2019 (<https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices>), provides, *inter alia*, the following definitions: (a) “Identifier is a value that identifies an element within an identification scheme⁷. A unique identifier is associated to only one element. It is often assumed in this report that unique identifiers are used, which are associated to personal data”; (b) “Pseudonym, also known as cryptonym or just nym, is a piece of information associated to an identifier of an individual or any other kind of personal data (e.g., location data). Pseudonyms may have different degrees of linkability (to the original identifiers). The degree of linkability of different pseudonym types is important to consider for evaluating the strength of pseudonyms but also for the design of pseudonymous systems, where a certain degree of linkability may be desired (e.g., when analysing pseudonymous log files or for reputation systems)”; (c) “Pseudonymization entity is the entity responsible of processing identifiers into pseudonyms using the pseudonymization function. It can be a data controller, a data processor (performing pseudonymization on behalf of a controller), a trusted third party or a data subject, depending on the pseudonymization scenario. It should be stressed that, following this definition, the role of the pseudonymization entity is strictly relevant to the practical implementation of pseudonymization under a specific scenario”.

More technically, i) “pseudonymization function, denoted P , is a function that substitutes an identifier Id by a pseudonym $pseudo$ ”; ii) “Pseudonymization secret, denoted s is an (optional) parameter of a pseudonymization function P . The function P cannot be evaluated/computed if s is unknown”; iii) “Recovery function, denoted R , is a function that substitutes a pseudonym $pseudo$ by the identifier Id using the pseudonymization secret s . It inverts the pseudonymization function P ”; iv) “Pseudonymization mapping table is a representation of the action of the pseudonymization function. It associates each identifier to its corresponding pseudonym. Depending on the pseudonymization function P , the pseudonymization mapping table may be the pseudonymization secret or part of it”.

¹⁶ Not by chance, Art. 89 of GDPR (‘Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes’) specifically mentions pseudonymization as a suitable technique to implement the minimisation of the processing operations (“Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data

identities into supporting the data protection goal of unlinkability, *i.e.*, reducing the risk of privacy-relevant data being linked across different data processing domains, and contributes to attain the key principle of data minimization under the GDPR, for example in cases where the controller does not need to access personal data relating to the data subjects, but only to their pseudonyms.

For this reason, “*pseudonymization can motivate the relaxation, to a certain degree, of data controllers’ legal obligations if properly applied*”.¹⁷

Without any prejudice to the aforementioned benefits, a key concept to keep in mind is that “*personal data which have undergone pseudonymization, which could be attributed to a natural person by the use of additional information, should be considered to be information on an identifiable natural person*” (Recital 26 GDPR). In other words, pseudonymous data are still personal data, falling into the scope of application of the GDPR and any other applicable data protection legislation.

Practically, when implementing pseudonymization, it is important to clarify the application scenario and the different actors involved and their roles, with particular respect to that of the pseudonymization entity,¹⁸ which can be attributed to different entities (*e.g.*, a data controller, a data processor, a Trusted Third Party or the data subject) depending on the case. Under each specific scenario, it is then required to consider the best possible pseudonymization technique and policy that can be applied, given the benefits and pitfalls entailed.

Obviously, there is not a one-size-fits-all approach and a case-by-case data protection risk analysis remains crucial, to consider all relevant aspects and variables (*e.g.*, privacy protection, utility, scalability, etc.).

3.1.1 TRADITIONAL AND ADVANCED PSEUDONYMIZATION TECHNIQUES

3.1.1.1 Basic techniques

According to recent studies by ENISA¹⁹, the most commonly used basic pseudonymization techniques are the following:

- a. **Counter:** the simplest pseudonymization method, where identifiers are replaced by sequential numbers generated by a monotonic counter (*e.g.*, 110, 111, 112, 113, etc.). Its simplicity makes it well-suited for small, uncomplicated datasets. The pseudonyms have no direct connection to the original identifiers, though the sequential nature can reveal information about the order of the data. However, this technique may face challenges with implementation and scalability in larger, more complex datasets;
- b. **Random Number Generator (RNG):** similar to the counter method, but with random numbers assigned to each identifier (*e.g.*, 110, 319, 818, 196, etc.). This provides stronger protection because randomness makes it harder to infer the original identifier, unless the mapping table is compromised. However, potential issues include "collisions" (where two identifiers receive the same pseudonym) and scalability concerns, especially in larger datasets;
- c. **Cryptographic hash function:** this method applies a one-way hash function to an identifier to generate a pseudonym. It is designed to be computationally infeasible to reverse or find two distinct inputs that result in the same output. While hash functions support data integrity, they

subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymization provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner”.

¹⁷ ENISA’s report on ‘Pseudonymization techniques and best practices’ referred to in Note no. 15.

¹⁸ See the definition in Note no. 15.

¹⁹ Reference is made, particularly, to the following reports by the ENISA (i) ‘Pseudonymization techniques and best practices’, dated November 2019 (Note no. 15); (ii) ‘Pseudonymization Advanced Techniques and Use Cases’, dated January 2021 (<https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases>); (iii) ‘Deploying pseudonymization techniques – The case of the health sector’, dated March 2022 (<https://www.enisa.europa.eu/publications/deploying-pseudonymisation-techniques>).

are generally weak for pseudonymization due to vulnerability to brute-force and dictionary attacks;

- d. **Hash-based Message Authentication Code (HMAC):** similar to a cryptographic hash function but includes a secret key to generate the pseudonym. Without this key, mapping identifiers to pseudonyms is not possible. HMAC is considered a strong pseudonymization technique for data protection. However, recovery of the original data may be problematic if the original identifiers are not stored. Different variations of this method may offer varying levels of utility and scalability;
- e. **Symmetric encryption:** a reversible cryptographic method that transforms personal data into pseudonyms using a secret key, which can later restore the data to its original form. Block ciphers (such as Auto-Encoders (AEs)) are used to encrypt identifiers with a secret key, serving both as the pseudonymization key and the recovery key. Symmetric encryption is considered a robust pseudonymization technique, with properties similar to HMAC. However, in terms of data minimization, the fact that pseudonyms can always be reversed may present a challenge, particularly if storing the original identifiers is unnecessary.






TECHNIQUE	EXAMPLE
 Counter	Progressive counter starting from 13, 14, 15
 Random number	Random values between 0000 and 9999 9701, 3069, 1454
 Hash function	MD5 has for "John" 527bd5b5de689e2c32ae974c6229ff785
 HMAC	MD5 has for "John" and key "1337" fb76bcf46a35e9c21168cd54e5d31ff
 Encryption	AES encryption for "John" and key "1337" WMaDIYzImXQFO92cs5hNQ==

Figure 6 - Application of basic pseudonymization techniques

Regardless of the choice as to which specific pseudonymization technique to apply, the policy (or mode) of implementation is also critical. Three main different pseudonymization policies can be listed (considering an identifier *Id* which appears several times in two datasets *A* and *B*):

- ✓ **Deterministic pseudonymization:** in all the databases and each time it appears, *Id* is always replaced by the same pseudonym *pseudo*;
- ✓ **Document randomised pseudonymization:** each time *Id* appears in a database, it is substituted with a different pseudonym (*pseudo1*, *pseudo2*, etc.); however, *Id* is always mapped to the same collection of (*pseudo1*, *pseudo2*) in the datasets *A* and *B*;
- ✓ **Fully randomised pseudonymization:** for any occurrences of *Id* within a database *A* or *B*, *Id* is replaced by a different pseudonym (*pseudo1*, *pseudo2*).

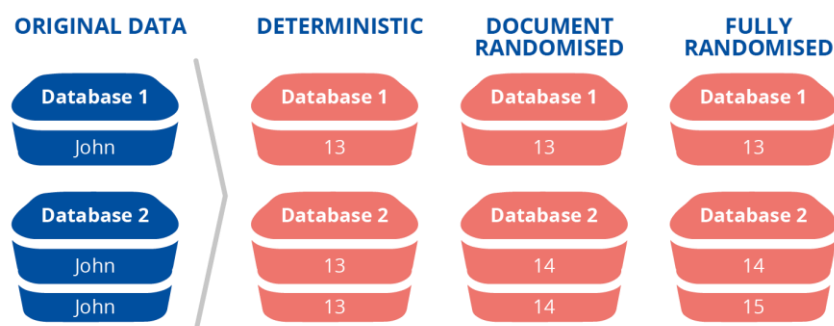


Figure 7 – Application of pseudonymization policies

3.1.1.2 Advanced techniques

The basic techniques described above, alongside with relevant policies and scenarios, can improve the level of protection of personal data, provided that the pseudonymization secrets used to create the pseudonyms are not exposed. However, in order to address some specific data protection challenges – even more in the healthcare domain – such typical solutions may not always suffice.

In this case, it is possible to address more complex situations, ensuring that the level of security is enhanced and that the risks of a personal data breach – including singling-out a specific individual – are properly minimized, thanks to more complex pseudonymization techniques, such as asymmetric and homomorphic encryption; ring signatures and group pseudonyms; chaining mode, pseudonyms based on multiple identifiers or attributes; pseudonyms with proof of ownership; secret sharing schemes.

Some advanced pseudonymization solutions, based on cryptographic techniques, also qualify as **Privacy-Enhancing Technologies (‘PETs’)** which will be further discussed below, aiming to enforce “*privacy principles in order to protect and enhance the privacy of users of information technology (IT) and/or of individuals about whom personal data are processed*”.²⁰

Among the many cutting-edge pseudonymization applications, it is worth focusing on the following.

a. Asymmetric encryption

This option enables the possibility to have two different entities involved during the pseudonymization process: (i) a first entity can create the pseudonyms from the identifiers using the Public pseudonymization Key (PK), and (ii) another entity is able to resolve the pseudonyms to the identifiers using the Secret (private) pseudonymization Key (SK). The entity who applies the pseudonymization function and the entity who can resolve the pseudonyms into the original identifiers do not have to share the same knowledge.

For example, a data controller (*e.g.*, a clinical centre) can make its public key available to its data processors (*e.g.*, a software house). The data processors can collect and pseudonymize the personal data using the PK, but the data controller is the only entity which can later compute the initial data from the pseudonyms, thanks to its SK. Such a scenario is strongly related to the generic scenario of a data processor being the Pseudonymization Entity, with the additional advantage, in terms of protecting individuals’ identities, that the processors do not have the pseudonymization secret (and that they do not store the mapping between original identifiers and the derived pseudonyms).

Similarly, a Trusted Third Party (TTP) may share its public key with one or more data controllers, remaining nonetheless the only one to be able to resolve any pseudonym created by any data controller using its private SK (*e.g.*, at the request of a data subject). Such scenario may also be

²⁰ Fischer-Hübner, S. (2009). *Privacy-Enhancing Technologies*. In: LIU, L., ÖZSU, M.T. (eds) *Encyclopedia of Database Systems*, pp. 2142–2147. Springer, Boston, MA.

relevant to cases of joint controllership, where a controller is performing the pseudonymization and another controller only receives the pseudonymized data for any admitted further processing²¹.

b. Homomorphic Encryption

Certain asymmetric encryption schemes support homomorphic operations, *i.e.*, a specific type of encryption allowing a third party to perform specific computations on the ciphertexts without having knowledge of the relevant decryption key. In other words, homomorphic encryption (HE) allows performing computations on encrypted data without first decrypting it: the computations themselves are also cyphered and, once decrypted, the outputs are identical to what would have been produced if the computations would have been performed on the original plaintext data.

HE uses a public key-generation algorithm to generate a pair of private and public keys, and an evaluation key which is needed to perform computations on the encrypted information, when it is shared with the entity that will perform them. This entity (*e.g.*, a research organization, in quality as data user under the EHDS) does not need access to the private key to perform the analysis: the client (*e.g.*, an hospital as data holder according to the EHDS), who retains the private key, can then decrypt the output to obtain the results it requires. Any entity that has only the public and the evaluation keys cannot learn anything about the encrypted data in isolation.

c. Zero-Knowledge Proof

A well-known cryptographic primitive is the so-called Zero-Knowledge Proof (ZKP), *i.e.*, any protocol by which a party (prover) is able to prove to another party (verifier) to be in the possession of a secret, without revealing any information about the secret itself. More generally, ZKP can be leveraged to prove that a statement is true, without revealing any details of the statement.

In the context of pseudonymization, if an individual associated with a pseudonym needs to prove that he/she is the owner of that pseudonym, without revealing his or her exact identity, a ZKP may provide the solution²².

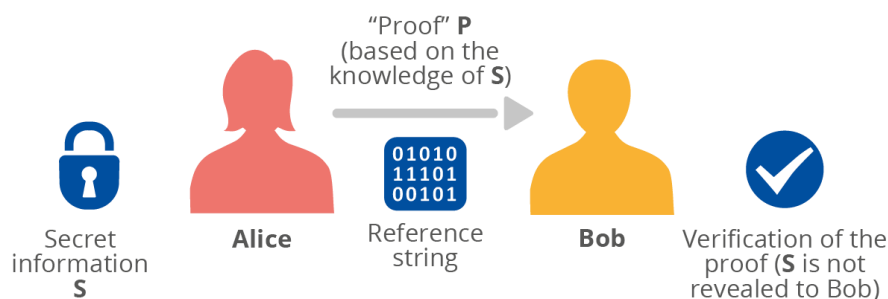


Figure 8 – Zero-knowledge proof for pseudonymization

d. Secure Multi-Party Computation

A Secure Multiparty Computation (SMPC) protocol allows the participating parties to jointly compute a function over their input data while keeping those input data private. Essentially, it removes the need for a trusted third party to view and manage the data. All parties (or a subset of

²¹ As pointed out by the ENISA in the aforementioned ‘Pseudonymization Advanced Techniques and Use Cases’ report, “Several pseudonymization schemes based on asymmetric encryption have already been proposed. A typical application is to make available healthcare data to research groups; more precisely, by using fully randomised pseudonymization schemes based on asymmetric cryptography (...), we may ensure that the identifiers (*e.g.*, social security number or medical registration number or any other identifier) of a given patient are not linkable. For instance, a participant may have different local pseudonyms at doctors X, Y, Z, and at medical research groups U, V, W – thus providing domain-specific pseudonyms to ensure unlinkability between these different domains; by these means, doctors will store both the real name/identity of their patients and their local pseudonyms, but researchers will only have (their own) local pseudonyms”.

²² An example of this scenario is provided by ‘anonymous transactions’ in cryptocurrencies: ZKP is used to allow verification of the transactions without the verifiers (miners) knowing anything about the transactions’ contents (and the senders and the receivers of the transactions are concealed).

the parties) may learn the result, depending on the nature of the processing and how the protocol is configured. SMPC uses a cryptographic technique called ‘secret sharing’, referring to the division of a secret and its distribution among each of the parties. This means that each participating party’s information is split into fragments to be shared with other parties. Secret sharing is not the only way to perform SMPC, but it the most common approach used in practice.

Each party’s information cannot be revealed to the others unless some proportion of fragments of it from each of the parties are combined (see *Figure 9* below, based on a highly simplified example made by the UK Information Commissioner’s Office – ‘UK ICO’ – in its guidance on ‘Privacy-enhancing technologies’, dated June 2023)²³. As this would involve compromising the information security of a number of different parties, in practice it is unlikely to occur. This limits the risks of exposure through accidental error or malicious compromise and helps to mitigate the risk of insider attacks and other types of data breaches.

Example

Three organizations (Party A, Party B and Party C) want to use SMPC to calculate their average expenditure. Each party provides information about their own expenditure – this is the “input” that will be used for the calculation. SMPC splits each party’s information into three randomly generated ‘secret shares’. For example, Party A’s input – its own total expenditure – is €10,000. This is split into secret shares of €5,000, €2,000 and €3,000. Party A keeps one of these shares, distributes the second to Party B and the third to Party C. Parties B and C do the same with their input data.

Party	Input data	Secret share 1 (to be kept)	Secret share 2 (to be distributed)	Secret share 3 (to be distributed)
A	€10,000	€5,000	€2,000	€3,000
B	€15,000	€2,000	€8,000	€5,000
C	€20,000	€7,000	€4,000	€9,000

When this process is complete, each party has three secret shares. For example, Party A has the secret share it retained from its own input, along with a secret share from Party B and another from Party C. The secret shares cannot reveal what each party’s input was (*i.e.*, Party A does not learn the total expenditure of Parties B or C), and so on.

Each party then adds together their secret shares. This calculates a partial result both for each party and the total expenditure of all three. The SMPC protocol then divides the total by the number of parties – three, in this case – giving the average expenditure of each: €15,000, as based on the original amounts, but no single party is able to learn what the other’s actual expenditure is.

Party	Input data	Secret share kept	Secret share Received	Secret share Received	Partial Sum
A	€10,000	€5,000	€4,000	€5,000	€14,000
B	€15,000	€2,000	€2,000	€9,000	€13,000
C	€20,000	€7,000	€8,000	€3,000	€18,000

Figure 9 - Simplified application of SMPC

²³ <https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies-1-0.pdf> .

Among several applications, SMPC can be used to enable privacy in both the inference and training phases of Machine Learning systems. Oblivious model inference allows a party to submit a request to a server holding a pre-trained model, keeping the request private from the server S and the model private from the client C. In this setting, the inputs to the SMPC are the private model from S, and the private test input from C, and the output (decoded only for C) is the model's prediction.

Practically, this means that SMPC can be also leveraged to secure the data and protect patients' privacy while allowing specific parties to train an AI model integrated in or connected with a Digital Twin, based on some given combined datasets, without exposing such data and so ensuring privacy-by-design.

3.2 PRIVACY-ENHANCING TECHNOLOGIES

Even though the concept of PETs is far from new and their relevant use is rapidly spreading, it has never had a universally accepted definition. As pointed out by the Organization for Economic Co-operation and Development (OECD) in its recent and detailed report on 'Emerging Privacy Enhancing Technologies - Current Regulatory and Policy Approaches', "[t]he absence of a stable definition in this field can hinder a concerted analysis by policy makers, and privacy enforcement authorities (PEAs) in particular, of the potential impacts of PETs on data protection and privacy assessments"²⁴.

ENISA refers to PETs as "software and hardware solutions, i.e., systems encompassing technical processes, methods, or knowledge to achieve specific privacy or data protection functionality or to protect against risks to privacy of an individual or a group of natural persons"²⁵.

According to the UK Information Commissioner's Office, in its turn, "PETs are technologies that embody fundamental data protection principles by minimising personal data use, maximising data security and/or empowering individuals"²⁶.

Like many major technological innovations, PETs are accompanied by significant underlying ambiguities. On one hand, they provide new functionalities and solutions that support the implementation of key privacy principles, such as data minimization, purpose limitation, privacy by design and by default, and data security, ensuring accountability for both data controllers and processors. On the other hand, however, PETs can also complicate the application of other privacy obligations, such as when a data controller using encrypted data processing tools may lose control over the data being fed into AI models. This could, in some cases, make it difficult to respond to data subject requests (such as access requests from patients) and to ensure that the processed data remains accurate, complete, and up to date.

For this reason, PETs should not be regarded as "silver bullet" solutions. They cannot replace legal frameworks but must function within and in alignment with them. Their use must be combined with legally binding and enforceable obligations to safeguard privacy and data protection rights, except where specific exemptions apply in certain predefined situations (e.g., scientific research to some extent).

After reviewing various studies and research – particularly from academic institutions like the Massachusetts Institute of Technology (MIT) and specialized agencies such as the US National Institute of Standards and Technology (NIST) – and taking into account significant developments in the private sector, the OECD suggests classifying PETs into the following four categories²⁷:

- ✓ **Data obfuscation tools** include zero-knowledge proofs, differential privacy, synthetic data and anonymization and pseudonymization tools. These tools increase privacy protections by altering the data, by adding "noise" or by removing directly or indirectly identifying details. Obfuscating data enables privacy-preserving machine learning and information verification,

²⁴ https://www.oecd.org/en/publications/emerging-privacy-enhancing-technologies_bf121be4-en.html .

²⁵ ENISA's 'Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies' dated 31 March 2016.

²⁶ The already mentioned and more recent update of the ICO's guidance on 'Privacy-enhancing technologies', dated June 2023.

²⁷ See the aforementioned report on 'Emerging Privacy Enhancing Technologies - Current Regulatory and Policy Approaches'.

without requiring personal data collection and processing. These tools can leak information if not implemented carefully, *e.g.*, (allegedly) anonymous data can still allow singling-out, with the help of data analytics and complementary data sets.

- ✓ **Encrypted data processing tools** include homomorphic encryption, secure multi-party computation, as well as trusted execution environments. Encrypted data processing PETs allow data to remain encrypted while in use (so called ‘in-use encryption’), so avoiding the need to decipher them before processing or computing. These kinds of PETs were widely deployed in COVID19 tracing applications, but their average computation costs tend to be high.
- ✓ **Federated and distributed analytics** allows executing analytical tasks upon data that are not visible or accessible to those executing the tasks. In federated learning, for example, data are pre-processed at the source level, meaning that they do not have to leave their repositories and systems, because the computation takes place locally and the model is then trained at central level thanks to already aggregated inputs. In this way, only the summary statistics/results are transferred to the parties that execute the tasks. Federated and distributed analytics requires reliable connectivity to properly operate.
- ✓ **Data accountability tools** include accountable systems, threshold secret sharing, and personal data stores. These tools do not primarily aim to protect the confidentiality of personal data at a technical level and are therefore often not considered as PETs in the strict sense. However, these tools seek to enhance privacy and data protection by enabling data subjects’ control over their own data, and to set and enforce rules for when data can be accessed. Most tools are in their early stages of development, have narrow sets of use cases and lack stand-alone applications.

PETs can be seen as the underpinnings of a new evolving paradigm of privacy and data protection, as they provide more control to data subjects and enhance social trust in the processing of big data, implementing data minimization.

In its data protection glossary, the European Data Protection Supervisor (EDPS) describes PETs as “*a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system*”²⁸.

From a technical perspective, they can be perceived as building blocks towards meeting data protection principles and the obligations on privacy-by-design set out by Art. 25 of GDPR.

For this reason, a growing number of policy makers and supervisory authorities are considering how to incorporate PETs in both domestic and cross-border regulatory frameworks. However, the highly technical and fast evolving nature of these technologies often presents a barrier to implementation by organizations and to their consideration in policy and legal regimes applicable to personal data and, more in general, to new technologies²⁹.

Notwithstanding this, the definition and regulation of at least some PETs in some horizontal legislation, or guidance adopted by competent Supervisory Authorities, is both necessary and urgent in the data-omnivore era of Artificial Intelligence.

A significant regulatory advancement could, however, be on the horizon.

Indeed, with a view to reinforcing the application of fundamental data protection principles and guaranteeing individual rights, the EDPB announced in its official ‘Work Programme 2024/2025’ that new (i) ‘Guidelines on anonymization’ and (ii) ‘Guidelines on pseudonymization’ will soon be adopted³⁰. Hopefully, **these new recommendations will revamp the outdated and rigid interpretation of anonymization that has been bequeathed by the former Article 29 Working**

²⁸ https://www.edps.europa.eu/data-protection/data-protection/glossary/p_en.

²⁹ This is one of the reasons why the ENISA suggested that “*Regulators (e.g., Data Protection Authorities and the European Data Protection Board) and the European Commission should promote the establishment of relevant certification schemes, under Article 42 GDPR, to ensure proper engineering of data protection*” (‘Data Protection Engineering – From theory to practice’ report, dated January 2022).

³⁰ https://www.edpb.europa.eu/our-work-tools/our-documents/strategy-work-programme/edpb-work-programme-2024-2025_en.

Party (now replaced by the EDPB) and that has conditioned the application of this concept by the authorities over the past 10 years (see below for a more detailed analysis on this).

Combined with a crucial judgement which was adopted by the Court of Justice of the European Union on 26 April 2023 – better examined below – which clarified when and by whom data can be considered anonymous³¹, the new standards and best practices laid down by the EDPB may help overcome the highly restrictive and patchy interpretation of anonymization and pseudonymization that many Supervisory Authorities have traditionally embraced and implemented so far.

That being said, given the relevance that some PETs may have in the world of AI, medical research and healthcare, it is worth analysing at least the most promising ones in more detail.

3.2.1 FEDERATED LEARNING

Federated Learning (FL) is a type of remote execution in which models are ‘sent’ to remote data-holding machines (e.g., servers) for local training. This allows researchers to process datasets residing – and remaining – at other sites for training models, without accessing those data. Thanks to FL, e.g., one or more parties – even in different jurisdictions – can build, develop and fine-tune AI models to be integrated in a DT with large volumes of health data in the availability of third parties, even as that data remains ‘invisible’ to each of them.

FL can thus be identified as a data-private machine-learning technique which makes it possible to collaboratively train AI models across multiple sites, without the need to exchange any local data. This strengthens both data security and minimization, aligning with the GDPR and other applicable data protection laws. Unlike traditional methods that require collecting data in a centralized location to train machine learning models, FL ensures that data remains within its original repositories, providing a higher level of privacy.

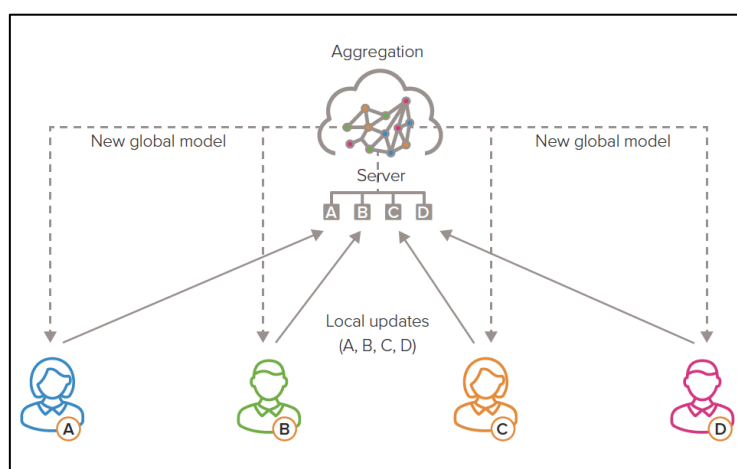


Figure 10 - Federated machine learning illustration³²

There are two approaches to accomplishing Federated Learning:

- ✓ **Centralized:** in this case, each site analyses its own data and builds a model, which is then shared to a remote, centralized location (a node) common to all researchers involved. This node then combines all models into a ‘global’ one and shares it back to each site, where researchers can use the new, improved model. Practically, a co-ordination server creates a model or algorithm and duplicate versions of that model are sent out to each distributed data source (i.e.,

³¹ SRB v. EDPS (Case T-557/20):

<https://curia.europa.eu/juris/document/document.jsf?text=&docid=272910&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=5822411>

³² This figure is taken from the report ‘From privacy to partnership: the role of Privacy Enhancing Technologies in data governance and collaborative analysis’, dated January 2023, prepared by The Royal Society in close collaboration with the Alan Turing Institute.

the clinical partners). The duplicate model trains itself on each local data source and sends back the analysis generated, so that it can be synthesized with all the others coming from other data sources and integrated into the centralized model by the coordination server. This process repeats itself to constantly refine and improve the model;

- ✓ **Decentralized:** in this different hypothesis, the model is built iteratively, as the remote node and local nodes take turns sending and returning information. Each participant sends a gradient on its dataset until the algorithm converges and the iterations use an optimization routine (such as stochastic gradient descent). Hence, there is no central co-ordination server involved in decentralized FL: each site communicates with the others, and they can all update the global model directly.

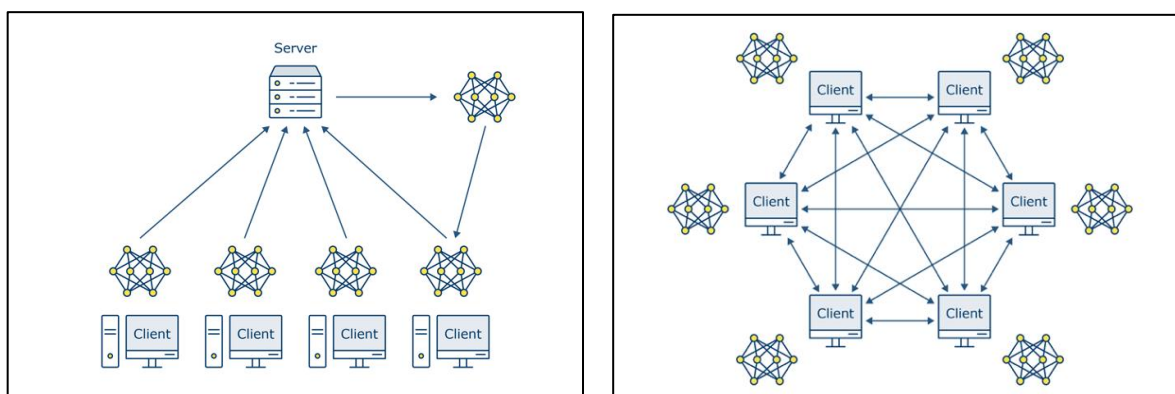


Figure 11 - Centralised and decentralized FL (from UK ICO's guidance on PETS)

In either approach, all the models which must be developed are improved by 'learning' from remote datasets, which are never revealed. This is because FL ensures that raw identifiable data are never pulled out of the data controllers' own repositories and thus shared, preventing the most common issues associated with data protection, such as the risk of a breach.

Nonetheless, models are still vulnerable to some advanced attack. FL applications, for instance, can leak information in the parameters that are sent back to the central node, as local models may preserve features and correlations from the training data samples that could then be extracted or inferred by attackers. For this reason, FL should in certain specific cases be reinforced and coupled with other PETs, such as SMPC to encrypt the shared model parameters and so ensure that they do not reveal their inputs, or Differential Privacy to add sufficient noise to make reasonably impossible singling-out any individual whose data were used in connection with any FL-orchestrated training task.

3.2.1 DIFFERENTIAL PRIVACY

Differential privacy (DP) is a property of a dataset or database, based on the randomized injection of noise, providing a formal mathematical guarantee about people's unidentifiability.

A crucial characteristic of DP is the concept of "epsilon" or ϵ (also known as the "privacy budget" or "privacy parameter"), which determines the level of added noise. More precisely, it represents the worst-case amount of information inferable from the result by any third party about someone, including whether or not they were included in the input dataset.³³

Therefore, noise allows for 'plausible deniability' of a specific person's personal information being in a record, implying that it is not possible to determine with reasonable confidence that data relating to that individual are included in a given set of information. That is to say that **DP allows for risk to be**

³³ Lee, J., and Clifton, C. (2011). *How Much Is Enough? Choosing ϵ for Differential Privacy*. In: Lai, X., Zhou, J., Li, H. (eds) Information Security. ISC 2011. Lecture Notes in Computer Science, vol 7001. Springer, Berlin, Heidelberg . https://doi.org/10.1007/978-3-642-24861-0_22.

quantified as the probability of reidentification, allowing the controller to ‘dial up or down’ and adjust for performance privacy trade-offs by referring to a set ‘privacy budget’.

This could be exemplified by thinking of a situation where someone asks someone else the question: “*Do you like icecreams?*”, which is a binary – yes or no – answer. However, it could be modified with the aid of a ‘coin toss’³⁴. Prior to answering, a coin is tossed: if a head is the result, the person answering the question tells the truth; if not heads, the person will give a ‘random’ answer (which in this case is another coin toss with a predefined “yes” if heads and “no” if not). Notably, though it is possible to deduce the probability of people who like ice-cream, the individuals answering this question now have a so-called ‘plausible deniability’. Indeed, although combining some basic facts about the independence of events may lead to derive a probability distribution, because of the introduction of randomness (*i.e.*, a person’s veracity depends on a coin toss) which produces deniability, the individuals are now permitted to say “I may or may not have answered truthfully”. This is the gravitas of differential privacy³⁵.

Now, to grasp the importance of Differential Privacy, imagine that instead of information about ice cream preferences, the data reveal a person’s health condition. Indeed, DP algorithms “*can provide assurance that after analysing a dataset of several individuals, the outcome of the analysis will not be affected and will remain the same, even if any individual’s data (up to ϵ) was not included in the dataset*”³⁶. Which means that this technique allows studying larger statistical trends in a dataset, while protecting personal information referring to the individuals who participate in it. There are two methods for the privacy budget to be enforced:

- ✓ interactive DP: where the noise is added to each query response and querying is terminated once the privacy parameter ϵ is met (*i.e.*, when the information obtained from queries reaches a level where personal data may be inferred);
- ✓ non-interactive DP: where the privacy budget is set *a priori*, as a property of the dataset itself. Non-interactive mechanism computes some function from the original database and releases the output once and for all, so that it can then be used by anyone to compute the answer to a particular class of queries, without requiring any further interactions with the DP curator. In this case, privacy protection algorithms are processed and published to the database, and users can process this database for any operation.

Noise can be added at the time of data collection (distributed DP), or at the central location before the data are released (centralised DP). In more detail:

- a. **Centralized (or global) DP:** it involves an “aggregator” having access to the real data (*e.g.*, a national Health Data Access Body under the EHDS). Each user of the system sends information to the aggregator without prior adding noise. The central node then applies DP, by adding noise to the output during computation of the final result, before it is shared with any third party. As evident, the key requirement – and so main disadvantage – of this approach is that all users have to trust the aggregator to act appropriately and protect people’s privacy, as it has to access the data in clear.
- b. **Distributed (or local) DP:** in this scenario, each user of the system (or a trusted third party on the latter’s behalf) must apply the DP mechanism before sending the data to the aggregator, preventing the issue relating to the need for trust in the central node. Since noise is added at the individual input data level, the total ‘amount’ of noise required is (much) larger than in global differential privacy, entailing a decrease in data accuracy. However, proper implementation of secure aggregation techniques, including SMPC, can help appropriately addressing this hurdle.

³⁴ Dwork, C., and A. Roth. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9 (3–4): 211–407. <https://www.nowpublishers.com/article/Details/TCS-042>.

³⁵ This example is taken from Bellovin, Steven M., Preetam K. Dutta, and Nathan Reiting. 2019. *Privacy and synthetic datasets*. *Stanford Technology Law Review* 22 (1): 2–52. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3255766

³⁶ ‘Data Protection Engineering – From theory to practice’ report by ENISA, dated January 2022.

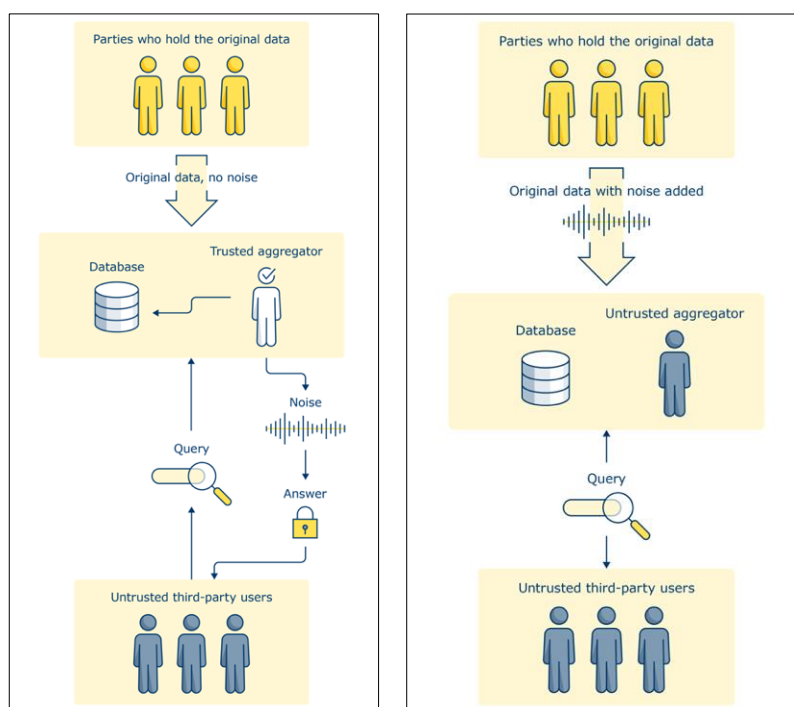


Figure 12 – Global and local DP in UK ICO's guidance on PETS

Both methods of DP can help obtain anonymous information as output, provided that a sufficient level of noise (privacy budget) is added to the data, it being understood that any original information kept by the aggregator in the global model, or by the individual parties in the local approach, represent personal information.

Another important propriety of DP is that post-processing is allowed, meaning that the result of the processing of differential private data through a fixed transformation maintains the same characteristic of DP (e.g., if a generative model is trained using an algorithm satisfying DP, the samples obtained can be published and processed without further privacy implications).

Finally, it must be highlighted that DP and Federated Learning can be combined in two ways: output perturbation (where noise is added to the output of an optimization algorithm) and objective perturbation (where noise is added at every step of the optimization algorithm), achieving high security levels and privacy-by-design

3.3 SYNTHETIC DATA

3.3.1 TECHNICAL SCENARIO

There are various forms of synthetic data, but the term essentially refers to the creation of artificial data that replicates the statistical characteristics of an original dataset. This process involves using AI-driven techniques to learn the relevant distributions of real data and then mimicking and sampling them to generate realistic, yet entirely fabricated, datasets that retain the same statistical properties as the originals. This approach enhances the protection of personal data and patient privacy while preserving the dataset's utility for statistical analysis and scientific research³⁷.

³⁷ The Synthetic Data 'Industry Connections Activity Initiation Document' adopted by the IEEE Standard Association (Version: 2.0, dated 13 November 2023) states that synthetic data "(...) is highly realistic and statistically representative to the original data and thus suitable to serve as a drop-in replacement for it (e.g., for AI training). Yet – when generated with appropriate privacy mechanisms – synthetic data is fully anonymous and impossible to re-identify". <https://ieee-sa.imeetcentral.com/p/eAAAAAASqnGAAAAAACXNJY>.

A more methodological-oriented definition has been provided in a report specifically dedicated to synthetic data which was commissioned by The Royal Society and the Alan Turing Institute, whereby this type of data is described as “*data that has been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s)*”³⁸. Similarly, the TechSonar Report 2022-2023 issued by the European Data Protection Supervisor (a report aimed to anticipate emerging technology trends and better understand future developments in the technology sector from a data protection perspective) defines synthetic data as “*artificial data that is generated from original data and a model that is trained to reproduce the characteristics and structure of the original data*”³⁹.

These data can be either **partially synthetic**, where only some variables are generated by a machine learning model, or **fully synthetic**, in case the entire dataset is algorithmically generated. In the former, original data is combined with replica-data, while in the latter, the dataset is exclusively synthetic. Additionally, **hybrid synthetic datasets** can be created by combining both real-world and fully synthetic data to more accurately reflect the characteristics of the original dataset. For instance, one approach is to select the closest point in the synthetic dataset for each real data point, allowing the reproduction of certain special cases of the source set (*e.g.*, a specific clinical feature), without directly using the real data.

Many practical alternatives exist for generating synthetic data⁴⁰.

The easiest option is drawing samples from a known distribution. In this case, the outcome does not contain any original (and personal) data and re-identification is unlikely to occur, mainly due to randomness. More complex methods rely on mixing real data and fake ones (with the latter being still sampled from known multivariate distributions, conditioned on the real observed data). In this case, some disclosure of personal data and re-identification is possible due to the presence of true values within the dataset, unless additional measures are implemented to prevent any reasonably foreseeable risk of singling-out.

The key distinction from traditional anonymization techniques is that data synthesis involves adding statistically similar information to the original data, rather than removing unique identifiers⁴¹.

Nonetheless, by maintaining the overall statistical properties, analysing synthetic data can yield the same research and mathematical conclusions as for the original data.

As said, methods to produce this type of data vary, but the underlying principle is that some or all of the values in the original dataset are substituted, by specific algorithms (such as GANs – Generative Adversarial Networks)⁴², with others taken from statistically equivalent distributions and structures, to create entirely new records with as little traceable relation to the originals as possible.

Synthetic data can hence help researchers prototype data-driven models and verify and validate machine learning pipelines, providing some assurance of performance and enforcing privacy-by-design.

Among the numerous advantages offered by this technique, the following are worth highlighting, when focusing first on population-based modelling for subsequent personalization in VHT applications:

- ✓ **de-biasing:** given biased training (personal or health) data, a natural approach is to train models using available data; these biases can then be seen in the output of the trained models. Rather than attempting to debias each trained model individually, a de-biased synthetic dataset could be generated and used to train each model, creating a unified approach for handling

³⁸ Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., and Weller, A. Synthetic Data – What, why and how? arXiv:2205.03257, 2022. <https://arxiv.org/abs/2205.03257>.

³⁹ https://www.edps.europa.eu/data-protection/our-work/publications/reports/2022-11-10-techsonar-report-2022-2023_en.

⁴⁰ The model can take many forms, from deep learning architectures such as the popular Generative Adversarial Networks (GANs), or Variational Auto-encoders (VAEs), through agent-based and econometric models, to a set of (stochastic) differential equations modelling a physical or economic system.

⁴¹ Bellovin, S.M., Dutta, P.K., Reitinger, N., *Privacy and synthetic datasets*. 2019, Stanford Technology Law Review 22 (1): 2–52.

⁴² A Generative Adversarial Network uses two models playing against each other: the ‘Generator’ learns to capture and recreate the data distribution, while the ‘Discriminator’ estimates the probability that a generated sample belongs to the original data distribution or rather has been created by the Generator, so determining whether the data is fake or not.

biases across an organization. Such data can be used then for training ‘black box’ machine learning pipelines, while mitigating the risk of historical biases being amplified;

- ✓ **data augmentation and imputation:** synthetic data can be generated with the specific aim to enlarge datasets that are too small, *e.g.*, to provide robustness against ‘outlier’ examples (when they are not needed for the purpose of a medical research). This is often referred to as semi-supervised learning. Synthetic data can both act as a ‘regulariser’, reducing variance in the learned downstream model, and be expanded for imputation (replacing missing values with substitutes), *i.e.*, filling gaps, correcting skewed value distributions, or removing spurious values in the original data. This addresses collection, formatting or normalization issues, which are pervasive especially in clinical datasets, and thus produces data that are actually more informative and realistic than the original ones⁴³. In addition, training machine (and in particular deep) learning networks and models – even more in the scientific area – requires vast amounts of correctly labelled data, which is often costly to produce. Synthetically generated labelled data offer a cost-efficient solution to this challenge;
- ✓ **data minimisation:** a large number of real-world examples demonstrate that high-dimensional, often sparse, datasets are inherently vulnerable to privacy attacks and that existing anonymization techniques too often do not provide adequate protection. By using synthetic data, “*a controller will be respectful of individuals’ confidentiality, since they differ from real data and the generation and processing of synthetic data does not invade the personal sphere of data subjects (in particular when real data refer to individuals’ sensitive characteristics, or to rare attributes that may be difficult to retrieve or may have a significant power of identification)*”⁴⁴.

In light of the above, it is not hard to see why synthetic data are gaining exponential attention.

The EDPS included data synthesis in both the (2021-2022 and 2022-2023) releases of its already mentioned TechSonar, endorsing the advantages described above.

After describing the pros, the EDPS also identifies the cons which must be properly addressed by data controllers (or by processors acting on their behalf):

- ✓ **output control:** especially in complex datasets, the best way to ensure the output is accurate and consistent is by comparing synthetic data with original data, or human-annotated data. However, for this comparison to be carried out, access to the original data is required;
- ✓ **difficulty to map outliers:** as synthetic data can only mimic, replicating specific properties of a phenomenon, but not duplicate real-world data, they may not cover some outliers in the original dataset, even though these can sometimes be more important than regular data points (*e.g.*, in clinical research)⁴⁵;
- ✓ **quality of the model:** quality of synthetic data is highly correlated with that of both the original dataset and the machine learning model used to generate the fake data, which thus may reflect and incorporate the same biases of the original data. Furthermore, the manipulation of datasets to create fair synthetic datasets might result in inaccurate data. Therefore, controllers will always need to reconcile the tension between different data protection principles, especially if the result of the processing entails consequences (*e.g.*, legal or health implications) for data subjects.

