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 3.1
Vision for the VHT and roadmap outline

Due date of delivery: 31 March 2023
Actual submission date: 1 April 2023

Start of the project: 01 October 2022
End date: 30 September 2024

Reference

Name	EDITH_D3.1_Vision_roadmap ouline_VPH_Mar2023
Lead beneficiary	Virtual Physiological Human institute (VPHi)
Author(s)	Liesbet Geris (VPHi)
Dissemination level	Public
Type	Report
Official delivery date	31 March 2023
Date of validation by the WP Leader	1 April 2023
Date of validation by the Coordinator	**1 April 2023**
Signature of the Coordinator	**1 April 2023**

Version log

Date	Version	Involved	Comments
23/02/2023	v1.0	Liesbet Geris	First draft
	v2.0		Internal review by VPHi (PC)
	v3.0		Review by UNIBO
	v4.0		Final review, formal checking by PC
	Final		Final submitted version

Executive summary

The present document is the first written presentation of the Virtual Human Twin (VHT) vision as it has been prepared by the EDITH consortium and discussed with select representatives of the wider ecosystem. After a brief statement on the genesis of the vision, the document is composed of two main parts: the outline of the VHT roadmap and the elaboration of the vision for the integrated Virtual Human Twin.

The **proposed outline of the VHT roadmap** starts with the state of the art and maturity of the different aspects of the VHT, including in silico medicine, Artificial Intelligence & Machine Learning, wearables and data-driven twins, as well as an overview of the relevant initiatives and actors (Scientific, clinical & industrial organisations; Infrastructures & platforms; Standardisation, Regulatory & HTA actors; public policy at EU and member state level). Next, the roadmap proposes the vision for the VHT, as discussed in detail in this deliverable. Subsequently, the required technological elements to realise this vision will need to be elaborated, including the data, the models, the integration of resources and the required infrastructure. This will be followed by a discussion on the required Standards, regulatory science, Health Technology Assessment, as well as the Ethical, Social and Legal aspects (including data privacy & protection as well as intellectual property management). Finally, the uptake of the VHT will be elaborated, including the user perspective, the organisation of the ecosystem and the sustainability of the VHT via private initiatives and a public infrastructure.

The section discussing the **vision for the VHT** starts by introducing a number of key concepts. The definition of digital twins in healthcare is provided, encompassing generic, population and subject-specific twins. Additionally, the life cycle of the digital twin in healthcare is explained briefly, from its inception to its credibility assessment and use in the context of human health and care. This is followed by an analysis of the main barriers for the development and adoption of digital twins in healthcare. With this background, a vision for the integrated Virtual Human Twin is proposed: *The Virtual human twin is an integrated multi-scale, -time and -discipline digital representation of the whole body enabling the comprehensive characterisation of the physiological and the pathological state in its heterogeneity and allowing patient-specific predictions for the prevention, prediction, screening, diagnosis and treatment of a disease, as well as the evaluation, optimisation, selection and personalisation of intervention options.* Subsequently, the realisation of the VHT is broken down into several key aspects: the community of practice, the infrastructure, the standards and the long-term sustainability. Each of these aspects is introduced and the related key concepts explained.

With the vision of the VHT elaborated, the next section goes further into the required technological developments to realise this mission. Although the full elaboration of this section – as well as the following ones defined in the roadmap outline - is out of scope of this deliverable, the **proposed organisation of the resources in the VHT** is discussed in more detail. Within the VHT, the atomic entities are data objects and model objects. The VHT framework can be imagined as an n-dimensional data space, which digital twin models constantly crawl. Taken together, the data and model objects define the location, content and conditions of use of the different resources in the data space. In the current proposal, six dimensions are proposed for the organisation of the data space (but other could be added as the VHT develops): the three dimensions of the Body (body height, width, and depth), as well as Age (time), Credibility and Clustering.

The final section of this deliverable briefly summarizes the EDITH actions of the coming months. After publication of the deliverable, the vision will be communicated broadly and discussed with a wide range of stakeholders in order to consolidate the vision into the first draft of the roadmap that is due 31st of July 2023.

Table of contents

List of figures

Acronyms

Acronym	Full name
AI	Artificial Intelligence
ASME	American Society of Mechanical Engineers
ATMP	Advanced Therapy Medicinal Product
B2B	Business-to-Business
BBCT	Bologna Biomechanical CT
BMD	Bone Mineral Density
CBER	Center for Biologics Evaluation and Research
CDRH	Center for Devices and Radiological Health
CE	Conformité Européenne
CoP	Community of Practice
CoU	Context of Use
CRO	Contract Research Organisation
CSA	Coordination and Support Action
DLT	Distributed Ledger Technology
DOP	Data Object Pose
DOT	Data Object Type
DT	Digital Twin
DTH	Digital Twin in Healthcare
DXA	Dual-energy X-ray absorptiometry
EC	European Commission
EMA	European Medicine Agency
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reproducible
FDA	Food and Drug Administration
FFR	Fractional Flow Reserve
GA	Grant Agreement
GDPR	General Data Protection Regulation
GPU	Graphic Processing Unit
HPC	High performance computing
HTA	Health Technology Assessment
ICT	Information and Communications Technology
ID	Identity
IHI	Innovative Health Initiative
IPR	Intellectual Property Rights
IST	In Silico Trials
ISW	In Silico World
ML	Machine Learning
MRI	Magnetic Resonance Imaging
nD	n-dimensional
NGO	Non-governmental Organisation
OEM	Original Equipment Manufacturer
PC	Project Coordinator
PM	Project Manager
QoI	Quantity of Interest
SaaS	Software as a Service
SaMD	Software as a Medical Device
SME	Small and Medium-sized Enterprises
SOP	Standard Operating Procedure
VHT	Virtual Human Twin
VV-40	Verification and Validation 40
WP	Work Package
yo	Years old

Acronyms

Acronym	Consortium partner
VHPi	Virtual Physiological Human Institute
LYN	Lynkeus SRL
ACC CYFRONET	Akademickie Centrum Komputerowe Cyfronet AGH
ATHENA	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis
BSC	Barcelona Supercomputing Center
DIGITALEUROPE	DIGITALEUROPE AISBL
EMPIRICA	Gesellschaft für Kommunikations- und Technologieforschung mbH
EPFL	École Polytechnique Fédérale de Lausanne
FORTH	Foundation for Research and Technology Hellas)
HITS	Heidelberg Institute for Theoretical Studie
INRIA	Institut National de Recherche en Informatique et Automatique
JÜLICH	Forschungszentrum Jülich GMBH
PI SCHOOL	PI SCHOOL
QMUL	Queen Mary University of London
RWTH	Rhine-Westphalia Technical University of Aachen
ULIÈGE	University of Liège
UNIBO	Alma Mater Studiorum – Università di Bologna
UvA	Universiteit van Amsterdam
VITO	Vlaamse Instelling Voor Technologisch Onderzoek

1 Introduction

The Virtual Human Twin (VHT) is an integrated multi-scale, multi-time, and multi-discipline representation of quantitative human physiology and pathology. Its realisation through a collaborative distributed knowledge and resource platform is specifically designed to accelerate the development, integration, and adoption of patient-specific predictive computer models, which will be used as clinical decision support systems for personal health forecasting or as methodologies for the development and de-risking of personalised medical products. The **vision of EDITH** is to facilitate the realisation of the opportunities presented by VHTs for the benefit of patients, healthcare providers, regulatory bodies and industry, both within Europe and globally.

EDITH is a Coordination and Support Action (CSA) funded by the European Commission, which will capitalise on the developments of digital technologies, high-performance computing, availability and access to research and healthcare data in Europe, with the **mission** of defining a roadmap to go from separated single organ systems to a data-driven and knowledge-driven fully integrated multi-scale and multi-organ whole-body twin. EDITH will facilitate this process by building an evolutionary ecosystem driven by a consensus among the relevant European communities and implemented through the aid of practical tools, such as a data/model repository and a simulation platform.

The objectives of the EDITH project are the following.

- To frame an ecosystem of digital twins in healthcare within the EU, EDITH is conducting a **mapping** of actors, initiatives, resources, and barriers in the digital twins, with the aim of ensuring adequate clinical representation and fostering the integration of all relevant stakeholders such as developers, technology and infrastructure providers, end-users, regulatory agencies, and HTA bodies.
- To build a **roadmap** towards an integrated Virtual Human Twin (VHT), identify the main research challenges and infrastructure needs and formulate clear policy recommendations. It will also address interoperability, computability and health information integration, identifying implementation needs/barriers and developing a strategy for the clinical deployment of the VHT model and its uptake in personalised clinical decision-making.
- To develop a **federated and cloud-based repository** of digital twins in healthcare (data, models, algorithms, and good practices), pooling together existing resources across Europe and providing access to relevant existing data and model repositories. The ecosystem will be leveraged to create a repository catalogue with available resources and recruit resources from the consortium and beyond.
- To outline a **simulation platform** supporting the transition towards an integrated VHT that will be implemented as a public infrastructure, providing a one-stop shop to design, develop, test, and validate single-organ digital twins and combine them with others for the integrated VHT models. Five use cases (cancer, cardiovascular, intensive care, osteoporosis, and brain) have been pre-selected to be developed as prototypes to show the added value of a simulation platform.

This document will start the procedural elements of defining the VHT vision and the preparation of the roadmap. Subsequently it will go into detail on the vision of the VHT as it is currently developed by the consortium with feedback from selected experts and stakeholders. Afterwards, the outline of the rest of the roadmap is provided, bearing in mind that changes in both sections will still occur as a consequence of the ongoing work by the consortium as well as the input obtained during the public consultation phase starting after summer 2023.

2   Genesis of the vision and roadmap outline

This deliverable results from work that started in the grant preparation phase and has continued through the review phase and into the project execution phase. Given the pressure of the first version of the deliverable to be ready by Month 10 (end of July 2023), the work has been carried out mainly within the community with input from specific external experts such as industrial colleagues in the EDITH Industry Advisory Board.

2.1   Meetings to discuss vision and roadmap

The main discussion meetings are recurring online meetings with the consortium and industry advisory board, complemented with a number of on-site consortium meetings.

