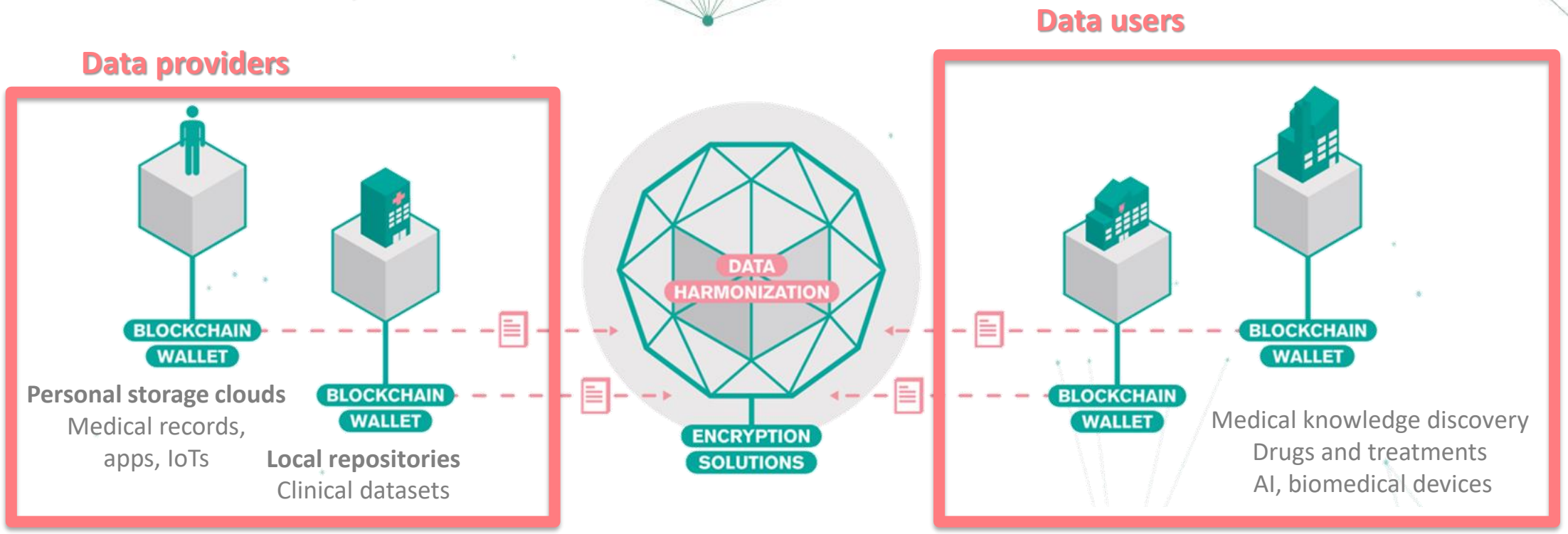
MY HEALTH
MY DATA

“MY HEALTH, MY DATA”



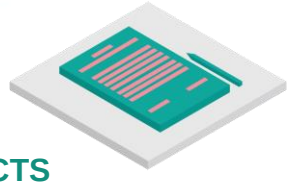
MyHealthMyData (MHMD) is a H2020 Research and Innovation Action (2016-2019) developing a **blockchain platform for sharing and exchanging HEALTHCARE DATA** for medical care, research and innovation

OBJECTIVES and INNOVATIONS

CITIZENS' EMPOWERMENT

DATA PRIVACY AND SECURITY

DATA VALUE ENHANCEMENT

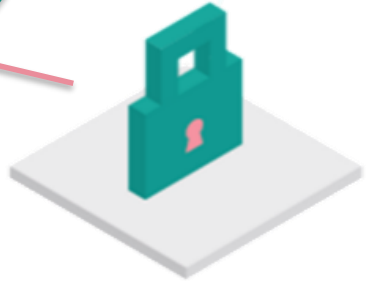


BLOCKCHAIN and SMART CONTRACTS

A private/permissioned blockchain architecture that **manages and authorizes the access and exchange of data** according to user-defined conditions

