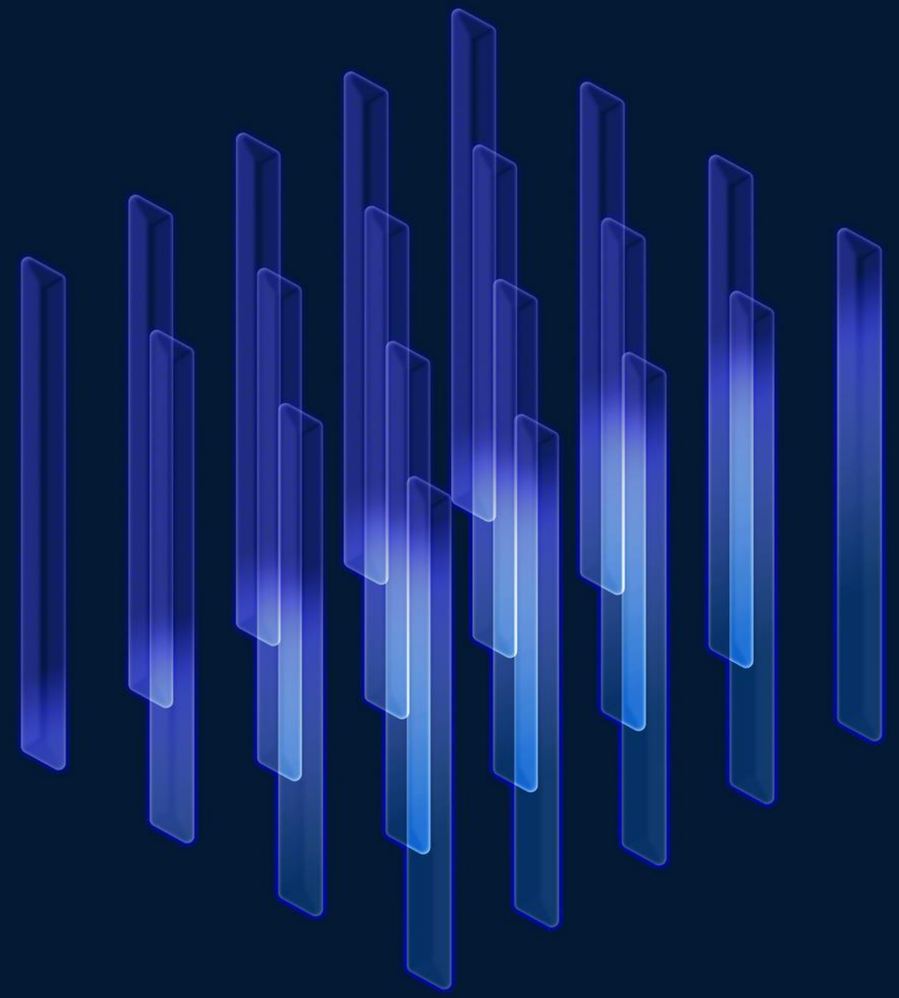
QCIF
Digital Research³

Federated Learning for eResearch.

Why, what, how, where to.



Peter Marendy \ 23 October 2025

Acknowledgement

I would like to acknowledge the Traditional Owners of the lands on which I am speaking to you today, the Turrbal and Jagera people.

I also acknowledge the Traditional Owners of the lands on which each of you are working from.

I acknowledge their continuing connection to their culture and pay my respects to their Elders past and present.

About us

Not-for-profit organisation dedicated to research excellence and catalysing innovation through cutting-edge data, and digital infrastructure & enterprise solutions.

We support researchers across Australia by providing expert guidance and training in a wide range of data and digital capabilities, impacting hundreds of projects and benefiting thousands of end-users.

The problem

- Traditional approaches to machine learning on distributed dataset involve:
  - Bringing the datasets together to form an aggregate at the model, or
  - Onsite access – bring the analyst/modeler to the various datasets
- This presents governance challenges
  - Different data owners/custodians – siloed data
  - Different access policies and procedures
  - Fear of data leakage/breaches
  - Data leaves sphere of control of data owners/custodians

