

# Federated learning to unlock biomedical research at scale

↳ Mathieu Andreux, PhD - Owkin  
Federated Learning Research Group Lead





Across industries,  
machine learning  
applications require  
to collect a critical  
mass of data



Healthcare



Pharma



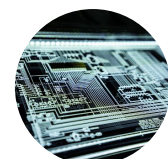
Manufacturing



Sustainable  
Energy



Aviation

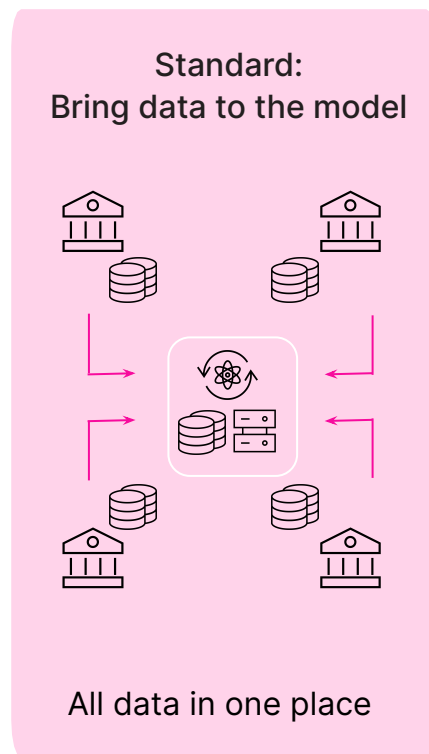


Cybersecurity



# Whenever data is regulated or sensitive, data sharing obstacles arise, reducing the impact of machine learning

The standard ML approach typically pools data in a central repository...

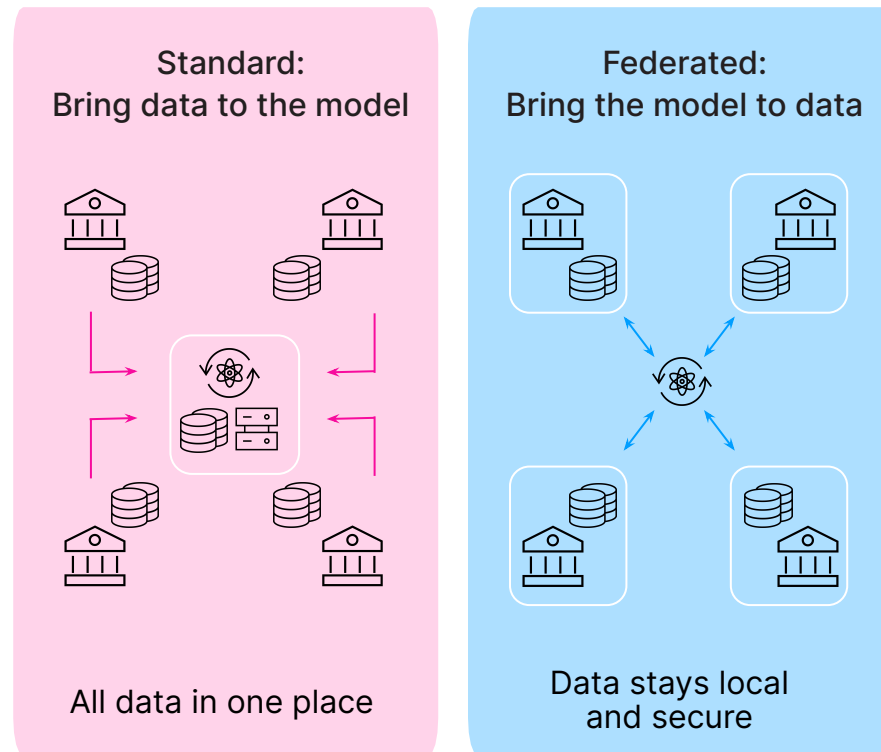


- ✓ **Volume** - critical data amount attained
- ✓ **Flexibility** - all data science can be run
- ✗ Data is sent once and hard to **refresh**
- ✗ Increased risk of data leaks and privacy breaches, compromising **compliance** (GDPR...)
- ✗ Data is a **strategic asset** that centers are not necessarily willing to share