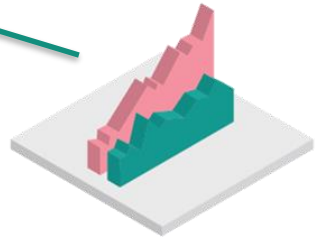
PERSONAL DATA ACCOUNTS and DYNAMIC CONSENT

Personal storage clouds to **aggregate personal data from disparate sources** (medical records, mobile apps, IoTs), access it for personal use or share it under conditions defined by a dynamic consent module implemented through a dedicated smart contract



DATA PRIVACY AND SECURITY TECHNOLOGIES

- PRIVACY-PRESERVING DATA PUBLISHING (Data anonymisation): exposing de-identified data
- SYNTHETIC DATA GENERATION: generation of synthetic datasets through machine learning
- SECURE COMPUTATION (Privacy-preserving data flow exécution): calculating algorithms on encrypted data



BIG DATA ANALYTICS

Exploring the feasibility on **de-identified** and **encrypted data** of

1. **DeepExplorer/DeepReasoner:** deep learning for AI configuration, data exploration, case-based reasoning, patient stratification
2. **Personalized physiological models for clinical decision support** (blood circulation model)
3. **Machine learning algorithms for knowledge discovery**
4. **Models for estimating the value of data**

DATA PRIVACY AND SECURITY TECHNOLOGIES

PRIVACY-PRESERVING DATA PUBLISHING *(Data anonymisation)*

Expose **de-identified data**

- **removing direct *identifiers***
  (i.e., name and surname, social security number)
- **transforming *quasi identifiers*** through generalisation (aggregation)
  (e.g., age, date/place of birth, zipcode)
- **randomisation** (statistical noise addiction, permutation)

k-anonymisation (generalisation)

ID	AGE	ZIPCODE	DIAGNOSIS
1	28	13053	Heart Disease
2	29	13068	Heart Disease
3	21	13068	Viral Infection
4	23	13053	Viral Infection
5	50	14853	Cancer
6	55	14853	Heart Disease
7	47	14850	Viral Infection
8	49	14850	Viral Infection
9	31	13053	Cancer
10	37	13053	Cancer
11	36	13222	Cancer
12	35	13068	Cancer