- October 11th 2022: EDITH kick-off meeting
- Since October 2022 (ongoing): 4 Working groups (Mapping, Vision, Repository/Platform, Sustainability), each meeting on a biweekly basis with dedicated agenda for each working group.
- November 29-30 2022: EDITH consortium meeting Leuven (1)
- Since December 2023 (ongoing): Industry Advisory Board meeting every 2 weeks
- January 30-31 2023: EDITH consortium meeting Leuven (2)

Additionally, several public events and community of practice meetings (e.g. Avicenna Alliance, VPHi) have taken place where EDITH consortium members have been invited to present the current status of the developing VHT vision.

2.2   Writing of the roadmap

The meetings serve to develop and discuss different elements of the vision and the roadmap. The written representation of these discussions is organised through shared documents in the EDITH google drive with the possibility of having written discussions using the ISW-CoP slack channel (edith_vision_roadmap). Finally, a manuscript summarising the vision articulated below is available in the ArXiv[1].

---

[1] Arxiv details when available.

3 Outline of the roadmap

The provisional lay-out of the roadmap is in 5 parts, starting with an overview of the state of the art in technologies, initiatives, structures and infrastructure. This is followed by the vision for the VHT proposed by the EDITH consortium. This will then be further elaborated in terms of technologies required to realize the VHT, models and data and their interaction, as well as the link to the public infrastructure. Subsequently, the users are discussed along with the evolutionary ecosystem and elements of sustainability. The roadmap will end with an overview of the most important aspects as well as tangible recommendations for the implementation, roll-out and uptake of the VHT. Specifically for the context of this deliverable, the part of the roadmap (vision) that will be further developed in the next section is shown in teal.

PART 1: State of the Art
- Maturity
    - In silico medicine
    - AI in health
    - Wearables & data-driven twins
- VHT initiatives & actors in Europe, identifying trends & game-changers
    - Scientific, clinical & industrial organisations
    - Infrastructures & platforms
    - Standardisation, Regulatory & HTA actors
    - Public policy at EU and member state level

PART 2 - VISION: Set the stage for the virtual human twin
- The VHT vision and barriers
    - Introduction
    - Digital Twins in Healthcare: from generic to subject-specific
    - Digital Twins in Healthcare: the life cycle
    - Barriers to the development of Digital Twins in Healthcare
    - Vision for the virtual human twin
- The VHT ecosystem
    - The VHT community of practice
    - The VHT infrastructure
    - The VHT Standards
    - The VHT Long-term sustainability
- Vision and mission of the VHT initiative and of the EDITH action

PART 3: Technology for the virtual human twin
- Organisation of resources
    - 6-dimensional framework as a backbone
    - Semantic annotation & Taxonomy
- Data
    - Data sources
    - Data management
    - Data reuse
    - Data transformation services
- Models
    - Model's exposure as data transformation services
    - Model's exposure as data generation services
    - Models' classification as data flow orchestrations
    - Models' classification by context of use

- Integration of resources
  - o Workflows
  - o Identification of possibilities for integration
  - o Bidirectional communication with users (including knowledge generation)
- Infrastructure
  - o Repository
  - o Simulation platform
  - o Computational resources
  - o Standard Operating procedures, algorithms

PART 4: Responsible Research & Innovation related to the virtual human twin

- Regulatory science and Standards
  - o Standards for data formats, data integration and data input into models
  - o Standardisation of modelling
  - o Standards for metadata of data and models
- Health Technology Assessment and Payers
- Ethical, legal and social aspects
- Legal aspects
  - o Data privacy & protection
  - o IPR management

PART 5: Uptake of the virtual human twin

- Users
  - o user profiles
  - o Clinical users
  - o Industrial users
  - o Academic users
  - o Individual users
- Evolutionary Ecosystem
- Sustainability
  - o Business models
  - o European infrastructure

PART 6: Conclusions & recommendations

In the following sections, we will elaborate the Part 2 (vision) and the first bullet of Part 3 (6-dimensional backbone for organisation of resources).

4 Vision for the virtual human twin

4.1 Vision and barriers

4.1.1 Introduction

The EC has recently launched the *Destination Earth* initiative. To use the words of the EC, "Destination Earth (DestinE) aims to develop – on a global scale - a highly accurate digital model of the Earth to monitor and predict the interaction between natural phenomena and human activities. As part of the European Commission's Green Deal and Digital Strategy, DestinE will contribute to achieving the objectives of the twin transition, green and digital". The vision of developing a comprehensive digital twin of the planet comes from the need to define policies that can guide an incredibly complex system, the ecosphere, into a state more desirable for the human species.

There is a strong parallel with human health. The need to define policies that guide the health of the citizens of the European Union toward an increase in the quantity and quality of life for each person clashes with the difficulty of making predictions for another incredibly complex system, the human body. We can imagine a *Destination Human* initiative that aims to develop a highly accurate digital model of human health to monitor and predict the interactions between its physiological and pathological phenomena and human healthcare interventions.

To date, a *Virtual Human Twin* (VHT), a digital twin of the human body capable of predicting how the health status of any single individual may change due to internal pathophysiological processes or external interventions, does not exist. What we do have are *Digital Twins in Healthcare* (DTH), subject-specific predictive models designed to support a narrowly specific clinical decision. But it is easy to imagine how accumulating specialised DTHs, capturing fragments of causal knowledge, and digitally stored quantitative data, capturing empirical knowledge, can progressively evolve into a full-blown VHT.

To realise the VHT, it is necessary to simplify, accelerate and standardise the development of DTHs and the systematic collection of quantitative observational data on the health status of individuals over time, under the effect of different diseases, and when exposed to a variety of healthcare interventions. The first necessary step in the definition of the VHT is thus an analysis of how DTHs are currently developed and what barriers are slowing down such development. After the discussion of the DTH definition, life cycle and barriers, the vision of the VHT will be presented along with the elements crucial to its realisation. The last section puts these elements in the perspective of the EDITH coordination and support action responsible for creating this roadmap.

4.1.2 Digital Twins in Healthcare: from generic to subject-specific

The management of human health in its broadest sense requires decision-makers to take **well-informed decisions that may affect the health status of single or groups of human beings** (hereinafter generically called *reference population*). Examples of this include clinicians making decisions on personalised therapeutic strategies for a patient; researchers making decisions on possible druggable targets to pursue in basic biomedical research; healthcare authority managers planning specific policies; biomedical companies seeking to refine, reduce and partially replace animal and human experimentation for the regulatory approval of new products; etc. This decision-making process usually involves the quantification of specific constructs that represent such health status called Outcomes, with selected metrics called **Quantity of Interest** (QoI), and then observing how such QoI develops in time, due to variations of internal (e.g., body weight) or external conditions (e.g., exposure to pollutants), or because of intentional interventions. The **Context of Use** (CoU) defines how the QoI informs a specific decision-making process relevant to human health and under which specific conditions such process occurs.

QoIs are usually measured experimentally, either directly on human volunteers or patients or indirectly on surrogates such as animal or in vitro models. But these experiments pose a long list of practical, ethical, legal, and socioeconomic challenges and are primarily responsible for healthcare services' high costs and limited capacity. Thus, there is intense ongoing research on developing new technologies that can refine, reduce, and partially replace the need for experimental measurements to estimate the QoIs necessary to support decision-making within specific CoUs.

A DTH is a computer simulation that predicts (as opposed to measuring experimentally) quantities of interest necessary to support decision-making within a specific context of use in healthcare.

Here, "computer simulation" refers to any software capable of predicting specific outputs given certain inputs. DTHs can be predominantly knowledge-driven predictive models built using existing knowledge about physics, chemistry, physiology and pathology or predominantly data-driven models built from large volumes of data using statistical modelling or artificial intelligence techniques, or any combination. In other industrial sectors, the term digital twin refers to a real-time computer simulation informed with sensor data. In the context of healthcare, this is not necessarily the case, although it might be. It should be stressed that while the DTHs can predict QoIs that are difficult or impossible to measure, they can do a lot more: they can predict how QoIs will evolve in time, how they will change depending on external actions, etc.

One of the most important features of digital twins is the accuracy with which they predict such quantities; thus, DTH can be divided into three broad categories:

- **Generic DTH**, for which the expected accuracy is that the predicted value is within the range of the values measured experimentally in the reference population;
- **Population-specific DTH**, for which the expected accuracy is that the predicted value is sufficiently close to some central property (typically mean or median) of the range of the values measured experimentally in the reference population;
- **Subject-specific DTH**, for which the expected accuracy is that the predicted value is sufficiently close to the value measured experimentally in each individual in the reference population.

By "sufficiently close", we mean that the predictive accuracy of the DTH is sufficient for its purpose as defined in the CoU. Thus, the same DTH can be sufficiently accurate for one CoU and insufficiently for another.

Currently, most DTHs are designed to predict just one or a small number of QoIs with the necessary accuracy only in a narrowly defined reference population (e.g., women over 55 with osteoporosis and no other conditions). This is because to develop DTHs, we need large volumes of detailed empirical observations and/or reliable mechanistic knowledge of the physiology and pathology/pathophysiology of the organs, tissues, and cells involved, as well as the mechanism of action of any intervention involved. Because of gaps in knowledge and data, the only way to manage this complexity today is to narrow the scope of the DTH, focusing on a particular process affecting a minimal portion of the human body, and to be used for narrowly defined CoUs. While these narrowly focused DTHs are extremely useful in specific cases, the time and cost required to develop DTHs with a broader scope and wider applicability are currently prohibitive. The Virtual Human Twin can provide such a framework to develop a new generation of DTHs, capable of predicting any QoI necessary for any relevant CoU and reference population.

