Unlocking patient-level data at scale with federated computing to drive collaborative research and advance science

NVIDIA FLARE Event 2024

Eric Boernert,
Product Manager Federated Open Science, Roche Informatics

Today's speakers

Eric Boernert
Product Manager, Federated Open Science

Jacek Chmiel
Product Expert, Federated Open Science

Fahime Sheikhzadeh
Principal Imaging Scientist, Digital Pathology

Roche at a glance
Who we are and what we do

128 years
founded in Basel in 1896

A leader in healthcare R&D
with CHF 13.2 billion invested in 2023

3 Nobel prizes and **44 Prix Galien**
since 1974

CHF 58.7 billion*
in Roche Group sales in 2023

45 Roche medicines & 90 diagnostics**
on the WHO List of Essential Medicines & Tests

>100,000
dedicated employees worldwide

>22 million people
treated with our medicines in 2023

29 billion tests
conducted with our Diagnostics products in 2023

*Unless otherwise stated, all growth rates and comparisons to the previous year are at constant exchange rates (CER; average rates 2023) and all total figures quoted are reported in CHF.
** Medicines and tests that have either been developed or acquired by Roche

Introduction

Making data Findable, Accessible, Interoperable, Reusable (FAIR)

Deep, longitudinal data from broad patient populations

Moving the analytics/algorithm to the data while preserving privacy instead of moving the data to the analytics/algorithm

Unlocking patient-level data at scale with federated computing to drive collaborative research and advance science

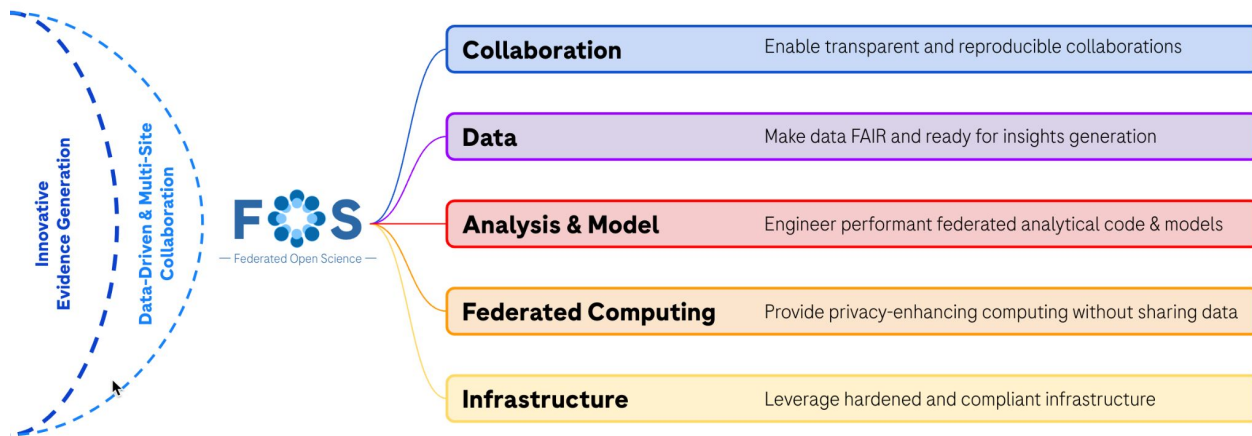Individual data separately recorded for each participant in a clinical study or in routine clinical practice

Building trust and bringing together researchers to generate new data-driven insights

Federated Open Science

#InvestingInTechForTomorrow

- Our focus is on next-generation and fit-for-purpose federation technology capability development and adoption for **real world healthcare settings** to:

  - Enable trusted and sustainable data collaborations among various stakeholders and across institutional/national borders

  - Increase data sovereignty for secure and data owner controlled access to meaningful, high-quality, multimodal and diverse patient-level RWD for secondary research purposes

Innovative Evidence Generation

Data-Driven & Multi-Site Collaboration

F✷S
— Federated Open Science —

Collaboration	Enable transparent and reproducible collaborations
Data	Make data FAIR and ready for insights generation
Analysis & Model	Engineer performant federated analytical code & models
Federated Computing	Provide privacy-enhancing computing without sharing data
Infrastructure	Leverage hardened and compliant infrastructure