# Federated Learning circumvents data sharing obstacles to enable ML across silos

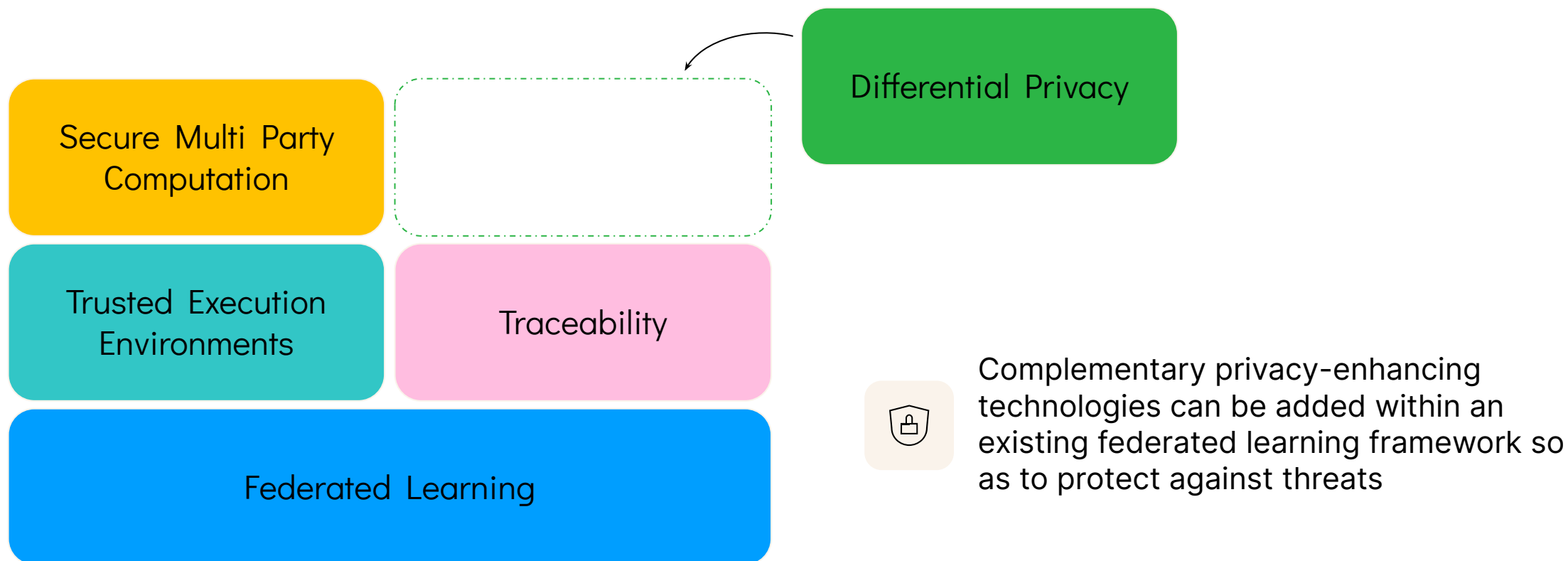
Building models on non-collocated datasets is possible with federated learning



- ✓ **Volume** - critical data amount attained
- ✓ **Flexibility** - most data science can be run
- ✓ Data can be **refreshed**
- ✓ Ensures **compliance** and increases **privacy**
- ✓ Centers keep **control** on data usage

