DPPH: Data Protection in Personalized Health

Personalized Medicine, Personalized Health Research Project
funded by the Strategic Focus Area Personalized Health and Related Technologies (PHRT) of the ETH Board
https://www.sfa-phrt.ch/

Jacques Fellay	Bryan Ford	Jean-Pierre Hubaux	Dimitar Jetchev	Effy Vayena	Olivier Verscheure
Head of Fellay Group, School of Life Science, EPFL & Head of Precision Medicine Unit, CHUV	Head of DEDIS lab, School of I&C, EPFL	Head of LCA1, School of I&C, EPFL (coordinator)	Head of GR-JET, School of Basic Sciences, EPFL	Head of Health Ethics and Policy Lab, Dept of Health Sciences and Technology, ETH	Executive Director, SDSC

P4 (Predictive, Preventive, Personalized and Participatory) medicine is called to revolutionize healthcare by providing better diagnoses and targeted preventive and therapeutic measures. However, to accelerate its adoption and maximize its potential, clinical and research data on large numbers of individuals must be efficiently shared between all stakeholders. The privacy risks stemming from disclosing medical data raise serious concerns, and have become a barrier that can hold back the advances in P4 medicine if effective privacy-preserving technologies are not adopted to enable privacy-conscious medical data sharing. The evolution of the regulation towards further guarantees (e.g., HIPAA in USA and the new GDPR in EU) reflects this urgent need.

Pairing privacy-conscious data sharing with recent advances in the field of *omics and, in particular, in high-throughput sequencing technology, leads to an explosive growth in the amounts of available data; this big data scale can usually not be handled with current hospital computing facilities, hence the need for elastic computing resources that can cope with huge amounts of data in a secure and privacy-aware infrastructure, supporting data processing and sharing.

This project seeks to address the main scalability, privacy, security and ethical challenges of data sharing for enabling effective P4 medicine, by defining an optimal balance between usability, scalability and data protection, and deploying an appropriate set of computing tools to make it happen. The target result of the project will be a **platform composed of software packages that seamlessly enable clinical and genomic data sharing and exploitation across a federation of medical institutions, hospitals and research laboratories across Switzerland in a scalable, secure, responsible and privacy-conscious way, and that integrates widespread cohort exploration tools such as i2b2, TranSMART or SHRINE** (see Fig. 1). This main goal materializes in the following outcomes that the project expects to deliver:

- A holistic **requirements analysis** of the medical data sharing ecosystem, from the standpoint of legal, ethical and medical stakeholders, including a roadmap for progressively addressing these requirements in the clinical and research practice. Aligned with this analysis, the project will keep liaisons with other PHRT/SPHN projects running in parallel, considering their scenarios and accounting for them in the developed security and privacy framework.

- A **scalable scientific computing infrastructure**, building on top of SDSC's data science framework[1], enabling elastic big data processing and powerful knowledge management enabling data traceability. For this purpose, scalable distributed filesystems will be revisited and adapted to work with homomorphically encrypted data; specifically, the employed big data solutions will rely on open-source standard tools such as Apache Spark and Hadoop Distributed File System (HDFS).

- Software-based solutions for **accountable and privacy-preserving data sharing** featuring trust distribution across a federation of sites with no single points of failure, leveraging existing prototypes such as UnLynx and ByzCoin, and improving on their performance by means of novel approaches based on lattice homomorphic encryption, secure multiparty computation, distributed ledger technologies (a.k.a. blockchains), and distributed access control systems. These solutions will be integrated in the scalable computing infrastructure of SDSC and incorporate and adapt widespread tools for cohort exploration and data analysis such as i2b2, TranSMART and SHRINE, and it will be prototyped and validated in a real-size scenario.

