The researcher's new path to clinical data

The BioMedIT project and the work of the SPHN Data Coordination Center

Dr. Katrin Crameri

Personalized Health Informatics Group
SIB Swiss Institute of Bioinformatics

September 2019

A project of SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences
SIB Swiss Institute of Bioinformatics

The dataflow

The dataflow **requirements and prerequisites**

Finding data

- Distributed, federated query system: Reliable estimation of the availability of data, samples and/or patients for planned research projects
- Data Catalogue: Identification of specific data sets, samples or patients for research projects

Requesting data

- Governance processes on the data-provider side
- Compatible IT infrastructure solutions at the hospitals
- Data processing, data extraction, coding (pseudonymization) and encryption of data

Receiving data

- Data Transfer and Use Agreement (DTUA)
- Ethics approval
- Data privacy and IT Security training
- Secure transfer of data

Using data for research

- Interoperability of data
- Secure and protected IT environment
- Compute capacity
- Cutting edge technologies and tools

The dataflow **coordination and implementation**

Personalized Health Informatics Group

- SPHN Data Coordination Center
- BioMedIT Project

Finding data

Requesting data

Receiving data

Using data for research

Collaboration Agreements between SPHN and UHs

DCC Working Groups

Hospital
- Clinical Data Semantic Interoperability WG
- Hospital IT WG

BioMedIT
- BioMedIT Interoperability WG
- BioMedIT Security WG

Research
- Bioinformatics & Data Analytics WG

BioMedIT network

BioMedIT nodes

1. Core-IT/Vital-IT
2. sciCORE
3. SIS

SPHN Infrastructure Dev. Projects
SPHN Driver Projects

The dataflow **requirements and prerequisites**

Finding data

- Distributed, federated query system: Reliable estimation of the availability of data, samples and/or patients for planned research projects
- Data Catalogue: Identification of specific data sets, samples or patients for research projects

Requesting data

- Governance processes on the data-provider side
- Compatible IT infrastructure solutions at the hospitals
- Data processing, data extraction, coding (pseudonymization) and encryption of data

Receiving data

- Data Transfer and Use Agreement (DTUA)
- Ethics approval
- IT Security training
- Secure transfer of data

Using data for research

- Interoperability of data
- Secure and protected IT environment
- Compute capacity
- Cutting edge technologies and tools

Finding data
Distributed federated query system

- The federated query system allows researchers to run simple queries against a subset of health related data of all university hospitals
- The platform provides information about the number of patients according to the query
- Basic set of search criteria, will be extended
- Helps researchers to determine the feasibility of conducting a research project and work through hypothesis generation by modifying the searches

Finding data
Distributed federated query system

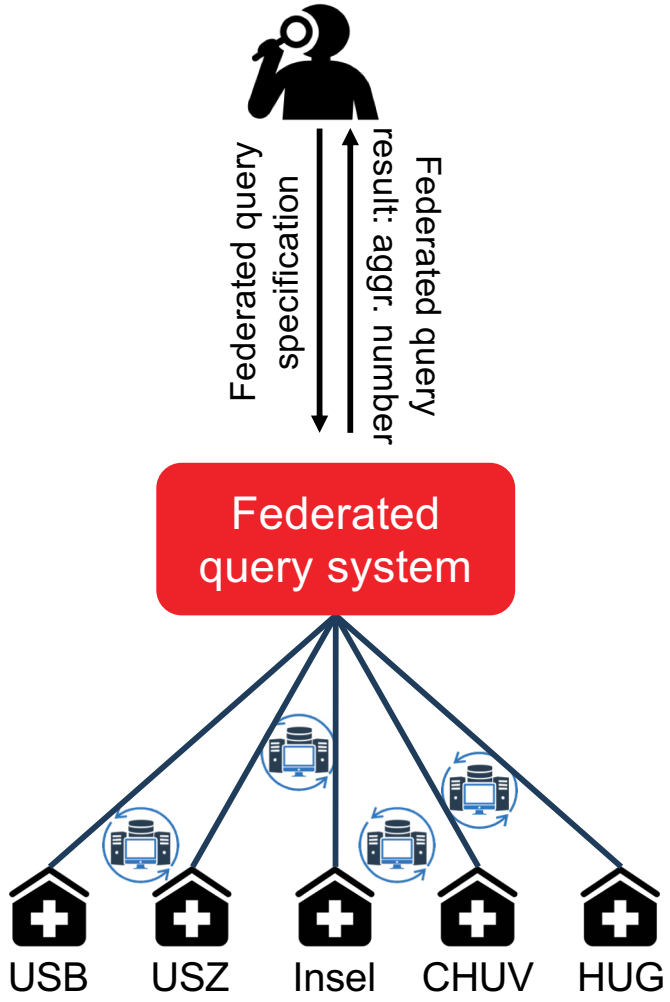
Status of implementation / timeline:

- Consortial collaboration agreement (6 of 7 parties approved)
- Installation completed at 2 sites, initiated at 3 sites
- Access and Use Policy to be adopted be the SPHN National Steering Board

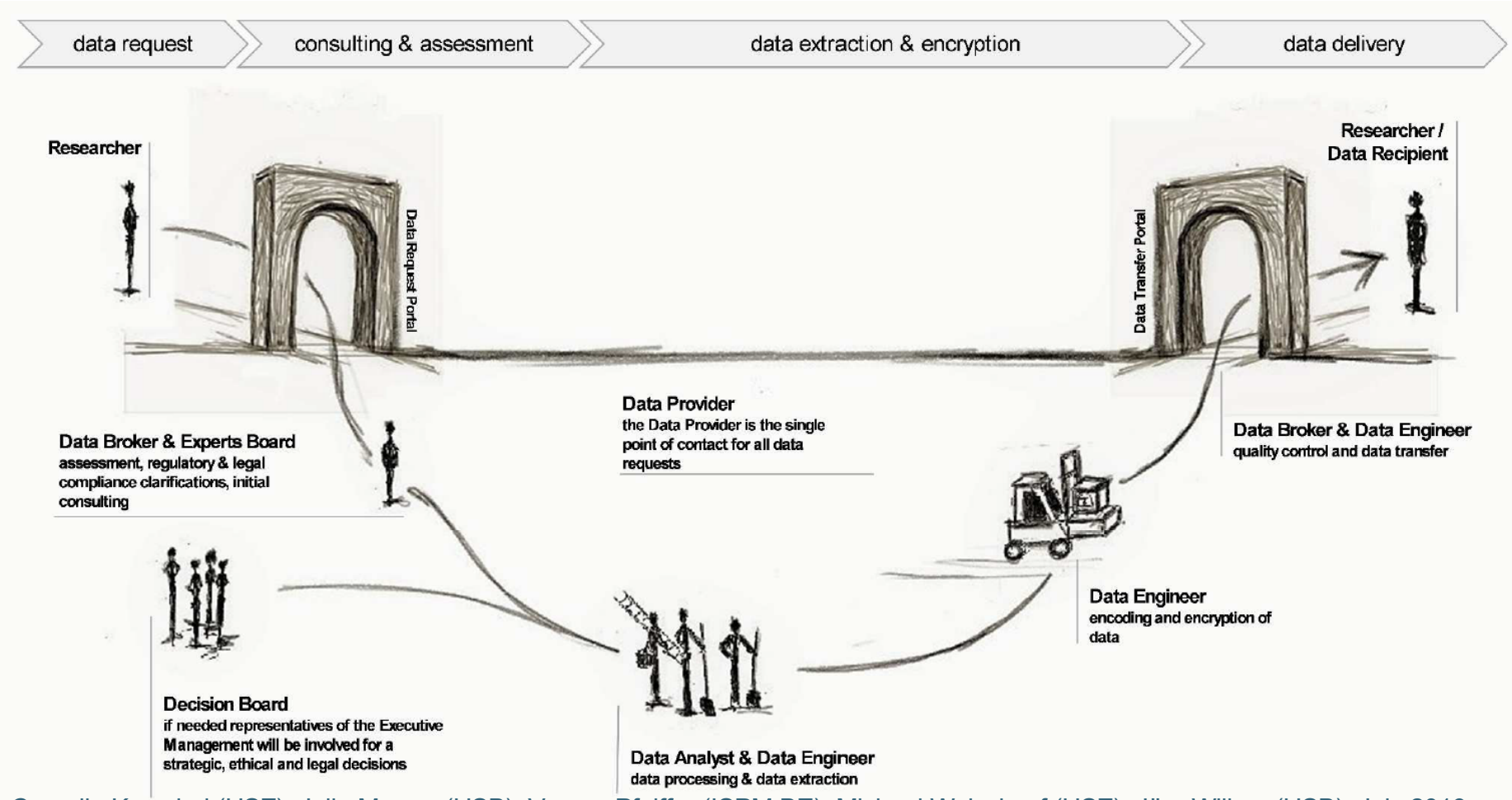
→ Test queries in Q4 2019

