

# Sustainability breakout session

## Building an Ecosystem for Virtual Human Twins

Edwin Morley-Fletcher, Minos Garofalakis, Sébastien Bratières

17/05/2023



EDITH is a coordination and support action funded by the Digital Europe program of the European Commission under grant agreement n° 101083771.



# EDITH objectives

Ecosystem



Roadmap



Repository



Simulation  
Platform



# One focus of EDITH CSA is on the ecosystem

Borrowed from biology, in economics the term ecosystem generally refers to a group of interacting entities that depend on each other's activities.

It is an organisational structure, different from both hierarchies and markets, characterised by:

- the emergence of specific pre-conditions facilitating the interaction of a set of actors
  - having varying degrees of complementarity
  - achieving some alignment though operating with significant individual autonomy

## In EDITH's *Vision for the VHT and roadmap outline*, Section 4.2.4 is dedicated to sketching *The VHT Long-term sustainability*

- An evolutionary transition in three phases, from a pre-competitive setting to a mature market system
- All phases presuppose the initial realisation, within EDITH's infrastructure, of a system of **Distributed Ledger Technology** (DLT)
  - allowing to permanently trace all types of assets exchanged on the DLT
  - also tracking their provenance
  - securing the findability, accessibility, semantic interoperability, and reusability of all activated resources.

# What is a DLT?

- A DLT is a network of computing devices which jointly operate a replicated distributed database where each local replication stores a set of transactions between actors operating through the same DLT protocol, preventing double-registering of transactions and manipulation of balances.
- By creating this guaranteed digital space, a new layer of **automation** becomes possible, based on process automation through software programmes called **Smart Contracts**.
- This way, without recourse to any central authority, the DLT can deliver **Trust, Transfer** and **Settlement**, minimising at the same time transaction costs
- Given that the boundaries of organisations are determined by **transaction costs**, especially **search** and **information** as well as **bargaining** and **enforcement** costs, a DLT-based ecosystem has the great advantage of operating on qualitatively reliable data-rich information systems without being paralysed by excessive transaction costs.

# How to envisage an evolutionary ecosystem?

Three phases:

- **Honour ledger**: the DLT infrastructure will host exclusively pre-competitive transactions and work on incentives based on forms of reputational scoring
- **Token ledger**: 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
- **Money marketplace**: the ecosystem will mature and specialise: while some entities dealing mainly with pre-competitive transactions will continue to exist, a growing number of entities will increasingly focus on competitive transactions in the form of B2B exchanges, with prices set in Euros and no-more in tokens.

# Reputation (*Honour ledger*)

- The ecosystem operates as **a barter mechanism facilitated by a technological precondition**: the creation of a DLT infrastructure which is initially largely EU-funded
- The DLT allows to trace:
  - all accesses
  - all transactions
  - all reputational outcomes
- The DLT makes it easy to:
  - share a dataset or a model in ways compliant with the FAIR principles
  - translate such data or model sharing transactions into:
    - a citation-based reward system
    - a reputational scoring mechanism
- The metric of success is how much data and how many models are shared by the actors who will be induced to operate through the EU-funded DLT platform (as a necessary precondition for accessing further VHT research and innovation EU-funding)

# Beyond simple reputation barter (*Token ledger*)

- In a second phase, by introducing **digital tokens** issued by the entity governing the DLT infrastructure, it will be possible to make the incentives for what one can get in return for his/her shared resources much fine-grained and flexible.
- In its basics, tokens will be issued to whoever contributes resources to the ledger.
- 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 VHT ecosystem will use its growing token economy to experiment with how it can become progressively self-sustained.



# A symbolic currency becoming in due course convertible

- A token is a symbolic **computerized “currency” tied to specific purposes**, that can be exchanged for assets or services within a community of practice, allowing such an ecosystem to experiment allocating and tracking symbolic value exchanges among its actors.
- Initially, tokens will only have a value only within the VHT DLT system.
- In a further phase, once there has been a sufficient uptake of the tokens, it will also become possible to purchase them in exchange for money: this will apply mainly to external entities not having contributed resources to the ledger but wishing to use the VHT 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.