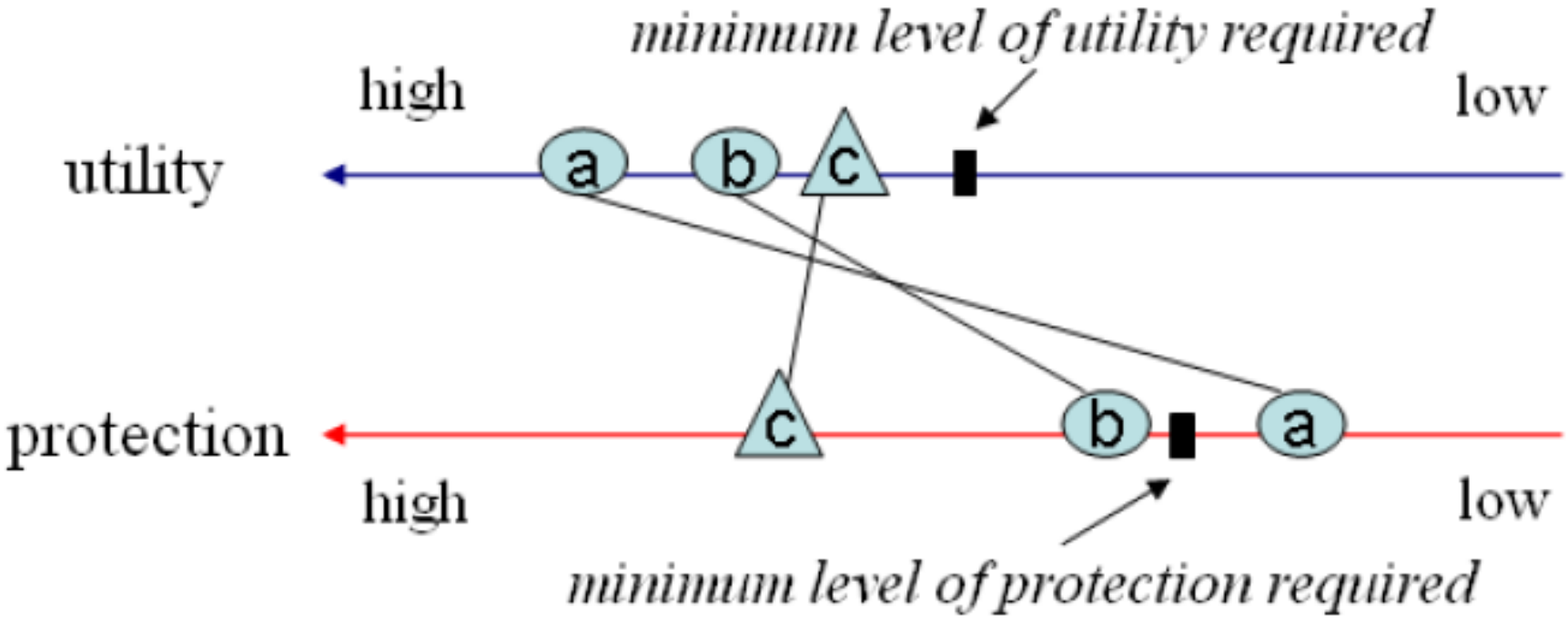
ID	AGE	ZIPCODE	DIAGNOSIS
1	[20-30]	130**	Heart Disease
2	[20-30]	130**	Heart Disease
3	[20-30]	130**	Viral Infection
4	[20-30]	130**	Viral Infection
5	[40-60]	148**	Cancer
6	[40-60]	148**	Heart Disease
7	[40-60]	148**	Viral Infection
8	[40-60]	148**	Viral Infection
9	[30-40]	13***	Cancer
10	[30-40]	13***	Cancer
11	[30-40]	13***	Cancer
12	[30-40]	13***	Cancer

ATHENA Research & Innovation Information Technologies

amnesia

OpenAIRE

The trade-off between privacy protection and data utility

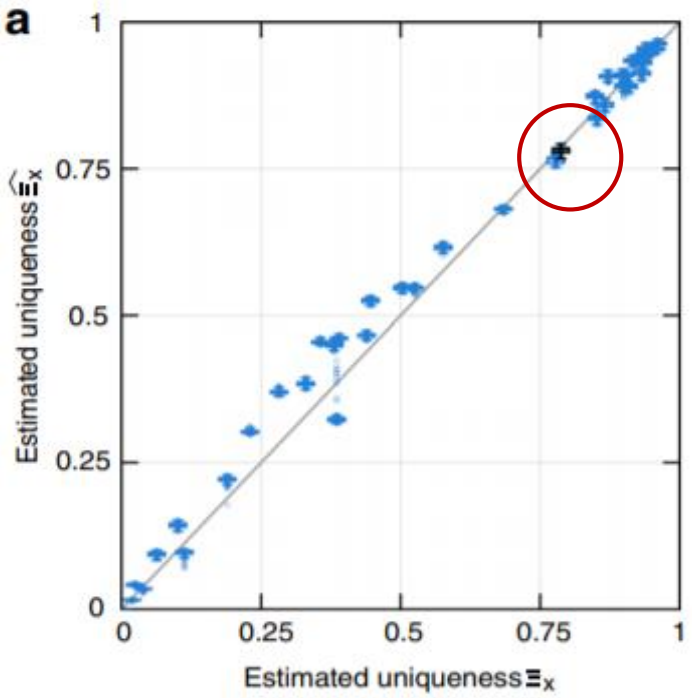


Loukides, G., & Shao, J. (2008, March). **Data utility and privacy protection trade-off in k-anonymisation.** In *Proceedings of the 2008 international workshop on Privacy and anonymity in information society* (pp. 36-45). ACM.

