CER-VD et APDI

Open Data

Sharing without Sharing – Privacy-Conscious Decentralized Data Analytics

Prof. Jean-Pierre Hubaux (EPFL)

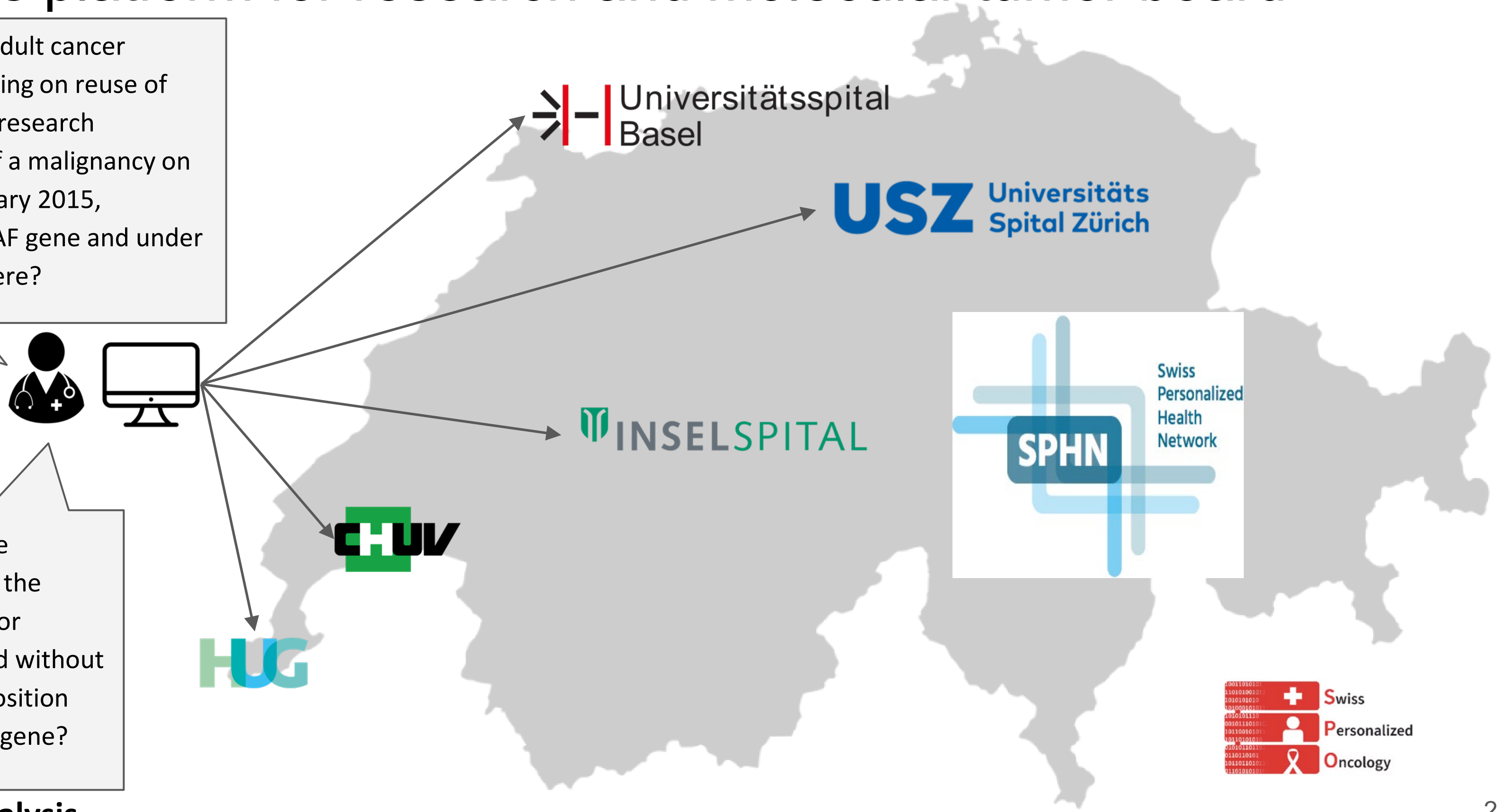
With gratitude to all the great colleagues and co-workers I have the privilege to collaborate with

7 October 2021

Use case for Swiss Personalized Oncology Project: federated analytics platform for research and molecular tumor board

Q1: How many adult cancer patients consenting on reuse of routine data for research with diagnosis of a malignancy on or after 1st January 2015, mutations in BRAF gene and under anti-PD-1 are there?

Explore



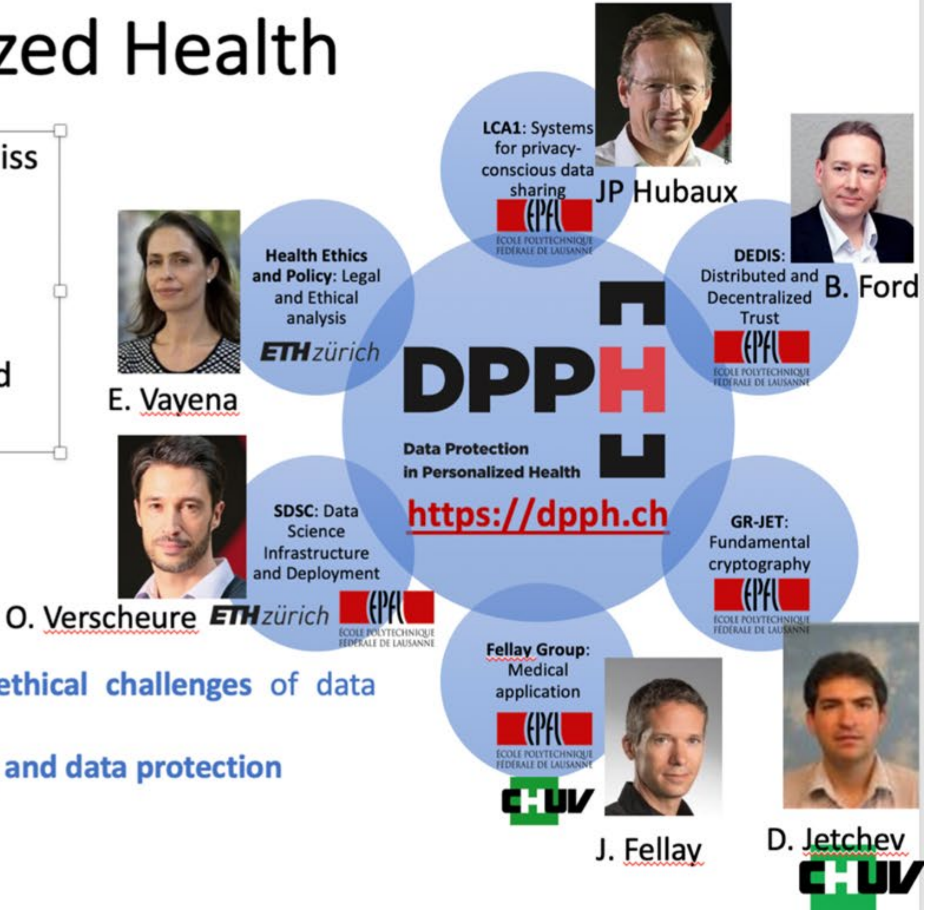
Q2: Among these patients, what is the overall survival for patients with and without a mutation on position 600 of the BRAF gene?

Analysis

DPPH – Data Protection in Personalized Health

- 5 research groups across the ETH domain + SDSC (Swiss Data Science Center)
- Funding: 3 Millions CHFrs
- Duration: 3 years (4/2018 - 12/2021)
- Funding Program: ETH PHRT (Personalized Health and Related Technologies)

Strategic Focus Area
Personalized Health and Related Technologies



Project goals:

- Address the main **privacy, security, scalability, and ethical challenges** of data sharing for enabling effective P4 medicine
- Define an optimal **balance between usability, scalability and data protection**
- Deploy an appropriate set of **computing tools**

EPFL

DPPH/MedCo - A highly interdisciplinary team

Core team

Leadership team


Prof. JP Hubaux
(Head of LDS, EPFL)


Dr. Juan Troncoso
(Senior researcher
LDS, EPFL)


Mickael Misbach
(Lead SW architect
LDS, EPFL)

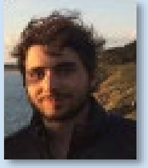

Nicolas Rosat
(Deputy CIO, CHUV)


Dr. JL Raisaro
(Data Science Lead, CHUV)

Development team


Dr. Francesco Marino
(Senior SW developer, LDS)


Joao Sa
(SW developer, LDS)


Jules Fasquelle
(SW developer, CHUV)


Nicolas Freundler
(SW developer, CHUV)

Hospital ambassadors


Prof. Alexandre Leichtle
(Inselspital)


Prof. Jacques Fellay
(CHUV/EPFL)


Dr. David Cavin
(HUG)


Solon Barraclough
(HUG)

SPO ambassadors


Prof. Olivier Michielin
(CHUV)


Dr. Michel Cuendet
(CHUV)

