ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Responding to the Security and Privacy Challenges Raised by Personalized Medicine

Challenges in Data-Driven Genomic Computing

Como, Italy, March 8th 2019

Juan Ramón Troncoso-Pastoriza

juan.troncoso-pastoriza@epfl.ch

DPPH
Data Protection
in Personalized Health

Strategic Focus Area
Personalized Health
and Related Technologies

Outline

- Privacy and Security Threats to Personalized Medicine
- Security and Privacy Notions and Technologies in Personalized Medicine
- Data Protection in Personalized Health
- MedCo: Privacy-Conscious Exploration of Distributed Clinical and Genomic Data
- Related Events and Publications
- Conclusions and Open Questions

Growing Concern: Medical Data Breaches

Around 5 declared breaches per week, each affecting 500+ people

https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

Under Investigation | Archive | Help for Consumers

As required by section 13402(e)(4) of the HITECH Act, the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals. The following breaches have been reported to the Secretary:

Cases Currently Under Investigation

This page lists all breaches reported within the last 24 months that are currently under investigation by the Office for Civil Rights.

Show Advanced Options

Breach Report Results

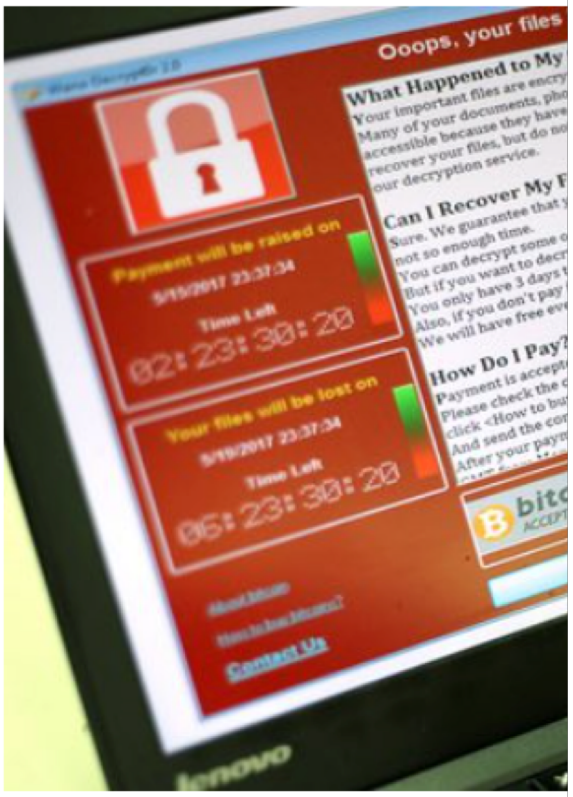
Expand All	Name of Covered Entity	State	Covered Entity Type	Individuals Affected	Breach Submission Date	Type of Breach	Location of Breached Information
	Humana Inc	KY	Health Plan	1447	02/27/2019	Hacking/IT Incident	Network Server
	West Virginia Public Employees Insurance Agency	WV	Health Plan	1400	02/27/2019	Hacking/IT Incident	Network Server
	Molina Healthcare	CA	Health Plan	895	02/22/2019	Hacking/IT Incident	Network Server

Recent targeted Attacks (2016-2018)

Do state institutions have the resources to fight hackers?

The Guardian,
14 May 2017

Public sector has lessons to learn as
from ransomware attack



A ransomware attack bought computers to a s
Tongo/EPA

FÉDÉRALE DE LAUSANNE

20. März 2016, 10:05 Uhr Klinikum Neuss

Wenn Cyberkriminelle Krankenhaus lahmleg

The Telegraph 5 Dec 2018

HOME | NEWS | SPO

News

UK | World | Politics | Science | Education | Health | Brexit | Royals | Investigat

News

NHS patients' genetic data targeted as foreign hackers attack high security MoD unit

share

Save 33

Another Major Concern: Re-identification Attacks against Genomic Databases

New Scientist

HOME NEWS TECHNOLOGY SPACE PHYSICS HEALTH EARTH HUMANS LIFE TOPICS EVENTS JOBS

Home | News | Health | Technology

DAILY NEWS 1 September 2008

Genetic data withdrawn amid privacy concerns

By Peter Aldhous

nature International weekly journal of science

Subscribe Login

nature news home | news archive | specials | opinion | features | news blog | nature journal

comments on this story

Stories by subject
- Genetics
- Health and medicine
- Policy

Stories by keywords
- NIH
- Wellcome Trust
- Broad Institute
- DNA databanks
- patient privacy
- genetics

This article elsewhere
- Blogs linking to this article
- Add to Digg
- Add to Facebook
- Add to Newsvine
- Add to Del.icio.us
- Add to Twitter

Published online 4 September 2008 | Nature | doi:10.1038/news.2008.1083

News

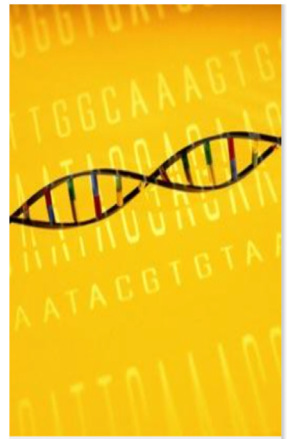
Researchers criticize genetic data restrictions

Fears over privacy breaches are premature and will impede research, experts say.

Natasha Gilbert

As fears over privacy prompt genetic databases in the United States and Britain to close public access to some of their data, scientists working in the field are complaining that the moves are premature and will impede research.

The US National Institutes of Health (NIH), the Broad Institute in Cambridge, Massachusetts, and the Wellcome Trust in London all decided to restrict access to data from genome-wide association (GWA) studies — which contain collections of thousands of people's DNA — after research



Could an individual's health details be extracted from pooled genetic data?