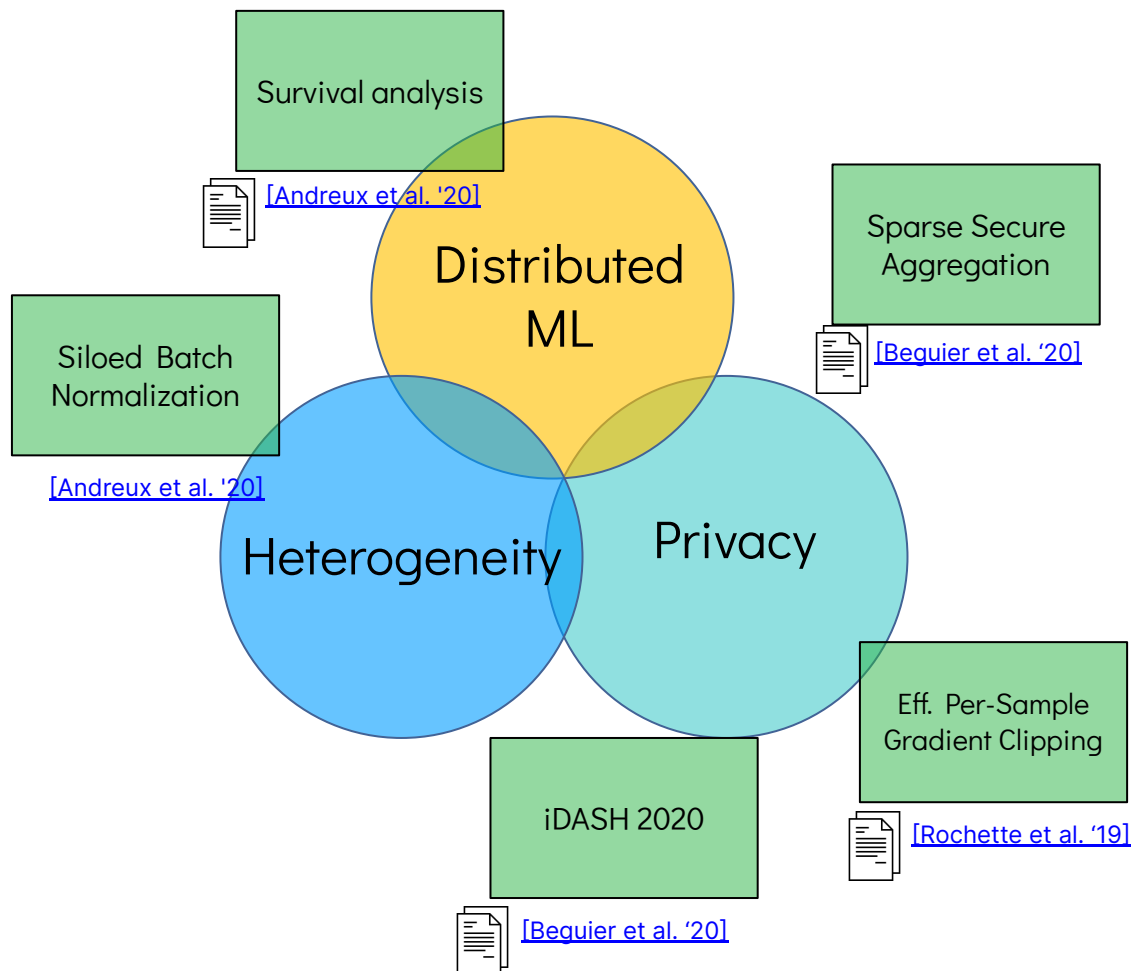


# Federated Learning can be used with other privacy-enhancing technologies to