

SHARE WITHOUT SHARING  
WORKSHOP



# Share without Sharing: How We Solved It

Prof. Jean-Pierre Hubaux, EPFL  
4 November 2021

With gratitude to my co-workers

3' panel talk

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

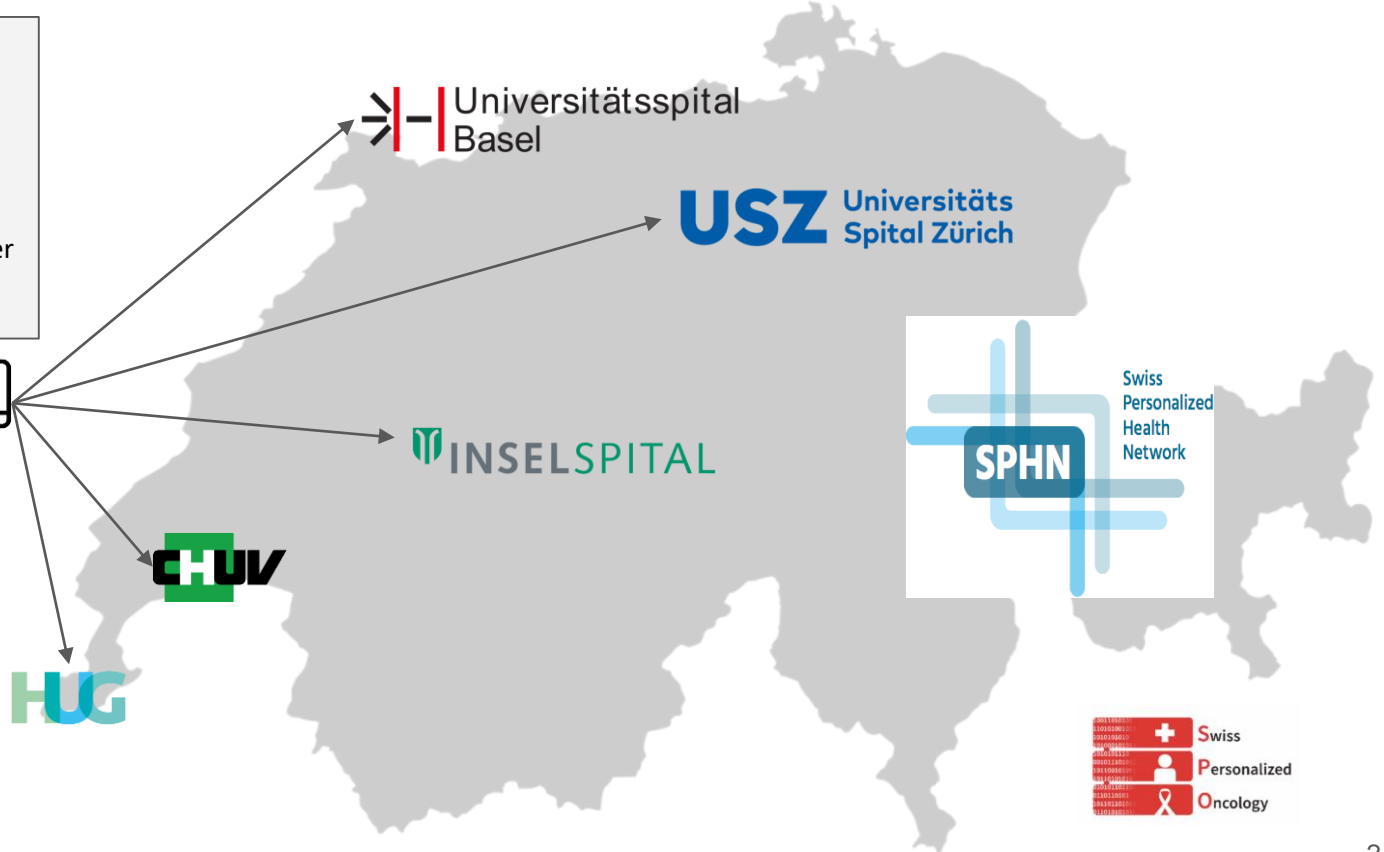
**Q1:** How many adult cancer patients consenting on reuse of routine data for research with diagnosis of a malignancy on or after 1st January 2015, mutations in BRAF gene and under anti-PD-1 are there?

