

Where to analyze sensitive human data?

BioMedIT: Switzerland's Trusted Research Environment

Dr Katrin Cramer, MPH

Director, SPHN Data Coordination Center and Personalized Health Informatics

SIB Swiss Institute of Bioinformatics

A project of



9 October 2023



The Swiss Personalized Health Network (SPHN)


Creation of a scalable and sustainable data-enabling environment

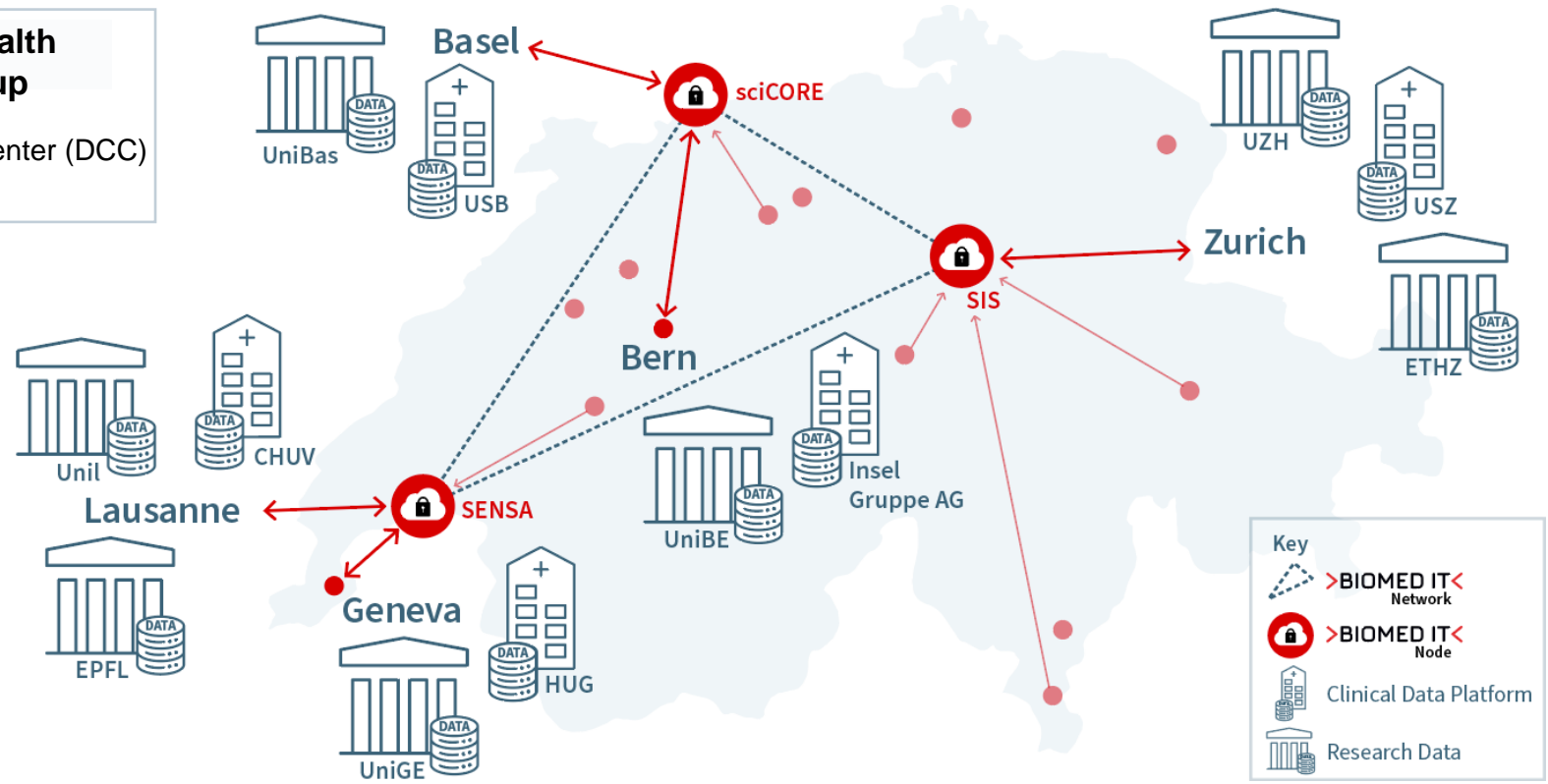
- Including routine health data, molecular / omics data, registry data, clinical research data, and other health-related data types
- Research infrastructure initiative funded 2017-2024 by the Swiss Government with CHF135 million
- Operating under a common Ethical Framework and one Information Security Policy, incl. the setup of a Trusted Research Environment