4.1.3 *Digital Twins in Healthcare: the life cycle*

The development of a DTH is a long and cumbersome process. It starts with the **identification of the clinical needs** expressed through epidemiological evidence that quantifies the limits of the current standard of care. HeartFlow is one of the first DTHs adopted in clinical practice[2], addressing a clear clinical need. Even though there is universal consensus among cardiologists that the best way to choose the most appropriate treatment for coronary stenosis is a Fractional Flow Reserve (FFR) measurement obtained through an invasive diagnostic test, only 20% of the UK patients with this condition are treated based on an FFR. Heartflow provides a quantification of the FFR based on medical images. Another clinical need addressed by a DTH is osteoporosis, where the standard of care requires specialists to decide whether to treat a patient using DXA-aBMD as a predictor of hip fracture risk. With this risk predictor, around one-third of the patients are not treated; of those, around 50% will experience a hip fracture in the following five years. Considering that current treatment can reduce the incidence of hip fracture by around 50%, with a better risk predictor, up to 7.5% of all hip fractures (60,000 per year only in Europe) could be avoided. The Bologna Biomechanical CT-Hip (BBCT-Hip) DTH estimates

---

fracture risk based on computer modelling and simulation on personalised patient data considering a wide range of fall scenarios[3].

The second step is **determining the causal relationship** between the QoI (hereinafter generically referred to as the model's outputs) and the parameters that control it (hereinafter generically referred to as the model's inputs). Living organisms are *entangled*, meaning that each internal state variable depends to some extent on all the other internal state variables. But some variables have a greater effect on the QoI than others. Ideally, for each QoI, one should run a sensitivity analysis on how the variation of any other possible QoI in the human body affects it. Since such systematic exploration is impossible, we use all available causal knowledge about the human body's physics, chemistry, physiology, and pathology to identify the minimum set of inputs that would allow a reasonably accurate prediction of the desired outputs. In some cases, the desired input cannot be measured on a patient-specific basis, and in such case we need to build a statistical model for such a quantity that describes how it varies across the population of interest, possibly as a function of the inputs that can be measured for each patient.

When the available causal knowledge is insufficient to build a reliable predictor, data-driven modelling techniques can be used to identify the best possible predictor from all available inputs. But **training a predictor** requires a large collection of data, both in depth (data from many diverse patients are required to train the predictor) and breadth (as we do not know *a priori* which quantities govern the QoI, we need to explore as many as possible).

The third step is the **model's implementation**. This is essentially a software development exercise that must be performed with the highest possible quality assurance level. A key factor here is the availability of accurate input data to build the benchmark problems used in the solver's verification. Another aspect is the definition of the model's execution environment. Depending on the nature of the data, there might be ethical-legal constraints imposing the data to be stored only at specific locations and under certain levels of cybersecurity; depending on the model implementation, there might be computational requirements that impose that the model executes only on specific computers.

The fourth step is the development of all necessary **pre-processing and post-processing tools**. Pre-processing tools are those that extract the necessary inputs from the available data. We might need the volume of a tumour, which can be measured on a 3D MRI dataset, but only once the tumour is segmented in the images. The accuracy and the degree of automation of pre-processing tools are critical, and frequently excellent models are poorly informed by sub-optimal pre-processing tools. Post-processing tools are required when the model's output is not the QoI required to support the clinical decision-making process optimally.

The fifth and most important step is the **model's credibility assessment**. This vast and long process requires first data from tightly controlled experiments to conduct the verification, validation, and uncertainty quantification. There is also a need for data that quantify the range of applicability of the model when used in clinical practice. Once the technical validation is completed, additional clinical validation might be required. This should be done independently from those who developed the DTH and using prospective clinical studies. However, in some cases, the regulator may accept studies on registry data (as far as publicly available) as evidence of clinical validity. In the regulatory space, there is also an ongoing discussion on the possibility of certifying a DTH by allowing the regulators to conduct validation studies against publicly available experimental data.

The last step is the **provision of (clinical) access**. DTH can be made available to the end-users as software embedded in the medical imaging consoles, installable software, software-as-a-service, etc. This is also related to the business models to make the DTH widely available, which should include not-for-profit modalities (for example, clinical end-users funding the further development of specialised DTH by no-profit organisations).

4.1.4 *Barriers to the development and adoption of Digital Twins in Healthcare*

A good starting point to conduct this analysis is a similar one conducted by the In Silico World Consortium, which aimed to identify the barriers slowing down the adoption of In Silico Trials (IST). In Silico Trials are digital twins of cohorts of patients, which are used to assess the safety and efficacy

---

[3] Keaveny TM, Clarke BL, Cosman F, Orwoll ES, Siris ES, Khosla S, Bouxsein ML. Biomechanical Computed Tomography analysis (BCT) for clinical assessment of osteoporosis. Osteoporos Int. 2020 Jun;31(6):1025-1048. doi: 10.1007/s00198-020-05384-2. Epub 2020 Apr 26. PMID: 32335687; PMCID: PMC7237403.

of new medical products before their widespread use. This analysis identified seven barriers: lack of advanced models, lack of independent validation collections, no clear regulatory pathways, poorly informed stakeholders, poor scalability and efficiency, lack of trained workforce, and lack of accepted business models. Re-analysing these barriers in light of the VHT, we propose the following list of important barriers that need to be addressed for the VHT to realise its full potential in advancing human health.

1. Lack of high-quality data available in open access for the development and validation of DTHs
2. Lack of software components for the development of DTH available under an Open-Source license
3. Difficulties related to multiscale/multisystem models by available data at all space-time scales and for all pathophysiological processes.
4. Lack of Open-Source software libraries and standard operating procedures (SOPs) that simplify the deployment and the certification of the ICT infrastructures for the provision of clinical access, supporting distributed storage and execution models.
5. Lack of consolidated existing regulatory pathways with the EU and lack of harmonisation of a technical standard equivalent to the ASME VV.40:2018.
6. Lack of legal clarity and certainty for the clinical deployment of DHTs
7. Lack of well-established operating procedures and supporting data to demonstrate the efficacy and cost-effectiveness of DTHs.
8. Need for recruitment of medical technology experts by healthcare providers and their professional recognition as co-decision makers in the hospital DTHs investments.
9. Need for exploring strategies supporting a smooth transition from the pre-competitive to the competitive development of DTHs and developing value propositions to attract the investments of SMEs and large med techs.
10. Need for creating training and re-training programs on developing, assessing, and using DTHs.
11. Lack of well-informed stakeholders

Barrier 1: Lack of high-quality data available in open access for the development and validation of DTHs

As explained in section 1.1.1, the entire development cycle for a DTH involves access to high-quality data. Since these data are generally not accessible, each development team must spend significant time designing and conducting the experiments to produce the necessary data. Most of these data are then kept private, as they are perceived as the team's main competitive advantage over the others. Sharing in open access the experimental data generated to develop AND validate models is essential for the faster development of DTH. The development of Open Access, high-quality validation collections should be funded by funders and regulatory agencies, which could use them as independent validation evidence when comparing DTH targeting the same clinical problem.

Barrier 2: Lack of software components for the development of DTH available under Open-Source license

The first DTHs were artisanal software artefacts, where the developers implemented each required function from scratch. But as the field matures, this approach causes "the same wheel to be re-invented many times". This is also caused by the very low re-usability of DTH academic software and limited use of Open-Source licensing (again seen as a way to lose competitive advantage over the other academic teams). In an ideal scenario, a DTH developer should focus on the core business of its model, reusing existing software for all other ancillary functions.

Barrier 3: Difficulties related to multiscale/multisystem models by available data at all space-time scales and for all pathophysiological processes.

Probably the most significant barrier is that the complexity of the effort required to develop a DTH increases exponentially. The first generation of DTH focused on problems that were very much confined to space-time and physiological sub-systems. The accurate prediction of the FFR of a stenotic coronary artery can be obtained by modelling the phenomenon at a single space-time scale and

considering only the physiology of the cardiovascular system (most parts of which can be lumped into some cleverly calibrated Windkessel models). Predicting the volumetric growth rate of a solid tumour as a function of chemotherapy is a much more challenging problem because it has to span multiple space-time scales (whole tumour, cell-to-cell interaction, single cell system biology, molecular dynamics of binding affinity). If the growth is assumed unconfined, the pathophysiology knowledge required to build the model is limited to the tumour. But suppose one wants to model how the tumour's growth interacts with the surrounding organs. How tumour biology interacts with the host biology, or how tumour cells penetrate, detach, are transported, and form metastases, we would need to account for the pathophysiology of a good part of the human body. With each scale and each physiological sub-system added, the amount of data and knowledge required to build the model increases exponentially, as does the challenge of assessing its credibility.

The only solution to this problem is a "divide et impera" approach, where each scale, each physiological sub-system, is modelled independently, and complex multiscale/multisystem models are built as orchestrations of such "atomic" component models. Some authors call these orchestrations *hyper-models* to stress that they are models of models.

In system biology, the reuse of models as components for more complex models has been addressed using standardised modelling languages and ontologies to ensure interoperability (e.g., CellML[4]). However, this extremely elegant approach works well only for models where the mathematical representation and its numerical solution can be completely separated (as it is for algebraic ordinary differential equations). In field problems where partial differential equations must be used, the complexity of adopting this approach becomes considerable and has yet to find widespread adoption. A second approach, explored in some research projects, is to define the hyper-models as orchestrations of remote procedure calls, such as web services. But here, the limiting factor is the complexity of developing, maintaining, and adopting the necessary orchestration libraries (i.e., MUSCLE2, or VPH-HF v1 or v2). The simplest approach is to build the orchestration only in terms of data flow. Traditionally, this approach assumes that intermediate data objects are not useful and, thus, are not stored permanently (i.e., Taverna Workflow Manager). But suppose one imagines a very large persistent dataspace; in that case, all DTHs could be reduced to atomic component models, as the data from other scales or sub-systems would be available as inputs. Of course, this would imply that the data are available in the dataspace at the space-time scale they were measured/predicted and homogenised/particularised at upper/lower scales.

Barrier 4: Lack of Open-Source software libraries and standard operating procedures (SOPs) that simplify the deployment and the certification of the ICT infrastructures for the provision of clinical access, supporting distributed storage and execution models.

A single DTH use needs to solve a single digital twin, whereas a single IST use may need to solve hundreds or thousands of digital twin models. Hence, where poor scalability and efficiency are significant challenges for advancing IST, for DTHs, the problem is more about providing *clinical access*. Each DTH solution needs to balance the need to keep sensitive data at prescribed locations and with the need to use appropriate computational resources usually unavailable within the hospital. Additionally, for the VHT to generate (clinical) impact, access to the compute infrastructure network needs to be streamlined and facilitated.