Collaborative spirit in action

Strategic thought partnership

- [Preventing Health Data Leaks with Federated Learning Using NVIDIA FLARE](#) - creating networks of trust supported by technology (unsafe jobs vs. safe jobs)
- Resilience and debuggability of NVFlare solutions
- New messaging backend
- Federated XGBoost
- Homomorphic Encryption
- Differential Privacy + privacy meters
- Confidential Computing
- Federated LLMs support
- New data scientist experience feedback incl. ML to FL transition
- Integration with Flower
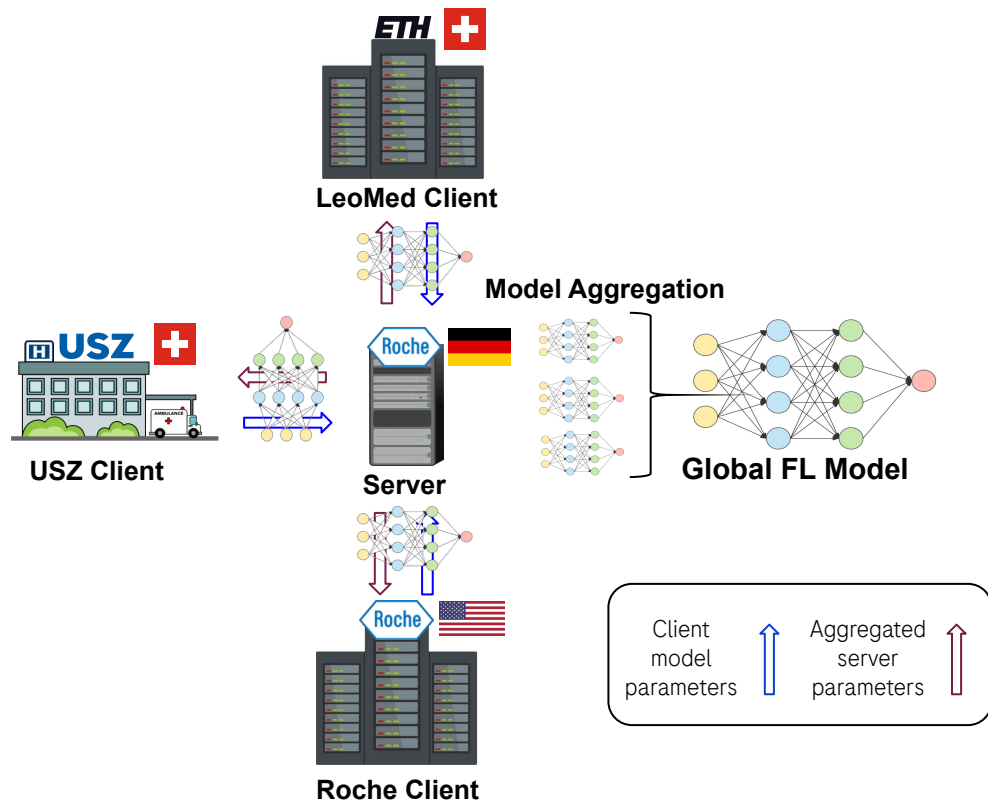- … and many more

Two-way exchange of design ideas

Valuable help and guidance

Enterprise viewpoint
→ meaningful feedback

Enthusiastic atmosphere with focus on moving forward

Use Case I

A Real-world Pilot Study with Roche Diagnostics Solutions, University Hospital Zurich & University of Zurich

LeoMed Client

Model Aggregation

USZ Client

Server

Global FL Model

Roche Client

Client model parameters   Aggregated server parameters

✅ Established Federated Learning Pipeline between Roche and USZ across different continents, using NVIDIA FLARE

The first ever real-world FL Pipeline involving Roche as a partner

Federated model learned data insights as effectively as centralized model
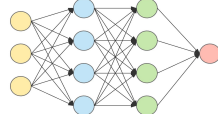

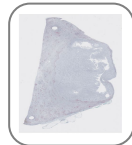## Centralized Model

TuPro Train All Dataset
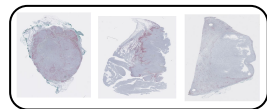
Server

CL Model

Test Score (TuPro Test Dataset)
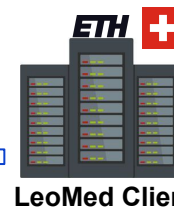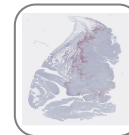Segmentation: **0.82**
Cell Detection: **0.77**

Federated Model

TuPro Train Dataset 1

TuPro Train Dataset 2
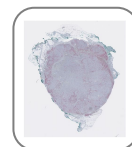
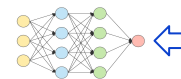LeoMed Client