**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**



# MedCO

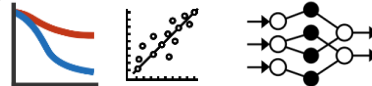
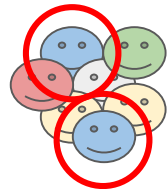
- **Distributed software platform** for federated cohort exploration and analytics of clinical and genomic data
- Co-developed by EPFL and CHUV
- Built on top of the i2b2 cohort explorer (i2b2 is used by 250+ hospitals worldwide)
- Relies on **advanced cryptographic techniques**  
→ Multi-party homomorphic 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



# April 2020: MedCo deployed at 3 hospitals



## EPFL software to enable secure data-sharing for hospitals



The MedCo system aims to facilitate medical research on pathologies – such as cancer and infectious diseases – by enabling secure computations on decentralized data. The unique software has recently been deployed at three Swiss hospitals.

02.04.20

### LINKS

- [MedCo](#)
- [LDS](#)
- [Video](#)



- First application:  
Swiss Personalized Oncology project:  
→ melanoma data and beyond
- Planned deployment at Zurich University Hospital
- Ongoing international deployments: USA, NL, Italy, France

# 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 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

# Feedback from Swiss authorities 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 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

# Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption

[David Froelicher](#), [Juan R. Troncoso-Pastoriza](#), [Jean Louis Raisaro](#), [Michel A. Cuendet](#), [Joao Sa Sousa](#), [Hyunghoon Cho](#), [Bonnie Berger](#), [Jacques Fellay](#) & [Jean-Pierre Hubaux](#) 

[Nature Communications](#) **12**, Article number: 5910 (2021) | [Cite this article](#)