Getty

Related stories
- Biomedical science: Betting the bank 23 April 2008
- Genome studies: Genetics by numbers 30 January 2008

Naturejobs

Gastroenterologist
Loyola University Chicago

Research Engineer / Research Scientist in Renewable Energy
King Fahd University of Petroleum & Minerals

More science jobs

Post a job

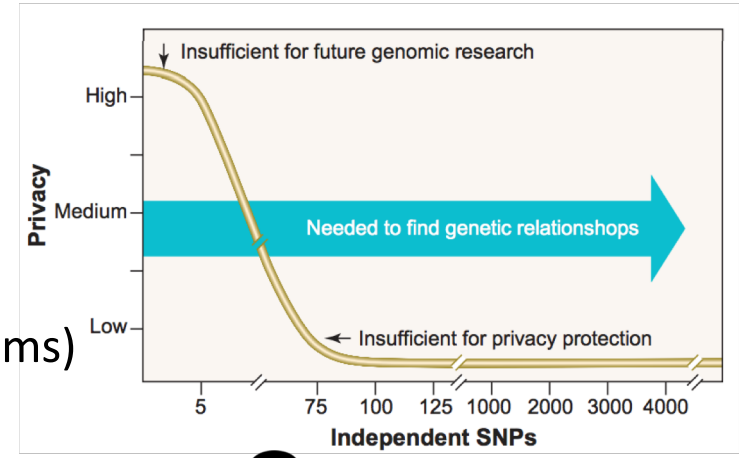
Resources
- Send to a Friend
- Reprints & Permissions
- RSS Feeds

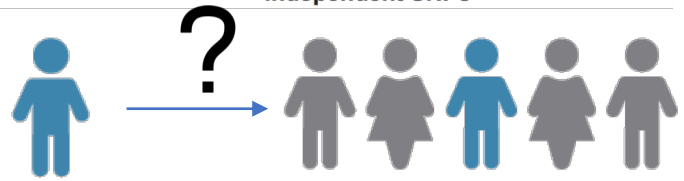
elsewhere on nature.com
- Genetics@nature.com

A Plethora of Attacks Against Genomic Privacy

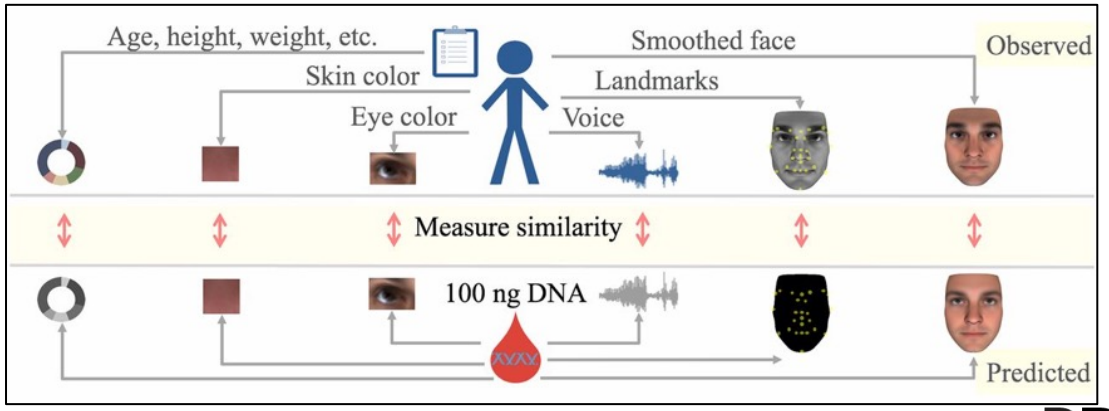
- ***Lin et al. 2004 Science***: 75 or more SNPs (Single Nucleotide Polymorphisms) are sufficient to identify a single person
- ***Homer et al. 2008 PLOS Genetics:*** aggregated genomic data (i.e., allele frequencies) can be used for re-identifying an individual in a case group with a certain disease
- ***Gymrek et al. 2013 Science:*** surnames can be recovered from personal genomes, linking "anonymous" genomes and public genetic genealogy databases
- ***Lipper et al. 2017 PNAS:*** Anonymous genomes can also be identified by inferring physical traits and demographic information

Standard de-identification and anonymization techniques are ineffective with genomic data





ysearch



Security / Privacy Requirements for Personalized Health

- Pragmatic approach, **gradual** introduction of new protection tools
- Different **sensitivity levels** of the data
- Different **access rights**
- Exploit **existing** data (electronic health records) and tools
- Be **future-proof** (and implement long-term protection)
- Awareness and enforcement of **patient consent**
- Secure also the **collection** of health data (via smartphones, wearable sensors,…)

Security and Privacy Notions

Technologies for Privacy Protection

Security and privacy notions

Security: Rightful access to data

- Access control (authentication and authorization)
- Availability (protection against Denial Of Service – DoS attacks)
- Auditability (logging)
- Accountability (verifiability)

Privacy: Rightful use of data

- Legal imperatives and expressed wishes of the data owner
- Consent and terms of use
- Difficult to tackle and evaluate in distributed data sharing scenarios
- Technical protection measures to limit risk (legal compliance)

Technologies for privacy and security protection

Traditional Encryption

- Protects data at rest and in transit
- Cannot protect computation

Homomorphic Encryption

- Protects computation in untrusted environments
- Limited versatility vs efficiency

Secure Multiparty Computation

- Protects computation in distributed environments
- High communication overhead

Trusted Execution Environments

- Protects computation with Hardware Trusted Element
- Requires trust in the manufacturer, vulnerable to side-channels

Differential Privacy

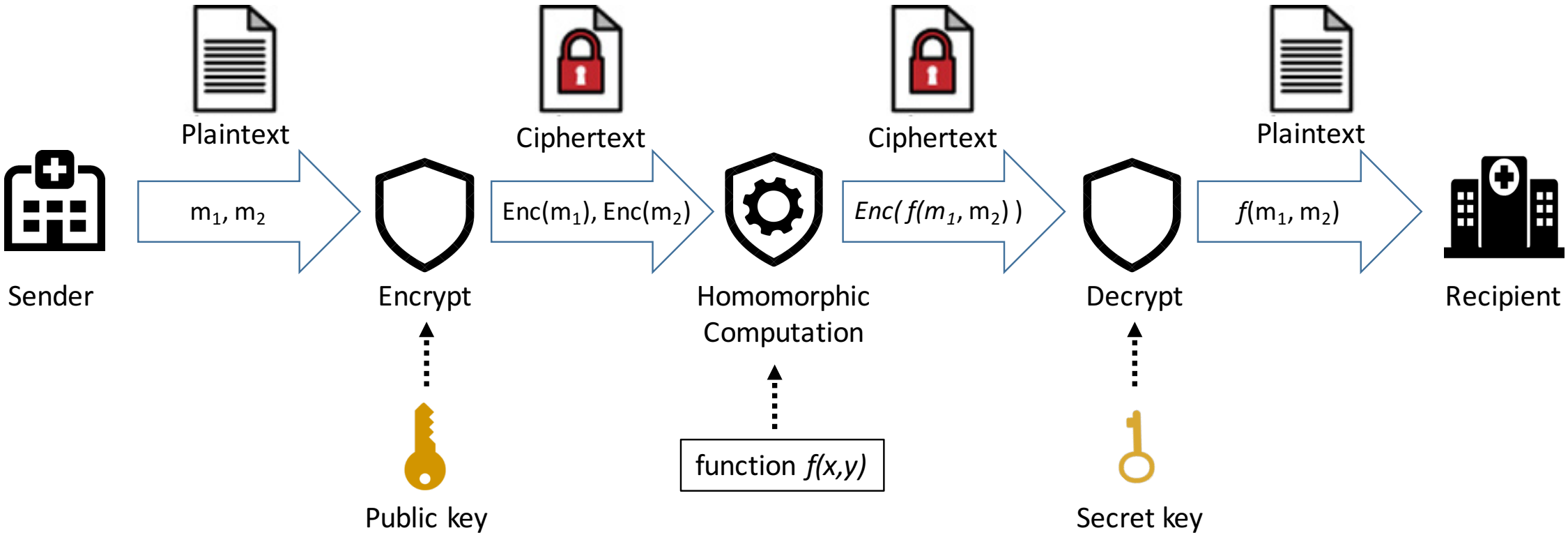
- Protects released data from inferences
- Degrades data utility (privacy-utility tradeoff)

