



Application of Artificial Intelligence & Machine Learning for Precision Medicine

February 17, 2023

Federated Learning at Owkin

↳ Joseph Lehár, SVP Strategy, Owkin



The volume of health data is surging



EMR, electronic medical record

1. EMC with Research & Analysis by IDC. The Digital Universe Driving Data Growth in Healthcare. [Link](#) (Accessed 3 December 2019); 2. IDC White Paper: The Digitization of the World - From Edge to Core. [Link](#) (Accessed 3 December 2019); 3. IDC White Paper: Healthcare: DATCON Level 3. [Link](#) (Accessed 28 February 2020).

Veeva ID:M-XX-00003260
Date of prep: October 2020

Image from Presentation by Asif Jan - F Hoffmann La Roche AG



Decentralized production

No natural monopoly for data production in healthcare
(different from mobile data)

Many actors generate data:

- Hospitals
- Companies
- Devices

Medical data are regulated

Data are **personal and sensitive** information. They fall under the EU **GDPR**:

- Need legal basis / consent from patient
- Only specific purpose / no reuse
- Patient opt-out

Medical data are valuable

Challenges

- **Cost** for data collection is high.
- **Perceived value** is high.
- Existing market with high **price** of data

Competitive ecosystem

- Academics
- Startups
- Pharmas



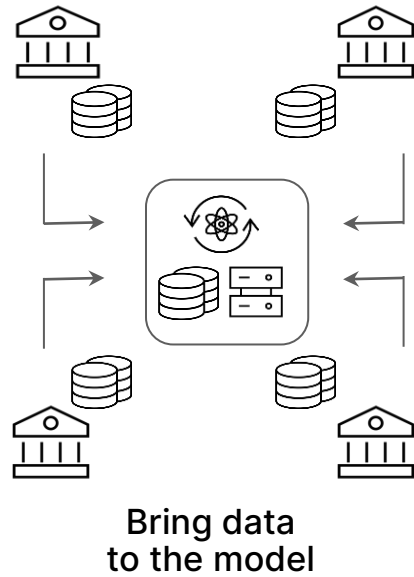
Decentralized production

Medical data are regulated