⁴³ Morley-Fletcher, E., ‘*New Solutions to Biomedical Data Sharing: Secure Computation and Synthetic Data*’, in *Personalized medicine in the making: philosophical perspectives from biology to healthcare*. Beneduce, C., Bertolaso M., Springer (2021), 173-189.

⁴⁴ ‘*Data Protection Engineering – From theory to practice*’ report by ENISA.

⁴⁵ For example, it would be very difficult to ‘hide’ someone affected by a very rare disease in a synthetic dataset containing information about people with cardiological pathologies. A synthetic data generator would either not accurately replicate statistics regarding that specific person, or would otherwise reveal potentially private information about him/her.

3.3.2 LEGAL SCENARIO

Both the EDPS and EDPB have been paying significant attention to synthetic data, organizing a dedicated webinar⁴⁶ and, as previously mentioned, featuring this technology in both the first and second editions of the TechSonar⁴⁷.

But more importantly, synthetic data have for the first time secured their rightful place in EU legislation, specifically within the AI Act. In more detail:

✓ **Article 10.5 of the AI Act** states that:

«To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems (...) the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to the provisions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met in order for such processing to occur:

- (a) *the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data;*
- (b) *the special categories of personal data are subject to technical limitations on the re-use of the personal data, and state-of-the-art security and privacy-preserving measures, including pseudonymisation;*
- (c) *(..)»*

✓ **Article 59.1, (b) of the AI Act** establishes that:

«In the AI regulatory sandbox, personal data lawfully collected for other purposes may be processed solely for the purpose of developing, training and testing certain AI systems in the sandbox when all of the following conditions are met:

- a) *(...)*
- b) *the data processed are necessary for complying with one or more of the requirements referred to in Chapter III, Section 2 [namely those applicable to high-risk AI systems] where those requirements cannot effectively be fulfilled by processing anonymised, synthetic or other non-personal data;*
- c) *(...)»*

In the legal realm, these seemingly modest words carry significant weight concerning the regulatory status of synthetic data.

Article 10 of the AI Act equates anonymized data with synthetic data, effectively treating them as equivalent and interchangeable options. This is further reinforced by: (i) the use of the word “or” between the two, and (ii) the clear separation from pseudonymized data, which is listed separately and categorized as distinct from synthetic data.

Article 59, in turn, clearly defines synthetic data by first grouping it with anonymous data and then explicitly classifying synthetic data as non-personal data. The phrase “or other” clearly indicates that both anonymous and synthetically generated data are categorized together as unidentifiable data, thereby placing them outside the scope of data protection legislation.

A fundamental principle of law is that legal provisions must be interpreted according to their literal meaning, based strictly on the wording chosen by the Legislator. In short, any subjective interpretation or alternative understanding must be excluded.

⁴⁶ Internet Privacy Engineering Network (IPEN) Webinar on the 16 June 2021, entitled “Synthetic data: what use cases as a privacy enhancing technology?”.

⁴⁷ See footnote no. 39.

In this context, **the EU Legislator’s intent to treat anonymous and synthetic data as equivalent** – especially concerning key obligations imposed on AI developers – **becomes unmistakably clear**.

Nevertheless, considering that a legal definition of synthetic data has not been provided by the legislator yet, a more cautious approach has been adopted by the European Data Protection Supervisor. The Authority does not exclude that, under specific circumstances, synthetic data could lead to the re-identification of the data subjects⁴⁸. Therefore, the requirement for data controllers to meet the 'reasonable non-identifiability' test, as outlined in Recital 26 of the GDPR, remains unchanged as a key condition for data to be considered anonymous (see further details below).

In practice, this means that market players must carefully assess several factors on a case-by-case basis – such as the type of generative model used, the context, methods, and purposes of the processing, the nature of the data, the category of data subjects, and the application of Differential Privacy or other PETs as additional privacy-preserving layers – to determine whether any risk of re-identification persists, despite the technical, organizational, and legal safeguards in place to mitigate foreseeable threats to individuals and their data⁴⁹.

Furthermore, *“Because risk is likely to evolve over time depending on the context, events, time, or agents, the very nature of risk affects the legal determination of synthetic data in the way that, under certain circumstances, synthetic data will be considered anonymous data, while in others, this will not be the case”*⁵⁰.

This assessment is inherently subjective, especially in the absence of precise rules, indicators or acceptability thresholds to evaluate the risk of re-identification, best practices include utilizing quantifiable statistical assessments whenever possible, along with performing penetration or motivated intruder testing.

Moreover, *ex-post* audit on re-identifiability and data protection impact assessment can assist in defining appropriate additional safeguards which may be required. Indeed, synthetic data requires validation of:

- a. utility: the property that the performance and predictive power of algorithms based on ML-generated data is not substantially lower than performance on the original data;
- b. obfuscation: the property that the synthetic data do not leak private information from the original data.

A comprehensive evaluation of various factors related to the data, the data environment, and applicable mitigations and safeguards, will determine whether synthetic data are classified as personal data or non-personal data.

The question – as well as the judicial and supervisory authorities’ divergent interpretations and academic debate as to what constitutes anonymous data – entirely turn on whether the risk of re-identification (i) must reach zero or not, and (ii) should be assessed in relation to everyone and in all contexts (as suggested by the Article 29 Working Party’s ‘Opinion 05/2014 on anonymization techniques’ – see below for more info), or only in relation to specific parties depending on their roles and contractual agreements.

Moreover, some more peculiar challenges should be borne in mind by regulators when dealing with synthetic data.

First, if it is genuinely established that synthetic data qualify as ‘personal data,’ considerable complexities would arise in figuring out how data protection rights and obligations should be applied. For example, how can the principle of data accuracy and right to rectification be enforced to synthetic

⁴⁸ EDPS, TechSonar report 2022-2023. https://www.edps.europa.eu/system/files/2022-11/22-11-10_techsonar_report_22_23_en.pdf.

⁴⁹ An article on synthetic data published by the EDPS and authored by Robert Riemann, reads that “A privacy assurance assessment should be performed to ensure that the resulting synthetic data is not actual personal data. This privacy assurance evaluates the extent to which data subjects can be identified in the synthetic data and how much new data about those data subjects would be revealed upon successful identification”. https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en.

⁵⁰ Fontanillo Lopez, C.A., Elbi, A., *On the legal nature of synthetic data*, Center for IT and IP Law, KU Leuven, NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research. <https://openreview.net/forum?id=M0KMbGL2yr¬eId=0mH-aK63WH>.

data, where it is not even clear if a relevant individual is the focus of the data? How does the right to object apply if an individual’s data have at some point been used to develop a model⁵¹?

Secondly, some practical and legal differences between pseudonymized or anonymized data, on the one hand, and synthetic ones on the other, support the notion that this latter type of AI-generated information represents a distinct third category of data, separate from the two above. Indeed, the nature of some data synthesis techniques and models results by-definition in almost negligible reidentification risks.

In conclusion, **it is high time** – also on the account of the more flexible interpretation provided in the aforementioned crucial ruling recently adopted by the Court of Justice of the European Union (CJEU) (SRB v. EDPS, Case T-557/20)⁵² – **for all Data Protection Authorities and regulators to adopt a clear shift in their approach regarding the conditions under which data can be deemed no longer identifiable, opening the door to advanced PETs, also combined together, and to ML-driven synthetic data generation.**

By simplifying and facilitating data reuse while fully respecting both the necessary levels of security and the rights vested to individuals, the path will be paved for the integrated and cross-border creation and broader dissemination of the VHT, that both require increasingly larger volumes of adequately protected data to train the associated AI models and systems, in the healthcare and medical research sectors.

3.4 ANONYMOUS DATA: STILL A BIG QUESTION MARK

In a widely recognized article published in December 2020, prominent researchers Aloni Cohen and Kobbi Nissim highlighted that “*There is a significant conceptual gap between legal and mathematical thinking around data privacy*”⁵³. This statement is especially true when it comes to data anonymization.

Despite the critical importance of distinguishing between personal and non-personal data, in most cases, it remains extremely challenging to clearly differentiate between these two categories. This difficulty is anchored in both technical and legal factors. From the first perspective, the increasing availability of data points and sources, as well as the continuing sophistication of data analysis algorithms – even more in connection with machine or deep learning – and performant hardware makes it easier to link datasets and infer personal information from ostensibly non-personal data.

From a legal standpoint, even after 29 years of data protection legislation (including 6 under the GDPR)⁵⁴, it is still unclear what the correct legal test is that should be carried out to correctly qualify data as anonymous or not⁵⁵. The already often-mentioned Recital 26 of the GDPR (see below) specifies that data are anonymous if it is ‘reasonably likely’ that they cannot – or can no longer – be linked to an identified or identifiable individual. By contrast, many national Supervisory Authorities and particularly the Article 29 Working Party (WP29, replaced by the EDPB since the entry into force of the GDPR) have, however, provided very rigid interpretations of the concept that conflict with this legislative text.

⁵¹ Mitchell C., Redrup Hill, E., *Are synthetic health data ‘personal data’?*; PHG Foundation, Cambridge University; 2023. <https://www.phgfoundation.org/report/are-synthetic-health-data-personal-data>.

⁵² See Note no. 31.

⁵³ Cohen A; Nissim K., *Towards Formalizing the GDPR’s Notion of Singling Out*, 117, Proceedings of the National Academy of Sciences, 8344, 2020. <https://www.pnas.org/doi/10.1073/pnas.1914598117>.

⁵⁴ The Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data became applicable in December of 1995, while the GDPR entered into force in May of 2018.

⁵⁵ Finck, M., Pallas, F. *They who must not be identified - distinguishing personal from non-personal data under the GDPR*. International Data Privacy Law, 2020, Vol. 10, No. 1. <https://academic.oup.com/idpl/article/10/1/11/5802594?login=false>.

Recital 26 – GDPR

«The principles of data protection should apply to any information concerning an identified or identifiable natural person (...) To determine whether a natural person is identifiable, account should be taken of all the means **reasonably likely to be used**, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means **are reasonably likely to be used** to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, **taking into consideration the available technology at the time of the processing and technological developments**. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable (...)

Regarding re-identification risks, the EU legislator highlights the importance of evaluating the following factors:

- ✓ **reasonability**, meaning that the relevant risks cannot not be completely and definitively ruled out, *i.e.*, in respect to anybody and any time. Threats are highly contextual and technology-dependent;
- ✓ **state-of-the-art technology**, meaning that data controllers and processors are not required to predict future technological advancements that could potentially reverse anonymization measures and identify individuals;
- ✓ **Available means and resources**, meaning that the assessment should take into account the tools and capabilities that could be leveraged by the parties involved, ensuring that risks are evaluated in relation to each individual's situation. Anonymization does not have to be universally foolproof, but should be assessed on a case-by-case basis

In contrast to a risk-based and flexible approach outlined by the applicable law, WP29 had taken a much stricter position in its 2014 guidelines on anonymization and pseudonymization⁵⁶, prescribing a zero-risk test. WP29 emphasized that anonymization should eliminate any and all risks of re-identification, in contrast to the more adaptable framework established by the law, by specifying that:

- ✓ “*anonymization results from processing personal data in order to irreversibly prevent identification*”;
- ✓ “*the outcome of anonymization as a technique applied to personal data should be, in the current state of technology, as permanent as erasure, *i.e.*, making it impossible to process personal data*”;

In addition, the WP29 formulation crucially indicated that “*when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data*”⁵⁷.

In addition to being unrealistic in most practical situations, this interpretation has faced substantial academic criticism. There are numerous instances where a controller may choose – or be required – to share anonymized data while retaining the original dataset. For example, a hospital might provide anonymized data for research purposes while keeping the original identifiable data for patients' care, demonstrating the impracticality of a zero-risk standard.

In short, this approach by WP29 – which has then been embraced also by some national data Protection Authorities – implies a rejection of the ‘reasonable re-identifiability’ approach set out by Recital 26 GDPR, as it considers the risk stemming from keeping the initial data to be intolerable in any event. Indeed, “*the concepts of irreversibility, permanence, and impossibility stand for a much stricter*

⁵⁶ Article 29 Working Party, *Opinion 05/2014 on Anonymization Techniques* (WP216), adopted on 10 April 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

⁵⁷ *Ibid.*

approach than that formulated by the legislative text itself”, with the effect that “These diverging interpretations have prevented legal certainty as to what test ought to be applied in practice”⁵⁸.

Confusion is also increased – and so reliance on anonymization is made riskier – due to:

- i. the formalization of the concept of pseudonymous data in the GDPR, because in some jurisdictions (e.g., in UK and Ireland)⁵⁹ and sectors (such as clinical trial), this type of data can, under certain specific circumstances, be considered anonymous, at least for some of the parties involved in the process;
- ii. discordant interpretations by national competent Supervisory Authorities as to the degree of irreversibility that individual de-identification must achieve so that data can be deemed anonymous and not pseudonymous⁶⁰.

In practice, this fragmentation leads to scenarios where a company seeking to carry out a project (such as a clinical study or scientific research) or launch a new service or technology (e.g., based on generative AI) across multiple Member States, entailing the use of anonymized data, must navigate a patchwork of rules, stemming from the diverging decisions made by national Data Protection Authorities, which greatly complicate and hamper cross-border compliance efforts.

As it has been highlighted, the solution to overcome this detrimental and still-locally-centred scenario depends on “whether an ‘absolute’ or ‘relative’ approach is adopted; i.e., whether identifiability is judged according to the abilities of anyone and everyone (including the data controller) to re-identify data or whether the relevant question is whether the data are ‘identifiable’ in the ‘hands’ of a specific actor”⁶¹.

As obvious, the evaluation as to when re-identification may be considered reasonable (Recital 26 GDPR) heavily varies based on a number of factors which may affect the processing. **The characterization of the health data is context-dependent, so that the personalization of risk “should not be seen as a property of the data but as a property of the environment of the data”⁶².**

It is easy to see how the level of risk can vary greatly depending on whether the entity attempting to re-identify individuals is a private individual, a law enforcement agency, or a large online platform. The capabilities, resources, and intent of these actors differ significantly, influencing the likelihood and potential consequences of re-identification.

As of now, regulators and Supervisory Authorities have not established – nor reached a consensus at the EU level – on (i) the standard of reasonableness that should be applied to uniformly assess risk of data anonymization across all Member States, and (ii) whether the assessment should follow an objective or subjective approach⁶³.

This lack of agreement continues to create uncertainty in the application of data protection rules.

⁵⁸ Finck, M., Pallas, F. *They who must not be identified - distinguishing personal from non-personal data under the GDPR*. <https://academic.oup.com/idpl/article/10/1/11/5802594>.

⁵⁹ In its draft ‘Anonymization, pseudonymization and privacy enhancing technologies guidance’ (Chapter I), the UK Information Commissioner’s Office points out that “(...) you will not always be able to state that a specific technique or set of controls will achieve these aims, particularly as technology changes over time. This means that even where you use anonymization techniques, a level of inherent identification risk may still exist. However, this residual risk does not mean that particular technique is ineffective. Nor does it mean that the resulting data is not effectively anonymized for the purposes of data protection law when you consider the context. Also, data protection law does not require anonymization to be completely risk-free. You must be able to mitigate the risk of re-identification until it is sufficiently remote that the information is ‘effectively anonymized’” (<https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/>). Also, the Irish Data Protection Authority deems that it is not “necessary to prove that it is impossible for the data subject to be identified in order for an anonymization technique to be successful. Rather, if it can be shown that it is unlikely that a data subject will be identified given the circumstances of the individual case and the state of technology, the data can be considered anonymous”. <https://www.dataprotection.ie/en/dpc-guidance/anonymisation-and-pseudonymisation>

⁶⁰ TEHDAS (the Joint Action Towards the European Health Data Space) specified in a number of reports that “there is a lack of common European interpretation of what constitutes ‘sufficient anonymization’ to transform personal data to non-personal data” and “what constitutes ‘pseudonymization’” (amongst others, <https://tehdas.eu/tehdas1/results/tehdas-identifies-barriers-to-data-sharing/>).

⁶¹ Mitchell C., Redrup Hill, E., *Are synthetic health data ‘personal data’?*; PHG Foundation, Cambridge University; 2023.

⁶² Stalla-Bourdillon S., Knight A. *Anonymous Data v. Personal Data - A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data* (2017) 34 Wisconsin International Law Journal 284, 301. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927945.

⁶³ Finck, M., Pallas, F. *They who must not be identified - distinguishing personal from non-personal data under the GDPR*.

3.4.1 JUDGEMENT IN CASE SRB V. EDPS (CASE T-557/20)

On 26 April 2023, the General Court of European Union (‘**General Court**’ or ‘**ECG**’)⁶⁴ adopted a crucial decision regarding the issue of data pseudonymization, in the Case T-557/20 (‘**Judgement**’ or ‘**Decision**’)⁶⁵.

The dispute arose from the Single Resolution Board’s (‘**SRB**’, the central resolution authority within the EU Banking Union) request to Deloitte to carry out certain assessments, aimed at determining whether the shareholders and creditors of *Banco Popular Español* would have received a better treatment if the bank had been subject to normal insolvency proceedings. In this context, the affected shareholders and creditors submitted five complaints under Regulation 2018/1725⁶⁶ to the European Data Protection Supervisor, alleging that the SRB failed to mention the transmission of data collected to third parties in its privacy statement, thereby violating its transparency obligations relating to the processing of personal data under the said Regulation.

In this case, the SRB argued that the disclosed data did not constitute personal data, as Deloitte received the information in pseudonymized form (represented by alphanumeric codes), which would not have enabled the firm to re-identify the shareholders. However, the EDPS concluded that the transmission involved pseudonymized data, and thus still qualified as personal data, due to the existence of additional information that could have allowed the re-identification of complainants, although such information was held and accessible only by the SRB. Therefore, the latter Board brought an action before the European General Court (EGC) seeking, *inter alia*, annulment of the decision of the EDPS.

Given that the concept of personal data within the meaning of Article 3(1) of Regulation 2018/1725, coincides with the definition provided by Article 4(1) of the GDPR, the General Court was essentially called upon to decide whether the information that was transmitted to Deloitte had to be qualified as personal data, *i.e.*, information relating to an identified or identifiable natural person.

Building on the principles established by the Court of Justice of the European Union in the well-known Breyer judgment (C-582/14), the EGC stated that “*in order to determine whether the information transmitted to Deloitte constituted personal data, it is necessary to put oneself in Deloitte’s position in order to determine whether the information transmitted to it relates to ‘identifiable persons’*” (Para. 97 of the Judgement). In this light, the General Court held that the sole alphanumeric codes received by Deloitte would not have allowed it to identify the complainants, as it had no access to the additional information needed to re-identify the data subjects, which was under the exclusive control of the SRB.

According to the Decision, the EDPS should have assessed whether Deloitte could re-identify the data subjects, rather than focus on the SRB’s ability to do so. The fact that the SRB held additional information allowing it to single out the data subjects was not sufficient to conclude that the data transmitted to Deloitte were personal in nature. For these reasons, the EGC upheld the SRB’s plea and annulled the decision of the EDPS.

This judgment, which is currently under appeal by the EDPS⁶⁷, has been widely praised by many European privacy professionals, particularly for its fresh perspective on the processing of pseudonymized data⁶⁸.

The principles outlined by the EGC represent a significant advancement compared to the hardline embraced by WP29 in its Opinion 05/2014 on anonymization techniques. **The Judgment indeed**

⁶⁴ Along with the Court of Justice, the General Court is one of the EU’s courts making up the Court of Justice of the European Union. The purpose of these courts is to ensure a uniform interpretation and application of EU law. Decisions of the General Court can be appealed to the Court of Justice, but only on a point of law. Before the Lisbon Treaty came into force on 1 December 2009, it was known as the Court of First Instance.

⁶⁵ The text of the ruling, which is currently under appeal by the EDPS (C-413/23 P):

<https://curia.europa.eu/juris/document/document.jsf?text=&docid=272910&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=5822411>.

⁶⁶ Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC.

⁶⁷ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62023CN0413>.

⁶⁸ A comment on this Decision can be found in the ‘Two decades of personal data protection. What next?’ publication prepared for the 20th anniversary of the EDPS. https://www.edps.europa.eu/system/files/2024-06/edps_20thanniversary-book_en.pdf.

clarifies that the risk of re-identifiability must be assessed concretely, focusing on the specific circumstances and position of the data recipient, rather than considering all parties involved. In this light, the much stricter approach stemming out of the WP29's traditional interpretation of anonymization – which affected tons of decision by some Supervisory Authorities after 2014 – should not be compatible with the more flexible criterion of 'reasonable re-identification' established by Recital 26 of the GDPR.

This shift in approach could have notable practical implications for the handling of pseudonymized data sharing. **If re-identification is practically impossible for the entity receiving such data, because it lacks the additional information or tools required for (or in any case allowing) singling-out the individuals, then, for that specific entity, the pseudonymized data may not be considered personal data.** As a result, data protection laws would not apply, and data would be *de facto* shared in anonymous form, exempting both the data controller and the processor(s) involved from abiding several burdensome obligations.

Additionally, when transferring pseudonymized data outside the European Economic Area, if the EU-based controller commits contractually not to provide the recipient (unless compelled by a competent public or judicial authority) with the means to re-identify the data (such as the private pseudonymization key), the controller will not need to adhere to the stringent rules on international data transfers outlined in Chapter V of the GDPR.

Hoping that the appeal will not reverse or affect in any manner the Decision, the rigid applications of data anonymization so far endorsed by some Supervisory Authorities – stemming from the interpretation offered by WP29 10 years ago – might be toned down in the light of this clear, up-to-date and more than commendable Decision by the General Court.

This would greatly simplify the sharing of unidentifiable data, while still protecting data subjects and particularly patients and their sensitive data by means of advanced pseudonymization techniques, also in combination with novel PETs, triggering highly positive effects in the healthcare, AI and scientific research fields.

Additionally – and equally significant – this 'subjective' interpretation of anonymous data represents a first European response to the growing demand from the research community across Member States for more standardized and consistent criteria to determine whether data can be classified as anonymous.

However, some additional aspects need as well to be taken into account, since the EGC Decision cannot be considered a silver bullet in all circumstances.

Regarding the key principles affirmed by the EGC, it should be noted that the Judgment does not establish broad presumptions or standard rules for assessing the re-identifiability of pseudonymized data. Instead, the General Court emphasized the need for a case-by-case analysis, focusing on the specific circumstances pertaining to each of the parties involved and, particularly, the resources they have to potentially trace-back data subjects. This implies that the same pseudonymized data might be considered anonymous for one party but re-identifiable for another, leading to different compliance obligations depending on the context.

From a purely judicial standpoint, the Judgement is not yet final: the EDPS brought an appeal before the CJEU, alleging that the General Court misinterpreted the provisions regarding privacy obligations laid down by Regulation 2018/1725.

There is no need to emphasize that the final ruling by the CJEU will have a huge impact on the entire technological and healthcare environment, including in connection with Artificial Intelligence and VHT.

4 General Data Protection Regulation

4.1 THE FRAMEWORK FOR VHT

The ability to gather data from the ‘physical counterparts’ is crucial for creating Virtual Human Twins and advance in clinical domains.

These data may be collected from a wide variety of sources, including *inter alia*: Electronic Health Records (*e.g.*, patient data such as medical history, lab results, diagnoses, treatments, and medications); medical imaging (*e.g.*, data from CT scans, MRIs, X-rays, and other imaging techniques); wearable Devices and Sensors (*e.g.*, Continuous health data such as heart rate, blood pressure, glucose levels, and activity metrics from devices like smartwatches or fitness trackers); Genomic Data (*e.g.*, Genetic information gathered through genome sequencing, which can be used to predict disease risks and personalize treatments); Patient Monitoring Systems (*e.g.*, data from ICU monitors, ventilators, and other hospital equipment that track real-time patient conditions, such as vital signs and oxygen levels); laboratory test results (*e.g.*, blood tests, urine tests, and other diagnostic lab data that provide insights into the biochemical and physiological state of the patient); behavioural and lifestyle data (*e.g.*, information about a patient’s habits, diet, physical activity, and other behavioural factors collected through surveys, apps, or self-reporting tools); pharmaceutical data (*e.g.*, information on prescribed medications, treatment adherence, and responses to treatments, including data from clinical trials); surgical and procedural data (*e.g.*, data from surgeries and other medical procedures, including pre- and post-operative outcomes); environmental and socioeconomic data (*e.g.*, factors such as geographical location, and socio-economic conditions that can affect a patient’s health).

Quite evidently, almost all the abovementioned data are of a personal nature and, for the most part, also fall into sensitive (so-called ‘special’) categories under the GDPR.

Collecting and processing such data in a socially responsible, legally compliant and ethically oriented manner, ensuring individual privacy, and minimizing algorithmic bias through inclusive and representative data collection, are essential preconditions for the development of digital twins. However, these priorities present complex challenges for the broader research and professional communities.

Due to the personalized nature, which is specifically tailored to a unique individual, the Virtual Human Twin inherently aggregates both data, that may be identifiable, or re-identifiable, and models providing customized insights about the specific examined condition. In medical contexts, a digital twin might even ideally encompass a patient’s entire health history.

Accordingly, as in their full ‘lifecycle’ DTs are associated with vast amounts of sensitive data, all highest safeguards afforded by the GDPR must apply, in order to protect the data subjects’ rights and confidentiality and ensure appropriate level of data security. Indeed, the advanced simulation and predictive capabilities of DTs require collecting and processing data at such a large scale and depth that this element alone inherently heightens the risk of privacy breaches.

Primarily, any and all data processing must be designed and carried out in accordance with the principles enshrined in Art. 5, the scope and relevance of which are such that they are in most cases also impactful from an ethical point of view:

- ✓ **lawfulness, fairness and transparency:** personal data connected to or used for DTs (such as medical histories, genomic data, and real-time health metrics) must be collected and processed in a lawful, fair, and transparent manner. Patients, and more generally individuals, need to be clearly informed about how their data are being used in connection with the VHT, while data controllers must ensure that a sound legal basis is in place which legitimizes the processing (such as the individual’s consent, the necessity for contract performance, compliance with legal obligations, protection of vital interests, performance of a task carried out in the public interest, or legitimate interests pursued by the data controller), and that a specific condition is met under Art. 9 GDPR which permits the collection and use of special categories of data.

Transparency is key to maintaining trust and enabling the data subjects to keep control over their personal data, how they will be exploited, for what purposes and under which conditions of lawfulness. This clearly has also an ethical reach, particularly when complex and not-easy-to-understand AI or machine learning systems and/or models come into play.

In addition, processing must not be misleading or deceptive, meaning that individuals should be treated fairly in all stages of the data processing cycle. For example, personal data should not be used in a way that disadvantages individuals in ways they were not informed about or that they would not reasonably expect.

- ✓ **Purpose Limitation:** the data must only be used to achieve the specific objectives for which they were initially collected (*e.g.*, providing health and care services to patients), as described in the information notice given to the data subjects. Reusing the data for unrelated purposes would violate this principle, in absence of the individual's consent, except for very limited cases which include scientific research. This ensures that organizations cannot collect data under a vague or broad premise and then use them in ways that are incompatible with the originally declared processing goals.

However, **secondary processing for, among others, scientific research purposes is considered compatible with the initial purpose, provided that appropriate safeguards are in place to ensure data minimization (see below for more details).**

Despite this significant regulatory opening aimed at supporting the research sector, we will see that the issues surrounding the secondary processing of health data will prove to be one of the major hurdles to overcome.

- ✓ **Data Minimization:** only data that are adequate, relevant and limited for attaining the intended purposes must be collected. Unnecessary or excessive data processing should then be discouraged and avoided. In this sense, this essential principle requires controllers to consider what the minimum amount of data is that is necessary to achieve the intended objective. The same reasoning applies to the number of people who, on a strictly-needed basis only, may be granted access to the data.

Data minimization helps to reduce the risk of misuse, ensures better compliance with privacy requirements, and limits the potential exposure of personal information.

- ✓ **Accuracy:** personal data must be accurate and kept up-to-date. Inaccuracies can lead to unfair treatment of individuals or flawed decision-making. Therefore, data controllers are responsible for taking reasonable steps to ensure the accuracy of the data they hold and, where necessary, to correct or delete inaccurate data.

This principle is even more important in the healthcare setting, where incorrect data could have severe consequences to patients, *e.g.*, since inaccurate or outdated information can lead to errors in the simulation or predictions made by the VHT, potentially compromising patient care.

- ✓ **Storage Limitation:** data should only be kept in a form that allows identification of individuals for as long as necessary for the purposes for which they were collected. Once such goals are fulfilled, the data should either be deleted or made anonymous, so that it can no longer be linked to individuals. This principle prevents indefinite retention of personal data, which would increase the risk of breaches and misuse. Unnecessary long-term storage of health data, which may increase risks of privacy breaches, should be avoided.
- ✓ **Integrity and Confidentiality:** personal data must be collected and processed in a manner that ensures their security. This involves implementing appropriate technical and organizational measures against unauthorized or unlawful processing, accidental loss, destruction, or damage. This principle encompasses data confidentiality (keeping data private), integrity (ensuring data is accurate and not tampered with), and availability (ensuring that data can be accessed when needed). For instance, encryption or pseudonymization, access controls, regular audits, and secure storage systems are vital to upholding this principle, particularly in connection with data falling into special categories. Data controllers and processors must also ensure that only (a

limited number of) duly authorized individuals can access the personal data, and that all individuals handling personal data are properly trained in data protection and privacy protocols.

- ✓ **Accountability:** this is one of the inspiring principles underpinning the whole European data protection legislation, requiring organizations to take responsibility for abiding all applicable obligations and to be able to demonstrate compliance. This means that both data controllers and processors must put in place appropriate technical, legal and organizational measures, such as implementing internal data protection policies, maintaining records of processing activities, appointing Data Protection Officers, where necessary, carrying out staff training, conducting data protection impact assessments, and implementing other tools to ensure due and demonstrable alignment with legal requirements.

By adhering to these GDPR principles, the development and the use of VHT in healthcare can be afforded while protecting patient privacy, maintaining data security, and ensuring that personal data are collected and processed responsibly and transparently.

In accordance with the above, each and every purpose for which personal data are intended to be collected and processed must be backed by a specific lawfulness ground, aimed to guarantee that the controller is duly permitted to undertake the envisaged activity. Under Article 6 of the GDPR, data processing is considered lawful if and to the extent that:

- a. **Consent of the data subject:** the data subjects have given their free and specific consent to the processing of their personal data for a well-determined purpose. Consent must be informed, unambiguous and revocable.
- b. **Performance of a contract:** it is necessary for the performance of a contract to which the data subject is a party (*e.g.*, for the provision of health services), or to take steps at the request of the data subject before entering into a contract.
- c. **Compliance with a legal obligation:** it is needed to fulfil a legal obligation to which the data controller is subject. This covers situations where the law requires certain data to be processed, such as for occupational health and safety requirements, retention of medical records, prescription monitoring, medical device incident reporting.
- d. **Protection of vital interests:** it is necessary to protect the vital interests of the data subject or another person. This typically refers to life-threatening situations, such as processing medical data during an emergency to provide care.
- e. **Public interest or exercise of official authority:** it is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the data controller. This often applies to government bodies or organizations acting under a legal mandate (*e.g.*, for public health research initiatives, government-mandated health screenings, national immunization programs, population health surveillance, managing national health registries).
- f. **Legitimate interests:** it is needed for the purposes of the legitimate interests pursued by the data controller or a third party, provided these interests are not overridden by the data subject's rights and freedoms, especially when they are children. This is a flexible basis but requires a careful balancing test to ensure the individuals' rights are not disproportionately affected. In the medical sector, this legal basis should be applied even more cautiously due to the sensitive nature of health data.

Provided that one of the legal bases mentioned above is relied upon, Article 9 of the GDPR affords further protection to certain types of personal data due to their inherently sensitive nature, which could lead to significant harm or discrimination if misused or anyway processed in violation of the applicable principles. These special categories of data include information revealing racial or ethnic origin, political opinions, religious beliefs, as well as biometric data, data concerning sex life or sexual orientation and, crucially for the research and medical sector, health and genetic data. The reasons for this heightened protection are, in brief:

- Increased risk of harm

Sensitive data, such as health and medical information, as well as genetic data, if accidentally or illegally disclosed or processed, can cause serious harm to individuals. For example, a breach of health data could lead to stigmatization, discrimination, or psychological distress. Health data might reveal private information about an individual's medical conditions, which could be misused for insurance, employment, or even social discrimination. This data can also be exploited for identity theft or fraud, as medical and genetic information are highly valuable in criminal markets.

- Fundamental rights and freedoms

Sensitive data relates closely to any individual's fundamental rights and freedoms. Health data, for example, touch on a person's physical and mental well-being, which is intimately connected to their right to privacy and human dignity. Misuse of such data can infringe on these fundamental rights, as it exposes highly personal and vulnerable aspects of an individual's most private sphere.

- Potential for discrimination

Sensitive data, especially genetic, racial, ethnic, and health data, can be used to unjustly discriminate against individuals. In healthcare and medical contexts, this could affect access to treatment or insurance coverage. These forms of data require higher protection to prevent individuals from being unfairly treated based on their personal characteristics or medical conditions. For instance, genetic data could be used by insurers to refuse coverage to individuals who are predisposed to certain medical conditions, or employers might discriminate against individuals based on their health records.

- Confidentiality in healthcare

Health data is among the most sensitive categories among all, as they may encompass an individual's medical history, treatments, diagnoses, and also genetic information. Maintaining the confidentiality of such data is crucial in healthcare, as it directly impacts the trust relationship between patients and healthcare professionals. If health data were not given stronger protection, patients might be reluctant to share important information with their healthcare providers, which could negatively impact diagnosis and treatment.

- Risk of profiling

Sensitive data, particularly genetic data, can be used for profiling individuals, which could even lead to invasive monitoring, surveillance, or behavioural predictions. In healthcare, this kind of data might be used to profile individuals for certain risks, affecting their access to services or treatments. Profiling based on sensitive data could also lead to predictive decisions that affect an individual's future opportunities, leading to inequities or further reinforcing discrimination. For this reason, this type of processing is prohibited in most of the cases.

- Public trust and ethical considerations

Higher protection is necessary to maintain public trust and social license in organizations that process sensitive data, particularly in sectors like healthcare. Patients must trust that their health data will be handled with the utmost care and confidentiality. Without this trust, people may be unwilling to engage in medical research, clinical trials, or share their data for public health purposes. Moreover, processing of sensitive data involves ethical considerations, especially when it concerns vulnerable groups, such as children, the elderly, or those with disabilities. These groups are often more susceptible to harm if their sensitive data is misused, and thus the GDPR imposes stricter safeguards to ensure ethical data processing.

In general, the processing of health (including medical) and genetic data is prohibited, unless one of the following exceptions – focusing on those which may be relevant in connection with the VHT environment – applies:

- **Explicit Consent** (Article 9(2)(a))

Health data can be processed if the data subjects give their explicit consent for one or more specific purposes (*e.g.*, a patient giving consent for his/her data to be used in a clinical trial, or for a particular treatment plan).

- **Provision of health or social care** (Article 9(2)(h))

Processing is permitted if it is necessary for the purposes of preventive or occupational medicine, medical diagnosis, the provision of health or social care or treatment, or the management of health or social care systems. This exception covers many standard healthcare activities (*e.g.*, doctors, hospitals, or healthcare providers processing patient data for diagnosis, treatment, or administrative purposes; or a hospital maintaining electronic health records or a clinic sharing medical data with specialists to provide care).

- **Public health grounds** (Article 9(2)(i))

Processing is lawful if it is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of healthcare (*e.g.*, Governments or public health Authorities collecting and processing health data to control the spread of infectious diseases, such as during COVID-19 pandemic).

- **Scientific research** (Article 9(2)(j))

Processing is permitted if it is necessary for scientific research purposes, as long as appropriate data minimization measures are in place in accordance with Art. 89 of GDPR, including pseudonymization, and subject to specific EU or Member States' law which must provide for appropriate safeguards for data subjects' fundamental rights and interests (*e.g.*, epidemiological research on disease patterns; longitudinal studies on chronic diseases; genetic research for personalized medicine; clinical trials for new drugs or treatments; AI-based medical diagnostics research).

In this latter specific regard, Recital 33 of the GDPR provides a good margin of flexibility for obtaining consent, when this is deemed the applicable legal ground, in the context of scientific research. It recognizes that at the beginning of a research project (*i.e.*, when the data are collected), it may not always be possible to fully identify the specific purposes for which personal data will be processed in connection with that specific research. In such cases, **Recital 33 endorses the possibility to request a broad consent referred to 'certain areas' of scientific research, provided that the research complies with recognized ethical standards.**

In brief, proving awareness that scientific research may evolve over time, and it might be difficult to specify all future processing purposes at the time of collecting consent, the EU Legislator conveniently allows the data subjects to provide their consent to general areas of research rather than highly specific purposes (for example, if a research study is focused on cancer research, data subjects may give broad consent to participate in related future studies without needing to know the precise details of every possible subsequent study), in keeping with established ethical standards for scientific research (Ethical Review Boards or Institutional Review Boards often play a key role in ensuring that research projects adhere to appropriate ethical standards).

This far-sighting concept of broad – or dynamic, as sometimes defined – consent aims to exempt data controllers from having to get the patients' green light for every novel use of the data, insofar connected to the same area of research, which was not specifically identified at the very outset of the data processing activities. This facilitates continued research while ensuring that participants are informed about the general scope of the study, thus supporting advances in fields like genomics, epidemiology, and chronic disease research.

In plain words, **Recital 33 introduces a mechanism for enabling controllers to lawfully have the data slightly repurposed during research pipelines in a transparent, ethical and controlled manner.**

Regarding the further processing of personal data (also called ‘secondary use’ of data, such as in the EHDS), Article 6(4) of the GDPR lays down the conditions that must be complied with by data controllers intending to process personal **non sensitive** data for purposes other than those described in the information notice (Privacy Policy) given to the data subjects – for which the data was originally collected.

In brief, when the further processing is not based on the data subject’s specific consent or on a EU or Member State law, the controller must evaluate whether the scope for the reuse of the data is compatible with the initial purpose, considering several factors:

- **link between the original and new purpose:** controllers must evaluate how closely related to the original one the new processing purpose is. If the secondary use is significantly different, it is less likely to be considered compatible.
- **Context of data collection:** the context in which the data were initially collected, as well as the relationship between the data subject and the data controller, plays a key role. If data were collected in a setting where individuals would reasonably expect them to be used for additional similar purposes, this supports compatibility. For instance, if data was collected for healthcare purposes, further processing for public health research may be viewed as compatible.
- **Nature of personal data and impact on data subjects:** the sensitivity of the data and potential impact on the rights and freedoms of the data subject must be evaluated. For example, special categories of personal data, such as health data, require a higher standard of justification for compatibility, as misuse or unexpected processing of sensitive data could greatly harm individuals’ privacy.
- **Safeguards in place for original processing:** the data controller must consider the safeguards applied in the original processing (*e.g.*, pseudonymization or encryption) and whether additional safeguards are necessary for the new purpose. Using such safeguards strengthens the case for compatibility by reducing the risk of harm to data subjects.

If the compatibility test fails (meaning the new purpose is not considered compatible), data controllers are required to rely on another legal basis under Article 6(1) to process the data for the new envisaged purpose, including obtaining fresh consent from data subjects, or fulfilling a legal obligation.

Nonetheless, as seen above (in connection with the ‘purpose limitation’ principle), **Art. 5 establishes a crucial presumption of compatibility of the reuse of data for carrying out scientific research activities, provided that appropriate data minimization measures are put in place**, such as pseudonymization or encryption of the data⁶⁹.

* * * * *

The framework outlined above – resulting from the joint application of the presumption of compatibility for the secondary use of data for scientific research (as an exception to the limitation of purpose principle), the opening to the processing of health, medical and genetic data for the same purpose (Article 9(2)(j)), and the endorsement to the concept of ‘broad consent’ in the research setting /Recital 33) – appears highly promising for facilitating and promoting medical and scientific research.

Nonetheless, **Article 9(4) of the GDPR then almost nullifies such flexibility, by allowing Member States to “maintain or introduce further conditions, including limitations”, with regard to the processing of health and genetic data.**

This means that while the GDPR provides a general and favourable framework and exceptions under which special categories of data can be lawfully processed within the EU research ecosystem, individual Member States have the chance to impose stricter national regulations, where they see fit, *de facto*

⁶⁹ Helpful indications are provided, with regard to the ‘compatibility presumption’ established by Art. 5 GDPR, in the ‘A Preliminary Opinion on data protection and scientific research’ adopted by the EDPS on 6 January 2020. https://www.edps.europa.eu/sites/default/files/publication/20-01-06_opinion_research_en.pdf

undermining and limiting the conduct of medical research across the EU in several ways, including:

✓ **Inconsistent regulations across Member States**

Article 9(4) leads to a lack of regulatory harmonization across the EU, making it more difficult for researchers and organizations carrying out multinational studies on the account of the need to comply with varying national laws. Indeed, researchers conducting cross-border clinical trials or medical studies may face challenges in navigating multiple sets of national regulations, which can create administrative burdens and delays.

✓ **Stricter conditions for health data processing**

Many Member States have imposed additional limits on sharing health data. For example, extra approvals or stringent data protection measures, including sometimes obtaining authorization from competent national Supervisory Authorities, may be required before health data can be shared for international research collaborations.

✓ **Limiting access to data for secondary research**

Some Member States have limited the reuse of previously collected health data, *e.g.*, for secondary research purposes or for developing a Digital Twin in health, without obtaining new consent or explicit green light from competent Data Protection Authorities, even when the GDPR provides for exceptions under certain conditions (*e.g.*, scientific research with safeguards like pseudonymization). Retrospective studies and related activities, which rely on existing health data, could be particularly affected by stricter national laws that make it difficult to access or process previously collected data.

✓ **Impact on genetic and biometric data research**

Some Member States have introduced additional restrictions or conditions for processing genetic data, given its highly sensitive nature, thus limiting research opportunities in fields such as genomics, personalized medicine, and pharmacogenomics, where large datasets of genetic information are essential.

✓ **Complexities in public health research**

Medical research and studies related to public health emergencies or pandemics face barriers if national laws impose stricter conditions on data collection and sharing, even when public health authorities have a strong interest in accessing this data for research purposes.

✓ **Potentially limiting scientific collaborations**

The introduction of varying national requirements discourages international collaboration in medical research, as institutions or researchers in Member States with stricter data protection rules might find it difficult to collaborate with researchers in countries with more flexible approaches, due to legal complexities around data sharing and compliance.

4.1.1 **GDPR AND EDITH'S USE CASES**

In consideration of the aforementioned framework governing the protection and circulation of personal data, the fragmentation derived from the uneven national rules disciplining data processing for scientific research can be highlighted in the analysis of practical cases. Specifically, five use cases were taken into account in EDITH, four of which allowed for an in-depth analysis of the GDPR and associated implications on data. This section 4.2 is therefore intended to dive into the criticalities stemming from the application of both the GDPR and national laws to the following sample of VHT use cases.