Barrier 5: Lack of consolidated regulatory pathways with the EU and lack of harmonisation of a technical standard equivalent to the ASME VV.40:2018.

DTHs are seen, from a regulatory point of view, as Software As Medical Device (SaMD) with predictive capabilities. FDA has recently clarified that also for those, the credibility can be assessed following the ASME VV-40:2018. IEC and ISO are working on a similar standard, which could be harmonised in the EU regulatory system. However, the VV-40 is recommended, even if not an EU-harmonised standard.

Barrier 6: Lack of legal clarity and certainty for the clinical deployment of DHTs

---

4 https://www.cellml.org/

The use and deployment of DHT technology in a clinical setting require stitching together the legitimate expectations of societal protection, the obligations of compliance with the different applications (and evolving) regulations and the ethical demands underlying the ongoing technological developments. Establishing a solid and uniform regulatory and legal framework will provide certainty and clarity to technology developers and providers and reassure the legal teams and investors. There is a need for a concrete and coherent environment to improve the development of DTHs by defining the legal criteria to be adopted to allocate risks and liabilities.

Barrier 7: Lack of well-established operating procedures and supporting data to demonstrate the efficacy and cost-effectiveness of DTHs.

Paramount to the uptake of DHTs in clinical practice is the ability to demonstrate their efficacy and cost-effectiveness. For clinicians, we need to provide clear, conclusive evidence that using the DTH significantly improves the current standard of care (efficacy). For Payers (healthcare authorities, insurance companies, etc.), we need to demonstrate the cost-effectiveness of the DTH when compared to the standard of care. Currently, only anecdotal information is available on either aspect.

Barrier 8: Need for recruitment of medical technology experts by healthcare providers and their professional recognition as co-decision makers in the hospital DTHs investments.

Depending on the role of the DHT in the medicinal product life cycle (as a generator of digital evidence or as the product itself), different business models need to be developed. SaMD-DHTs could be considered health technologies that hospitals buy as instrumentations or services through competitive bids. But providing a clinical access model plays a significant role in deciding the commercial positioning. DTH developers can position themselves as Original Equipment Manufacturers (OEM) for established medical technology providers, as medical technology providers that sell instrumentation to hospitals, or as service providers with various business models linked to this.

Barrier 9: Need for exploration of strategies supporting a smooth transition from the pre-competitive to the competitive development of DTHs and develop value propositions to attract the investments of SMEs and large med techs.

Key players are clear that, even though DTH models are often still at relatively low TRLs, the substantial current R&D investments into DTH science are insufficient to establish an economic sector. Instead, the need is for an economic value chain and ecosystem around DTH, with goods and service producers and consumers, prescribers, payers, supporting players such as certifiers, data or computational resource providers, and consolidated distribution channels. Implicit here is an entrepreneurial understanding of the pathways leading from scientific innovation to monetisation. Yet, it is not surprising that this understanding is difficult to achieve, because of several peculiarities of the DTH sector, given the scarce amount of experience that can be carried over from other sectors. Equally, it is impossible to provide a shared understanding of business mechanisms specific to DTH by analysis of existing business ventures alone because there simply are too few, and many are still regularly pivoting their models. In close concertation with the industry, value propositions for the VHT must be developed and a clear understanding of the balance between public investments/infrastructure and market/commercial dynamics.

Barrier 10: Need for creating training and re-training programs on developing, assessing, and using DTHs.

We need experts in the companies developing the DTHs, in the notified bodies that CE-mark DTHs, in the HTA authorities that evaluate their cost-effectiveness, and in the healthcare providers to plan and steer the necessary technological investments toward this area. Linked to this last point is a delicate issue of where the decisional procurement power lies within healthcare providers.

Barrier 11: Lack of well-informed stakeholders

Adopting silico medicine solutions involves a long list of stakeholders, from the citizens/patients to policymakers. Virtually none of these non-technical stakeholders is well informed on the potentials and limitations of in silico methodologies. Various polls suggest opinions oscillating from the excessive

expectations caused by the over-hype for AI to a principled rejection of the whole concept of predicting health. We must ensure all these stakeholders receive accurate, factual and adequately balanced information.

4.1.5 *Vision for the virtual human twin*

From the analysis of the life cycle of a DTH, and of the main barriers to its wider adoption, we can start to formulate some elements of the Vision of what the Virtual Human Twin could be. Other valuable elements come from a critical review of how the available knowledge on the human body is currently used in healthcare decision-making. Two significant issues stand out: not all knowledge is actionable, and too much knowledge is produced and used under severe reductionist idealisations. Also, a lot of data are collected cross-sectionally, losing the dynamics over time of the processes.

The simplest way to form knowledge on human health is to collect observational data on individuals at various points in their life span, in healthy and diseased status, and with/without the effect of specific healthcare interventions. But this is limited by several methodological, ethical, and legal difficulties: as a result, only a small fraction of this knowledge is *actionable*, e.g., can be acted upon to solve practical problems such as healthcare decision-making. Tentative knowledge is considered actionable when its *credibility* is high enough to be helpful in solving that problem. But this is frequently an issue:

- Many available observational **datasets are qualitative or semi-quantitative**, even though the phenomenon of interest can, in principle, be expressed quantitatively. This makes the quantification of the credibility of the resulting tentative knowledge problematic.

- Much data is of **low quality or, worse, of unknown quality**. The latter case makes it impossible to estimate the credibility of the resulting tentative knowledge.

- Most **observational data is obtained from model organisms** such as rats, mice, fruit flies, worms, zebrafish, etc. But these represent empirical knowledge only for the organism they have been observed in. These organisms become models when we use them to infer knowledge about human health. Before considering the information, they produce as tentative knowledge, the *analogy* (functional similarity) between the animal model and the human target, must be demonstrated. which is rarely done (one example being Data Resource Center of the the NIH-funded SPARC project[5]).

- Another problem that we need to overcome are the **too many reductionist compromises** that impair medical sciences:

  o The first is the concept of the **average patient**. In the VHT, all knowledge should be referred to individuals at a given point in their life; population knowledge, where necessary, should be generated dynamically by averaging individual information. Same for probabilities over a time span, as required, for example, by epidemiology. This would enable a truly personalised medicine, where every bit of information (not only a subset, i.e., the genome) is used to inform the medical decision.

  o The second is that of **scale separation**. For each physiological or pathological process, available knowledge stored in the VHT should cover from the atomistic scale to the whole organism scale, and sometimes, where the interaction with the environment is a paramount, even beyond. Homogenisation and particularisation models would capture the knowledge of how the data change depending on the space-time scale at which we observe them, providing a continuum representation of the knowledge over space-time. Finally, we could link the clinical signs at the organism level with the relevant microscopic events at the cellular and molecular scales.

  o The third is that of **system separation**. Medical knowledge is organised into specialities[6], defined by the conventional separation of the body into 12 organ systems (cardiology, neurology, respiratory, etc.), by specific diseases (oncology, venereology), or by organisational roles (emergency, general). In the VHT, knowledge should be accumulated for all organ systems, all disease processes, and every phase of the health

---

care process. This opens to a truly holistic medicine still rooted in the principles of the scientific method and western medicine.

- ○ The fourth reductionist compromise is that of **age separation**. Medical knowledge tends to separate life span in paediatric (< 16yo), adult, and geriatric (>65yo). But there is solid evidence of *lifelong health*: what happened in our childhood influences adult and geriatric health. In the VHT, the entire clinical history of an individual, from birth to death, should be captured.

From these sparse considerations, we can attempt a first definition of vision:

The Virtual human twin is an integrated multi-scale, -time and -discipline digital representation of the whole body enabling the comprehensive characterisation of the physiological and the pathological state in its heterogeneity and allowing patient-specific predictions for the prevention, prediction, screening, diagnosis and treatment of a disease, as well as the evaluation, optimisation, selection and personalisation of intervention options.

More practically, the *Virtual Human Twin is an ever-growing accumulation of all quantitative knowledge on how individual subjects' health status changes over time. All knowledge is digitally stored in the form of adequately annotated data and predictive models. Observational data capture empirical knowledge, whereas predictive models capture causal knowledge. Data and models must be annotated with enough information to assess their credibility.*

Who is going to realise this vision? As this challenge is beyond the capabilities of any single organisation or even community (e.g., in silico medicine, wearables), we are convinced the VHT can only be realised by establishing and engaging the entire ecosystem. The notion of ecosystems captures the complex set of interlinkages among sectors and industries. The ecosystem encompasses all players operating along a value chain: the smallest start-ups and the largest companies, the research activities, the services providers and suppliers, the users and all other stakeholders. There are several key elements related to the ecosystem that requires further elaboration, the first of which is the ***community of practice*** (CoP), which includes academic and industrial researchers, developers of DTH solutions, clinical users, industrial users, CROs, regulators, health authorities and payers, policymakers, etc. A stub for this community already exists: the In Silico World project has recently transferred to the VPH Institute an online community of practice called ISW_CoP, based on Slack, which hosts the conversations of over 600 experts on best practices for in silico medicine. The VHT community must define best practices to make the VHT vision a reality through consensus processes such as this road-mapping effort. While the definition of these best practices is a continuous process, some educated guesses can be already made, which help formulate the VHT initiative's mission.

To realise the VHT vision, the VHT CoP needs to articulate a mission formed by three additional key elements:

- The ***Infrastructure*** provides a concrete operational space within which the VHT knowledge is generated, stored, annotated, revised, integrated, exchanged, etc.
- The ***Standards*** capture the consensus on how things must be done to ensure the highest possible level of interoperability and reuse.
- ***Long-term sustainability*** defines how the ecosystem collaborates and evolves to fulfil this shared vision.

Thus, the VHT is a vision that can only be realised by following an ecosystem approach, including establishing a CoP, an infrastructure, a set of standards, and a tangible outlook on long-term sustainability.