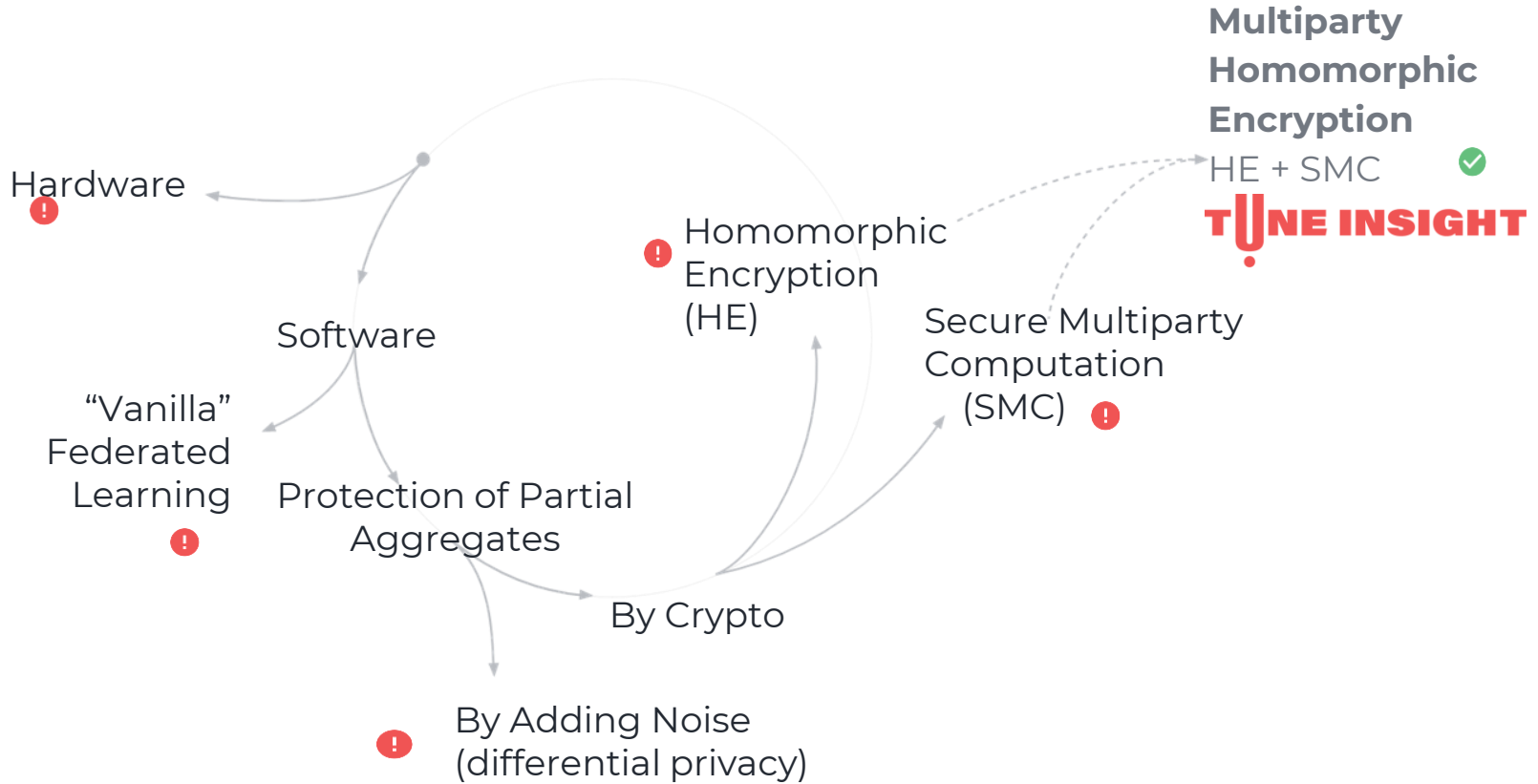
[Metrics](#)

## Abstract

---

Using real-world evidence in biomedical research, an indispensable complement to clinical trials, requires access to large quantities of patient data that are typically held separately by multiple healthcare institutions. We propose FAMHE, a novel federated analytics system that, based on multiparty homomorphic encryption (MHE), enables privacy-preserving analyses of distributed datasets by yielding highly accurate results without revealing any intermediate data. We demonstrate the applicability of FAMHE to essential biomedical analysis tasks, including Kaplan-Meier survival analysis in oncology and genome-wide

# Share without Sharing: Available Options





## Enterprise Data & Analytics



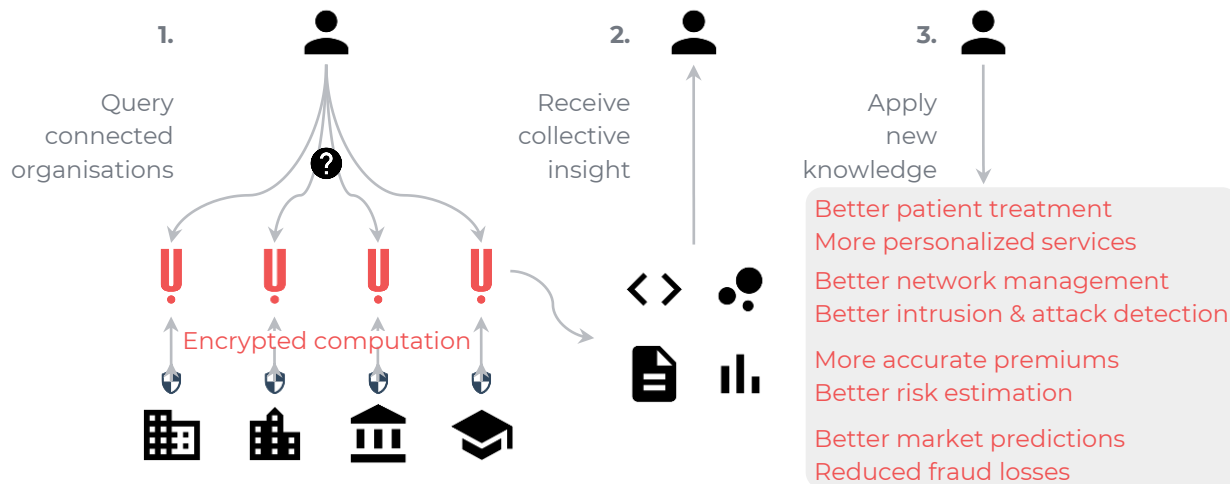
However, organizations are **prevented** to enter valuable data collaborations due to fear of **data leaks** and **data protection regulations**