A. BOLOGNA BIOMECHANICAL COMPUTED TOMOGRAPHY (UNIBO)

Bologna Biomechanical Computed Tomography (**BBCT**) is a Digital Twin in healthcare solution that predicts, starting from a CT scan of the thigh region, the Absolute Risk of Hip Fracture at the time of

the scan (ARF0), and under some assumptions of disease progression, also the risk of conventional time spans (ARF5 or ARF10). In preclinical studies on cadaver preparations, the Digital Twin showed an accuracy in predicting the force required to fracture a proximal femur in side-fall conditions of 85%. In a retrospective clinical validation on a cohort of 100 women, half of which with a hip fracture, ARF0 showed a stratification accuracy of 87%, compared to 75% of the clinical standard of care (areal Bone Mineral Density of the proximal femur region measured with Dual X-ray Absorptiometry, aBMD). In a qualification advice from the EMA, relative to the possible use of BBCT as an in-silico drug development tool for the optimization of the dose-response for new osteoporosis treatments, the team received indication that, to achieve a positive qualification assessment, a significantly larger and more important prospective validation study would be required.

BBCT is a model based on the processing of patients' data currently under development in Italy. Both the GDPR and the Italian Privacy Code (Legislative Decree 30 June 2003, n. 196) apply to the data processing taking place in connection with BBCT, since almost all the data collected are of a personal nature. More specifically, since the model is based on the recognition and analysis of osteoporosis conditions from CT scans, it is safe to assume that most data can be considered “data related to health” as defined by Art. 4, no. 15) of GDPR. Therefore, a legal basis for processing must be found under both Articles 6 and 9 of GDPR.

As seen above, the conditions for the lawful processing of health data can be considered either consent under Art. 9.2, a), or scientific research under Art. 9. 2, j), of GDPR.

Should consent be considered the preferred legal ground, then the requirements set out by Art. 7 of GDPR should be fulfilled before collecting data. In other words, the patients should be previously and specifically informed, by means of adequate privacy notice, about (among additional information to be provided under Art. 13 of GDPR) the envisaged scientific research purpose, and they should give free, explicit, unambiguous consent about the processing of their personal and health data for this specific objective. In this respect, it should be noted that the conditions for the validity of individual consent are subject to various interpretations by the national Data Protection Authorities. Considering that BBCT is a model currently being developed in Italy, consent for scientific research purposes should be asked considering the interpretation given by the Italian Data Protection Authority (‘**Garante**’). Specifically, this latter Authority has regularly defined a consent as valid under the GDPR, only when specific to a single identified purpose. In other words, the Italian DPA tends to reject the view of broad consent given by the data subject considering non-specified purposes of “scientific research”. Such a stringent interpretation implies that, during the development of the use case, it might be necessary to ask the patients to re-consent every time the purpose for data processing varies from the original one (as described in the Privacy notice initially given to the data subjects).

As an alternative to consent, Art. 9.2, lett. j) of GDPR allows for the processing when it is “*necessary for [...] scientific [...] research purposes [...] in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject*”. Art. 110-bis of the Italian Privacy Code states that the Italian Garante can authorize further data processing – namely, data reuse – for scientific research purposes when informing the individuals proves impossible or would prejudice the scope of scientific research. Moreover, it empowers the Italian DPA to define the conditions and measures that the data controller must implement to protect the data subjects when re-purposing the processing of their personal data, including to ensure appropriate security⁷⁰.

In terms of privacy-by-design and by-default, a Data Protection Impact Assessment should be conducted and continuously updated, in order to prevent all reasonably foreseeable risks, including data breaches. Moreover, considering the principle of accountability, a privacy-oriented approach in the development of BBCT should be adopted. In this respect, pseudonymization has already been implemented as a security and data minimization measure. Another step in this direction could involve

⁷⁰ For the sake of completeness, it should be noted, however, that data processing conducted by Scientific Hospitalization and Treatment Institutions (IRCCS – *Istituti di Ricovero e Cura a Carattere Scientifico*) does not fall in the definition of “further processing” under Art. 110-bis of the Italian Privacy Code.

assessing whether non-personal data might be processed in specific cases or evaluating the risk of re-identification when pseudonymized data is combined.

B. **ATRIAL MODELLING TOOL KIT (QMUL)**

Atrial Modelling Tool Kit ('**Atrialmtk**') is an in-silico trial solution that enables personalized investigation of the effects of fibres and fibrosis on fibrillatory dynamics in the atria. The computer model takes segmentation masks of the atria, which can be produced from raw MRI or CT data by an expert, or input directly in those cases in which the user already has them, or otherwise existing atrial surface meshes can be taken as input. After the identification of some key anatomical landmarks, the model automatically produces a simulation grade atrial mesh that incorporates transmatability, atrial regions and atrial fibres, which are input into an electrophysiological simulation. Atrial fibrillation, characterized by irregular activation of the heart's top two chambers, can be treated with ablation to target tissue sustaining the arrhythmia. However, since only one ablation strategy can be applied to each patient, determining the optimal strategy that treats the arrhythmia while minimizing tissue ablation remains challenging. Atrialmtk successfully predicted long-term atrial fibrillation recurrence in individual patients following ablation therapy by combining outcome data with patient-specific simulation responses. Importantly, classifiers trained on a combination of clinical data and simulation data outperformed those trained on clinical data alone.

Atrialmtk is currently deployed exclusively for scientific research, as it is used only for the improvement of the models itself and not for healthcare purposes.

Since the case is based on health data related to patients in the UK, both the Data Protection Act 2018 and the UK GDPR must apply. Following the collection by the clinical team, the data undergo a pseudonymization process by the hospital before being transferred to the research team. The hospital team maintains a linked list of data, allowing at any moment the singling out of each patient, whenever needed. Atrialmtk is based on patients' consent, serving as a legal ground for the processing under Art. 9.2, lett. a) of UK GDPR as long as all requirements set out in Art. 7 of UK GDPR are properly complied with. In any case, patients must be informed about the data processing both before granting (or denying) their consent, by means of an appropriate privacy notice, and afterwards: this is why they must be put in condition of asking for information about the processing and exercising their rights, including in particular (but not only) withdrawing their consent, without any condition and at any moment.

The competent data controller must adopt all the organizational and technical measures necessary to avoid risks and data breaches. In this sense, the pseudonymization applied to the data constitutes a measure capable of guaranteeing an adequate level of security; however, given that special categories of data are at stake, additional measures (*e.g.*, strong authentication, segregation of duties, written authorization to the personnel allowed to access the data on a strictly-needed-basis) must be put in place to safeguard the patients. A Data Protection Impact Assessment (DPIA) needs to be carried out by the controller, because health pseudonymized – and so identifiable – data are processed under this use case on a large scale⁷¹.

C. **CANCER DIAGNOSIS BASED ON OMICS INFORMATION (BSC)**

'Cancer diagnosis based on omics information' is the use case developed under the PerMedCoE project⁷², intended to demonstrate the ability of software to identify appropriate cancer treatments based on a patient's clinical information, omics data and personalized cell level models. This use case relies on a publicly accessible dataset, including anonymized somatic mutation data, copy number alterations, DNA methylation and expression data, generated from several hundred patients with Chronic Lymphocytic Leukemia at a Spanish hospital. The processing of data is supported by an HPC centre in Spain.

⁷¹ More documents about the data management workflow of this use case are provided here:

https://www.cemrg.co.uk/models_files/Cardiac%20Modelling%20Data%20Management%20Workflow%20and%20Data%20Types_V1.0.0_21-12-21.pdf.

⁷² HPC/Exascale Centre of Excellence for Personalised Medicine in Europe. <https://permedcoe.eu/>.

The use case at hand is based on CT scans and genetic information, both qualifying as health data under the GDPR, so falling into special categories according to relevant Art. 9. In this scenario, the applicable legal ground for processing the data could be found either in the patients' consent or in the controller's need to process such data in order to carry out scientific research purposes (Art. 9.2, lett. j of GDPR). In this latter case, the specific conditions for the processing are established by the Spanish data protection law (*Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales*). Specifically, the 17^o additional disposition (*Disposición adicional decimoséptima*), entitled “*Tratamientos de datos de salud*”, establishes that when the individuals' consent was obtained for a specific purpose, the reuse of the data for health and biomedical research purposes shall be considered lawful and compatible, insofar as the data are used for purposes or areas of research related to the area in which the initial study was scientifically integrated⁷³. It is worth pointing out that, whenever obtaining the data subject's consent proves impossible or would otherwise result in a disproportionate effort by the controller, the legal basis for reusing the data can be found in Spanish Law 14/2007 (*‘Investigación biomédica’*), which allows for scientific research in case of positive approval by the Ethical Committee.

According to the principle of accountability, the hospital, in its capacity as data controller, is required to demonstrate compliance with the principles laid down in Art. 5.1 of GDPR. In particular, the processing of data must not only rely on a legitimate basis but also adhere to principles of transparency and fairness. This means that the patients must be informed about how, when, why and which data will be processed. Additionally, only the minimum amount of personal data necessary to achieve the defined and specified purpose should be collected and processed (data minimization and purpose limitation). Finally, the data must be accurate and precise enough to ensure the validity of scientific research outcomes while safeguarding the rights of the data subjects.

In view of the phase of development of the use case, it is assumed that the hospital involved will act as data controller, as it determines the purposes and the modalities of the processing of the patients' data for the primary reason of healthcare and the secondary use for scientific research. In this role, the hospital needs to rely on legal ground under both Articles 6 and 9.2 of GDPR and ensure utmost transparency towards the data subjects about the processing of their data, according to Art. 13 of GDPR. Should the controller opt to leverage the data subjects' consent, then it will have to guarantee that it is informed, specific, explicit and free. After completing these steps, the hospital will be in condition to start sequencing the patients' biopsies and transforming them into pseudonymized data. In case the purposes of scientific research should change during the execution of the research, a new legal basis will have to be identified (*e.g.*, consent needs to be re-obtained). With respect to the principle of processing limitation, the retention period of the personal data should be no longer than needed to achieve the intended research proposes.

Once pseudonymized, the data are transferred to a HPC centre, where they are stored in encrypted form only when they undergo further elaboration. This implies that some processing operations are delegated by the data controller, to the HPC centre which will thus take the role of data processor on behalf of the hospital. Accordingly, an agreement will have to be entered into between the controller and the processor under Art. 28 of GDPR, to set out the subject matter, duration, nature and purpose of the processing, the type of personal data, the categories of data subjects and the obligations and rights of the controller.

Once the results are generated by the HCP centre, they will be communicated to the hospital. Different requirements will apply depending on the pseudonymous or anonymous nature of such outputs.

D. *POD (VITO)*

POD is a tool that can be used for analysing patients' data in a distributed but secure manner. The foundational technology behind this demo is Solid (Social linked data), a solution designed to securely store data in decentralized data containers called pods, while still enabling access for valid use cases.

⁷³ “Se considerará lícita y compatible la reutilización de datos personales con fines de investigación en materia de salud y biomédica cuando, habiéndose obtenido el consentimiento para una finalidad concreta, se utilicen los datos para finalidades o áreas de investigación relacionadas con el área en la que se integrase científicamente el estudio inicial”.

Patients can store and process personal and medical information in their personal pods and share the data via a consent system. Processed data and analysis results are then written back to the pod.

Data in Solid pods are stored as linked data according to the Resource Description Framework (RDF), a standard model for data interchange on the web, enabling data to be linked and queried across different sources⁷⁴. RDF represents information using subject-predicate-object triples, making it flexible and powerful for representing complex data relationships. It enables interoperability, scalability and allows data to be linked to other data sources. Access and access grants to the pods follow the Inrupt documentation⁷⁵. Agents are identified using webIds (*e.g.*, Agent A identified by webId <http://agentA/webId> and Agent B identified by webId <http://agentB/webId>). If Agent B wants to access a resource in Agent A's pod (*i.e.*, <http://agentA/pod/231467543>), Agent A must approve the request. An access request is sent to Agent A, and a verifiable credential (formatted as a JSON file) is created. This request is reviewed by Agent A. If approved, another verifiable credential is generated, allowing Agent B access to the resource (<http://agentA/pod/231467543>) in Agent A's pod. Authentication of users/owners of a set of linked data resources follow the Solid-OIDC principles⁷⁶.

In this case, a patient or citizen either already has a Solid pod or is prompted to create one. Through the pod's online interface, the individuals are asked to complete a brief medical questionnaire. After completing it, they are explicitly asked for consent to share the information with a data analysis service, such as a public healthcare system. The data analysis service will review the questionnaires it has access to and identify individuals eligible for a Virtual Human Twin analysis. Consider, for example, the case of the BBCT, which assesses the risk of hip fracture based on a CT scan of the pelvic area. If the individual has an appropriate CT scan and meets the criteria for the VHT, they are prompted to share their medical data, clearly specifying what data will be shared, with whom, for what purpose, and for how long. The individual can revoke this access at any time via a private history of their data access grants. Once the analysis is complete, the data analyser (*e.g.*, public hospital, doctor, or medical software company) is authorized to write the results to a designated location in the individual's pod.

The POD is therefore a case study aimed at securely storing patients' health data for purposes of scientific research. The use case in question concerns the processing of personal data of European patients/citizens, which is why the GDPR must apply. The personal data being processed in the POD case can be considered of two categories: "ordinary" personal data (*e.g.*, the name, email address etc.) and data related to health. In this specific case, a number of data processing operations come into play (*e.g.*, the processing of the credentials used for granting access to the data subject; the collection and analysis the data collected through the medical questionnaire; the storage of the data; the sharing of the data; the processing of data related to history logging), some of which will need to rely of the individual's consent, while other can be based on different lawfulness conditions under Articles 6 and 9 of GDPR.

In case the data subject's consent is required, it will have to be informed, specific, freely given and unambiguous, as well as withdrawable at any time and without any hindrance by the controller. On the other hand, should the processing be grounded on Art. 9.2, lett. j) of GDPR, namely the need for the controller to process the data to carry out scientific research, the processing will need to be carried out based on a Union or a Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subjects. Consequently, in the absence of such a law, the processing cannot be based on lett. j) of Art. 9.2 of GDPR. Moreover, if personal data from individuals living in different Member States will be collected and processed in the context of this use case, the data controller will also have to comply with national applicable legislation.

As a decentralized system, the PODs ensure data minimization avoiding the risks which are typically connected with central servers or cloud storage. Additionally, the strong authentication system prevents unauthorized access and major risks of data breaches by third parties. In terms of security, the data

⁷⁴ More detailed information: <https://www.w3.org/RDF/>.

⁷⁵ <https://docs.inrupt.com/>

⁷⁶ Specifications: <https://solidproject.org/TR/oidc>.

undergo a pseudonymization process being shared with the entity entrusted with the data analysis service.

4.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

Considering the complex and, in some cases, inconsistent regulatory framework explored in Sections 3 and 4.1 above, several barriers emerge which need to be properly addressed to create homogeneous, competitive and scalable legal conditions for using and re-using both personal and sensitive data to nurture the EU VHT ecosystem.

Finally translating into various (inter)dependencies conditioning the implementation of some key EU objectives and goals, including in particular the European Health Data Space, these hindrances are explored below.

A) Update and standardize anonymization procedures

Barrier

Divergent interpretations by competent national Data Protection Authorities as to what constitutes anonymous data are no longer acceptable, **even from a legal competition and antitrust perspective**, if implementing the European strategy for data, together with all the prominent regulations which underpin this initiative, is to become a concretely achievable goal. This is consideration clearly reported, *inter alia*, by TEHDAS⁷⁷.

Subject to the obligation for all data controllers (included, in particular, all those operating in the sensitive data-driven VHT ecosystem) to assess on a case-by-case basis what specific circumstances and contextual variables may affect the level of risk of reidentification of data subjects (and evaluate, on this basis, what technical, legal and operational measures are most suitable for minimizing those risks), **a EU-wide consensus needs to be reached on standard anonymization procedures**.

To this end, the full potential of Privacy Enhancing Technology is to be unleashed, **defining novel regulatory paths which leverage by-default these tools and methodologies to enforce data minimization and security in relation with the secondary use of personal and health data, especially in the scientific research environment**.

Guidelines must be adopted that predefine the conditions, taking into account the main use-cases emerging from the market and from the almost 30-years of application experience of data protection legislation, both technical and legal, whose fulfilment allows the data controller, still under its own responsibility (namely, accountability), to consider that a certain dataset which has undergone a specific de-identification processing can safely be considered anonymous. **The margin of uncertainty, which significantly hampers the health and research sectors, must be minimized by reinforcing the principle of accountability and dispelling the persistent fear of regulatory sanctions whenever data anonymization is addressed**.

Informed by key market use cases and nearly 30 years of data protection experience, appropriate guidelines must be developed to predefine the technical and legal conditions, that – when met – allow the data controllers, under their accountability, to reliably consider a dataset as anonymous following specific de-identification processing. Furthermore, legitimate interest should be enhanced as the sole applicable legal ground when data are processed with the exclusive purpose of anonymizing them, in order to prevent any misinterpretation or national misalignment on this crucial aspect⁷⁸.

⁷⁷ A resounding case, in this specific respect, is the Italian Guarante's decision on the do-called 'THIN' research project, in which the same advanced anonymization techniques were deemed suitable in many member States, but not in Italy. The decision is currently under appeal. A brief description of the provision by the Italian DPA is available here: https://www.ipinitalia.com/data-protection/did-i-really-anonymize-that-data-the-controversial-italian-decision-on-the-thin-case/#_ftn3.

⁷⁸ This is already established in the 'Opinion 05/2014 on anonymization techniques', whereby the WP29 states that anonymization: «*as an instance of further processing of personal data, can be considered to be compatible with the original purposes of the processing but only on condition the anonymisation process is such as to reliably produce anonymised information (...)*».

Proposed action

- **By the European Data Protection Board:** refer to and incorporate, as fully applicable, globally recognized technical standards (such as in particular ISO/IEC 27559:2022 and ISO/IEC 20889:2018) in the new (i) ‘Guidelines on anonymization’ and (ii) ‘Guidelines on pseudonymization’ that are in the process of being defined (and which are due to replace the outdated ‘Opinion 05/2014 on anonymization techniques’), possibly involving ENISA.
- **By the European Data Protection Board (jointly with the EHDS Board):** describe in detail, in the abovementioned set of guidelines, how some widely-tested PETs like Secure Multi-Party Computation and Homomorphic Encryption may help both data controllers and processors (acting on the latter’s behalf) to implement data minimization and security, reinforcing accountability, when personal and especially health and genetic data undergo a secondary processing for any purpose permitted by Art. 34 EHDS, including in connection to any stage of a DT lifecycle.
- **By the European Data Protection Board (together with the EHDS Board and aligning with the European Data Innovation Board):** set out a guidance to help stakeholders in the healthcare and medical areas to fully leverage the great potential of synthetic data, particularly when paired with additional techniques for minimizing reidentification risk, such as Differential Privacy above all. This further guideline could be incorporated into the guidelines mentioned above or, preferably, addressed in a separate provision that takes into account all the technical specificities of synthetic data.
- **By the EU Commission (in coordination with all competent Boards):** better monitor to ensure that legislations adopted by Member States in relation to the primary and secondary processing of health data do not undermine or in any manner prevent the European objectives and public interest concerning the secure circulation and especially the reuse of health data for permitted predetermined purposes.

B) Shift the approach from absolute to relative anonymity**Barrier**

The WP29 Opinion 05/2014 sets an exceptionally high threshold for anonymization, stating that data are only considered anonymous if irreversibly modified so that no individual can be directly or indirectly re-identified by anyone and by using any available means or information. In other words, as explicitly stated by WP29, the results of the anonymization procedure must eventually be the same that would derive from the cancellation of the data.

This absolute approach limits data usability, cross-border research and innovation, as organizations struggle to meet such rigid anonymization criteria, incurring higher costs and facing technical challenges, along with the risk of being sanctioned by supervisory Authorities.

Proposed action

- **By the European Data Protection Board:** promote in the above-mentioned guidelines, as well as in the ‘*Guidelines on the processing of data for scientific research purposes*’ which were further announced in the EDPB ‘Work Programme 2024/2025’, a more reasonable and contextual approach to distinguish pseudonymous and anonymous data, building on the commendable principles established by the ECG in the Judgement explored in Section 3.4.1 above (Case T-557/20). **An urgent shift is needed from a so-called ‘absolute anonymity’ approach, where data is considered anonymous only if re-identification is completely impossible for anyone and under any circumstances, to a more context-specific ‘relative anonymity’ perspective, where data can be deemed anonymous as long as re-identification is reasonably unlikely by a specific party holding the data, given the resources and additional information available**

solely to that party. In other words, the same dataset may be pseudonymous for one party while being anonymous for another. The EMA should be involved, or at least consulted, for drafting such guidance affecting the scientific and medical research sectors.

C) Enhance a ‘broad consent’ applicable to the research sector

Barrier

Broad consent is not widely used for scientific research due to legal uncertainties under the applicable data protection legal framework. Although Recital 33 provides some flexibility, the GDPR’s strict requirements for specific, informed, and unambiguous consent prove challenging, as broad consent covers general areas of research rather than precise uses. This leads to concerns about data subjects’ rights and regulatory compliance, making researchers hesitant to rely on it. Further complications arise – also in this case – from variations in national regulations across EU Member States, causing a fragmented legal environment, especially for cross-border research. Additionally, Ethical Review Boards often require specific consent, further discouraging the use of broad consent.

Proposed action

- **By the European Data Protection Board:** lay down, in the forthcoming ‘*Guidelines on the processing of data for scientific research purposes*’, precise indications to define the scope and acceptable use of broad consent under the GDPR. The crucial chance offered by Recital 33 must be adequately exploited, building on the experience of those Member States where broad consent is already admitted and a building block of the national research environments (*e.g.*, Germany, Finland, Austria, and the Netherlands, that have legislative frameworks or guidance that permit the use of broad consent, particularly in the context of medical research and biobanks). Fragmentation at local level in this respect should be avoided and prevented.
- **By the European Medicines Agency:** define guidance, in cooperation with, *e.g.*, the EDPB, the European Group on Ethics in Science and New Technologies (‘EGE’), and the European Network of Research Ethics Committees (‘EUREC’), setting out **appropriate ethical safeguards uniformly applicable all across the EU**, such as regular reviews by ethical boards and **requiring researchers to provide periodic updates to participants on how their data are being used.** This would help build trust and transparency, ensuring patients feel informed and empowered. Transparency might be ensured, by way of example, developing digital tools or online platforms where participants can easily access information about how their data are being used, change their consent preferences (through not only broad, but also ‘dynamic consent’), or withdraw consent if desired, even only in relation to specific areas of research.

5 European Health Data Space

5.1 THE FRAMEWORK FOR VHT

The (proposal for a) European Health Data Space Regulation has crucial implications for the VHT ecosystem⁷⁹.

The EHDS is the first of the sector-specific data spaces proposed by the EU Commission in 2020, in its communication ‘A European strategy for data’⁸⁰.

This legislation establishes common rules, standards, infrastructures and a governance framework with a view to facilitating access to electronic health data for the purposes of primary and secondary use of these data.

The main goals of the EHDS are:

- empowering individuals through better digital access to their personal health data;
- supporting free data circulation, by ensuring that health information (either personal or non-personal) follow people;
- set up strict rules for the use of anonymized, or pseudonymized health data for research, innovation and policy-making.

The EHDS is a structured environment where health-related data can be securely stored, accessed and shared among authorized stakeholders. Practically, this means that HCPs across the EU can access a patient’s medical history to provide better care and that medical researchers can access and pool data from various sources to accelerate the discovery of groundbreaking treatments and innovations, also based on AI.

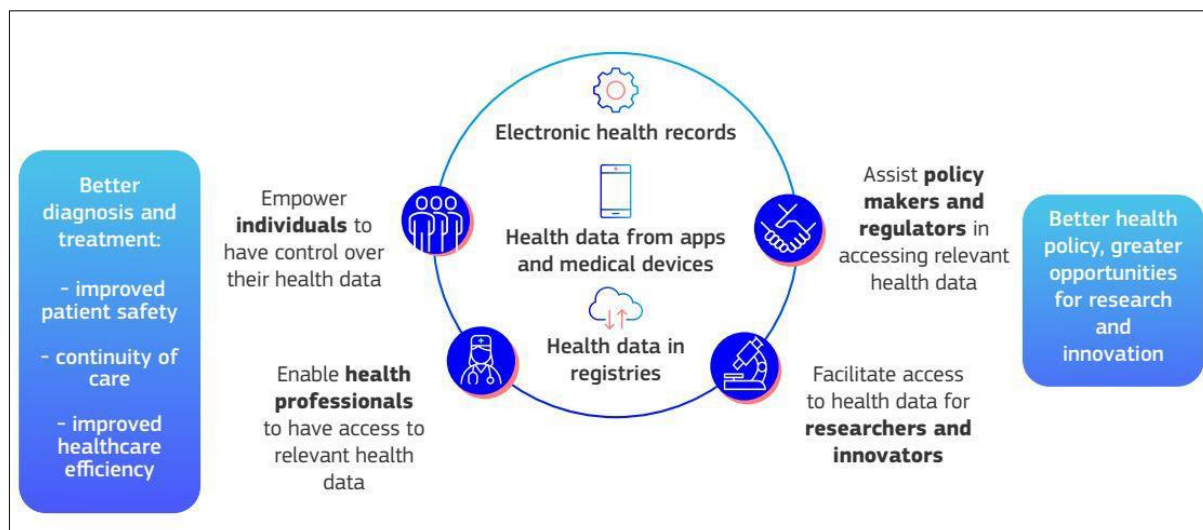


Figure 13 - Source: EHDS Fact Sheet⁸¹

Focusing on the VHT, the main fundamental benefits deriving from the EHDS may be:

- **Improved data accessibility and integration**

The EHDS will create a unified, interoperable data environment across EU Member States, allowing Digital Twins to access and integrate health data from various sources such as

⁷⁹ In spring 2024, the European Parliament and the Council reached a political agreement on the Commission proposal for the EHDS.

⁸⁰ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>

⁸¹ https://ec.europa.eu/commission/presscorner/detail/en/fs_22_2713

electronic health records, clinical trial databases, biobanks, medical devices, wellness applications, public health registries, registries for medicinal products, research cohorts, and others. This level of data integration is essential for building accurate and comprehensive Digital Twin models that require diverse, real-world patient data for effective simulations and predictions.

- **Enhanced data interoperability**

By standardizing data formats and promoting the European Electronic Health Record Exchange Format (EEHRxF), the EHDS will make health data more compatible across systems. This interoperability allows Digital Twins to utilize a consistent data structure across borders, making it easier to scale and refine models for wider populations, thus improving the applicability and accuracy of the VHT in clinical trials and personalized medicine.

- **Facilitating secondary data use for research and innovation**

Crucially, one of the EHDS's goals is to support secondary use of health data for research and innovation, which includes creating simulations and predictions with Digital Twins. By providing regulated, secure access to de-identified data, the EHDS enables Digital Twins to safely use patient information for developing, testing, and validating models without compromising privacy. This access accelerates innovation and the application of Digital Twins in predictive health, personalized treatment planning, and clinical trials.

- **Supporting cross-border health research**

By enabling cross-border health data access, the EHDS allows Digital Twins to leverage data from diverse populations across Europe. This improves the scalability and robustness of DT models, which can incorporate variations in genetics, lifestyle, and health conditions unique to different regions, enhancing the relevance of Digital Twins in clinical research and personalized care solutions.

- **Accelerating innovation in clinical trials and personalized medicine**

With greater access to interoperable, secure health data, Digital Twins can be more effectively used in clinical trial optimization, virtual control arms, and adaptive trial designs. These applications help in refining treatment protocols, reducing trial costs, and improving patient safety, all of which align with the goals of the Clinical Trial Regulation for advancing innovation in clinical research.

It must be highlighted that 'Electronic health data' to which the EHDS's applies refer to both:

- **personal data:** data concerning health and genetic data, as defined in the GDPR, processed in an electronic form; and
- **non-personal data:** data that have been anonymized, so that they no longer relate to an identified or identifiable individual, and data that have never been related to a data subject.

Almost all actors of the health environment are encompassed within the EHDS scope:

- **Individuals** (patients and citizens): who will gain rights to access, control, and share their health data across borders within the EU, enhancing their ability to manage their electronic health data securely and use them for personalized healthcare.
- **Health services suppliers and health products manufacturers:** companies that create digital health solutions, like Electronic Health Record (HER) systems, wellness Apps and medical devices, can leverage the EHDS's standardized and interoperable framework, which facilitates the development and deployment of digital health tools across the EU single market.
- **Healthcare providers and institutions:** the EHDS facilitates the exchange of patient data among healthcare providers, enabling them to access health records of patients from other Member States to support cross-border healthcare and continuity of care.

- Researchers and innovators: researchers and companies in the healthcare and pharmaceutical sectors can access health data for secondary purposes, such as clinical research, policy development, and innovation. This access is regulated to ensure privacy and data protection, supporting advancements in personalized medicine and public health research.
- Public health authorities and policy makers: the Regulation enables public health authorities to access aggregated health data for health policy-making and resource allocation, allowing them to use evidence-based data to respond to public health needs and improve healthcare services.

The health data holders, under the EHDS, are all natural or legal persons, public authorities, agencies or other bodies operating in the healthcare or care sectors (including reimbursement services), or developing products or services intended for the health or care sectors, or developing or manufacturing wellness applications, or performing research in relation to the healthcare or care sectors, or which is a Union institution, body, office or agency, whenever such natural or legal person has:

- in its capacity as a data controller (or joint controller) under the GDPR, the right or obligation, in accordance with applicable law, to process personal electronic health data for the provision of healthcare or care, or for public health, reimbursement, research, innovation, policy making, official statistics, patient safety or regulatory purposes; or
- the ability to make available, including to register, provide, restrict access or exchange non-personal electronic health data, through control of the technical design of a product and related services.

Therefore, data holders shall act as controllers in relation to electronic health data of personal nature, meaning that the EHDS will not apply to data holders that collect and process such data on behalf of others, in the quality of data processors. By way of example hospitals, as data controllers, are data holders of their EHRs. Similarly, pharma companies are data holders of their clinical trial data and biobanks. Medical device manufacturers may be data holders of non-personal data generated by their devices, if they have access to those data and the ability to generate them, while they would not qualify as data holders in cases where they merely process personal electronic health data on behalf of a hospital.

To complete the list of parties involved in data exchange under the EHDS, it is also important to mention:

- ‘Data user’, *i.e.*, any natural or legal person, including EU institutions, bodies or agencies, that is granted lawful access (by a health data access body) to electronic health data for secondary use pursuant to a data permit, a data request, or an access approval by an authorized participant in HealthData@EU (the cross-border infrastructure for secondary use of electronic health data);
- ‘Health data access body’: Member States’ national authorities that will control the use of and access to EHD in their national Secure Processing Environments (*e.g.*, Findata, French Data Hub, DaTraV – German Research Data Centre, Healthdata.be, Norwegian Health Data Program, Danish Health Data Authority). Their remit is mainly to issue permits, manage access time limits and address data breaches.

In its proposal for the EHDS, regarding primary use, the EU Commission stressed that patients cannot yet fully exercise their right to access and control their health data. Additionally, the design and deployment of new digital health products and services is hindered by the fragmentation of standards and specifications, within and across Member States. Interoperability of health information systems comprises complementary components: legal (the same rules must apply to all the parties involved), organisational (similar policy and care processes), semantic (similar way of coding the data to be processed) and technical interoperability (for applications and information technology infrastructure)⁸².

⁸² More detailed info are available here:

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733646/EPRS_BRI\(2022\)733646_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733646/EPRS_BRI(2022)733646_EN.pdf).

To overcome this issue, the EHDS specifies a list of “priority categories of Electronic Health Data” that patients must always be able to access and exercise their rights over, and to which healthcare professionals must have unrestricted access for primary use, free from any territorial or technical limitations. Such priority categories include:

- **patient summaries:** *i.e.*, personal details, contact information, information on insurance, allergies, medical alerts, vaccination/prophylaxis information, possibly in the form of a vaccination card, current, resolved, closed or inactive problems, textual information related to medical history, medical devices and implants, procedures, functional status, current and relevant past medicines, social history observations related to health, pregnancy history, patient provided data, observation results pertaining to the health condition, plan of care, information on a rare disease such as details about the impact or characteristics of the disease;
- **electronic prescriptions;**
- **electronic dispensations:** information on the supply of a medicinal product to a natural person by a pharmacy based on an electronic prescription;
- **medical images and image report:** electronic health data related to the use of or produced by technologies that are used to view the human body in order to prevent, diagnose, monitor, or treat medical conditions;
- **laboratory results:** Electronic health data representing results of studies performed notably through in vitro diagnostics such as clinical biochemistry, hematology, transfusion medicine, microbiology, immunology, and others, and including, where relevant, reports supporting the interpretation of the results;
- **discharge reports:** Electronic health data related to a healthcare encounter or episode of care and including essential information about admission, treatment and discharge of a natural person.

The EU Commission is required to lay down the technical specifications for the priority categories of personal EHD, **setting out the European Electronic Health Record Exchange Format (EEHRxF), which must be commonly used, machine-readable and allow transmission between different software applications, devices and healthcare providers, of both structured and unstructured health data**⁸³. The EEHRxF must include the following elements:

- harmonized datasets containing EHD and defining structures, such as data fields and data groups for the representation of clinical content and other parts of the EHD;
- coding systems and values to be used in datasets containing electronic health data;
- technical interoperability specifications for the exchange of electronic health data, including its content representation, standards and profiles.

Member States must ensure that appropriate digital services are established and made available: (i) to individuals (notably patients), or their authorized representatives, enabling them to access to their EHD free of charge and exercise their other rights under the EHDS, and (ii) to health professionals, at least with reference to the priority categories of EHD, for the provision of healthcare, including at cross-border level (through MyHealth@EU).

MyHealth@EU is a platform for improving the interoperability, accessibility, and quality of cross-border healthcare in the European Union, ensuring that individuals can receive appropriate medical care and that their health data are securely exchanged and protected throughout their journey across member States. Each member State must designate one national contact point for digital health, as an organizational and technical gateway for the provision of services linked to the cross-border exchange of personal electronic health data in the context of primary use. The local contact point will have to

⁸³ The already mentioned EU Commission’s ‘Recommendation on a European Electronic Health Record exchange format’ builds on the work done by the eHealth Network, which has so far issued a number of technical guidance (*e.g.*, on HL7 Clinical Document Architecture (CDA) Release 2, HL7 FHIR and DICOM for medical imaging).

connect to all other national contact points for digital health and to the cross-border infrastructure MyHealth@EU.

Member States may use this platform to provide supplementary services that facilitate telemedicine, mobile health, access to translated health data, exchange or verification of health-related certificates, including vaccination card services supporting public health and public health monitoring or digital health systems, services and interoperable applications, to achieve a high level of trust and security, enhance continuity of care and ensure access to safe and high-quality healthcare.

Also EHR systems lie at the heart of the EHDS, as they are the central technical prerequisite for fulfilling the objective of ensuring secure and smooth processing and cross-border transfer of health data. An EHR system under the EHDS must consist of two core elements which form an integral part of the software (together, the ‘harmonization components’):

- ✓ the **interoperability component**: EHR systems must have the ability to interact with software applications and devices from the same or different manufacturers, thanks to the EEHDxF, in order to transfer and receive personal EHD;
- ✓ the **logging component**: EHR systems must be able to record logging information about access to personal EHD by users of the system. As a minimum standard, such information must include, for each time the data is accessed: (i) identification of the health provider or other individuals having accessed personal electronic health data, (ii) categories of data accessed, (iii) time and date of access, (iv) data source.

In this respect, both manufacturers of medical devices and in vitro diagnostic medical devices and providers of high-risk AI systems as defined in the AI Act, which do not qualify as medical devices, must comply with the essential requirements set out by the EHDS for harmonized components, when claiming interoperability of those medical devices or high risk AI (HRAI) systems with such components of EHR systems.

Patients are vested with several rights, building on those granted under the GDPR, concerning their personal electronic health data falling in the scope of primary use:

- right to access: patients can always access their personal health data, by means of specific digital access services to be built up at national level, especially data in designated ‘priority categories’ essential for healthcare provision. This access includes the ability to view their data through secure electronic health data access services and to download an electronic copy free of charge.
- right to insert data: patients can add relevant health information to their own electronic health record through designated electronic health data access services or linked applications, allowing for more comprehensive and personalized healthcare management.
- right to request rectification: patients have the right to request corrections to their personal health data if they identify errors or inaccuracies, ensuring that their health records are correct and up-to-date.
- right to data portability: patients can request that their personal electronic health data be transferred to another healthcare provider of their choice. This right applies even across borders within the EU, where the data must be shared in a standardized format through the MyHealth@EU platform, facilitating seamless healthcare transitions.
- right to restrict access: patients are entitled to limit the access of healthcare professionals and providers to all or part of their EHD. However, patients should be informed that restricting access may affect the quality or scope of healthcare services provided to them.
- right to information on data access: patients can obtain information on who has accessed their personal health data, ensuring utmost transparency. They may also receive automatic notifications about any access to their EHD by healthcare professionals, keeping them informed of how their data is used in healthcare.

Last, but not least, Member States may establish by national laws that patients have also the right to opt-out from digital access to their personal EHD by health professionals. It remains unclear whether exercising this right also restricts access across Europe. The rule's wording suggests it does not, but this would lead to a contradictory outcome, as patients who opt-out would, by default, likely want their data access restricted in other Member States as well.

One of the primary challenges the EU Legislator aims to tackle through the EHDS – and the most relevant for the VHT ecosystem – is promoting health data reuse by addressing the national fragmentation stemming from Article 9.4 of the GDPR. Therefore, focusing on the secondary use of electronic health data, it must first and foremost be pointed out that access to specific categories of data (enumerated below) must only be granted by health data access bodies to health data users where the processing by the latter is necessary for one of the following specific purposes, many of which may be highly relevant for the VHT:

- i. limited to public sector bodies and Union institutions, bodies, offices and agencies exercising their public tasks (including where the data are processed for carrying out these tasks by a third party on behalf of that public sector body or of Union institutions, agencies and bodies):
 - a) public interest in the area of public and occupational health, such as activities for protection against serious cross-border threats to health and public health surveillance or activities ensuring high levels of quality and safety of healthcare, including patient safety, and of medicinal products or medical devices;
 - b) policy making and regulatory activities, to support public sector bodies or Union institutions, agencies and bodies, including regulatory authorities, in the health or care sector to carry out their public tasks;
 - c) statistics, such as national, multi-national and Union level official statistics related to health or care sectors;
- ii. for data users operating both in the public and private sectors:
 - d) education or teaching activities in health or care areas;
 - e) **scientific research related to health or care sectors, contributing to public health or health technology assessment, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices, with the aim of benefitting the end-users, such as patients, health professionals and health administrators, including:**
 - **development and innovation activities for products or services;**
 - **training, testing and evaluating of algorithms, including in medical devices, in-vitro diagnostic medical devices, AI systems and digital health applications.**
 - f) **improving delivery of care, treatment optimization and providing healthcare, based on the electronic health data of other natural persons.**

The categories of electronic health data that health data holders are required to make available for reuse (or 'further processing', as named in the GDPR) in connection with one of the purposes described above are:

- EHRs;
- human genetic, epigenomic and genomic data;
- other human molecular data such as proteomic, transcriptomic, metabolomic, lipidomic and other omics data;
- electronic health data automatically generated by medical devices and wellness applications;
- data from clinical trials, clinical studies and clinical investigations;

- population-based health data registries (public health registries);
- data from medical registries and mortality registries;
- other health data from medical devices;
- data from registries for medicinal products and medical devices;
- data from research cohorts, questionnaires and surveys related to health;
- health data from biobanks and associated databases.

In addition to the above, also the following data must be shared by the health data holders:

- data on professional status, specialization and institution of health professionals involved in the treatment of a natural person;
- data on factors impacting on health, including socio-economic, environmental and behavioural determinants of health;
- aggregated data on healthcare needs, resources allocated to healthcare, the provision of and access to healthcare, healthcare expenditure and financing;
- pathogen data, impacting on human health;
- healthcare-related administrative data, including dispensation, claims and reimbursement data.

Given the wide-ranging purposes for secondary use and the extensive pool of electronic health data available, it is clear how Digital Twins (DT) and the Virtual Health Twin (VHT) infrastructure stand to gain numerous benefits from the EHDS, such as

✓ **Enhanced precision and personalization**

Leveraging comprehensive data from Electronic Health Records (EHRs), genomic and epigenomic profiles, and other personalized health datasets, Digital Twins can achieve a highly accurate representation of a patient's unique biological and health characteristics. This integration of multi-dimensional health data enables Digital Twins to simulate individualized health scenarios and predict patient-specific responses to treatments, thus supporting highly tailored therapeutic strategies and personalized intervention planning.

However, while the VHT may be used in research, translating insights into clinical decisions still requires regulatory validation of the Digital Twin applications separately.

✓ **Improved predictive modelling and preventive care**

Data on socio-economic, environmental, and behavioural factors enables Digital Twins to simulate broader influences on health outcomes, enhancing the accuracy of predictive models. Access to pathogen data and aggregated public health data supports Digital Twins in identifying patterns, allowing healthcare providers to predict potential health risks and focus on preventive care, which can reduce the incidence of disease.

Predictive models can provide insights into health trends and risks, but using them as actionable decision-making tools for individual patient care will need regulatory approval.

✓ **Support for clinical trials and drug development**

Digital Twins benefit significantly from access to data on clinical trials, clinical studies, and investigations, enabling them to simulate patient responses to new drugs or treatments. By testing Digital Twin models in a virtual environment before real-world trials, researchers can optimize clinical trial protocols, identify suitable patient cohorts, and potentially accelerate drug development processes.

However, fully integrating Digital Twins into the regulatory process for drug approvals is challenging due to a lack of standard validation frameworks for digital models (see various Sections below for more detailed info).

✓ **Optimization of treatment pathways and resource allocation**

Aggregated data on healthcare needs, resource allocation, and healthcare expenditures can help Digital Twins simulate and optimize healthcare workflows and treatment pathways. This can be used by healthcare providers to improve resource management, reduce costs, and allocate healthcare resources more efficiently.

Nonetheless, regulatory reliance on conventional approaches can limit the role of Digital Twins in approved treatment protocols.

✓ **Enhanced training and decision support for health professionals**

With healthcare-related administrative data and insights into healthcare professional specializations and roles, Digital Twins can simulate complex treatment scenarios and provide decision support to healthcare professionals. This data supports the development of virtual training modules and assists healthcare professionals in making informed, data-driven decisions, thereby improving patient care quality.

✓ **Increased interoperability and cross-border continuity of care**

The EHDS framework enables cross-border access to health data, which is valuable for Digital Twins used in the EU's diverse healthcare systems. By standardizing data access and formats, the EHDS allows Digital Twins to function consistently across Member States, supporting the continuity of care and enabling patients to benefit from Digital Twin insights when receiving care abroad.

However, regulatory acceptance varies from one Member State to another, which may limit the use of Digital Twins in patient-facing scenarios across the EU.

✓ **Population health management and public health research**

Population-based health registries and other public health data provide Digital Twins with the data needed to analyse health trends at a population level, facilitating public health research. This enables the development of Digital Twins for entire communities, helping policymakers simulate health outcomes under various interventions, anticipate public health needs, and improve overall health outcomes across populations.