4.2 The VHT ecosystem

4.2.1 The VHT community of practice

The stakeholder groups can be defined by looking at the typical value chain for medical technologies:

- **Researchers** who develop new knowledge and new methodologies and test them in pre-clinical and clinical settings;
- **Innovators** who translate research results into potential solutions for clinical or industrial unmet needs;
- **OEMs** (Original Equipment Manufacturers), typically software developers that provide libraries and solvers used to implement the Digital Twins;
- **Business angels and investors** who support the creation of new companies that want to sell Digital Twins;
- **CROs** (Contract Research Organisations) that assist in the conduction of clinical studies for the validation of digital twins but also that can use digital twins to design and optimise clinical studies of new treatments;
- **Medical Device regulators** that provide marketing authorisation for digital twins that are used as clinical decision support systems, but also qualification for digital twins used as medical device development tools;
- **Drug/ATMP regulators** that provide qualification for digital twins used as drug/ATMP development tools;
- **Conformity assessment agencies** (e.g. Notified bodies) designated by an EU country to assess the conformity against relevant regulations and applicable standards of digital twins for healthcare or other medical products developed with digital twins before they are placed on the market.
- **HTA authorities** that evaluate the effectiveness, cost-effectiveness, and reimbursement level of new digital twins for healthcare or other medical products developed with digital twins;
- **Digital Twin sellers**, which can be categorised depending on their business model:
    - *Biomedical instrumentation sellers* that sell the digital twins as software embedded in their hardware;
    - *Medical device sellers* that sell digital twins as algorithms embedded in their medical devices or that complement them;
    - *Medical software sellers* that sell digital twins as stand-alone products or services;
    - *Broker sellers* that sell digital twins developed by third parties, usually as software as a service (SaaS);
- **Data brokers** that curate and resell use licenses for data collections usually generated by third parties;
- **GDPR officers**, including data controllers, data providers, DPOs, supervisory authorities, and their legal advisors, involved with the handling of sensitive data used to develop, validate, or use digital twins.
- **Buyers and Payers** that buy digital twin in healthcare technology to provide healthcare. This is a variegated galaxy of stakeholders linked to the national or regional healthcare provision model. It includes private and public healthcare providers, insurance, healthcare authorities, group buyers, etc.
- **Medical Product Developers** who use digital twins to design, optimise, or assess the safety and efficacy of new medical products, both in the pre-regulatory and regulatory phases;
- **Healthcare policymakers** who may develop specific policies linked to the use of digital twins but also who may use digital twins to support policy making, usually with the mediation of experts.

For each of these stakeholders, the VHT represents a different value proposition to which each stakeholder can contribute with different added value.

For researchers, the VHT represents the opportunity to continue focusing on their narrow specialism while at the same time developing DTHs that encompass various organ systems, pathophysiology processes, etc. The slogan is: "holism by collaboration". In the early stage of the ecosystem (see sec 1.3), researchers will be the main contributors to data and models. Still, researchers-run repositories will play the lion's share even when the VHT becomes a more mature and industrially relevant infrastructure.

Innovators are probably the ones with the most significant potential benefit. Taking to the market a DTH research prototype today takes between five and ten years; with the VHT, this time to market could be cut in half. They will contribute to the VHT by consolidating available resources into validated and accepted pathways for innovation, regulatory approval, etc.

OEM will use the VHT to better position their software layer and ultimately increase the revenues from the healthcare domain. They will provide solvers and libraries with a high level of maturity, technical and certification.

Business angels and investors will use the VHT as the primary source of information to guide their investments in the sector. Navigating the VHT, they will find the innovators with the most profitable products, develop new business opportunities as the sector structures and specialises, and guide start-ups and spin-offs toward the aspects of the business that maximise revenues. They will give the VHT business wisdom and guide its development toward sustainable business models.

All Contract Research Organisations (CRO) will benefit from the VHT as DTHs reduce animal experimentation and refine and reduce human experimentation. But the most innovative CROs will be able to offer their customer a different level of service, where they support product development from discovery to post-marketing surveillance by maintaining data and models that support the entire development cycle of whole families of similar products. CROs will provide the VHT with the first level of commercial use and contribute to its long-term sustainability.

Medical Device regulators such as the FDA CDRH or conformity assessment agencies such as the European notified bodies will use the VHT to consolidate real-world data and in silico testing frameworks that will speed up but also drastically improve the efficacy of regulatory surveillance. They will contribute to the VHT by providing the highest levels of credibility through appropriate certification/qualification pathways.

Regulators of medicinal products such as FDA CDER and EMA will initially use the VHT to handle well-known regulatory challenges: better-powered dose-response and safety studies, placebo arms where the placebo is unethical, testing treatments for rare diseases, etc. DTHs may also provide non-invasive, more ethical alternatives to quantifying important biomarkers and refining human experimentation. But we hope they also use the VHT to slowly move their regulatory epistemology to less phenomenological paradigms. Drug regulators will contribute to the VHT with high complexity challenges around multi-organ, multi-system processes.

Regulators of Advanced Therapeutic Medicinal Products such as CBER, and in general of all products that fit poorly into the classic device/drug separation, will finally have the opportunity, using the VHT, to develop a regulatory science more specific for these complex classes of medical products. They will provide the VHT with a drive to expand its limits to modelling all kinds of interventions, including gene and cell therapies, tissue engineering, etc.

Also, HTA authorities can harvest huge benefits from the VHT. In the short term, they will benefit from developing an in-silico-assisted HTA, where many questions can be first explored in silico. But in the long run, a fully integrated in silico development and quality assurance of biomedical products will enable a whole different HTA paradigm, where the transition from efficacy to effectiveness becomes a smoother, more continuous process, and the costs of post-marketing surveillance are drastically reduced (and thus its scope can be expanded without damaging innovation). The VHT will also make the pricing/reimbursement process more factual. HTA experts contribute to the VHT by expanding the initial focus on clinical use to a broader view of human health, which is also made of social, economic, and organisational determinants.

New companies that commercialise digital twins will see an explosion of marketing opportunities. They will monitor the VHT for pre-commercial solutions and guide the developers to their commercial exploitation according to their preferred business model. Biomedical instrumentation sellers will use the VHT as the innovation emporium, where to pick the next new thing to add to their product line. The VHT will make it easier to develop smarter medical devices that embed predictive capabilities and are supplemented by them. Medical software sellers and brokers will have an easy way to expand their portfolios, but also which solutions are worth investing in, in terms of credibility and popularity. Their role is key to the VHT, and they will drive to a more sustainable state.

The VHT will drive and consolidate the emerging business of health data brokers. Data brokers will use the VHT to find valuable collections or compose them by merging data collected by separate entities

and assist them in transforming these resources into revenues. They will also contribute substantially to the long-term sustainability of the VHT.

The development, validation and use of DTHs currently pose several challenges for those legally responsible for treating sensitive data. The VHT will provide them with new solutions for these problematic activities regarding data protection and associated legal risks. Privacy experts will be essential in developing the ethical/legal framework within which the VHT will operate. They also advise the policymakers on any legislative intervention that might be necessary to make this aspect less critical than it is now.

Buyers and payers face significant challenges in handling disruptive innovations like DTHs. Leaving to med-tech giants to guide it will prevent exploitation that changes business models and may reduce costs while improving service. The VHT will make it easier to test and explore innovations in healthcare delivery models and build public procurements, even in the form of trans-European consortia. These experts will be critical in showing where the VHT can and does impact the European healthcare system.

The use of digital twins to develop and de-risk biomedical products, what we call In Silico Trials, has been and remains one of the biggest opportunities for medical companies involved with the VHT. The VHT will make the development, validation, and qualification of In Silico Trials methodologies easier, cheaper, and faster. Medical companies will drive the industrial development of the VHT.

Last but not least, healthcare policymakers at all levels (regional, national, union) have a tremendous opportunity with the VHT to tap into a community of experts that can inform policy decisions with data, facts, and a broad spectrum of specialist expertise. Policymakers may become essential for developing the VHT, where legislative barriers or the lack of clear legislation may cause the VHT to develop more slowly and thus have a smaller impact than expected.

4.2.2   The VHT infrastructure

What we describe here is a preliminary general vision of what, in the long term, the VHT infrastructure should be. This will inspire the proof of concept that the CSA EDITH is developing, but for obvious reasons, it will not be even close to this level of ambition. The **VHT infrastructure** should be developed around five strategic pillars considered essential for the global effort to drive forward **VHT** development.

1. Distributed/federated architecture
2. Governance
3. Openness
4. User Engagement
5. Industry collaborations and partnerships

Distributed/federated architecture

The heart of the VHT infrastructure will consist of a distributed/federated platform. The different services of the VHT platform will be semantically mapped to different types that occur within the context of the VHT. The platform will include centralised **core elements**, *i.e.*, the "hidden" elements required to run the platform. The second level will include the **platform and science-specific services**, *i.e.,* generic platform elements necessary to the end-users (e.g., wiki, collaborative documents), scientific services like the repository, tools for running the workflows etc. These might include services for semantic re-annotation or services to promote resources along the Credibility axis. The last level (domain-specific **services**) is end-user facing and includes all the federated services relevant to the VHT.

A future benefit of administratively treating **domain-specific** products and **services** separately from **Platform and Science specific services** is that it facilitates an easy onboarding path for more domain-specific tools and services and even tools from other scientific domains. The onboarding is further facilitated by the federated -more flexible- nature of the VHT platform. Component owners will enjoy full *autonomy* regarding the services and applications they provide. New services will be developed, mature, and be provided as federated services that will be part of the VHT platform, following integration/quality/interoperability requirements.

The VHT infrastructure needs to be "**as open as possible, as close as necessary**", *i.e.,* very accommodating of various existing components, formats, and protocols, but on the other hand, it needs

to provide a unified and intuitive user experience that does not expose all the "sharp edges" of the underlying machinery and duct-tape. This follows the Robustness Principle (Postel's Law) made famous during the specification of the TCP protocol: "be conservative in what you do, be liberal in what you accept from others".

The implementation as a distributed/federated platform will be more **flexible** and **adaptable** to changing requirements and user needs. Different entities can choose which technologies and standards to use and evolve their systems independently. Moreover, its distributed nature will allow the platform to easily **scale** to handle increased traffic or user demands by adding more servers or nodes and to be deployed across **multiple locations**, allowing for a wider reach and better performance for users in different regions. Given the distributed nature of the storage services archiving data and the computing services elaborating on them, they need to be connected by high-speed geographic networks.