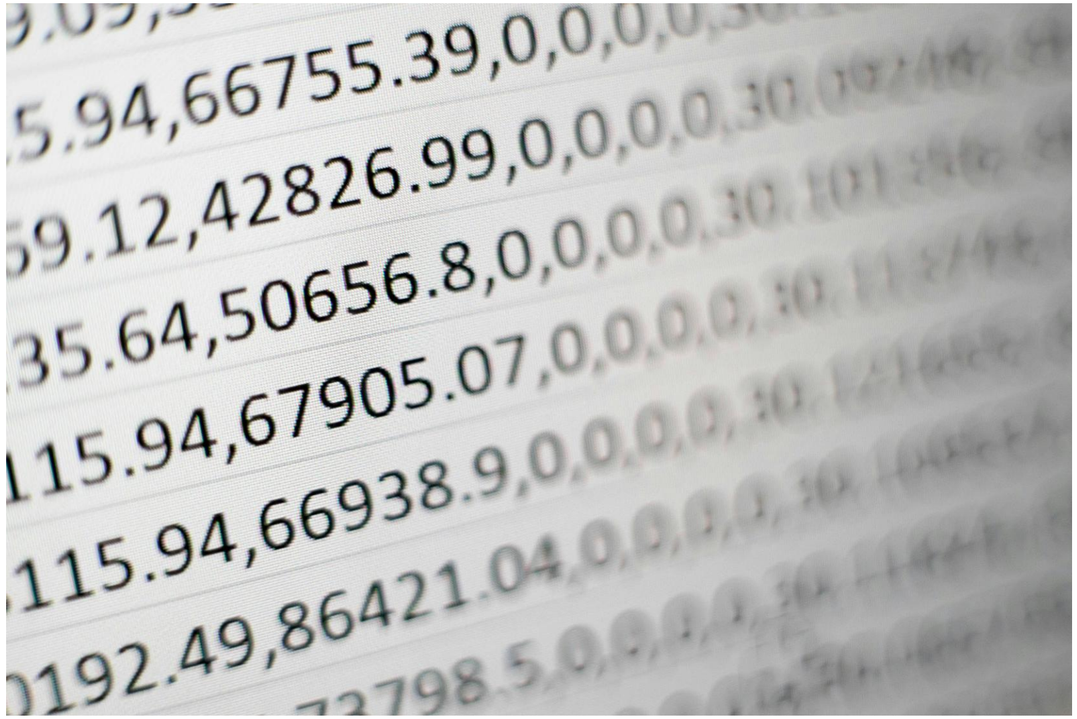


Photo by Mika Baumeister on Unsplash

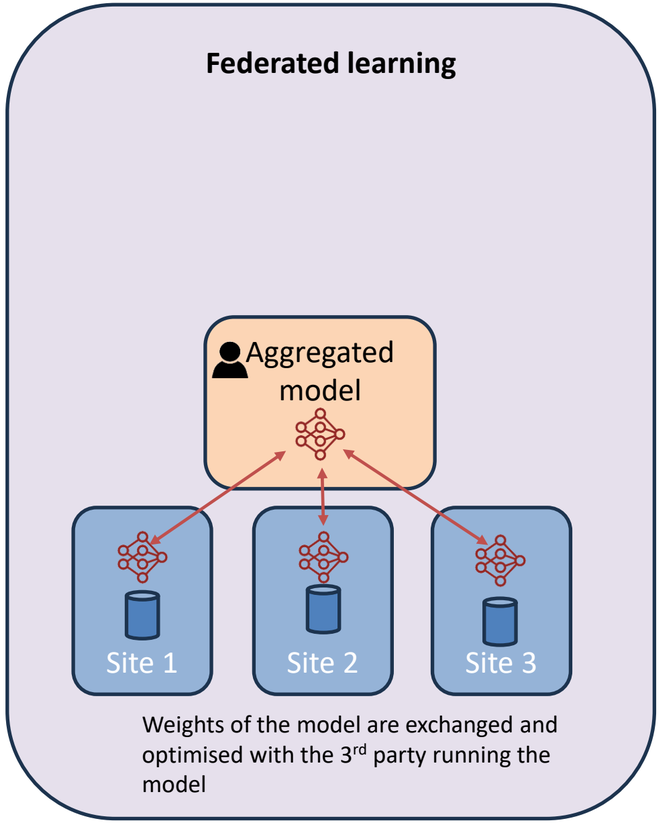
Why should we care?

- The governance hurdle blocks, discourages, and prevents valuable datasets from being utilised in research.
- E.g in health:
    - Australia's lack of data infrastructure hinders large-scale research efforts for chronic diseases, which have significant long-term effects, comorbidities, and economic burdens.
    - Traditional approaches require data to be merged centrally, resulting in significant governance issues with data existing across many providers.
    - Chronic conditions place a significant burden on individuals, their families, and the health system. People with chronic conditions are at risk of premature death and poor quality of life, while multimorbidity can make treatment more complex and costly to the health system.

Photo by National Cancer Institute on Unsplash

A solution?

- Federated Learning
  - Bring the model to the data
  - Data remains in sphere of control of data owner/custodian
  - Aggregation of outcomes from many trained models
- Types of Federated Learning
  - Horizontal Federated Learning
  - Vertical Federated Learning
  - Federated Transfer Learning

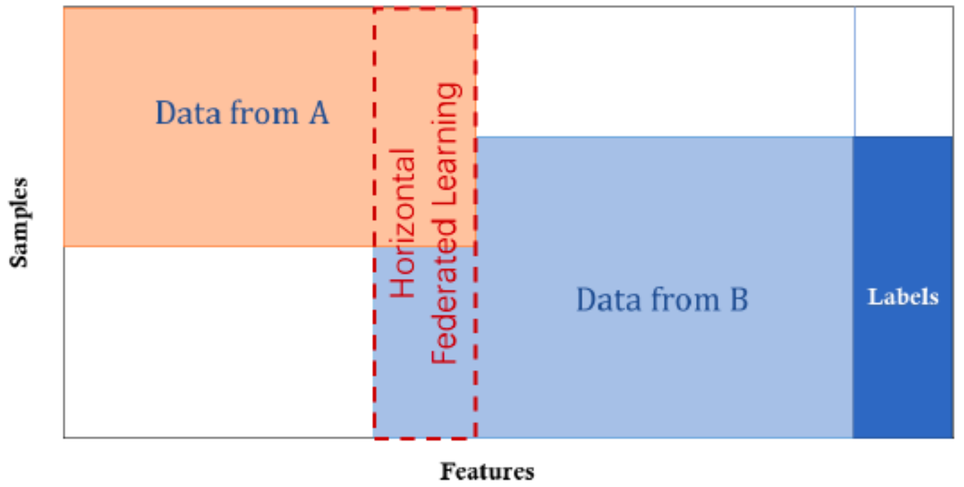


Three core federated learning strategies

Category 1: (most common)

