SYNTHEMA + ERN-EuroBloodNet

Joint Training Programme on Synthetic Data Generation in SCD and AML

Funded by the European Union

SYNTHEMA - Secure Federated Learning via P2P Secure Multi Party Computation

Dr. Sofia Tsekeridou, Netcompany S.A.

E-mail: sofia.tsekeridou@netcompany.com

Netcompany

#1

Webinar 2: Regulatory on Medical Devices, AI Act, Privacy-Preserving and Ethical Considerations

15/05/2025

Privacy in Federated Learning

Siloed Data Centres

- For **data and AI-driven health innovation acceleration**, rise **of Federated Learning**
- Particularly needed in **clinical contexts, heavily regulated** for personal/ sensitive data protection and privacy guarantees (GDPR, EHDS, data protection regulations, etc.)

Privacy in FL

- FL enhances **privacy and data protection by design**
- **No need to share data**, as algorithms move to data

Privacy Issues in FL

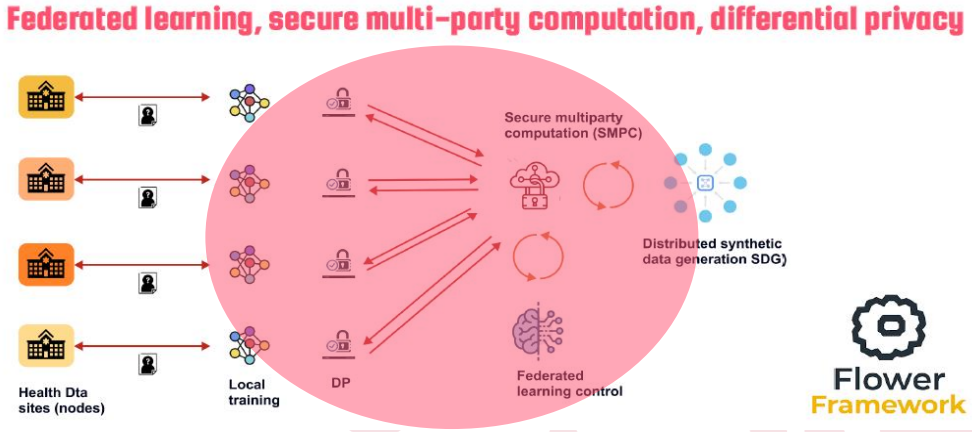
- FL susceptible to **data leakage** or **inference attacks** during **network communication** for **model updates & aggregation**
  - Membership attacks, gradient leakage, etc. (*)
- **Centralized aggregation**: single point of trust – what if aggregation server **not trusted**?

(*) J. Geiping et. al., "*Inverting Gradients - How easy is it to break privacy in federated learning*", 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020

PETs: Secure Multi-Party Computation (SMPC) Protocols to the Rescue

- Computation of functions (i.e. computation of the **model learned parameters**) over **shared secrets held** by **different peers without revealing** them directly
  - **Privacy-preserving aggregation of model updates**

- **Key properties:**
  - **Privacy**: No party learns others' raw secrets
  - **Correctness**: Mathematical guarantees that computation is performed faithfully
  - **Robustness**: provides guarantees even with some malicious participants

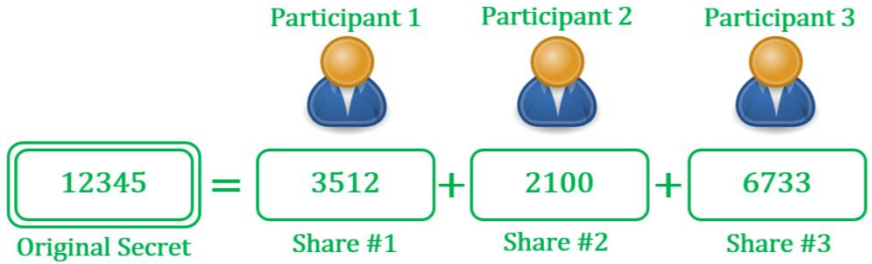


C. Zhao et al, "*Secure Multi-Party Computation: Theory, practice and applications*", Elsevier Information Sciences, Vol. 476, 2019, pp. 357-372.