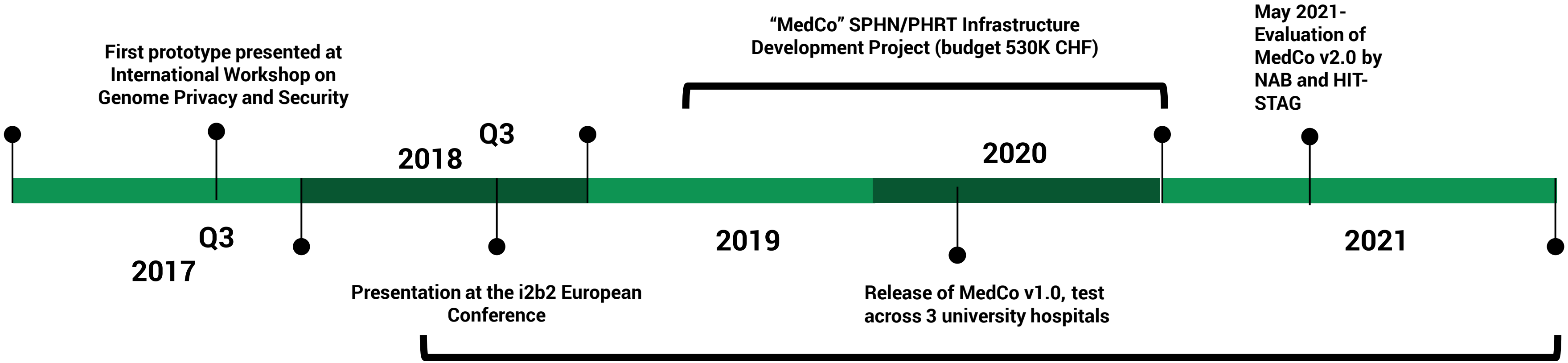

Dr. Sylvain Pradervand
(CHUV)

Unfolding

SPHN Swiss Personalized Health Network

Strategic Focus Area
Personalized Health
and Related Technologies

CHUV HUG INSEL EPFL



"Data Protection in Personalized Health (DPPH)" PHRT Project
(budget 3M CHF)

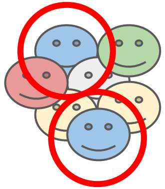
Strategic Focus Area
Personalized Health
and Related Technologies

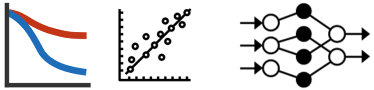
EPFL ETH zürich SDSC DPPH Data Protection in Personalized Health

MedCo in one slide...

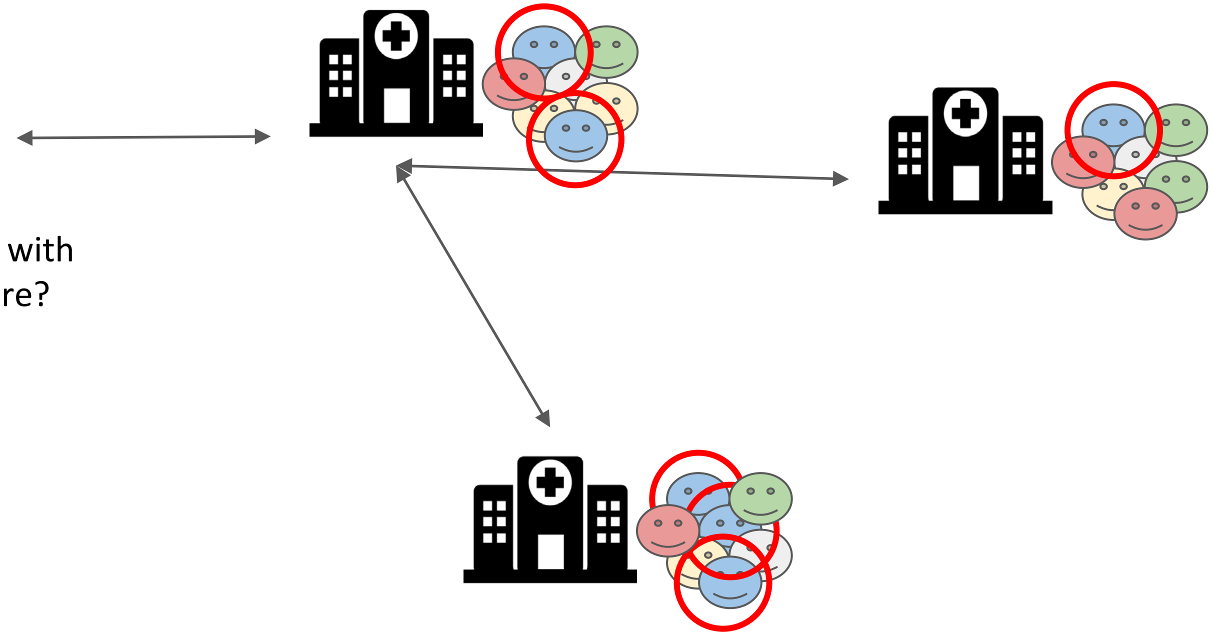
- **Distributed software platform** for federated cohort exploration and analytics of clinical and genomic data
- Co-developed by EPFL and CHUV
- Built on top of the i2b2 cohort explorer (i2b2 is used by 250+ hospitals worldwide)
- Relies on **advanced cryptographic techniques** → Multi-party homomorphic encryption (MHE)
- Code-reviewed and pen-tested by third-party industrial companies, compliant with hospitals' information security policies

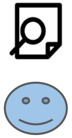
- Main functionalities
    - **MedCo-Explore: cohort exploration**
        - Obtaining cohort sizes for clinical research studies based on inclusion/exclusion criteria,
    - **MedCo-Analysis: federated analytics**
        - Survival analysis
        - ML training and testing (under development)





MedCo-Explore: cohort exploration/selection

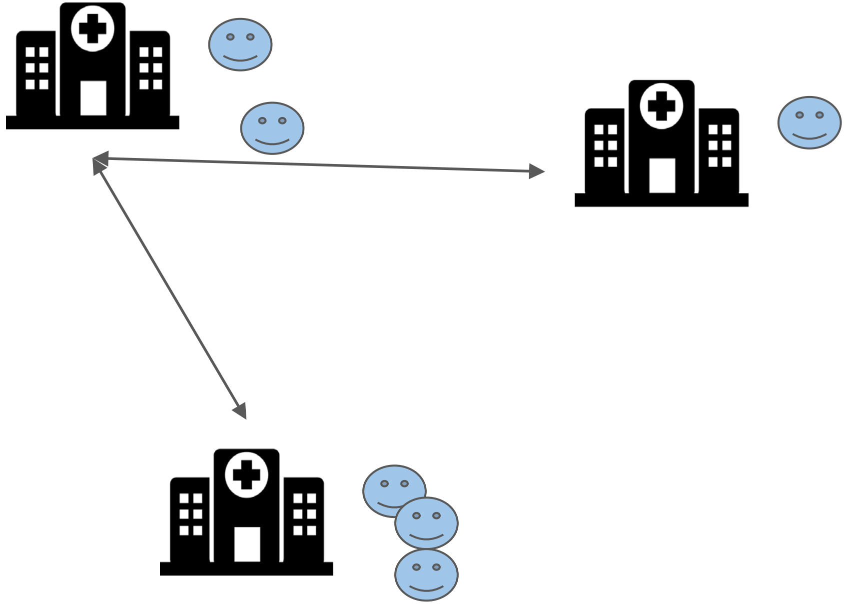




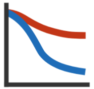
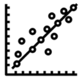
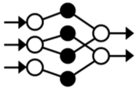
How many individuals with
these features are there?

Result ⇒ 6

MedCo-Analysis: distributed analytics



Distributed analysis on virtual cohort
(the data stay where they are):

- **Overall survival (Kaplan-Meier)** 
- **Linear/logistic regression** 
- **Neural networks** 

Positioning of MedCo with respect to similar distributed platforms

	Criteria \ Platform	CLINERION Real World Data Solutions	SHRINE	DataSHIELD	VANTAGE	HP	MedCo
Functionality	Federated cohort exploration	Yes	Yes	No	No	Yes	Yes
	Federated analytics	No	No	Yes	Yes	Yes	Yes
Privacy/Security	Protection of intermediate results and distributed computation	No	No	No	No	No	Yes
	Protection of end result from inference attacks	No	Yes	No	No	No	Yes
	Fine-grained role management	No	No	No	No	No	Yes
Usability	Graphical user interface	Yes	Yes	No	No	Yes	Yes
	Public API	No	Yes	Yes	Yes	Yes	Yes
	Service & support	Yes	No	No	No	No	Yes
	Extensibility	No	Yes	Yes	Yes	Yes	Yes

GDPR legal compliance: partial aggregates are not personal data anymore, they are anonymous

JMIR Publications
Advancing Digital Health & Open Science

Articles | Search articles

Journal of Medical Internet Research | Journal Information | Browse Journal | Subm

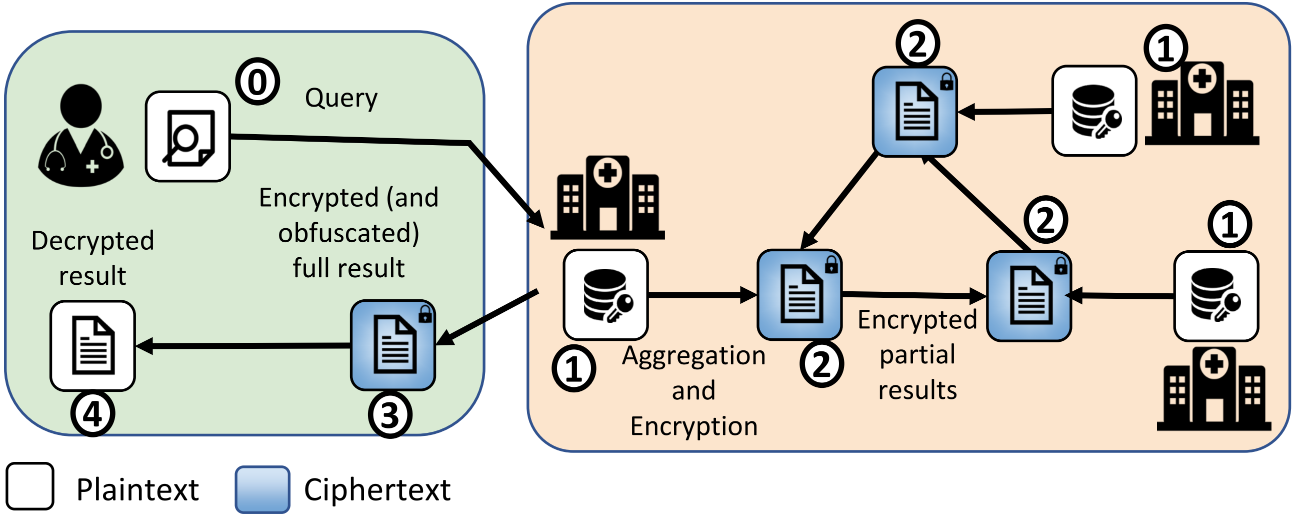
Published on 25.2.2021 in Vol 23, No 2 (2021): February

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/25120, first published October 19, 2020.

Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis

James Scheibner [1,2]; Jean Louis Raisaro [3,4]; Juan Ramón Troncoso-Pastoriza [5]; Marcello Ienca [1]; Jacques Fellay [3,6,7]; Effy Vayena [1]; Jean-Pierre Hubaux [5]

Legal qualification of data processed through MedCo



- Data transferred/processed with MedCo can be considered **anonymized data**
- **No need of bilateral data transfer agreements** between institutions to perform federated analytics
- **Patient consent not required for MedCo-Explore** (out of scope of HRA)
- Patient consent might not be required for MedCo-Analysis (under some circumstances)

	1 : Local data	2 : Encrypted partial results	3 : Encrypted full result	4 : Decrypted full result
Data status	Individual-level data	Locally aggregated	Globally aggregated	Globally aggregated
Legal qualification	Personal data	Anonymized data	Anonymized data	Anonymized (if proper protection in place) Personal data (if proper protection not in place)

Data Protection Impact Assessment (DPIA) for multisite medical data analysis (June 2021)

Centralized approach with standard pseudonymization

Threat	Threat likelihood	Threat impact	Risk	Risk level
Unlawful access to the system	Unlikely	High	Loss of data confidentiality	Moderate
Malicious use of the system	Possible	High	Loss of data confidentiality	High
Loss of data	Unlikely	Minor	Loss of data integrity, data unavailability	Minor
Data leak of host/cloud	Possible	High	Loss of data confidentiality	High
Collusion of host/cloud	Possible	High	Loss of data confidentiality	High
Corrupted or malicious host/cloud	Possible	High	Data unavailability, loss of data integrity, loss of data confidentiality, loss of data correctness	High
Unavailability of host/cloud	Possible	Minor	Data unavailability, loss of data correctness	Moderate
Re-identification/attribute inference	Possible	High	Loss of data confidentiality	High

Federated approach enhanced **with MedCo**

Threat	Measure introduced with MedCo	Threat likelihood	Threat Impact	Risk	Risk level
Unlawful access to the system	1	Unlikely	Minor	Loss of data confidentiality	Low
Malicious use of the system	1, 2, 4, 10	Possible	Minor	Loss of data confidentiality	Low
Loss of data	3, 5	Unlikely	Minor	Loss of data integrity, data unavailability	Low
Data leak	4, 5, 8, 9, 10	Unlikely	Minor	Loss of data confidentiality	Low
Collusion between nodes	4, 9	Unlikely	Moderate	Loss of data confidentiality	Moderate
Corrupted or malicious nodes	2, 5, 6, 7, 8, 9	Unlikely	Moderate	Data unavailability, loss of data integrity, loss of data confidentiality, loss of data correctness	Moderate
Unavailability of of nodes	6, 7	Possible	Minor	Data unavailability, loss of data correctness	Moderate
Re-identification or attribute inference	1, 2, 4, 9, 10	Unlikely	Minor	Loss of data confidentiality	Low

The DPIA was elaborated notably with the help of Valérie Junod and Sylvain Métille

Feedback from EDÖB on MedCo DPIA

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Data Protection and Information Commissioner

“... the threat impact of most risks with the MedCo system shows to be clearly lower than with traditional systems. Since data processed within the Medco framework remain encrypted at rest and during computation, an attacker would cause little damage. **As no entity has the full decryption key, it seems indeed unlikely that he could decrypt and abuse the stolen data**. (...)”

13 September 2021

The “Holy Grail” for SPHN secure federated analytics: FAMHE

Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption

David Froelicher[1], Juan R. Troncoso-Pastoriza[1], Jean Louis Raisaro[2,3], Michel A. Cuendet[4], Joao Sa Sousa[1], Hyunghoon Cho[5], Bonnie Berger[5,6,7], Jacques Fellay[2,8], and Jean-Pierre Hubaux[1,*]

[1]Laboratory for Data Security, EPFL, Lausanne, Switzerland
[2]Precision Medicine Unit, Lausanne University Hospital, Lausanne, Switzerland
[3]Data Science Group, Lausanne University Hospital, Lausanne, Switzerland
[4]Precision Oncology Center, Lausanne University Hospital, Lausanne, Switzerland
[5]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[6]Computer Science and AI Laboratory, MIT, Cambridge, Massachusetts, USA
[7]Department of Mathematics, MIT, Cambridge, Massachusetts, USA
[8]School of Life Sciences, EPFL, Lausanne, Switzerland
[*]jean-pierre.hubaux@epfl.ch

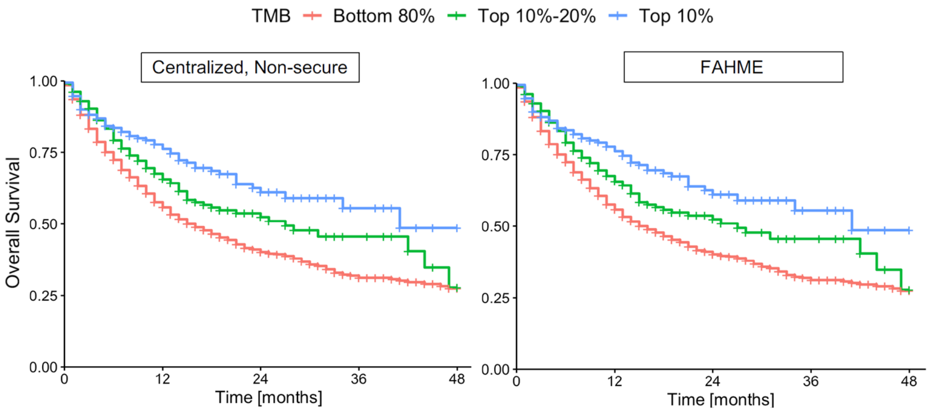
Accepted for publication in Nature Communications
→ will appear on 11 October

https://doi.org/10.1101/2021.02.24.432489

- Idea: train and run ML models on decentralized datasets without “seeing” the data
- Initially CHUV + EPFL, then Broad Inst. + MIT researchers joined the effort

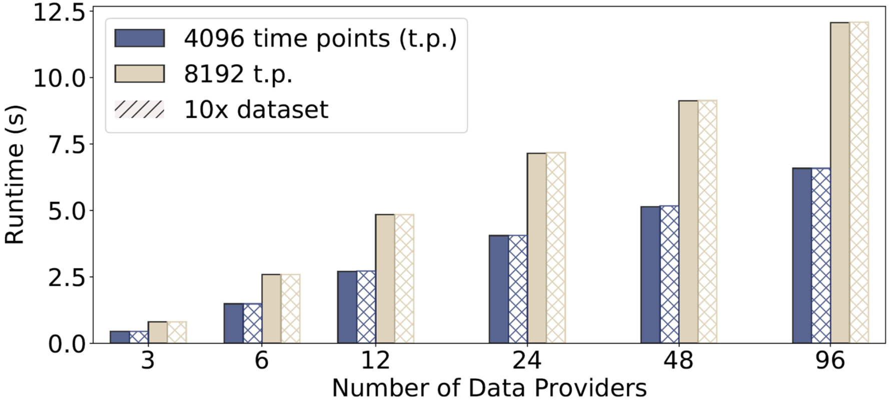
FAHME: Privacy-Preserving Federated Analytics for Precision Medicine with MHE - Survival curves (Kaplan-Meier)

Data split among 3 data providers:



Number at risk

Time [months]	0	12	24	36	48
Bottom 80%	1302	584	230	85	33
Top 10%-20%	186	103	41	17	4
Top 10%	172	99	41	13	4



[Centralized] Samstein, R. M. et al. Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. Nat. genetics 51, 202–206 (2019).
[FAHME] Froelicher et al. Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption.

FAHME: Privacy-Preserving Federated Analytics for Precision Medicine with MHE - GWAS

Use case: 1857 patients spread among 12 data providers.