The problem of re-identification

- **99.98% of US citizens** correctly identified in 'anonymised' datasets with just **15 demographic attributes including** age, gender, and marital status

- *Example test:* through **date of birth, location (PUMA code), marital status and gender** the model uniquely identifies **78.2 ± 0.5% of the 3 million people in this population**



ARTICLE

https://doi.org/10.1038/s41467-019-10933-3 OPEN

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher [1,2,3], Julien M. Hendrickx[1] & Yves-Alexandre de Montjoye[2,3]

23 July 2019

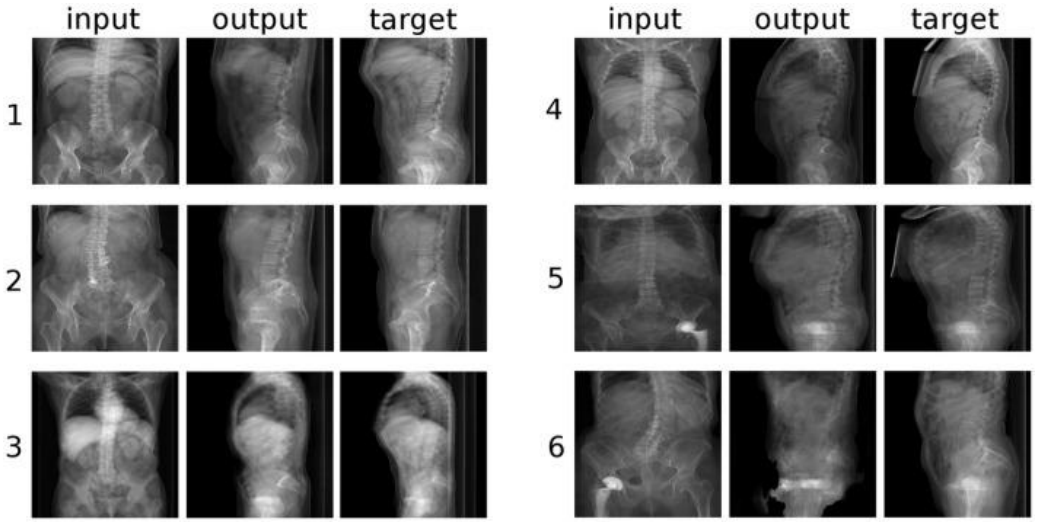
DATA PRIVACY AND SECURITY TECHNOLOGIES

SYNTHETIC DATA GENERATION

«Machine learning enabling machine learning»

Generation of **synthetic datasets** through machine learning algorithms that recreate pre-defined characteristics of a target population for one or more clinical modalities

- ***Synthetic structured clinical data (cardiac function)***
- ***Cardiovascular magnetic resonance images (MRI)***



Galbusera, F., Niemeyer, F., Seyfried, M., Bassani, T., Casaroli, G., Kienle, A., & Wilke, H. J. (2018). **Exploring the Potential of Generative Adversarial Networks for Synthesizing Radiological Images of the Spine to be Used in *In Silico* Trials**. *Frontiers in bioengineering and biotechnology*, *6*, 53.

- **Scalable, GDPR-compliant**
- **Complementary**: cannot replace patient data but can be a useful adjunct to accelerate testing and algorithm development;
- **Risks of attribute disclosure** still present