Figure from Webinar 1 – F. Alvarez, "Infrastructure for SDG in ERN-EuroBloodNet"

SMPC Protocol Implementation

- **Additive secret sharing**
  - A cryptographic technique that **divides a secret value** into **multiple parts** (**shares**)
  - **Each share** of the original secret is **distributed to several participants**, ensuring that **control over the secret value is shared** and not held by a single entity.



Additive secret sharing
Image courtesy of geeksforgeeks.org

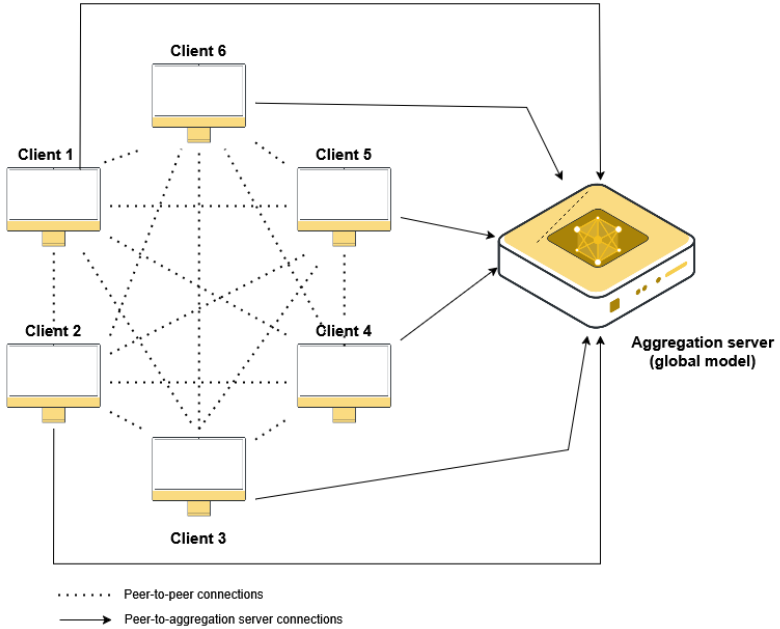
Issues with conventional SMPC-FL Implementations

- Each client **communicates directly** with the **Aggregation Server** after local client model training to **share model parameters**
    - Hence, most SMPC - Federated Learning implementations rely on **a centralized or semi-trusted aggregator**
    - This introduces a **single point of failure** and a **trust bottleneck**
    - It further exposes the system to **insider threats, collusion,** or **node compromise**.

Novel P2P SMPC Protocol within Flower FL framework (*)

- Leverages **additive secret sharing** and **peer-to-peer (P2P) message passing <u>across peers</u>** to construct a **<u>fully decentralized aggregation protocol</u>** compatible with existing FL workflows

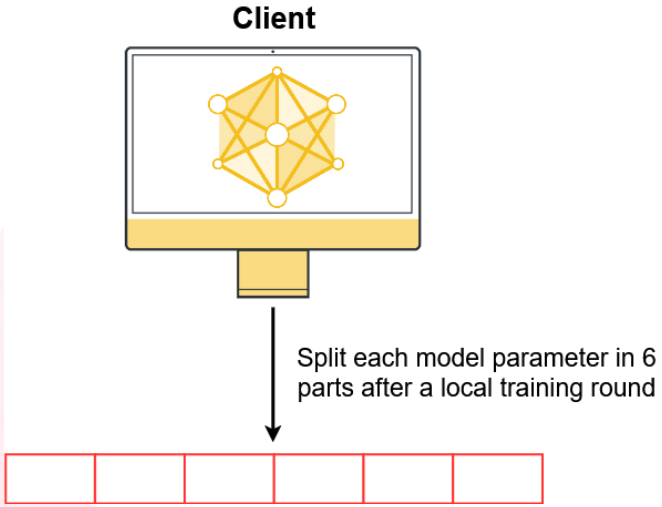
- **Each client** contributes to computation by **securely exchanging and combining secret shares** (i.e. the model learned parameters) **with its peers** and not only the Aggregation Server