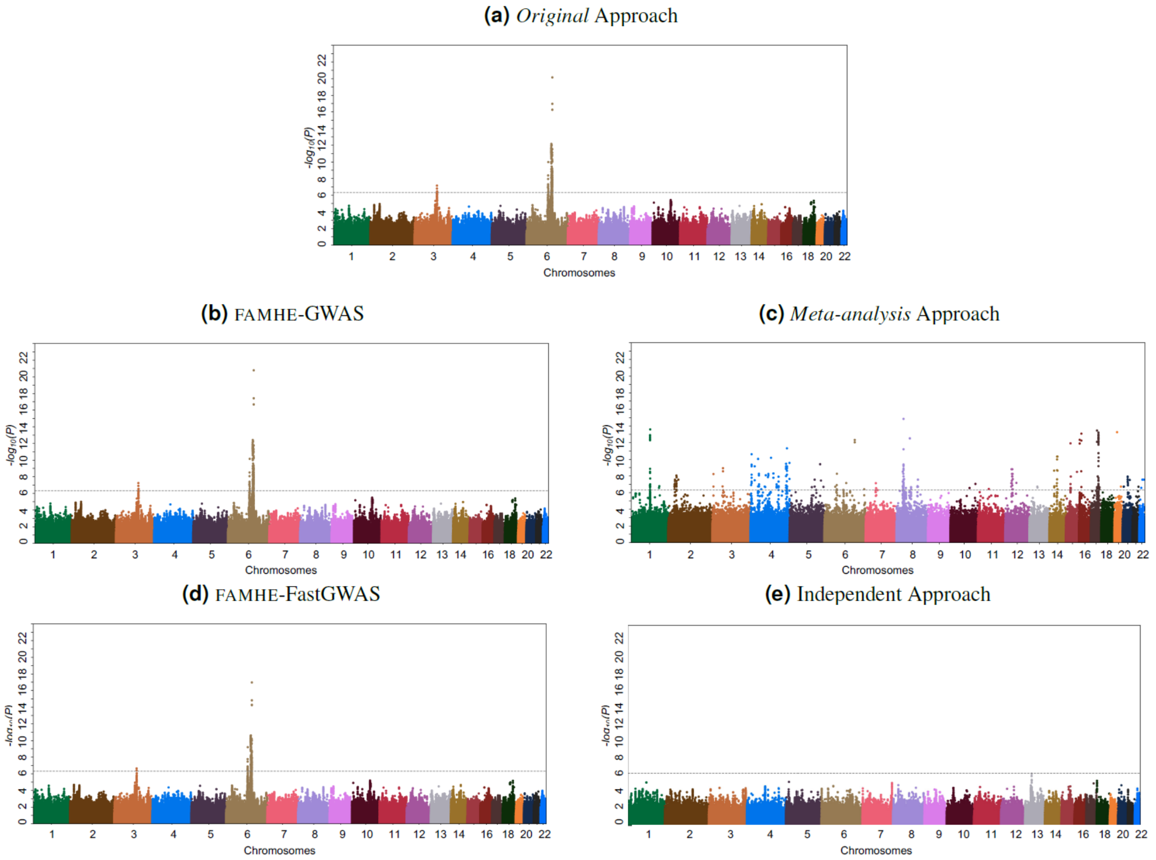
Original = Centralized approach

FAMHE-GWAS = Exact secure federated approach

FAMHE-FastGWAS = Efficient secure federated approach

Meta-analysis = distributed non-secure non-iterative approach

Independent = 1 DP alone



[Original approach] McLaren, P. J. et al. Polymorphisms of Large Effect Explain the Majority of the Host Genetic Contribution to Variation of HIV-1 Virus Load. Proc. Natl. Acad. Sci. 112, 14658–14663 (2015).

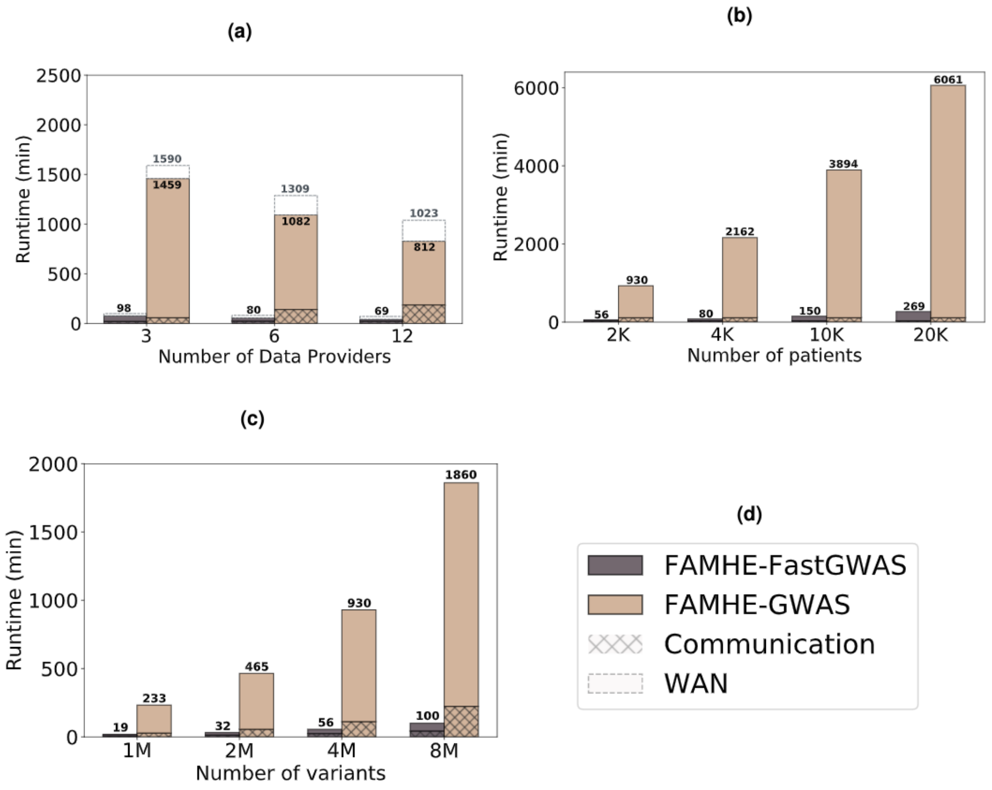
[FAHME] Froelicher et al. Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption.

FAHME: Genome-wide association study

Default: 1857 patients spread among 12 data providers.

→ **scale in all dimensions**

a. *With the number of data providers*
b. *With the number of patients*
c. *With the number of variants*



[Centralized] McLaren, P. J. et al. Polymorphisms of Large Effect Explain the Majority of the Host Genetic Contribution to Variation of HIV-1 Virus Load. Proc. Natl. Acad. Sci. 112, 14658–14663 (2015).
[FAHME] Froelicher et al. Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption. Submitted to Nat. comms. https://doi.org/10.1101/2021.02.24.432489

Enterprise Data & Analytics

$100B Market

(GlobalData, Enterprise Data and Analytics Market, 2020))

56%

say own data is not enough, expanding ability to source external data

(Forrester, The Insights Professional's Guide To External Data Sourcing, 2019)

However, organizations are **prevented** to enter valuable data collaborations due to fear of **data leaks** and **data protection regulations**

TUNE INSIGHT

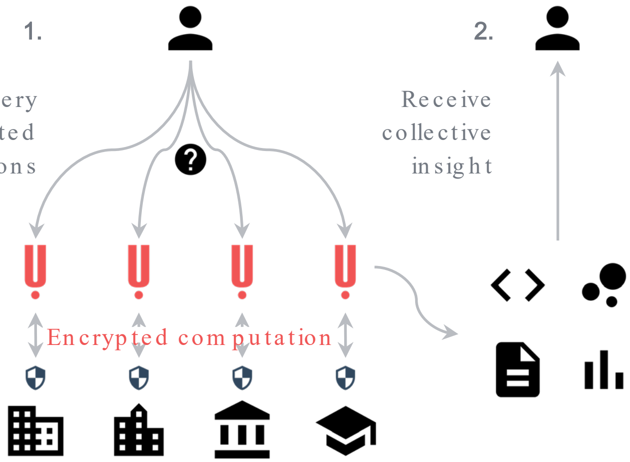
Cross-vertical enterprise SaaS enabling organizations to make better decisions, together, by orchestrating secure collaborations around their sensitive data.

- Incubating at Lab for Data Security (LDS)
- CHF400k in customer-paid projects including with Swiss Re, Armasuisse
- Pilot deployed at Swiss hospitals
- CHF100k EPFL Innogrant
- State-of-the-art post-quantum encryption technology

1. Query connected organisations

2. Receive collective insight

3. Apply new knowledge



Better patient treatment
More personalized services

More accurate premiums
Better risk estimation

Better market predictions
Reduced fraud losses

Access to insights

Immediacy

Scalability

Compliance

start up ticker .ch

NEWS CALENDAR ASSETS INTERACTIVE

TUNE INSIGHT

Tune Insight secures pre-seed round from Wingman Ventures

22.09.2021

> FINANCING

More news about

Tune Insight SA >

1

Tune Insight B2B software enables organizations to make better decisions by collaborating securely on their sensitive data to extract collective insights. Incubated at the EPFL Laboratory for Data Security, with a deployment in Swiss university hospitals and customer-funded projects in the insurance and cybersecurity businesses, Tune Insight will use the funds to accelerate product development, strengthen the team, and onboard more customers.