→ Enable institutions to responsibly share interoperable health data

→ Enable researchers to access, integrate, and analyze data

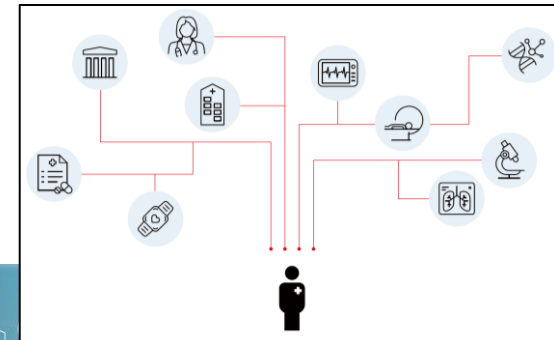
The Swiss Personalized Health Network


Personalized Health Informatics Group
 SPHN Data Coordination Center (DCC)
 BioMedIT Network



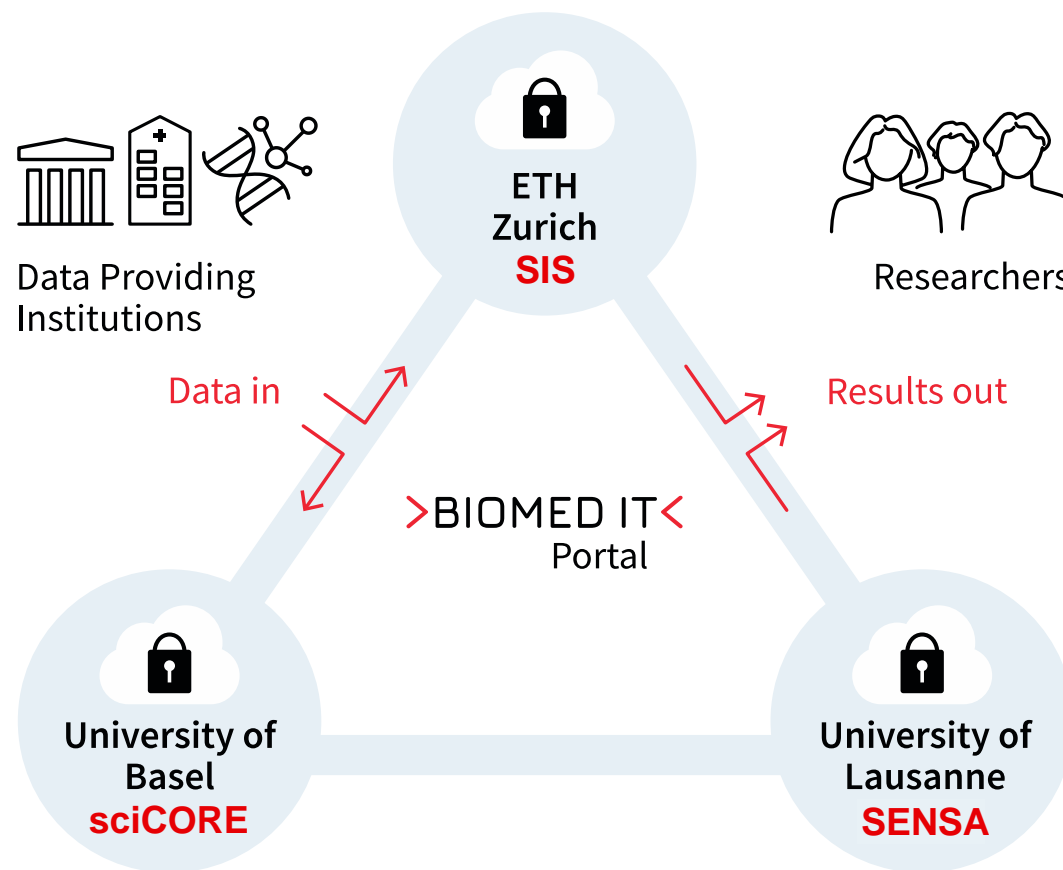
Systematic use of large amounts of health and human omics data: What does it take?