→ Optimisation of the system

→ Real data queries in 2020



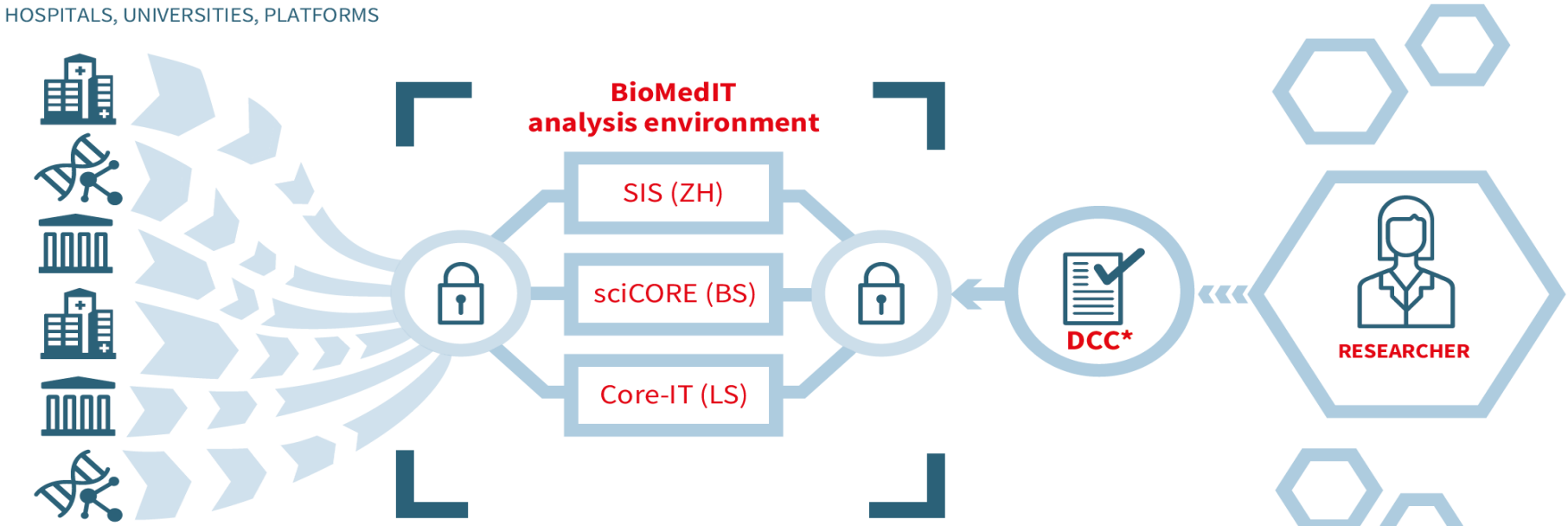
Requesting data
Workflow in the University Hospitals



Cornelia Kruschel (USZ), Julia Maurer (USB), Verena Pfeiffer (ISPM BE), Michael Weisskopf (USZ), Jörg Willers (USB), July 2019