Conclusion

Achievements:

- We have solved the problem of GDPR-compliant federated analytics for medical data
- We provide MedCo, a fundamental building block for SPHN and beyond
- In September 2021, SPHN gave green light for further experimental deployments of MedCo
- Worldwide leading project on secure, privacy-conscious medical data sharing
- Solution for economic viability beyond 2021: Tune Insight

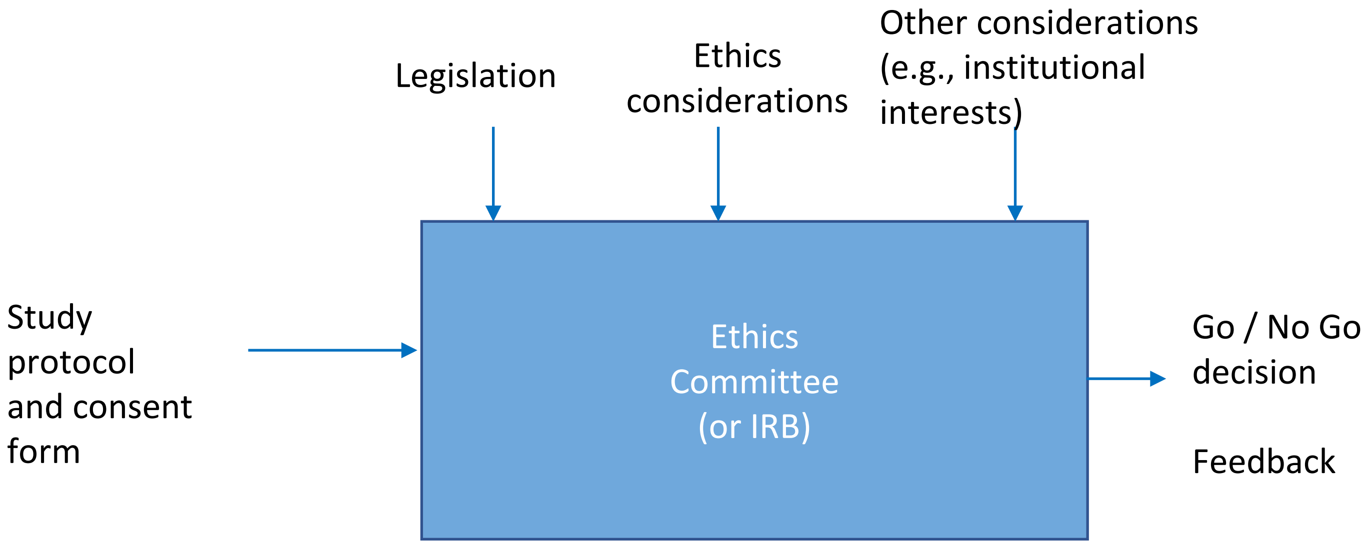
Ongoing and future steps:

- Assessment by CER-VD and APDI
- Transfer of people and know-how from EPFL to Tune Insight
- Further deployments in Swiss hospitals and beyond

https://dpph.ch https://medco.epfl.ch

Back-up slides

Ethics Committees

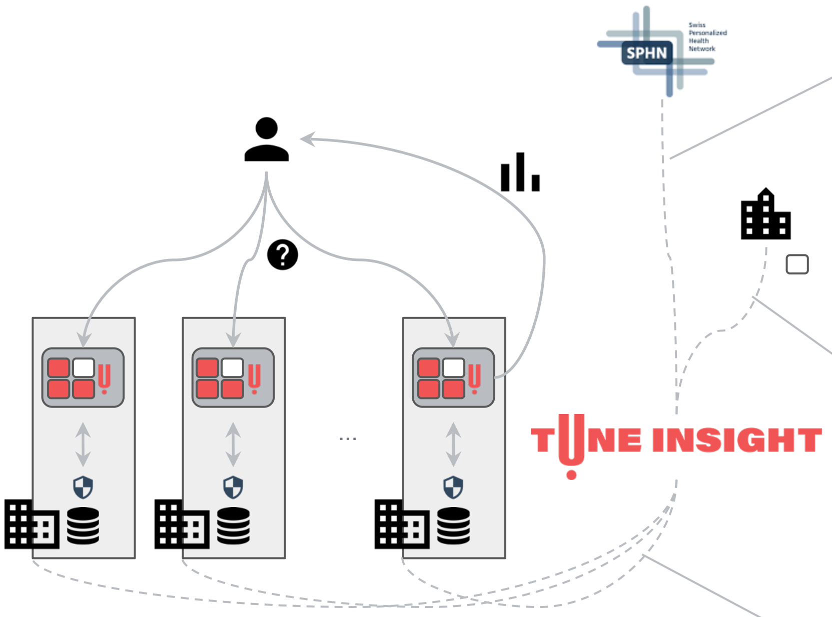


- Very slow, manpower-hungry and tedious process to check the proposed data-protection measures
- Need to obtain informed consent; diversity of consent forms
- Ethics committees make an on-paper *a priori* evaluation, with little control on what happens afterwards
- Risk of “race to the bottom”: the researchers that obtain permissions to see more data will extract more value → competitive advantage

ñ ļ ¡ >ή ЁЎ ļ P Ж О ή¡ ļ Ex ЎФЍ‡ЬPļ uPЖЎ ЍФ“ !

Legend

- Authorized User
- Query
- Third Party
- SPHN Hospital Environment
- ! ſ ſ Ў ſ %Ф¡ ..ļ ЖБЮ
- Third Party Module
- Secure Modules (certified by Tune Insight)
- Tune Insight Secure Platform (Open APIs)



Project contract(s)*

- Project-specific developments
- Health-specific platform modules

CLA contract*

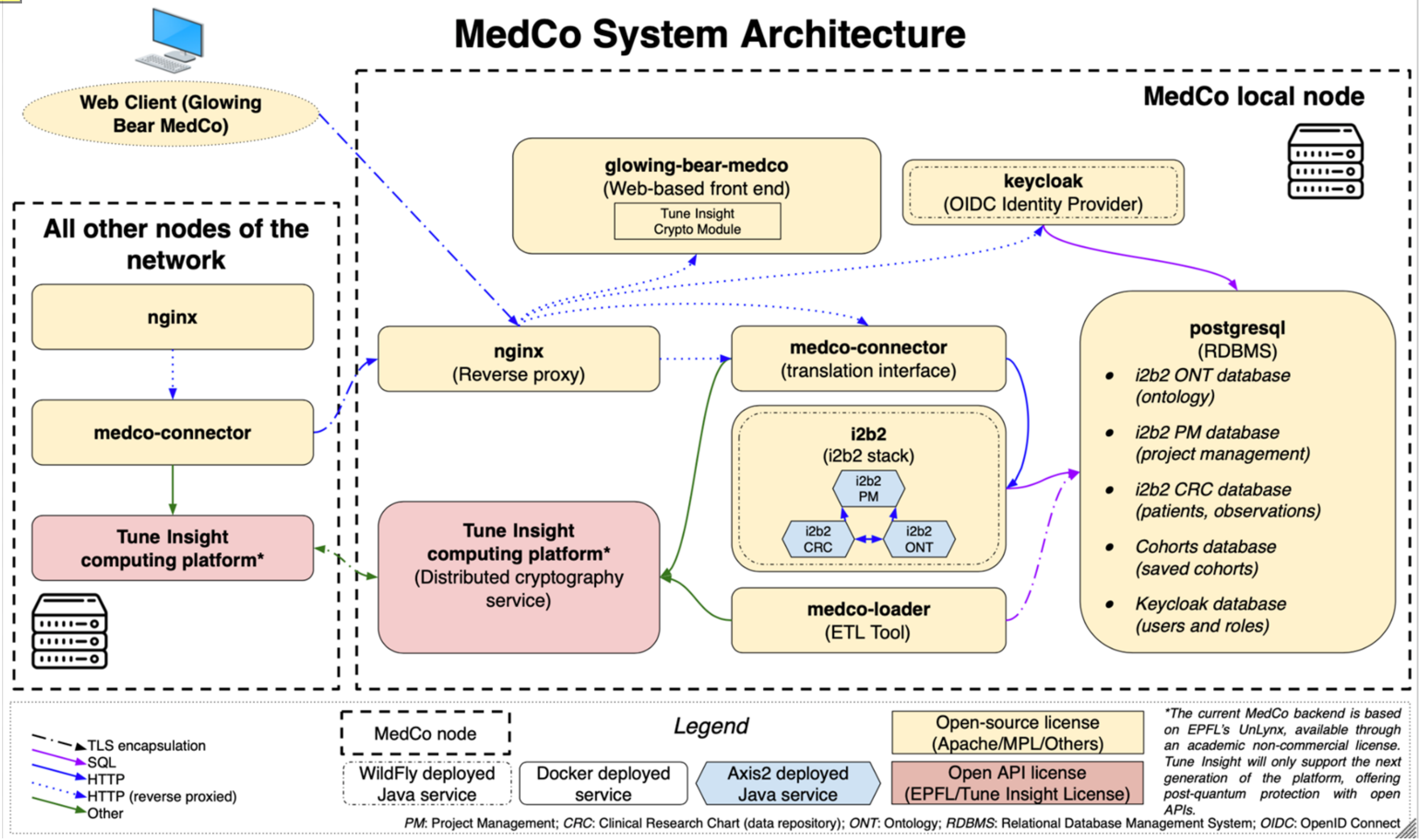
- Third party developed modules
- Security/privacy certification by
- Potential platform Integration by

Use license contract

- Support
- Maintenance
- Updates
- Upgrades (new ML functionalities)

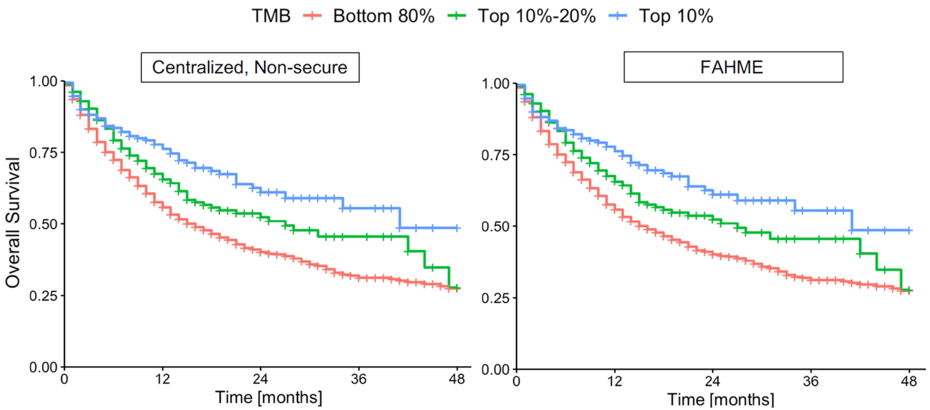
* Optional

Third Party: Non-Tune Insight and non-EPFL developers of platform modules. Examples: CHUV, Insel,...
Third Party Module: Statistical computation modules that are registered and deployed on top of the platform, but not certified or coded by Tune Insight.