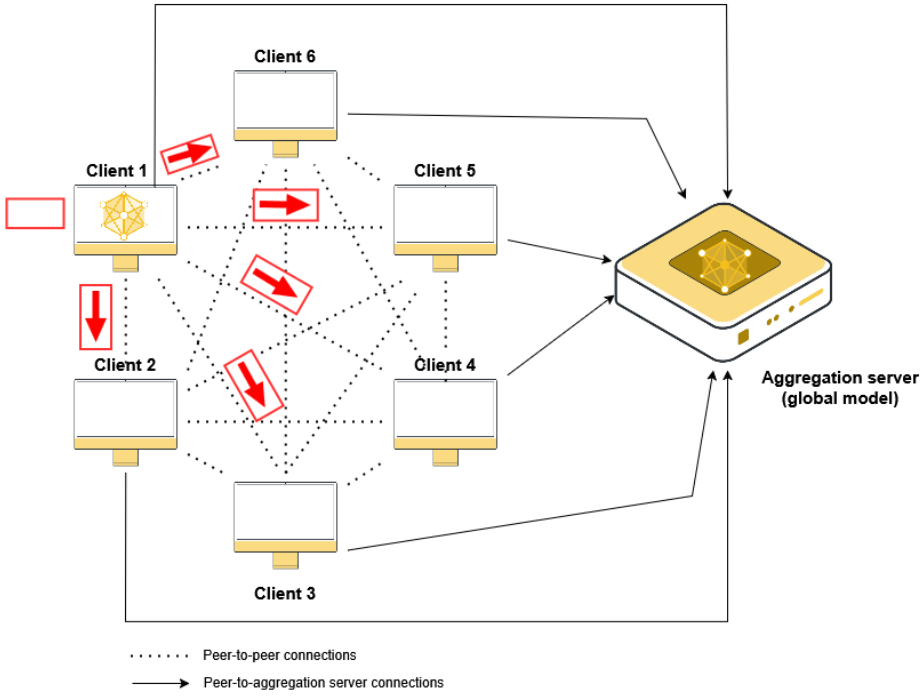
(*) P. Demetrakopoulos, T. Anagnostopoulos and S. Tsekeridou, "*Distributing trust: A P2P SMPC protocol for secure federated learning*," 2025 3rd International Conference on Foundation and Large Language Models (FLLM), Vienna, Austria, 2025, pp. 547-552, doi: 10.1109/FLLM67465.2025.11390879.
https://ieeexplore.ieee.org/document/11390879