# Differences in governance models

- In the reputation phase, the governance may remain minimal:
  - day-by-day operations can be ensured by a small consortium paid to do so and to adopt basic technical decisions
  - any major decision on the infrastructure is taken with recourse to a direct democracy model, in which all contributors can participate in decision-making with an assembly process.
- In the token phase, the DLT infrastructure should substantially remain a largely public-funded resource, but its governance will require a more structured representative democracy.
- The main focus will be on the quality of service for its users. Systems must scale to extensive collections and handle distributed solutions relying on multiple hardware and software providers.
- When the development strategies of research groups and companies that have purchased tokens and are willing to use the DLT will become a key factor, the metric of success will be how important the VHT becomes in the development strategies of public and private developers.

# Ensuring long-term sustainability

- The VHT infrastructure will have to be run by a legal entity, possibly organised as a joint undertaking between the EC and major European industrial players, as already experimented in other similar cases.
- An EU-funded segment of the VHT should operate for the not-for-profit researchers, where most interoperability technicalities standards will be tested and standardised.
- Other segments, certified for interoperability by the legal entity, will be fully commercial and operating on a private basis, pursuing sustainable business-to-business models.

# More complex and broader valuation mechanisms

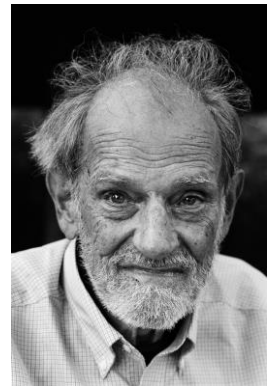
- In a more advanced stage of development, the VHT DLT infrastructure will possibly also engage in analysing how incentives linked to automated assignment and distribution of value and quality validation can be determined by ML mechanisms valuing different attributes or even through Shapley-value mathematical methods determining the only distribution satisfying a collection of properties within a coalition game.

# Towards a sustainable DTH ecosystem: Resource Valuation

- Economic incentives ensure principled, fair valuation of resources, and allocation of value to participants
- “Valuation” is not necessarily tied to economics, e.g., it may provide means to measure the *quality* of a resource
- Critical in a vibrant ecosystem with a multitude of resources addressing the same/similar needs
  - e.g., quality-based ranking can quickly identify “best” resources for a given task

# A principled data valuation tool: Shapley value

- Several ad-hoc data valuation mechanisms
  - Query-based (on-demand/flat pricing), auction-based (using bids to price), reputation-based (user feedback), ...
- **Shapley value:** ML as a coalitional game [*Lloyd Shapley'1953*]
  - Data sources seen as players in a coalition
  - Usefulness of data characterized via ML utility function
  - Distribute value generated by coalition based on sources' marginal usefulness
  - ONLY distribution satisfying some key properties



# Key Shapley value properties

- **(Task-dependence)** Data has a value only when used and the value of data depends on the task it helps to fulfill
- **(Cumulativity)** The value of data accumulates when used for multiple times
- **(Inter-dependence)** The utility of one data source hinges on others when jointly used for solving a task
- **(Fairness)** Data's value should fairly reflect the utility of different data sources

# Shapley value: Math definition and properties

- The Shapley value of  $i$ th player is  
$$s_i = \frac{1}{N} \sum_{S \subseteq I \setminus \{i\}} \frac{1}{\binom{N-1}{|S|}} [U(S \cup \{i\}) - U(S)]$$

$S$  is a subset of all players excluding  $i$        $I = \{1, \dots, N\}$  is the set of all players

Marginal contribution of player  $i$       Utility function

- Properties

- Group Rationality:**  $\sum_{i=1}^N s_i = U(I)$
- Fairness:**
  - (1) If  $U(S \cup i) = U(S \cup j)$  for all  $S \subseteq I$ , then  $s_i = s_j$
  - (2) If  $U(S \cup i) = U(S)$  for all  $S \subseteq I$ , then  $s_i = 0$
- Cumulativeness:**  $s_i^{U+V} = s_i^U + s_i^V$