# TUNE INSIGHT

juan@tuneinsight.com

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

- 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
- Raised pre-seed with Wingman Ventures



**Access to insights**  
**Personalization**




**Immediacy**  
**Scalability**



**Compliance**  
**Control**

# MHE: mathematical proofs instead of vendor lock-in and side-channel attacks

	Software-based solutions (MHE) 	Hardware-based solutions (e.g., Intel SGX)
<b>System and trust model</b>	<b>Decentralized</b> (federated computing, edge computing) or <b>centralized</b> (outsourced) systems	<b>Only centralized</b> systems (data has to be transferred to the TEE)
<b>Assumptions</b>	Protection against passive adversaries with quantum computing power: <b>processing infrastructure (including side-channels) and other data providers</b>	Protection against passive adversaries (other tenants); <b>limited protection against the processing infrastructure</b> ; protection against side-channels is implementation-dependent
<b>Implementation cost</b>	<b>Tailored solution</b> ; application-specific design; composition of cryptographic building blocks; limited range of efficient functionalities	<b>Available SDKs</b> ; relatively easy conversion to secure enclave; general-purpose solutions; limited libraries and memory inside the enclave
<b>Performance and overhead</b>	<b>Less than 10x</b> overhead when full packing capacity is utilized (federated training of GLMs and NNs). Up to 4-5 orders of magnitude overhead for non-optimized or non-packed solutions	<b>Negligible</b> overhead for <b>regular instructions</b> ; <b>4x overhead for memory</b> copy operations; <b>35x overhead for syscalls</b> to/from enclave
<b>Response to newly discovered vulnerabilities</b>	<b>Software patch</b> with protocol update; usually, no re-encryption of the data is needed	<b>Firmware</b> patch with variable <b>performance impact</b> (1x to 20x slow-down); <b>architecture change and hardware replacement</b> ; <b>enclave code update</b> (update signatures, keys, and require new attestation)

GLM : Generalized Linear Model  
 MHE : Multi-party homomorphic encryption  
 NN : Neural Network

SDK : Software Development Kit  
 SGX : Software Guard eXtensions  
 TEE : Trusted Execution Environment

- We have solved the problem of GDPR-compliant federated learning
- Solution: Multi-party homomorphic encryption (MHE)
  - Perform computations without “seeing” the data
  - Rely on decentralized trust and mathematical proofs
  - No need to transfer the data
- Scalability with the number of data providers and the size of the datasets
- Green light from the federal data protection authority
- With Tune Insight:
  - We take these research breakthroughs to market with great initial traction
  - We have raised the money → 8 highly qualified specialists already hired
  - Contact us/me at [contact@tuneinsight.com](mailto:contact@tuneinsight.com)

D. Froelicher, J.R. Troncoso-Pastoriza, A. Pyrgelis, S.Sav, J.S. Sousa, J.P. Bossuat and J.P. Hubaux, **SPINDLE: Scalable Privacy-Preserving Distributed Learning**. PoPETS, 2021

S.Sav, A. Pyrgelis, J.R. Troncoso-Pastoriza, D. Froelicher, J.P. Bossuat, J.S. Sousa and J.P. Hubaux, **POSEIDON: Privacy-Preserving Federated Neural Network Learning**. NDSS,

2021