Still, using Digital Twins to guide public health interventions would require validation and regulatory approval.

✓ **Wellness and preventive health applications**

Data from wellness Apps contribute to a broader view of a patient's lifestyle and daily health habits, which is crucial for building accurate Digital Twins. This data allows Digital Twins to provide insights into preventive health strategies, enhancing long-term wellness by encouraging healthier lifestyle changes tailored to individual profiles.

Member States must designate one or more Health Data Access Bodies (HDAB) responsible for overseeing access to health data for secondary use. They may establish new public bodies or use existing ones, provided they meet the EHDS requirements. If multiple HDABs are designated, one must act as a coordinator within the Member State and in collaboration with HDABs across the EU.

Each HDAB must contribute to the uniform application of the EHDS across the EU and cooperate with all other HDABs, as well as with the EU Commission and competent Data Protection Authorities, to ensure compliance with the GDPR.

Member States must ensure that each HDAB has sufficient human, financial, technical, and ethical resources to perform its tasks effectively. Where necessary, ethics bodies should support HDABs, either by providing expertise or being integrated into the HDAB itself.

Very importantly, HDABs must avoid conflicts of interest, especially regarding their various functions (e.g., data assessment of data access applications, anonymization, and data provision in secure

processing environments). Functional segregation or other organizational safeguards may be used to prevent conflicts within HDABs.

They have several key responsibilities:

- **decide on data access applications and data requests, as well as issue permits for secondary use of health data;**
- **provide secure access to health data for users** and ensure compliance with the EHDS;
- **process and manage data requests, including pseudonymizing or anonymizing the data;**
- make publicly available a standardized and machine-readable national dataset catalogue including details, in the form of metadata, about source and nature of electronic health data, as well as all information relating to data applications, permits, requests and results communicated by the data users;
- preserve confidentiality of IP rights and trade secrets, and ensure high level of data protection;
- **maintain systems to record data access applications, data requests, the decisions on those applications and the data permits issued and data requests answered;**
- cooperate at national and EU levels to lay down common standards, technical requirements and appropriate measures for accessing EHD in a secure processing environment;
- facilitate cross-border access to health data for secondary use through HealthData@EU and publish relevant information.

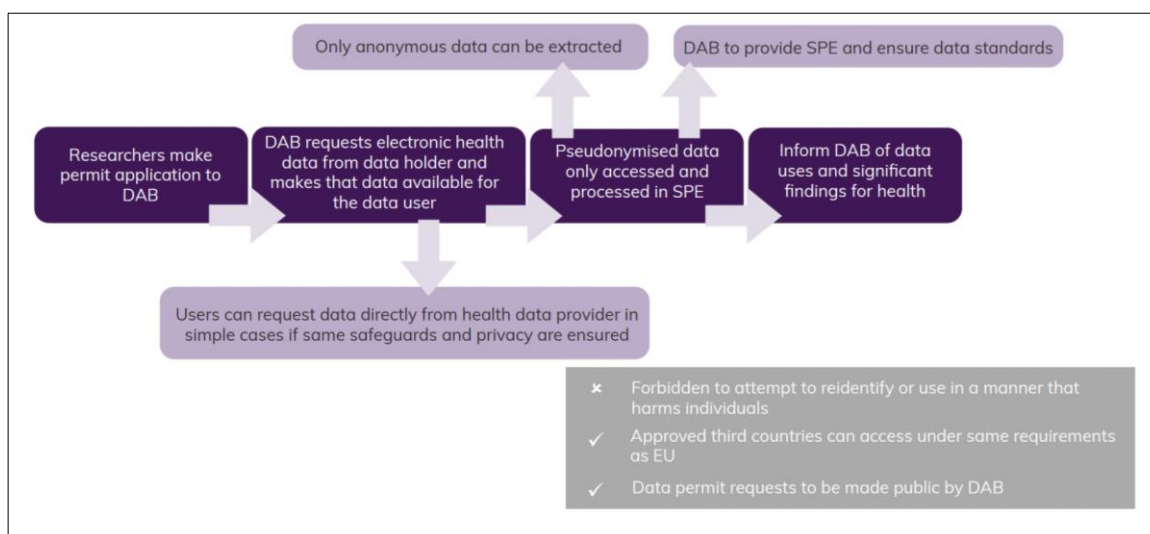


Figure 14 - Data access process for secondary use

Health data users can access and process electronic health data for secondary use only by virtue of a valid data permit (Article 46), a data request (Article 47), or data access approval (per Article 45.3) by the relevant authorized participant in HealthData@EU.

Health data users are **strictly prohibited from attempting to re-identify individuals** linked to the data accessed for secondary use. Additionally, such users must make the results or outputs of their secondary data use publicly available, by **appropriately anonymizing the personal data which may be contained among relevant information**.

Article 44 of the EHDS, in promoting data minimization and purpose limitation in accordance with the GDPR, poses one of the most significant barriers to the implementation of data access for permitted purposes under secondary use scope.

More precisely, Health Data Access Bodies must ensure that the Data Users can only access electronic health data that are adequate, relevant and limited to what is necessary in relation to the purpose of processing indicated in the data access application and in line with the data permit granted. For this reason:

i) Primary option: anonymization

In line with the GDPR and data minimization and purpose limitation principles, **HDAB must provide electronic health data only in an anonymized format**, where the purpose of processing by the data user can be achieved in this manner, taking into account the information provided in the access application.

ii) Secondary option: pseudonymization

To the extent that the health data user has sufficiently demonstrated that the purpose of processing cannot be achieved with anonymized data, the access bodies shall grant access to electronic health data in pseudonymized format. **The information necessary to reverse the pseudonymisation shall be available only to the health data access body or a body that acts as trusted third party in accordance with national law (this requirement appears to ground its rationale in the same ‘subjective’ approach endorsed in the Judgement analysed in 3.4.1 above, as the data would qualify as anonymous for the data user).**

Recital 43 EHDS reads that the **EU Commission should set out the procedures and requirements, and provide technical tools, for a unified procedure for anonymising and pseudonymising the electronic health data** under the EHDS for secondary use. **It goes without saying that everything discussed in Section 3 and later in 4.2, a) and b), is relevant here as well.**

A data permit can be issued by HDAB to a Data User, within 3 months of receiving the application, only when the following cumulative criteria are fulfilled:

- the purpose must be among those explicitly permitted by the EHDS;
- the health data requested must be necessary, adequate, and proportionate, respecting data minimization;
- processing must comply with GDPR, justifying pseudonymization if anonymization is not an option;
- the applicant must have appropriate qualifications and expertise in connection with the purpose to be carried out;
- adequate technical and organizational safeguards must be in place to prevent misuse and protect data subjects’ rights;
- ethical assessments must comply with national law;
- any exceptions to opt-out rights must follow national legal requirements;
- all other legal requirements must be met;
- if no risks may arise for national defence, security, public security and public order, or in relation to confidential data in governmental databases of regulatory authorities.

HDABs may (thus meaning that they are not obliged to) **charge health data users with fees for the services they carry out in relation to secondary use. Such fees must be transparent, non-discriminatory and proportionate to the costs incurred in making the data available, should not hinder competition, and cover all or part of the expenses** associated with assessing data access applications or requests, granting, refusing, or modifying data permits or responding to data requests. **These costs may include costs related to consolidating, preparing, anonymizing, pseudonymizing and provisioning electronic health data.**

Member States may offer **reduced fees for specific types of data users** within the Union, such as public sector bodies or Union institutions, offices, agencies and bodies operating in the field of public health, **university researchers or micro-enterprises**. Fees may also include compensation for costs incurred by health data holders –required to provide an estimate of such expenses to the health data access body – in preparing and compiling data for secondary use.

If the data holders or data users do not agree on the level of the fees within 1 month of the data permit being granted, the HDAB may set the fees in proportion to the cost of making available electronic health data for secondary use. Before issuing a data permit or responding to a data request, HDABs must inform applicants of the expected fees and provide the option to withdraw the application (in this case, applicant can only be charged the costs that have already been incurred).

The EU Commission, through implementing acts, will establish principles for fee policies and structures to ensure consistency and transparency across Member States.

The EHDS final version grants the patients a reversible right to opt-out⁸⁴. Under this provision, Member States must establish an accessible and clear opt-out mechanism. Clearly, this is an obligation for national Legislators and not a choice as for opt-out in primary use. Once an individual opts out, insofar their personal health data can be identified in a dataset, they must no longer be shared with data users, even in anonymized form. Member States may introduce exceptions to this rule, permitting data use despite opt-out, but only for specific secondary uses by public authorities and under strict conditions.

Finally, each Member State must designate one national contact point for secondary use of electronic health data, which will be an organisational and technical gateway, enabling and responsible for making electronic health data available for secondary use in a cross-border context. National contact points must connect to the cross-border infrastructure for secondary use of EHD, **HealthData@EU**, acquiring the needed technical capability to do it.

The EU Commission must develop, deploy and operate a central platform for HealthData@EU by providing IT services needed to support and facilitate the exchange of information between health data access bodies as part of the cross-border infrastructure for the secondary use of electronic health data.

At the same time, following a positive compliance check, the same Commission may, by means of implementing act, take decisions to connect individual authorized participants to HealthData@EU, including:

- **Health-related research infrastructures** or similar structures whose functioning is based on Union law and which support the use of EHD for research, policy making, statistical, patient safety or regulatory purposes;
- **third countries or international organizations**, as long as they prove compliance with the requirements of the EHDS and provide access to data users located in the Union, on equivalent terms and conditions, to the electronic health data available to their health data access bodies, under strict privacy safeguards.

A new independent advisory and regulatory body, the European Health Data Space Board, is established to facilitate the exchange of information and cooperation among Member States and the EU Commission. The EHDS Board will be composed of two representatives per Member State, one nominated for primary use and the other for secondary use. Market surveillance authorities for EHR systems and services, the European Data Protection Board and the European Data Protection Supervisor, EMA, ECDC, ENISA, must be invited to attend the Board's meetings, where the issues are of relevance according to the Board.

The main tasks of the EHDS Board are, both in regard of primary and secondary use:

⁸⁴ See Digital Europe 'Joint Statement: health organisations define EHDS's opt-out required for life-saving research', dated 8 June 2023. <https://www.digitaleurope.org/news/joint-statement-health-organisations-define-ehds-opt-out-required-for-life-saving-research/>.

- assist Member States in coordinating national authorities to ensure a consistent application of the EHDS;
- issue contributions and exchange best practices on the implementation of the EHDS on matters related to the coordination of the implementation of the EHDS at Member State level;
- facilitate cooperation between national health authorities operating under the EHDS through capacity-building and information sharing;
- create guidelines, in consultation and cooperation with relevant stakeholders, including representatives of patients, health professionals and researchers, in order to help health data users to fulfil some specific obligations;
- promote discussions on electronic health data use with advisory forums, regulators, and policymakers.

5.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

In general, the EHDS is both to be welcomed for strengthening patients' control over their electronic health data and their rights in relation to their sharing and is much needed to overcome the current barriers for the secondary use of these data.

It offers significant benefits by facilitating improved healthcare delivery and medical innovation across the EU. Through streamlined access to electronic health data, it can support better-informed and patient-centred care, enabling healthcare providers to deliver more accurate and timely treatments, especially for cross-border patients. Additionally, the EHDS is set to accelerate health research, innovation and public health policy-making by providing researchers and policymakers with access to a broad pool of data, fostering innovation in areas like personalized medicine and public health.

Notwithstanding the above, several substantial loopholes and hurdles remain unsolved, with the likely effect of preventing the EHDS from fully realizing all its objectives. These hurdles are explored here below.

A) National fragmentation is still on track: prevent national limitations

Barrier

Recital 37 states that *“For the purpose of processing electronic health data for secondary use, one of the legal bases set out in Article 6(1), points (a), (c), (e) or (f), of Regulation (EU) 2016/679 [GDPR] combined with Article 9(2) of that Regulation should be required. This Regulation provides a legal basis in accordance with Regulation (EU) 2016/679 (...) for the secondary use of personal electronic health data including the safeguards to permit the processing of special categories of data, in accordance with Articles 9(2), points (g), (h), (i) and (j), of Regulation (EU) 2016/679 (...) Consequently, Member States may no longer maintain or introduce under Article 9(4) of Regulation (EU) 2016/679 further conditions, including limitations and specific provisions requesting the consent of natural persons, with regard to the processing for secondary use of personal electronic health data under this Regulation, except as referred to in Article 33(5)”*.

In all likelihood, this is one of the most important considerations regarding the EHDS. Nonetheless, it has not been reflected in any legal provision in the ‘operative part’ of the Regulation. The consequence is that, **due to the lack of any kind of specific reference to the prevalence of the obligation, posed by Art. 33 EHDS, to make the categories of electronic health data described above available for secondary use, Art. 9.4 of the GDPR – which allows Member States to maintain or introduce**

further conditions, including limitations, with regard to the processing of genetic and health data – will continue to apply in full and unconditionally⁸⁵.

In practice, this means that national legislators retain full discretion to establish significantly more restrictive conditions for data reuse (such as in Italy, among many, where secondary use for scientific research is subject to the prior authorization of the Data Protection Authority).

It is, therefore, a critically decisive question.

Proposed actions

By the EU Commission: any step by the European Data Protection Board would not be sufficient to overcome this impasse, as the Board does not have such powers as to be able to prevent the enforcement of a legal provision (Art. 9.4 GDPR in this case). Therefore, two different paths can be pursued:

- during the very final stage of approval of the Regulation, include in it a specific reference to the non-applicability of Art. 9.4 to data sharing processes for secondary use;
- make the Commission’s checks on compliance with the minimum conditions for the implementation of the processes of secondary use under the EHDS much stricter and more frequent, ensuring that no Member State maintains or adopts additional limitations or further conditions according to Art. 9.4 of the GDPR (opening, in case, infringement procedures against Member States which should continue to get in the way of implementing the EHDS).

B) Direct the HDAB’s task to anonymize and pseudonymize the data

Barrier

As seen above, Art. 37.1(d) EHDS entrusts the Health Data Access Bodies with, *inter alia*, with the task of anonymizing or pseudonymizing the data, in accordance with the principles of data minimization and purpose limitation (before making them available to the data users based on the information and declarations provided in the submitted application), particularly specifying the purpose for secondary use and the justification of why this purpose cannot be achieved with anonymized data.

Given the number of complexities arising around data anonymization, this appears as a very challenging task, if not properly directed. All the more so considering that, as per Art. 51 EHDS “*The health data access body shall be deemed controller for the processing of the personal electronic health data when fulfilling its tasks pursuant to this Regulation*”. This means that HDABs will bear responsibility – and ultimately liability – for the adequacy of the anonymization or pseudonymization techniques they implement.

Proposed action

By the EU Commission and European Data Protection Board, in collaboration with the EHDS Board: establish, as outlined in Recital 43 of the EHDS, specific guidelines and procedures to ensure consistent and comprehensive uniformity across Member States in how HDABs will have to apply data anonymization and pseudonymization to electronic health data (under Articles 37 and 44). In absence of such a guidance, as pointed-out also in Section 4 above, data access authorities will need to navigate the interpretative and practical uncertainty that still surrounds these legal concepts, particularly regarding the boundary that separates pseudonymization from what can confidently be classified as data anonymization (*i.e.*, without significant risk of sanctions or claims). Best practices tailored to the health and care sectors already developed by the ENISA should be integrated and built on.

⁸⁵ See the EDPB-EDPS ‘Joint Opinion 03/2022 on the proposal for a regulation on the European Health Data Space’, dated 12 July 2022. https://www.edps.europa.eu/system/files/2023-04/22-07-12_edpb_edps_joint-opinion_europeanhealthdataspace_en.pdf.

C) Outlining the exact boundaries of research-related purposes for secondary use

Barrier

From the perspective of industry and research stakeholders, the opening offered by the EHDS to the reuse of electronic health data for (i) development and innovation activities for products or services, and (ii) training, testing and evaluation of algorithms, including in medical devices, in-vitro diagnostic medical devices, AI systems and digital health applications (Art. 34.1(e) EHDS), stands as crucial.

Among other effects, this reuse may enable the creation of advanced healthcare products and services, fostering breakthroughs in treatment methods, diagnostics, and patient care technologies, as well as support the training, testing, and fine-tuning of algorithms for AI-driven healthcare technologies, improving their accuracy, reliability, and safety, ultimately leading to more effective and personalized healthcare solutions.

Notwithstanding that these activities must be strictly connected to “*scientific research related to health or care sectors, contributing to public health or health technology assessment, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices*”, in the absence of any definition of what ‘scientific research’ exactly is from a legal standpoint and of a precise drawing of the boundaries of the above purposes, as it already happens in relation to health research under Art. 9.2.(h) GDPR, stakeholders may find that the benefits are not commensurate with the risks that may arise (*e.g.*, blocking of activities, sanctions).

Proposed action

By the European Health Data Space Board, together with the European Data Protection Board: adopt guidelines and best practices regarding the purposes of secondary use identified by Art. 34.1(e) of the EHDS, to exactly pinpoint what type of research-related activities is in scope or not, what categories of product and services must be excluded from this provision (*e.g.*, health-monitoring Apps), and to what extent the electronic health data can be reused to train, test and fine-tune AI system, if qualified as high risk according to the AI Act, and algorithms, if used for HRAI models or systems. Solely for those cases when access is allowed by the health data access bodies to the electronic health data in pseudonymized form, these guidelines should also set out further technical, organizational and security measures which may help a data user to better comply with the principles of data minimization and privacy-by-design.

6 Artificial Intelligence Act

6.1 THE FRAMEWORK FOR VHT

Artificial Intelligence plays a crucial role in the concept and the future of the VHT. This set of methods – which includes, among others, machine learning, logic- and knowledge-based and statistical approaches – are consistent with and enable the primary aim to offer fundamental descriptive insights into both the healthy and unhealthy human body, as well as to enable actionable model-based predictions regarding the impact of interventions on the health and illnesses of individuals.

The ability of AI to uncover unnoticed connections and hidden patterns within vast datasets, consequently enhancing the understanding of individual health conditions and treatment responses, is one of the main reasons explaining and displaying the potential impact of AI in the development of Digital Twins. AI also enables continuous data analysis and learning from different and various data sources (including wearables and medical devices), allowing Digital Twins to be updated in real-time and to continuously increase their predictive accuracy and relevance, thus ensuring timely interventions and improved patient outcomes. This set of technologies is also capable of enhancing predictive modelling in the healthcare domain, thanks to the fast and efficient analysis of historical data and patient behaviours, providing health professionals with the possibility to anticipate complications and adjust treatment plans proactively. Other benefits deriving from the application of AI in the context of Digital Twins concern automation of *routine* tasks (*i.e.*, this technology helps automating data collection and analysis, so that healthcare professionals can focus on others and more complex activities), support for clinical decision-making (*i.e.*, AI can help provide evidence-based and personalized recommendations for each Digital Twin, improving the quality of care and limiting variability in treatment approaches), simulations (*i.e.*, AI enables the simulation of multiple treatment options within the Digital Twin, helping doctors to visualize possible outcomes and thus make more informed clinic decisions), patient engagement (*i.e.*, thanks to AI, health data can be visually presented through Digital Twins, allowing patients to be able to better understand their health status and treatment options, which results in improved engagement and adherence to treatment plans), personalization at scale (*i.e.*, with AI support, Digital Twins can help generate highly personalized treatment plans for a larger population, thus rendering customized medicine more affordable).

AI technologies can be utilized in numerous ways to develop the VHT and associated DTs:

- **Enhancing system assessment:** AI can significantly increase the sheer number of systems that can be assessed;
- **Efficiency in operator-intensive tasks:** AI models can accelerate some operator-intensive tasks within the procedure, such as replacing manual image segmentation with precise models, like convolutional neural networks;
- **Development of surrogate models:** AI can facilitate the development and incorporation of surrogate models or machine learning-based PDE solvers, replacing some features of multiscale models, to replace more expensive and computation-intensive aspects of these simulations with a more efficient approach;
- **Model personalization:** AI can aid in parameterizing and personalizing mechanistic models;
- **Data-driven predictive models:** the application of AI can result in fully data-driven predictive models that, for instance, would provide decision support for treatment options or patient-specific prognosis predictions;
- **Hypothesis generation:** AI could assist in generating hypotheses from data, informing and guiding the development of mechanistic multilevel models.

Therefore, AI is a crucial component in creating Digital Twins in the healthcare and clinical domains. On this regard, the implementation of AI systems in the context of the VHT should now and in the

future be addressed following the new legal and ethical framework set forth by the recently adopted Artificial Intelligence Act (AI ACT)⁸⁶.

This landmark piece of legislation stands at the forefront of the EU digital future, creating an innovative and comprehensive regulatory framework designed to ensure that AI technologies are developed and deployed safely, securely ethically, and in line with EU fundamental rights and values. The AI ACT aimed at creating a balanced regulatory environment to foster innovation while at the same time ensuring public safety and fundamental rights. This was accomplished by introducing a risk-based classification for AI systems, parametrizing bans and requirements accordingly, as well as providing for several measures to support innovation, in order not only to protect individual rights and freedoms, but also to stimulate economic growth and competitiveness within the EU market. By prioritizing human-centric AI solutions, the AI ACT set a global standard for AI governance with the ambition of positioning the European Union as a world leader in the ethical development of AI, increasing trust in AI technologies among both citizens and businesses.

These new innovative and unprecedented regulatory approaches to AI pose, however, complex challenges for the VHT community of professionals, due to the disruptive growth in the use of AI for the creation of Digital Twins.

The first challenge is to understand whether, and if so in which cases, the technologies generally associated with the concept of artificial intelligence employed in the creation of Digital Twins – and/or the VHT itself – could be qualified as “*artificial intelligence systems*” for the purposes of the AI ACT.

From a technological point of view, artificial intelligence is not a monolithic concept. Instead, it encompasses a wide variety of technologies, methods, and applications, each of which has specific features and implications. In this regard, in attempting to identify a common, all-encompassing definition of AI to be used as the cornerstone of a regulatory framework, several challenging elements need to be considered. Among others, the AI field is in permanent and dynamic evolution: new approaches and techniques are emerging at an unprecedented rate, also and in particular when dealing with the healthcare and clinical domains. Such dynamicity compromises any attempt to establish a static definition capable of effectively encompassing all existing and forthcoming scenarios. Moreover, AI systems can considerably differ in both complexity and capabilities, and such variability calls into question the feasibility or appropriateness of any one-size-fits-all definition of AI.

From a legal point of view, there are substantial implications associated with the establishment of a definition of AI. A narrow and clear definition is essential to determine a well-defined and predictable allocation of responsibilities, accountability and compliance efforts. In the absence of such a definition, ambiguity, opacity and arbitrariness are introduced into the legal framework, leading to enforcement problems, risks of circumvention, potential gaps in the protection of individuals and detrimental effects on the public economy and EU competitiveness.

The AI ACT recognized this complexity⁸⁷ and sought to shed light on this ever-changing landscape by proposing a specific definition of “*artificial intelligence system*”.

⁸⁶ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

⁸⁷ The first attempt to normatively define an “*AI system*” was developed by the EU commission in the *Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts* (COM/2021/206 final): «*‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*». This proposal for a definition was then followed by the proposal of the European Council («*‘artificial intelligence system’ (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts*») and the proposal of the European Parliament («*‘artificial intelligence system’ (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments*»).

Article 3(1)(1) – AI Act

«*‘AI system’ means a **machine-based system** that is designed to operate with **varying levels of autonomy** and that may exhibit **adaptiveness after deployment**, and that, for explicit or implicit objectives, **infers**, from the **input** it receives, how to generate **outputs** such as predictions, content, recommendations, or decisions that can **influence physical or virtual environments**».*

This concept is «*based on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches and [...] not cover systems that are based on the rules defined solely by natural persons to automatically execute operations*»⁸⁸. It is also clarified that, for the purpose of the definition adopted by the AI ACT, «*AI systems can be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serves the functionality of the product without being integrated therein (non-embedded)*»⁸⁹.

In order to assess the relevance of the adopted definition in relation to the technologies connected to the creation of Digital Twins (and/or the VHT area itself), each of the elements that make up the above-mentioned definition should be individually analysed. In this regard, a first important guideline for the interpretation and application of this pivotal rule is provided by Recital 12 of the AI ACT, where several of the building blocks of the definition are detailed:

- **machine-based:** «*refers to the fact that AI systems run on machines*»;
- **varying levels of autonomy:** «*meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention*»;
- **adaptiveness after deployment:** «*refers to self-learning capabilities, allowing the system to change while in use*»;
- **explicit or implicit objectives:** «*underscores that AI systems can operate according to explicit defined objectives or to implicit objectives. The objectives of the AI system may be different from the intended purpose of the AI system in a specific context*»;
- **capability to infer:** «*refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data. The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved. The capacity of an AI system to infer transcends basic data processing by enabling learning, reasoning or modelling*»;
- **environments:** «*should be understood to be the contexts in which the AI systems operate, whereas outputs generated by the AI system reflect different functions performed by AI systems and include predictions, content, recommendations or decisions*».

Further keys to interpreting and applying the definition of “AI system” adopted by the AI ACT might be found in the “*Explanatory memorandum on the updated OECD definition of an AI system*” drawn up by the Organisation for Economic Cooperation and Development (OECD)⁹⁰.

⁸⁸ Recital 12 AI ACT.

⁸⁹ Recital 12 AI ACT.

⁹⁰ OECD (2024), “Explanatory memorandum on the updated OECD definition of an AI system”, OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, <https://doi.org/10.1787/623da898-en>.

OECD definition of “AI system”

«An AI system is a **machine-based system** that, for explicit or implicit **objectives**, **infers**, from the **input** it receives, how to generate **outputs** such as predictions, content, recommendations, or decisions that can **influence physical or virtual environments**. Different AI systems vary in their levels of **autonomy** and **adaptiveness after deployment**».

Among other specifications, a list of useful clarifications translate into the following concepts, which are common to the definition of “AI system” provided by the AI Act:

- **Autonomy:** «means the degree to which a system can learn or act without human involvement following the delegation of autonomy and process automation by humans. Human supervision can occur at any stage of the AI system lifecycle, such as during AI system design, data collection and processing, development, verification, validation, deployment, or operation and monitoring. Some AI systems can generate outputs without these outputs being explicitly described in the AI system’s objective and without specific instructions from a human»;
- **Adaptiveness:** «is usually related to AI systems based on machine learning that can continue to evolve after initial development. The system modifies its behaviour through direct interaction with input and data before or after deployment»;
- **Environment or context:** «is an observable or partially observable space perceived using data and sensor inputs and influenced through actions (through actuators). The environments influenced by AI systems can be physical or virtual and include environments describing aspects of human activity, such as biological signals or human behaviour. Sensors and actuators are either humans or components of machines or devices»;
- **Objectives:** «AI system objectives can be explicit or implicit; for example, they can belong to the following categories, which may overlap in some systems: (-) Explicit and human-defined [...] (-) Implicit in (typically human-specified) rules [...]; (-) Implicit in training data [...] (-) Not fully known in advance [...]. AI systems can operate according to one or more types of objectives. In addition, user prompts can supplement design objectives when the system is in operation [...]»;
- **Input:** «Input is used both during development and after deployment. Input can take the form of knowledge, rules and code humans put into the system during development or data. Humans and machines can provide input. During development, input is leveraged to build AI systems, e.g., with machine learning that produces a model from training data and/or human input. Input is also used by a system in operation, for instance, to infer how to generate outputs. Input can include data relevant to the task to be performed or take the form of, for example, a user prompt or a search query»;
- **Inference:** «The concept of “inference” generally refers to the step in which a system generates an output from its inputs, typically after deployment.⁵ When performed during the build phase, inference, in this sense, is often used to evaluate a version of a model, particularly in the machine learning context. In the context of this explanatory memorandum, “infer how to generate outputs” should be understood as also referring to the build phase of the AI system, in which a model is derived from inputs/data»;
- **Output(s):** «The output(s) generated by an AI system generally reflect different functions performed by AI systems. AI system outputs generally belong to the broad categories of recommendations, predictions, and decisions. These categories correspond to different levels of human involvement, with “decisions” being the most autonomous type of output (the AI system affects its environment directly or directs another entity to do so) and “predictions” the least autonomous».

Given the building blocks of the definition identified above, it is likely that a DT can be qualified as an “AI system” under the AI ACT. A DT is a machine-based system that is provided with varying levels of autonomy, being able to operate with a degree of independence from the intervention of humans

(e.g., healthcare professionals). The ability of a DT to learn on an ongoing basis, either through new data sources (constantly/periodically) being updated (e.g., wearable and medical devices) or by learning from its previous performance, makes this system endowed with sufficient adaptability after deployment. In addition, most DTs operate according to clear objectives and is based on evident inferential capabilities: from the large amount and variety of input data, relating even to an individual patient’s medical history, the system can obtain outputs such as visual representations and recommendations for healthcare treatment. This also makes the VHT capable of influencing the relevant health domain environment, both doctor-side and patient-side. In addition, other systems that compose and/or take part in the process of developing and deploying a DT could also be qualified as “AI systems” under the AI ACT where all the above conditions are met.

In addition to the definition of “AI system”, the notion of “*general-purpose AI model*” adopted by the AI ACT might also be relevant in the context of the use of AI technologies for the implementation of Digital Twins.

Article 3(1)(63) – AI Act

«*‘general-purpose AI model’ means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*».

This concept «*should be clearly defined and set apart from the notion of AI systems to enable legal certainty*» and «*should be based on the key functional characteristics of a general-purpose AI model, in particular the generality and the capability to competently perform a wide range of distinct tasks*»⁹¹. The AI ACT also highlights that «*these models are typically trained on large amounts of data, through various methods, such as self-supervised, unsupervised or reinforcement learning*» and that «*general-purpose AI models may be placed on the market in various ways, including through libraries, application programming interfaces (APIs), as direct download, or as physical copy. These models may be further modified or fine-tuned into new models*»⁹². Regarding the relationship between AI systems and AI models, the AI ACT also clarifies that «*although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems*», subsequently specifying that «*when a general-purpose AI model is integrated into or forms part of an AI system, this system should be considered to be a general-purpose AI system when, due to this integration, this system has the capability to serve a variety of purposes. A general-purpose AI system can be used directly, or it may be integrated into other AI systems*»⁹³.

The second challenge affecting the application of the AI ACT to AI systems employed in the context of the VHT – where they fall under the definition analysed above – concerns the scope of application of this EU regulation. In fact, while the scope of the AI ACT tends to encompass virtually every area of AI application, certain exceptions have been introduced. The latter, in some cases, found justification in the opportunity to foster the development and deployment of AI in specific contexts.

In the Digital Twins domain, the most important provision is the one set forth by Article 2(6) of the AI ACT, which states that «*this Regulation does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development*». This rule has been adopted on the basis that this piece of legislation «*should support innovation, should respect freedom of science, and should not undermine research and development activity*». At the same time, the lawmaker limited the scope of application of this clause to the sole “scientific” field («*without prejudice to the exclusion of AI systems specifically developed and put into*

⁹¹ Recital 97 AI ACT.

⁹² Recital 97 AI ACT.

⁹³ Recital 97 AI ACT.

service for the sole purpose of scientific research and development, any other AI system that may be used for the conduct of any research and development activity should remain subject to the provisions of this Regulation»⁹⁴).

The rationale behind the above-mentioned exclusion is the same reason why this EU regulation does not apply to *«to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service»⁹⁵*. Incidentally, in view of the necessity of conducting research and development focused on AI systems and models for the creation, validation, and use of DTs, this second area of exclusion could also be relevant in the context under consideration. However, the AI ACT sets a boundary to this exception, clarifying that it is *«without prejudice to the obligation to comply with this Regulation where an AI system falling into the scope of this Regulation is placed on the market or put into service as a result of such research and development activity and to the application of provisions on AI regulatory sandboxes and testing in real world conditions»⁹⁶*.

Another key challenge in assessing the application of the AI ACT in the context of the Digital Twins is to figure out whether and how AI systems and models deployed in this field – or the VHT considered as an AI system under this EU regulation – rank against the risk classification adopted by the AI ACT for defining bans, obligations and requirements.

The main reference to be considered is the classification rules for high-risk AI systems set forth by Article 6 of the AI ACT: *«Irrespective of whether an AI system is placed on the market or put into service independently of the products referred to in points (a) and (b), that AI system shall be considered to be high-risk where both of the following conditions are fulfilled: (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I; (b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I»*.

Considering that the list of EU harmonisation legislation in Annex I includes *Regulation (EU) 2017/745 on medical devices* and *Regulation (EU) 2017/746 on in vitro diagnostic medical devices*, is it likely that DTs will be qualified as high-risk AI systems in connection with their classification as medical devices pursuant to the above-mentioned legislations.

At the same time, it cannot be ruled out that some of the AI models and/or AI systems that are components of a DT or that contribute to its implementation may fall into one of the (other) risk classes identified by the AI ACT.

For instance, this is the case of the so-called “limited risk” AI systems, *i.e.*, those AI systems in relation to which this EU regulation lays down certain transparency obligations due to the risks of impersonation, manipulation or deception associated with them. On this regard, it is possible that a DT might include AI systems intended to interact directly with natural persons (*e.g.*, a doctor and/or patient conversation interface). Moreover, it should be assessed whether some of the possible outputs of AI systems that are part of a DT (*e.g.*, representations of human organs) could be qualified as synthetic audio, image, video or text content, or as image, audio or video content constituting a deep fake (where the AI ACT qualifies as “*deep fake*” any *«AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful»*). In these cases, the labelling requirements provided for in Article 50 of the AI ACT would apply.

Similarly, given the potential need of models with the capability to perform a wide range of distinct tasks, as well as the availability of large number of datasets from a variety of different sources that are used to create Digital Twins, it cannot be ruled out the relevance of the provisions laid down by the AI

⁹⁴ Recital 25 AI ACT.

⁹⁵ Article 2(8) AI ACT.

⁹⁶ Recital 25 AI ACT.

ACT for general-purpose AI models. Once again, there would emerge the necessity to also consider the application of the requirements set forth in this EU regulation for this specific category of AI systems⁹⁷.

To properly appreciate which obligations should apply to AI systems that are part and/or are involved in the creation and deployment of DTs, taking into account the level of risk relevant in each case (and without prejudice to the exceptions to the scope of application outlined above), the subjective scope of application of the AI ACT should also be duly considered.

Notwithstanding the aim of establishing accountability for the entire AI value chain, this EU regulation places most of the obligations and requirements on the “providers” of AI systems.

Article 3(1)(3) – AI ACT

«‘provider’ means a natural or legal person, public authority, agency or other body that **develops** an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge».

Alongside providers of AI systems, the AI ACT also gives special consideration to the “deployers” of AI systems.

Article 3(1)(4) – AI ACT

«‘deployer’ means a natural or legal person, public authority, agency or other body **using** an AI system under its authority except where the AI system is used in the course of a personal non-professional activity».

Qualifying a stakeholder as a provider or deployer of AI systems is fundamental in view of defining obligations and responsibilities under the AI ACT. In the context of the Digital Twins, there are several entities, both public and private, that play a qualified role in the development and deployment of AI systems. These include, among others, hospitals, doctors, research centres, universities, technology partners specialized in specific stages of AI system development (e.g., retrieval and processing of training data, model training, validation, model customization). Each of them could qualify as a provider or deployer of one or more AI systems (or of the VTH as an AI system, and in such cases there could be sharing of roles and related responsibilities between two or more entities to be assessed, e.g., in the case of a consortium). In addition, under the rules of the AI ACT, a deployer could also turn into a provider, if specific conditions are met⁹⁸.

From the combination of the role held by a stakeholder in the AI value chain and the level of risk of an AI system pursuant to the rules on classification laid down by this EU regulation, it is subsequently possible to identify the applicable obligations and requirements under the AI ACT. Although in the case of a DT being considered as an AI system – or an AI system used for its implementation – all the relevant provisions set forth in the AI ACT would apply, some of the requirements set forth for high-risk AI systems assume special relevance in this context.

Regarding providers, reference should be made to data and data governance obligations. One of the problems linked to AI systems, especially when applied in the medical domain, is that of ensuring data accuracy and quality, not only for the input data, but also in connection with the logic and functioning of the AI system and its final outputs. This is why Article 10 of the AI ACT provides that the datasets used for training, validation and testing of high-risk AI system must comply with specific quality criteria. In particular, training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system. In addition,

⁹⁷ See Chapter V of the AI ACT.

⁹⁸ «Any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system for the purposes of this Regulation and shall be subject to the obligations of the provider under Article 16, in any of the following circumstances: (a) they put their name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are otherwise allocated; (b) they make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in such a way that it remains a high-risk AI system pursuant to Article 6; (c) they modify the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such a way that the AI system concerned becomes a high-risk AI system in accordance with Article 6. Article 25 of the AI ACT» (Article 25 AI ACT).

training, validation and testing data must be relevant, sufficiently representative and, as much as possible, free of errors, unbiased and complete, as well as possess appropriate statistical properties, considering the particular characteristics or elements of the specific geographical, behavioural or functional context within which the system is to be used.

Also relevant are the obligations of the provider regarding human oversight. Article 14 of the AI ACT requires that high-risk AI systems should be designed to ensure effective human oversight throughout their entire use. To facilitate oversight, the high-risk AI system should enable users to understand its capabilities and limitations, allowing for monitoring of its performance, be aware of potential automation bias and the risks of over-relying on its outputs, especially in decision-making, accurately interpret its outputs using available tools and methods, choose not to use the AI system or override its outputs when necessary, intervene or halt the AI system's operation safely if needed.

Regarding the deployer, the requirement to perform a fundamental rights impact assessment should be mentioned. Pursuant to Article 27 of the AI ACT, in some circumstances before deploying a high-risk AI system, public entities and private providers of public services (and other specific entities) must assess the potential impact on fundamental rights of the AI systems. This could be particularly relevant for public entities, such as hospitals, that might want to employ a DT as deployers.

Finally, among all the other AI ACT provisions relevant in the Digital Twins framework, it is necessary to highlight some of the measures in support of innovation that were introduced by this EU regulation.

In particular, the rules on AI regulatory sandboxes assume special relevance in this context.

Article 3(1)(55) – AI ACT

«'AI regulatory sandbox' means a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, pursuant to a sandbox plan for a limited time under regulatory supervision».

In the light of the AI ACT, the establishment of AI regulatory sandboxes aims to achieve several purposes, regarding legal certainty (*i.e.*, improving compliance with relevant regulations), best practices (*i.e.*, promoting cooperation and sharing of best practices among authorities), innovation and competitiveness (*i.e.*, support the development of a robust AI ecosystem), regulatory learning (*i.e.*, contributing to evidence-based regulatory insights) and market access (*i.e.*, facilitating faster access to the EU market for AI systems, especially from SMEs and start-ups). For these reasons, pursuant to Article 57 of the AI ACT, each EU Member State must ensure that their competent authorities establish at least one AI regulatory sandbox by mid 2026.

As an additional innovation measure connected with the establishment of regulatory sandboxes, the AI ACT provides for specific rules on further processing of personal data for developing certain AI systems in the public interest in the AI regulatory sandbox (including AI systems developed for safeguarding substantial public interest by a public authority or another natural or legal person in the area of public safety and public health, including disease detection, diagnosis prevention, control and treatment and improvement of health care system).

6.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

Using AI technologies to support the creation and the implementation of Digital Twins constitutes a significant advancement in the way healthcare could be approached, fostering more efficient, proactive and personalized methods and systems in this domain. The several advantages connected with the adoption of AI in the VHT framework highlight how this set of technologies can enhance the capabilities of Digital Twins, thus having the potential to transform the overall healthcare experience for both patients and medical professionals.

Notwithstanding the above, AI is now being governed by a recently adopted and particularly innovative legal framework. This can bring to regulatory uncertainty, also due to the lack, as yet, of specific

guidelines. Therefore, this regulatory scenario could pose significant barriers to the full exploitation of AI for Digital Twins.

These barriers need to be properly and proactively addressed in advance to enable the full exploitation of AI in the VHT ecosystem in full compliance with the legal and ethical framework of the AI Act.

A) Clarify the definition of “AI system”

Barrier

The definition of “AI system” is crucial to the application of the AI ACT. The extensive discussions that have been carried out during the process of approval of the AI ACT, with different definitions having been proposed over time, highlight both the complexity and sensitive nature of this issue and the need to reach a definition that is certain, clear and with predictable effects.

Currently, the definition of “AI system” adopted by the AI ACT appears particularly broad and likely to encompass a large and undefined number of technologies and applications. However, in the absence of further guidance and interpretation elements, this might undermine the purposes of this piece of legislation. On the one hand, the potential vagueness of the definition could lead to uncertainty among public and private stakeholders, including those acting within the VHT framework. On the other hand, this same vagueness could encourage circumvention attempts aimed at excluding the application of this EU regulation.

The above-mentioned scenario is particularly relevant in the context of DTs. The potential uncertainty around the concept of “AI system” could limit or compound the use of AI for research and medical application, preventing the EU community from benefiting from the highly significant advances that this technology can bring in this specific area.

In light of such a framework, the EU Commission’s task to adopt guidelines on the implementation of the AI ACT becomes crucial. In particular, pursuant to Article 94 of the AI ACT, *«[t]he Commission shall develop guidelines on the practical implementation of this Regulation, and in particular on [...] the application of the definition of an AI system as set out in Article 3, point (1)»*. In these cases, the Regulation requires the European Commission to *«pay particular attention to the needs of SMEs including start-ups, of local public authorities and of the sectors most likely to be affected by this Regulation»* and clarify that such guidelines *«shall take due account of the generally acknowledged state of the art on AI, as well as of relevant harmonised standards and common specifications that are referred to in Articles 40 and 41, or of those harmonised standards or technical specifications that are set out pursuant to Union harmonisation law»*.

These guidelines should be adopted in a way that reduces any room for doubt regarding the scope of application of the definition of “AI system” to a reasonable minimum.

Proposed action

- **By the EU Commission:** adopt, as a matter of priority, the guidelines on the application of the definition of “AI system” as set out in the AI ACT. It is deemed appropriate to suggest that these guidelines should include practical examples of the application of the definition of an “AI system” referring to strategic domains. As such, the qualification of DTs as “AI systems” under the AI ACT should be explicitly addressed, where appropriate by consulting with public and private stakeholders operating in the VHT framework. This could also be a useful benchmark for other areas of scientific research.
- **By the European Artificial Intelligence Board:** in application of the tasks defined by Article 66 of the AI ACT, on its own initiative, issue its recommendations and/or its written opinions on the EU Commission’s guidelines on the application of the definition of an “AI system” as set out in the AI ACT, following the approach and the aim described above, as well as taking into account the VHT framework.

B) Define the boundaries of the exception for scientific research

Barrier

The AI ACT does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development.

This provision is particularly relevant to the VHT field, potentially allowing the use of AI for the benefit of this specific domain of scientific research without the application of the AI ACT, in accordance with the express intention of the EU lawmaker.

However, the aim of supporting innovation, respecting freedom of science and not undermining research and development activity behind this exception might be thwarted by a wording which is not particularly accurate and comprehensive.

On the one hand, the concept of scientific research is not explicitly defined in the AI ACT. The absence of a definition on this point results in a situation of uncertainty with respect to which research activities might qualify as having a scientific nature.

On the other hand, the reference to the «*sole purpose*» of scientific research and development introduces a factor of ambiguity in defining what may or may not fall within the scope of application of this exception.