match confidentiality requirements



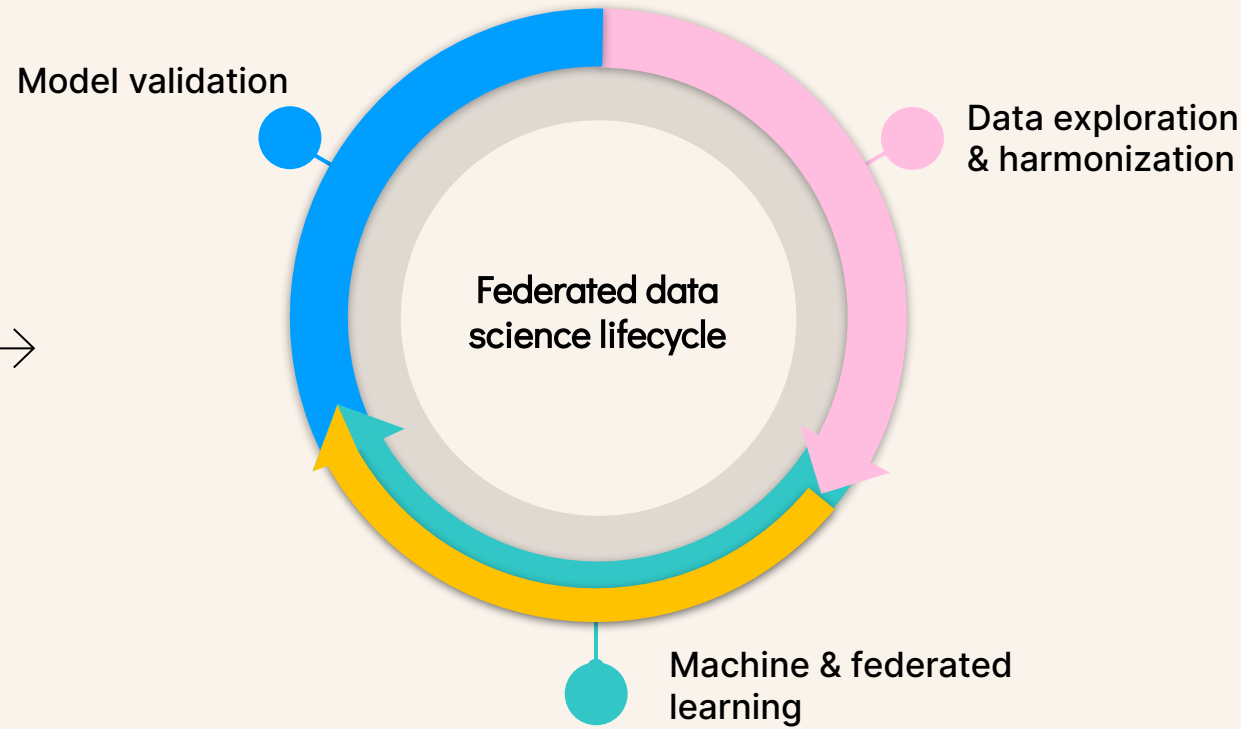
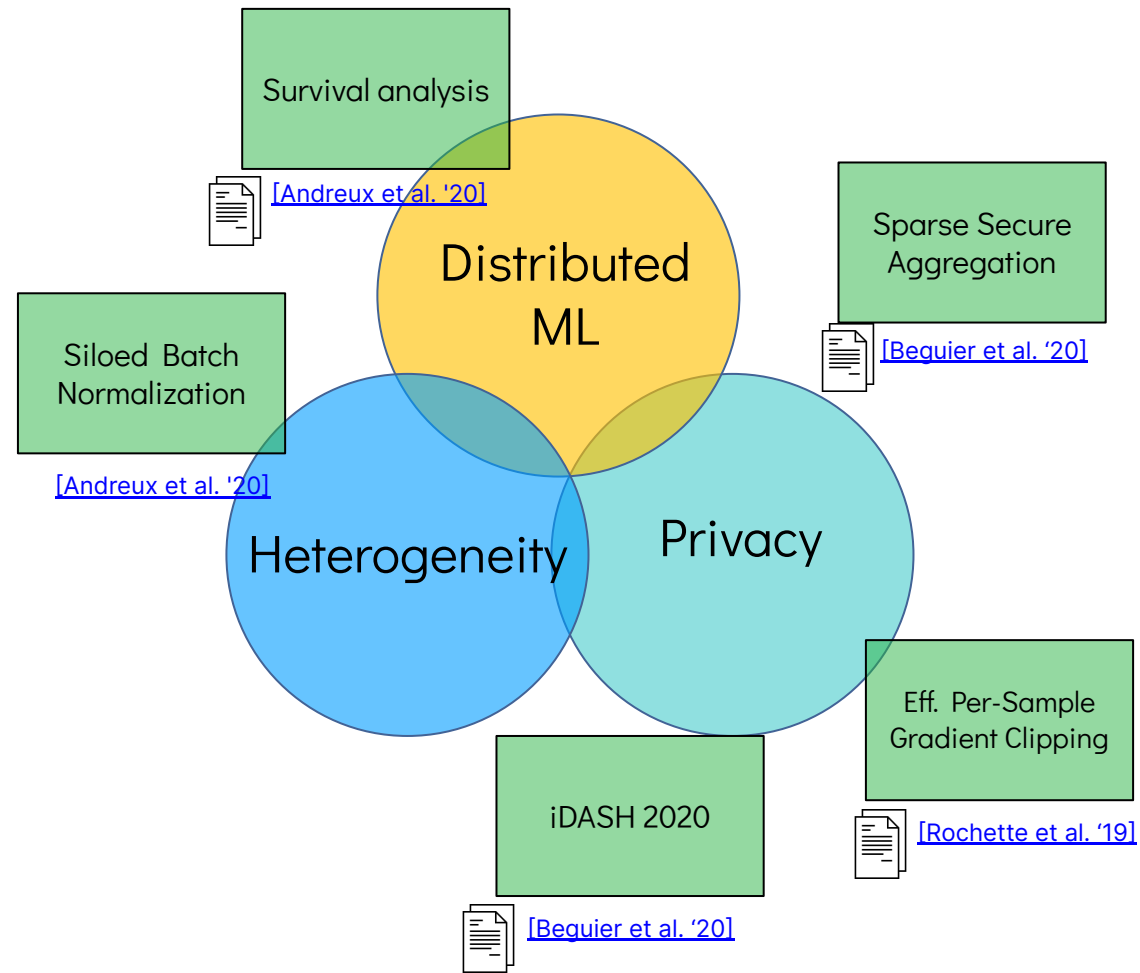