Distributed Ledger Technologies (Blockchains)

- Strong accountability and traceability in distributed environments
- No privacy by default

Homomorphic Encryption

- Homomorphic encryption
  - It enables computations directly on encrypted data
  - Ii is probabilistic (semantic security is a requirement for enabling homomorphic computation)
  - **problem**: it can only efficiently enable polynomial operations



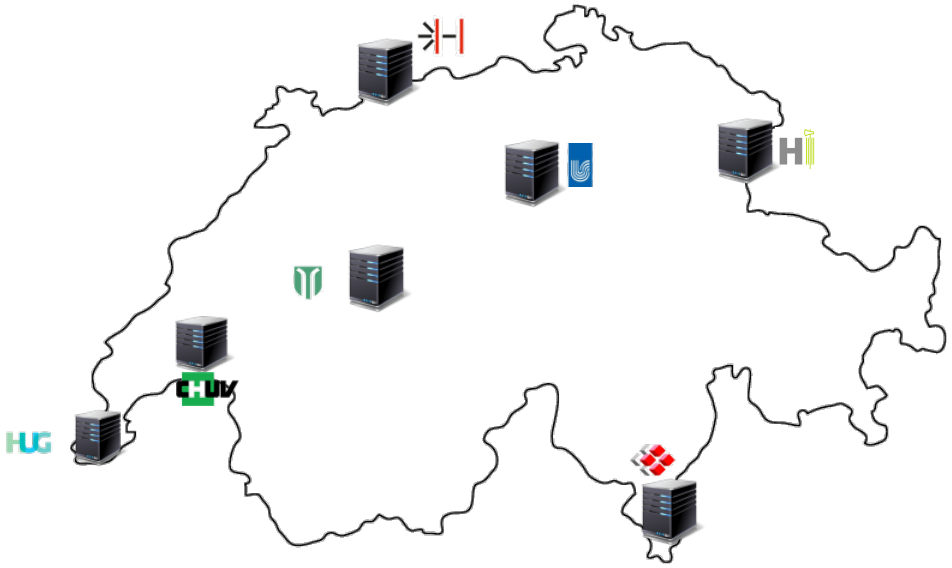
Homomorphic Encryption

- Homomorphic Encryption and Fully Homomorphic Encryption
  - ➢ (Practical/Somewhat) Homomorphic Encryption ≠ Fully Homomorphic Encryption

Somewhat Homomorphic Encryption	Fully Homomorphic Encryption (FHE)
Can perform polynomial operations of limited degree	Can perform any operation represented as a binary circuit
Need to set-up security parameters tailored for the specific computation	Generic security parameters
Arithmetic circuits and integer inputs	Logical circuits and binary inputs
Constrained computational overhead (several orders of magnitude speed-up w.r.t. FHE)	Very high computational overhead
Expansion controlled by the ciphertext/plaintext cardinality ratio	Very high expansion
Based on lattice cryptography	
Promise of post-quantum resilience	

DPPH

Data Protection
in Personalized Health



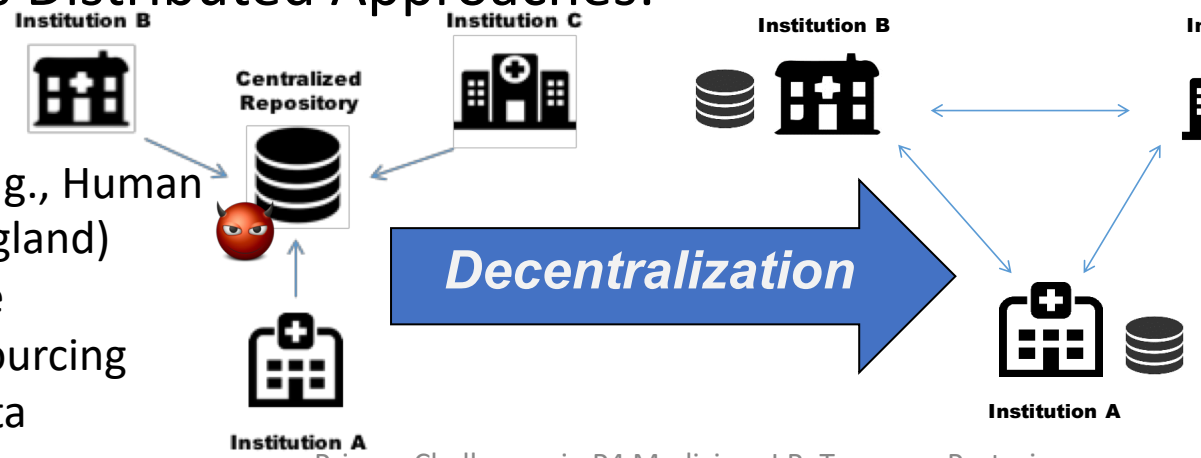
Data Protection in Personalized Health at the Swiss nation-wide scale

DPPH: Data Protection in Personalized Health

- P4 (Predictive, Preventive, Personalized and Participatory) medicine
  - Revolutionize healthcare by providing better diagnoses and targeted preventive and therapeutic measures
- Challenges:
  - Interoperability
  - Efficiency and usability in data sharing
  - Scalability/Big Data
  - Mitigation of privacy risks and compliance with data protection
- Centralized vs Distributed Approaches:

Swiss Personalized Health Network

Strategic Focus Area
Personalized Health and Related Technologies



Centralized approach (e.g., Human Longevity, Genomics England)