Proposed action

- **By the EU Commission / European Artificial Intelligence Board:** adopt an opinion, recommendation or guideline clarifying what is meant by «*scientific research and development*» and how the «*sole purpose*» limit should be interpreted for the purpose of applying the exception provided by the AI ACT.
- **By the European Data Protection Board:** suggest a definition of “*scientific research*” within the *Guidelines on the processing of data for scientific research purposes*, which are expected to be adopted in the next year following the Work Programme 2024-2025, and/or align the definition that will be used in this context to the meaning of “*scientific research*” relevant in view of the application of the exception provided by the AI ACT.

C) Harmonize and coordinate risk classification rules among different regulations

Barrier

Since Annex I of the AI ACT includes *Regulation (EU) 2017/745 on medical devices* and *Regulation (EU) 2017/746 on in vitro diagnostic medical devices* among the relevant EU harmonization legislation, it is likely that DTs will be qualified as high-risk AI systems in connection with their classification as medical devices pursuant to the above-mentioned legislations.

However, on this regard Recital 51 of the AI ACT clarifies that «*the classification of an AI system as high-risk pursuant to this Regulation should not necessarily mean that the product whose safety component is the AI system, or the AI system itself as a product, is considered to be high-risk under the criteria established in the relevant Union harmonisation legislation that applies to the product. This is, in particular, the case for Regulations (EU) 2017/745 and (EU) 2017/746, where a third-party conformity assessment is provided for medium-risk and high-risk products*».

The mere fact that DTs – just like other products under other EU harmonized legislation – is subject to multiple EU regulations constitutes a factor of complexity to be considered. In addition, DTs may be classified with different levels of risk depending on the EU regulation under consideration.

In this scenario, the different risk classification rules adopted by the AI ACT and the *Regulation (EU) 2017/745 on medical devices* and the *Regulation (EU) 2017/746 on in vitro diagnostic medical devices*

and the potential different qualification of the risk level for the DTs between the AI ACT and these other pieces of legislation could represent a significant barrier to the development of the VHT.

In the absence of coordination between these different EU regulations, there is the risk of subjecting the VHT to over-regulation, with a consequent waste of resources and duplication of requirements, which would be detrimental to the goals that this innovative technology aims to achieve.

Proposed action

- **By the EU Commission:** adopt guidelines containing detailed information on the relationship of the AI ACT with the EU harmonization legislation listed in Annex I, and in particular with *Regulation (EU) 2017/745 on medical devices* and *Regulation (EU) 2017/746 on in vitro diagnostic medical devices*, pursuant to Article 96(1)(e) of the AI ACT. The relationship between the risk classification criteria adopted by these EU regulations should be specifically assessed.
- **By the European Artificial Intelligence Board / Medical Device Coordination Group:** issue recommendations and/or written opinions on the EU Commission’s guidelines containing detailed information on the relationship of the AI ACT with *Regulation (EU) 2017/745 on medical devices* and *Regulation (EU) 2017/746 on in vitro diagnostic medical devices*, following the approach and the aim described above.

D) Clarify the concepts of provider and deployer and coordinate them with other regulations

Barrier

The AI ACT allocates responsibilities and compliance duties especially to providers and deployers of AI systems. However, the high-level definitions of “*provider*” and “*deployer*” provided in this EU regulation do not seem equipped to resolve in advance borderline situations that may occur in practice.

This is particularly true in the domain of the VHT, where a large variety of public and private stakeholders carry out different and complementary activities in the development and deployment of AI systems in the healthcare domain.

The potential gap between the definitions abstractly defined by the AI ACT and concrete situations that may occur in real-life scenarios could lead to uncertainty in relationships between different players involved in the AI value chain. This is likely to raise the risk of misallocation of responsibilities as required by law, as well as to discourage AI development and deployment activities, especially in the healthcare domain.

In addition, and for the same reasons, the definitions of “*provider*” and “*deployer*” might require to be coordinated with the definitions of other subjective categories that are central to other legislation. In particular, this is the case of the concepts of “*data controller*” and “*data processor*” pursuant to the GDPR.

Proposed action

- **By the EU Commission/European Artificial Intelligence Board:** adopt an opinion, recommendation or guideline clarifying the concepts of “*provider*” and “*deployer*”, declining them in practical terms based on real-life scenarios.
- **By the European Data Protection Board:** amend the *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*, adopted on 7 July 2021 (the version 2.1), including a section aimed at clarifying – with practical examples – how the concepts of “*data controller*” and “*data processor*” under the GDPR relate to the recently adopted definitions of “*provider*” and “*deployer*” in the AI ACT.

E) Support the adoption of guidelines and standards for high-risk systems requirements and obligations

Barrier

Placing on the market, putting into service, and using high-risk AI systems in the European Union is now subject to several requirements and obligations pursuant to the AI ACT. Although the importance of these measures is generally acknowledged, there is a risk that they could become a barrier to the development and use of AI systems, especially in the healthcare domain, including DTs.

In fact, these obligations and requirements are particularly demanding and are often described only in general terms by the AI ACT. At the same time, for the fulfilment of these obligations, previous harmonization experience cannot be relied upon.

The situation just described is particularly significant in areas such as the development and deployment of DTs in the healthcare domain. In this scenario, such complex obligations – like those on data and data governance and human oversight – require to be accompanied by regulatory tools that can facilitate compliance, so that the latter does not become an obstacle to scientific research and implementation of AI in favour of public health.

Proposed action

- **By the EU Commission:** adopt the guidelines on the application of the requirements and obligations referred to in Articles 8 to 15 and in Article 25 of the AI ACT (pursuant to Article 96 of the AI ACT). These guidelines should also devote specific attention to the application of these measures in strategic domains, like the healthcare field.
- **By the European Artificial Intelligence Board:** in application of the tasks defined by Article 66 of the AI ACT, issue its recommendations and/or its written opinions on the EU Commission's guidelines on the application of the requirements and obligations referred to in Articles 8 to 15 and in Article 25 of the AI ACT, following the approach and the aim described above, as well as taking into account the specificities applying to the healthcare framework.
- **By the standardization bodies:** move forward and quickly conclude the work on the creation of standards referring to obligations and requirements for high-risk AI systems, taking into consideration the needs of strategic domains, such as the healthcare field.

F) Promote a specific regulatory framework for VHT in the EU

Barrier

The potential and beneficial effects that AI can derive are countless and impact on almost every area of the economic and individual lives of citizens and businesses operating in the European Union. The AI ACT established a framework of rules for legal and ethical development and deployment of AI systems in the European Union, without privileging specific fields of AI application, while also declining requirements and prescriptions in specific areas according to a risk-based approach. At the same time, this EU regulation has introduced several relevant measures aimed at supporting innovation.

This scenario, however, while surely positive, could still result in having areas where AI is already being applied not finding sufficient consideration in EU and national legislative policies, due to several factors (*e.g.*, investment concentration, regulatory limits, secondary interests).

To prevent such a situation from affecting the VHT domain, with respect to which AI is a crucial driver, the goal of policies which ensure the development of a specific regulatory framework for the VHT in the EU becomes a matter of priority.

Proposed action

- **By EU Member States:** promote the establishment of AI regulatory sandboxes on VHT by competent authorities at national levels in conjunction with competent authorities of other Member States.
- **By National competent Authorities:** establish of AI regulatory sandboxes on VHT (as stated in the previous point), also allowing for the involvement of private and public actors within the VHT AI ecosystem.
- **By the AI Office and EU Member States:** encourage and facilitate the drawing up of codes of conduct specifically addressed to the VHT, pursuant to Article 95(2) of the AI ACT, considering the exemplificative area listed in the AI ACT «(a) *applicable elements provided for in Union ethical guidelines for trustworthy AI*; (b) *assessing and minimising the impact of AI systems on environmental sustainability, including as regards energy-efficient programming and techniques for the efficient design, training and use of AI*; (c) *promoting AI literacy, in particular that of persons dealing with the development, operation and use of AI*; (d) *facilitating an inclusive and diverse design of AI systems, including through the establishment of inclusive and diverse development teams and the promotion of stakeholders' participation in that process*; (e) *assessing and preventing the negative impact of AI systems on vulnerable persons or groups of vulnerable persons, including as regards accessibility for persons with a disability, as well as on gender equality*»).

7 Data Act and Data Governance Act

7.1. THE FRAMEWORK FOR VHT

The introduction of the Data Governance Act (DGA, applicable from September 2023) and the Data Act (that will apply as of September 2025) by the European Union is a fundamental step towards a robust data governance and data sharing infrastructure for the EU single data market, with significant implications for the VHT.

The DGA established a comprehensive legal framework to enhance data sharing and foster a trusted environment for the reuse of data across the Union. As part of the EU’s broader data strategy, it aims to unlock the value of data for innovation while ensuring robust safeguards for privacy, security, and proprietary information. In this sense, it supports the setup and development of common European Data Spaces in strategic domains, involving both private and public players, including health and research. Mechanisms are laid down to, *inter alia*, regulate access to data held by public sector bodies, create a framework for trusted **data intermediation services**, and encourage the formation of **data altruism** organizations, which facilitate voluntary data sharing for the public good. It also delineates clear roles for key stakeholders, such as public sector bodies, data holders and data users, ensuring transparency and accountability in data exchanges. Both personal and non-personal data are in scope of the DGA. By harmonizing rules on data governance and promoting the development of a secure and interoperable data economy, the DGA seeks to bolster innovation in areas such as health, artificial intelligence and smart manufacturing.

When it comes to a foundational element driving the development and functionality of the VHT, such as sharing and reusing personal health data, the DGA serves as the overarching framework, or the ‘genus’, while the EHDS represents a more tailored application, or the ‘species’, within this broader structure. Given that the EHDS establishes a more specific and detailed regulatory regime for health data, it takes precedence over the general provisions of the DGA. In essence, the EHDS can be viewed as the first sector-specific iteration of the DGA, designed to address the unique challenges and requirements associated with health data governance.

The DGA establishes a framework that includes various actors critical to its implementation, particularly concerning the reuse of data. Among these, public sector bodies, defined broadly as “*State, regional or local authorities, bodies governed by public law or associations formed by one or more such authorities, or one or more such bodies governed by public law*” (Art. 2(17)), play a foundational role as they are often tasked with granting or denying access to specific categories of data within the DGA’s scope.

They fall into the category of **data holders**, defined as any legal person, including public sector bodies and international organizations, or a natural person who is not the data subject with respect to the specific data in question, that, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data.

On the other hand, **data users** are natural or legal persons who have lawful access to certain personal or non-personal data, as well as the right to use such data for commercial or non-commercial purposes.

The most relevant players under the DGA are data intermediation service providers (so-called ‘**data intermediaries**’), namely entities that facilitate data sharing and access between data holders and data users. Specifically, these services are those which aim to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exerting the rights of data subjects in relation to personal data, excluding at least the following:

- services that aggregate, enrich, or transform data to add substantial value and license the resulting data to users, without creating a direct commercial link between the original data holders and data users;
- services that focus on mediating copyright-protected content;

- services exclusively enabling data sharing within a single data holder’s ecosystem or among a closed group of entities (*e.g.*, in Internet of Things (IoT) networks or contractual collaborations), as these do not involve open or neutral data-sharing facilitation;
- data-sharing services offered by public sector bodies that do not aim to create commercial relationships. These services are typically for public tasks rather than market-based intermediation.

Many companies are reasonably hesitant to share their data due to concerns about losing competitive advantage and the potential for misuse. To overcome this barrier, a complex framework of rules is established in the DGA (especially Art. 12) to ensure that data intermediaries operate as reliable and trustworthy facilitators of data sharing and pooling within the Common European Data Spaces (*e.g.*, the Health Data Access Bodies under the EHDS). To foster trust in data sharing, this approach emphasizes the **neutrality and transparency of data intermediaries** while empowering individuals and businesses to maintain control over their data.

In practice, these entities act as neutral third parties that facilitate connections between individuals or companies and data users. While they may charge fees for their intermediation services, they are prohibited from directly using the data they handle for financial profit, such as selling the data to another entity or developing proprietary products based on it. To ensure their neutrality and prevent conflicts of interest, data intermediaries must adhere to strict requirements under the DGA, including above all a clear **structural separation between the data intermediation service and any other service that they may provide**. For instance, the services provided as data intermediaries must be legally distinct from any other business activities, and commercial terms, such as pricing, must not depend on whether the data holder or user also utilizes additional services offered by the same intermediary. **Moreover, any data or metadata obtained during their operations can only be used to improve the quality of the intermediation service itself, not for external purposes.**

Both stand-alone organizations providing data intermediation services and multi-service companies offering intermediation alongside other activities can serve as trusted intermediaries. **In the case of the latter, the data intermediation function must be legally and financially separated from all other business operations, ensuring transparency and neutrality.**

To operate as data intermediaries under the DGA, providers are required to **notify the competent national authority** of their intention to offer such services⁹⁹. This notification process is designed to be fair, non-discriminatory, and free from anti-competitive practices. Once the competent authority confirms that the notification includes all the necessary information, the data intermediary is officially recognized and may begin operations.

Data intermediation services may involve various forms of data exchange, including **intermediation between data holders and potential users** through bilateral or multilateral data sharing, the creation of platforms or databases to enable data exchange or joint usage, or the establishment of infrastructures to connect data holders with users. They may also support **intermediation between natural persons**, such as data subjects or individuals seeking to share non-personal data, and potential data users. Additionally, data intermediation services can include **data cooperatives**, which are organizational structures composed of data subjects, one-person undertakings or SMEs, aimed at supporting its members in relation to specific aspects of data sharing and processing¹⁰⁰.

The types of data covered by the DGA include data protected under the following grounds:

- commercial confidentiality, encompassing business, professional, and company secrets;

⁹⁹ According to Art. 13 of DGA, each Member State must designate one or more competent authorities to carry out the tasks related to the notification procedure for data intermediation services and then notify the EU Commission of the identity of those competent authorities.

¹⁰⁰ Data Cooperatives support their members “*in the exercise of their rights with respect to certain data, including with regard to making informed choices before they consent to data processing, to exchange views on data processing purposes and conditions that would best represent the interests of its members in relation to their data, and to negotiate terms and conditions for data processing on behalf of its members before giving permission to the processing of non-personal data or before they consent to the processing of personal data*” (Art. 2(15) of DGA).

- statistical confidentiality;
- intellectual property rights belonging to third parties;
- personal data, provided such data are not within the scope of the Open Data Directive.

According to Art. 5 of DGA, public sector bodies authorized under national law to grant or deny access to the re-use of data categories are required to make publicly available the conditions for such re-use and the procedures to request it via a dedicated single information point. Member States must ensure that public sector bodies are provided with adequate resources to meet these obligations. **The conditions for re-use must be fair, transparent, proportionate, and objectively justified concerning the type and purpose of the data and must not be used to restrict competition.** Public sector bodies are also obligated to preserve the protected nature of the data in compliance with Union and national law. This may include requirements to ensure that:

- a) personal data are properly anonymized, and commercially sensitive information such as trade secrets or intellectual property are aggregated, modified, or otherwise safeguarded against risks of unauthorized disclosure;
- b) both the access and the re-use of data occur in secure processing environments controlled by the public sector body itself;
- c) both the access and the re-use of data occur within the physical premises in which the secure processing environment is located in accordance with high security standards, insofar the remote access cannot be allowed without jeopardizing third parties' rights and interests.

As also established in the EHDS, **data users shall be prohibited from re-identifying any data subject to whom the data relate and shall take technical and operational measures to prevent such singling-out.**

If reuse cannot be allowed due to legal or regulatory constraints, and there is no legal basis for data transmission under the GDPR, public sector bodies must make reasonable efforts to assist potential re-users in obtaining consent from the data subjects or permissions from data holders.

One of the main novelties introduced by the DGA is the concept of **data altruism**, defined as “*the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available, for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest*”. In brief, it is the **free choice to ‘donate’ the data for the common good.**

Member States may implement organizational or technical arrangements (or both), to facilitate data altruism, including establishing national policies to this end (to be notified to the EU Commission), to assist data subjects in making their data held by public sector bodies available for data altruism purposes, and set out the necessary information concerning the reuse of their data in the general interest.

The DGA establishes the framework for certain entities to qualify as **data altruism organizations**, which must meet specific criteria to be officially recognized. To obtain this designation, an entity must:

- i) be registered in the public register of data altruism organizations;
- ii) perform data altruism activities;
- iii) be established as legal entities under national law with objectives aligned with the general interest;
- iv) operate on a not-for-profit basis, and maintain legal independence from any for-profit entities;

- v) carry out the data altruism activities through a structure that is functionally separate from its other activities;
- vi) comply with some additional requirements to be laid down by the EU Commission, by means of delegated acts, in a dedicated rulebook.

Data altruism organizations are also required to uphold transparency and ensure that the rights and interests of data subjects and data holders are protected. This includes providing clear information on the purposes for which data will be used, ensuring such purposes align with general interest objectives, and implementing robust methods for obtaining data altruism consent. To support the collection of data for altruistic purposes, the EU Commission is tasked with adopting implementing acts to introduce a **European data altruism consent form**. This standardized form will facilitate consent collection, ensuring clarity and consistency across Member States while upholding the principles of trust and data protection.

Clearly, in parallel with the secondary processing framework offered by the EHDS, data altruism can play a significant role in addressing data availability challenges which are inherent in developing and deploying the VHT. Through data altruism organizations, individuals and entities can voluntarily share personal and non-personal data for general interest objectives, such as improving disease modelling, optimizing treatment protocols, or advancing public health outcomes. To ensure trust, data altruism organizations are required to operate under strict transparency and accountability standards. At the same time, by prioritizing security and data protection, the DGA helps build confidence among data subjects and stakeholders, encouraging wider participation in data altruism initiatives. In the context of research, the data shared through altruistic mechanisms can be aggregated and anonymized to enhance the development of predictive models through Digital Twins-driven models without compromising individual privacy. At a more general level compared to the EHDS, the DGA establishes a pathway for ethical and effective data sharing that may support scientific discovery, improve healthcare outcomes, and drive the development of next-generation medical technologies and policies.

Finally, this Regulation establishes the **European Data Innovation Board**, an advisory body in charge of fulfilling a large number of tasks (Art. 30 of DGA), including supporting the EU Commission in developing EU-wide consistent best practices, in particular on data intermediation, data altruism and the use of public data that cannot be made available as open data, as well as on the prioritization of cross-sectoral interoperability standards. The Board includes representatives from the following entities: (a) Member States' competent authorities for data intermediation, (b) Member States' competent authorities for data altruism; (c) the European Data Protection Board; (d) the European Data Protection Supervisor; (e) the European Union Agency for Cybersecurity; (f) the EU Commission; (g) the EU SME Envoy/representative appointed by the network of SME envoys¹⁰¹; (h) other representatives of relevant bodies selected by the Commission through a call for experts.

* * * * *

The Data Act is a complementary piece of legislation to the DGA and represents another pillar of the European data strategy. While the latter Regulation focuses – as seen above – on strengthening trust and increasing the availability of data, also thanks to secure intermediaries and data altruism organizations, the Data Act aims at enhancing the data economy by establishing clear rules on data access and use. It seeks to ensure fairness in the digital environment, stimulate data-driven innovation, and promote a competitive data market.

The Data Act empowers users of connected products, whether businesses or individuals that own, lease, or rent such products, to have greater control over the data they generate, while preserving incentives for entities that invest in data technologies. It also establishes general conditions governing cases where a business is legally required to share data with another business.

¹⁰¹ More information is available here: https://single-market-economy.ec.europa.eu/smes/sme-strategy-and-sme-friendly-business-conditions/sme-envoys-network_en.

By clarifying who can use what data and under which conditions, the Data Act aims to foster technological advancement and increase data availability, ensuring fairness in the allocation of data value among all actors in the data economy, thereby enhancing the EU's competitiveness in the global digital landscape. More in detail, it establishes harmonized rules regarding:

- granting users access to **data generated through the use of a connected product¹⁰² or related service¹⁰³**;
- facilitating the **sharing of data from data holders to data recipients**;
- enabling the **transfer of data from data holders to public sector bodies** or EU institutions, agencies, or bodies, when an exceptional need arises for performing tasks in the public interest (business-to-government).

Beyond data holders, recipients and users, the Regulation also applies to **manufacturers of connected products** and to the suppliers of related services placed on the market. In essence, they need to comply with a number of technical requirements aimed enable data sharing under the Data Act: “*Connected products shall be designed and manufactured, and related services shall be designed and provided, in such a manner that product data and related service data, including the relevant metadata necessary to interpret and use those data, are, by default, easily, securely, free of charge, in a comprehensive, structured, commonly used and machine-readable format, and, where relevant and technically feasible, directly accessible to the user*” (so-called principle of ‘access-by-design’). In the VHT ecosystem, e.g., medical devices manufacturers fall in this category.

If users cannot directly access data from a connected product or related service, data holders are required to provide prompt access to the data, along with the necessary metadata for their interpretation and use. This access must be of the same quality as available to the data holder, as well as provided easily, securely, and free of charge. The data must be delivered in a comprehensive, structured, widely used, and machine-readable format. Where applicable and technically feasible, access should also be continuous and in real-time.

The definition of ‘data holder’ is not aligned with that established under the DGA and is very different from the one provided for by the EHDS¹⁰⁴. This is likely to create legal uncertainty between stakeholders.

The Data Act applies to all raw and pre-processed data generated through the use of a connected product or related service that are readily available to the data holder (such as the manufacturer of the product or the service provider). This refers to data that can be accessed with reasonable effort, beyond just simple operations. Both personal and non-personal data, along with relevant metadata, are in scope. This includes data collected from individual sensors or interconnected groups of sensors, such as measurements of temperature, pressure, flow rate, audio levels, pH values, liquid levels, position, acceleration, or speed. However, **inferred or derived data do not fall under the application of this Regulation.**

For example, thinking about a connected glucose monitor used by patients to manage diabetes, this means that the data directly generated from the device, such as glucose levels, timestamps of readings, or device battery status, would fall within the scope of the Data Act (including necessary metadata such as calibration settings or sensor status). By contrast, derived data such as a complete diagnostic report predicting long-term trends in the patient’s health based on glucose readings or enriched datasets that

¹⁰² ‘Connected product’ is defined as “*an item that obtains, generates or collects data concerning its use or environment and that is able to communicate product data via an electronic communications service, physical connection or on-device access, and whose primary function is not the storing, processing or transmission of data on behalf of any party other than the user*” (Art. 2(5)). Examples include connected cars, medical and fitness devices, industrial or agricultural machinery, etc.

¹⁰³ ‘Related service’ is defined as “*a digital service, other than an electronic communications service, including software, which is connected with the product at the time of the purchase, rent or lease in such a way that its absence would prevent the connected product from performing one or more of its functions, or which is subsequently connected to the product by the manufacturer or a third party to add to, update or adapt the functions of the connected product*” (Art. 2(6)), i.e., anything that would make a connected product behave in a specific manner, such as an App to adjust the brightness of lights, or to regulate the temperature of a fridge, or to operate a gym product.

¹⁰⁴ ‘Data holder’ means “*a natural or legal person that has the right or obligation, in accordance with this Regulation, applicable Union law or national legislation adopted in accordance with Union law, to use and make available data, including, where contractually agreed, product data or related service data which it has retrieved or generated during the provision of a related service*” (Art. 2(13 of Data Act).

combine glucose data with other sources (*e.g.*, dietary intake logs or physical activity records), would fall outside the scope. Similarly, proprietary algorithms used to generate advanced health predictions from raw data would remain protected under intellectual property laws and would not be subject to the Data Act's provisions.

If the user wishes to share these data with another entity or individual, they can either do so directly or they can ask the data holder to share it with a third party of their choice (excluding gatekeepers under the Digital Markets Act)¹⁰⁵. The data holder must have a contract in place with the user (*e.g.*, sales contract, rental contract, related service contract, etc.) defining the rights regarding the access, use and sharing of the data that is generated by the connected product or related service.

To ensure businesses are not discouraged from investing in data-generating products, the data obtained cannot be employed by data users to create a competing connected product. However, the Data Act does not restrict competition in related or aftermarket services.

At the same time, in order to safeguard trade secrets while supporting the Data Act's objective of increasing data availability, the data holder and the user (or the third party indicated by the latter) can agree on specific measures to maintain the confidentiality of trade secrets: if such measures are violated, the data holder is entitled to withhold or suspend data sharing. However, the data holder may lawfully refuse to share data only if it can demonstrate a significant likelihood of incurring serious economic harm due to the disclosure of its trade secrets.

Terms unilaterally imposed on a take-it-or-leave-it basis, particularly those related to making data available, may be subject to an unfairness test under the Data Act. The regulation provides a non-exhaustive list of terms that are always deemed unfair, such as clauses that exclude or limit the liability of the party imposing the term for intentional acts or gross negligence. It also outlines terms presumed to be unfair, such as those that unduly restrict remedies for non-performance of contractual obligations, limit liability for breaches, or unjustly extend the liability of the party upon whom the term is imposed.

If a term is found to be unfair, it becomes invalid and, where feasible, is severed from the contract without affecting the remaining provisions. For terms presumed to be unfair, the party that imposed them has the opportunity to demonstrate that the term is fair and justified, thereby avoiding its invalidation.

Standards and interoperability are critical for enabling the seamless use of data from various sources within and across Common European Data Spaces, supporting research and the development of new products and services. To achieve this, the Data Act establishes essential requirements for participants in data spaces, with further specifications to be developed by the European Commission through delegated acts. These requirements aim to address the so-called 'data lock-in' problem, facilitating data flows and ensuring that data structures, formats and vocabularies are publicly accessible, where available. Participants must also enable the interoperability of data-sharing agreements, including mechanisms like smart contracts.

The Data Act emphasizes the importance of interoperability between data processing services, which is crucial for allowing customers to easily switch between different providers with ease. It prepares the groundwork for harmonizing standards and developing open interoperability specifications, ensuring that systems can effectively interact across data spaces.

The European Commission is tasked with identifying barriers to interoperability and prioritizing standardization needs. Where necessary, the Commission may request European standardization organizations to draft harmonized standards that meet the requirements of the Data Act. If these efforts fail to produce adequate standards, the Commission may adopt common specifications as an alternative, ensuring their development is inclusive and reflective of feedback from the European Data Innovation Board. These measures aim to create a robust and interoperable data ecosystem across the EU.

¹⁰⁵ Gatekeepers under the DMA are listed here: https://digital-markets-act.ec.europa.eu/gatekeepers_en.

7.2. BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

The Data Governance Act and the Data Act are designed to enhance data sharing and innovation and aim to establish a trustworthy framework for data exchange, including in the healthcare sector. The achievement of such ambitious objectives is pursued by imposing a wide range of both legal, procedural and technical requirements that stakeholders must navigate. For instance, the DGA imposes conditions to ensure that data sharing does not compromise privacy or intellectual property rights, which could necessitate additional safeguards when it comes to the development and deployment of VHT. Similarly, the Data Act’s provisions on data accessibility and fairness may require medical devices manufacturers and healthcare organizations to adjust their data management practices.

On account of their rationale, both these regulations stand as enablers rather than barriers: by establishing clear rules and fostering trust in relation to the sharing of protected data or data generated through connected products and related services, they create a more favourable environment for the adoption of the VHT. Moreover, the emphasis on interoperability and standardization further supports the integration of such technologies into existing systems.

It is crucial to consider, in addition to the above, that the EHDS serves as a specialized framework of rules within the overarching structure established by the DGA, providing tailored regime specifically for health data. In essence, the EHDS functions as the first sector-specific implementation of the DGA, designed to meet the particular needs of data management for primary and secondary purpose in the healthcare sector. **For this reason, to identify the main barriers relating to data governance, reuse and interoperability in connection with the generation, integration and implementation of the VHT, reference must be made to the regulatory hurdles and dependencies described in Section 5.2.** Likewise, the analysis of the regulatory issues surrounding the concepts of data anonymization and pseudonymization, which are often referred to in both the DGA and the Data Act, is set out in Section 3 of this document.

A) Better alignment of definitions in practice

Barrier

The definitions of ‘data holders’ provided for, respectively, in the DGA, the Data Act and the EHDS, are not perfectly aligned. The relevant variations reflect the distinct scopes and objectives of each regulation but determine in practice the possibility of inconsistent interpretations and applications. Considering the critical role of the ‘data holder’ concept across all three frameworks, such divergences risk generating **legal uncertainty and operational challenges**. This lack of uniformity could result in fragmented approaches to data governance, particularly when entities operate under multiple regulatory frameworks, as in the case of the VHT (*e.g.*, medical devices manufacturers).

Proposed action

- **By the European Data Innovation Board and the European Health Data Space Board:** cast some light, by means of guidance or specific provisions, as to how the definition of ‘data holder’ outlined in the DGA and Data Act must be construed in the light of the more complex one provided under the EHDS, and (even more) vice versa. **Harmonizing these definitions or providing clear, detailed guidance on how the term should be interpreted in different regulatory contexts** would be essential to mitigate the risks of legal ambiguity, fostering clarity and enhancing the effectiveness of the EU regulatory data strategy in the health domain.

8 Clinical Trials Regulation

8.1 THE FRAMEWORK FOR VHT

Digital twins have introduced a transformative change in clinical trials, reshaping how studies are conducted, and patient care is approached. By generating virtual patient models that accurately simulate real individuals or populations, Digital Twins allow researchers to virtually test treatments and interventions, leading to significant time and resource savings. These models harness predictive analytics, integrating data from multiple sources to predict treatment outcomes, identify potential risks, and tailor interventions for each patient. Furthermore, the ability to remotely monitor patients and gather real-time data through wearables and sensors enhances the precision and efficiency of trials.

Digital twins also support adaptive trial designs, enabling researchers to make real-time adjustments based on new data, which helps optimize patient outcomes. With improved safety monitoring, cost reductions, and shorter trial durations, digital twins have paved the way for a new era of precision medicine and more efficient clinical trials¹⁰⁶.

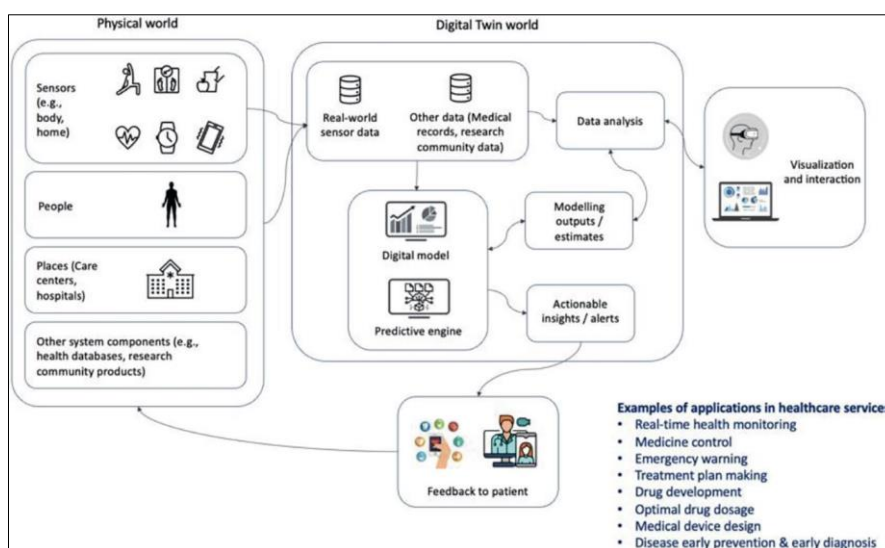


Figure 15 - High-level component view of a DT for precision medicine¹⁰⁷

The availability of biomedical data has expanded considerably, thanks to diverse sources such as large biobanks, electronic health records, medical imaging, wearable devices, biosensors, and affordable genome and microbiome sequencing. This increased data accessibility has been essential in advancing health DT solutions, allowing researchers to predict how a drug or intervention will behave under various scenarios. Additionally, data collected from personal digital devices, along with patient-generated health information – such as self-reported symptoms, physical markers, demographic details, and lifestyle data over time – further contributes to the creation of more comprehensive and accurate DTs.

Although the Clinical Trial Regulation (CTR) no. 536/2014 does not explicitly mention or apply to Digital Twins as standalone entities, its provisions may apply to the use of these technology under certain circumstances, namely **when Digital Twins are used within or alongside interventional clinical trials**, e.g., to inform drug development, predict outcomes, or support regulatory submissions.

¹⁰⁶ Armeni, P.; Polat, I.; De Rossi, L.M.; Diaferia, L.; Meregalli, S.; Gatti, A. Digital Twins in Healthcare: Is It the Beginning of a New Era of Evidence-Based Medicine? A Critical Review. *J. Pers. Med.* 2022, 12, 1255. <https://www.mdpi.com/2075-4426/12/8/1255>.

¹⁰⁷ Armeni, P., Polat, I., De Rossi, L. & Diaferia, L., Visioli, G., Meregalli, S., Gatti, A.. (2023). Digital Twins for Health: Opportunities, Barriers and a Path Forward. <http://doi.org/10.5772/intechopen.112490>

In all these cases, the CTR's requirements would need to be abided, including adherence to ethical guidelines, data integrity, and patient safety standards.

Clearly, full replacement of human trials with fully AI-driven technologies is not yet within the scope of the CTR. Nonetheless, these latter models serve as complementary tools that can be used to refine trial designs, select patient populations, and reduce the number of participants needed in some cases.

The CTR would be relevant to the VHT, *inter alia*, in case of:

- **Direct involvement in clinical trials**

If DTs are integrated as part of the design or execution of a clinical trial process for testing and validating new drugs or therapies. This may include scenarios where the VHT is employed to simulate patient responses to treatments, optimize trial protocols, or predict safety and efficacy outcomes (*e.g.*, if a Digital Twin helps determine dose ranges or patient subgroups for a trial). In these cases, sponsors must ensure that VHT models are validated and that any data derived from them is consistent with the CTR's requirements for scientific integrity, data reliability, and reproducibility¹⁰⁸.

- **Data use and supplementary evidence**

The CTR allows for the submission of supplementary data alongside traditional clinical trial results, primarily to support regulatory decision-making and to provide a more comprehensive understanding of the medicinal product's safety, efficacy, and quality. This means the VHT can be leveraged to provide additional insights and predictive models, simulate scenarios, or support findings from traditional trials, enhancing the overall data package submitted for regulatory review (*e.g.*, helping to demonstrate safety, efficacy, or dosing patterns).

- **Adaptive trial designs and real-time adjustments**

The CTR encourages adaptive trial designs, where ongoing data allow researchers to make real-time adjustments to the trial. The VHT can prove particularly helpful in this context, by continuously analysing patient data and simulating responses, thus enabling trial adjustments without waiting for real-world results. Evidently, when DTs are used to inform real-time decisions, the application of the CTR ensures that these models are scientifically validated and any adjustments they prompt are ethically and legally sound. Researchers must therefore demonstrate that the relevant outputs are reliable, particularly when making decisions that impact patient safety.

- **Post-marketing surveillance**

Digital Twins can be used in Phase IV of the clinical trials for post-marketing surveillance to monitor long-term safety and effectiveness. The CTR provisions for continuous safety monitoring apply here, ensuring that the VHT helps simulate patient outcomes, track adverse effects, or predict long-term risks. Moreover, if this technology is employed to monitor a drug's post-market performance, it must be integrated into the overall safety management system required by the CTR, ensuring that data generated from VHT models are accurate, traceable, and transparent, and that any predictive analytics or simulations are used ethically to safeguard ongoing patient safety.

While setting out a harmonized and streamlined regulatory framework to facilitate approving, conducting and supervising clinical trials across Member States, the CTR only applies to interventional clinical trials that involve testing medicinal products intended for human use, such as drugs, to assess their safety, efficacy, and quality (this includes phase I to IV trials, ranging from early-stage trials in humans to post-marketing studies). Observational studies focusing on real-world evidence, epidemiological research, or post-market surveillance, where no specific treatments or interventions (such as new drugs or therapies) are administered, fall out of the scope of the CTR.

¹⁰⁸ Vidovszky AA, Fisher CK, Loukianov AD, Smith AM, Tramel EW, Walsh JR, Ross JL. Increasing acceptance of AI-generated digital twins through clinical trial applications. *Clin Transl Sci.* 2024 Jul;17(7):e13897. <https://doi.org/10.1111/cts.13897>.

Crucially, the Regulation aims to make it easier to carry out cross-border trials within the EU, by providing a single electronic portal – named Clinical Trials Information System (‘CTIS’) – for trial applications, approvals and management. This helps to simplify the process, especially for trials conducted across multiple Member States.

Listed below are some of the main provisions of the CTR which may be relevant for the purposes of this D6.2, in all cases when DTs are integrated into the clinical trial process:

a) Centralized application and approval process

When DTs are used as part of a clinical trial’s design or data analysis, precise information must be provided within the trial application to be submitted in the CTIS regarding how the relevant models will be integrated and validated. **This ensures that the use of Digital Twin technologies is clearly documented and evaluated during the trial approval process.**

b) Ethical screening

Ethical standards are a core aspect of the CTR, requiring that all clinical trials be approved by competent ethics committees before proceeding. If DTs are employed to simulate patient responses or predict adverse effects, these digital models must undergo ethical review to ensure they do not compromise patient safety or the integrity of the study.

c) Safety monitoring

DTs can prove helpful in continuous safety monitoring throughout the trial by providing real-time simulations and predictions. However, **these applications must align with the risk-based monitoring approaches defined in the CTR, ensuring that they enhance, rather than replace, traditional safety measures.**

d) Data integrity and transparency

The CTR emphasizes the importance of data integrity and requires that all data used in clinical trials be reliable, reproducible, and transparent. Where DTs are used for predictive modelling or supplementary analyses, they must generate results that need to be traceable and verifiable to meet these standards. Additionally, it is required that any data generated by, or used to build, VHT models must be made available to regulatory authorities and, where appropriate, to the public. This ensures that Digital Twins contribute to the trial’s dataset in a way that is consistent with regulatory expectations for transparency.

e) Informed consent

Informed consent represents one of the fundamental principles underpinning the EU legislation on clinical trials. Accordingly, should the VHT be used, *e.g.*, in the trial design phase to simulate patient outcomes, or to analyse data from real patients, then the latter must be fully informed about how their data will be used, including the role of digital technologies in the study and the relevant impacts.

f) Use of innovative trial designs¹⁰⁹

The CTR includes many provisions that promote flexibility and efficiency in the clinical trial process, encouraging new methodologies that can adapt to real-time data. In this sense, the CTR acknowledges the need for innovative and adaptive trial designs, which open up opportunities to integrate DT technology as part of a hybrid trial strategy, to help design trials by identifying patient subgroups, simulating various treatment scenarios, or optimizing trial parameters. **When used to this end, DTs must be validated and scientifically justified as part of the**

¹⁰⁹ This approach is also promoted in the ‘Pharmaceutical Strategy for Europe’ adopted by the EU Commission in November 2020 (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0761>), whereby it is stated that the CTR needs to “address new developments such as adaptive and complex trials, and the use of in-silico techniques and virtual approaches”.

clinical trial application to ensure that the model-based approaches meet the same standards as traditional trial methods¹¹⁰.

g) Risk management and continuous monitoring

The VHT can be particularly valuable for risk management, by using it to simulate potential adverse effects before they occur in real patients, allowing to proactively address safety concerns. The CTR's emphasis on continuous safety monitoring can be enhanced with Digital Twin technology, provided these digital tools are validated and integrated into the risk management plan (Annex I of the Regulation, which lists the required documentation for clinical trials, implies that any novel methodologies, including Digital Twins models, must be sufficiently detailed and justified to ensure that they do not introduce risks to participants).

Digital Twin technologies hold the promise of revolutionizing connected care by transforming how lifestyle, health, wellness, and chronic diseases are managed. They have the potential to enable more personalized and proactive healthcare, offering real-time monitoring, predictive analytics, and tailored treatment strategies. However, the full potential of Digital Twins in the healthcare sector has yet to be realized due to significant barriers. These include technical challenges, such as data integration and model validation; regulatory hurdles, particularly around data privacy and the lack of clear guidelines for their use; and ethical concerns related to patient autonomy and consent. Addressing these obstacles will be essential to unlock the transformative capabilities of Digital Twins in the future.

The EMA has on some occasions recognized the potential of digital methodologies and computational modelling, by establishing 'Innovation Task Force' to facilitate the uptake of innovative methods in the process for development and approval of medicines. It has also published some guidance, such as:

- 'Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation'¹¹¹;
- 'Qualification of novel methodologies for drug development: guidance to applicants'¹¹².

Nonetheless, this is far from being sufficient to set out the needed standards and procedures.

That being said, without prejudice to the other findings detailed in the Chapters above, some of the major barriers which emerge to date in connection with the use of the VHT in clinical trials are summarized below:

✓ **Lack of specific regulatory framework**

As already specified, the CTR does not explicitly address the use of Digital Twins and *in-silico* technologies, leading to significant regulatory uncertainty. While the regulation governs traditional interventional trials, **there are no specific guidelines on how DTs can be integrated, validated, or accepted as part of the trial process.** Other relevant legislations as well, such as those for medicinal products (Regulation 726/2004) and medical devices (MDR/IVDR), fail in turn to comprehensively address these digital solutions. **This lack of clarity makes it difficult for companies to know how to justify and present the use of Digital Twins when submitting clinical trial applications, and it may discourage their use due to concerns about regulatory acceptance.**

✓ **Regulatory acceptance of virtual evidence**

One of the core barriers under the CTR is the **acceptance of virtual evidence generated by computer models.** Indeed, even if they can accurately simulate patient outcomes, the evidence generated from these simulations may not be considered as robust as data from traditional trials. Regulators need clear criteria on how to assess and interpret data from Digital Twins. **Without**

¹¹⁰ In this respect, see Katsoulakis, E., Wang, Q., Wu, H. *et al.* Digital twins for health: a scoping review. *npj Digit. Med.* 7, 77 (2024). <https://www.nature.com/articles/s41746-024-01073-0>.

¹¹¹ https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpc-modelling-and-simulation_en.pdf

¹¹² https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf

well-established and homogeneous frameworks, the risk exists that evidence from Digital Twins may be viewed with scepticism, making it harder for companies to use this technology as a core part of their trial designs. Therefore, there is an urgent need to further develop uniform requirements for acceptance of evidence generated via digital models, and to set out harmonized standards and best practices to ensure that Digital Twins can be integrated into the clinical trial process all across the EU. Without such cohesion and consistency, there is a risk of fragmented approaches that could hinder cross-border collaboration and innovation.

✓ **Validation and standardization challenges**

A key requirement under the CTR is that data used in trials must be reliable, reproducible, and scientifically validated. With Digital Twins, it can be challenging to ensure that the models and simulations they generate meet these criteria. To date, **there is no standardized approach to validating Digital Twin models, which means that the methods used to build and test them might vary widely.** Accordingly, clear guidance and standardized practices for the verification and validation of these models need to be defined and adopted.

✓ **Voluntary Scientific Procedures**

Although there are voluntary guidelines and scientific procedures available (such as those set by the European Medicines Agency in connection to the so-called ‘qualification process’) to establish the regulatory acceptability for novel methodologies, these are not mandatory or standardized across the EU. **They offer some support for companies wanting to use Digital Twins and *in-silico* models, but do not provide the same level of regulatory certainty as the established processes for traditional trials.** Stakeholders might use these procedures to seek early feedback from regulators on their digital models, but without a formal framework in place, the path to regulatory acceptance remains unclear, inconsistent, and case-by-case.