Queen Mary University of London

SBA Research

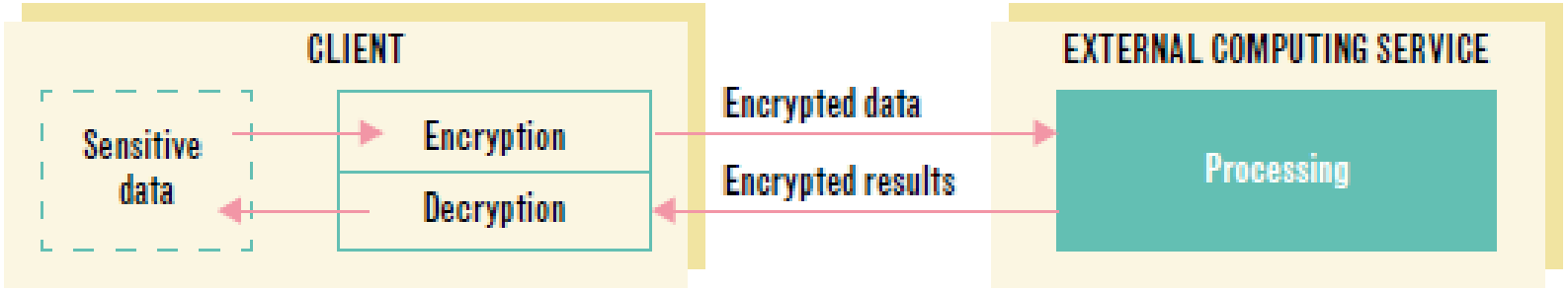
DATA PRIVACY AND SECURITY TECHNOLOGIES

SECURE COMPUTATION *(Privacy-preserving data flow execution)*

Calculate algorithms on encrypted data – revealing RESULTS only

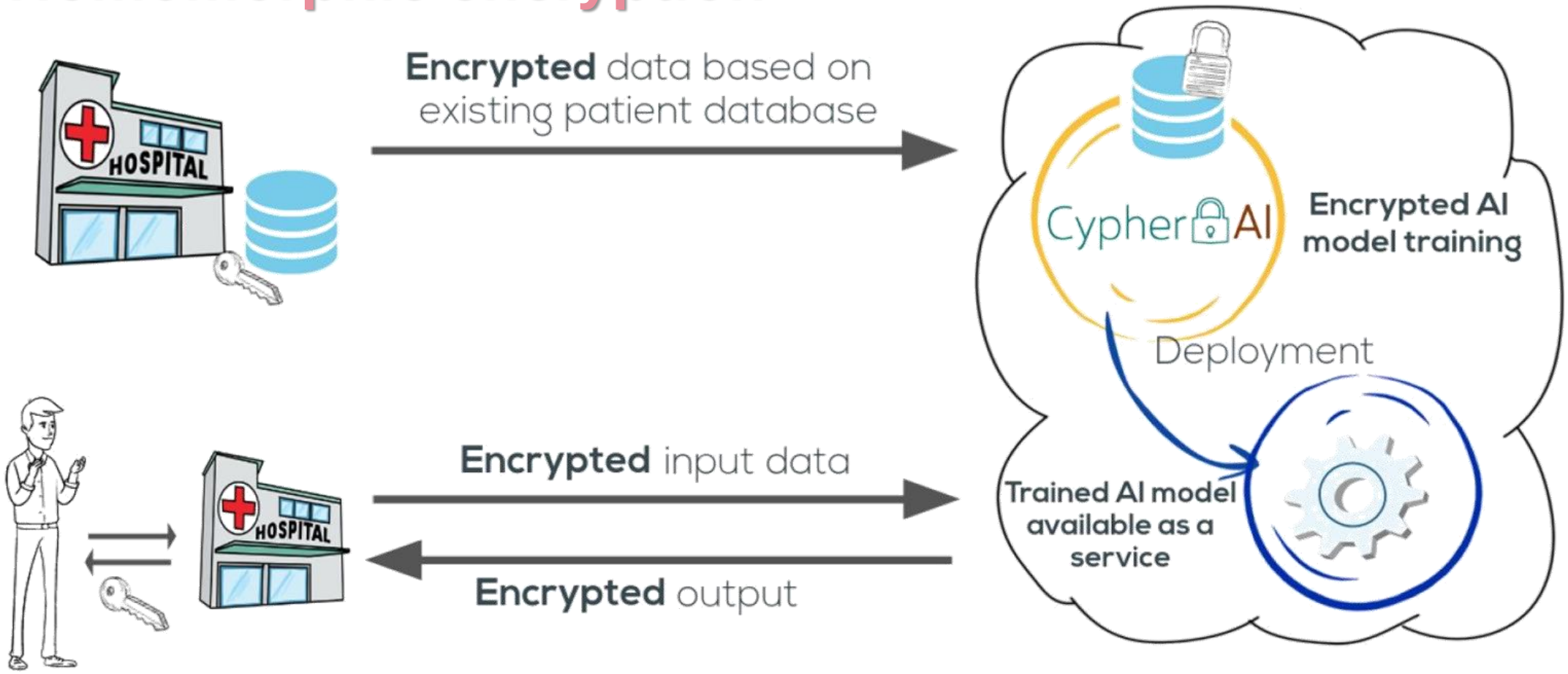
- **Secure multi-party computation:** parties jointly compute a function on entries in a distributed manner, keeping the inputs private
- **Homomorphic encryption:** data is encrypted before being sent to the computing service, and calculations are made on encrypted data