A federated/distributed infrastructure can also promote **interoperability** between different systems, allowing users to communicate and share data across different platforms, and fostering innovation. At the same time, it also improves the **performance** of the platform by reducing latency and increasing bandwidth. However, it may also require more coordination and **governance** to ensure interoperability and maintain quality.

Governance

Governance is essential since it defines the management structure, roles, and decision-making procedures. The **governance framework** will identify/establish the policies, governing roles and responsibilities (admin, provider, and user profiles) and decide the standards to adopt (section 4.2.3). It will also ensure the clarity of the business models and access policies (tiers, pricing policies, commercial agreements). It will build a detailed roadmap following the evolutionary ecosystem approach (section 4.2.4).

Openness

Openness in the VHT infrastructure will allow users and developers to share their work and collaborate with others, bringing together people with different backgrounds and expertise, leading to more diverse perspectives and insights, reducing, at the same time, duplication of effort and resources by allowing users and developers to build on each other's work. The research findings will be available to the entire CoP: the scientific community, the policymakers, and the stakeholders. An open VHT infrastructure will also facilitate the sharing of data and resources, encouraging, at the same time, uniformity of protocols and formats, standardised wherever possible and standardisable in hopefully all other cases (section 4.2.3).

User Engagement

A user-friendly and visually appealing platform will facilitate the CoP's interaction with the VHT infrastructure. The design will be user-centred, and user surveys will be used to gather insights that inform the design and functionality. The layout will be easy to navigate. The platform will have a dashboard that will guide the user through the different offerings of the VHT, allowing the users to customise their profiles and add the services and tools that are useful to them. Clear documentation will be available, not only for the main functions of the platform but also for the different services. The platform will combine tools and services from different sources, all valuable for developing the VHT. Users will use the platform also for collaboration and interaction with other users. The ecosystem will provide user incentives for sharing data, models, or other content (section 4.2.4), while it will consider the feedback from the users, working to improve itself. Finally, user engagement should be measured! Analytics tools will be used to monitor user behaviour and identify improvement areas.

Industry Collaborations and Partnerships

As mentioned in the previous section, the industry can provide valuable guidance and feedback for designing and developing the VHT infrastructure. This collaboration can help ensure that the following steps focus on developing technologies and solutions that are relevant and useful to the real world while advancing the field of virtual human twins. By providing real-world context, the industry can help to better understand the challenges and opportunities associated with VHT. At the same time, the industry

can provide expertise and resources (access to data, software tools, etc.). Their feedback will help develop more relevant and valuable solutions for external users while advancing the general field of VHT.

4.2.3   The VHT Standards

The VHT will heavily rely on standardisation to ensure the highest level of interoperability. Where possible, we will support formal technical standards developed by international standardisation bodies; where these are missing, we will use *de facto* standards. The Community of Practice will also produce standardisation by generating Standard Operating Procedures (SOPs) that codify the good practices and the rule of use of the VHT.

Lastly, the VHT will be part of a European ecosystem of information technology services such as the European Health Data Space. Here we commit to achieving the highest possible integration and interoperability with such services, including adopting specific standards that these services support.

This said, our view on the use of standards is pragmatic. It is essential to state that the goal is to ensure the broadest possible adoption for the VHT. So, when standards help this by simplifying access and ensuring high levels of interoperability, they are welcome; when the support of particular standards involves a significant overhead that complicates the adoption of the VHT, they will not be supported. Similarly, if the community is split relatively evenly between two *de facto* standards, we will support both, asking the promoters to work on translation tools that facilitate moving from one to the other.

4.2.4   The VHT Long-term sustainability

The interactions within the VHT ecosystem are driven by rules, policies, and standard operating procedures that regulate how the community of practice members exchange data, models, and services as they co-develop the VHT.

Given the complexity of the community of practice involved and the variety of value propositions they expect from the VHT, we cannot imagine the ecosystem as rigidly fixed. At the risk of oversimplifying, we can distinguish three phases, which presuppose the initial realisation, within EDITH's infrastructure, of a system of Distributed Ledger Technology (DLT) allowing to permanent trace all types of assets exchanged on the DLT, also tracking their provenance and securing the findability, accessibility, semantic interoperability, and reusability of all activated resources.

In the early stages, the DLT infrastructure will host exclusively pre-competitive transactions and work on incentives based on forms of reputational scoring (Honour ledger); in the second stage, pre-competitive and competitive transactions will coexist, and exchanges will be facilitated through the issuance by the DLT infrastructure, of digital tokens with no direct monetary value, but operating as the scaffold on which symbolic prices can emerge through supply and demand of all assets traded, included the DLT services (Token ledger); in the third and final stage, the ecosystem will mature and specialise: while some entities dealing mainly with pre-competitive transactions will continue to exist, a growing number of subjects will increasingly focus on competitive transactions in the form of business-to-business exchanges, with prices set in Euros and no-more in tokens (Money marketplace).

Honour ledger phase

The currency used in this phase is only reputation, based on the reciprocal assignment of quality scoring by all the partners involved. The ecosystem de facto operates as a barter mechanism facilitated by the operation of the DLT infrastructure. Tracing all transactions and reputational outcomes, the ledger will be entirely funded with public money.

The main focus in this phase will be on interoperability, e.g., how the participants can barter data and models with the smallest possible effort. The DLT infrastructure should make it very easy to share a dataset or a model in ways compliant with the FAIR principles.

The main motivations for sharing data and models will be because they are forced to do so by the funders. Still, also because it allows them to barter access to their resources with others they can use, and because the use of resources by others will be tracked, possibly with mechanisms that translate them into citation-based reward systems and reputational scoring.

The main motivation for reuse will be the simplification of developing and validating new models. This should be particularly true in developing multi-scale and multi-system models, where the work of one on a sub-system can be entirely re-used.

In this phase, the metric of success is how much data and how many models are shared. This translates into another sub-metric: the pain-gain ratio for contributors. The easier it is to share resources, and the more rewarding it is, the better it is.

The governance will require a reverse "T" model: the day-by-day operations will be ensured by a single organisation or a small consortium that is paid to do it and take technical decisions in a fairly autocratic way. Any other decision on the infrastructure is taken with a direct democracy model, where all contributors can participate in the decisional processes with an assembly process.

Token ledger phase

Eventually, the ecosystem will see the flourishing of competitive transactions, aside from the pre-competitive ones. The value of both will be expressed in digital tokens issued by the legal entity governing the DLT infrastructure. The "cashing-in" of tokens, by operators contributing resources to the ledger exchange system and the "paying-out" of tokens, by operators purchasing assets offered by others, will allow the development of an increasing nexus of token prices for all transactions taking place through the DLT, which will also begin to charge a token-fee for its services.

The ledger infrastructure will remain a substantial public resource, but its governance will require a more articulated democracy with representation.

The VHT ecosystem will use its growing token economy to experiment with how it can become progressively self-sustained. In a more advanced stage of development, the DLT infrastructure will possibly also engage in analysing how incentives linked to automated assignment and distribution of value can be determined by ML mechanisms valuing different attributes or even through Shapley-value mathematical methods (inspired by Nobel Laureate Lloyd Shapley) determining the only distribution satisfying a collection of properties within a coalition game.

In its basics, the token-ledger phase will be characterised by token-based exchanges replacing barters, and tokens will be issued to whoever contributes resources to the ledger. However, it will also be possible to purchase tokens in exchange for money: this will apply mainly to external entities not having contributed resources to the ledger but wishing to use the DLT facilities.

Further tokens will be gained anytime shared resources will be used, while everybody will pay with the tokens they have accumulated for being allowed to use somebody else's resources.

The main focus in developing the infrastructure will be the quality of service for its many users. Systems must scale to extensive collections and handle a truly distributed system based on various hardware and software providers.

The main motivations for sharing will remain the same as in the previous phase. But the move from an honour ledger to a tokens ledger will render the incentives for what one can get in return for his/her shared resources much more fine-grained and flexible.

The main motivation for reuse will change by the extent to which the VHT ecosystem will now be influenced by the development strategies of research groups and companies that have purchased tokens and are willing to use the DLT.

This, of course, implies guarantees of persistence for the infrastructure. In this phase, the metric of success is how important the VHT becomes in the development strategies of public and private developers.

Because of the need to ensure long-term sustainability, the VHT infrastructure will have to be run by a legal entity, possibly organised as an NGO (e.g., a foundation) or as a joint undertaking between the EC and the major European industrial players, similar to EuroHPC, IHI, etc. This organisation will need to ensure the existence of a public segment of the VHT for the not-for-profit researchers, where most interoperability technicalities standards are tested and standardised. But it will also have to favour the creation of fully commercial segments of the VHT, which are certified for interoperability by the leading legal entity. Beyond that, they are operated entirely in a private way, pursuing sustainable business-to-business models.

Market phase

Eventually, the number of transactions triggered by research groups and commercial companies having monetised their access to the DLT-operated ecosystem will possibly end up changing the nature of the latter, attracting a growing number of subjects out of the ledger and into a fully-fledged money marketplace. Academics will remain the artisans who explore the borders of the VHT territory, while the merchants and entrepreneurs increasingly tend to privatise the ecosystem. Will thus the ecosystem dissolve into a mature industrial sector, or will the DLT infrastructure maintain sufficient resilience and attractiveness because of the advanced qualities of the services it will provide?

In any case, one should expect there to be a public VHT for academic research and early pre-competitive developments, supported by the EC like any other research infrastructure, various VHTs for not-for-profit activities, supported by various charitable mechanisms, and several commercial VHT infrastructures that provide B2B services to an ever-growing industry of in silico medicine.

4.3 Vision and mission of the VHT initiative and the EDITH action

Virtual Human Twin (VHT) is an integrated multiscale, multi-time, and multi-discipline representation of quantitative human physiology and pathology. Its realisation through a collaborative distributed knowledge, resource and simulation platform is specifically designed to accelerate the development, integration, and adoption of patient-specific predictive computer models, which will be used as clinical decision support systems, for personal health forecasting or as methodologies for development and de-risking of personalised medical products.