✓ **Data integration and interoperability issues**

Digital Twins rely on data from multiple sources, such as electronic health records, wearable devices, lab tests, and more. Ensuring that this data can be integrated seamlessly and used to create accurate digital models is challenging, especially under the strict data integrity requirements of the CTR. In this sense, as already highlighted for the EHDS, **the lack of interoperable data standards across the EU further complicates this issue, making it difficult to aggregate and use data from different systems or countries.** This is particularly problematic for multi-national trials, which the CTR aims to facilitate.

✓ **Ethical concerns and informed consent**

The CTR poses very strict requirements for informed consent, to ensure that trial participants are aware of how their data will be used. When Digital Twins are involved, particularly if they use real-world patient data to generate simulations, there are ethical concerns about data privacy, patient autonomy, and consent. Individuals must be informed that their data will be used to create digital simulations, and it may not always be clear how these models will be used over the course of the study.

✓ **Need for harmonization across Member States**

While the CTR aims to harmonize clinical trial processes across the EU, implementation and interpretation can still vary between Member States. This lack of uniformity can create complications for trials involving or integrating this technology, as companies may need to address different (and so cost-ineffective) regulatory expectations in each country. Harmonized guidelines that explicitly include Digital Twin technology would help ensure that companies can easily conduct multi-national trials without facing inconsistent regulatory requirements.

8.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

Digital Twins – and more generally in-silico trials – in healthcare offer significant benefits under the Clinical Trial Regulation by enabling virtual simulations of patient responses, which can reduce the need for large physical control groups, enhance adaptive trial designs, and improve efficiency in drug development. By providing predictive models and real-time data integration, the VHT supports more personalized and cost-effective trials while maintaining high standards of safety and data integrity, aligning with the regulatory requirements for robust and reliable evidence.

Despite their enormous potential, regulatory uncertainty and the lack of specific guidelines pose significant barriers to the full exploitation of Digital Twins under current EU clinical trial and connected pharmaceutical legislation.

A) New guidelines to establish uniform V&V criteria and clear regulatory pathways

Barriers

The CTR does not specifically regulate digital technologies used in connection with clinical trials, such as for *in-silico* applications including Digital Twins, as opposed to traditional *in vivo* (in living organisms) or *in vitro* (in a lab environment) trials. This generates uncertainty for the VHT ecosystem because **there is no clear, established framework on how to justify and present the use of Digital Twins when submitting clinical trial applications and to present data from these digital models in a sound way that regulators can accept.**

For traditional trials, the CTR outlines a well-defined pathway for gaining approval. This includes submitting data that meets specific standards, undergoing a centralized review process (via the CTIS), and following structured safety and efficacy guidelines. By contrast, digital technologies do not have this kind of clear regulatory pathway, which is why stakeholders in the VHT ecosystem must face questions about how to validate the relevant models, how to demonstrate that simulations can accurately predict real-world patient outcomes, and how to present this evidence as part of a marketing authorization application.

Given that these technologies rely on computational models that simulate biological processes, competent authorities – primarily the European Medicines Agency – need to be convinced of the reliability, reproducibility, and robustness of both models employed and data generated through Digital Twins. Notwithstanding this, there is no standardized approach for validating the software and algorithms used, as well as data and models, making regulatory acceptance challenging.

To properly overcome this barrier, **operators need more clarity about the applicable legal pathways and on how to present evidence from the use of Digital Twins during the drug development and approval process.** Indeed, while some voluntary scientific procedures exist, there is no established framework akin to traditional trials, which complicates the process of gaining regulatory acceptance.

Likewise, the existing pharmaceutical legislation does not fully accommodate digital innovation, including the use of novel technologies in clinical trials. **This creates a mismatch between the pace of technological advances and the regulatory environment, delaying the integration of Digital Twins related and, more generally, *in-silico* methodologies.**

Proposed actions

- **By the European Medicines Agency:** issue guidelines that specifically address the use of Digital Twins technologies in clinical trials, **laying down clear criteria, applicable at an EU-wide level, on validation, data integration, and model reliability**, to help companies understand how to effectively incorporate these models and simulations into the trial process, ensuring regulatory compliance. Such guidelines should cover aspects like **model verification and validation, ensuring that their predictions are reliable, reproducible, and scientifically credible, as well**

as including standards to demonstrate that a Digital Twin accurately predicts real-world patient responses, in a robust and scientifically validated manner.

- **By the European Medicines Agency:** build on the framework provided by the ASME VV-40 in the area of medical devices (which provides guidelines for assessing the credibility of computational models used in the design, evaluation, and regulatory approval of medical devices)¹¹³, to set a **European standardized approach for validating Digital Twins models, defining how to verify their accuracy, validate their predictive capabilities, and certify their integration into the clinical trial process**, thus supporting wider regulatory acceptance and reinforcing trust among regulators, clinicians, and patients.
- **By the European Medicines Agency:** given that the EMA already offers scientific advice for novel trial methodologies, creating a **regulatory sandbox** would be a natural extension of its role and tasks. This would provide a controlled environment where companies may **test innovative technologies like the VHT, helping to assess validation standards and allowing for iterative testing and direct feedback from regulators**. Similarly, **pilot programs** and collaborative **mock submission schemes** could demonstrate the practical benefits and potential of Digital Twins models, offering case studies to guide future regulation and promote adoption across the EU.

B) Enhance data integration and interoperability

Barriers

Digital Twins in clinical trials must integrate with real-world patient data from various sources, including pre-clinical data, electronic health records, and real-world data gathered through biosensors, and wearable devices. Nonetheless, **there is no standardization across the EU to ensure that all this data can be effectively and securely integrated into computational models** (as in ASME VV-40). Ensuring that the data used in connection with the VHT are **consistent and interoperable** across different systems and countries remains a major hurdle. Without standardized data formats and clear guidelines, companies face difficulties in pooling data, which is essential for robust digital simulations.

Proposed action

- **By the EU Commission:** this barrier is common to the EHDS. Establishing common data standards across the EU is essential for ensuring that data from different systems can be seamlessly integrated into Digital Twins. This involves setting standard formats, terminologies, and protocols for data sharing, particularly in healthcare. The EHDS (Art. 6) requires the EU Commission to set out the **European Electronic Health Record Exchange Format (EEHRxF)**¹¹⁴, which “*shall be commonly used, machine-readable and allow transmission of personal electronic health data between different software applications, devices and healthcare providers. The format should support transmission of structured and unstructured health data*”. The development of the *EEHRxF* is also supported by collaboration between Member States and entities like the eHealth Network (a voluntary network that connects national authorities responsible for eHealth designated by the Member States), which facilitates coordination to ensure national health systems adopt and integrate these interoperable standards. Well-known approaches like the **Fast Healthcare Interoperability Resources (FHIR)** and **Health Level Seven (HL7)** can serve as frameworks to promote data standardization.

¹¹³ <https://www.asme.org/codes-standards/find-codes-standards/assessing-credibility-of-computational-modeling-through-verification-and-validation-application-to-medical-devices>. While ASME VV-40 is not a regulatory requirement in itself, it has been recognized as a best practice for model validation by regulatory bodies like the FDA. Its principles can also be used to align with EU regulations on clinical trials and medical devices, ensuring that Digital Twin models are validated in a way that meets international standards.

¹¹⁴ Commission Recommendation (EU) 2019/243 of 6 February 2019 on a European Electronic Health Record exchange format (OJ L 39, 11.2.2019, p. 18). <https://eur-lex.europa.eu/eli/reco/2019/243/oj>.

9 Medical Device Regulation / In Vitro Diagnostic Regulation

9.1 THE FRAMEWORK FOR VHT

The ‘twin’ EU Regulations 2017/745 (Medical Device Regulation (MDR)) and 2017/746 (In Vitro Diagnostic medical device Regulation (IVDR)) establish comprehensive legal frameworks to ensure the quality, safety, performance, security and efficacy of medical devices and in vitro diagnostic medical devices within the European Union. Both regulations explicitly extend their scope to include software, including stand-alone applications, provided that some specific conditions are satisfied.

Therefore, on the account of their nature, Digital Twins¹¹⁵ may qualify as medical devices under the MDR or IVDR, if their intended purpose and use align with the perimeter set forth in Article 2 of the respective regulations¹¹⁶. Very briefly (and hyper-simplifying), **a digital twin may be classified as a ‘Software as medical device’ under the MDR if it (i) is intended by the manufacturer to be used for a specific medical purpose¹¹⁷; (ii) has the capability to edit and process the data that it collects, thus autonomously providing outputs different from input data; (iii) is used on humans, extending this concept to all cases in which the software provides information that is subsequently used to make decisions for diagnostic or therapeutic purposes (i.e., software can be a medical device notwithstanding the fact that it does not itself act in or on the human body)¹¹⁸**. Similarly, according to the IVDR, a Digital Twin may qualify as an in vitro diagnostic device if it is intended by its developer to be used, alone or in combination with other ‘components’, for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information relevant to the diagnosis or monitoring of medical conditions¹¹⁹.

The MDR introduced major changes to the legal regime applicable to medical devices, with the aim of ensuring a high level of patient safety and increased transparency in the use of medical devices within the European single market. The key objectives of the MDR include improving the quality, safety and reliability of medical devices; ensuring the transparency and traceability of devices within the EU; strengthening post-market surveillance and clinical evaluation of medical devices; and establishing strict criteria for the classification of devices and their conformity assessment.

The main actors envisaged by the MDR include:

- ✓ the ‘manufacturer’, namely the natural or legal person responsible for the design and marketing of the device under its own name or trademark¹²⁰. It must ensure that devices comply with safety and performance requirements, carry out a clinical evaluation and draw up the necessary technical documentation, affix the CE mark, assign a UDI code to the device and register it in the European database on medical devices (‘Eudamed’)¹²¹, while post-market obligations include adopting procedures for reporting incidents and monitoring the device over time.

¹¹⁵ In this section the focus is on individual Digital Twins. A separate discussion will be required on the VHT infrastructure itself

¹¹⁶ Both the MDR (Recital 19) and the IVDR (Recital 17) indicate that: “*It is necessary to clarify that software in its own right, when specifically intended by the manufacturer to be used for one or more of the medical purposes set out in the definition of*” a medical device or, respectively, of an in vitro diagnostic medical device, “*qualifies as a medical device*” or in turn “*as an in vitro diagnostic medical device, while software for general purposes, even when used in a healthcare setting, or software intended for well-being purposes is not a medical device*”, or “*an in vitro diagnostic medical device*”. In any case, “*The qualification of software, either as a device or an accessory, is independent of the software’s location or the type of interconnection between the software and a device*”.

¹¹⁷ Art. 2(1) of the MDR identifies the following medical purposes: “- *diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease; - diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability; - investigation, replacement or modification of the anatomy or of a physiological or pathological process or state; - providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations*”.

¹¹⁸ K Minssen T, Mimler M, Mak V. When Does Stand-Alone Software Qualify as a Medical Device in the European Union?-The CJEU’s Decision in Snitem and What it Implies for the Next Generation of Medical Devices. *Med Law Rev.* 2020 Aug 1;28(3):615-624. <https://doi.org/10.1093/medlaw/fwaa012>.

¹¹⁹ Art 2(2) of the IVDR specifies that the information provided by the in vitro diagnostic medical device must: (a) concern a physiological or pathological process or state; (b) concern congenital physical or mental impairments; (c) concern the predisposition to a medical condition or a disease; (d) determine the safety and compatibility with potential recipients; (e) predict treatment response or reactions; (f) define or monitor therapeutic measures.

¹²⁰ Art. 2(30) of the MDR defines the ‘manufacturer’ as “*a natural or legal person who manufactures or fully refurbishes a device or has a device designed, manufactured or fully refurbished, and markets that device under its name or trademark*”.

¹²¹ <https://webgate.ec.europa.eu/eudamed/landing-page#/>.

- ✓ ‘Importers’ (“any natural or legal person established within the Union that places a device from a third country on the Union market”, Art. 2(33)) and ‘distributors’ (“any natural or legal person in the supply chain, other than the manufacturer or the importer, that makes a device available on the market, up until the point of putting into service”, Art 2(34)) are responsible for verifying the compliance of devices with the MDR; they must cooperate with the competent authorities to mitigate the associated risks and ensure the presence of the correct CE marking.
- ✓ The ‘authorized representative’ is the natural or legal person appointed by the non-EU manufacturer to act on its behalf in the EU territory. The representative is responsible for registering both the manufacturer and its medical device(s) in the Eudamed and, like the manufacturer, may be held liable for defective devices under certain circumstances.
- ✓ The ‘person responsible for regulatory compliance’ is a new role introduced by the MDR to ensure that the manufacturers comply with its requirements and obligations. It plays an internal control function and is obliged to report any non-compliance.

Article 2(1) of the MDR defines a medical device as any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more predetermined medical purposes, and “*which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means*”.

The key aspect of this definition is the **intended medical purpose** as determined by the manufacturer. A product qualifies as a medical device if it is specifically designed to contribute to medical and healthcare activities such as diagnosis, therapy, or disease monitoring, regardless of whether it achieves its intended effect through physical, chemical, or technological means.

Software can be defined as Software as a Medical Device (SaMD)¹²², both when it is included in another medical device-hardware (embedded) or when it operates independently (stand-alone), as long as it supports healthcare professionals in diagnosis and therapy, such as in the case of software used for the interpretation, analysis and processing of medical data or producing diagnostic information¹²³.

Manufacturers must comply with a comprehensive set of requirements under the MDR before their medical devices can be placed on the market. Central to these obligations is the **establishment and maintenance of robust risk and quality management systems, which are critical for ensuring the safety and performance of the devices throughout their entire lifecycle**.

The MDR also mandates that manufacturers conduct thorough **clinical evaluations to substantiate the safety and effectiveness of their devices**, demonstrating compliance with the applicable essential requirements. This process involves systematic collection, appraisal, and analysis of clinical data, which must be carefully documented and included in the technical documentation for the device. This technical documentation serves as evidence that the product meets all regulatory standards and is fit for its intended medical purpose.

Before placing devices on the market, manufacturers are required to undergo a **conformity assessment procedure, to demonstrate that all safety and performance requirements laid down in the MDR have been appropriately fulfilled. This process, depending on the risk classification of the device, may involve independent evaluation by competent notified bodies**. In this respect, medical devices that comply with the relevant harmonized standards, or the relevant parts of those standards, shall be presumed to be in conformity with the requirements of the MDR. Once the procedure is successfully concluded, the manufacturer is entitled to obtain a **CE marking** for the medical devices, as a proof of certification.

¹²² While SaMD is defined in the International Medical Device Regulators Forum (IMDRF) guidelines (<https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf>) and the term is broadly used within the EU, the correct terminology in an EU context is Medical Device Software (MDSW), as defined by MDR rule 11.

¹²³ Ludvigsen, K., Nagaraja, S., & Daly, A. (2022). When is software a medical device? understanding and determining the “intention” and requirements for software as a medical device in European Union law. *European Journal of Risk Regulation*, 13(1), 78-93. <https://doi.org/10.1017/err.2021.45>.

Notified bodies are independent organizations designated by competent national authorities to conduct conformity assessments, including evaluating technical documentation, reviewing clinical evidence, auditing quality management systems and, where required, performing on-site inspections. Their responsibility extends to ensuring that medical devices consistently comply with the regulation throughout their lifecycle, which involves ongoing surveillance of manufacturers and their post-market performance. Notified Bodies must meet the stringent criteria set out in Annex VII of the MDR, which require proven competence, impartiality and independence. They must have sufficient technical expertise, resources, and procedures in place to assess devices accurately and effectively, and they must remain free from conflicts of interest to safeguard the objectivity of their evaluations. Designation is monitored by the national authority and the EU Commission, ensuring that Notified Bodies maintain high standards of performance and integrity in their operations.

Once the devices are on the market, manufacturers retain ongoing responsibility for their performance and safety. This includes the **implementation of post-market surveillance systems to monitor the device's real-world use and promptly address any issues that may arise**. These obligations reflect the MDR's objective to maintain high standards of patient safety and product quality throughout the product's entire lifecycle¹²⁴.

Performance evaluation is governed by Annexes I and IX of the MDR which, respectively, set out the requirements to demonstrate that the devices are designed and manufactured to ensure a high level of safety and performance in each phase of their operation, and the procedures for conformity assessment, which may include examination of the technical documentation and verification by a notified body.

According to Art. 51 and Annex VIII of the MDR, devices shall be divided into classes I, IIa, IIb and III, on account of their intended purpose and inherent risks. The classification determines the level of control required and the stringency of the conformity assessment. In more detail:

- **Class I** (low risk): devices in this class present the **lowest risk** to patients and users (*e.g.*, non-invasive devices such as bandages, wheelchairs, and simple surgical instruments). They are non-invasive or minimally invasive, intended for temporary or transient use, and generally of low complexity. Self-declared conformity by the manufacturer without the need for a Notified Body (except for Class I devices with a measuring function, sterile components, or reusable surgical instruments).
- **Class IIa** (medium-low risk): devices falling in this class pose a **moderate level of risk** and are typically used for short-term applications or for managing relatively low-risk physiological conditions (*e.g.*, hearing aids, dental fillings, or diagnostic devices like urine test strips). They may be invasive but generally for short-term use and are employed to support or influence physiological functions without critical intervention.
- **Class IIb** (medium-high risk): this class involves a higher level of risk due to their greater interaction with the human body and so invasiveness, or critical function (*e.g.*, infusion pumps, ventilators, or implants that are not life-sustaining but have significant physiological interaction).
- **Class III** (high risk): in this case, the devices are the highest risk category, involving complex and critical technologies, often used in life-sustaining or high-risk medical applications (*e.g.*, pacemakers, heart valves, and implantable devices with medicinal components). They are invasive, often implanted, and intended for long-term use, as well as critical in sustaining life, preventing severe health deterioration, or directly interacting with vital organs.

Specifically for software, **Rule 11 of Annex VIII** states that software intended for diagnostic or therapeutic purposes is generally included in Class IIa, but may be reclassified to Class IIb or III if its

¹²⁴ Moti, D., Kirani, A., Navigating the Regulatory Landscape of Software as a Medical Device (SaMD) Compliance Challenges, Best Practices, and Future Trends, International Journal of Engineering, Management and Humanities (IJEMH) Volume 5, Issue 2, Mar.-Apr., 2024 pp: 300-309. <https://www.ijert.org/papers/IJERT22A6973.pdf>.

use may have significant health consequences¹²⁵. For example, software designed to monitor vital physiological parameters in real time may be included in Class IIb because of its potential ability to prevent immediate risks to the patient. Therefore, **most of SaMDs – including DTs – are likely to fall into Class IIa or IIb**, thus requiring evaluation by notified bodies. This aims to strengthen regulatory oversight for high-risk software, increasing the safety of devices placed on the market¹²⁶.

Art. 61 of the MDR imposes rigorous clinical evaluation standards to confirm adherence to general safety and performance requirements, covering aspects such as reliability, safety, and an acceptable benefit-risk profile relative to the state-of-the-art (Annex I, MDR). Notably, these evaluations must include robust clinical evidence collection as post-market release.

In this latter respect, manufacturers must implement a post-market surveillance system for the continuous collection of data, analysis and corrective actions. These data must feed into risk-benefit assessments, clinical evaluations and safety updates throughout the entire lifecycle of medical devices (Article 83(3) MDR). The surveillance plan, required as part of the technical documentation, outlines procedures for data collection, complaint investigation, incident management and post-market clinical follow-up (Annex III, MDR). The post-market clinical follow-up process must ensure that new side effects, risks and off-label uses are identified and managed over time and that the benefit/risk ratio remains favourable (Annex XVI, MDR).

Regularly updated post-market reports, which include a summary of findings and corrective actions, are mandatory and depend on the class of the devices. Manufacturers are required to draw up ‘Periodic Safety Update Reports’ for class IIa, IIb and III devices, to be regularly updated and made available to notified bodies and competent authorities. Article 87 requires the timely reporting of any serious incidents that may involve the devices, in order to ensure a rapid response and the prevention of future risks.

In addition, the MDR introduces a **Unique Device Identification (UDI) system**, namely a framework aimed to enhance the traceability, safety and transparency of medical devices within the European Union and support the effectiveness of post-market surveillance. It requires that every medical device placed on the EU market be assigned a unique identifier to ensure consistent identification throughout the device’s lifecycle. The system includes two different components:

- **UDI-DI (Device Identifier)**: a static code specific to a manufacturer and a device model;
- **UDI-PI (Production Identifier)**: a dynamic code identifying production-related information, such as lot number, serial number, and expiration or manufacturing date.

Crucially, the MDR (Art. 103) established a **Medical Device Coordination Group (MDCG)** to ensure uniform application of the regulation across the EU. Comprised of representatives from the Member States, the MDCG is chaired by the European Commission, which also provides administrative and logistical support. Its overarching goal is to promote harmonization and cooperation between national competent authorities and other stakeholders in the regulation of medical devices.

The MDCG plays a pivotal role in overseeing the implementation of the MDR by providing guidance, advice, and interpretations on regulatory requirements. It facilitates consistency in areas such as conformity assessments, designation and monitoring of notified bodies, and market surveillance activities. The group ensures that regulatory decisions are aligned across Member States, fostering a unified approach to safety and performance standards for medical devices in the EU market. Additionally, it develops and disseminates non-binding guidance documents and recommendations to clarify aspects of the MDR for manufacturers, notified bodies, and regulators. These documents are

¹²⁵ Rule 11 of Annex VIII of the MDR establishes that: “Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause: - death or an irreversible deterioration of a person’s state of health, in which case it is in class III; or - a serious deterioration of a person’s state of health or a surgical intervention, in which case it is classified as class IIb. Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb. All other software is classified as class I”.

¹²⁶ Lal A, Dang J, Nabzdyk C, Gajic O, Herasevich V. Regulatory oversight and ethical concerns surrounding software as medical device (SaMD) and digital twin technology in healthcare. *Ann Transl Med.* 2022 Sep;10(18):950. <http://doi.org/10.21037/atm-22-4203>.

instrumental in addressing complex issues such as the classification of devices, clinical evaluation requirements, and post-market surveillance obligations. Through its coordination efforts, the MDCG enhances transparency and predictability in the regulatory environment, supporting the safe and effective use of medical devices across the EU. In its “Guidance on cybersecurity for medical devices”, the group emphasizes the importance of incorporating strong security standards and measures at every stage of the device lifecycle, including a risk assessment that extends to post-market surveillance, and promotes a collaborative approach between manufacturers, suppliers, healthcare professionals and patients to reduce the risks associated with the use of medical devices, including SaMDs.

In vitro diagnostic devices (**IVD**) are defined by the IVDR as “*any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, piece of equipment, software or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information on one or more of the following: (a) concerning a physiological or pathological process or state; (b) concerning congenital physical or mental impairments; (c) concerning the predisposition to a medical condition or a disease; (d) to determine the safety and compatibility with potential recipients; (e) to predict treatment response or reactions; (f) to define or monitoring therapeutic measures*”. Examples of IVMD include pregnancy tests or SARS-CoV-2 tests, HIV tests, blood glucose meters and blood typing reagents.

IVDs are classified into four risk-based categories, from class A (low risk) to class D (high risk). This classification framework is designed to align the regulatory requirements with the potential risk posed by the device to patients and public health. The risk assessment considers factors such as the intended use of the device, the consequences of incorrect results, and the device’s impact on critical medical decisions.

Analogous to the MDR, the IVDR establishes comprehensive rules governing the placement of IVDs on the market, their availability, and their use for human health purposes, in order to ensure a high level of safety, performance, and reliability for these devices, thereby fostering confidence among healthcare providers and patients. **Conformity assessment procedures under the IVDR are tiered based on the classification of the relevant device:** while class A devices can generally be self-certified by manufacturers without the involvement of a notified body, higher-risk classes – Class B, Class C, and Class D – require a more stringent evaluation process, with the manufacturers having to seek certification from competent notified bodies. This assessment includes a review of the IVD’s technical documentation, quality management system, and compliance with performance and safety requirements. Class D, representing the highest-risk category, often entails additional scrutiny, including batch testing or validation by an external reference laboratory, to ensure reliability and accuracy for critical applications.

Manufacturers are also required to carry out **performance evaluations** and, where necessary, **performance studies** to provide robust evidence of the device’s analytical and clinical validity. These evaluations must demonstrate that the IVD meets the performance and safety criteria specified in the Regulation. The evidence from these studies is incorporated into the device’s technical documentation and is subject to review during the conformity assessment process. In addition, in vitro devices must comply with the ‘General safety and performance requirements’ outlined in Annex I of the IVDR. These include detailed stipulations for device labelling, technical documentation and the establishment of a quality management system that governs all aspects of the devices’ lifecycle. Labels must include precise information to ensure the safe and effective use of the device, while technical documentation must provide comprehensive evidence supporting its compliance with the IVDR.

Upon positive completion of the conformity assessment procedure, the manufacturer can obtain the CE marking, which signifies that the device complies with the IVDR’s safety, performance, and quality requirements. Once placed on the market, IVDs are subject to rigorous **post-market surveillance and vigilance requirements**. Manufacturers must implement systems to monitor the performance of devices in real-world use, identify any safety concerns, and report adverse events to competent authorities. These obligations ensure that any potential risks are promptly addressed, safeguarding patient safety and maintaining public trust in the regulatory framework.

In light of the above, **in all cases when the SaMD is AI-driven, or qualifies itself as an AI system, falling in the scope of the AI Act, the following should be considered about the associated level of risk:**

- **if the device, in the context of the process of acquiring CE marking as SaMD, is subject to the conformity assessment of a notify body (falling in class IIa, IIb, III with regard to MDR and class B,C,D with regard to IVDR), it must qualify as ‘High-risk’ under the AI Act;**
- **if the device, during process of acquiring CE marking as SaMD, is not subject to the conformity assessment by a notify body (class I-devices in the MDR and class A-devices in the IVDR), it must not be considered ‘High-risk’ according to the AI Act.**

Beyond the intricate implications of the AI Act, the MDR and IVDR present substantial obstacles to the adoption of digital twins as medical devices in the medical and healthcare sector. These challenges arise largely from the difficulty of aligning cutting-edge technologies with rigorous regulatory frameworks originally designed for conventional equipment and tools.

One major barrier lies in demonstrating compliance with the ‘General Safety and Performance Requirements’ under both regulations. Indeed, DTs often involve machine learning algorithms and adaptive systems, which require continuous updates and recalibration. However, **the regulatory emphasis on pre-market evaluations, such as clinical investigations and performance studies which are static in nature, does not easily accommodate the iterative development and (real-time) adaptability characteristic of Digital Twins.** This creates uncertainty about how manufacturers can provide sufficient evidence of safety and performance for such evolving technologies.

Similarly, **post-market surveillance and vigilance requirements** are designed for medical devices with stable configurations. As some DTs evolve through ongoing data collection and algorithmic adaptation, the current surveillance framework does not provide adequate mechanisms to account for their dynamic nature. This creates a regulatory mismatch that may hinder the ability to track, assess, and mitigate risks effectively throughout Digital Twins’ lifecycle under the MDR and the IVDR.

Another legal hindrance relates to the definition of medical device and intended purpose. Both the MDR and the IVDR require that a device must have a specific medical purpose as defined by its manufacturer, such as diagnosis, treatment, or monitoring. **Digital Twins, however, often serve broader and evolving purposes, including predictive modelling, simulations, and decision support, which may not neatly fit the regulatory definitions.** This lack of clarity can create uncertainty about whether a DT qualifies as a medical device (precisely a SaMD), complicating its pathway to market.

As specified above, medical devices are classified on a risk-basis, with higher-risk ones requiring stricter assessments. Digital Twins, particularly those predicting critical health outcomes, could be classified in higher risk categories (Class IIb or III under the MDR or Class C or D under the IVDR). However, **the criteria for applying this classification are not easily adaptable to Digital Twins, which in most cases function as software with indirect (rather than direct) interaction with patients. This ambiguity makes it difficult to determine the appropriate regulatory pathway.**

Moreover – and crucially – **regulatory requirements for clinical validation and performance studies are also difficult to fulfil for Digital Twins, as they often involve predictive simulations rather than direct interventions, challenging traditional metrics of clinical efficacy and safety grounded in real-world patient outcomes.** DTs operate within a different paradigm, leveraging advanced algorithms, virtual modelling, and large datasets to predict medical outcomes or optimize treatment strategies. This divergence makes it difficult to apply established validation methods, which often require evidence generated through randomized controlled trials or similar studies that focus on measurable effects observed in patients. In contrast, the outputs of DTs are often probabilistic or theoretical, making it challenging to fit these results into conventional safety and efficacy benchmarks.

Additionally, ensuring the reproducibility and reliability of predictive models, particularly those that evolve through machine learning, complicates the demonstration of compliance with regulatory expectations. **The dynamic and adaptive nature of some DTs further adds to this complexity, as**

the technology may change over time, requiring ongoing validation that traditional frameworks are not well-equipped to address. These factors collectively highlight the need for updated or tailored regulatory approaches, in the EU medical device single market, to assess the unique attributes of VHT effectively.

* * * * *

The MDR/IVDR and the Health Technology Assessment regime share an interconnected focus on ensuring the safety, efficacy and value of medical devices in healthcare. These frameworks, while distinct in purpose, overlap significantly in their objectives and processes, providing opportunities for a cohesive approach to evaluating and introducing new medical technologies.

As detailed above, the medical device legislation lays down the requirements for medical devices to enter the European market, emphasizing compliance with General Safety and Performance Requirements, as well as clinical evaluation and post-market surveillance obligations. By requiring rigorous clinical evidence of safety and performance, the MDR ensures that devices meet a baseline standard for patient care and functionality. This evidence serves – *inter alia* – as a foundation for HTA, which evaluates medical technologies’ broader impact, including their clinical and cost-effectiveness, and societal benefits¹²⁷. HTA processes complement the MDR by considering not only the safety and performance of devices but also their value within specific healthcare systems. For instance, data generated through MDR-required clinical evaluations and performance studies can feed into HTA reports, providing evidence of a device’s effectiveness compared to existing technologies. Similarly, post-market surveillance offers real-world data that HTA bodies can analyse to assess long-term outcomes, cost-effectiveness and adoption challenges.

The integration of MDR processes with HTA ensures a comprehensive lifecycle approach to medical device evaluation. This connection underscores the importance of aligning regulatory frameworks with healthcare system needs, facilitating informed decision-making about resource allocation, reimbursement and patient care improvements: by bridging the regulatory and health policy dimensions, the MDR and HTA together enhance the quality, accessibility, and sustainability of healthcare technologies across the EU.

HTAs compile information from various fields, while also considering efficacy, safety, economic and organizational aspects of methods, solutions, tools and tech advancements in the health sector. This process is conducted in a systematic, transparent, unbiased and rigorous manner. The purpose is to inform decision-making by using explicit methods to promote an equitable, efficient, high-quality health system. This kind of assessment can be applied at different points in the lifecycle of health technology, *i.e.*, pre-market, during market approval, post-market, and through the disinvestment of health technology. The approach and methods used at each phase will differ and depend on the available evidence (whether primary or secondary data) and the decision about the technology.

While licensing approval is mainly focused on the technical and safety profile of the medical device, HTA bodies have different interests and, therefore, varying evidence requirements which, normally, aim to inform policymakers (and decision-makers in general) of the rationale allocation of resources within finite budgets to the funding (or using) of healthcare interventions.

As seen above, the European regulatory landscape for medical devices is highly fragmented due to the decentralized nature of the CE marking system. Unlike the pharmaceutical sector, which has a central authority in the European Medicines Agency, conformity assessment and approvals under the MDR are managed at Member States national level by independent certification organizations known as notified bodies. These entities often lack the capacity or resources to keep up with all relevant technological advancements and innovative approaches, such as Digital Twins and in-silico methodologies.

¹²⁷ A novel Regulation (EU) 2021/2282 on health technology assessment entered into force on 11 January 2022 and will apply from 12 January 2025. ‘Health technology assessment’ is defined, in Art. 2(5), as “a multidisciplinary process that summarises information about the medical, patient and social aspects and the economic and ethical issues related to the use of a health technology in a systematic, transparent, unbiased and robust manner”.

Despite these challenges, EU regulations do not explicitly prevent the use of DTs and in-silico evidence for regulatory submissions. Some companies have successfully used finite element simulation model results to support CE marking applications, particularly in orthopaedics. In the absence of a harmonized standard, these companies often relied on the ASTM F2996-20 standard to support the credibility of their in-silico evidence¹²⁸.

The already mentioned ASME VV-40 standard is undoubtedly the most established document in this field, creating a consistent framework for assessing the credibility of computational models in healthcare. The inclusion of references to VV-40 for assessing the credibility of predictive models sets a crucial precedent. This recognition implies that any medical device company could theoretically produce in-silico evidence supported by a VV-40 credibility assessment and refer to this ISO 21535:2023 as a precedent, potentially broadening the acceptance of Digital Twin-based methodologies across various medical device categories.

9.2 BARRIERS, DEPENDENCIES AND PROPOSED ACTIONS

Although highly promising also in the medical device field, the VHT still face significant hurdles under the applicable MDR and IVDR, which are conceived to embrace and encompass traditionally designed and operating medical devices. Therefore, the requirements for clinical validation, performance studies, and conformity assessments under these regulations are difficult to fulfil, as they rely on established conventional metrics of safety and efficacy that do not easily apply to virtual simulations. Additionally, the evolving nature of some Digital Twins raises questions about post-market surveillance, continuous compliance, and the management of cybersecurity risks, creating further regulatory uncertainty.

Clearly, many obstacles are closely related to, overlap with, or are almost identical to those outlined above concerning the clinical trial sector.

A) Divergence in risk-based classification and requirements duplication¹²⁹

Barrier

As already extensively explained, the AI Act introduced a risk-based classification for AI systems and models, dividing them into four levels of (i) unacceptable risk (AI systems that pose a clear threat to safety or fundamental rights are prohibited), (ii) high risk (AI systems that significantly affect health, safety, or fundamental rights are subject to stringent requirements), (iii) limited risk (AI systems with specific enhanced transparency obligations), (iv) minimal risk: AI systems with minimal or no risk, subject to basic regulatory intervention.

By contrast, the MDR and IVDR classify medical devices based on their potential risk to patients and users, using a different framework: (a) MDR distinguishes the devices into four classes: I (low risk), IIa, IIb, and III (high risk), based on factors like duration of use, invasiveness, and level of interaction with the human body; (b) IVDR, in turn, identifies four classes: A (low risk), B, C, and D (high risk), considering aspects such as the intended purpose and potential impact on public health.

Many DTs qualifying as Software as medical devices may classify as both high-risk AI systems under the AI Act, and high-risk devices according to the MDR or the IVDR. For this reason, manufacturers will be required to comply with all respective requirements, undergoing different conformity assessments, risk evaluation and post-market surveillance processes, and implementing all required security, data governance and mitigation measures. Because of the overlap between these

¹²⁸ A significant milestone in promoting these methodologies within the regulatory community is the revision of the ISO 21535:2023 standard for hip-joint replacement implants. This standard explicitly recognizes the role of theoretical analysis and modelling, including finite element analysis, in selecting appropriate component sizes for testing worst-case scenarios.

¹²⁹ Very helpful indications emerge, in this specific regard, from the questionnaire on “Artificial Intelligence in medical devices” which has just been published (25 November 2024), by the German Notified Bodies Alliance for Medical Devices (IG-NB, *Interessengemeinschaft der Benannten Stellen für Medizinprodukte in Deutschland*) and the European Association of Medical Devices Notified Bodies. <https://www.team-nb.org/medical-devices-ai-questionnaire-jointly-published-by-ig-nb-team-nb/>.

legislations, in absence of specific guidance by competent Authorities, there is the tangible and costly risk of duplication of some requirements, creating unnecessary burdens.

Proposed action

- By the Medical Device Coordination Group / European Artificial Intelligence Board:** adopt guidance aimed to (i) guidelines designed to standardize the application of the risk classes established by all the regulations involved and, *a fortiori*, the obligations that are laid down by each of them; and so (ii) address the risk of double-requirements arising from the overlap between the MDR/IVDR and the AI Act, ensuring clarity, reducing uncertainty, and streamlining compliance pathways for manufacturers of AI-enabled medical devices, such as some Digital Twins, ultimately fostering a more efficient regulatory environment. For instance, the MDCG could specify how conformity assessments or clinical evaluations performed under the MDR/IVDR could satisfy related obligations under the AI Act, particularly regarding the safety, performance, and transparency of AI systems. Similarly, the AI Office could outline how existing MDR/IVDR documentation and evaluations, such as post-market surveillance or risk management plans, align with the AI Act's requirements for monitoring AI systems' robustness and accountability. Moreover, **mechanisms for reciprocal assessments, acknowledgement and endorsement can be put forward, thereby avoiding redundant evaluations. Coordinated efforts may also promote a more unified interpretation of overlapping requirements, such as those relating to data governance, risk classification, and algorithm validation.** In essence, proactive and coordinated guidelines from competent authorities is essential to harmonize the interplay between the MDR/IVDR and the AI Act, fostering a regulatory environment that supports innovation while safeguarding safety and efficacy of AI-driven SaMD.

B) Unclear regulatory pathways for CE marking and post-market evaluations

Barrier

The regulatory frameworks under the MDR and IVDR place significant emphasis on pre- and post-market evaluations, such as clinical investigations and performance studies, which are designed to assess medical devices in a static, fixed state. These evaluations typically involve one-time or periodic assessments that measure the device's safety and efficacy based on its intended use and defined technical specifications at a particular point in time. Some Digital Twins operate fundamentally differently, as they are dynamic, evolving systems that adapt in real time to new data inputs and environmental changes, using machine learning algorithms and predictive modelling. **This iterative development and real-time adaptability present a challenge for traditional regulatory processes, which are not designed to accommodate technologies that evolve continuously after deployment.** The static nature of conventional evaluations may fail to capture the ongoing risks, performance changes, or evolving outputs of some Digital Twins technologies, making it difficult to ensure their compliance over time.

Furthermore, the regulatory frameworks set out under the MDR and IVDR are built around assessing medical devices through real-world clinical outcomes that are tangible, measurable and directly observable in patients. These metrics are well-suited for static devices or technologies with a clearly defined and fixed functionality. In contrast, as said, Digital Twins operate on the principle of simulating physiological processes, disease progression, or treatment responses using advanced algorithms and virtual modelling. **These predictive capabilities provide valuable insights into patient health but do so by generating probabilistic data rather than empirical results obtained through direct interaction with the patients. This divergence challenges regulators to assess the reliability, validity, and safety of these technologies using traditional methodologies that may not fully capture the complexity and adaptability of DTs.**

Proposed action

- **By the Medical Device Coordination Group (involving the European Artificial Intelligence Board):** adopt new methodologies, guidance and qualification procedures, also leveraging and adapting the framework already provided by the ASME VV-40 (setting out guidelines for assessing the credibility of computational models used in the design, evaluation, and regulatory approval of medical devices), to define a standardized European pathway that better accommodates the iterative and adaptive nature of some Digital Twin models, while ensuring safety, performance, and compliance over their lifecycle. The MDCG can, when relevant, outline processes for real-time performance monitoring and periodic re-validation of DTs that reflect their evolving algorithms and data-driven outputs. Additionally, the Group could propose a framework for adaptive conformity assessments, allowing for incremental updates to a Digital Twin’s functionalities and/or predictive capabilities, without requiring complete re-certification. This approach would mirror the iterative design and deployment cycles of these technologies, enabling manufacturers to refine their models while maintaining regulatory oversight. Additionally, the MDCG could foster the EU VHT ecosystem, encouraging collaboration between manufacturers, regulators and clinicians to build and nurture trust in the field. This might involve piloting new validation methods, creating repositories for shared best practices, or establishing advisory panels with expertise in machine learning and predictive modelling.

10 Intellectual Property

10.1 THE FRAMEWORK FOR VHT

The management of Intellectual Property Rights (IPR) in the context of a public VHT infrastructure is primarily concerned with the identification of the following three key elements: (i) the different stakeholders and their respective interests, (ii) the (broadly conceived) knowledge and information (including datasets, models, methods, tools and other forms) that each party may contribute to the collaborative ecosystem, and (iii) the assets that will emerge from the further elaboration of such knowledge.

In order to ascertain which information can be protected by exclusive IP rights and which IP rights would be applicable, it is essential to consider all of the aforementioned elements collectively. Furthermore, it is crucial to determine how these can be managed by the consortium partners in a manner that not only benefits the wider community, but also ensures the continued viability and financial sustainability of the ecosystem.

Identifying those with an interest in the proper management of IPR within the context of the VHT infrastructure may appear to be a relatively straightforward undertaking. However, a more forward-thinking approach would also entail considering the potential for new subjects to emerge and become involved in the infrastructure in the future. At the moment, and for the purposes of this document, the main categories whose interests shall be taken into account are:

- a. **Researchers:** two core, contrasting interests are at play. On the one hand, there is a desire to ensure a high degree of openness, as this facilitates further research. Conversely, however, facilitating the acquisition of IPR on the products of their research can serve as an incentive, particularly if they are able to derive financial benefit from said IPR. This second element would become increasingly relevant when considering the long-term, whole-system perspective of retaining talent in the research field.
- b. **Contract Research Organizations (CROs):** similarly, CROs' interests centre on securing IP ownership, protecting proprietary algorithms, ensuring robust data privacy, and defining clear licensing rights. Effective IPR management enables CROs to safeguard competitive advantage, ethically handle patient likeness, manage joint development rights, and avoid infringement risks, fostering innovation in clinical research.
- c. **OEMs (Original Equipment Manufacturers):** it would be prudent to ensure that the relevant IPR are maintained for the products provided for implementation in digital twins.
- d. **Business angels and investors / Digital twin vendors / Medical Product Developers:** it can be reasonably assumed that effective IPR management is a key factor in safeguarding the financial investment of investors. In this sense, investors would expect the companies they finance to be able to protect their products and utilize them in a productive manner.
- e. **Buyers and payers (both private and public) of digital twins / Hospitals:** this category is sufficiently expansive to encompass healthcare providers and insurance companies, among other entities. It is reasonable to posit that such entities would wish to ensure that the products they procure do not infringe upon the IPR of others. Furthermore, it is probable that they would advocate for transparency in order to facilitate research and development.
- f. **Healthcare professionals:** it seems probable that they would advocate for the development of openly accessible products, as this would facilitate the deployment of novel digital twin-based technologies.
- g. **Healthcare policy makers:** it is important to strive for a balance between openness and innovation, with the ultimate goal of achieving a net societal benefit.
- h. **Patients (both as individuals and organizations/advocates):** it is possible that they may not be particularly interested in IP protection for their data, as the majority of their concerns would

instead relate to matters of privacy. In a more general sense, it is probable that they would favor an open approach, as this would facilitate further research and, subsequently, more favorable outcomes for all patients.

Just as the subjects falling into the aforementioned categories are stakeholders of different interests, so in different ways they can contribute, first of all, to the construction of the database (the Repository) and the accumulation of knowledge, which collectively form the basis of the ecosystem. In this regard, it is imperative to direct attention not so much towards the individual pieces of health data (understood here as defined in Art. 2.2 EHDS: “*data concerning health and genetic data as defined in Regulation (EU) 2016/679, as well as data referring to determinants of health, or data processed in relation to the provision of healthcare services, processed in an electronic form*”), but rather towards the manner in which these data are structured and the technological instruments are employed for their processing.

Indeed, there is considerable debate surrounding the potential for identifying IPR on **individual health data** which, as previously discussed, are rather a significant area of concern in relation to data protection legislation.