Horizontal federated learning

- Overlapping features between nodes, different samples
- Same model run on each node
- e.g. Same observations made at each node, but for different patients.



Horizontal Federated Learning *

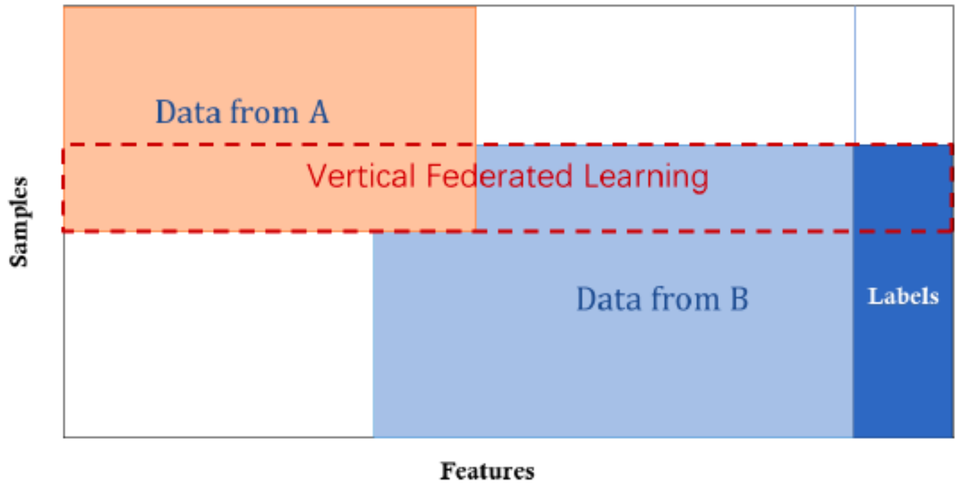
* Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 10, no. 2, pp. 1-19, 2019

Three core federated learning strategies

Category 2: (some uptake)

Vertical federated learning

- Overlapping samples between nodes, different features
- Creates an aggregate data set across nodes, but data remains in place.
- More complicated to implement
- e.g. Overlapping patients at each node, but different observations.



Vertical Federated Learning *

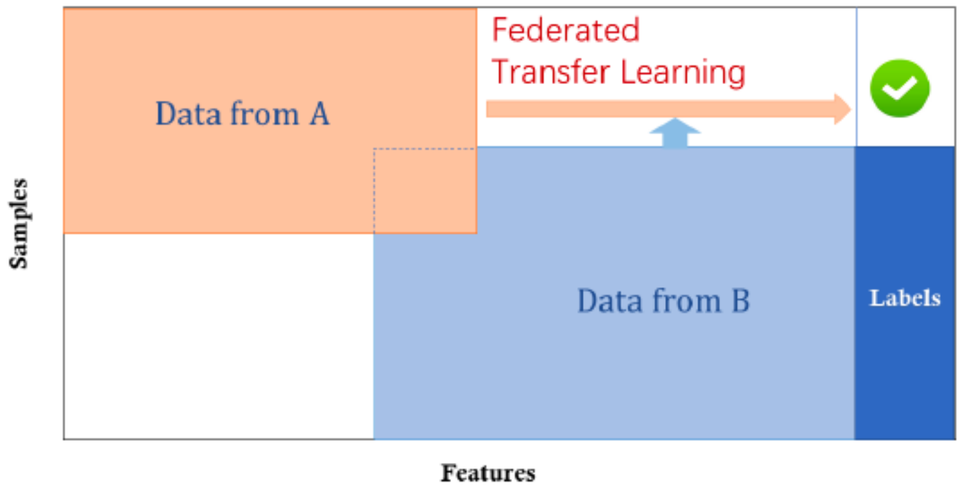
* Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 10, no. 2, pp. 1-19, 2019

Three core federated learning strategies

Category 3: (emerging approach)