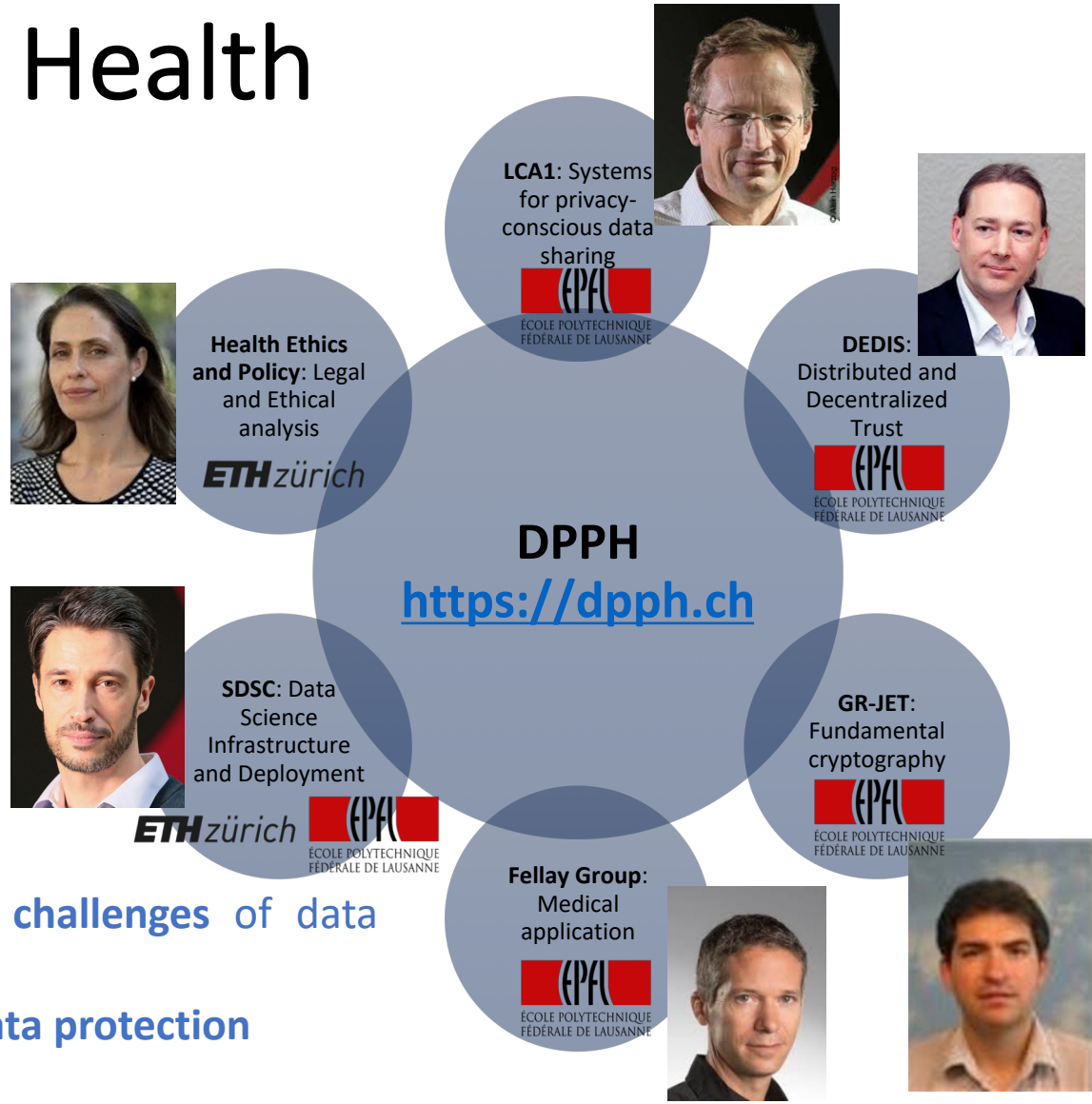
- Single point of failure
- Impossibility of outsourcing individual patient data

Distributed approach (e.g., Beacon GA4GH, Swiss Personalized Health Network - SPHN)

- Distributed trust
- Take computation to the data

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 - 3/2021)
- Funding Program: ETH PHRT (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**

DDPH: Key Enablers

The integration of the shown tools and frameworks provides **crucial benefits for medical research**

Renku

Scalable distributed scientific computing infrastructure (SDSC)

Scalability and Reproducibility

tranSMART FOUNDATION

i2b2 Informatics for Integrating Biology and the Bedside

Ease querying and aggregating distributed medical data

Accessibility and Usability

UnLynx

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Secure and privacy-conscious data sharing and processing for medical data (LCA1)

Privacy and Verifiability

OmniLedger

Robust support for atomic immutable distributed transactions (DEDIS)

Auditability and Access Control

COMPLIANCE

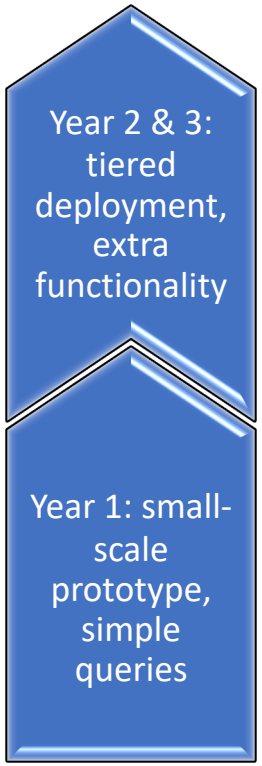
Framework for inference risks and countermeasures, and ethical analysis of distributed platforms for medical data sharing