It is challenging to ascertain which data would be protected under the aegis of **copyright** law, particularly envisaging how the items that constitute health data could ever fulfil the prerequisite of creativity. This type of data could only be considered to constitute “real-world facts”¹³⁰, which are clearly devoid of any creative input and thus not subject to copyright. It can be argued that only data of a similar nature to medical reports, namely those that constitute the creative output of the healthcare professional who wrote them, could be subject to copyright. In the case of medical images (such as x-rays and ultrasounds), a common argument is that they are photographs and therefore potentially subject to copyright; still, medical images are obtained through technical means, which can appear as a circumstance further supporting the argument that they lack creativity.

It is possible to propose different reasonings with regard to a specific category of data, namely that of synthetic data. Synthetic data may be defined “*as artificially created data such as fabricated numerical values, text, images, and videos*”¹³¹. The question of authorship in relation to AI-generated content is currently a topic of ongoing debate. It is argued that an AI application cannot hold IPRs in itself. Conversely, it could be argued that in instances where a human has provided creative input to an AI machine that is sufficiently understandable and predictable, the human could be considered the author (and the AI could be regarded as a mere tool). It is evident that the concept of authorship is confronted with a novel challenge in the context of synthetic data. In the event that authorship is established, it may be possible to qualify synthetic data as “fictional facts” (rather than “real-world facts”), which could satisfy the requisite level of creativity for copyright protection¹³². For these reasons, synthetic health data may be eligible for copyright protection, although the current uncertainty regarding authorship raises significant doubts about the reliability of this form of protection. Still, while there may be potential for future developments in the field of synthetic data, it seems unlikely that copyright protection for individual items of health data will be granted in the near future.

It is more probable that copyright will assume a significant role in the protection of health data when considered as a **database**. Indeed, as the Database Directive (Directive 96/9/EC) establishes in Article 3, databases may be subject to copyright if the author’s intellectual creation is expressed in the selection and arrangement of data. It is crucial to highlight that it is precisely this selection and arrangement that constitute the author’s creation, rather than the data itself. This implies that the copyright pertaining to the database does not extend to the underlying health information.

Apart from what already mentioned above, it is essential to highlight that databases also receive a *sui generis* protection, as provided by the Database Directive. This *sui generis* protection is consequential to an “*investment of considerable human, technical and financial resources*” (Recital 7, Database Directive). The owner of this right to protection would be the person who “*takes the initiative and the*

¹³⁰ Lee, Peter, Synthetic Data and the Future of AI (February 10, 2024). 110 Cornell Law Review (Forthcoming). <https://ssrn.com/abstract=4722162>.

¹³¹ *Ibid.*

¹³² *Ibid.*

risk of investing” (Recital 41, Database Directive). Since the aforementioned investments would certainly be provided in the context of health data databases, the provisions of the Directive would find application. This *sui generis* protection may apply separately and irrespective of the protectability of both the data items and the database itself by other rights (e.g., copyright) as provided by Art. 7 of the Directive; therefore, database protection does not imply a choice that limits other IPR management strategies.

It can therefore be concluded that the interest in health data in the IP field is predominantly focused on the management of the databases that collect and organize them.

Highly related to this matter are the recently introduced rules on **data mining** activities of copyrighted contents, which could be of significant relevance for the purposes of building a VHT infrastructure. According to the Directive (EU) 2019/790 “*on copyright and related rights in the Digital Single Market*” (DSM), text and data mining (TDM) means “*any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations*” (Art. 2, point 2). TDM is fundamental to the development of AI technologies. These techniques may have a twofold importance: if it is true that the parties involved in the development of the VHT infrastructure would likely perform such activities on the items provided in the Repository, it should not be underestimated the potential advantage of having the ability to extract and use materials available outside the same infrastructure. Moreover, such rules are relevant to assess whether other parties that are external to the EDITH CSA may lawfully adopt said techniques on the resources that are provided in the Repository or that are produced thanks to it.

Art. 3 of the DSM Directive prescribes an exception to certain copyrights in case of “*reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access*”. This exception would thus not be applicable to external parties who have not acquired “lawful access”, which shall be understood as access based on an open access policy, on contractual arrangements, on other lawful means, or to content that is publicly available online (Recital 14, DSM Directive). On the other hand, where lawful access does occur, the exception would certainly be applicable to a variety of actors that fit into the definition of research organization outlined in Art. 2, point 1. This includes, in particular, those who conduct scientific research: (a) on a non-profit base or by reinvesting all the profits in their scientific research; or (b) pursuant to a public interest mission recognized by a Member State. It is clear that numerous research organizations that are part of the VHT ecosystem could rely on these provisions, although it is not to be taken for granted that further results of such elaboration could be used for commercial purposes. It is important to note, as well, that those that act under this rule are required to abide by appropriate levels of security in storing copies of the materials (Art. 3, co. 2).

Art. 4 of the DSM Directive also provides a significant exception for reproductions and extractions of lawfully accessible works, when done by any entity (not necessarily a research organization), if the text which was “mined” was accessed lawfully and the copyright owner has not expressly prohibited the use of the text/work for the purpose of data mining. This exception might find an even wider application than the previous one (since it covers commercially motivated TDM), though it is dependent on the IPR holder’s choice as to prohibiting TDM or not. While there is clearly an opt-out option for Art. 4, it is not possible to exclude the application of Art. 3 through contractual arrangements (Art. 7, DSM Directive).

Finally, the regulatory framework governing **trade secrets** shall be taken into consideration, given that this legal notion is expressly referred to in the proposed EHDS Regulation. Also to this end, it would be useful to draw a distinction between individual items of health data and health data databases. According to the Directive on the Protection of Trade Secrets (Directive 2016/943 – TSD), in order to qualify as such, there are three main requirements: i) secrecy, ii) existence of commercial value because of secrecy, iii) existence of reasonable steps to keep the secrecy. It seems highly unlikely that individual

data could fulfil these requirements, especially for what concerns commercial value¹³³. When considering medical images specifically, it could also be noted that the data elements extractable from said images are “*readily observable to anyone looking at the image*”¹³⁴, so the fundamental requirement of secrecy would be lacking.

The database itself, instead, could likely fall under the scope of trade secrets, as it could hold significant commercial value, as long as it is of course kept secret¹³⁵.

Upon conclusion of this regulatory overview, the proposal for the EHDS Regulation merits particular attention. In fact, it provides, in Art. 33a, that “*electronic health data entailing protected intellectual property [and] trade secrets [...] shall be made available for secondary use*”. To this end, it has been argued that, following Article 5(d) TSD, “*public health can serve as a legitimate interest for the purpose of which an application for the measures, procedures and remedies provided for in the TSD is dismissed. The EHDS as such can be regarded as a clarification of this exception, or, in other words, as a specific outcome of this balancing exercise*”¹³⁶.

For the time being, arguably even more pertinent, although no less intricate to oversee, is the function of IPR in safeguarding the instruments employed to process shared data and the outcomes derived from such processing.

In this regard, any predictive computer models, as well as the infrastructure constituted by the “trinity” of Catalogue-Repository-Platform, must first be considered **software** subject to the application of Directive 2009/24/EC “on the legal protection of computer programs”. The objective of this piece of legislation is to eliminate discrepancies in the protection of computer programs across Member States. To this end, it affords them the same copyright protection as that extended to literary works. The form of protection in question encompasses the expression of a computer program in any form, rather than the underlying ideas and principles or any constituent elements.

In order to achieve the greatest possible degree of openness, it may be advantageous to consider the option of making one’s software “open-source”. Open-source software (OSS) is defined as software that is distributed in a manner that allows the source code to be accessed freely, modified, run, adapted, compiled and distributed. The licensee is granted rights that extend beyond mere use, although they may also be subject to specific obligations (*e.g.*, in regard to redistribution). A variety of licenses align with the tenets of the open-source movement, necessitating a tailored selection based on the specific interests at hand. One notable distinction can be made between “copyleft” and permissive licenses. A permissive free software license imposes minimal requirements regarding the redistribution of the software in question. One may cite the MIT license as a typical example of a permissive license. Furthermore, “copyleft” requires that all modified and extended versions of the program be free as well. The most prominent example of “copyleft” licensing is the Global Public License (GPL).

Predictive computer models, or certain functions therein, as well as (more broadly) computer simulations, such as the VHT environment and the various digital human twins involved, may also be eligible for **patent** protection.

At the European level patents may be granted through (i) the national route, which involves filing separate applications with each national patent office, resulting in individual national patents, (ii) the European Patent Convention (EPC) route, which allows filing a single application with the European Patent Office (EPO), which, if granted, can be validated in multiple European countries, and lastly (iii) the Unitary Patent (UP) route, which also begins with an EPO application, but upon grant, it can be

¹³³ Aplin, T., Radauer, A., Bader, M.A. *et al.* The Role of EU Trade Secrets Law in the Data Economy: An Empirical Analysis. *IIC* 54, 826–858 (2023). <http://doi.org/10.1007/s40319-023-01325-8>.

¹³⁴ Tschider, Charlotte and Ho, Cynthia M., Artificial Intelligence and Intellectual Property in Healthcare Technologies (July 16, 2024). Ch. 11: Artificial intelligence and intellectual property in healthcare technologies, in *Research Handbook on Health, AI and the Law* (Edgar, ed. Barry Solaiman & I. Glenn Cohen). Available at SSRN: <https://ssrn.com/abstract=4972529>

¹³⁵ Aplin, T., Radauer, A., Bader, M.A. *et al.* (2023). <http://doi.org/10.1007/s40319-023-01325-8>.

¹³⁶ De Noyette, E., Trade secrets in the EHDS (June 13, 2024). Available at <https://www.law.kuleuven.be/citip/blog/trade-secrets-in-the-ehds/>.

converted into a Unitary Patent for uniform protection across participating EU states, with disputes handled by the Unified Patent Court (UPC), offering centralized enforcement and simplified litigation.

Under the EPC, a computer program product, defined as a “*computer-implemented invention*”, in order not to be excluded from patentability, must fulfil the patentability requirements of novelty, inventive step and susceptibility of industrial application. Moreover, it must have a “*technical character*”, meaning that the software must produce a “*further technical effect*” going beyond the “normal” physical interactions between the program (software) and the computer (hardware) on which it is run¹³⁷.

Accordingly, whether computer-implemented simulations and predictive computer models are patentable at the EPO depends on whether the claimed invention has any technical features that prevent the claim as a whole from being excluded from patentability.

In this context, it is important to consider the 2021 G1/19 decision of the Enlarged Board of Appeal (EBA) of the EPO¹³⁸. The rationale behind this decision can be traced back to the 2006 Infineon case (T 1227/05), in which the EPO had previously addressed the question of patentability for computer simulations. The traditional difficulty with computer simulations is that they are frequently regarded as merely mental or mathematical techniques, which are precluded from patentability in accordance with Article 52 of the EPC. In the Infineon case, the EPO reached the conclusion that the computer simulation in question was patentable on the grounds that it was deemed to fulfil an adequately defined technical purpose. For nearly two decades, this case has served as the benchmark for determining the patentability of computer simulations.

In the G1/19 decision, the question was posed as to whether the criteria set out in Infineon would be applicable even in cases that lacked a direct link to physical reality. The case in question concerned a simulation of the movement of a pedestrian through an environment. The EBA’s conclusion was that computer simulations lacking a real-world link should not be excluded *a priori* from patent protection. The decision stated that the crucial factor is not whether the simulated system itself is technical, but whether the simulation contributes to the resolution of a technical problem. In this sense, the EPO has clearly diverged from the approach taken in the Infineon case.

In the context of a VHT infrastructure, it is nevertheless challenging to refute the assertion that simulations can facilitate the resolution of technical issues, even in instances where a tangible link to the real world may be absent. In light of the aforementioned considerations, the path of patentability emerges as a viable and promising avenue.

The same approach shall apply to computer-implemented inventions related to Artificial intelligence and machine learning. AI is based on computational models and mathematical algorithms which are *per se* of an abstract nature. Nevertheless, patents may be granted when AI leaves the abstract realm by applying it to solve a technical problem in a field of technology. “*For example, the use of a neural network in a heart-monitoring apparatus for the purpose of identifying irregular heartbeats makes a technical contribution. The classification of digital images, videos, audio or speech signals based on low-level features (e.g., edges or pixel attributes for images) are other typical technical applications of AI*”¹³⁹.

In relation to the possibility of patenting AI models, algorithms and systems, the fact that the patent application shall disclose the invention in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art (Art. 83 EPC) can also lead to difficulties. Indeed, since AI models may operate like “black boxes”, the inscrutability of such inventions may preclude a sufficient disclosure, with the effect of undermining the whole patenting process or, in any case, offending the

¹³⁷ Guidelines for Examination in the EPO, G-II, 3.6 Programs for computers. Available at: https://www.epo.org/en/legal/guidelines-epc/2024/g_ii_3_6.html.

¹³⁸ ECLI:EP:BA:2021:G000119.20210310 (March 10, 2021). Available at: <https://www.epo.org/en/boards-of-appeal/decisions/g190001ex1>.

¹³⁹ Guidelines for Examination in the EPO, G-II, 3.3.1 Artificial intelligence and machine learning. Available at: https://www.epo.org/en/legal/guidelines-epc/2024/g_ii_3_3_1.html.

societal *quid pro quo* by which inventors must disclose their inventions to receive twenty years of exclusive rights¹⁴⁰.

Another topic worth of consideration in this context is in relation to the patentability of inventions consisting in the results obtained thanks to the use of an AI system, which poses questions related to the inventorship and ownership of the invention itself¹⁴¹. On the one hand, it has already been made clear that “*a machine is not an inventor within the meaning of the EPC*”¹⁴². On the other hand, the solution to the ownership problem inevitably attribution relies on the respective contributions of the various persons in the chain of creation, in relation to and in connection with the AI system(s) that played a role in the invention-making process. In this regard, however, there is a high degree of uncertainty in relation to both the subjects that potentially may be considered owners of the patent and the threshold determining which contributor shall prevail.

In light of all the considerations above, it is imperative that the management of IPR pertaining to both background data and knowledge, as well as the outcomes of their processing, is explicitly delineated in a comprehensive contract or policy. To this end, it will be essential to:

- Define **rules for accessing and using background IP**, including datasets, models, and tools provided by the stakeholders. This shall include information related to licenses, terms of access, usage rights, attribution, security measures and dependencies solution measures.
- **Establish guidelines for the creation, ownership, and exploitation of new IP** generated through the VHT Platform. This shall include information related to ownership (and joint ownership) principles, commercial exploitation or non-commercial use, IP protection methods or reliance on open-source standards, disclosure requirements and patenting strategy.
- Set up a **framework for the establishment of IP ownership arrangements and related responsibilities**. This should include consideration of potential licensing agreements and remuneration options linked to the use of IP resulting from the use of the services encompassed by the infrastructure.

In the implementation of such a policy, the decision regarding the use and dissemination of information, and therefore the selection of licensing models (amongst both infrastructure users and external parties), is largely contingent upon a comprehensive evaluation of the diverse interests represented by contributors and the IPR associated with the assets in question.

As mentioned above, management of certain copyrighted software components could rely on “non-IP” mechanisms which, in many cases, may be more significant than exclusive rights for promoting innovation. Adoption of an open-source model would require software to be distributed with its source code and be subject to licenses in which the copyright holder grants subsequent users rights to use, modify, and distribute such software. Open-source software facilitates commons-based “peer production” in which large numbers of unconnected programmers contribute to massive, collective software projects.

It is also important to note that the protection of software as a literary work under copyright law is somewhat limited. Copyright law affords protection to the specific form of expression (primarily the source code) from identical (either total or partial) copies. It can be reasonably deduced that for those types of programs, including DTs themselves, which are the result of the collaboration of the users of the infrastructure resources, patent protection would be a more valuable asset, where the requisite conditions exist.

¹⁴⁰ Lee, Peter (2024).

¹⁴¹ Shemtov, N., A study on inventorship in inventions involving AI activity (February 2019). Available at: https://link.epo.org/web/Concept_of_Inventorship_in_Inventions_involving_AI_Activity_en.pdf.

¹⁴² ECLI:EP:BA:2021:J000820.20211221 (December 21, 2021). Available at: <https://www.epo.org/en/boards-of-appeal/decisions/j200008eu1>.

It is therefore important to ensure that the patenting of VHT-related software is not unnecessarily hindered, while also maintaining the goal of allowing a certain degree of openness. To this end, it is worth to recall that even models, algorithms and software that could be eligible for patenting and meet the necessary requirements may rely on use and distribution models that benefit a significant number of contributors and users.

Indeed, inventions deemed essential for the implementation of specific technology standards may be designated as Standard Essential Patents (SEP). Standards are developed by Standards Developing Organizations (SDOs), such as the International Organization for Standardization (ISO), with the objective of ensuring interoperability, compatibility, and quality across products from different manufacturers. In accordance with the relevant regulations, holders of SEP patents are obliged to license their invention under FRAND (Fair, Reasonable, and Non-Discriminatory) terms. This guarantees that other companies can access and implement the standard without encountering excessive costs or unfair restrictions. However, the determination of what constitutes “fair” or “reasonable” can give rise to disputes, frequently involving litigation over licensing fees and infringement claims.

Even in cases where patents relate to inventions that are not essential for the implementation of a standard, but which nevertheless represent an asset for the wider community, it may be possible to implement a general licensing program that establishes uniform conditions for all participants and/or third parties interested in using or further developing such technologies and instruments. It is of the utmost importance to conduct comprehensive planning for the IPR management policy and to implement a process of regular updates.

10.2 CHALLENGES FOR IPR MANAGEMENT

In consideration of the IPR framework delineated in Section 10.1, it is evident that the attraction and utilization of background IP, in addition to the protection and monetization of foreground IP, are primarily within the purview of the aforementioned interested parties and, consequently, the partners of the VHT ecosystem.

Nevertheless, a number of significant issues have been identified that require further investigation in order to gain a deeper understanding of the challenges (if not obstacles) involved in establishing an IPR management policy that:

- enables the lawful exploitation of third-party data, and/or the continuous attraction of new partners carrying new data and knowledge to the benefit of the whole ecosystem;
- enables different forms of remuneration for the efforts, both in terms of sharing of information and of further elaboration thereof, of the partners;
- ultimately, and as already mentioned, enables partners to create a VHT ecosystem that not only benefits the wider community, but also ensures the continued viability and financial sustainability of the same.

10.2.1 EHDS

The first topic which deserves further analysis is the proposal for the EHDS Regulation¹⁴³.

Art. 33a provides that “*electronic health data entailing protected intellectual property [and] trade secrets [...] shall be made available for secondary use in accordance with the principles set forth in this Regulation*”, also specifying that in such case “*all specific appropriate and proportionate measures, including legal, organisational, and technical ones, [...] necessary to preserve the protection of intellectual property rights [and] trade secrets*” shall be taken.

It essentially mandates that data holders, including private entities, make certain categories of electronic health data available for secondary use, even if such data entails protected intellectual property and trade secrets. As previously noted, this constitutes an exception to the exclusivity of IPR, which may

¹⁴³ Please refer to Section 5 of this document.

have a significant impact on health datasets that are subject to copyright and database protection, or that constitute trade secrets.

While this may facilitate the development of robust DTs by enabling access to diverse and comprehensive datasets, it can also raise some issues, as it triggers some tension between data accessibility and IPR protection, to the point that the obligation to share data protected by IPR may even deter investment in research and development, as companies might be reluctant to develop new technologies or data-driven solutions if they are required to disclose proprietary data.

Moreover, the broad definitions and requirements in Article 33a may lead to legal uncertainties regarding the scope of data to be shared and the extent of IP protection. This ambiguity complicates compliance efforts for data holders and may result in inconsistent interpretations across Member States. In this regard, from the perspective of a private company, the decision to allow the Health Data Access Bodies (HDABs) to have sole responsibility in determining which data are to be protected by IPR or protected as trade secrets may lead to controversial outcomes.

It is notable that the allowed purposes for secondary use of electronic health data under the EHDS Proposal do not expressly exclude revenue-generating commercial purposes, such as product development. This means that, in principle, commercial organizations, such as technology companies, pharmaceutical companies, and medical device manufacturers can all take advantage of the EHDS rules on secondary use to gain access to data held by third parties. In the context of the VHT ecosystem, this aspect could be particularly relevant, and to some extent concerning, as questions arise about whether commercialization aligns with the Regulation's intent if the VHT infrastructure is developed using health data accessed under the EHDS.

In this regard, however, partners must also consider at least two aspects:

- ✓ access to datasets under the EHDS does not negate the necessity of considering the respective IPR owners. This is particularly relevant when assessing the IPR that may be exploited in the secondary use of the data, such as the possibility of co-ownership of a patent resulting from the innovation.
- ✓ Article 41a, par. 3 provides that "*health data users shall make public the results or output of the secondary use of electronic health data*". This implies that the use of data acquired in accordance with the EHDS Regulation may have an impact on the options that the partners have to protect the output or the results of their research. This, in turn, may become an important consideration for data users before submitting a data application or a data request.

In light of the aforementioned considerations, it is recommended that the EHDS should ideally provide for, on the one hand, uniform guidelines at EU level regarding the scope of data to be shared and the extent of IP protection; on the other hand, it should also provide for specific measures, including legal, organizational and technical measures necessary to maintain the confidentiality of protected data. It is similarly imperative that clear guidelines be established regarding the processes and criteria employed by HDABs in evaluating the protection of IPR.

Furthermore, in order to prevent the deterrence of investment in research and development by data holders, adequate remuneration must be provided. This should entail that the monetary compensation scheme for making data available for secondary use reflects the costs associated with providing the data and all related services, including the collection of data and their preparation, anonymisation (synthetization) and pseudonymization. Furthermore, it must account for the costs associated with the risks inherent to the mandatory sharing of data covered by IPR.

Concurrently, the partners of the VHT ecosystem shall implement a policy that, first and foremost, contains specific guidelines for the submission of data applications and requests, as well as the utilization of data accessed under the EHDS. This policy will consider both the advantages and disadvantages of data usage, including, for example, the rights of data holders and transparency obligations. Moreover, it shall establish specific conditions within this context, thereby limiting the discretion of participants in the VHT ecosystem. Finally, licensing models may be established whereby commercial entities can access data under terms that require reinvestment into public health or shared

benefits with public institutions; such an approach would contribute to public health objectives and comply with ethical standards in a manner consistent with the spirit of the EHDS Regulation.

10.2.2 OPEN SOURCE

It is evident that certain issues associated with the utilization of protected material would not be applicable in the context of freely accessible knowledge, such as the utilization of freely accessible datasets and tools.

It is important to note that open source is not solely associated with software; it can also be applied to datasets, including those containing healthcare data that are essential for training AI systems in the VHT infrastructure. The distribution of such datasets without commercial restrictions has the potential to foster greater accessibility and transparency in AI development, including within the context of the VHT ecosystem. However, the open-source paradigm for AI is fundamentally more complex than for traditional software, as it requires the alignment of multiple elements that are essential for the functioning of AI systems.

The concept of open source was originally conceived with the intention of guaranteeing that developers would be able to utilize, examine, modify and disseminate software without encountering any restrictions. Nevertheless, the tenets of open source do not align seamlessly with those of AI. In the context of software, developers typically require access to the source code. In contrast, working with AI frequently requires access to a trained model, its training data, the preprocessing scripts, the code governing the training process, and the model's underlying architecture, among other components. The multiplicity of components involved creates significant obstacles to the implementation of open-source principles in AI systems, particularly in the context of healthcare data, where issues of privacy, security, and intellectual property converge.

The definition of what constitutes “open source” in respect to AI remains an open question, and institutions could play a pivotal role in providing clarity and guidance on this matter. One illustrative example of an approach to addressing these concerns is the AI Act, which requires that generative AI models disclose the training datasets and methodologies used in their development. In particular, Articles 53 and 54 provide that the obligations set out in the Regulation “*shall not apply to providers of general-purpose AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available*”.

Notwithstanding the aforementioned ambiguities, there is a general consensus regarding the advantages that an open-source framework can offer in the context of AI development. Such benefits include:

- creating standards that enhance interoperability: open-source models advance research, reduce costs, offer flexibility and customization, and empower developers to foster innovation;
- fostering transparency to build trust: the datasets and codes of open-source models can be audited and verified by third parties, which helps to ensure their quality and reliability, thus making the products safer and easier to identify the source of bias in the training data or the model architecture itself;
- ensuring control over development processes: open models facilitate the assessment of AI systems by regulators and civil society for compliance with laws protecting civil rights, privacy, consumers, and workers;
- simplifying compliance with regulations.

This latest point is worthy of particular attention. As previously indicated, the AI Act provides an illustrative example in this regard. Indeed, the Regulation expressly provides that “*This Regulation does not apply to AI systems released under free and open-source licences*” (Article 2), with some notable exceptions (such as for high-risk AI systems or prohibited AI practices). This is particularly significant in light of the considerations set forth in Recital 102, which states that “*software and data, including*

models, released under a free and open-source licence that allows them to be openly shared and where users can freely access, use, modify and redistribute them or modified versions thereof, can contribute to research and innovation in the market and can provide significant growth opportunities for the Union economy”.

It would be beneficial for institutions to consider actively encouraging the adoption of open-source licenses by providing incentives or regulatory measures that are similar to those outlined in the AI Act. As an illustration, datasets and tools distributed under open-source licenses could be deemed eligible for compliance exemptions or could benefit from reduced oversight requirements. Such measures would serve to reduce the barriers to entry for developers, thereby encouraging the use of open standards and fostering a more collaborative and innovation-friendly environment.

Nevertheless, the choice to utilize open-source materials must also consider the downstream implications. For example, the utilization of open-source tools may impose limitations on the manner in which the outcomes of subsequent development can be commercialized or disseminated, contingent upon the conditions of the licenses involved.

The selection of an appropriate licensing model is of paramount importance in such circumstances. Permissive licenses such as Apache 2.0 are among the most popular licenses in the AI community , allowing significant freedom in the use and modification of open-source materials, even for commercial purposes. Nevertheless, more restrictive licenses may impose obligations on derivative works, thereby limiting the scope of commercial exploitation. It is therefore imperative that the adoption of open-source principles in AI development be subject to careful regulation within the IPR management policy of the VHT infrastructure. It is essential that such policies strike a balance between the need for openness and collaboration on the one hand, and the protection of proprietary interests and compliance with legal frameworks on the other hand, in order to ensure the sustainability of the ecosystem.

10.2.3 PATENTABILITY OF AI MODELS AND ALGORITHMS

As previously mentioned, given the potential limitations of copyright protection, particularly with regard to the software components of the VHT infrastructure, patent protection may offer a more valuable asset, provided that the necessary conditions are met. Nevertheless, pursuing effective patent protection for the outcomes of research and innovation activities conducted through the VHT infrastructure (which, by way of example, may encompass both proprietary tools utilized for data elaboration/production and the final results of the activity) could present a number of challenges. Balancing these interests with the necessity of fulfilling public health objectives may increase the overall complexity.

A primary aspect to focus on regards the requirement that the patent application must disclose the invention in a sufficiently clear and complete manner for it to be carried out by a person skilled in the art (Art. 83 EPC). As already highlighted, this sufficiency of disclosure represents a fundamental tenet of the patent system, ensuring a societal balance by which inventors disclose their innovations in order enriching the public knowledge base, in exchange for a time-limited monopoly.

However, when dealing with AI models, as an integral part of the VHT, this requirement presents unique challenges. As mentioned above, a case in point is given when AI systems operate as “black boxes”, whereby the internal workings of algorithms, particularly those utilizing machine learning, remain opaque even to their developers. This opacity risks implying inadequate disclosure, which could render the patent invalid or impede its enforcement. Conversely, the stipulation does not require the complete disclosure of an AI system’s source code. Applicants must provide sufficient detail to enable the reproduction of results without disclosing information that is not essential for understanding the technical contribution of the invention. This flexibility can be advantageous for innovators who wish to protect their core intellectual assets while fulfilling legal obligations.

To gain a deeper insight into the magnitude of this issue, it is instructive to examine the European case law, such as the EPO Boards of Appeal decision in T 0161/18¹⁴⁴, which underscores the imperative of

¹⁴⁴ ECLI:EP:BA:2020:T016118.20200512, dd. 12 May 2020. <https://www.epo.org/de/boards-of-appeal/decisions/t180161du1>

disclosing particular technical details of AI systems, including the training data or processes, when these elements are pivotal to the invention’s functionality: in that case, it has been held that as “*the application does not disclose which input data are suitable for training the artificial neural network according to the invention, or at least one data set suitable for solving the present technical problem*”, therefore “*the training of the artificial neural network cannot be reworked by the person skilled in the art*”. Consequently, those seeking patent protection must tread a fine line, disclosing sufficient information to satisfy the sufficiency requirement while avoiding any compromise of proprietary knowledge or competitive advantage.

Given the opaque nature of AI systems, a useful comparison can be drawn with patents on biological material, which often involve similarly complex and sensitive disclosures. In accordance with Rule 31 of the EPC Implementing Regulations on the Grant of European Patents, the deposit of biological material in a recognized depository may satisfy the disclosure requirement in instances where the material is not reproducible by conventional means. By way of analogy, in the context of AI and VHT patents, innovative approaches such as depositing detailed descriptions of training methodologies or providing access to representative datasets might be explored in order to meet sufficiency standards while safeguarding sensitive data.

It would be advantageous for European institutions to address the disclosure challenges inherent to AI patents by adopting measures analogous to those set forth in Rule 31. The creation of secure repositories for AI-specific materials, including trained models, datasets, and algorithms, would allow inventors to fulfil the sufficiency of disclosure standards while safeguarding sensitive information. The EPO could issue supplementary guidelines that elucidate optimal methodologies for delineating AI-related innovations and delineating the circumstances under which deposits are required. It would be beneficial to establish rules that address the opaque nature of AI and mandate clear explanations of the technical effects involved and their correlated performance metrics. International collaboration through the World Intellectual Property Organization (WIPO) could facilitate the harmonization of standards, thereby ensuring global consistency. By continuously updating these frameworks, it may be possible to achieve a balance between robust intellectual property protection and societal transparency, thereby fostering innovation and ethical AI development.

Such a trade-off pertains to the dual role of patents in incentivizing and potentially inhibiting innovation, which may be particularly significant in the context of the VHT. The conferral of exclusive rights through the protection of intellectual property encourages investment in the development of sophisticated AI systems, as the ownership of these rights ensures a return on the substantial costs incurred during the research and development phase. In the context of the VHT, such protection serves to incentivize advancements in personalized medicine, simulation-based diagnostics, and education. Nevertheless, exclusivity, particularly in regard to foundational AI technologies, may give rise to monopolistic practices that impede competition and restrict the accessibility of critical innovations.

To achieve a balance between these concerns, it would be beneficial for the VHT ecosystem to implement an IPR management framework that is inspired by the Standard Essential Patent (SEP) model. To facilitate broad adoption while safeguarding the rights of inventors, SEPs require licensing that is fair, reasonable, and non-discriminatory (FRAND). A comparable approach could be applied to the VHT ecosystem, where specific foundational AI tools or methods are essential for interoperable and scalable systems.

The implementation of a FRAND-like model within the VHT domain would facilitate the licensing of proprietary innovations on equitable terms. Such a framework would prevent the over-concentration of power among a few patent holders while ensuring the widest possible access to critical technologies. To illustrate, proprietary algorithms or synthetic data generation techniques – which are central to VHT applications – could be made available through licensing mechanisms that promote accessibility without undermining patent incentives.

The European Commission’s recent proposal for a Regulation on SEPs provides a timely and relevant framework to inform this approach. By introducing mechanisms for transparency, essentiality checks, and FRAND dispute resolution, the regulation seeks to balance innovation incentives with accessibility.

Adopting similar principles for VHT technologies could ensure robust IP protection while fostering a collaborative and scalable ecosystem. Such an IPR policy would not only mitigate innovation-inhibiting effects of exclusivity but also align the VHT ecosystem with broader public health goals by fostering collaboration, transparency, and scalability.

11 Ethical implications

The use of Digital Twin technology in healthcare introduces a groundbreaking shift toward deeply personalized, predictive, and dynamic patient care. DTs, furthermore, can vastly improve clinical precision and enable proactive management of health conditions, ensure better diagnostics, less invasive treatments and faster medicine discovery, as well as help fine-tune and speed-up regulatory pathways for medical devices and clinical trials.

However, with these advantages come intricate ethical challenges that must be critically addressed. Ensuring the responsible development and deployment of DTs requires examining concerns around, *inter alia*, privacy, data integrity and quality, patient autonomy, human oversight, equality, and the potential for over-reliance on AI-driven recommendations.

The figure below – taken from a renowned publication¹⁴⁵ – provides an overview of some of the main socio-ethical issues arising from Digital Twins in health.

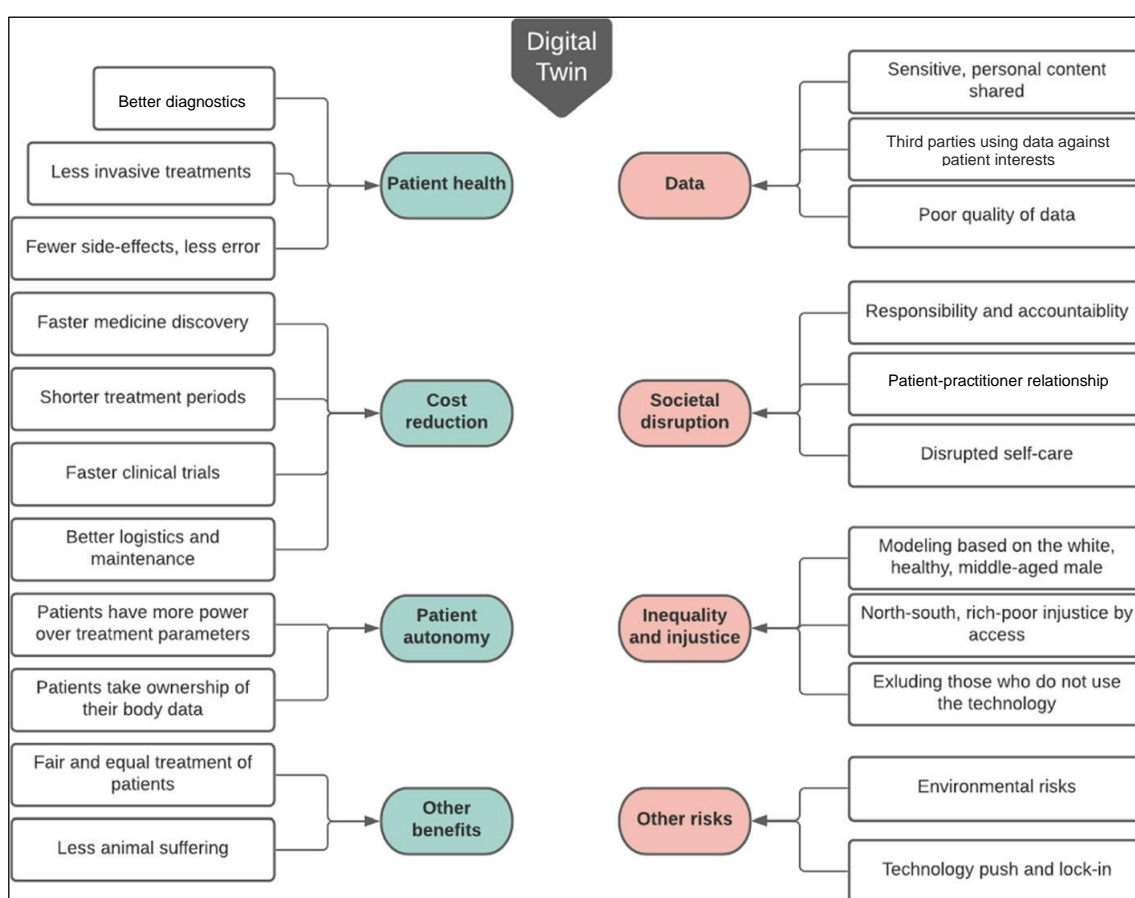


Figure 16 – Overview of the ethical issues surrounding Digital Twins in health¹⁴⁵

Moreover, the scenario is further complicated as the VHT integrates several emerging technologies, including AI, Internet of Things, wellness Apps, big data and robotics, with each of these components contributing its own set of socio-ethical challenges to the resulting creation.

¹⁴⁵ Popa, E.O., van Hilten, M., Oosterkamp, E. *et al.* The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks. *Life Sci Soc Policy* 17, 6 (2021). <http://doi.org/10.1186/s40504-021-00113-x>.

The other table set out below (source¹⁴⁶) summarizes the risks emerging from the ethical challenges associated with the VHT.

Ethical Implications	Emerging risks
Technical Limitations and Malfunctioning	Biases and reproduction of socially embedded stereotypes Reliability issues Limitations of the representation based on data Level of maturity and machine autonomy Deceiving against performance expectations Opacity of the system and how it operates Portability and operability issues
Privacy related issues	Personal data, including remote real-time collection, use and repurposing Data minimization negates the holistic approach to representation Privacy preferences and requirements, regulation or applying personalization Consent mechanism and data sharing Consent withdraw and data storage
Security related vulnerabilities	Cybersecurity Data theft Identity theft and biometrics hijacking
Human autonomy and self-determination	Influence over decision-making processes Nudging behavior Undermining trust Devaluation of human capacities
Control over the system	Stakeholders interplay Human agency Shared responsibility Machine autonomy
System and data ownership	Property rights Transfer of ownership Liabilities and accountability
Production/Operation costs	Sustainability Economic costs Costs for individuals and society
Inequalities in access	Affordability issues for the end user Capacity of existing infrastructures Deepening social inequalities Forced adoption to avoid disadvantageous conditions
Lack of regulation and accountability uncertainties	Specificities of HDTs and legal gaps Accountability issues Law enforcement issues
Deceptive, exploitative and/or manipulative practices	Dual use Profiling and targeting Exploiting specific vulnerabilities Oppression, coercion, theft and extortion
Mass surveillance, social control, erosion of democratic values and social cohesion	Abusive and misuse Social control Increased power imbalances
Paradigm shifts	Erosion of the value of privacy Disruption of economy Devaluation of human capacities Disruption of human-human relationships Extension of human life

Figure 17 – Risks emerging from ethical implications of Digital Twins in health¹⁴⁶

The scope of ethical assessment is to identify and potentially prevent any negative societal impacts of technology, while also enabling stakeholders from diverse sectors to consider the costs and benefits involved. This approach aligns with a broader cultural view that science and technology should be open to public engagement, where significant changes brought about by technological advances undergo critical examination and dialogue to assess their anticipated socio-ethical and legal impacts.

Ethical principles play an essential role in encouraging stakeholders to protect and advance key human values, complementing the more rigid demands posed by applicable laws and regulations. While legislators establish enforceable rules, ethics often serves as the driving force behind laws, urging individuals and organizations to pursue actions that respect and uphold societal principles. Additionally,

¹⁴⁶ Fontes, C., Carpentras, D. & Mahajan, S. Human digital twins unlocking Society 5.0? Approaches, emerging risks and disruptions. *Ethics Inf Technol* 26, 54 (2024). <https://link.springer.com/article/10.1007/s10676-024-09787-1>.

ethics frequently lays the groundwork for legislation, as many laws arise from addressing ethical dilemmas. Despite this alignment, however, legal frameworks alone cannot prevent all morally questionable outcomes. By offering direction and establishing mindful boundaries, ethical values remain essential, fostering responsible conduct in ways that laws might not strictly enforce, in those areas where societal well-being deeply depends on moral accountability¹⁴⁷.

Clearly, evaluating the socio-ethical consequences of new technologies is especially crucial at the early stages of their development. In these initial phases, decisions shaping the path from foundational research to practical applications are still being made, offering greater flexibility to guide their outcomes. Additionally, early ethical scrutiny allows for intervention before these technologies begin to significantly affect society: when the designs are not yet solidified, greater opportunity exists to adjust their course, making it possible to manage potential societal impacts more effectively. For emerging technologies, some assessment methods take an exploratory approach, delving into potential socio-ethical implications even when they are not fully clear¹⁴⁸.

This section aims to briefly explore the ethical dimension of the VHT, based on scientific literature and building on the recommendations provided in (i) the Ethics guidelines for trustworthy AI developed by the High-Level Expert Group on Artificial Intelligence (AI HLEG) set up by the EU Commission¹⁴⁹, and (ii) the FUTURE-AI guideline, which provide for a comprehensive, consensus-driven framework aimed at establishing trustworthy and deployable AI in healthcare¹⁵⁰. This latter international guideline emphasizes that trustworthy AI in clinical settings hinges on both technical and ethical rigor, addressing issues ranging from safety and robustness to transparency, fairness, and social accountability. A central component of FUTURE-AI is its emphasis on six guiding principles designed to foster trust and acceptance among patients, clinicians, and health institutions. These principles – Fairness, Universality, Traceability, Usability, Robustness, and Explainability – create a structured foundation for developing AI that aligns with ethical and societal expectations¹⁵¹.

The issues discussed below do not specifically include compliance with applicable regulations and the corresponding responsibilities of those involved, as these topics are extensively covered in a significant portion of this document. In this section, legal accountability shall be regarded as the foundation for adhering to the other applicable ethical principles outlined below, as it ensures that actions, decisions, and systems not only adhere to moral and societal standards but also comply with the legal frameworks in force. By grounding ethical behaviour in legal compliance, accountability fosters social trust and helps mitigate the risks deriving from technological advancements.

A) Privacy and data protection

Privacy is one of the most pressing ethical concerns in the context of Digital Twins in health, primarily due to the vast and sensitive range of data that relevant models need gather, process and analyse. The VHT is designed to aggregate data from multiple sources, including both clinical and non-clinical data. While this comprehensive collection enables the delivery of highly personalized insights, it simultaneously creates a repository of sensitive information that is highly vulnerable to misuse and breaches. A violation within a VHT framework could lead to unauthorized access or exposure of a

¹⁴⁷ See Elisabetta Biasin, ‘In Silico World D9.1 Legal and Ethical Inventory’. <https://zenodo.org/records/7104079>.

¹⁴⁸ Popa *et al.*, <http://doi.org/10.1186/s40504-021-00113-x>

¹⁴⁹ The full version of the ‘Ethics guidelines for trustworthy AI’ is available here. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai/>

¹⁵⁰ Lekadir, K., Feragen, A., Fofanah, A. J., Frangi, A. F., Buyx, A., Emelie, A., Lara, A., Porras, A. R., Chan, A.-W., Navarro, A., Glocker, B., Botwe, B. O., Khanal, B., Beger, B., Wu, C. C., Cintas, C., Langlotz, C. P., Rueckert, D., Mzurikwao, D., ... Starmans, M. P. A. (2023). *FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare*. arXiv. <https://arxiv.org/abs/2309.12325>.

¹⁵¹ FUTURE-AI’s lifecycle approach spans the design, development, validation, and deployment of AI in clinical practice. For each stage, it outlines specific practices aimed at aligning AI development with ethical standards, addressing critical risks such as privacy breaches, bias, lack of interpretability, and over-reliance on AI outputs. In particular, the framework incorporates regular auditing, governance measures, and risk management strategies to monitor AI’s real-world performance and maintain continuous alignment with regulatory standards and ethical norms. Additionally, the guideline encourages stakeholder engagement at every stage, promoting a collaborative approach that involves clinical, technical, ethical, and regulatory expertise to refine AI for safe and effective healthcare integration.

patient’s entire medical and lifestyle history, causing identity theft, reputational damage and significantly harmful misuse of data.