Explaining the P2P SMPC Protocol – Assumption: 6 clients

Step 1: Client splits parameters after local training round in 6 parts

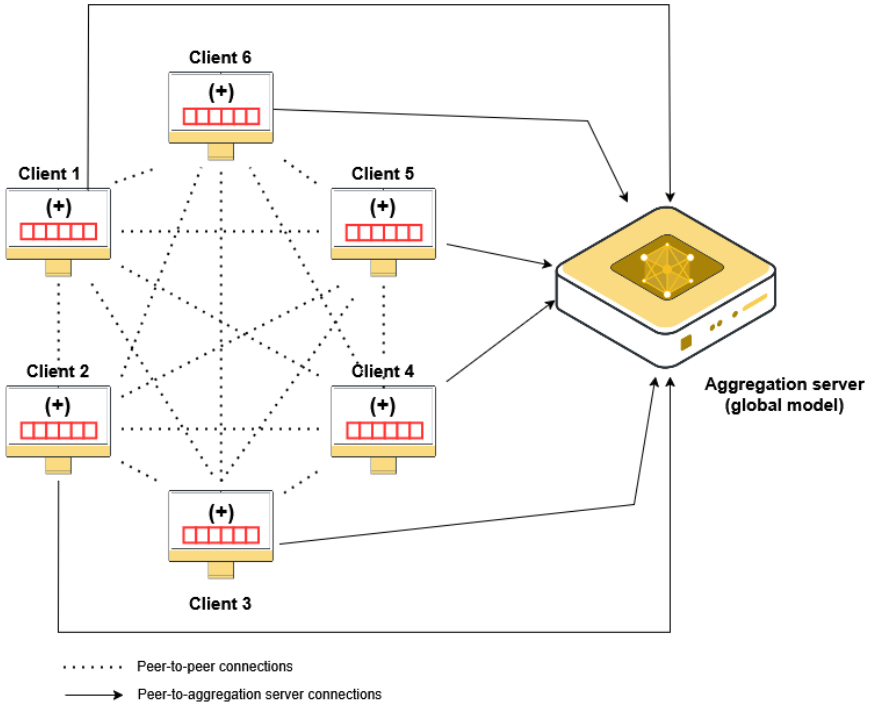


Step 2: clients distribute shares to all other clients in P2P mode

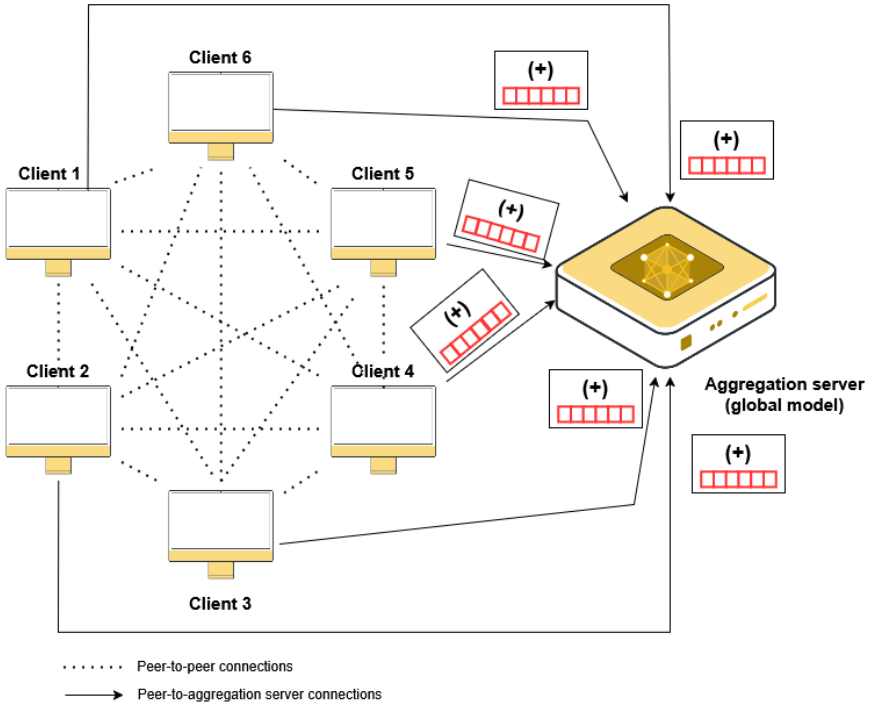


Explaining the P2P SMPC Protocol

Step 3: Each client holds one share per parameter from other peers. Local aggregation performed at each client