Owkin conducts methodological R&D to enable distributed machine learning under statistical heterogeneity and privacy constraints



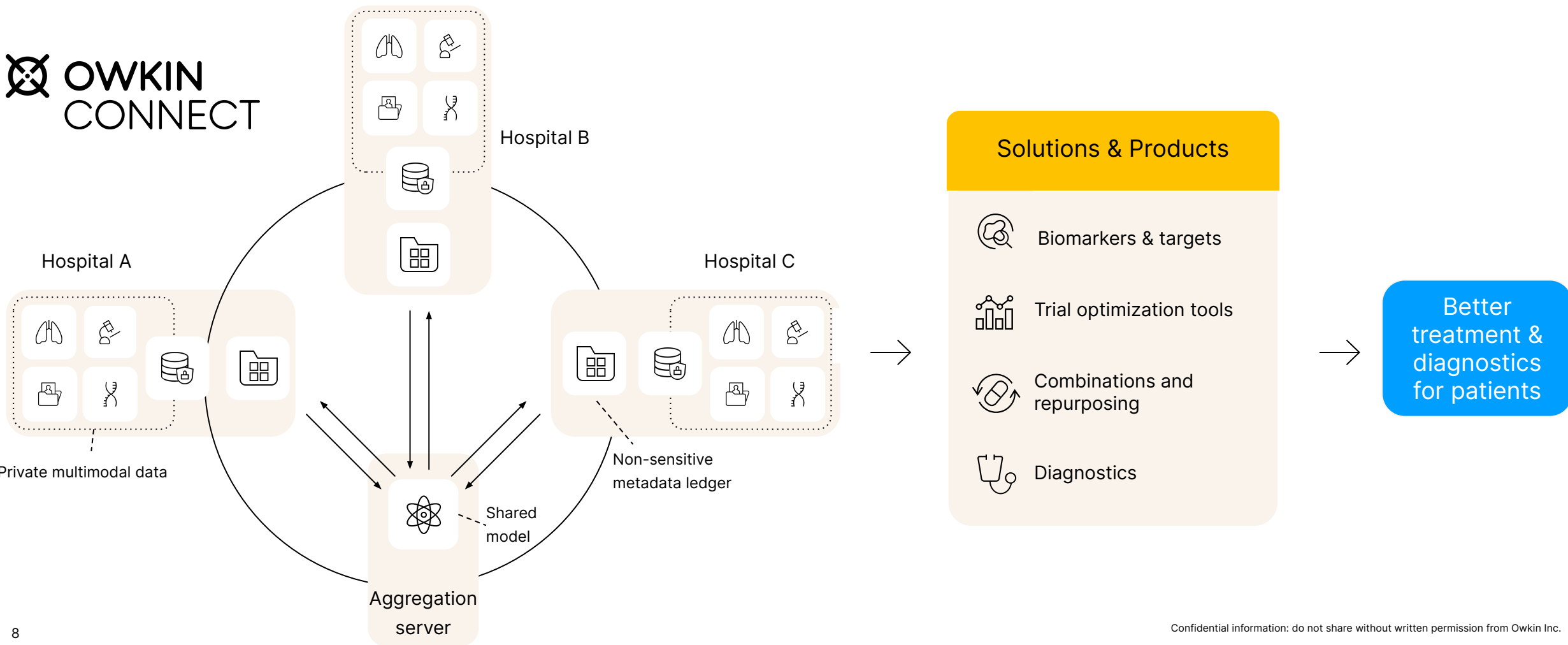


Owkin conducts methodological R&D to enable distributed machine learning under statistical heterogeneity and privacy constraints across the full data science lifecycle.





# Owkin develops Owkin Connect, a federated learning software, to enable biomedical research at scale across diverse data modalities





# MELLODDY, an Owkin Connect real-world deployment across 10 pharmaceutical companies

In 3 yearly runs, MELLODDY aims to show **predictive benefits of modelling across tasks, data types and partners** at the largest achievable scale, based on:

- 10 million annotated small molecules;
- 1 billion assay biological activity labels;
- multiple highly complex phenotypes at high throughput



PHARMA PARTNERS



PUBLIC PARTNERS



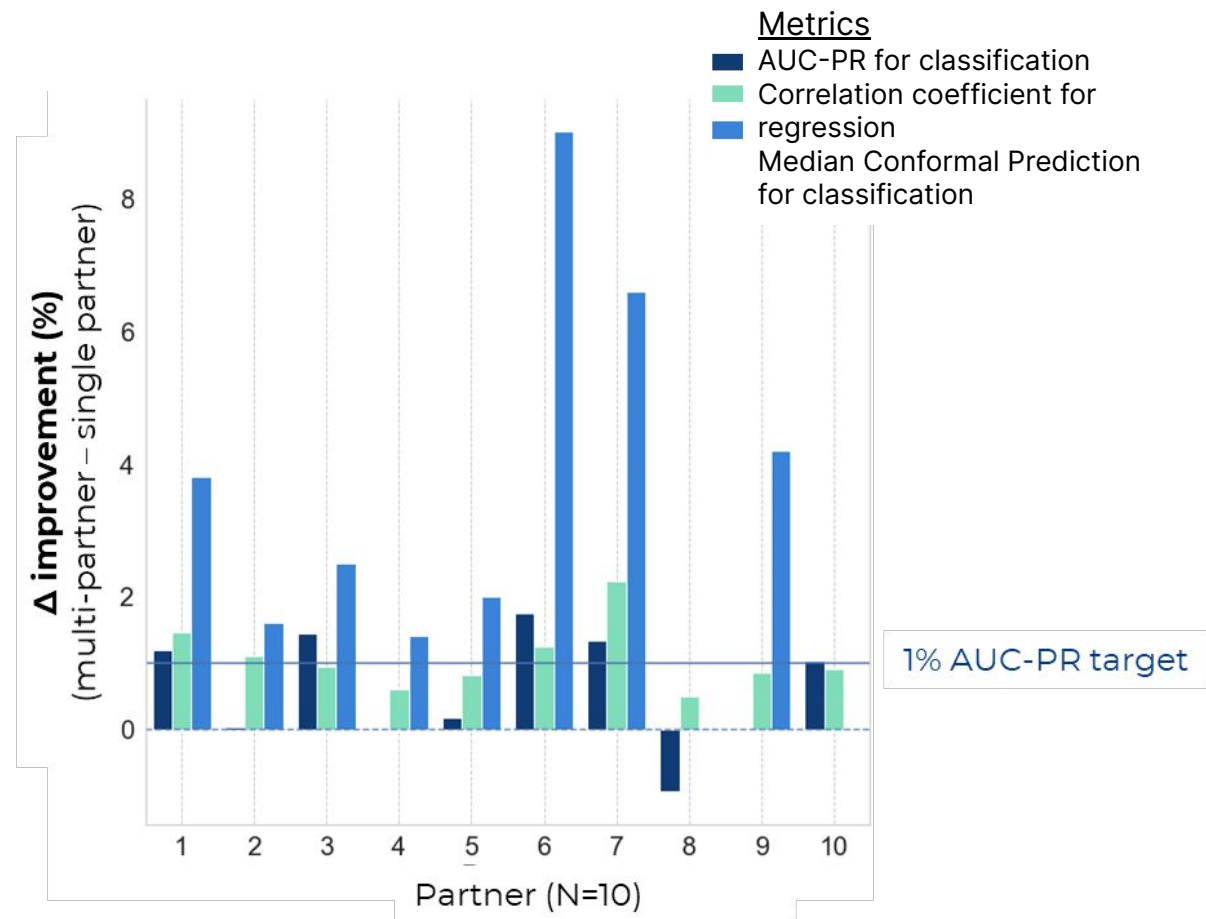
This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement N° 831472. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA

18M€





# MELLODDY's year 2 results show early evidence of federated learning benefits on models' performance.



*“In demonstrating federated multi-task learning across more than 100,000 machine learning tasks representing more than 40,000 concentration response assays we are excited to see **early evidence** that it indeed **boosts the predictive performance and chemical applicability of models** used to inform drug discovery programs”*

Hugo Ceulemans, Scientific Director,  
Janssen Pharmaceutica, NV and  
MELLODDY Project Leader

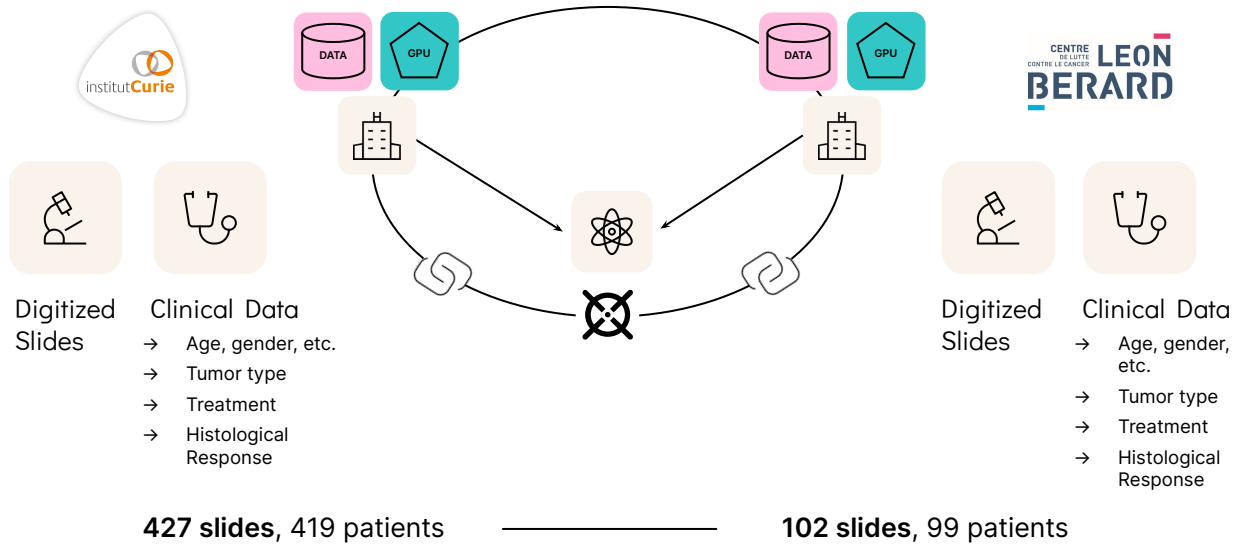
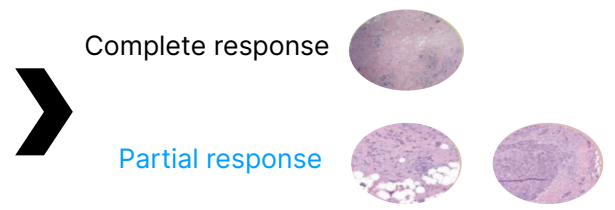


# Healthchain-Breast is a research project to study treatment response in breast cancer using federated learning

- **Medical Question:** Can we predict the rate of pathological complete response (pCR) to neoadjuvant chemotherapy (NACT) from slides at diagnostic for early triple negative breast cancer patients?
- **Implications:** if AI predicts pCR, therapeutic de-escalation (less cardiac toxicity) can be considered, if not patients can be included in clinical trials for PARP inhibitors or immunotherapies