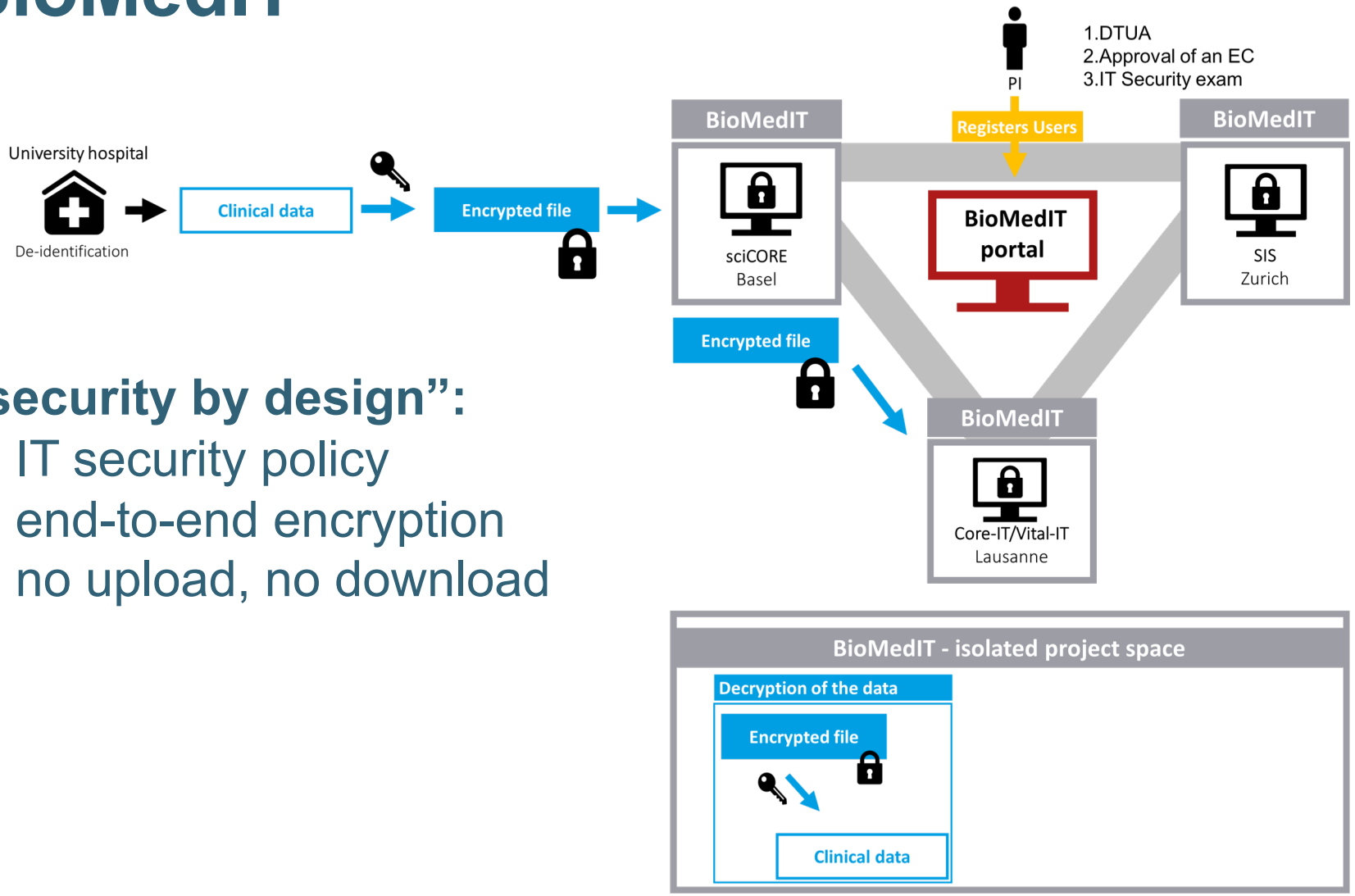
Receiving data
The BioMedIT network

BioMedIT provides researchers with access to a secure and protected computing environment for analysis of sensitive data without compromising data privacy