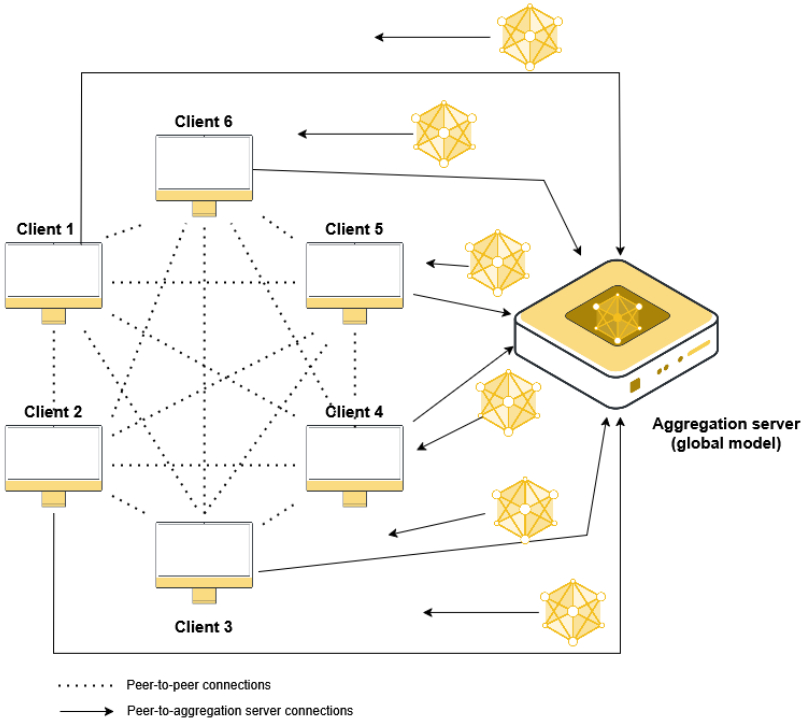


Step 4: Clients send locally aggregated parameters to the Central Aggregation Server

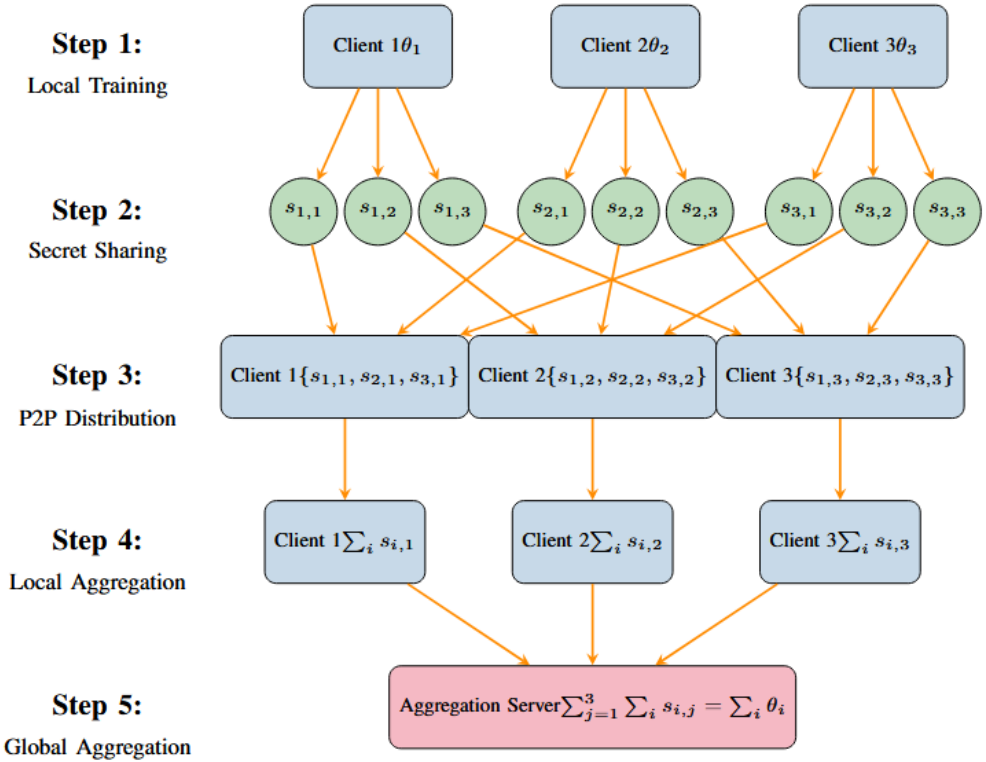


Explaining the P2P SMPC Protocol

Step 5: Aggregation server **aggregates all received local parameters**, computes **global model**, **sends it back to clients** and a **new training round** begins