# Shapley value: Computational challenges

- The Shapley value of  $i$ th player is

$$s_i = \frac{1}{N} \sum_{S \subseteq I \setminus \{i\}} \frac{1}{\binom{N-1}{|S|}} [U(S \cup \{i\}) - U(S)]$$

$\downarrow$   
 $|S| = 2^{N-1}$

- Complexity of computing the exact value could require exponential time
- In the ML context, evaluating utility itself is computationally expensive

**"Towards Efficient Data Valuation Based on the Shapley value."** Jia\*, Dao\*, Wang, Hubis, Gurel, Li, Zhang, Spanos, Song. AISTATS 2019  
**"Efficient Data Valuation for Nearest Neighbor Algorithms."** Jia, Dao, Wang, Hubis, Gurel, Hynes, Li, Zhang, Spanos, Song. VLDB 2019.  
**"An Empirical and Comparative Analysis of Data Valuation with Scalable Algorithms."** Jia, Sun\*, Xu\*, Zhang, Li, Song. arXiv:1911.07128

# DTH model valuation



- Large number of model validation/evaluation metrics
  - Speed, memory, accuracy (training/generalization), cross-validation metrics, complexity, explainability, composability, ...
- For relative model valuation, *standard DTH model benchmarks* will need to be developed by the community
  - Similar to need for community benchmarks in ML, DBs, computer architecture, ...
- Valuation measures can also draw on other factors, e.g., *model provenance*
  - Another key use for DLT: tracking model and data provenance

Join at  
**slido.com**  
**#2309 239**



# IPR & licensing

# We have open-source licenses + selection help

## Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

{ Which of the following best describes your situation? }



**I need to work in a community.**

Use the [license preferred by the community](#) you're contributing to or depending on. Your project will fit right in.

If you have a dependency that doesn't have a license, ask its maintainers to [add a license](#).



**I want it simple and permissive.**

The [MIT License](#) is short and to the point. It lets people do almost anything they want with your project, like making and distributing closed source versions.

[Babel](#), [.NET](#), and [Rails](#) use the MIT License.



**I care about sharing improvements.**

The [GNU GPLv3](#) also lets people do almost anything they want with your project, *except* distributing closed source versions.

[Ansible](#), [Bash](#), and [GIMP](#) use the GNU GPLv3.

{ What if none of these work for me? }

**My project isn't software.**

[There are licenses for that.](#)

**I want more choices.**


[More licenses are available.](#)

**I don't want to choose a license.**

[Here's what happens if you don't.](#)



Ecosystem for Digital  
Twins in Healthcare



(Re)usable Data Project


[Learn more](#)
[Evaluation](#)
[Explore](#)
[About](#)

# The (Re)usable Data Project

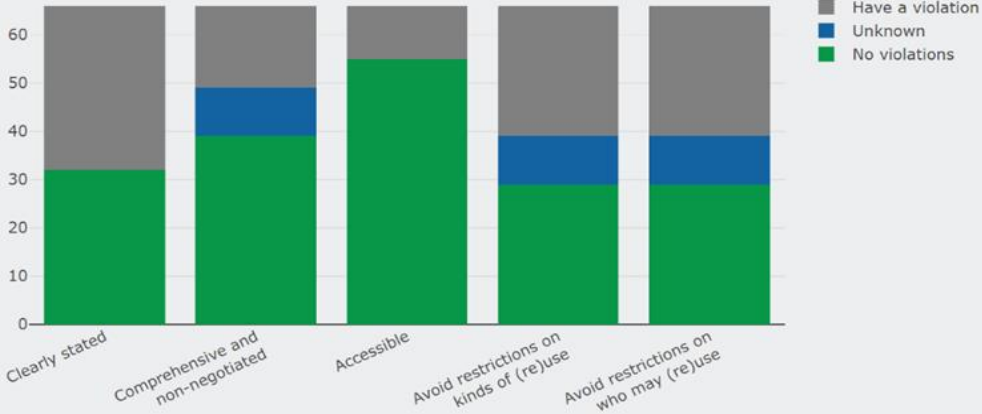
Inspired by the efforts of scientists around the world and the game-changing efforts of projects like the Creative Commons, the Wikipedia Foundation, and the Free Software movement, we hope to engage the larger community in an open and fruitful discussion on issues concerning the use and reuse of scientific data, including the balance of openness and how to make ends meet in an increasingly competitive environment.