*DCC: Data Coordination Center

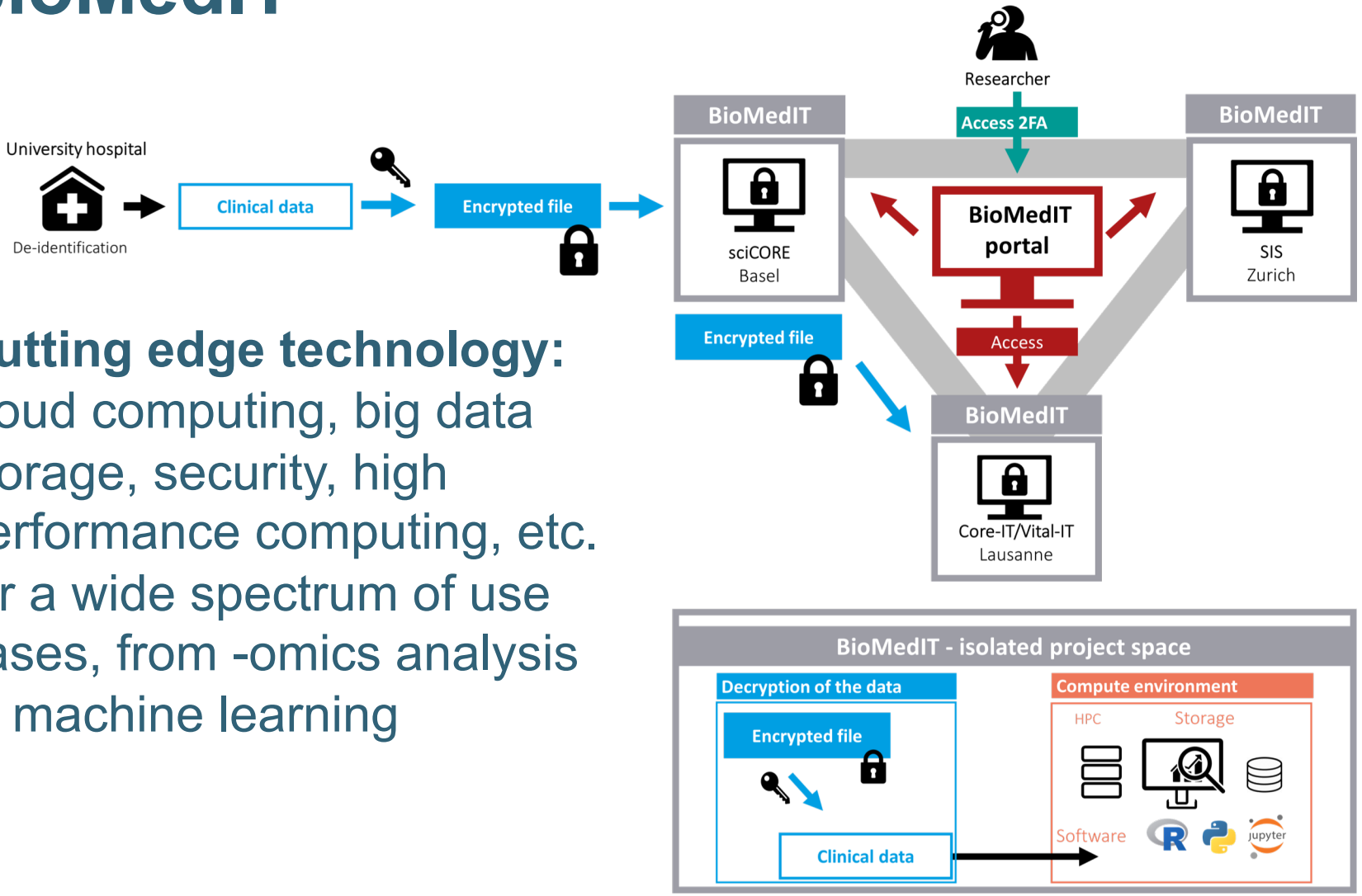
Receiving data
BioMedIT



“security by design”:

- IT security policy
- end-to-end encryption
- no upload, no download

Using data for research
BioMedIT



Cutting edge technology:
cloud computing, big data storage, security, high performance computing, etc. for a wide spectrum of use cases, from -omics analysis to machine learning

Using data for research
Technical and semantic interoperability

Interoperable local clouds, one IT security policy, connection to all data providers (PHRT Technology Platforms, hospitals and university laboratories);

Semantics (definition, understanding, standards), description formalism (transport and storage), data models (exchange and analytics);

Investment in future data use: Federated computing (bring the analysis to the data, containerized workflows)

Conclusion

- Current state: Challenges identified, most projects in the prototype phase, integration into an overall infrastructure ongoing
- Process-innovation: SPHN as a research infrastructure initiative impacts also the care system
- Biggest challenges: Scalability and sustainability

Swiss Personalized Health Network

dcc@sib.swiss

A project of

SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences

SIB Swiss Institute of Bioinformatics