EDITH is a Coordination and Support Action funded by the European Commission. The **vision of EDITH** is to facilitate the realisation of the opportunities presented by VHT for the benefit of patients, healthcare providers, regulatory bodies and industry, both within Europe and globally. EDITH will capitalise on the developments of digital technologies, high-performance computing, availability and access to research and healthcare data in Europe, defining a roadmap from separated single organ systems to data-driven and knowledge-driven fully integrated multiscale and multiorgan whole-body twins. EDITH will facilitate this process by building an evolutionary ecosystem driven by a consensus among the European community of practice and implemented through the aid of practical tools, such as a data/model repository (within the scope of EDITH), and a simulation platform (to be implemented after EDITH).

The goals of EDITH are to tangibly foster a sustainable ecosystem. Starting from a comprehensive roadmap of the current landscape, EDITH will implement a federated cloud-based repository, gathering human digital twin resources (models, data sets, algorithms, good practices), and design the architecture of a simulation platform to facilitate the transition towards the use of comprehensive Virtual Human Twin (VHT) models in personalised medicine.

The Virtual Human Twin will facilitate the development and adoption of digital twins in healthcare of any complexity at reasonable costs at reasonable times. As data acquisition and computing power technologies evolve, the scope for digital twins will become broader and the knowledge deeper. This will boost scientific research and technological innovation, creating massive business opportunities in the European Union. The Virtual Human Twin is science, not science fiction; it is the future of medicine.

5   Technology for the Virtual Human Twin

The elements discussed in the previous section provide the general outline for the blueprint of the Virtual Human Twin framework and infrastructure that allows to bring together resources across scales, levels, types, time and organ systems. This section provides a high-level description of the essential characteristics of this VHT framework.

As described above, *the Virtual Human Twin infrastructure is a collaborative distributed knowledge repository and simulation platform of quantitative human (patho)physiology, designed specifically to accelerate the development, the integration, and adoption of patient-specific predictive computer models as clinical decision support systems or as methodologies for the development and derisking of new medical products*. This framework can be imagined as an n-dimensional data space, which DTH models constantly crawl. This data space represents the totality of our current collective, quantitative knowledge of the human physiology and pathology. Its primary goal is to simplify DTH models' development, validation, integration, and adoption.

The framework is built on a number of pillars discussed briefly below and further elaborated in this chapter of the roadmap (most of which is outside the scope of this deliverable).

Model pillar

Tools and guidelines to establish which specific model can be considered "actionable" within the VHT simulation platform, namely, to verify the compliance of the actual implementation of a model with respect to the data model(s) (data pillar) and in terms of portability and compatibility with available HPC, cloud and other compute resources (computation pillar). Tools will include cover types of models with different intended use in the context of personalised decision support systems, targeting various end users, and presenting specific technical and regulatory challenges.

Data pillar

Ensuring the connection with the VHT federated cloud-based repository and strategies for accessing various types of data and other resources, such as access to databases and access to continuous (real-time) monitoring devices. Addressing issues related to sensitive data storage and visualisation, the privacy of datasets for platform users and federated learning using metadata from wearable systems and clinical repositories.

Model integration towards VHT

Model development, integration of different levels/scales and organ systems. Integration of different types of data and integration of data and models. Integration of non-co-located resources. Orchestration of integration through data streams. Implementing knowledge-discovery processes on the platform allows for the continued development of individual digital twins and their integration towards the VHT.

Computation pillar

Access to the computing infrastructure (cloud, edge & HPC). Evaluation of HPC readiness of DTH applications and workflows, including aspects like scalability, portability, use of parallelisation, use of accelerator technologies (GPUs), networking requirements, data access, and application packaging (e.g., containerisation). Increase HPC readiness of DTH applications and development of new workflows, e.g. for integrating models or models and data, and the (automated) finding of compute resources.

Access pillar

Based on end-user needs & profiles, creation of a blueprint with technical specifications for data & model access, to be used in the simulation platform deployment. Inclusion of (but is not limitation to) implementation of authentication methods for user profiles to customise the DTH platform for each login. Creation of a personal data repository to pull personal data from bulk data sources (hospital databases, e-health, etc.) linked to the user profile (doctor, researcher, etc.) for running a VHT model,

and implementation of consent workflows for selected DTH applications requiring access to the predefined types of data from the personal data repository for creating each sensitive data-fed personal DTH.

The technical design of the platform architecture needs to include all the elements above, taking into account the platforms-state-of the art. The EDITH roadmap will provide guidelines for interoperability, quality assurance, and documentation of resources in the platform. It will also establish integration criteria for the inclusion and interoperability of tools and services into the platform. The further elaboration of this section (technology for the VHT) is outside the scope of this deliverable. However, given that the first section, the organisation of resources in a 6-dimensional framework is an important element that requires public input and discussion, it has been included in this deliverable.

5.1 Organisation of resources in a 6-dimensional framework

Within the VHT, the atomic entities are data objects and model objects that together define the location, content and conditions of use of the different resources in the platform. The concept of the 6-dimensional framework is currently under revision for publication in a peer-reviewed international journal and is available on ArXiv[7].

5.1.1 The data objects

Predicted data in the context of the VHT is defined as data obtained as the result of running an in silico model (be it a data-driven one, a knowledge-driven one or a combination of both). Each VHT data object is a digital dataset, stored and annotated according to some basic rules.

The dataset must contain quantitative information on human pathophysiology, whether measured or predicted. It must be stored and curated according to the FAIR principles to be findable, accessible (possibly through authentication and authorisation), interoperable and reusable. The dataset must be annotated with a minimum metadata set, including information on the data object type and its position in the data space. The Data Object Type (DOT) is a unique identifier associated with enough information to decide if and to what extent that data object is suitable input for a DTH model. This includes information on the dataset regarding its semantics (what the data mean), its syntax (in which standardised, interoperable formats the dataset is accessible), and its accessibility (how the dataset can be accessed). Eventually, DOTs will be selected from a list of standardised types, possibly organised in a well-structured taxonomy or ontology. But for some time, the list of supported DOTs might be a *folksonomy*, a user-generated way of organising content, which is periodically scrutinised and consolidated into proper ontologies.

In computer vision and robotics, the pose of an object is the combination of the object's position and orientation. Pose estimation determines a detected object's pose relative to some coordinate system. This information can then be used, for example, to allow a robot to manipulate an object or to avoid moving into the object. The Data Object Pose (DOP) includes all information to define the position of the data object in the VHT six-dimensional reference system and the scale information, such as the grain and range of the dataset.

Grain is defined as the larger of the minimum distance (or time span) that can be distinguished by the instrumentation or as the characteristic distance (or time span) of variation of the smallest (or fastest) feature of interest measured using this instrumentation. Extent is defined as the smaller of the maximum distance (or time span) that the same instrumentation can measure as the characteristic distance (or time span) of variation of the largest (or slowest) feature of interest measured using this instrumentation.

The six dimensions of the data space and represented in Figure 1 and defined in the following sections. The concept of grain and range as scale representation applies well to datasets that define the variation of a quantity in space and time. But since we assume by convention, as described above, that also scalar values are associated with a point in the 6D reference system of the VHT, in that case, the grain represents the least significant digit of the measurement/prediction (reproducibility of the measurement,

---

[7] (add reference in footnote when available)

uncertainty of the prediction). In contrast, the range could be used to represent the uncertainty of positioning in space and time for that scalar quantity.

Whenever a new DOT is added, it should also be provided with the transformation functions required to calculate the DOP for each data object with that DOT. See figure 1 for a schematic overview of the DOT/DOP and below for a more detailed description of the data space and its six dimensions.
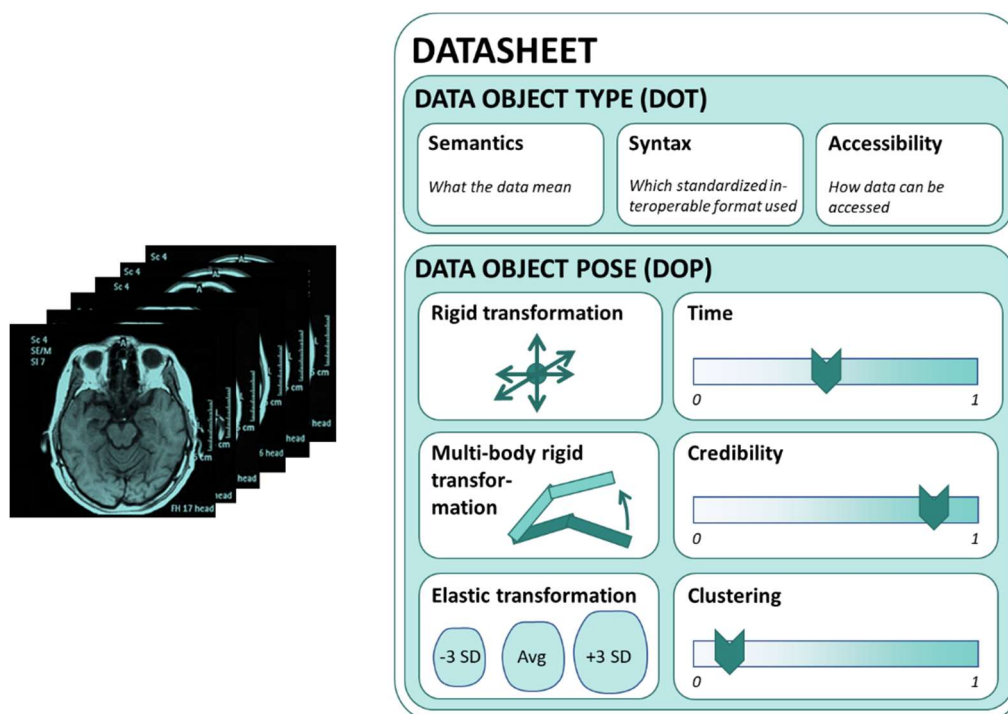

Figure 1: A graphic representation of the data sheet describing the relevant information of a data object.