If you would like to join our efforts to highlight the use and reuse of data in the sciences, please feel free to contact us on our [tracker](#), create a pull request against our [repository](#), or join our [forum](#).





### High-level summary of curated data resources



Category	No violations (Green)	Unknown (Blue)	Have a violation (Grey)
Clearly stated	32	0	28
Comprehensive and non-negotiated	39	10	11
Accessible	55	0	15
Avoid restrictions on kinds of (re)use	29	10	21
Avoid restrictions on who may (re)use	29	10	21

Learn more »

See our data »

Join us »

# Our sources data

You may also explore our data with simple visualizations [here](#).

Show  entries

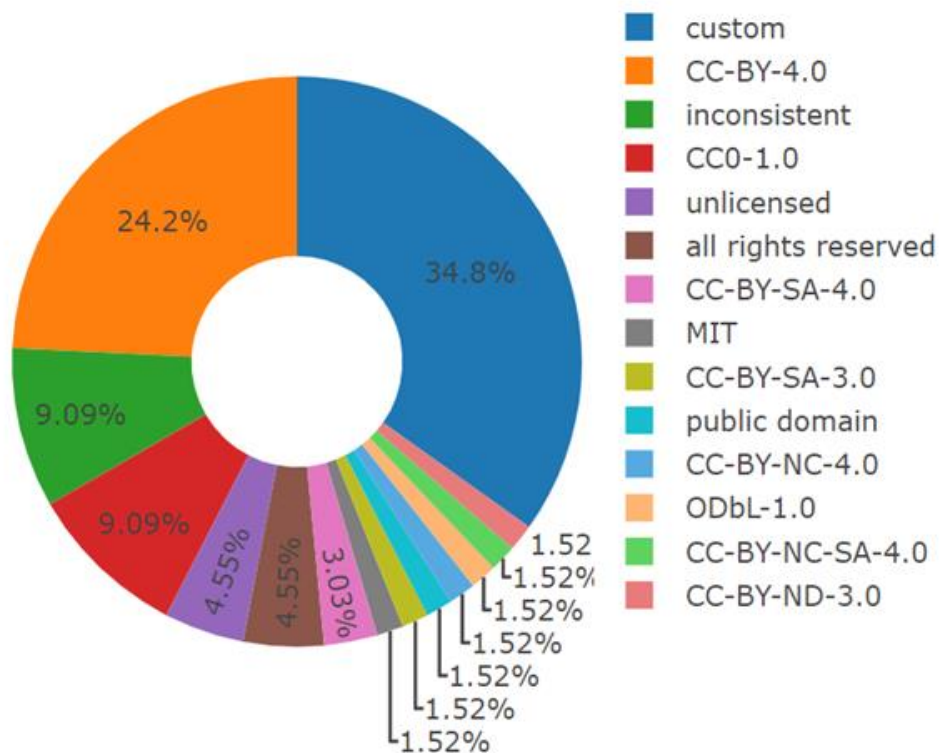
Search:

Name	Tags	Grade	Description	License Info	License Issues
<a href="#">Alliance of Genome Resources (AGR)</a>	biology, MOD, functional annotation, disease-gene association, orthology, phenotype and disease models	★★★★★	The primary mission of the Alliance of Genome Resources (the Alliance) is to develop and maintain sustainable genome information resources that facilitate the use of diverse model organisms in understanding the genetic and genomic basis of human biology, health and disease.	permissive	<a href="#">Commentary</a>
<a href="#">ArrayExpress</a>	biology, microarray experiments, functional genomics, high-throughput, microarray, sequencing	★★★★½	ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.	permissive	<b>Criteria A.2.2</b> Minimal custom permissive terms. <a href="#">Commentary</a>
<a href="#">Bgee</a>	biomedical, x-species, expression data	★	Bgee is a database to retrieve and compare gene expression patterns in multiple animal species, produced from multiple data types (RNA-Seq, Affymetrix, in situ hybridization, and EST data).	unknown	<b>Criteria A.1.1</b> While the CC0 tool is explicitly being invoked in main areas, there is some inconsistency, i.e. different ontology licensing and downloads that obviously encompass non-CC0 data. <a href="#">Commentary</a>
<a href="#">BioCyc Database Collection (BioCyc, public)</a>	biology, genomic resource, sequence, gene structure, pathways, reactions, functional annotation	★★★	BioCyc is a collection of 20,028 Pathway/Genome Databases (PGDBs) for model eukaryotes and for thousands of microbes, plus software tools for exploring them. BioCyc is an encyclopedic reference that contains curated data from 130,000 publications.	restrictive	<b>Criteria A.2.2</b> Non-standard/custom license. <b>Criteria B.1</b> One term of the license is that you must "Notify SRI that you are making BIOCYC DATABASES available in this manner"; this, combined with somewhat bulky access (see comments), I believe rises to a barrier to reuse as a

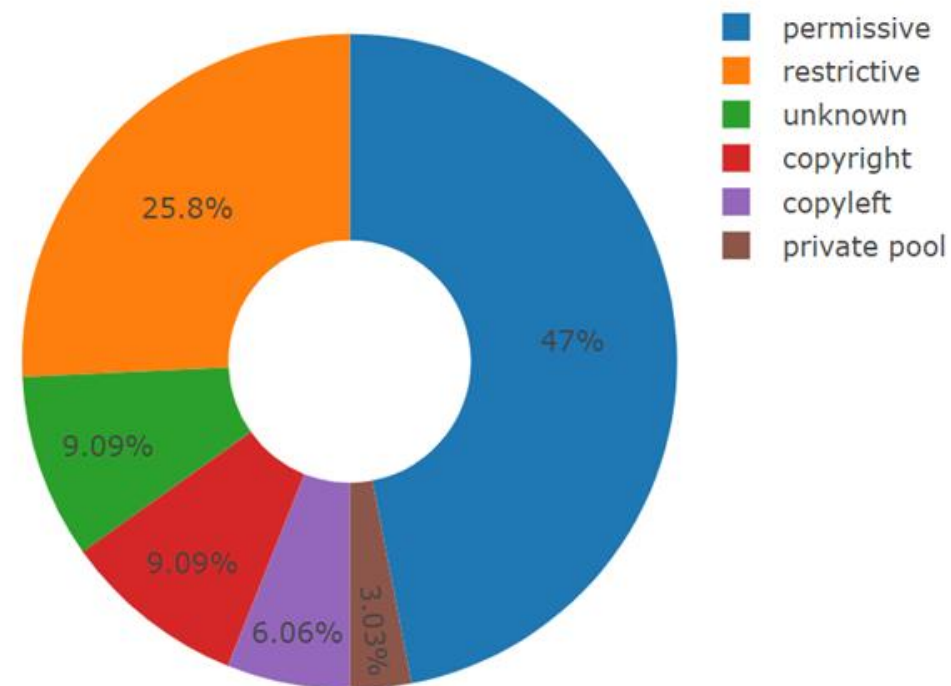




Licenses used



Overall license reuse categories





# How to tackle IPR & licensing issues?

**The problem: How to share, buy, sell, re-use? Need legal certainty.**

- assets
  - models
  - data
  - model/data adaptors
  - computation service
  - combinations and pipelines
- interfaces
  - patient to research team (consent/reconsent/dynamic consent)
  - university to spin-off
  - startup to distributors, to integrators, to healthcare providers
  - industry access to training/model selection/evaluation data/specific VHT services

## **Questions for today**

- typical needs?
- what assets/licences/types of protection are used today?
- shortcomings? are current IPR tools adequate?
- does this community have recommendations for extensions of IPR laws and regulations?
- should EDITH produce specific guidelines/templates for certain usages?