Federated transfer learning

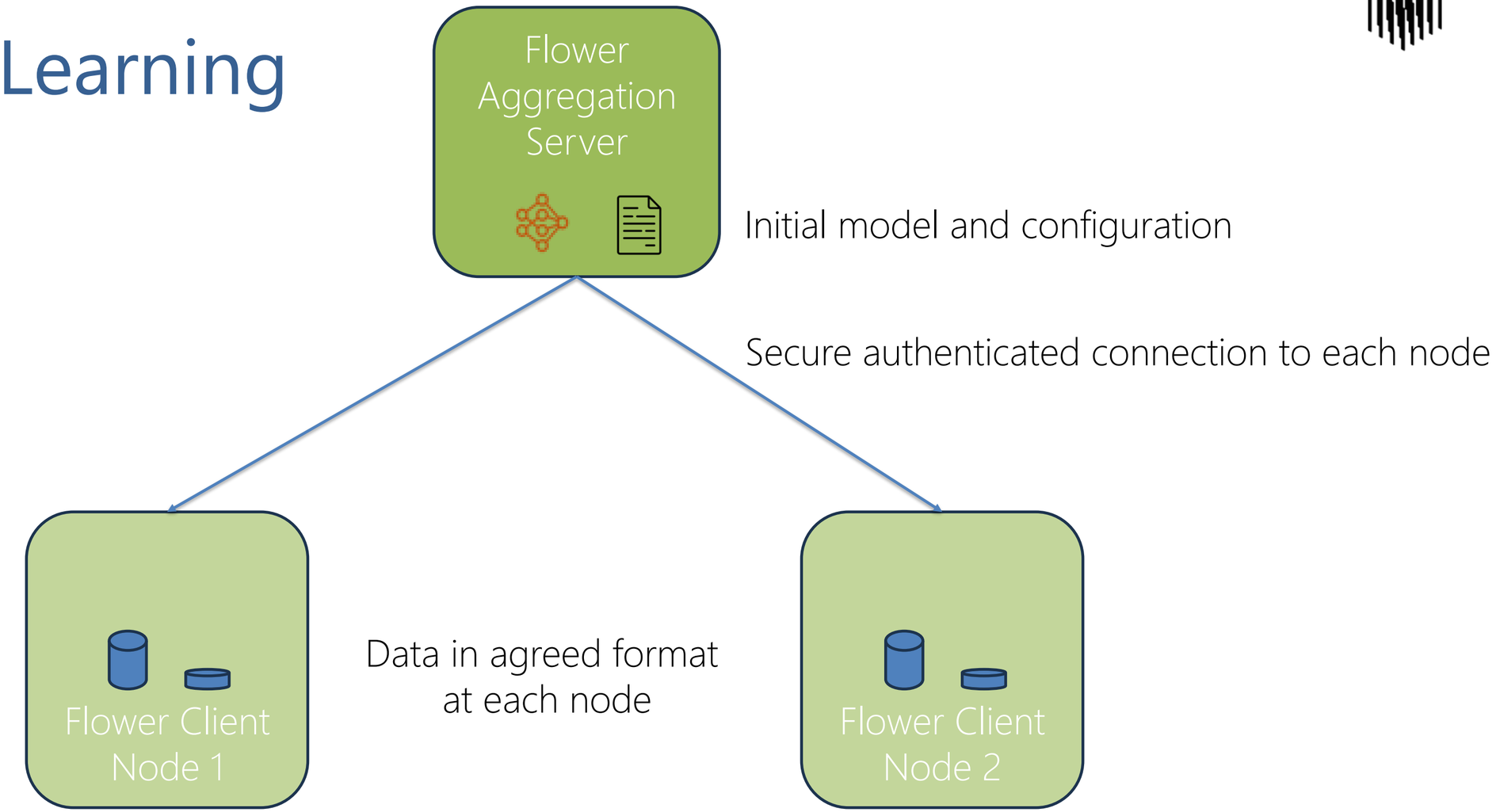
- Common samples between nodes, different features
- Model is trained on the feature rich dataset, then transferred to the feature poor dataset.
- Allows feature poor dataset to be trained against.
- e.g. Same images at each node, but with different resolutions and metadata.