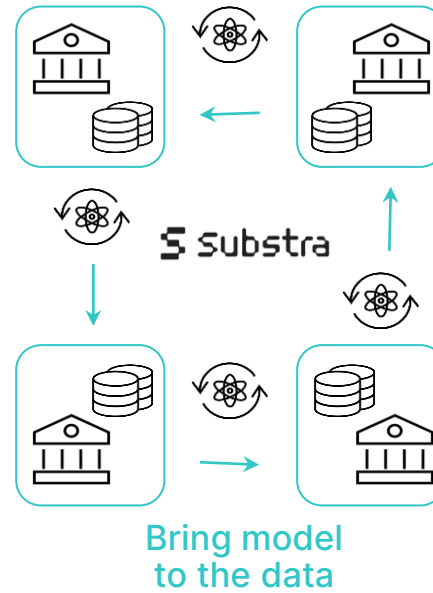
Medical data are valuable



**Collaboration between data managers is essential,
but privacy and security are paramount**

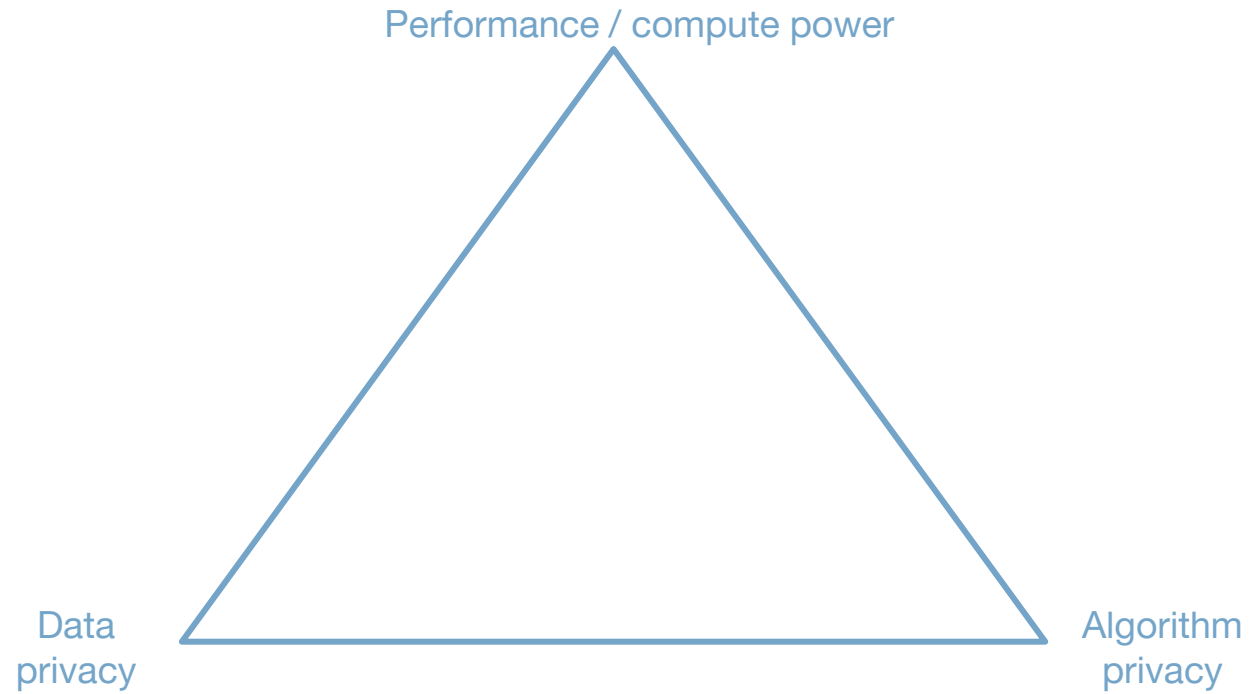
Aggregation

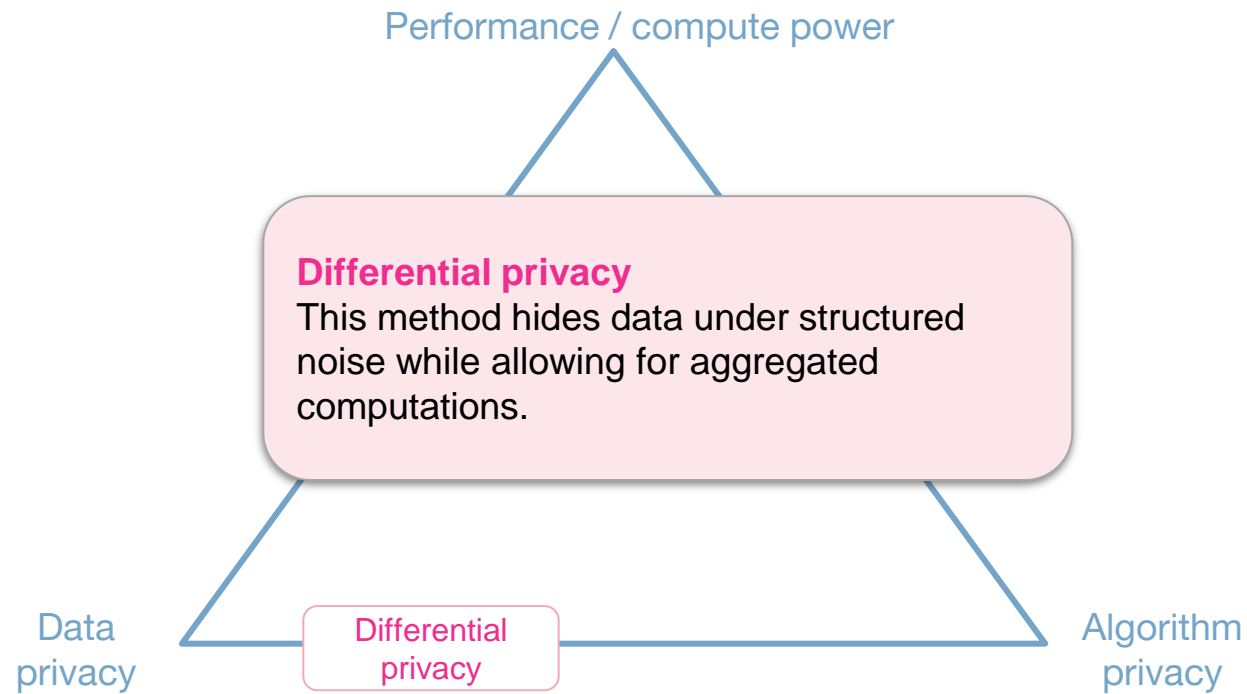


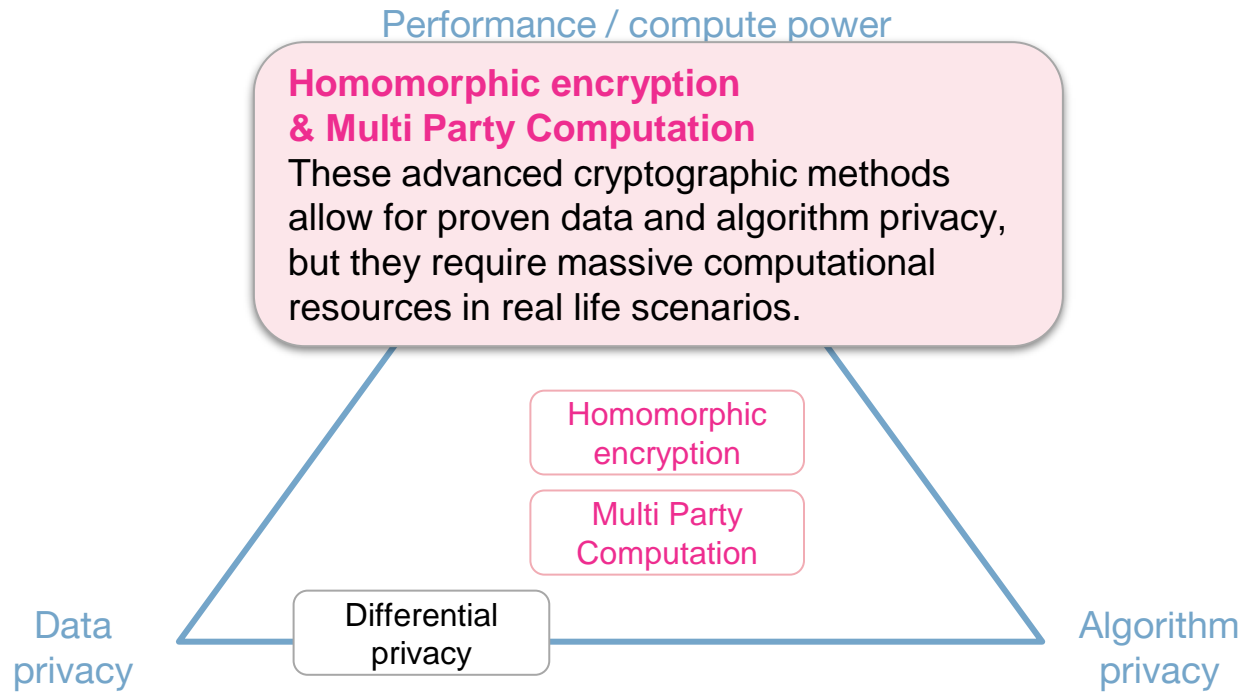
Federated

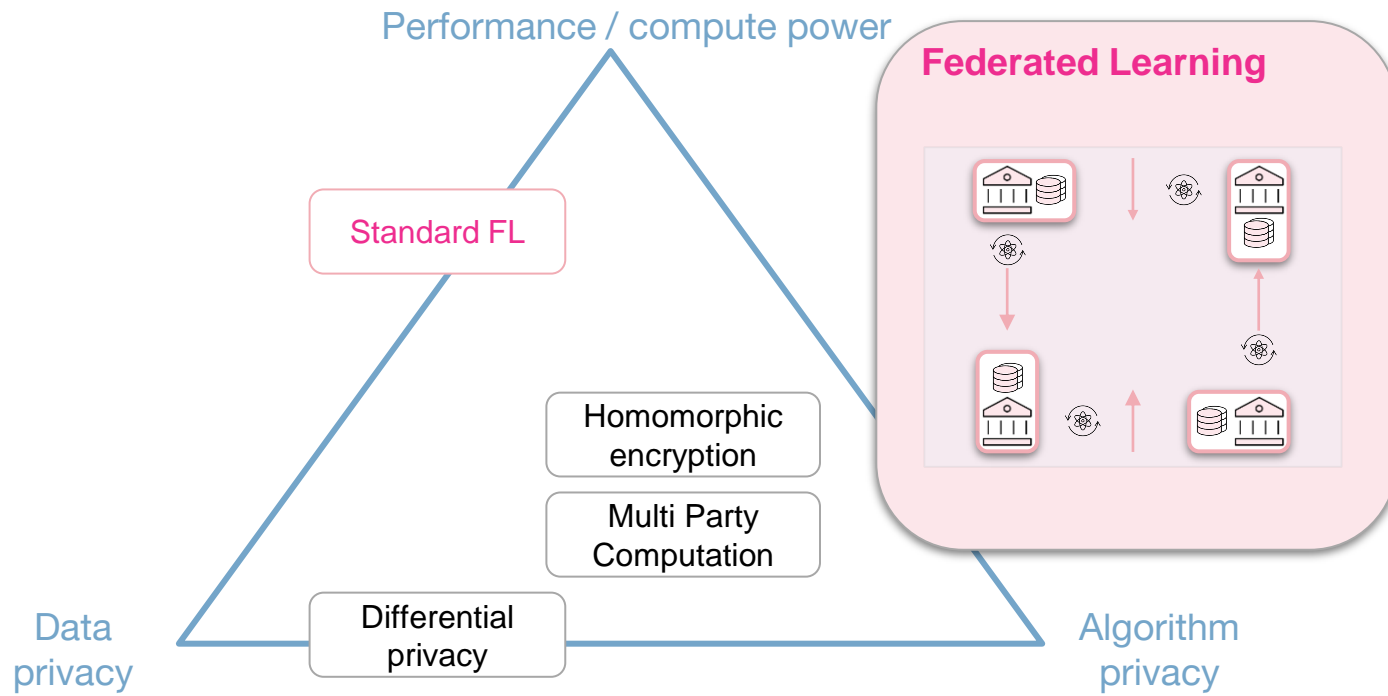


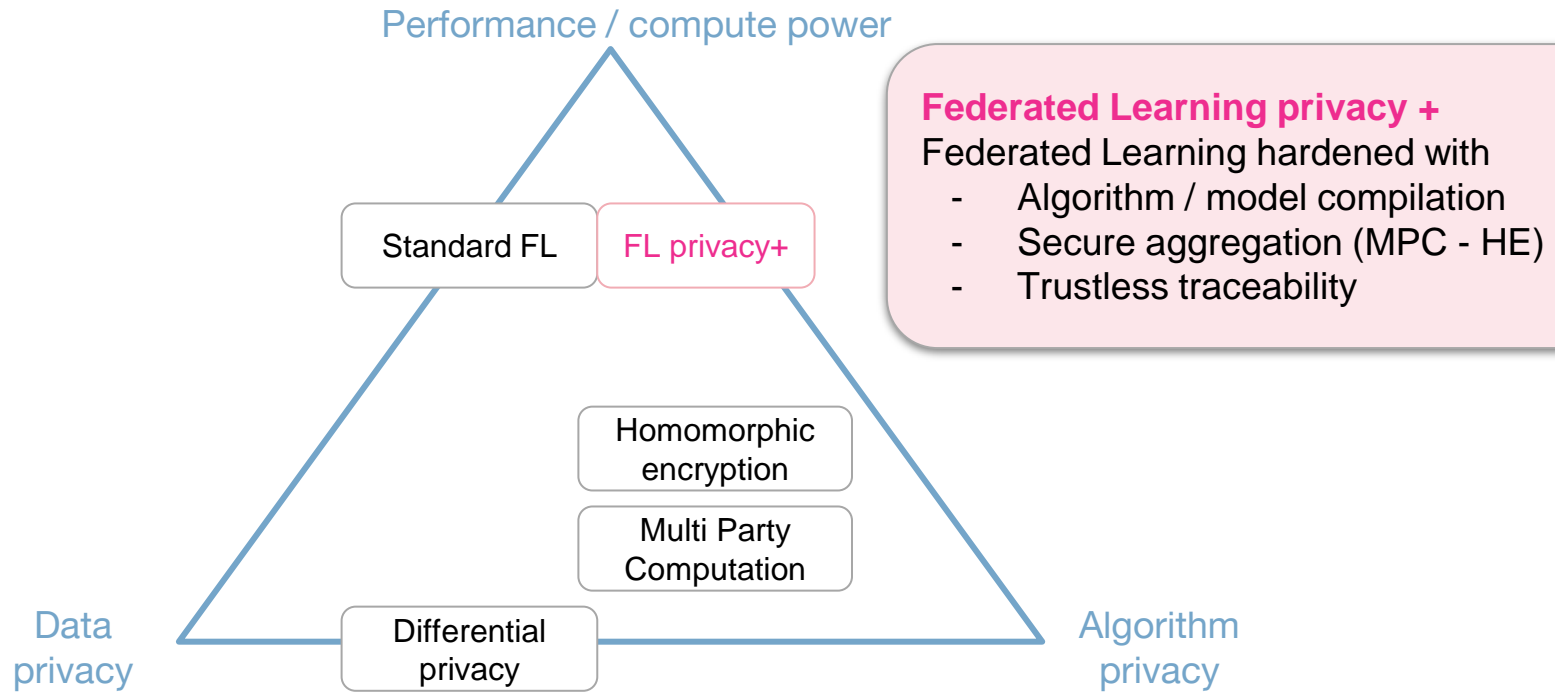
-  **Secure & private**
-  **Institute controls data**
-  **Efficient & scalable**