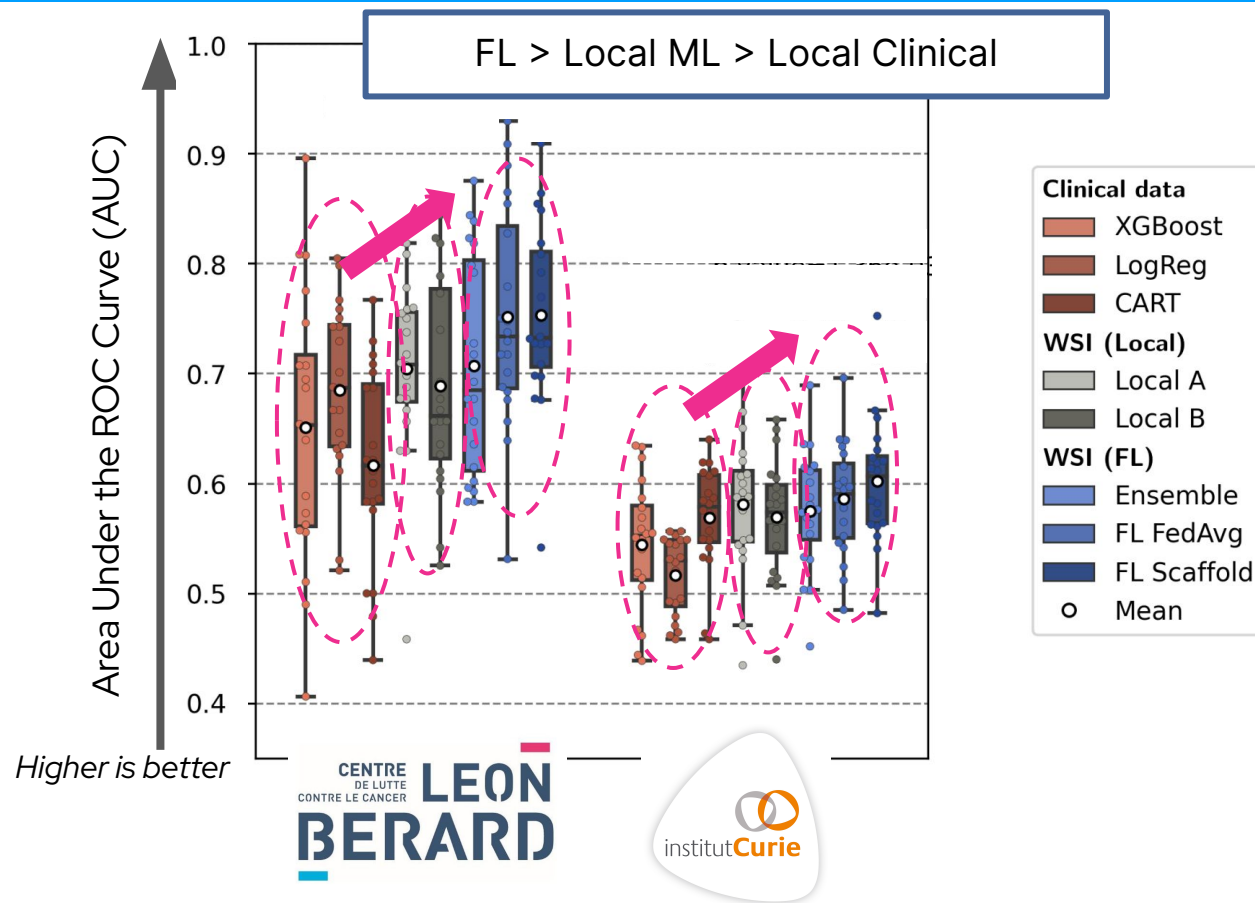


H&E slide





# Healthchain results indicate benefits of using slides to predict treatment response over clinician annotations, as well as of federated learning collaboration



- ✓ Deep Learning models on WSI are as good or better than clinical models relying on human annotations
- ✓ FL models on WSI outperform local WSI models and have better transfer performance
- ✓ First FL project in real conditions on large histopathology images



Pre-publication

Ogier du Terrail, J., Leopold, A., Joly, C., Beguier, C., Andreux, M., Maussion, C., ...& Bataillon, G, Heudel, P. E. (2021). Collaborative Federated Learning behind Hospitals' Firewalls for Predicting Histological Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. medRxiv.