USZ Client

Server

TuPro Train Dataset 3

Roche Client

FL Model

✅ **FL ~= CL**
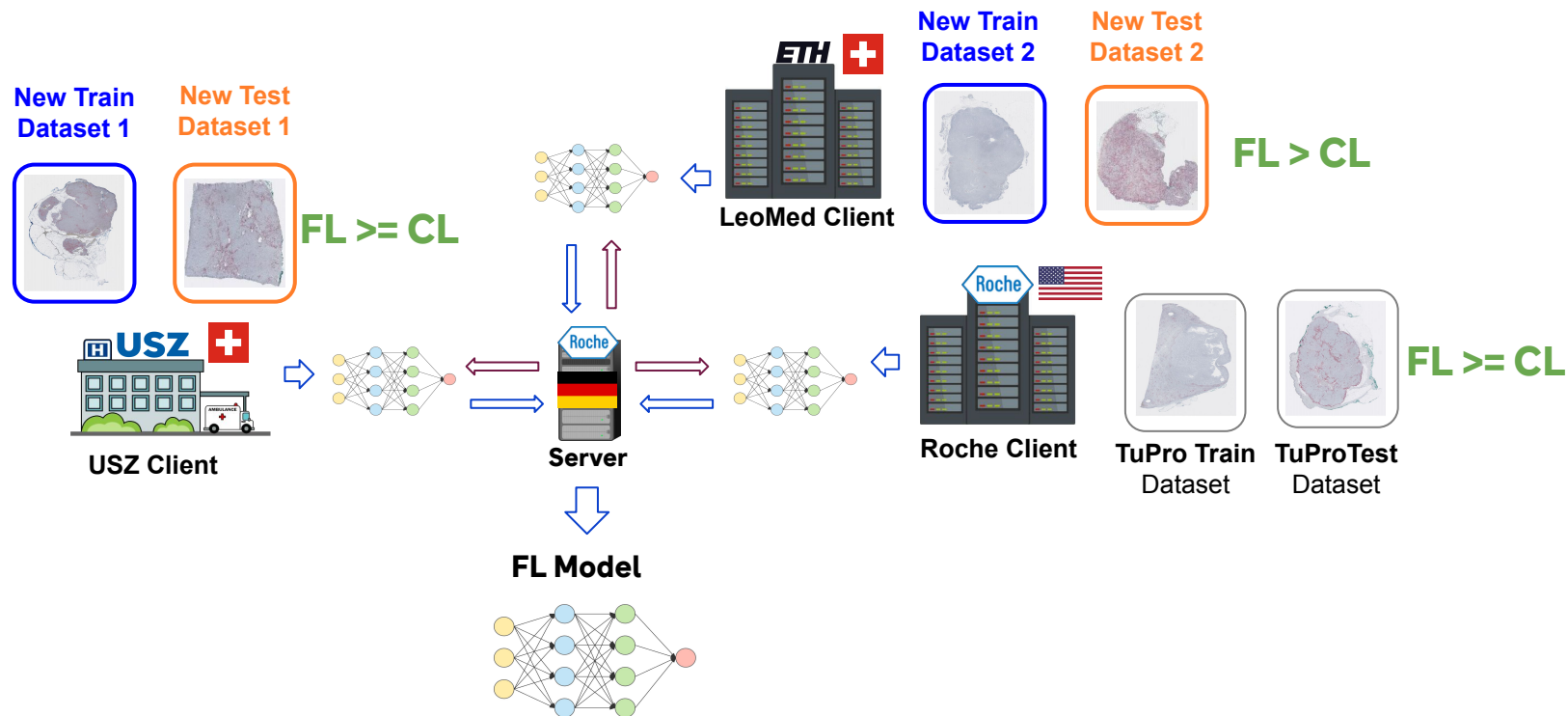
Test Score (TuPro Test Dataset)
Segmentation: **0.78**
Cell Detection: **0.78**

Federated model gained new insights from new client data sets which Roche has no access

✅ The FL model achieved better or equivalent performance across all three client test sets!

Use Case I
Conclusions

FL model achieves equal or better performance compared to the centralized model.

FL model generalizes better than the centralized model across clients

Challenges:

- Experiment duration
- Manual intervention ("babysitting") required
- Each client site requires interdisciplinary team (algorithm design, medical and IT infrastructure expertise)

Learnings:

- Real-world FL is possible, but challenging and relies on an interdisciplinary team!
- Optimize experiment duration first
- Investigate model performance on client datasets for guidance on domain alignment

Use Case II
INTONATE-MS academia-industry collaborative federated research network

- **In**nova**t**ive disc**o**veries in chronic **n**eurological diseases – coll**a**bora**t**ions in partn**e**rship

    - Heinz Wiendl, Univ.-Prof. Dr. med: ....*An innovative and visionary initiative in the area of academic (public) industry collaborations with a potential to revolutionise our usage and understanding of clinical and research data, biosamples and their respective analyses in neuroscience...*[1]