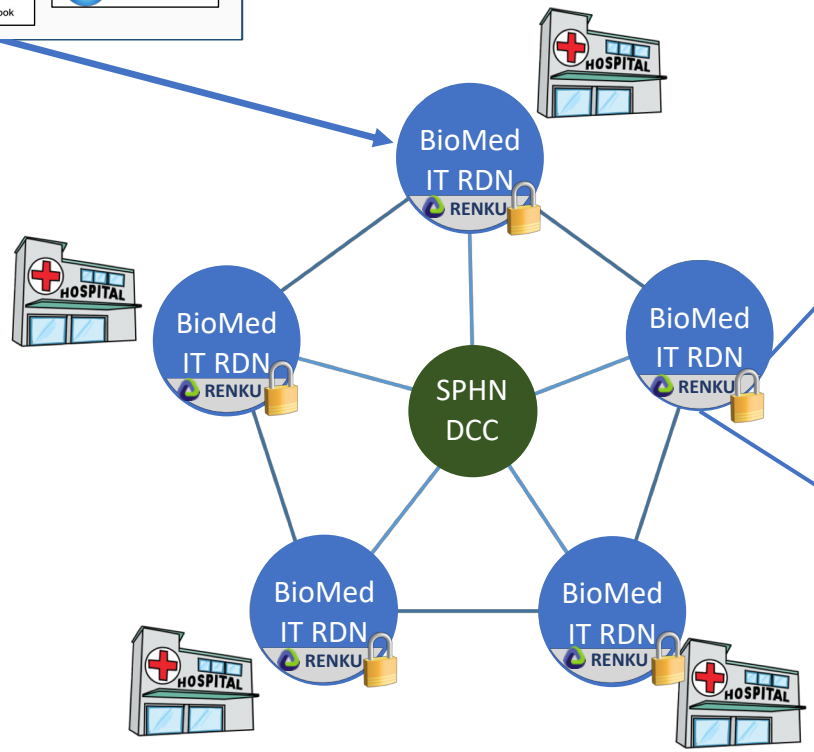
Legal and Ethical compliance

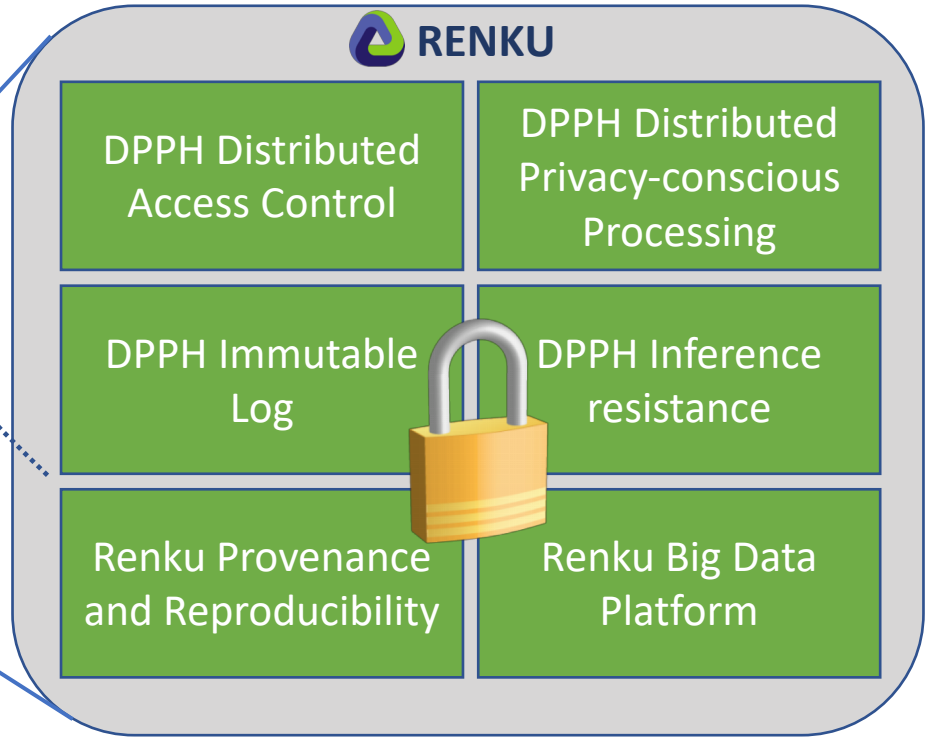
i2b2 and TranSMART are just examples of tools. The DPPH project is *tool-agnostic

Envisioned secure infrastructure for privacy-conscious medical research in Switzerland

DPPH/Renku Blockchain







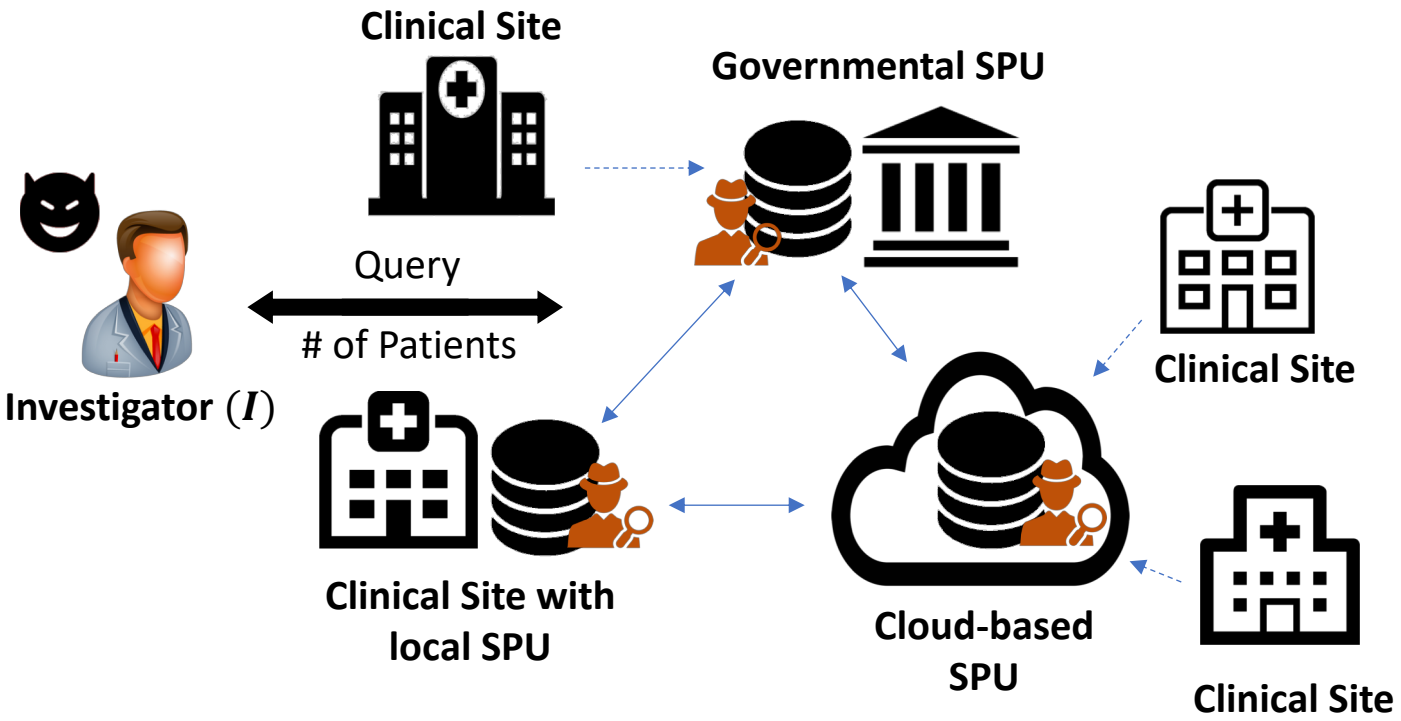
RDN: Regional Data Node
DCC: Data Coordination Center

MedCo

Enabling Privacy-Conscious Exploration of Distributed Clinical and Genomic Data

Raisaro et al., IEEE TCBB 2018

System and threat models



SPU: Storage and Processing Unit

Honest-but-curious adversary:

- honestly follows the protocol
- tries to infer sensitive data from the different steps of the protocol

Malicious-but-covert adversary:

- can tamper with the protocol
- tries to infer sensitive data from the query end-result

Main Privacy and Security Concerns

- **Loss of data confidentiality** due to illegitimate access to the data
  - External (hacker) or internal (insider) attacker stealing the data