Thence, ensuring suitable levels of data security and protection is a primary challenge for these data-driven innovations widely gain social consent and license. Indeed, privacy still emerges as one of the most critical issues, when considering value alignment and ethical implementation of systems that rely on personal data, as well as a factor undermining public trust and acceptance of technologies leveraging AI¹⁵².

In view of the EHDS, for instance, a key challenge to address (especially for data users) will be maintaining data anonymization, as per applicable obligation, while keeping the Digital Twins highly tailored and so personalized to each patient.

Additionally, DTs spark complex questions around data ‘ownership’ and informed consent. Many patients may not fully grasp the scope of the data being collected, nor the potential for this data to be used in unforeseen ways (*i.e.*, repurposed for permitted purposes, such as research, healthcare optimization and policymaking)¹⁵³. Accordingly, ethical principles connected to the VHT require the utmost transparency *vis-a-vis* the patients, providing clear and detailed explanations about data collection and processing purposes before their data are collected, empowering them to make informed choices and exploit their full degree of autonomy over healthcare decisions, including the right to opt-out or withdraw their consent¹⁵⁴.

Long-term data usage further complicates informed consent, as individuals may initially consent to specific purposes but later change their preferences. As already specified in this D6.2, more adaptive consent models are necessary (*e.g.*, ‘broad consent’ or ‘dynamic consent’) to ensure ongoing alignment with ethical standards, allowing patients to granularly adjust their consent based on evolving preferences¹⁵⁵.

B) Accuracy and misrepresentation

A core feature of the VHT is their ability to represent a patient’s current health status dynamically and predictively, with the accuracy and reliability of this representation being crucial. Errors or limitations in the model could lead to misinterpretations of an individual health condition, possibly resulting in inappropriate or harmful treatment recommendations.

In the context of Digital Twins in health, data are inherently complex and multi-dimensional, including not only quantitative metrics such as heart rate, body temperature, and other physiological indicators, but also more subtle elements like cognitive abilities and emotional states. As a result, diverse data collection approaches are essential, involving sensors, human input, psychological assessments, and insights derived from machine learning¹⁵⁶. Additionally, the data underlying the model need to be regularly updated, as obsolete or incomplete information can hinder effective decision-making and may even lead to counterproductive outcomes in healthcare, personalized services, and human-machine interactions.

Patients may want to have direct access or influence over their Digital Twin’s data, including the ability to update or delete personal information, potentially impacting the accuracy and reliability of the associated models. The exercise of these rights according to the applicable legislation (particularly the GDPR and Member States’ national data protection laws) can, under certain specific circumstances connected to scientific research, be restricted when likely to render impossible or seriously impair the achievement of the objectives of data processing. Aside from these cases, data controllers who build up

¹⁵² *Ibid*. The authors point out that the privacy risks are also “*amplified by the fact that the digital twin is arguably a data intense scenario [which] can multiply the digitalization of information to an unprecedented degree*”.

¹⁵³ For more details, please refer to the previous sections dedicated to the GDPR and the EHDS.

¹⁵⁴ Huang PH, Kim KH, Schermer M. Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study *J Med Internet Res* 2022;24(1):e33081. <https://www.jmir.org/2022/1/e33081>.

¹⁵⁵ Iqbal JD, Krauthammer M, Biller-Andorno N. The Use and Ethics of Digital Twins in Medicine. *J Law Med Ethics*. 2022;50(3):583-596. <https://doi.org/10.1017/jme.2022.97>.

¹⁵⁶ Venkatesh, K. P., Raza, M. M., & Kvedar, J. C. (2022). Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *NPJ Digital Medicine*, 5(1), 150. <https://doi.org/10.1038/s41746-022-00694-7>.

Digital Twins are required to timely and unconditionally implement patients' requests to exercise their rights (e.g., to 'be forgotten', to withdraw the consent, to opt-out, to request data portability or data rectification).

Clearly, the effectiveness of a DT largely relies on the quality and comprehensiveness of the data it is fed with¹⁵⁷. Should any disparity between the digital model and the patient's actual health status exist, then healthcare decisions could be misguided (e.g., a Digital Twin which does not account for recent lifestyle changes might inaccurately predict disease progression, leading to unnecessary or even harmful interventions).

Therefore, ethical principles must emphasize rigorous validation processes to ensure DTs accurately represent the patient's health at every stage, minimizing the risk of errors.

Moreover, as DTs are designed to enhance personalized health care, the VHT face ethical risks also from the epistemic side. The quality of the data collected may fall short of meeting intended objectives, such as achieving a thorough understanding of an individual's health or accurately predicting their risk of developing specific diseases.

Threats to data quality and so accurate individual health representation further increase when it comes to wearables and wellness Apps, which now enable the collection of a wide range of bio-signals. While non-medical-level wearables provide an accessible way for the public to track and manage lifestyle habits, the data they generate often lacks the precision required for clinical use, as they are not required to undergo the strict quality standards which apply to medical devices and SaMD. Rather than advancing personalized health care, relying on data from commercial wearables could lead to an individual distorted digital representation. From both ethical and legal standpoints, developers are required to carefully assess the level of data accuracy necessary for the services they intend to create¹⁵⁸.

C) Patient autonomy

Patient autonomy, in the context of the VHT, takes on new dimensions that are well illuminated by Beauchamp and Childress's principles of biomedical ethics. Accordingly, respect for individual autonomy is a foundational ethical guideline that ensures individuals have the freedom to make informed, voluntary choices about their own healthcare. In the context of Digital Twins technologies – where virtual representations of patients are created using vast amounts of sensitive health data to simulate, predict and personalize medical interventions – upholding patient autonomy requires careful attention to issues of informed consent, transparency and control over digital data representations.

Autonomy is traditionally rooted in patients' right to make decisions about their body and medical treatments, free from any coercion. With the VHT, this concept is extended to include decisions about the use, access, and modifications of the digital replica of their health profile.

Informed consent, a cornerstone of autonomy, becomes vastly more complex when patients must authorize not only direct medical interventions, but also the continuous and often sophisticated ways their data will be used. Given that DTs integrate (real-time, in some cases) health data and predictive analytics, patients must understand the ongoing nature of this technology, which can be challenging due to the complex algorithms and data inputs involved. As already said above, to effectively respect individual autonomy, healthcare providers and VHT developers must ensure patients fully comprehend what they are consenting to, including how their digital representations will be used, updated and possibly shared with third parties.

The principle of autonomy in Beauchamp and Childress's framework also emphasizes the importance of freedom from undue influence or manipulation. The VHT, however, may introduce subtle pressures on patients to accept recommended treatments derived from these models, potentially compromising their ability to make free and independent decisions. For instance, if a DT's predictive analysis suggests

¹⁵⁷ Fontes, C. et al: "Without data that encompasses the multifaceted attributes of a human – ranging from physical and physiological aspects to cognitive and emotional traits – an HDT remains a skeleton devoid of actionable insights". <https://link.springer.com/article/10.1007/s10676-024-09787-1>.

¹⁵⁸ Huang et al. <https://www.jmir.org/2022/1/e33081>.

a high risk for a particular condition, a patient might feel compelled to follow specific medical recommendations to mitigate that risk. Even if these insights can be beneficial, the perceived ‘authority’ of the Digital Twin’s predictions could lead patients to feel as though they have no choice but to comply, subtly undermining their autonomous decision-making. Moreover, there is an inherent risk that healthcare providers might prioritize DT-generated insights over patient-reported symptoms or preferences, especially when these insights are backed by large data sets and complex algorithms. Thus, respecting patient autonomy involves careful communication that supports patients in making choices that align with their own values, rather than feeling obligated by a data-driven model based on machine learning¹⁵⁹.

Clearly, transparency and the right to access and manage personal data play a key role in supporting autonomy. Patients who do not have access to or precise understanding of the data fed to their DT cannot make truly self-governed choices about their processing, *i.e.*, consenting or not. This results in the obligation of ensuring that patients are not only informed about the initial purposes for which their data are collected and any envisaged secondary use, but also empowered to access and adjust their data at any moment. This approach aligns with the concept of autonomy as it allows data subjects to influence the ongoing development and application of their digital replica, ensuring they accurately reflect their health and personal preferences¹⁶⁰.

Finally, patients’ autonomy must address potential conflicts between their right to make personal health decisions and broader ethical concerns related to beneficence and non-maleficence. For instance, DTs are designed with the intent to maximize health outcomes (beneficence) and avoid any kind of harm (non-maleficence), sometimes through predictive and preventative interventions that may not align with the patient’s wishes. Respecting autonomy in this situation means acknowledging the individual right to decline interventions, even if the recommendation of the VHT suggests otherwise. Therefore, the ethical use of Digital Twins requires a balance between the predictive capabilities of this technology and the respect for individual preferences and freedoms, ensuring that the patient’s autonomy remains at the centre of any decision-making process.

D) Ownership and control

The concept of ‘data ownership’ – which, it should be highlighted, is never mentioned as such in data protection or other applicable legislation – is complex and multifaceted in the context of DTs, raising crucial questions about control, rights and responsibilities over the data and digital representations of individuals¹⁶¹.

Ownership in this setting first touches on who has the primary rights to control and manage the data within the VHT model. Patients may feel that, as the source of the data, they should retain ownership and have the final say in how their digital replica is used, updated, or even deleted¹⁶². This perspective aligns with concepts of personal autonomy and privacy, reinforcing the idea that individuals should have control over their personal data, especially when it includes intimate health information. However, because DTs are typically managed by healthcare providers, research institutions, or even private companies specializing in digital health technology, these entities often assert their own claims over the data¹⁶³. They may argue that their technical investment in creating and maintaining the VHT models and infrastructure – along with the proprietary algorithms that generate insights from the data – entitles them to a degree of ownership or control over the digital assets.

Furthermore, some considerations made above with reference to data access, rectification and cancellation are especially pertinent in this context, as disputes over data (and Digital Twin) ownership

¹⁵⁹ Bruynseels K, Santoni de Sio F, van den Hoven J. Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Front Genet.* 2018 Feb 13;9:31. <https://doi.org/10.3389/fgene.2018.00031>: “‘Dataism’ in this context might become a new form of medical paternalism”.

¹⁶⁰ Popa *et al.* <http://doi.org/10.1186/s40504-021-00113-x>.

¹⁶¹ Hummel P, Braun M, Dabrock P. Own data? Ethical reflections on data ownership. *Philos Technol* 2021;34(3):545-572. <https://doi.org/10.1007/s13347-020-00404-9>.

¹⁶² Teller M. 2021 Legal aspects related to digital twin. *Phil. Trans. R. Soc. A* 379: 20210023. <https://doi.org/10.1098/rsta.2021.0023>.

¹⁶³ Popa *et al.* <http://doi.org/10.1186/s40504-021-00113-x>.

may also impact on the rights to access and modify the DT itself. Healthcare providers may want to control access to ensure that data within the Twin remains accurate and clinically relevant. Conversely, patients may claim their right to review, alter, or remove specific information they find inaccurate or irrelevant. This can create a conflict of interests, as modifying or omitting data may impact the efficacy of the DT and its predictive capabilities.

Questions also arise around the ownership of derivative insights generated by a Digital Twin. For instance, if a patient's DT is used to discover a pattern or risk factor valuable for medical research, should the patient retain rights to that insight, or do only healthcare providers or researchers hold the ownership, and related IP rights, of that new knowledge?

These issues become even more intricate when third parties (especially when commercial) are involved. By way of example, health insurers or pharmaceutical companies may wish to access data within the VHT infrastructure or within a specific DT, or even the DT itself, to tailor services, assess risk, or personalize treatment plans. On their turn, patients may have concerns about losing control over their data to these entities, fearing potential exploitation or misuse, such as discriminatory practices in insurance pricing or eligibility. In such scenarios, ownership becomes entangled with privacy rights and concerns about how patient data may be commercially leveraged. Establishing legal protection that limits how VHT data can be used, especially in non-clinical contexts, is essential to prevent discrimination and protect patient rights.

Ultimately, the concept of ownership intersects broader ethical and legal frameworks around data governance and digital identity. Many people advocate for a model where ownership remains firmly with the patients, allowing them to retain full control over their DT and the ability to consent to any use or sharing of their data. Others propose co-ownership or stewardship models, where healthcare providers and patients share responsibilities and rights over the DT, balancing patient autonomy with clinical or research benefits. As of today, the applicable legislation does not seem to admit the implementation of this latter option.

As the use of Digital Twins in healthcare evolves, establishing clear, ethical, and legally sound standards for data ownership will be crucial to protecting patient rights while allowing the technology to fulfil its potential in personalized and predictive medicine.

E) Fairness and equality

The deployment of Digital Twins in healthcare raises ethical questions around fairness and equality, especially concerning access to this advanced technology. As a sophisticated system of systems, the VHT requires – *inter alia* – substantial technological infrastructure, data storage capabilities and computational resources, which may be more accessible to wealthier healthcare facilities and patients in higher socio-economic groups. Consequently, the advantages of the VHT could be unequally distributed, leading to a healthcare system where access to personalized, data-driven treatment is a privilege rather than a right.

This disparity in access risks deepening existing inequalities in healthcare. Patients in underserved communities, who may already face barriers to quality care, could miss out on the benefits of VHT technology due to financial, cultural, geographical, or infrastructural limitations¹⁶⁴. Ethical integration of the VHT in the healthcare sector demands policies that aim for equitable distribution, such as public subsidies or grants that help lower-income patients and communities gain access to Digital Twin-supported treatments. Additionally, regulators and policymakers should work with healthcare providers to explore scalable VHT models that can operate effectively in resource-limited settings.

There are other ways in which the VHT can lead to inequality and other forms of injustice. For instance, misuse of relevant data poses a potential risk of discrimination¹⁶⁵. Digital Twins that incorporate predictive algorithms based on a range of data sources might inadvertently flag individuals as high-risk

¹⁶⁴ Braun M. Represent me: please! Towards an ethics of digital twins in medicine. J Med Ethics. 2021 Mar 15;medethics-2020-106134. <https://doi.org/10.1136/medethics-2020-106134>.

¹⁶⁵ Popa *et al.* <http://doi.org/10.1186/s40504-021-00113-x>.

due to factors outside their control, such as genetic predispositions or socio-economic background. Insurers, employers, or even healthcare providers could misuse this information, creating stigmas or unjust barriers for individuals based on their health predictions¹⁶⁶. Establishing legal protection that limits how VHT data can be used, especially in non-clinical contexts, is essential to prevent discrimination and protect patient rights.

Moreover, algorithmic inequalities may arise if DTs are trained on biased datasets: if a digital replica is built on data from predominantly affluent, homogeneous populations, its predictions and recommendations may not accurately reflect the health realities of diverse patient groups. A fair approach to the VHT requires that the algorithms behind these systems are free from biases that could lead to discrimination. To this end, it is essential to rigorously test algorithms to identify potential biases that may emerge in the course of development. Utilizing fairness metrics is one effective way to measure how algorithms perform across various demographic groups, helping to ensure that the VHT models offer equitable outcomes for all users¹⁶⁷. When disparities are identified, models should be updated and refined to provide a more balanced result. Scientific protocols, developed after weighing-up the ethical framework, should mandate diversity in dataset sources, emphasizing that the design and implementation of the VHT accounts for the health variances across different demographics – as also required by the AI Act – thereby minimizing biases that could reinforce inequalities.

Fairness is also deeply connected to transparent decision-making processes that empower both patients and healthcare providers to understand and trust how the VHT and its DTs can influence care. When these models are used to inform clinical decisions, it is ethically essential to clearly communicate their role, potential limitations, and the ways in which they might impact treatment options. This transparency ensures that patients have a full picture of their care process and can make informed decisions. By fostering such openness, healthcare providers and VHT developers not only promote autonomy – allowing patients to actively participate in their healthcare choices – but also build trust in the technology itself, making individuals more comfortable with the role DTs play in their treatment.

F) Technological over-dependence

This issue is intrinsically linked to the entire field of artificial intelligence, presenting varying yet consistently high levels of risk. To mitigate its potential consequences, the GDPR, with notable foresight, addressed this concern through Article 22, prohibiting fully automated individual decision-making when such decisions could have significant impacts on individuals. This provision finds its most evident and critical application in the healthcare sector, where the stakes are particularly high.

Building on this principle, the AI Act has then imposed the obligation for human oversight, requiring that AI systems, particularly high-risk ones, operate under human-in-the-loop mechanisms to ensure accountability and ethical compliance. As detailed in the dedicated section above, this means human operators must have the ability to intervene, override decisions, or evaluate outcomes where necessary, particularly in sensitive domains like healthcare. Such oversight safeguards against potential harms, biases, or errors, ensuring AI systems align with human values and legal standards.

Indeed, while DTs offer significant benefits in several areas, including predictive healthcare, their integration may foster an over-reliance on technology, which poses ethical risks for patient care. The powerful analytics and prognostic capabilities of DTs may overshadow the nuanced understanding that comes from direct human interaction in healthcare. As digital models increasingly influence diagnostic and treatment decisions, there is the concrete ethical risk that clinicians and patients alike may overly trust AI-driven outputs, potentially disregarding the importance of empathy, communication and individualized care¹⁶⁸.

This technological dependency could shift healthcare dynamics toward a system that prioritizes efficiency and predictive accuracy over relational care. For instance, a doctor might feel pressured to follow VHT-driven recommendations, fearing that diverging from data-based insights could be seen as

¹⁶⁶ Iqbal *et al.*

¹⁶⁷ Teller M.

¹⁶⁸ Iqbal *et al.*

neglectful or unscientific. To prevent such a scenario, ethical guidelines should advocate for a balanced approach that combines VHT-based insights with traditional, compassionate patient care.

Furthermore, AI-based recommendations are not infallible, as there is always the potential for misdiagnosis or suboptimal treatment due to reliance on algorithms that may lack context-specific understanding. Complex or atypical cases might be oversimplified within a VHT framework, leading to inaccurate conclusions. To mitigate these risks, healthcare institutions should adopt protocols that emphasize the importance of human oversight in such technology-influenced decisions, ensuring that healthcare professionals remain engaged and critical in assessing all outputs deriving from the VHT.

12 Technical standards

The standards, guidelines, as well as domain-specific terminologies and ontologies relevant for the VHT are described in detail in the EDITH document “*Standardization landscape, needs and gaps for the virtual human twin (VHT)*”¹⁶⁹. Based on this landscaping document, the corresponding “*EDITH standards Implementation guide*” provides a practical guideline for using and implementing standards, terminologies, and metadata guidelines when setting up, executing, and archiving virtual human twins and their parts¹⁷⁰. This is complemented by a comprehensive and interactively searchable listing of relevant standards in the **EDITH FAIRsharing collection**¹⁷¹.

These building blocks describe in detail the existing landscape of standards, *e.g.*, for data formats, data integration and data input into models, the modelling itself, for metadata, model quality and model validation, as well as data provenance and interoperability with both research and clinical data. In addition, there are standards and recommendations for the bundling of data from an individual or patient under consideration of data privacy and data protection laws. Most of these standards were defined by research communities and are well established in the sector.

Notwithstanding this, there are still some gaps in the standardization scenario, and standards for some specific modelling approaches are still under discussion or development, *e.g.*, the **Multi-Cellular Modelling Language (MultiCellML)**¹⁷² or the **Open Virtual Tissues (OpenVT)** standard for multicellular agent-based virtual tissue models¹⁷³, which are defined under the roof of the **COMBINE (computational modelling in biology network)** consortium¹⁷⁴.

In addition to the community-driven standards, some standards exist which are defined by technical committees of official standard defining organizations (‘SDOs’) like **ISO (International Organization for Standardization)**, *e.g.*, in **ISO/TC 276 on Biotechnology** and **ISO/TC 215 on Health Informatics**¹⁷⁵, or **IEC (International Electrotechnical Commission)**, *e.g.*, in **IEC/TC 62 on Medical equipment, software, and systems**¹⁷⁶, sometimes even in joint committees between them.

In contrast to standardization efforts driven by specific communities, these standards defined by SDOs have the advantage of reflecting the consensus of larger and international communities. Moreover, they are maintained and updated, if needed, by the established SDO committees also on a long-term basis and, therefore, they are seen as stable and reliable standards, particularly suitable also for validation and regulatory purposes.

As opposed to often openly and royalty-free available community standards, standards published by SDOs are typically bound to a license and corresponding royalties. This fees, which on the average are not high, should not be considered as commercial fee, but rather a contribution to cover the costs of the SDO that are usually non-profit.

12.1 ASPECTS OF INTEREST AND RELEVANT IMPACTS/CURRENT USE:

The most relevant official standards for the VHT are the best practices in data acquisition, integration and analysis, which are defined in **ISO 20691:2022 (“Biotechnology - Requirements for data formatting and description in the life sciences”)**¹⁷⁷. On the one hand, it is a reference framework for the development of interoperable data and metadata standards for many life science domains (including the VHT domain), especially regarding standards for data and model formatting, preparation, structuring and integration. On the other hand, it also provides recommendations and guidelines for the concerted

¹⁶⁹ <https://zenodo.org/doi/10.5281/zenodo.10492795>

¹⁷⁰ <https://zenodo.org/doi/10.5281/zenodo.10524794>

¹⁷¹ <https://fairsharing.org/4787>

¹⁷² <https://multicellml.org/>

¹⁷³ <https://www.openvt.org/>

¹⁷⁴ <https://co.mbine.org/>

¹⁷⁵ Respectively <https://www.iso.org/committee/4514241.html> and <https://www.iso.org/committee/54960.html>

¹⁷⁶ https://www.iec.ch/dyn/www/f?p=103:7:::::FSP_ORG_ID:1245

¹⁷⁷ <https://www.iso.org/standard/68848.html>

use of such data formats as well as of descriptive metadata standards, terminologies and ontologies in the life sciences.

The ISO/TS 9491 standard series “*Biotechnology - Recommendations and requirements for predictive computational models in personalized medicine research*” was prepared by the EU project EU-STANDS4PM (European standardization framework for data integration and data-driven in silico models for personalized medicine)¹⁷⁸ and consists of two parts:

- i) ISO/TS 9491-1:2023 (“*Guidelines for constructing, verifying, and validating models*”), released in 2023, defines specific recommendations and requirements for data preparation to integrate health data into computational models, as well as for model formatting, validation, simulation, storing, sharing and their application in clinical trials and research. In addition, ethical requirements for modelling are addressed¹⁷⁹.
- ii) ISO/CD TS 9491-2 (“*Guidelines for implementing computational models in clinical integrated decision support systems*”)¹⁸⁰ lays down recommendations for implementing computational models into clinical decision support systems.

The ISO 23494 series (“*Biotechnology – Provenance information model for biological material and data*”)¹⁸¹ describes how to document the provenance information for biological data and samples that is important to trace elements of processed data and specimens back to the original source of the data or biological material.

Further important standards and regulations for DTs and SaMD are:

- the Medical Device Regulation¹⁸².
- the In Vitro Medical Device Regulation¹⁸³.
- ISO 13485:2016 about “*Medical devices — Quality management systems — Requirements for regulatory purposes*”.
- ISO/IEC 27001 about “*Information security, cybersecurity and privacy protection*”.
- ISO 14971:2019 about “*Medical devices - Application of risk management to medical devices*”

12.2 CURRENT OPEN AND CHALLENGING POINTS

12.2.1 MODEL CREDIBILITY ASSESSMENT STANDARD

Besides these standards which are already in use, there is an **urgent need for an international model validation and credibility assessment standard**. As already pointed out in Sections 8 and 9, such standard should comprise the scope of the ASME VV-40 standard (‘Assessing Credibility of Computational Modelling through Verification and Validation: Application to Medical Devices’)¹⁸⁴, defined by the American Society of Mechanical Engineers for assessing the quality and model credibility of medical devices. The model quality assessment according to ASME VV-40 starts with a clear definition of the scientific / medical question of interest (QoI) and the context of use (CoU). The CoU is a complete description of the planned modelling use and defines the role and scope of the model used to address the question of interest. In the next step, the model risk (with its two components model influence and decision consequence) is assessed. The applicability of a model is given by the evidence to support the use of the model in the defined CoU, considering the risk. Then a risk-informed credibility assessment, which encompasses model verification, validation, and uncertainty quantification (VVUQ) can be performed. The use of such a credibility assessment standard is important for the evaluation of

¹⁷⁸ <https://www.eu-stands4pm.eu/>

¹⁷⁹ <https://www.iso.org/standard/83516.html>

¹⁸⁰ <https://www.iso.org/standard/87403.html>

¹⁸¹ <https://www.iso.org/standard/87403.html> and <https://www.iso.org/standard/80715.html>

¹⁸² Please refer to Section 9 of this D6.2.

¹⁸³ Please refer to Section 9 of this D6.2.

¹⁸⁴ <https://www.asme.org/codes-standards/find-codes-standards/assessing-credibility-of-computational-modeling-through-verification-and-validation-application-to-medical-devices>

new drugs and for getting relevant regulatory approval by competent authorities like FDA, which is already using ASME VV-40 for regulatory purposes in the USA, and EMA in the EU.

Since the credibility assessment should not be possible only for knowledge-based mechanistic models, but also for the assessment of data-driven AI models, rule-based and agent-based models (ABM), a **joint working group of ISO/TC 276 and IEC/TC 62** is currently established (already existing as an *ad-hoc* group in IEC/TC 62) to define an international standard for model validation and credibility assessment encompassing also AI- and ABM models, extending and internationalizing the ASME VV-40 standard.

Elements from the Total Product Lifecycle (TPLC) model, Good Machine Learning Practice (GMLP), Predetermined Change Control Plans (PCCP's), comprising the SaMD Pre-Specifications (SPS) and the Algorithm Change Protocol (ACP), as well as Medical Device Development Tools (MDDT), should be included as well into the new credibility assessment standard [US-FDA 2019].

12.2.2 BEST PRACTICES AND CLINICAL PRACTICE GUIDELINES

Beside the standards described above, there are many reporting and best practices / clinical practice guidelines, consensus statements and minimal clinical core datasets. Most of them are specific to a medical field and/or a disease. These best practice guidelines are a foundational pillar for evidence-based medicine and guide clinicians in their daily work decisions. As of today, many of these guidelines are relevant for the use and application of AI models in medical diagnosis, treatment, or prognosis¹⁸⁵.

Other best practices like **PRISMA** ('Preferred Reporting Items for Systematic reviews and Meta-Analyses') are reporting guidelines for systematic reviews of a field¹⁸⁶. Whereas most best practices guidelines are very domain specific, there are also some general statements, rules and frameworks defining what should be reported in such best practice guidelines, such as the **PICAR** ('Population, Intervention, Comparison, Attributes of eligible CPGs, Recommendation characteristics')¹⁸⁷ and the **PICOR** ('Population, Intervention, Comparator. Outcome and Recommendation')¹⁸⁸ statements.

12.3 POLICY RECOMMENDATIONS

Despite the large range of standards and guidelines – some of which have been explored above – there are still areas that are not covered by the existing standards. Examples are harmonized 'Data Access Agreements', a general standard for reporting simulation output data, a standardized way for the biomedical integration of multiscale models and a standard for assessing the quality of data and for the access to data for model validation. Another challenge is the life-cycle management of data belonging to a patient, which in principle can be implemented based on the patient-centred clinical decision support framework¹⁸⁹.

¹⁸⁵ Wang Y, Li N, Chen L, Wu M, Meng S, Dai Z, Zhang Y, Clarke M. Guidelines, Consensus Statements, and Standards for the Use of Artificial Intelligence in Medicine: Systematic Review. *J Med Internet Res*. 2023 Nov 22;25:e46089. <https://www.jmir.org/2023/1/e46089>.

¹⁸⁶ Page et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ*. 2021 Mar 29;372:n71. <https://doi.org/10.1136/bmj.n71>.

¹⁸⁷ Johnston et al., Systematic reviews of clinical practice guidelines: a methodological guide, *J Clin Epidemiol*. 2019 Apr;108:64-76. <https://doi.org/10.1016/j.jclinepi.2018.11.030>.

¹⁸⁸ Mancin et al., Systematic review of clinical practice guidelines and systematic reviews: A method for conducting comprehensive analysis, *MethodsX*. 2023 Dec 21;12:102532. <https://doi.org/10.1016/j.mex.2023.102532>.

¹⁸⁹ Sittig, D.F. et al. (2023). A lifecycle framework illustrates eight stages necessary for realizing the benefits of patient-centered clinical decision support. *J Am Med Inform Assoc*. 2023 Aug 18;30(9):1583-1589. <https://doi.org/10.1093/jamia/ocad122>.

13 Conclusions

In the light of such an extensive analysis of the EU legal framework applicable to the Virtual Human Twins and relevant stakeholders ecosystem, it is now appropriate to outline conclusions that chart the path toward establishing a Code of Conduct aimed, alongside several much needed policy-making initiatives around data reuse and in-silico medicine, health systems and data interoperability, AI-driven SaMDs, IPR management, technical standards and ethics-by-design approach in the field, to facilitate a more informed, cohesive and widespread adoption of this transformative and pioneering technology.

A Code of Conduct under Article 40 of the GDPR, applicable across the EU, offers numerous benefits by fostering consistency, trust and compliance for data-related activities. This crucial accountability tool is capable to set out a framework of rules tailored to the specific needs and challenges of a sector, promoting clarity and harmonization in how GDPR and relevant data governance principles are implemented across Member States. **For industries involving innovative technologies, like Virtual Human Twins, an EU-wide Code can bridge the gap between general legal provisions and sector-specific realities, ensuring that regulatory requirements are both practical and effective.**

By detailing best practices and defining acceptable standards for data use and reuse, a Code of Conduct enhances transparency and accountability, benefiting both stakeholders and data subjects. **Organizations adhering to such a Code might substantiate compliance steps more effectively, reducing uncertainty and fostering trust among regulators, patients and society itself.** Furthermore, it provides a structured approach to addressing complex issues such as data minimization, purpose limitation and the management of sensitive health data, streamlining compliance efforts and reducing the risk of violations.

Furthermore, from an operational perspective, a unified Code of Conduct would simplify cross-border activities, eliminating inconsistencies in the interpretation of GDPR provisions across EU jurisdictions. This would not only **reduce administrative burdens and associated compliance costs for organizations operating in multiple Member States** but would also **ensure that data subjects receive consistent levels of protection**, regardless of both where DTs are developed, put on the market or used and so of where the data are collected and re-purposed for building the needed models. Ultimately, **an EU-wide Code of Conduct under Article 40 would promote a balanced approach to innovation and regulation, enabling technological advancements while safeguarding fundamental rights.**

Under a procedural standpoint, associations and other bodies representing specific categories of stakeholders – acting either as data controllers or processors – may draft and propose codes of conduct, laying down elective rules regarding at least:

- a) fair and transparent processing of personal data in a vertical or horizontal sector of the market;
- b) the legitimate interests pursued by controllers in specific contexts;
- c) the (primary and secondary) processing of personal data;
- d) pseudonymization, anonymization and application of PETs and data synthesis to personal data;
- e) transparency obligations *vis-a-vis* the data subjects;
- f) the exercise of the individual rights enshrined in the GDPR (and in many other applicable regulations, such as the EHDS);
- g) the need to reinforce the protection of children and the manner in which the consent of the holders of parental responsibility over children is to be obtained;
- h) the measures and procedures to implement accountability, privacy-by-design and by-default, and appropriate levels of data security;
- i) the procedure for managing the notification of personal data breaches to competent supervisory authorities and the associated communication, when applicable, to data subjects;

- j) the transfer of personal data to third countries or international organizations;
- k) out-of-court and other dispute resolution procedures for resolving litigations between controllers and data subjects.

Association and bodies representing specific stakeholders are entitled to put forward proposals for sectoral codes of conduct, submitting the relevant drafts to the supervisory authority of the EU Member State where they are established. The competent DPA must then issue its opinion on whether the proposed code complies with the applicable legislation and approve it, upon confirmation that the code provides sufficient appropriate safeguards.

Should the draft code relate to processing activities in several Member States, the competent national supervisory authority, before giving its final approval, is required to submit it to the European Data Protection Board, which must provide an opinion on whether the code complies with the GDPR both in terms of data protection and governance and security measures. In the event of a positive opinion, the Board shall submit it to the EU Commission which may, by way of implementing acts, decide that the approved code of conduct must have general validity within the Union.

In brief, if a proposed code aims for EU-wide applicability, the supervisory authority must forward it to the EDPB for a qualified and binding opinion. The Board is required to evaluate whether the code meets GDPR requirements and ensures harmonization across Member States. Following a positive opinion by the EDPB, the EU Commission may grant the code binding status across the EU, solidifying its role in unifying data governance and protection standards for the relevant sector.

Many of the legal and regulatory barriers highlighted in this document can be overcome, or at least mitigated, through a dedicated Code of Conduct for the VHT which establishes clear and, above all, consistent and common rules for the use and secondary processing of data, particularly health data, for the development and adoption of Digital Twin-based technologies in the research and healthcare fields.

Such a Code would offer a sector-specific framework that provides clarity and uniformity across Member States, addressing the fragmentation that currently (over)complicates compliance and regulatory procedures for organizations operating in multiple jurisdictions. These divergent interpretations and lack of cohesion often lead to inconsistent application of the principles, obligations and requirements applicable to data and associated data-driven technologies, particularly concerning the use and secondary processing of sensitive information in the health and medical contexts.

The Code would define specific standards for critical issues such as lawful processing bases, data minimization, purpose limitation, data quality and validation, the requirements for anonymization or pseudonymization, or the use of synthetic data, ensuring that health data can be responsibly used for developing and safely integrating the VHT in the upcoming scenarios stemming from the AI Act, the EHDS and the DGA, as well as in connection with the regulatory pathways established for clinical trials and medical devices, also in view of properly safeguarding and enhancing the connected IPR and promoting an ethical approach to increase trust among regulators, researchers, clinicians, and especially patients¹⁹⁰.

Despite the considerable time and effort required to establish a comprehensive Code of Conduct for the VHT, it is essential that these long-term endeavours proceed in tandem with more immediate actions by EU regulators and policymakers. The fragmented and overly complex regulatory landscape is already exerting significant negative effects on the sustainability and investment prospects of Europe’s technology, AI, and data-driven markets. If left unaddressed, these challenges risk continuing to undermine the Union’s competitiveness on a global scale, as clearly pointed out in Mario Draghi’s report explored in Section 2.1, also considering that “EU

¹⁹⁰ Two significant references exist to date: (i) the “Code of Conduct for Service Providers in Clinical Research” promoted in France by the European CRO Federation (EUCROF), approved by the EDPB on 18 June 2024, by means of the opinion 12/2024, and then by the French supervisory Authority, the CNIL (*Commission Nationale de l’Informatique et des Libertés*) on its plenary meeting of 12 September 2024, by resolution no. 2024-064 (the full text of the EUCROF Code of conduct is available via https://www.cnil.fr/sites/cnil/files/2024-10/2024codeeucrof-partie1_eng.pdf); (ii) the ‘Code of conduct regulating the processing of personal data in clinical trials and other clinical research and pharmacovigilance activities’ promoted in Spain by Farmaindustria and approved by the Spanish supervisory Authority, the AEPD – *Agencia Española de Protección de Datos* (the full text of the Farmaindustria Code of conduct is available via <https://www.aepd.es/documento/farmaindustria-code-conduct-regulating-processing-personal-clinical-en.pdf>).

regulation imposes a proportionally higher burden on SMEs and small mid-caps than on larger companies”¹⁹¹.

Immediate steps are needed to address inconsistencies and streamline regulatory processes that currently hinder innovation and deter investment. This includes reducing the overlapping layers of regulation that create legal uncertainty for businesses, making it harder for them to navigate compliance requirements efficiently. **While the Code of Conduct offers a promising avenue for long-term harmonization and clarity, short-term measures are critical to mitigating the current barriers to market entry and growth. These efforts must focus on fostering a more cohesive regulatory environment that encourages innovation, protects data subjects, and supports Europe’s position as a leader in the global tech and AI landscape, including for the VHT.**

In this sense, the forthcoming implementation of the European Health Data Space presents a unique opportunity to create a dedicated ‘**Strategic Committee**’ within the EU Commission, **envisioned as a coordination task force that would be tasked with promoting, overseeing, and ensuring the harmonization of guidelines, policies, procedures, and innovative regulatory pathways by all involved policymakers and competent authorities.** This Committee, which on account of its nature might ideally benefit from the joint involvement of the Commissioners for (i) Tech Sovereignty, (ii) Health and animal welfare, (iii) Economy and productivity, implementation and simplification, (iv) Startups, research and innovation (v) Cohesion and reforms, should comprise members from all competent:

- *Directorates-General*, including particularly those for:
 - Communications Networks, Content and Technology
 - Research and Innovation (together with the Joint Research Centre);
 - Health and Food Safety;
 - Internal Market, Industry, Entrepreneurship and SMEs;
 - Competition;
- and *Executive Agencies*, such as the:
 - Health and Digital Executive Agency;
 - European Research Executive Agency
 - European Research Council Executive Agency.

More than any other legislation, **the EHDS has profound interconnections and dependencies with all other current or upcoming regulations:** (i) the **GDPR**, including *inter alia* about primary and especially secondary use of electronic health data, the obligation for data anonymization or pseudonymization before making the dataset available to data users, the management of reinforced data subjects’ rights, (ii) the **AI Act**, for AI-driven technologies interacting with, or integrated into EHR systems or medical devices, (iii) the **MDR** and **IVDR**, *e.g.*, because certain components of EHR systems could qualify as medical devices or in-vitro diagnostic devices, or in view of the interoperability of medical devices with the harmonized components of EHR systems, or the need for data holders to make all electronic health data generated through, or in any case coming from, medical devices or relevant registries, available for reuse, (iv) the **CTR**, since data deriving from clinical trials, studies and investigations published in the CTIS and EUDAMED databases will be open for secondary use. Additionally, the EHDS implies the establishment of a forward-looking **Intellectual Property management framework** to effectively safeguard the rights and trade secrets linked to electronic health data subjected to secondary processing by a potentially vast and diverse range of stakeholders, as well as alternatively enhancing, when possible, also faster-growing open-source approaches.

¹⁹¹ ‘The future of European competitiveness’ report. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en.

Given the pressing need to re-evaluate and streamline established legal concepts and regulatory practices, the forthcoming implementation of the EHDS represents a pivotal opportunity for EU policymakers to adopt a more progressive and innovation-driven approach. **This moment could mark a significant shift in pace, enabling a regulatory environment that not only accommodates but actively promotes advancements in research, healthcare and technology sectors. By prioritizing measures that simplify legal pathways and ensure ethical integrity, policymakers can create conditions that encourage investment and foster transformative developments across these fields.**

To this end, **the support for the broader and more seamless adoption of the VHT should be the primary area of focus** by the Committee, as well as a litmus test for the more ambitious achievement of both the European Research Area and the European Health Union.

VHT models in general rely heavily on the secondary use of health data, which the EHDS is uniquely positioned to enable. By facilitating access to high-quality, interoperable, and ethically managed health data, the EHDS can provide the foundation for these technologies to thrive, delivering unparalleled benefits in predictive modelling, treatment optimization, and patient care.

Through targeted policy adjustments and a focus on creating a cohesive and accessible data ecosystem, by adopting far-sighted guidance, setting EU-wide uniform best practices and technical standards, the EHDS ‘strategic coordination unit’ can serve as a catalyst for growth, competitiveness, and sustainability in Europe’s research and healthcare landscape.

The Committee should engage, as necessary, with the relevant EU and national authorities, offices and bodies, along with stakeholders’ groups, where provided by applicable regulations (*e.g.*, the Medical Device Coordination Group, the EHDS Stakeholder Forum and Advisory Forum under the AI Act). Such involvement should be tailored to address the specific needs and issues at hand, as identified in the “Barriers, dependencies, and proposed actions” outlined in Sections 4 to 9 of this document.

After all, the successful implementation and full realization of the Health Data Space depend fundamentally on addressing and overcoming the existing barriers that hinder or, in some cases, completely obstruct the advancement of medical research and innovation. These hurdles, whether regulatory, technical, or ethical, act as significant roadblocks to creating a seamless and efficient VHT ecosystem. **Without resolving these challenges, unlocking the EHDS’s full transformative potential will prove unfeasible, particularly enabling groundbreaking initiatives such as the VHT¹⁹².**

Furthermore, **a pioneering approach must be outlined for IPR management in the context of the VHT ecosystem, focusing on three main aspects: copyright and database protection, trade secrets, and patentability.** Focusing on the EHDS, a tension arises between data-sharing obligations, which can benefit and fuel the VHT ecosystem, and IPR protection, which is necessary to incentivize private investment. **Open-source approaches could increase accessibility and foster collaboration but require both compliance incentives and careful regulation to balance transparency with protection of sensitive data.** On the other hand, the challenges of ‘sufficiency of disclosure’ for AI innovations need to be addressed (since “black box” features of AI complicate patentability under European Patent Convention requirements) by creating secure repositories for relevant models, training data, and methodologies. **Licensing frameworks inspired by Standard Essential Patents, operating under FRAND (Fair, Reasonable, and Non-Discriminatory) terms,** might be put forward to balance innovation and accessibility, ensuring broad use of foundational technologies while providing fair compensation. Finally, a harmonized **IPR management policy** specifically designed for the VHT ecosystem will have to be defined and enforced, to prioritize transparent and equitable licensing frameworks, provide clear guidelines for IPR use in AI contexts and encourage collaborative efforts to align public health objectives with innovation incentives.

The gradual implementation of the measures outlined in this document, culminating around the 2028/2030 period – when the obligations under the AI Act will fully apply and the initial exchanges of data for both primary and secondary uses under the EHDS will commence – represents a

¹⁹² As highlighted, with reference to many of the barriers identified, also by TEHDAS. See Section 2.2.

critical opportunity for the EU to position itself as a global leader in the Virtual Health Technology industry, also in connection with the medical devices and clinical trial fields.

Achieving this requires the coordinated execution of the proposed actions, alongside the parallel development of a comprehensive Code of Conduct focused on addressing key issues related to data governance and the secure, efficient circulation of health data.

Such a Code of Conduct would establish EU-wide, consistent, ethically grounded standards for the management and sharing of data, ensuring alignment with GDPR and the AI Act while fostering trust among stakeholders and patients. **By integrating these initiatives with the EHDS infrastructure, under the coordination and supervision of the suggested *ad hoc* strategic Committee, the EU would nurture a robust and financially attractive ecosystem that would support the responsible adoption of cutting-edge technology initiatives like the VHT.** This dual-track approach would not only enhance regulatory clarity and efficiency but also attract significant public and private investments, drive research innovation, and reinforce the EU’s competitiveness in the rapidly evolving global AI-driven and VHT market.

The European Union needs, in fact, to adequately benefit from the **profound transformative impact** that is implied by a growing uptake of the Virtual Human Twin concept and products. By allowing personalised simulations of alternative multiannual individual health pathways, **the VHT can be particularly effective in shifting healthcare towards primary, secondary and tertiary prevention.** They very persuasively stimulate the adoption of interventions and behaviours that can avoid the onset and development of diseases; they allow for early detection of a medical condition through timely screening techniques; they monitor and optimally treat, so long as meaningful, a disease progression. Policy-wise, the VHT can thus constitute, both at a Union level as well as in all Member States, a very apt “**growth-enhancing expenditure**” within the **EU new economic governance framework**, capable of **significantly supporting the future economic and social sustainability of Europe’s ailing welfare and healthcare systems.**