*What is your experience? Where are blockers? We value your first-hand experience on these issues.*

Active poll

6



Join at  
**slido.com**  
**#2309 239**

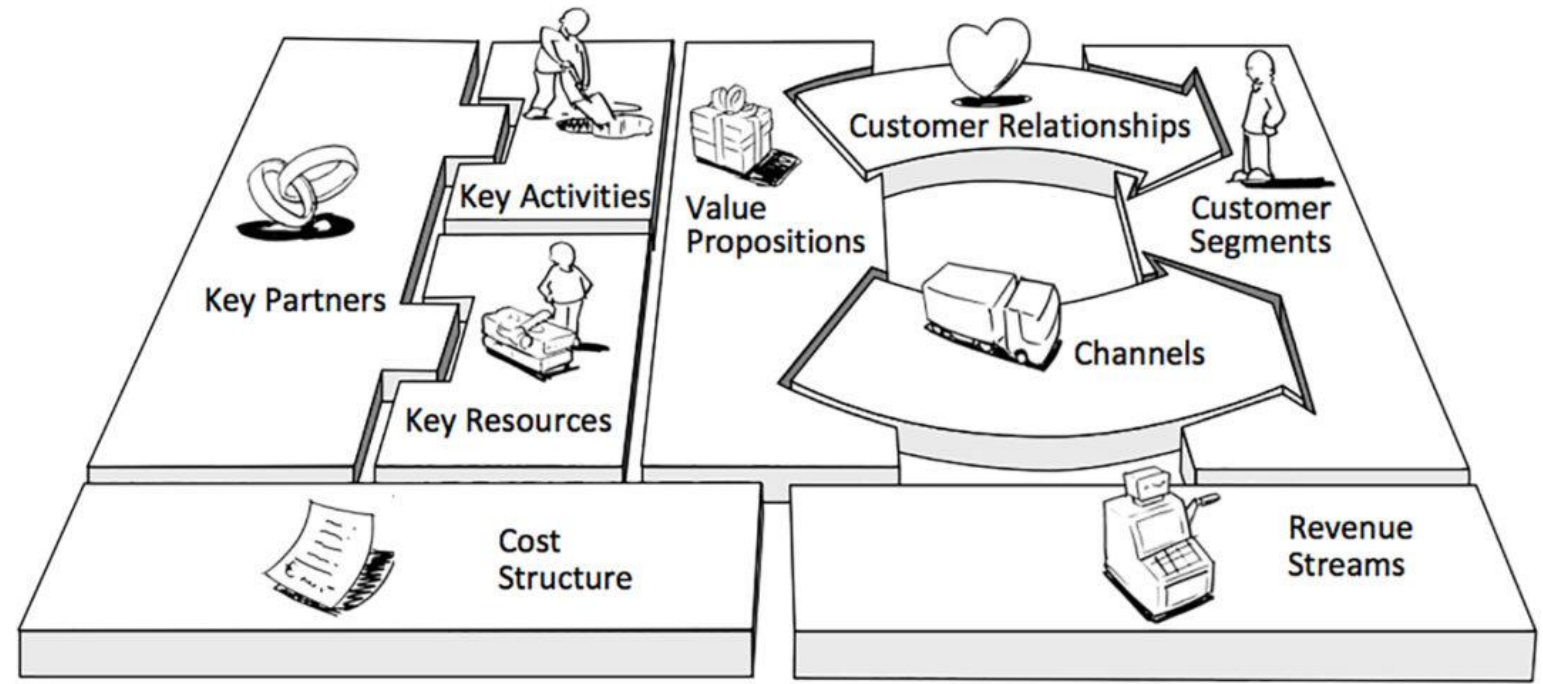
## Typical current needs for IPR solutions?