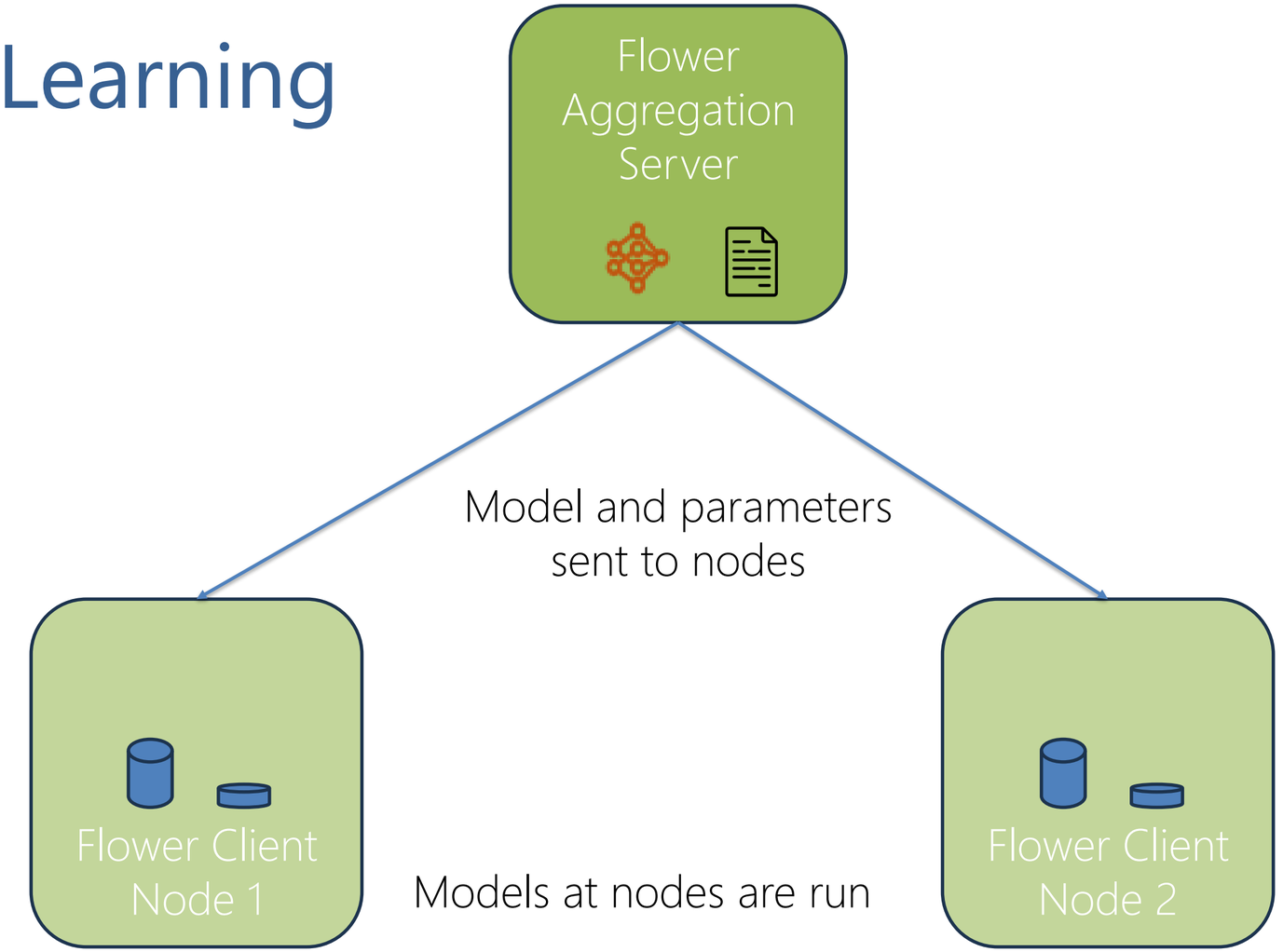
Federated Transfer Learning *

* Q. Yang, Y. Liu, T. Chen and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 10, no. 2, pp. 1-19, 2019