MedCo System Architecture
Web Client (Glowing Bear MedCo)
MedCo local node
glowing-bear-medco
(Web-based front end)
Tune Insight Crypto Module
keycloak
(OIDC Identity Provider)
All other nodes of the network
nginx
medco-connector
Tune Insight computing platform*
nginx
(Reverse proxy)
medco-connector
(translation interface)
i2b2
(i2b2 stack)
i2b2 PM
i2b2 CRC
i2b2 ONT
Tune Insight computing platform*
(Distributed cryptography service)
medco-loader
(ETL Tool)
postgresql
(RDBMS)
• i2b2 ONT database (ontology)
• i2b2 PM database (project management)
• i2b2 CRC database (patients, observations)
• Cohorts database (saved cohorts)
• Keycloak database (users and roles)
Legend
TLS encapsulation
SQL
HTTP
HTTP (reverse proxied)
Other
MedCo node
WildFly deployed Java service
Docker deployed service
Axis2 deployed Java service
Open-source license (Apache/MPL/Others)
Open API license (EPFL/Tune Insight License)
*The current MedCo backend is based on EPFL's UnLynx, available through an academic non-commercial license. Tune Insight will only support the next generation of the platform, offering post-quantum protection with open APIs.
PM: Project Management; CRC: Clinical Research Chart (data repository); ONT: Ontology; RDBMS: Relational Database Management System; OIDC: OpenID Connect

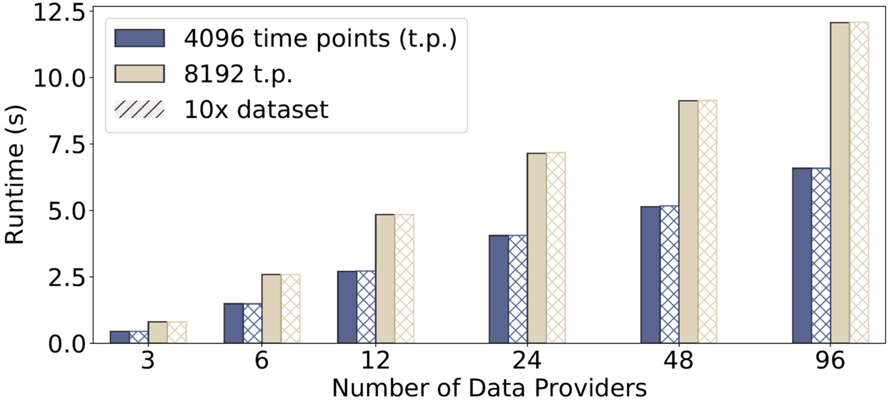
FAHME: Privacy-Preserving Federated Analytics for Precision Medicine with MHE - Survival curves (Kaplan-Meier)

Data split among 3 data providers:



Number at risk

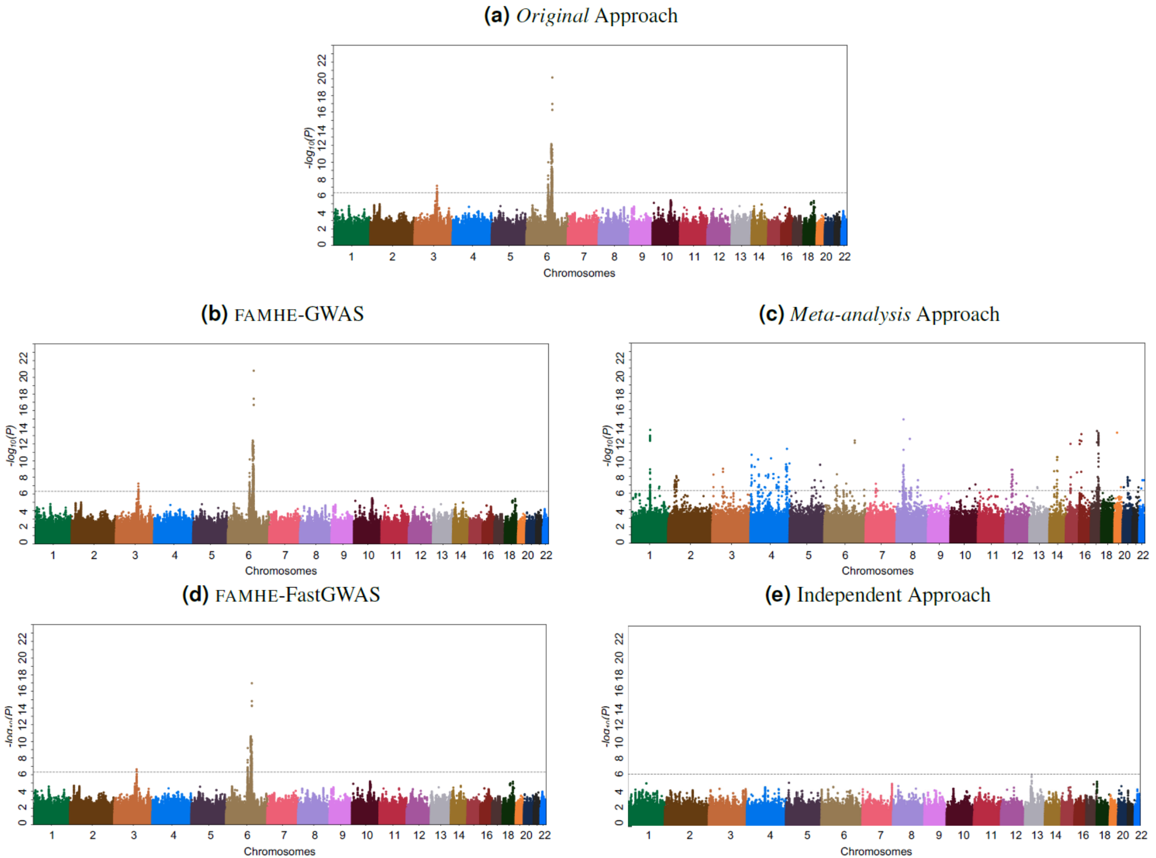
Time [months]	0	12	24	36	48
Bottom 80%	1302	584	230	85	33
Top 10%-20%	186	103	41	17	4
Top 10%	172	99	41	13	4



[Centralized] Samstein, R. M. et al. Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. Nat. genetics 51, 202–206 (2019).

[FAHME] Froelicher et al. Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption.

FAHME: Privacy-Preserving Federated Analytics for Precision Medicine with MHE - GWAS

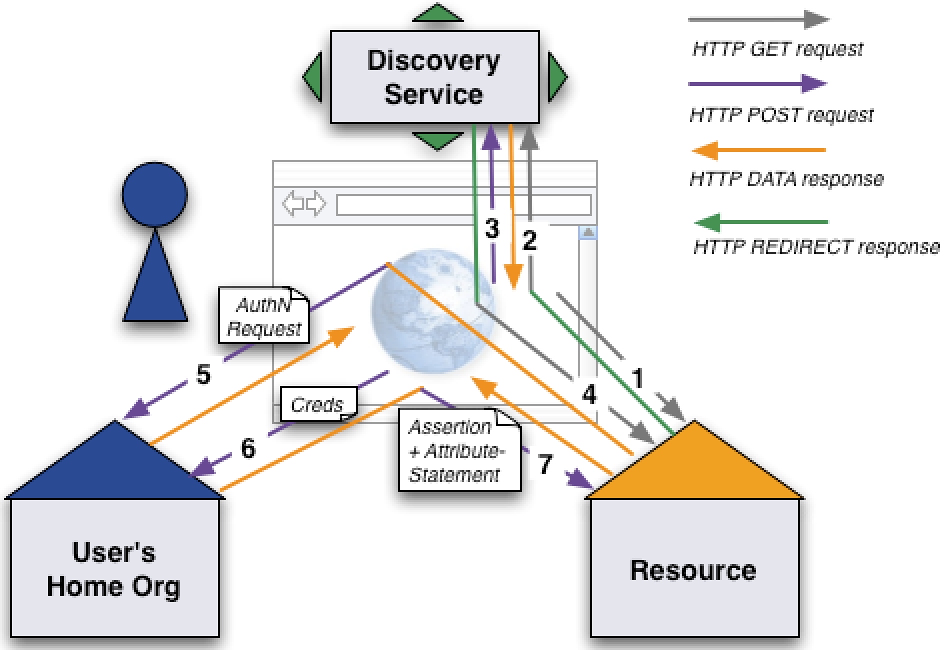


[Original approach] McLaren, P. J. et al. Polymorphisms of Large Effect Explain the Majority of the Host Genetic Contribution to Variation of HIV-1 Virus Load. Proc. Natl. Acad. Sci. 112, 14658–14663 (2015).

[FAHME] Froelicher et al. Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption.