Entire Process (3 clients)



Security and Privacy Guarantees

No individual share leaks model updates

- This protects the system from various kinds of attacks

Aggregation server **cannot infer the origin** (peer) of parameters

- Hence, it does **not need to be trusted / be semi-trusted**

Resilience to collusion

- Requires a **large number of compromised / malicious parties** to break security

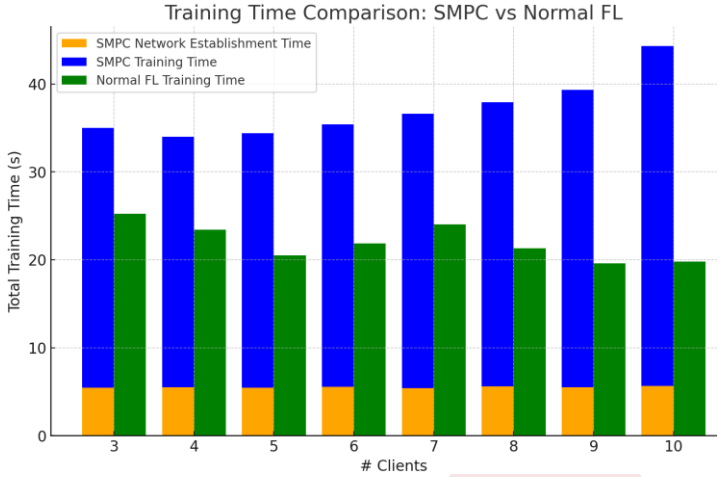
When combined with **differential privacy (as implemented in SYNTHEMA)**

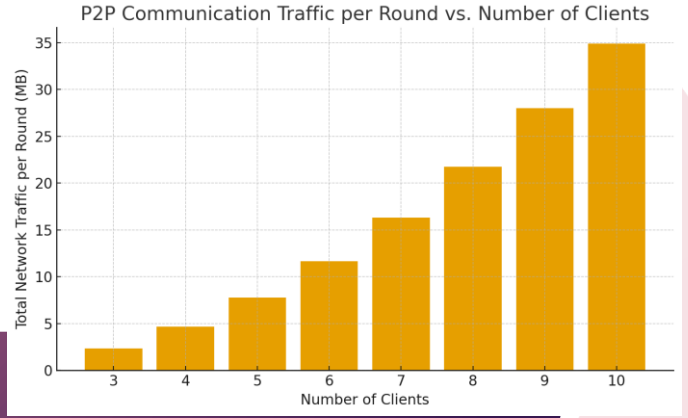
- **security and privacy** are significantly **enhanced**

Experimental Evaluation - Performance

- Two configurations:
  - **Standard FL**: Clients send plaintext model updates to centralized aggregator.
  - **P2P-SMPC FL**: Clients split model updates into additive secret shares, exchange them peer-to-peer, aggregate locally and send masked updates to centralized server.

- Performance metrics:
  - **Model accuracy and loss** over training rounds
    - Comparable accuracy achieved by both
  - **Training time per training round** as the number of clients increases for both configurations
    - P2P-SMPC configuration: consistent increase in runtime with more clients, attributed to P2P exchange of shares and need for synchronization.
  - **Total network traffic per round** as number of clients increase.
    - Total amount of data exchanged across network grows approximately quadratically with number of clients