Horizontal Learning

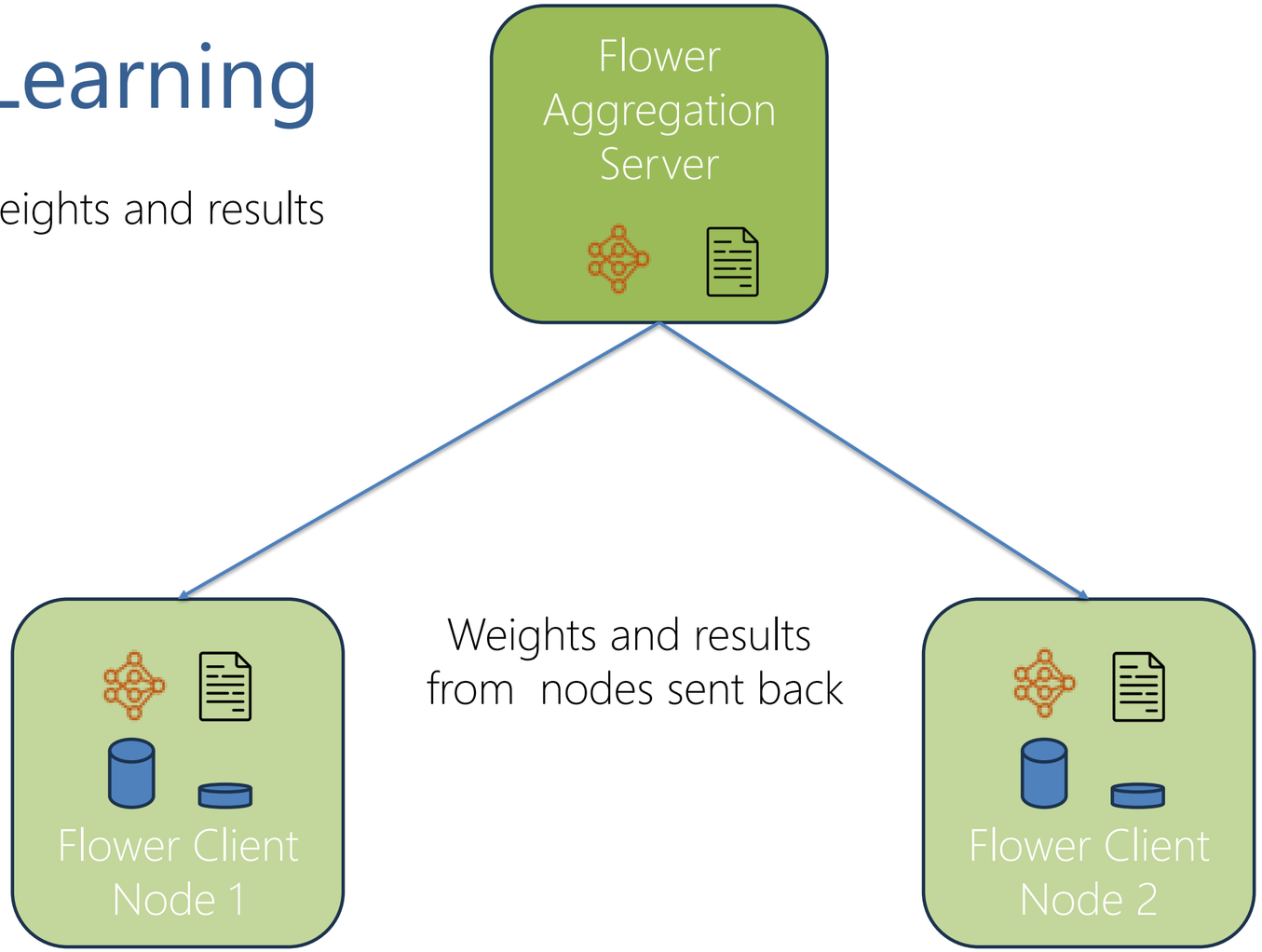


Horizontal Learning



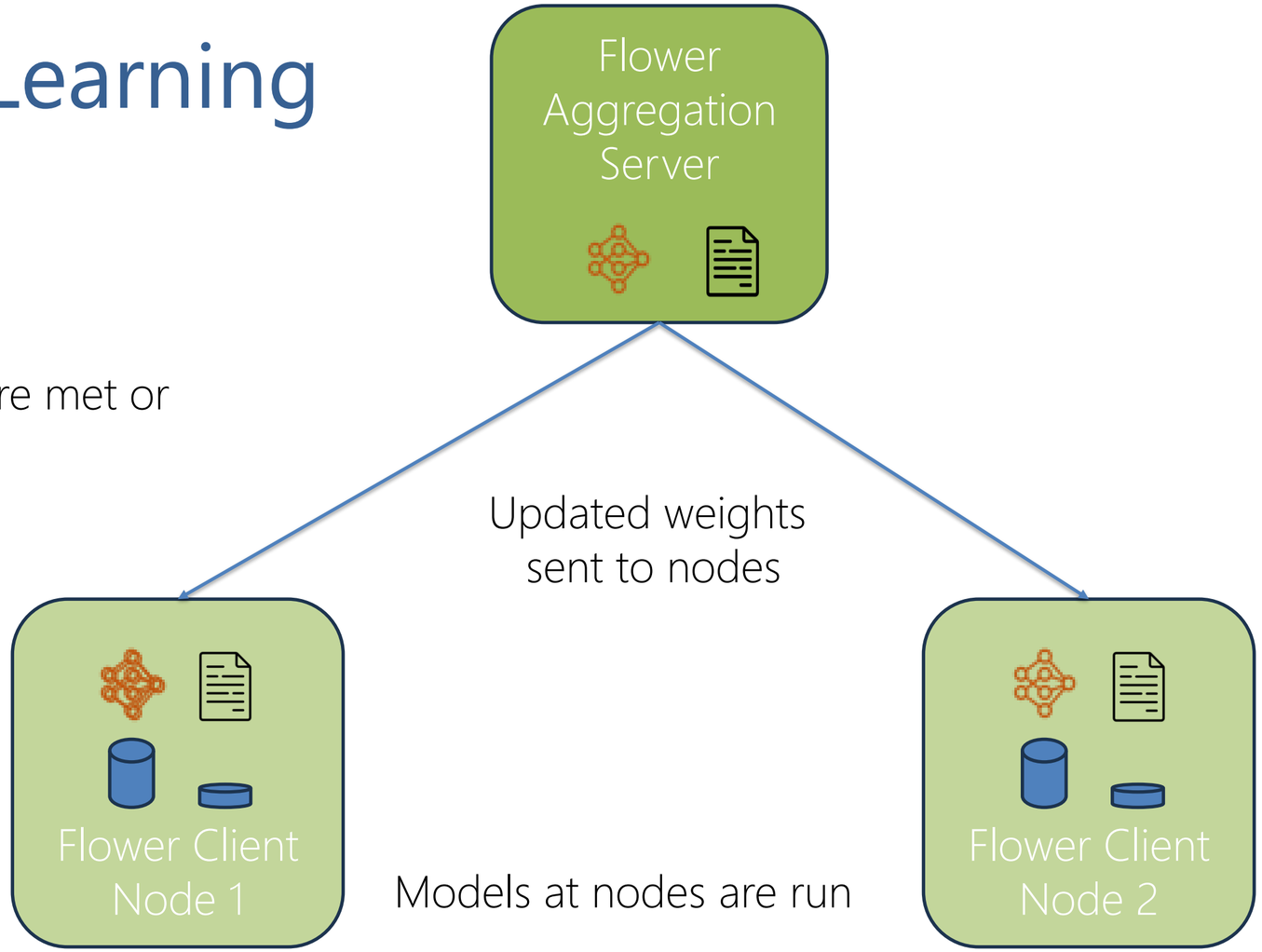
Horizontal Learning

Aggregation of weights and results

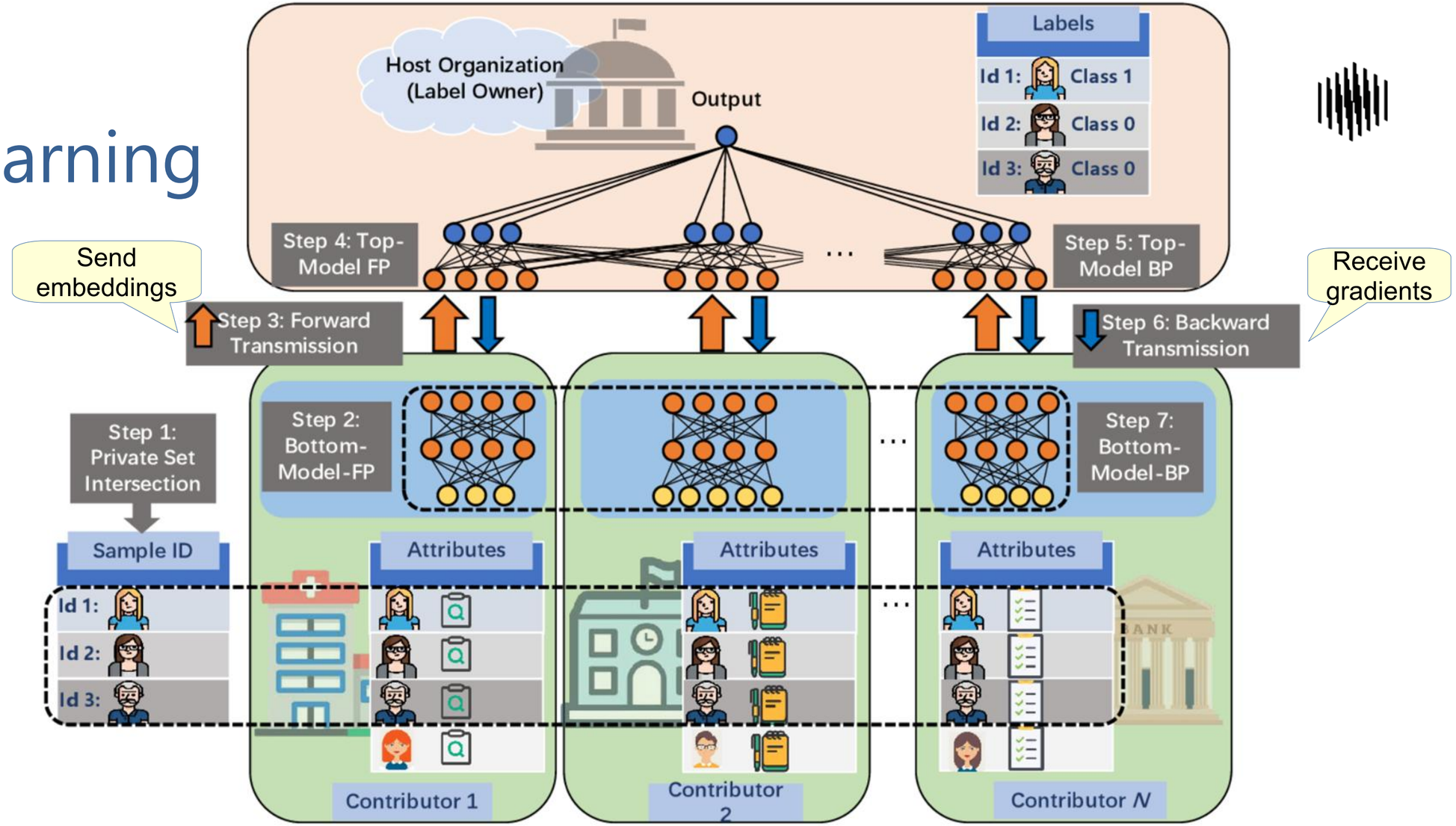


Horizontal Learning

The cycle continues until
set number of iterations are met or
parameters converge



Vertical Learning



Vertical Federated Learning: Challenges, Methodologies, and Experiments, K. Wei, et.al.

Selecting a framework

Initial Criteria

- Maturity (in years)
- Community (1 – 5)
- Documentation (1 – 5)
- Development (1 - 5)
- Ecosystem (1 – 5)
- Federated Learning Algorithms (1 – n)
- Machine Learning Agnostic (1 – n)
- Heterogenous (T/F (1/0))
- Differential Privacy (T/F (1/0))

Other Considerations

- Performance
- Deployment
- Administration
- Cost
- Supported platforms
- Security