'Industrial & Enabling Tech' prize category
(2018 and 2019 finalist)



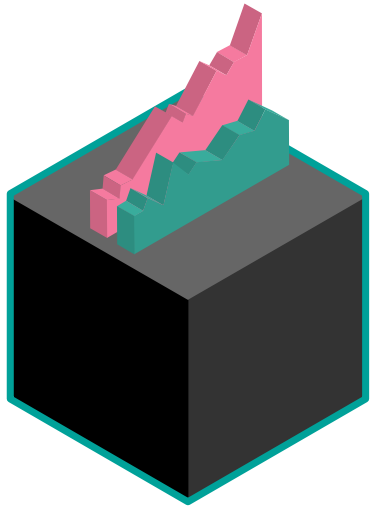
DATA PRIVACY AND SECURITY TECHNOLOGIES

Homomorphic encryption

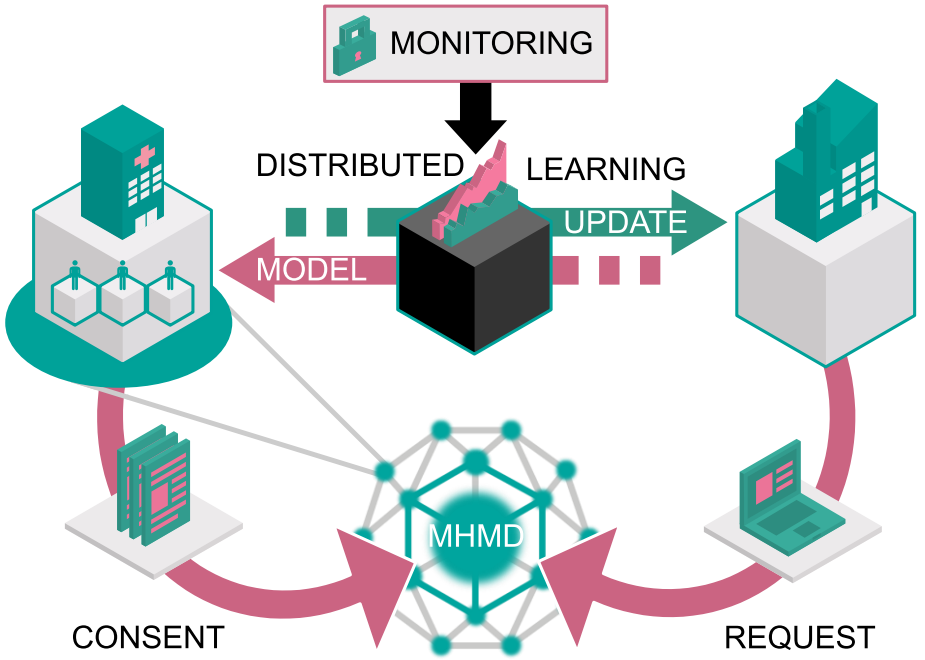


DATA PRIVACY AND SECURITY TECHNOLOGIES

SECURE COMPUTATION *(Privacy-preserving data flow execution)*



Distributed deep learning:
federated and split machine learning where the training is distributed among trusted data providers with “differential privacy” guarantees



ATHENA Research & Innovation Information Technologies

SIEMENS Healthineers

For more info:

Contact

info@myhealthmydata.eu

Website

http://www.myhealthmydata.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732907