Flower

Collaboration with Flower Labs

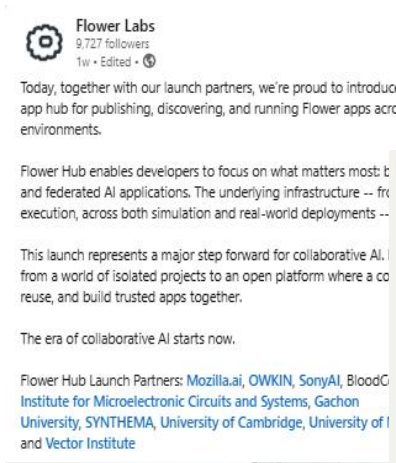
SYNTHEMA **Flower Next Pilot Application**

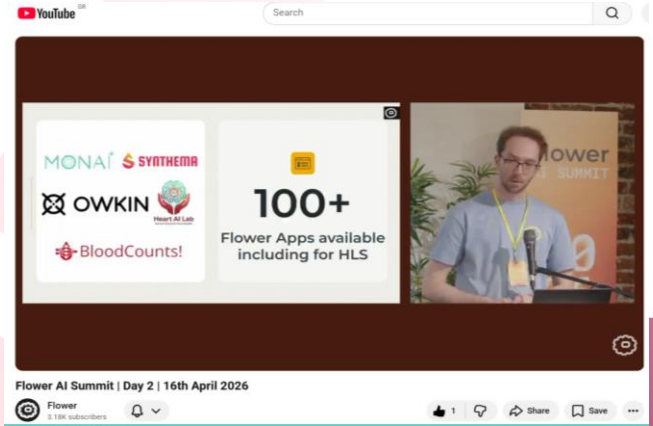
- started in January 2025

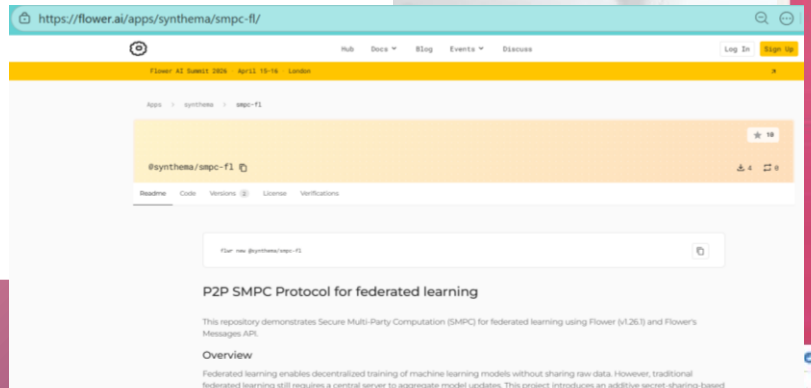
SYNTHEMA P2P SMPC Presentation at **Flower.AI Summit 2025**, London

SYNTHEMA **one of the early Flower Hub Launch partners**

- publishing the P2P open source SMPC App: @synthema/smpc-fl - Flower Hub







Other SYNTHEMA Related Work in PETs

- **Anonymization / Federated Anonymization** –Webinar 3 next Friday!

- SYNTHEMA as core use case and Netcompany as lead editor of **BDVA White Paper on Synthetic Data in Healthcare**
  - Healthcare - BDV Big Data Value Association

Thanks!

Any questions?

Keep in touch!

eurobloodnet.eu /ERNEuroBloodNet @ERNEuroBloodNet @erneurobloodnet.bsky.social

synthema.eu /synthema @SYNTHEMA_EU @synthemaeu.bsky.social

Funded by
the European Union

Acknowledgements

European Reference Network
for rare or low prevalence complex diseases
Network Hematological Diseases (ERN EuroBloodNet)

Funded by the European Union

This project is supported by the European Reference Network on Rare Haematological Diseases (ERN-EuroBloodNet)-Project ID No 101085717. ERN-EuroBloodNet is partly co-funded by the European Union within the framework of the Fourth EU Health Programme.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

SYNTHEMA

Funded by the European Union

SYNTHEMA is an initiative funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101095530.