- Situation in Multiple sclerosis (MS)

    - MS is a disease of the central nervous system, estimated to affect 2.3 million people worldwide[2]
    - MS is a condition that has a great degree of variability across people and across time
    - There is no cure for MS yet and research continues to better understand and treat the disease

- Big data & Machine Learning in MS

    - Overcoming the complexity of the disease and its treatment, specifically in addressing challenging topics such as diagnosis, progression and personalized treatments
    - Heterogeneous disease course requires large-scale quality datasets incl. combining real-world with clinical trial data

Use Case II
From pathfinding to scaling

- Federated study 1
  - Enhance understanding and utility about novel concepts such as lack of disease stability in clinical practice and for new biomarker development
  - Applying a Federated Analysis on a combined cohort from real-world clinical settings and clinical trials

- Federated study 2
  - Towards automated and data-driven MRI analyses for monitoring disease and progression
  - i.a. Evaluating a high-precision brain lesion segmentation algorithm against real-world imaging data

- Federated study 3
  - Exploring the clinical and social-economic benefits of high treatment persistence in an international setting
  - Model development and validation investigating statistical inference in presence of heterogeneity in data distributions across sites

- Federated study 4
  - Explore the incidence of silent disease progression in a multi-center, multi-country, retrospective study
  - Overcoming limited patient pool challenge in MS and establishing faster, cheaper and sustainable access to a large-scale, heterogeneous and longitudinal patient cohort

Use Case III
EU IHI IDERHA

IDERHA Vision

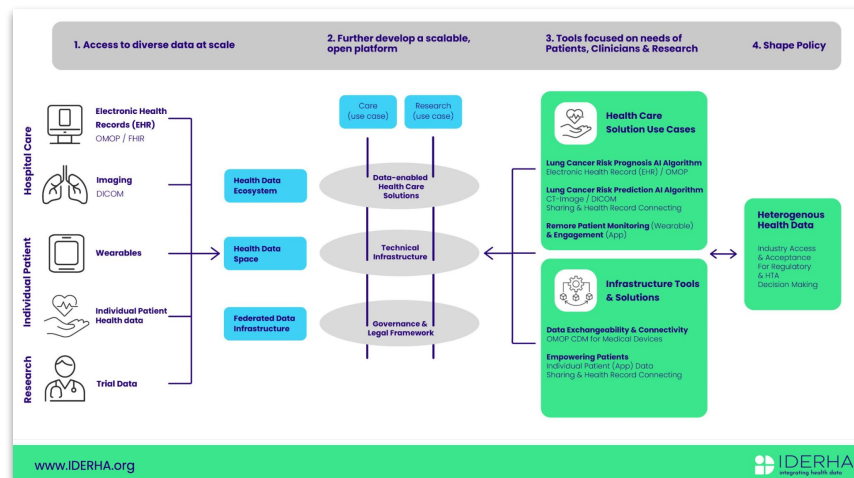*Integration of heterogenous Data and Evidence towards Regulatory & HTA Acceptance*

'In IDERHA we will be an open, disease agnostic, federated data space which enables connectivity, access, use and reuse of digital health data, and develop consensus policy recommendations on health data access and heterogeneous health research (e.g. RWE) for regulatory and HTA decision-making.'

IDERHA
integrating health data

🌐 www.iderha.org

in www.linkedin.com/company/iderha

IDERHA
integrating health data

1. Access to diverse data at scale
2. Further develop a scalable, open platform
3. Tools focused on needs of Patients, Clinicians & Research
4. Shape Policy

Hospital Care
- Electronic Health Records (EHR) — OMOP / FHIR
- Imaging — DICOM

Individual Patient
- Wearables
- Individual Patient Health data

Research
- Trial Data

Health Data Ecosystem

Health Data Space

Federated Data Infrastructure

Care (use case) — Research (use case)

Data-enabled Health Care Solutions

Technical Infrastructure

Governance & Legal Framework

Health Care Solution Use Cases
- Lung Cancer Risk Prognosis AI Algorithm — Electronic Health Record (EHR) / OMOP
- Lung Cancer Risk Prediction AI Algorithm — CT-Image / DICOM Sharing & Health Record Connecting
- Remote Patient Monitoring (Wearable) & Engagement (App)

Infrastructure Tools & Solutions
- Data Exchangeability & Connectivity — OMOP CDM for Medical Devices
- Empowering Patients — Individual Patient (App) Data Sharing & Health Record Connecting

Heterogenous Health Data
Industry Access & Acceptance For Regulatory & HTA Decision Making

www.IDERHA.org

13

Acknowledgements

Doing now what patients need next