Federated 2-factor authentication based on Switch-AAI or Switch edu-ID

- Federated authentication and authorization mechanism compatible with Switch-AAI (Shibboleth login procedure)
- Already used by Swiss hospitals and universities



https://www.switch.ch/aai/demo/medium/

Positioning of MedCo with respect to similar distributed platforms

	Criteria \ Platform	CLINERION Real World Data Solutions	SHRINE Shared Health Research Information Network	DataSHIELD Secure Bioscience Collaboration	VANTAGE	HP	MedCo
Functionality	Federated cohort exploration	Yes	Yes	No	No	Yes	Yes
	Federated analytics	No	No	Yes	Yes	Yes	Yes
Privacy/Security	Protection of intermediate results and distributed computation	No	No	No	No	No	Yes
	Protection of end result from inference attacks	No	Yes	No	No	No	Yes
	Fine-grained role management	No	No	No	No	No	Yes
Usability	Graphical user interface	Yes	Yes	No	No	Yes	Yes
	Public API	No	Yes	Yes	Yes	Yes	Yes
	Service & support	Yes	No	No	No	No	Yes
	Extensibility	No	Yes	Yes	Yes	Yes	Yes

Clinerion

+ Already deployed in all Swiss university hospitals
+ Proven track record

- Mainly designed for “patient recruitment” and pharmas’ needs
- Rigid data model enabling queries across only 5 variables (diagnosis, procedures, labs, treatments, demographics)
- Customized ETL (hard to maintain without Clinerion support)
- Needs central trusted third-party
- Closed API (vendor lock-in) => expensive customizations, no extension or integration with other software components possible
- Limited to cohort exploration based on patient counts => no distributed analyitcs
- Lacks data protection guarantees for aggregated data leaving hospitals’ IT infrastructure

MedCo

+ Designed for “data recruitment” based on SPHN standards and hospitals needs
+ Fully integrated with i2b2, thus enabling distributed and privacy-preserving fine-grained queries for rich cohort exploration based on SPHN ontology and semantic interoperability standards
+ Open API and free license for academic and non-commercial purposes => extensible with new modules and improvements
+ Possibility to run distributed analytics (beyond patient counts) for hypotheses generation without compromising patients’ privacy
+ Hospital IT security and legal compliance
+ State-of-the-art data protection technologies
+ Support by driver projects
+ Deployed and tested in 3 out of 5 Swiss university hospitals

- Not used in operational environments yet

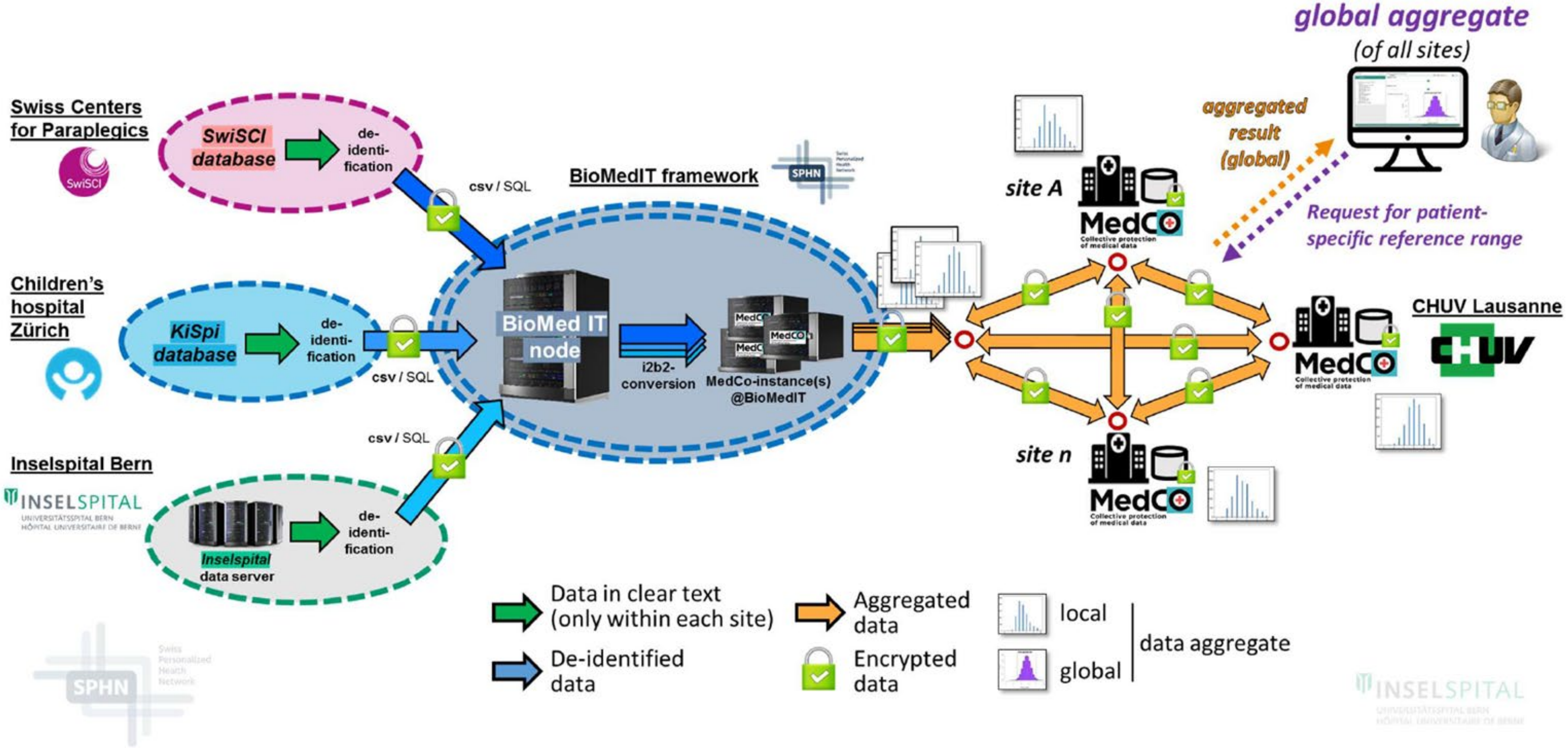
What we have accomplished

- Shown that MedCo works, on data sets 1000 times larger than the ones currently used with Clinerion
- Addressed legal/ethical issues; produced the Data Privacy Impact Analysis (as requested by GDPR)
- Live tests by the NAB and HIT-STAG of MedCo for 3 weeks in May 2021; no feedback received so far
- Extensively demoed to the Swiss oncology community
- Thorough comparison with alternative solutions, including Clinerion
- Found the “Holy Grail” of secure federated analytics
- Traction from outside SPHN
    - Ophthalmology
    - UT Health (Houston), IKNL (Cancer Center, NL), Fondazione Maugeri (Pavia, I)
    - Cybersecurity, insurance → launch of start-up Tune Insight

About medical data

- The current situation of medical data is appalling and is a worldwide embarrassment (lack of standards, poor quality, etc.)
- Switzerland is no exception, unfortunately
- **It is not DPPH/MedCo mission to clean the mess**
- But once MedCo is deployed, it will be an **incentive** for hospitals and clinics to work together → this should bootstrap a virtuous circle
- MedCo uses i2b2 to facilitate adoption

BioMed IT-MedCo-Hybrid for Swiss BioRef



Credit: Alexander Leichtle and Harald Witte, Swiss BioRef, Insel

MedCo: Secure, Privacy-Conscious Federated Analytics Infrastructure for Precision Medicine

Was presented and demoed at multiple meetings of the Swiss Personalized Oncology (SPO) and Swiss Molecular Pathology Platform (SOCIBP) Projects

Attendees included:

Olivier Michielin (CHUV), Mohamed Ben Tires-Alj (USB), Marc Rubin (Insel), Christian Britschgi (USZ), Simon Haefliger (Insel), Sacha Rothschild (USZ), Pedros Tsantoulis (HUG), Andreas Wicki (USZ), Sylvain Pradervand (CHUV)

What MedCo will bring to SPHN

FEDERATED FEASIBILITY ANALYSIS

ETHICS + LEGAL APPROVALS

PROJECT DATA ANALYSIS

Current SPHN model

Federated query (powered by Clinerion)
- Patient count per hospital are sent in the cleartext (no protection for low numbers or consecutive queries)
- Need of a central trusted party for counts aggregation
- Only few variables can be queried

Raw data access and analysis
- Any analysis
- Data in clear in the BioMedIT secure zone
- Needs DTUA and ethics approval

BioMedIT

SPHN model enhanced by MedCo

Federated query (powered by MedCo)
- compatible with RDF and SPHN semantic dataset => any variable can be queried
- Patient counts per hospital encrypted => no need of trusted central party
- Utmost security and privacy protection

MedCo Collective protection of medical data

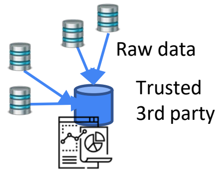
Federated data analysis (powered by MedCo)
- Possibility of generating scientific hypothesis without transferring data for a set of predefined analyses
- Data confidentiality and privacy preservation
- No DTUA
- Streamlined ethics approval

BioMedIT

Raw data access and analysis
- Any analysis
- Patient-level at the BioMedIT secure zone
- Needs DTUA and ethics approval

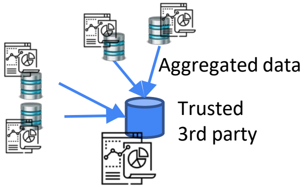
Federated Learning - Current Approaches

(a) Fully centralized



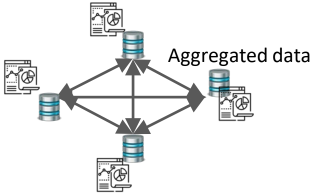
All of Us
EGA
Genomics England

(b) Meta-analysis



https://covidclinical.net/

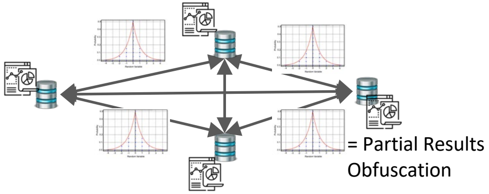
(c) Decentralized



http://www.datashield.ac.uk

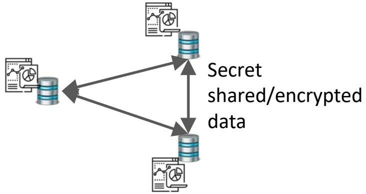
Personalized Health Train (PHT)

(d) Differential Privacy Decentralized



- M. Kim et al. "Secure and Differentially Private Logistic Regression for Horizontally Distributed Data," TIFS 2019
- M. Abadi et al. Deep learning with differential privacy. In ACM CCS, 2016.
- Chaudhuri and C. Monteleoni. Privacy-preserving logistic regression. In NIPS, 2009.