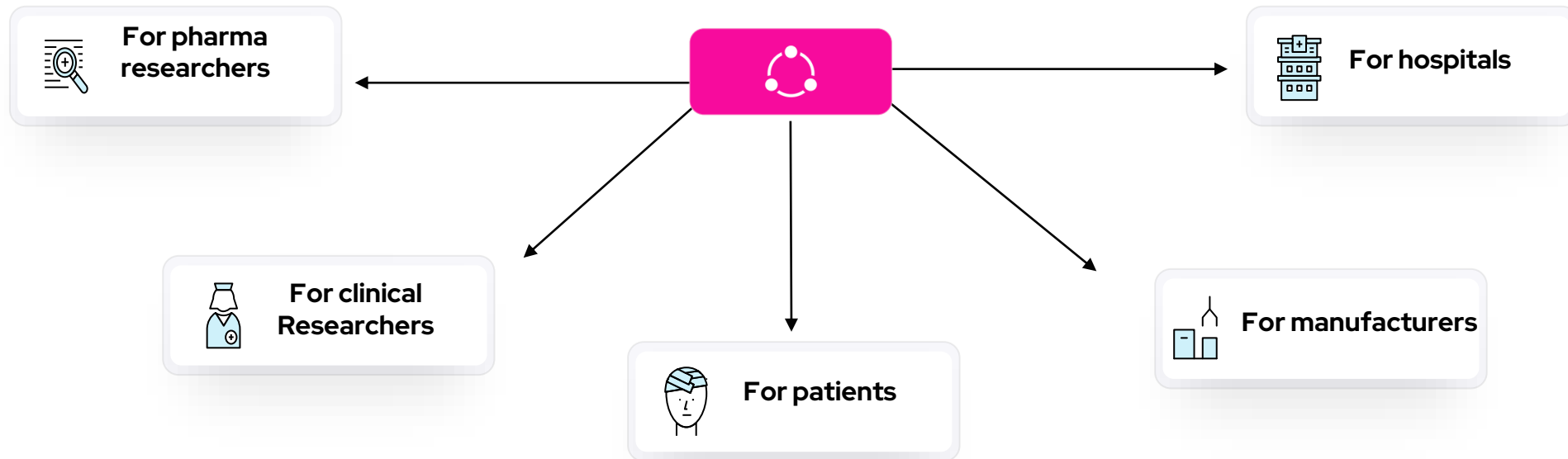






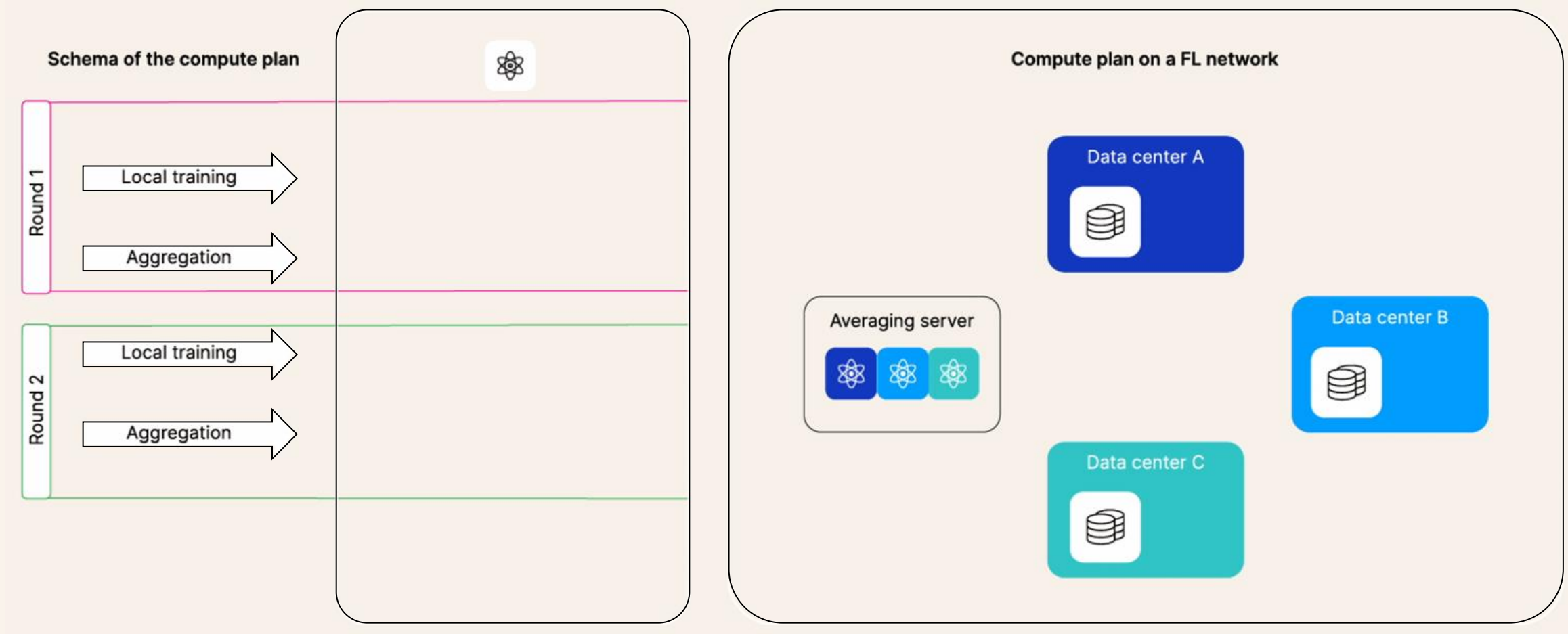


- + Cooperative consortia
- + International collaboration
- + Grouping of heterogeneous data