➔ Standard encryption can protect data ONLY at rest or in transit BUT NOT during processing (e.g., in the memory)

- **Patient re-identification** due to legitimate access to the data
  - Malicious users performing “smart” data requests in order to re-identify patients in a specific dataset (e.g., patients with HIV)

➔ De-identification or anonymization is ineffective with genomic data

MedCo: Combining the best of Information Security and Medical Informatics

Data model

Interoperability layer
Meta API

Privacy-preserving computing framework

Modern GUI

MedCo

Collective protection of medical data

DISCLAIMER

MedCo is a generic concept and it is not fundamentally tied to these technologies, but can be adapted and integrated to other ones

Main features of MedCo

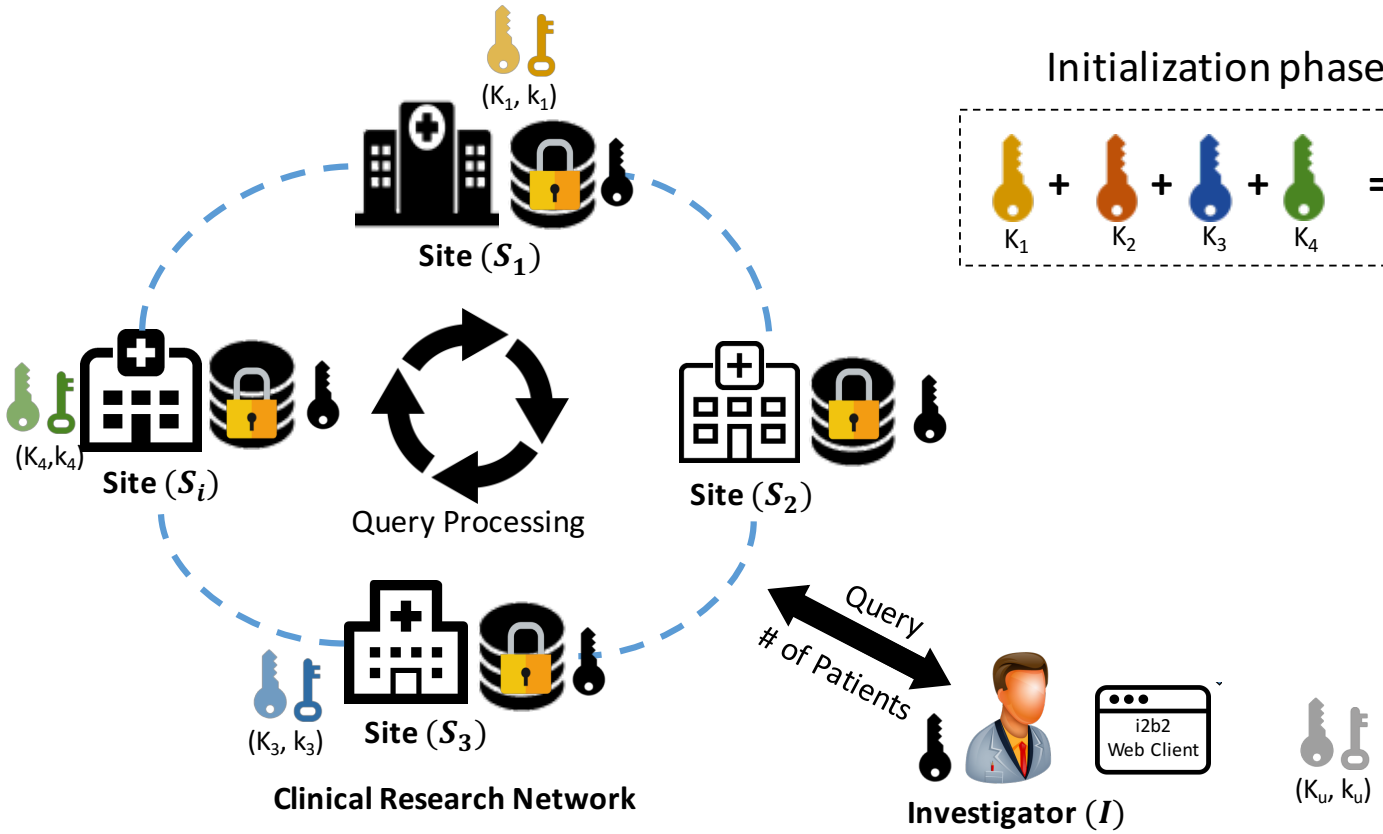
Functionality:

- Cohort exploration functionalities (same as a federated i2b2):
  - **Count the number of research subjects** based on a set of inclusion and exclusion clinical and genetic criteria
  - **Retrieve the identifiers** of such subjects

Security/Privacy:

- Protection of data confidentiality at rest, in transit and **during computation**
- no single point of failure
- only the investigator can obtain the query end-result
- (optional) unlinkability
- (optional) differential privacy