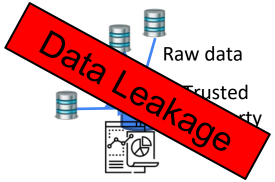
(d) Cryptographic (SMC, HE) Decentralized



- A. Gascón et al.. Privacy-preserving distributed linear regression on high-dimensional data. PETS, 2017.
P. Mohassel and Y. Zhang. SecureML: A system for scalable privacy-preserving machine learning. In IEEE S&P, 2017.

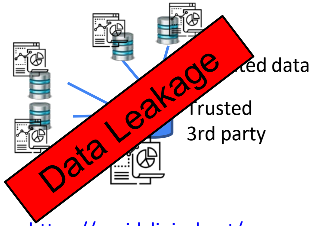
Federated Learning - Current Approaches

(a) Fully centralized



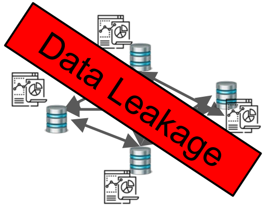
All of Us
EGA
Genomics England

(b) Meta-analysis



https://covidclinical.net/

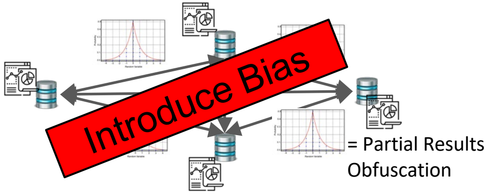
(c) Decentralized



http://www.datashield.ac.uk

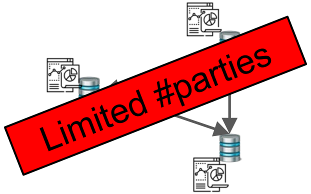
Personalized Health Train (PHT)

(d) Differential Privacy Decentralized



- M. Kim et al. "Secure and Differentially Private Logistic Regression for Horizontally Distributed Data," TIFS 2019
- M. Abadi et al. Deep learning with differential privacy. In ACM CCS, 2016.
- Chaudhuri and C. Monteleoni. Privacy-preserving logistic regression. In NIPS, 2009.

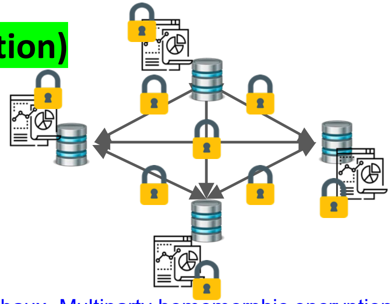
(d) Cryptographic (SMC, HE) Decentralized



- A. Gascón et al.. Privacy-preserving distributed linear regression on high-dimensional data. PETS, 2017.
P. Mohassel and Y. Zhang. SecureML: A system for scalable privacy-preserving machine learning. In IEEE S&P, 2017.

Our solution (Secure Multiparty Computation + Homomorphic Encryption)

- Data Confidentiality
- Not data outsourcing
- Scale with #parties
- Exact results

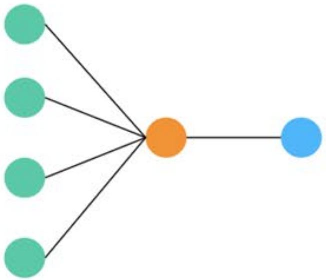


C. Mouchet, J. R. Troncoso-pastoriza, J.-P. Bossuat, and J. P. Hubaux. Multiparty homomorphic encryption: From theory to practice. PETS'21. https://eprint.iacr.org/2020/304, 2020.

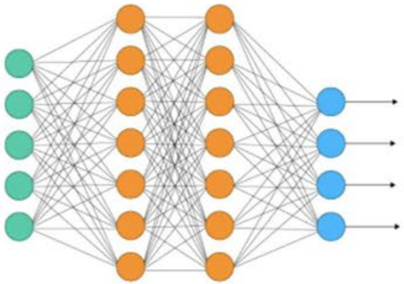
Multi-Party Homomorphic Encryption (MHE) Efficient Functionalities

- Statistical computations (aggregations, histograms, moments,...)
- Machine learning/AI: Increasing set of models can be efficiently trained and evaluated with MHE

Generalized Linear Models
(Represented as a simple neural network)



Deeper Neural Network



Input Layer Hidden Layer Output Layer

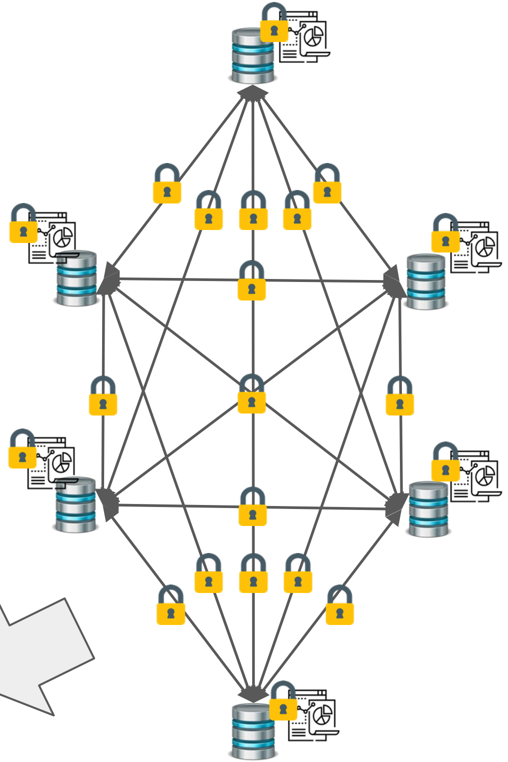
Froelicher, J.R. Troncoso-Pastoriza, A. Pyrgelis, S. Sav, J.S. Sousa, J.-P. Bossuat, J.-P. Hubaux, "Scalable Privacy-Preserving Distributed Learning". PETS'21 https://arxiv.org/abs/2005.09532

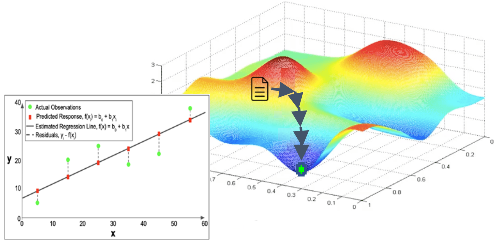
S. Sav, A. Pyrgelis, J.R. Troncoso-Pastoriza, J.-P. Bossuat, J.S. Sousa, J.-P. Hubaux, "POSEIDON: Privacy-Preserving Federated Neural Network Learning". NDSS'21 https://arxiv.org/abs/2009.00349

SPINDLE: Scalable Privacy-preservINg Distributed LEarning

Generic Secure Federated Learning that ensures **Data Confidentiality + Model Confidentiality by building on:**

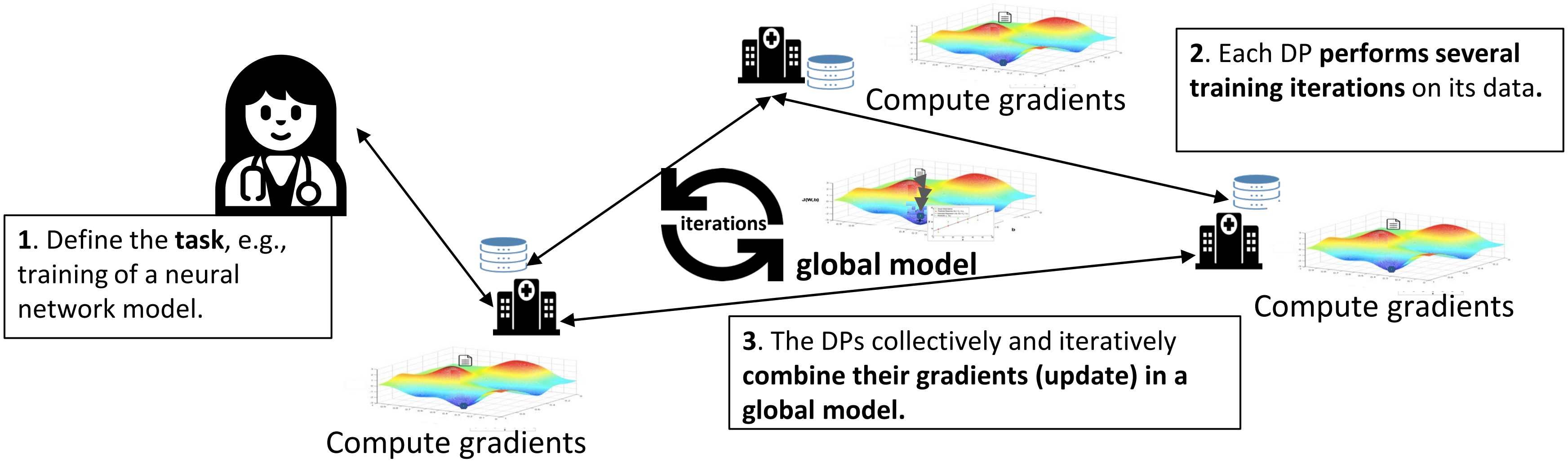
Extended MapReduce Abstraction	
Cooperative Gradient Descent	Multiparty Homomorphic Encryption (MHE)
→ instantiated for Generalized Linear Models	





POSEIDON: Privacy-Preserving Federated Neural Network Learning

Solution: The data providers (DPs) collaborate to enable a joint gradient descent while protecting their privacy and **obtain a global and accurate model**



4. After the training, **the model** is **kept secret for oblivious predictions**

Parameterization:
Strong interdependencies between learning parameters and cryptographic parameters

International collaborations

- GA4GH Data Security Work Stream
- MedCo now part of the i2b2 official community projects
- Prof. Shawn Murphy, HMS, and the ACT Network
- Broad Inst. + MIT
- Hôpital ophtalmique de Lausanne + Hôpital Rothschild (P