- + In situ validation at scale
- + Critical mass for rare disease R&D
- + Non-duplication of data

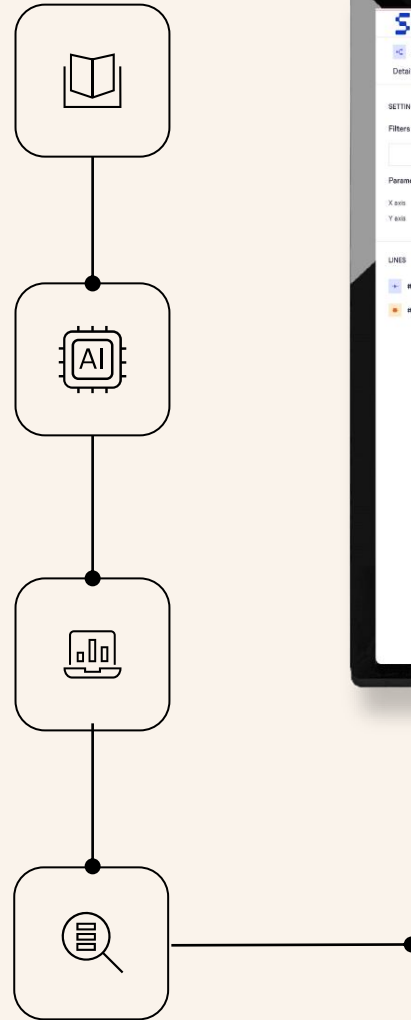
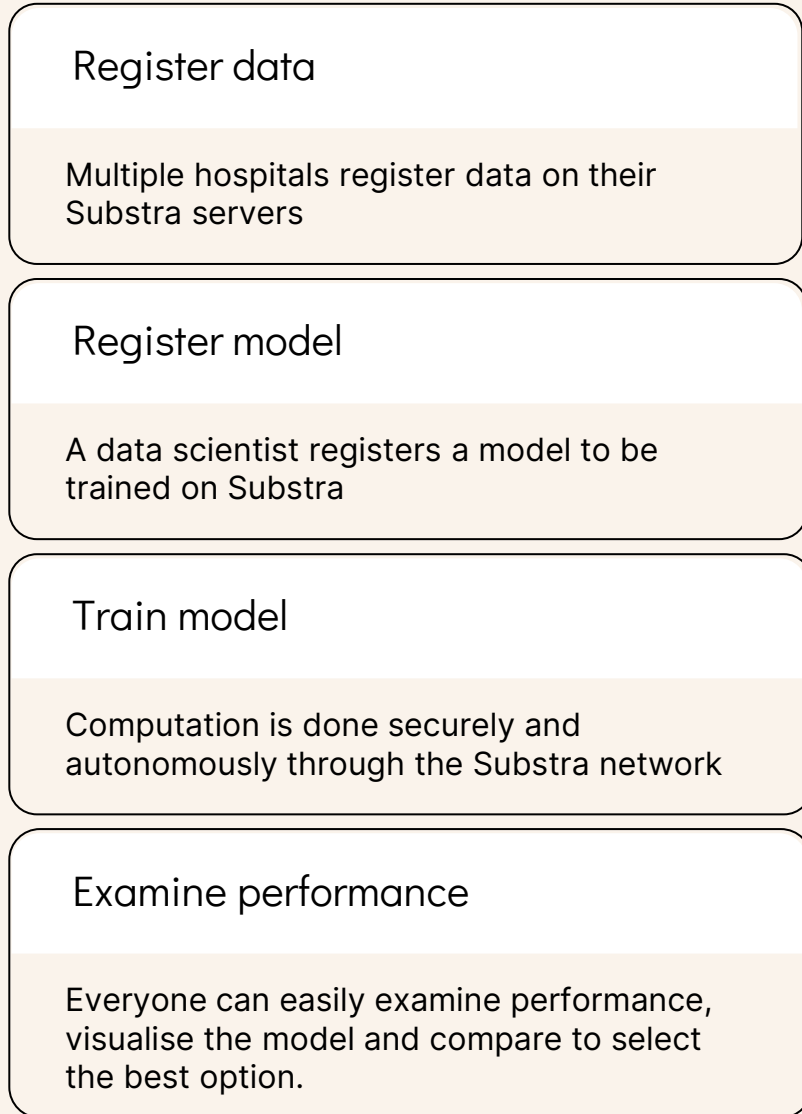