# Many other federated learning consortia are starting:



Central Repository for Digital Pathology  
(2021 - 2027)

IMI consortium - 70M€  
45 partners



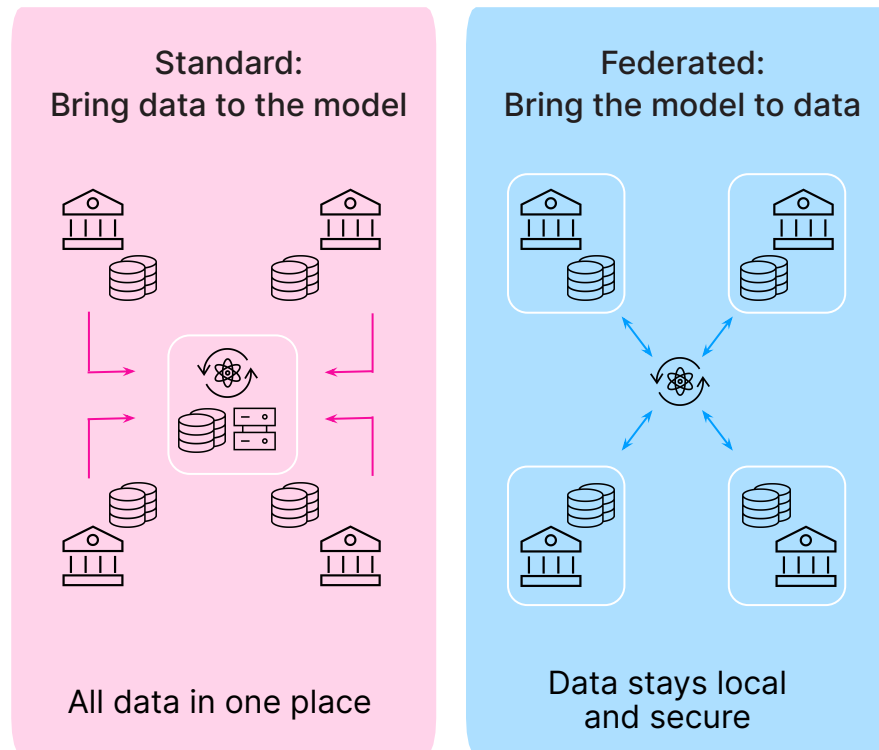
Optimal treatment for patients with solid tumours in  
Europe through AI (2021 - 2026)

IMI / H2020 consortium - 21M€  
36 partners

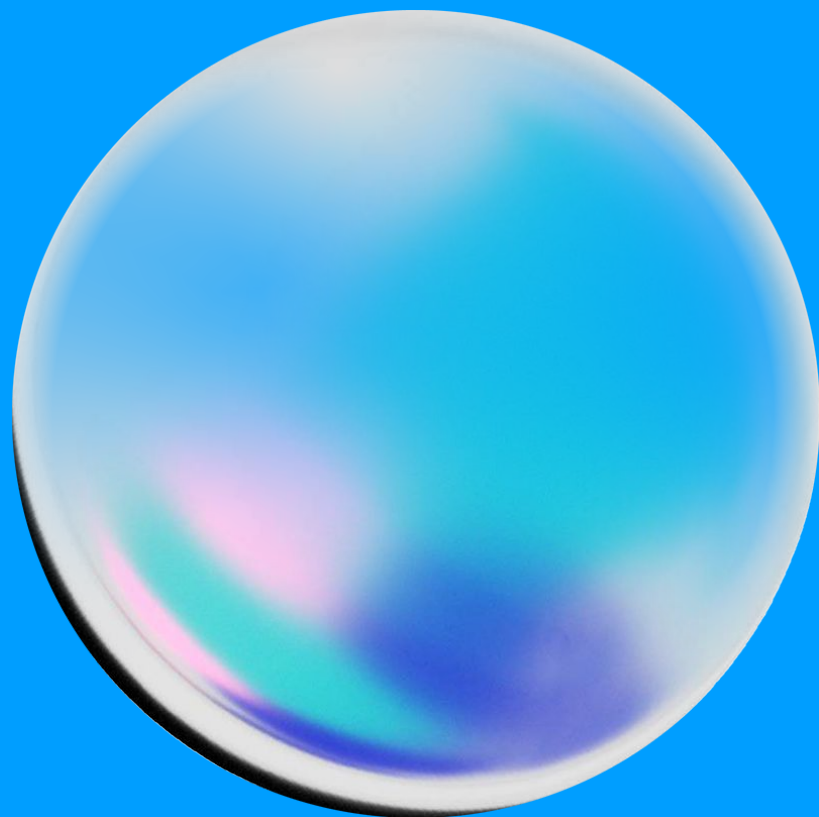




# Federated Learning can break silos to enable machine learning applications at scale



- ✓ Federated learning ensures **compliance** and **increases privacy** level
- ✓ Federated learning opens new **opportunities** to apply **machine learning across silos**
- ✓ Owkin uses federated learning to **unlock biomedical research at scale** in multiple research consortia



# Thank you

 [www.owkin.com](http://www.owkin.com)

 [hello@owkin.com](mailto:hello@owkin.com)  
[mathieu.andreux@owkin.com](mailto:mathieu.andreux@owkin.com)

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

 [@mathieu.andreux](https://www.linkedin.com/in/mathieu.andreux)

 New York, Boston, London, Paris, Nantes