- Anonymous**  
Following open science mandates and conceptualizing/articulating them in the license terminology.
- Anonymous**  
IPR are often regulated at national level, so we will need to take into account this level of complexity, maybe including exemption for prevailing local laws
- Anonymous**  
Legal capacity and understanding
- Anonymous**  
Web need yo properly understand the limits of a particular license
- Anonymous**  
Ensure a "proper" use of the assets I share
- Anonymous**  
Licensing technology developed at universities

Join at  
**slido.com**  
**#2309 239**



# Business Model Canvas (Osterwalder & Pigneur 2010)



Adapted from 'Business Model Generation', Alexander Osterwalder, Wiley 2012.  
[www.businessmodelgeneration.com](http://www.businessmodelgeneration.com)  
Licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

# Business modelling and Bizmod4VHT

**The problem: With very few computational clinical models in the market, how to predict what business models can function?**

- what will be the future driver of an emerging digital twin model market?
  - digital twin models sold to clinics/hospitals/diagnostics centers bundled with diagnostics/therapy equipment: medical imaging hardware, radiotherapy,
  - models sold directly to patients together with wearable sensors or consumer-grade health and fitness equipment: smart scale, exercise equipment, kitchen scale, fitness bands
  - standalone software built around a single suite of models, sold as package to healthcare providers, with a unique user interface; e.g. by EHR management software makers
  - mix-and-match model libraries which can run on several platforms
- what are the largest roadblocks for the development of a digital twin model market?
  - performance, clinical accuracy
  - patient trust
  - too few models
  - ...