System and query protocol overview



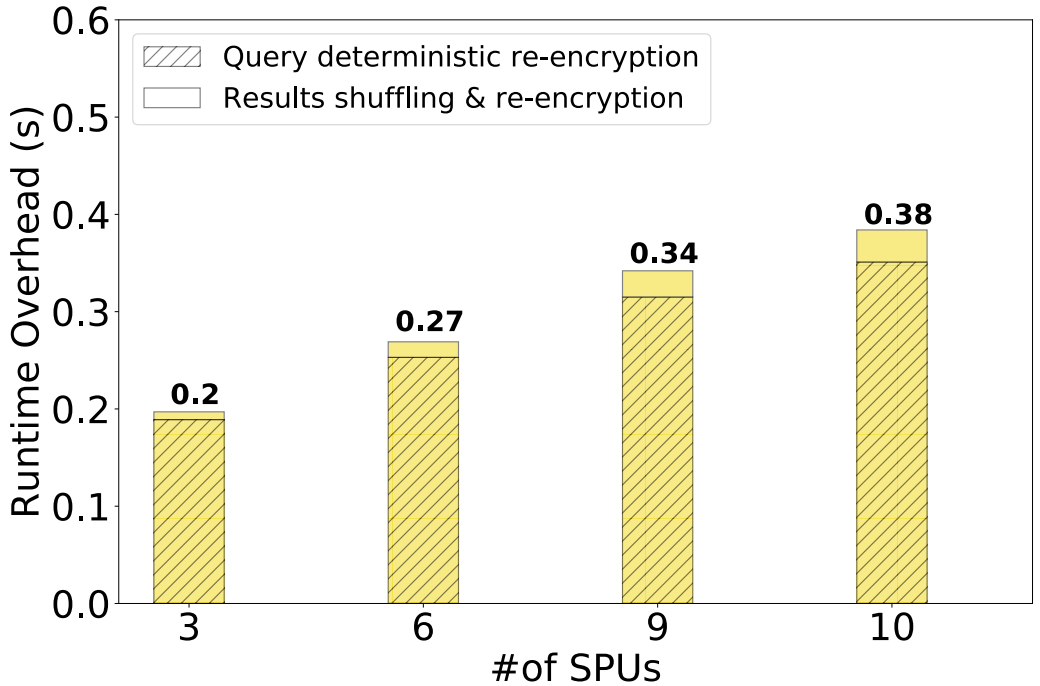
A, B) ETL & Encryption Phase

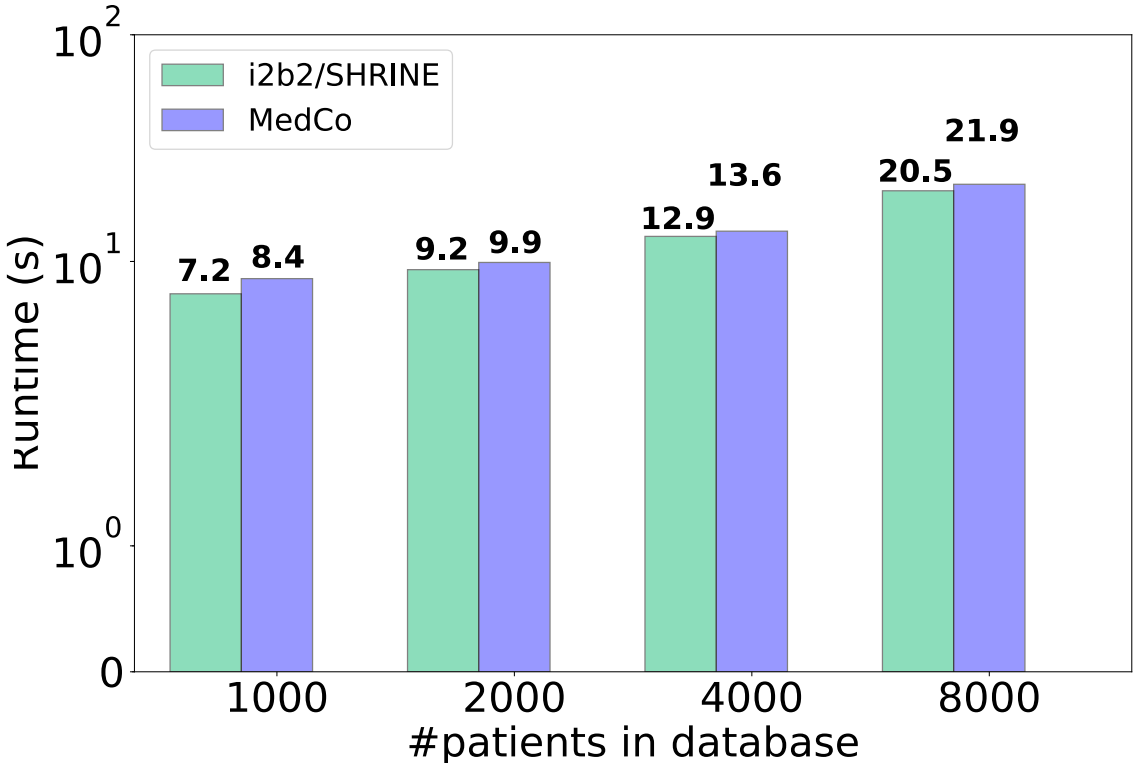
1) (user) Query Generation
2) (distributed) Query Tagging
3) (local) Query Processing
4) (local) Result Aggregation
5) (local) Result Obfuscation
6) (distributed) Results Shuffling
7) (distributed) Results Re-Encryption
8) (user) Result Decryption

 Sites' public keys  Sites' secret keys  Collective public key

Test on clinical oncology use case

- Experimental Setup:
  - 3 clinical sites with one SPU each
  - 8,000 patients at each site with >1 million clinical and genetic attributes from The Cancer Genome Atlas





Query: "Number of patients with skin cutaneous melanoma AND a mutation in BRAF gene affecting the protein at position 600."

HW/SW configuration: 3 servers (2.5.GHz Intel Xeon E5-2680 v3 CPU with 12 cores); 256GB of RAM; Postgres SQL; ElGamal on Ed25519 with 128 bits security; UnLynx library (Froelicher et al.'17)

MedCo: Demo Video

Events and Publications

Genomic Privacy and Security

Events on Genome Privacy and Security

- **Dagstuhl** seminars on genome privacy and security 2013, 2015
- **Conference on Genome and Patient Privacy (GaPP)**
  - **March 2016, Stanford School of Medicine**
- **GenoPri**: International Workshop on Genome Privacy and Security
  - July 2014: Amsterdam (co-located with PETS)
  - May 2015: San Jose (co-located with IEEE S&P)
  - November 12, 2016: Chicago (co-located with AMIA)
  - October 15, 2017: Orlando (co-located with Am. Society for Human Genetics (ASHG) and GA4GH)
  - **October 3, 2018: Basel, Switzerland (co-located with GA4GH)**
- **iDash**: integrating Data for Analysis, Anonymization and sHaring (already in previous years)
  - **October 15, 2018: San Diego, CA**
- Inst. For Pure and Applied Mathematics (IPAM, UCLA)