Owkin's federated learning platform, powered by Substra

↳ Ready-to-use, open source federated learning software developed by Owkin, now hosted by the Linux Foundation AI & Data Foundation



How Substra is applied at Owkin



Case study: MELLODDY collaboration, Substra in action

Objectives

Leverage the world's largest collection of small molecules with known biochemical or cellular activity to enable more accurate predictive models and increase efficiencies in drug discovery

Results

Successful runs of training traceable, privacy preserving, federated multi-task models on the MELLODDY platform using Substra

MELLODDY notably improved the performance of drug discovery models via privacy-preserving and federated learning for all 10 pharmaceutical partners.

Project

Consortium

Owkin led European IMI consortium – €18.4M Budget. Project duration: 2019-2022

Research questions

- Prove feasibility of a trustless & privacy preserving federated learning platform
- Develop & train multi-task machine learning models for drug discovery

Data

- 10 million annotated small molecules
- 1 billion assay biological activity labels
- Multiple high-complexity phenotypes from high throughput screening

Partners

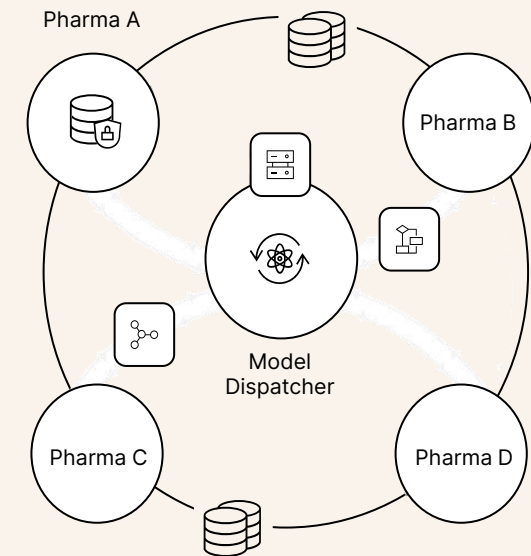
Pharmaceutical partners



Other partners



Project Coordinators



Why use Substra to connect and build models from healthcare data

↳ Substra is the only federated learning software proven in real production environments powering medical research



Data agnostic

Can be used with many different data modalities such as video, images, audio, time series or tabular data



Framework agnostic

Compatible with any machine learning model from any framework including TensorFlow, Torch and Scikit-Learn



Infrastructure agnostic

Works seamlessly with any IT infrastructure setup - on premise or in the cloud



Flexible

Can be deployed in a federated production environment or used on a single machine to perform simulations and debug code



Traceable

Monitoring, traceability and fine grained permissions for complete control of data manipulations



Secure

Proven in strict privacy settings to meet global compliance requirements for hospitals and pharmaceutical companies