[1] https://datascience.ch

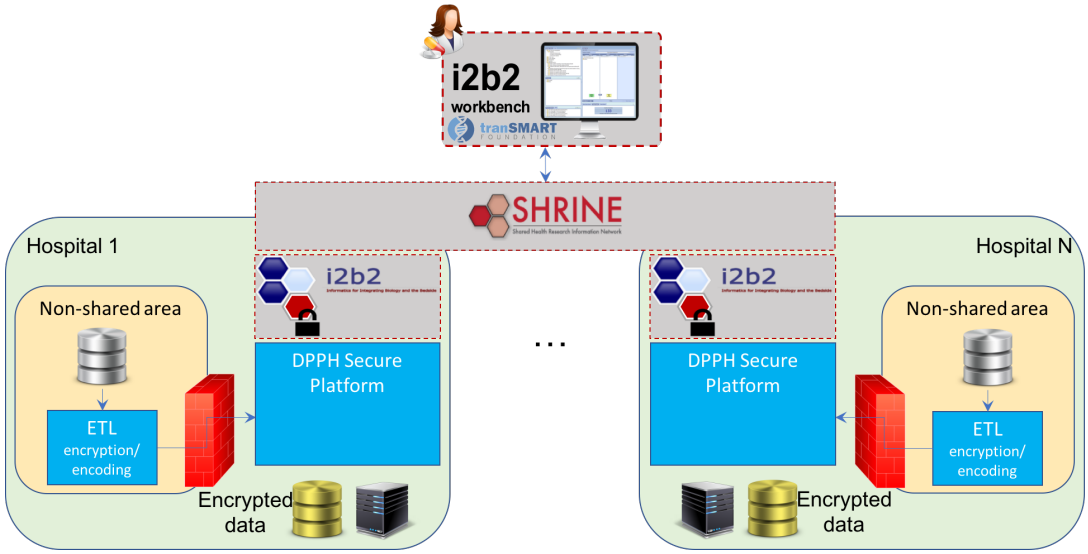

Figure 1. *Architectural view of the proposed system. ETL: "Extract, Transform, Load"*

- A quantitative **analysis of inference risks, and countermeasures** for addressing them when releasing aggregated results on patient data, enabling decision-support systems in medical and *omics scenarios. The patient dimension will also be accounted for by exploring the privacy and security risks of **mHealth technologies**, by proposing and evaluating solutions aimed at avoiding that pervasive apps can discover the usage of a health-related sensitive app in patients' devices. Additionally, a comprehensive **ethical analysis** of distributed platforms for medical data sharing from a normative point of view and through qualitative research.

The first milestone of the project comprises the outcomes of the requirements analysis after the first six months, which will drive an agile development methodology to produce an early prototype to be deployed at the first year (second main milestone), followed by the development and integration of the aforementioned privacy, security and scalability solutions into a fully functional prototype (third milestone), which will be deployed and validated by the end of the project (fourth and final milestone) in real and practical use cases aligned with the liaised projects.

By bringing together the extensive expertise of the involved groups in genomic privacy and protection of medical data (Hubaux), secure distributed systems (Ford), big-data and knowledge management (Verscheure), cryptographic techniques for decentralized machine learning (Jetchev), ethics of biomedical research (Vayena) and genomic research and precision medicine (Fellay), this project is meant to combine knowledge from the data science, computer science, ethics, medicine and genomics communities to effectively tackle the challenges currently thwarting data sharing for P4 medicine. By establishing liaisons with other PHRT/SPHN projects, DPPH seeks to cover the Swiss national level, targeting prototypes at a national scale, and by leveraging on already established connections with the Global Alliance for Genomics and Health (GA4GH) and its Software Security Group, DPPH also guarantees international relevance and consistency, strengthening the impact of the project.

In summary, this project will produce **innovative technical solutions** to many privacy-related issues in close collaboration with clinicians, medical researchers and hospital IT specialists in various clinical settings, and **provide the personalized health community with optimal resources to perform cutting-edge research in an ethical and privacy-conscious way**.

Project Factsheet:

- ***Consortium***: LCA1 (EPFL, coordinator), DEDIS (EPFL), GR-JET (EPFL), Fellay Group (EPFL), Health Ethics and Policy (ETHZ), SDSC (EPFL/ETHZ)
- ***Funding***: CHF 2,984,500
- ***Duration***: 3 years
- ***Kick-off Event***: February 15th 2018
  Workshop on Secure, Privacy-Conscious Data Sharing
  http://dpph18.epfl.ch