  Algorithmic Challenges in Protecting Privacy for Biomed Data

  January 10-12, 2018
- DPPH Workshop, 17 February 2018

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

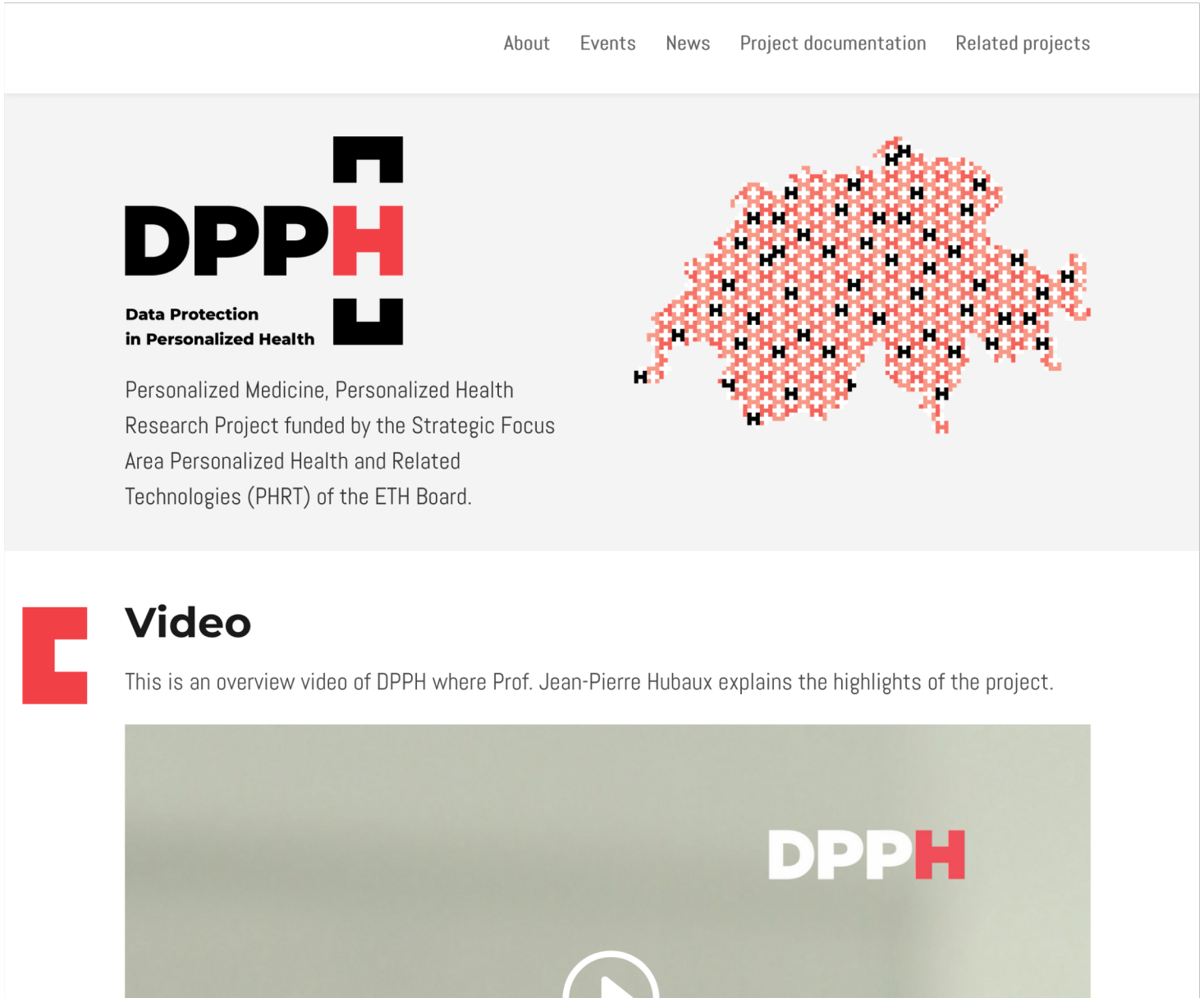
iDASH

ipam

DPPH18

DPPH18.epfl.ch

DPPH: Data Protection in Personalized Health



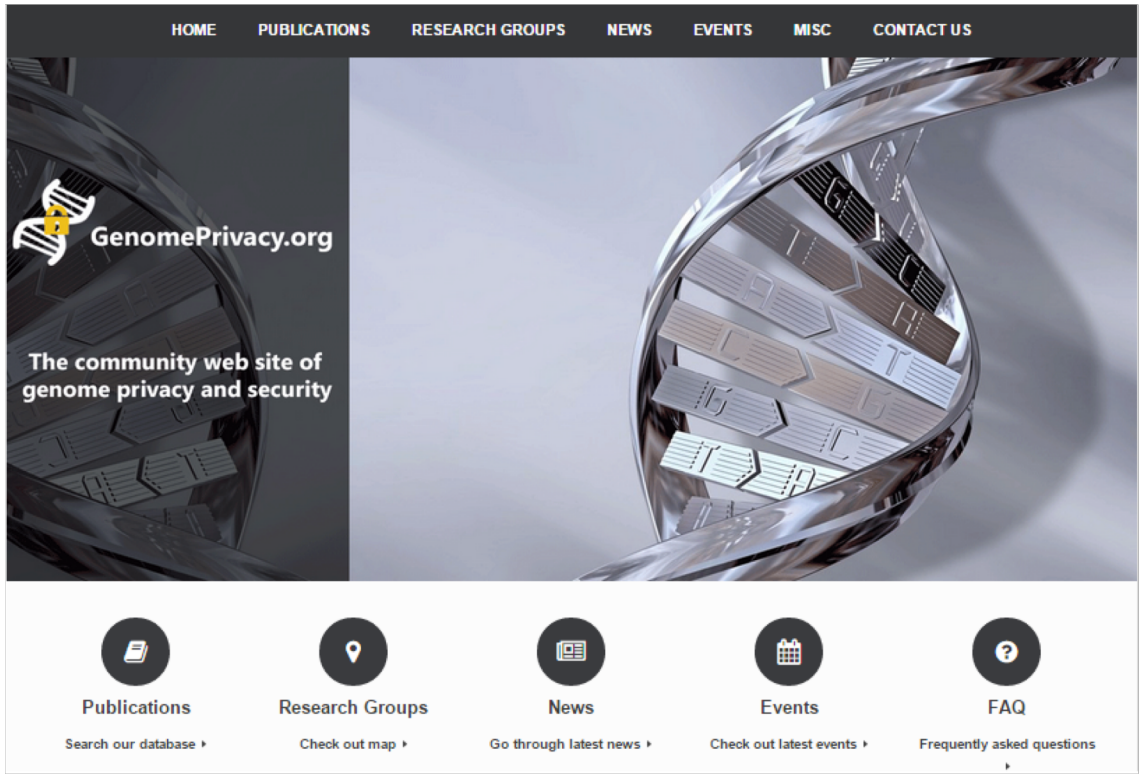
https://dpph.ch

MedCo: Collective protection of medical data

https://medco.epfl.ch

https://dpph.ch

Genome Privacy



https://genomeprivacy.org

https://medco.epfl.ch

https://dpph.ch

Community website

- Searchable list of publications on genome privacy and security
- News from major media (from Science, Nature, GenomeWeb, etc.)
- Research groups and companies involved
- Tutorial and tools
- Events (past & future)

Conclusions

- Worldwide, the confidentiality of health data is **in jeopardy**
- Effective personalized medicine dramatically **scales up the amount of data**, involving also *omics data
- IoT and ubiquity of **wearables** collecting health-related data will further increase the risk
- **Post-quantum** cryptographic mechanisms for secure processing, further long-term secure processing needed.
- ***omics data** requires additional protection mechanisms
- The project *Data Protection for Personalized Health* addreses data sharing in medical environments, applying **practical and effective** tools such as MedCo

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Responding to the Security and Privacy Challenges Raised by Personalized Medicine

Challenges in Data-Driven Genomic Computing

Como, Italy, March 8th 2019

Juan Ramón Troncoso-Pastoriza

juan.troncoso-pastoriza@epfl.ch

DPPH
Data Protection
in Personalized Health

Strategic Focus Area
Personalized Health
and Related Technologies