OWKIN: AI for precision medicine



Our data network



Selected academic partners



Consortia



Published collaborative research



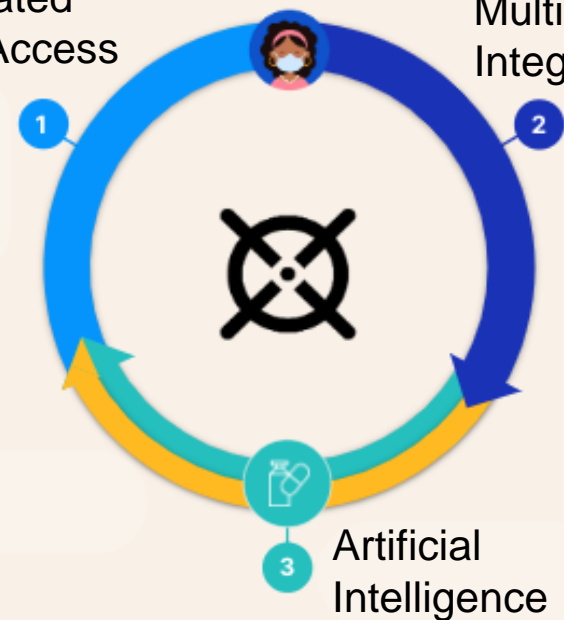
What we do

Multimodal Patient Data

INPUT

Federated Data Access

Multimodal Integration



OUTPUT

Diagnostics

Pharma Solutions




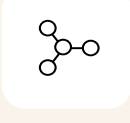
Drug Discovery

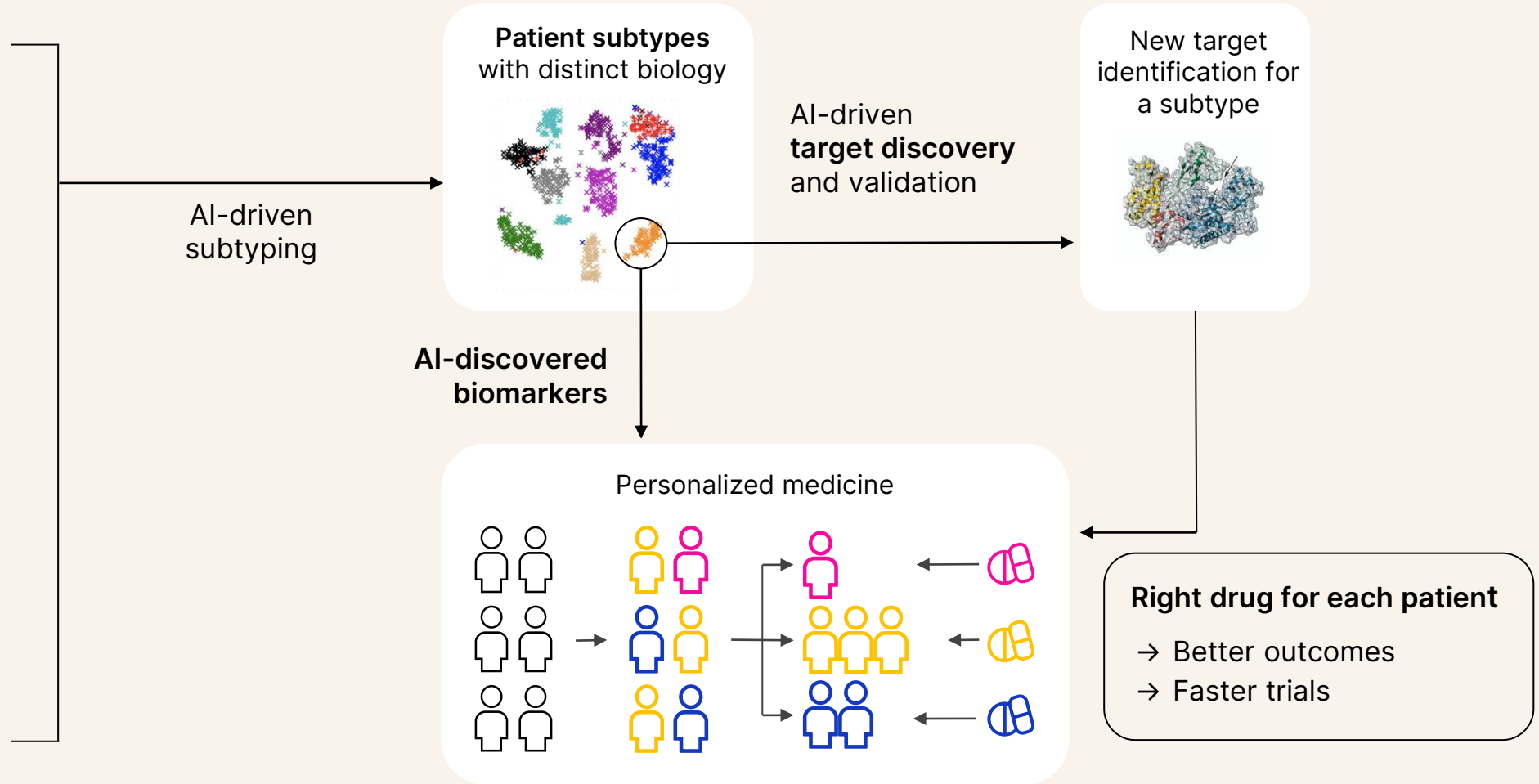
Owkin's reverse translational approach to precision medicine



→ We apply AI to multimodal patient data to enable discovery and development

Multimodal patient data

-  Clinical
-  Histology
-  RNA-seq
-  DNA-seq/WES
-  Spatial -omic / Single cell





Join our revolution towards precision medicine



www.owkin.com



joseph.lehar@owkin.com



[@OWKINscience](https://twitter.com/OWKINscience)



[@owkin](#) | [josephlehar](#)



New York, Boston, London, Berlin, Nantes, Paris



Humanity: The people at the heart of AI-augmented healthcare
Published by Owkin (2023)