5.1.2   *The model objects*

In VHT, model objects are defined as data space crawlers. A VHT model requires a finite number of inputs, described in terms of DOTs and DOPs, and produces, upon successful execution, a certain number of outputs, also described in terms of DOTs and DOPs. When a model is active, every time a new data object with the necessary DOT is added to the data space; the VHT model is automatically executed. Its outputs are also added to the data space in the appropriate DOP. This is why they are defined as data space crawlers: we can imagine model objects like little insects that crawl the honeycomb of data objects, "eat" some data objects from certain honeycomb cells and "lay" some new data objects in other cells. Thus, every time we add to the VHT a group of data objects that constitute a valid input for a model object, the dataspace will automatically enrich with new predicted data. This implies that VHT models must execute in batch mode; however, human interaction is still possible using the Mechanical Turk paradigm.

Two crucial technical aspects need to be addressed: remote execution and orchestration. In the simplest scenario, the VHT will run on a single computer cluster with some storage. All data objects are stored in this storage, and all model objects execute on the computer cluster. But as soon as we imagine more complex architectures, we might have a situation where the storage that contains the data objects, and the computer that executes the model objects, are not co-located. To ensure maximum flexibility, we can imagine a scenario where data objects and model objects are portable, the first using data replication services and the second using container architectures. This would allow the creation of a rule-based system that decides case by case if it is better to move the data or the models.

The second issue is model orchestration. As we mentioned above, many problems require the orchestration of multiple models, where the outputs of one model are the inputs of another. In many cases, model orchestration can be formulated exclusively in terms of data flow. Model A reads its inputs and calculates its outputs. Model B reads these outputs as inputs and calculates its outputs. And so on.

This type of orchestration can be provided as a by-product of the proposed data space crawler architecture. Essentially model B execute as soon as in the data space appear the outputs of model A, which model B consider as valid inputs. The complexity of the orchestration topology is not critical since the data flow can handle virtually any topology. The only exceptions are the so-called "strongly coupled" models. These are models where executing the next step in the calculation of model B requires the results of the current step of the calculation of model A. Strongly coupled models run simultaneously, exchanging data as they go through a shared file system or through the computer memory. These models can be orchestrated using specialised libraries. In this case, for the VHT, the whole orchestration would be seen as a single model object.

5.1.3   The six dimensions of the data space

We have currently identified six dimensions for the data space, although others might be added as the VHT is being developed. These are the three dimensions of the Body (body height, width, and depth), as well as Age, Credibility and Clustering. All this information combined for a given data set amounts to the DOP.

The VHT Anatomical Template

We use the hypothetical generation of the VHT Anatomical Template to illustrate how these six dimensions are defined. Let us imagine having a large collection of 3D body scans of humans of all ages, genders, etc. In theory, all scans were taken with the subject in the same pose (standing with the feet slightly apart, arms along the sides with the palms forward).

Each dataset is expressed with respect to an implicit reference system specific to the type of scanner used. We position the data object on the time axis in correspondence to the age in years of the subject at the time of the scan, assuming a time normalisation where 0 corresponds to birth and 1 is the longest human life ever lived.

Assuming the scans were all performed with fully certified 3D scanners, we place all datasets at value 1 on the credibility axis (which ranges from 0 for non-qualified data to 1 for fully certified measured data). Since each dataset refers to an individual, we place all of them at 0 on the clustering axis (the degree of clustering k is defined as $k = 1/x$, where x is the number of clusters: the homo sapiens sapiens cluster has $k = 1$; male-female clustering has $k = 0.5$; and individual datasets have $k = 0$, assuming an infinite number of human beings).

If we now select all datasets for individuals of a certain age, we can perform some spatial normalisations. The first normalisation operation assumes the body is a rigid object. We define an anatomical reference system (e.g., origin in the projection of the centre of mass on the floor, X oriented from posterior to anterior, Y from medial to lateral, and Z from feet to head) and calculate for each dataset the rigid transformation so that they are all aligned to the anatomical reference system. The second normalisation operation assumes the body is a kinematic chain, e.g., a set of rigid bodies articulated through idealised joints. We define in the anatomical reference system an ideal body posture. Then we calculate the multi-body rigid transformation for each dataset that aligns each scan to this ideal body posture. The third and last spatial normalisation assumes the body is an elastic object. We use statistical atlas techniques to calculate for each time point the average body shape and then calculate the transformation of each dataset to this average body shape. The vector of average body shapes at different ages is the VHT anatomical template. Each new VHT data object must be posed to this anatomical template.

Body

Data objects can be defined over 0, 1, 2, or 3 spatial dimensions. For example, the systolic blood pressure of a subject is a 0D data object; how the blood flow velocity varies along the length of an artery is a 1D data object; the distribution of temperature over a region of the skin is a 2D data object; the distribution of bone mineral density in a bone is a 3D data object.

Each data object (except 0D objects) represents the spatial variation of its values using an implicit reference system; so, in a 3D data object, the value corresponding to the coordinates (0, 0, 0) places such value at the origin of this implicit reference system. In addition, each data object is referred to a specific individual and their anthropometry. But to simplify the automatic annotation, the clustering,

and other similar operations, it is convenient that each data object is mapped to a conventional anatomical space by posing it with respect to the VHT anatomical template (which provides the DOP). 3D objects can be easily posed in the anatomical space; the bone mineral density distribution of a patient's femur can be posed in the VHT Space region corresponding to the femur of the anatomical template. 2D and 1D objects can also be posed with respect to the anatomical template, with some cautions. However, 0D objects do not have an anatomical location. But because all data objects in the VHT must have one, all 0D data objects are, by convention, mapped on a 3D point, which is located in a conventional point in the anatomical space. So, for example, the systolic blood pressure value could be posed at the centre of the heart region in the anatomical template or in the arm region where the sphygmomanometer was applied.

Two spatial transformation functions should be available for a given DOT: rigid roto-translation and elastic registration. The rigid roto-translation can consider the data object as a single rigid body or a kinematic chain depending on the data type. This transformation function is used to calculate the transformation required to align the data object to the VHT anatomical template and is stored in the DOP of the data object. For example, the elastic registration transformation function is used in particular cases when we need to compute the average geometry for a sub-population (which corresponds to a coordinate on the clustering axis).

Among the essential metadata for each DOT, one must include the spatial range and grain of the data object, which facilitates the definition of the spatial scale in which the data object is defined. It should be noted that this grounding of the data to the anatomy poses some challenges. A major one is how to handle datasets that refer to multiple anatomical locations, for example, the recordings of a multi-lead electrocardiogram. In such cases, one could position the dataset in correspondence with the heart centre or at the centre of the chest region. Or, if the anatomical location of each lead is available, one could decompose the dataset into multiple data objects, one for each channel, and place them at the anatomical location of their lead.

Age

The data object is positioned in the Age axis according to the age of the subject when the data were collected. For each DOT, the temporal transformation function will calculate a scaling factor that transforms the time and date when the data object was generated into a coordinate on the Time axis and store it in the DOP. Each time, a scaling factor vector is calculated to coordinate for time-varying data objects. Since the time of birth is not generally available information, we will assume that all subjects born on a given day were born at 12:00 (noon) because sixty per cent of babies are born during the day, between 6 A.M. and 6 P.M.

Among the essential metadata for each DOT, one must include the temporal range and the grain of the data object, which makes it possible to define the time scale in which the data object is defined.

Credibility

When a new data object is added, it is placed at the lowest level of credibility (non-qualified data). The data owner can submit a data object to the credibility transformation function. The higher the credibility of a data object, the higher its value. Depending on the level of credibility that the owner is requesting, the application must be informed by a smaller or greater amount of information that captures the provenance, the quality, the metrological properties (or computational credibility properties if the data are computed), and the certifications of the instrumentation/software used. For high levels of credibility, the request might be evaluated by a panel of experts, possibly in coordination with regulatory agencies.

Clustering

Each DOT must include among its transformation functions an averaging function that enables clustering. For data objects defined in space, this is typically an elastic registration function; for time-varying objects, it might involve a synchronisation function; for data objects not defined in space-time, these are more properly averaging functions in the statistical sense.

When added to the VHT, each data object is placed at clustering k = 0 (no clustering). To ensure irreversible anonymisation, the metadata includes a unique data object ID and a unique PatientID, which

is not associated with the individual identity. Where necessary, a LocalPatientID can be used to support pseudo-anonymisation schemes.

All data objects are automatically added to one default cluster: homo sapiens (k = 1). Specific research projects may calculate other sets, which are stored with enough metadata to inform the number of groups and the criteria used for clustering. This means that on the Clustering axis, there might be in the same coordinate multiple data objects for the same DOT type, each obtained with different clustering criteria. So, for example, under k = 0.5, we could have a male-female clustering but also a healthy-diseased clustering or a clustering over-below 55 years of age.

6   Next steps

After publication of this deliverable, it will be disseminated through the established EDITH communication channels (website, social media, newsletter). A form will be available on the project website to collect feedback (www.edith-csa.eu/materials). Several public discussion meetings will also be organised (May 2nd, June 1st) to discuss the content of the deliverable. In the coming months the remaining parts of the roadmap will be elaborated, reflecting the ongoing discussions within in the consortium and with the external advisory groups- and experts (industry, clinical etc.) and the wider community. The **first full draft of the roadmap will be published 31st of July 2023**.

Next steps (until summer)
- Internal working group meetings (Mapping, Vision, Repository/Platform, Sustainability), each meeting on a biweekly basis with dedicated agenda for each working group.
- Industry Advisory Board meeting every 2 weeks
- April (TBD): meeting with the Advisory Group of Stakeholders
- May 16-17 2023: EDITH deep thinkers meeting in Rome
- April 6 (3-4), May 2 (12-1), June 1 (12-1): online public meetings providing information on EDITH after publication of deliverable.
- June (TBD): meeting with EU large infrastructures
- July 31: submission of the first version of the roadmap

After **summer 2023, the public phase** of the project will start. This entails a number of on-site ecosystem meetings in Q4 2023 and 2024 and other public events. Additionally, detailed documentation will be provided on how to contribute use cases and resources into the repository. Communication on these initiatives and events will be done through social media, the EDITH website as well as newsletters and other communication channels from all partner institutions.