	Maturity (Years)	Community	Documentation	Development	Ecosystem	FL Algorithms #	ML Agnostic #	Security	Heterogenous (T/F)	Differential Privacy	Performance	Deployment	Administration	TOTAL
PySyft	4	5	5	5	5	5	2		1	1				33
Flower	4	4	4	5	3	5	3		1	1				30
TF Federated	7	4	5	4	1	5	1		0	1				28
FATE	4	4	3	5	5	3	2		0	1				27
IBM FL	3	4	4	1	3	5	4		1	1				26
NV Flare	5	3	3	3	2	4	3		1	1				26
OpenFL (intel)	3	3	3	3	2	4	3		0	1				22
substra	4	2	2	2	1	1	3		1	1				18
FederatedScope	3	1	1	2	2	3	3		1	1				17
Fedscale	3	1	1	1	1	3	3			1				14
FedML	1	1	2	3	1	1	1		1	1				12
Plato	3	1	2	1	2	1	1			1				12
EasyFL	1	2	4	0	0	1	0		0	1				8
FEDn	3	3		0						1				4
XFL	1	1		0						1				3
PaddleFL	2	1		0						1				2
FedLearner	3	1		0						1				2
pyVertical				0						1				1

Implementation challenges

- Frameworks are largely command line
  - Not user friendly for many stakeholders
- Need to choose model type and federated learning approach that suits the datasets
- Federated learning approach influences the client node setup
- Need to be able to deal with varying deployment infrastructure
- Need for good representative synthetic data for initial model development
- Lack of understanding from data owners/custodians – still governance hassles!!!

Photo by Yogendra Singh on Unsplash

Future directions

- Security is the primary focus of current research
  - Framework security
  - Data security
  - Data re-identification
- Some approaches
  - Differential Privacy
  - Homomorphic encryption
  - Secure muti-party computation

- User friendly interfaces
- Federated model discovery and accessibility

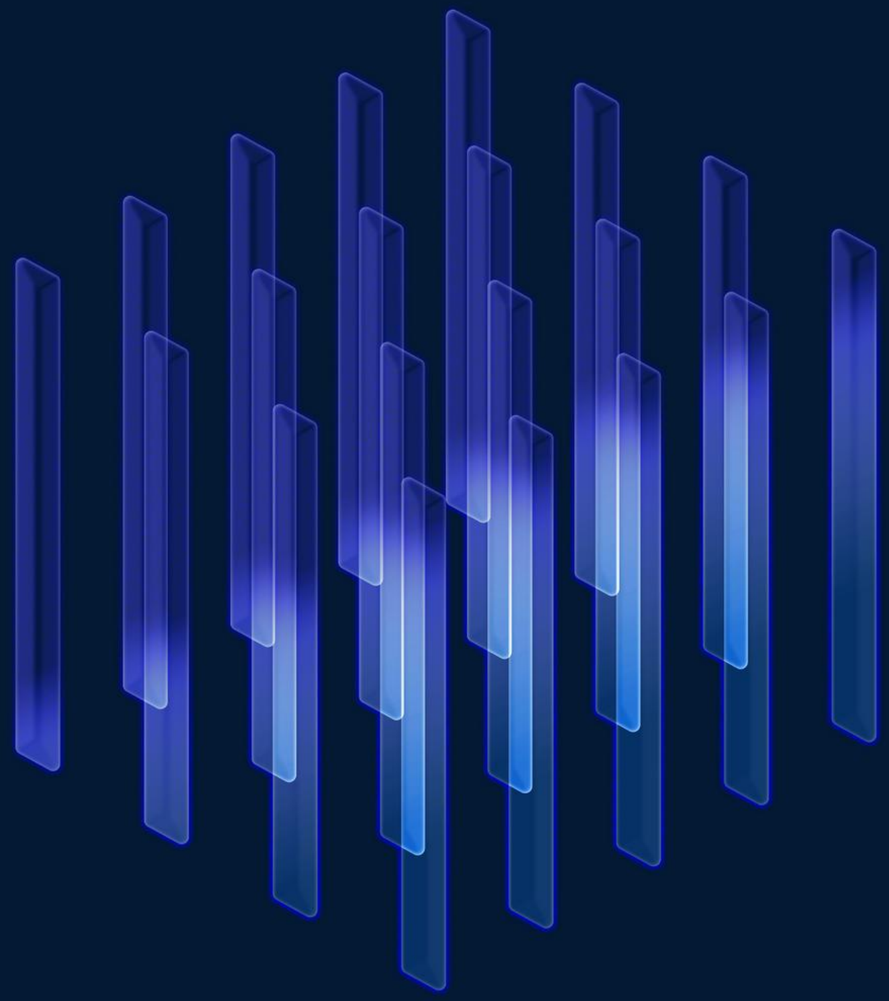
Photo by Stefan Steinbauer on Unsplash

Implementation challenges

- The approach bypasses traditional data governance challenges with fresh approaches to research using Federated Learning.
- By harmonizing data in situ and sending algorithms to the data, federated learning enables:

    - Owners of the data to analyse their own datasets while maintaining privacy principles
    - Data to remain in its place of origin, adhering to best practice privacy principles
    - Democratising AI expertise by empowering healthcare organizations to use FL
    - Overcoming data silo-related impediments and addressing privacy concerns across jurisdictions with different legislation

Photo by Giorgio Trovato on Unsplash



QCIF Digital Research³

Thank you!

Peter Marendy

peter.marendy@qcif.edu.au

linkedin.com/in/peter-marendy-64b4591/

0000-0002-5499-9863

www.qcif.edu