Join at  
**slido.com**  
**#2309 239**



# Returning to the evolutionary three-phases hypothesis



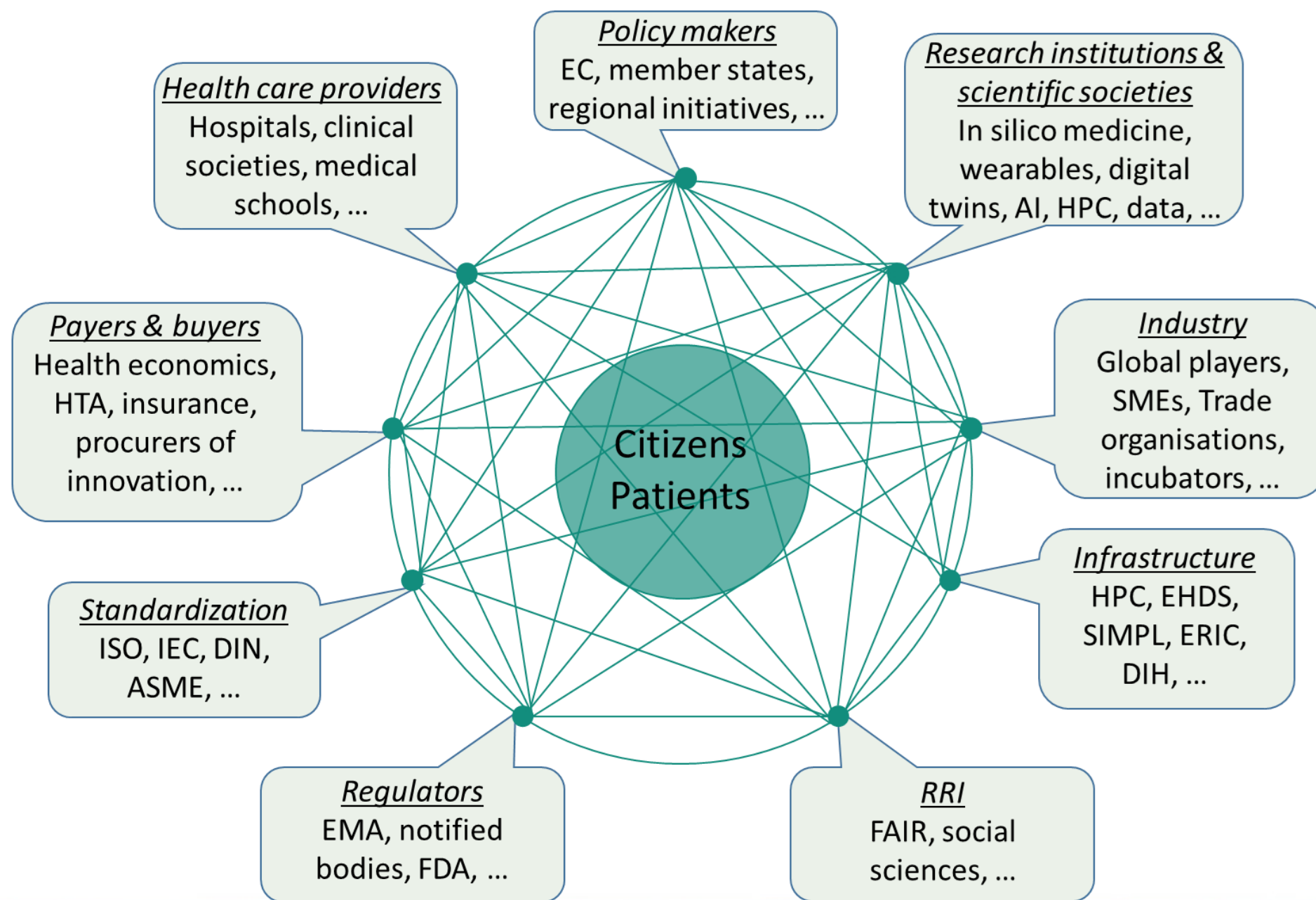
# Money marketplace

- Eventually, the number of transactions triggered by research groups and commercial companies having monetised their access to the DLT-operated ecosystem will start to attract 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 will 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.



# Centrality of the evolutionary ecosystem

- Our thought until now has been led by the persuasion that the VHT can only be realised by establishing, engaging, and keeping engaged a vibrant ecosystem.
- Believing that EDITH must expand the involved stakeholders and build bridges with the molecular dynamics, bioinformatics, and systems biology communities, we have been trying to figure out how key elements of the VHT ecosystem can be synergised by aligning:
  - People: the VHT Community of Practice
  - Assets: the VHT Infrastructure
  - Rules: technical standards and standard operating procedures
  - Motivations: long term sustainability



# Is there the need of a partial reorientation?

- Should the announcement that the European Commission is launching a **preliminary market consultation**, in the context of the upcoming procurement delivering a platform for advanced virtual human twin models integration and validation under the **DIGITAL Work programme 2023-24** significantly modify the perception of what are the priorities to take into account?
- An indicative global budget of **20 million €** for an indicative duration of **24-36 months** will be available for procuring a **Platform for advanced virtual human twin (VHT) models**
- The platform will comprise of the following components:
  - the **federated repository** of data and VHT model assets
  - an open-source reference implementation access **federator for building and visualising simulations**
  - an **orchestrator of computational services** for running these simulations on shared resources.

# More detailed features

- The main objective is to develop a distributed platform making available to users:
  - a federated repository of VHTs related resources
  - a combined set of open source software toolkits, deployed and hosted by each user organisation
  - access to computational services, enabling them to develop, test and integrate VHT models.
- The platform will be used by researchers, developers, engineers, practitioners, innovators in the health and care domain, including for professional training and educational purposes.
- It will provide controlled and secure access to an environment of simulation and visualisation tools open access and proprietary data and model assets for advanced modelling.
- The platform will be used for developing, testing and integrating existing and new VHT models, based on reference datasets, other research outputs and user resources, and will be fully interoperable with augmented and virtual reality environments.
- Its use will be based on access to computational services enabled by strategic digital capabilities (e.g. HPC, cloud, edge computing, AI), with links to suitable testing and experimentation facilities and other resources becoming available in the context of the European Health Data Space.

# Do these goals imply for EDITH a change of approach?

- Do an evolutionary ecosystem and correlative incentives seem to be no more directly in focus?
  - Still, by referring to a federated repository, a federator for building and visualising simulations, and an orchestrator of computational services, there is indirectly the assumption of having to deal with a varied ecosystem.
  - For sure, there should be room for a DLT-based platform.
  - What about incentives?
  - What about a «gaming» sandbox?
  - What about experimentations in value allocation without having to cope with straightforward monetisation?

# Further thoughts...

# Thank you

<http://www.edith-csa.eu>



EDITH is a coordination and support action funded by the Digital Europe program of the European Commission under grant agreement n° 101083771.