- Strong capabilities in clinical bioinformatics, computational biology, and computational service infrastructure
- Secure data mobilization
- High-performance IT infrastructures for big data computing and storage
- Security measures for ICT systems to protect confidential information



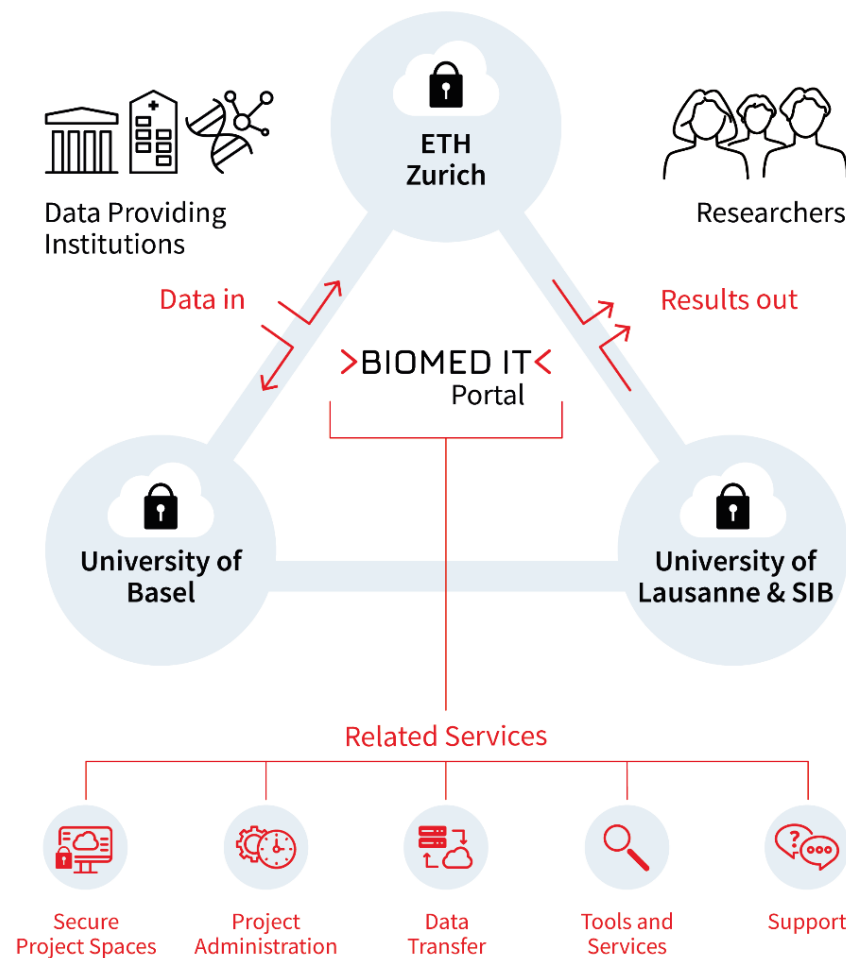
BioMedIT: Switzerland's Trusted Research Environment

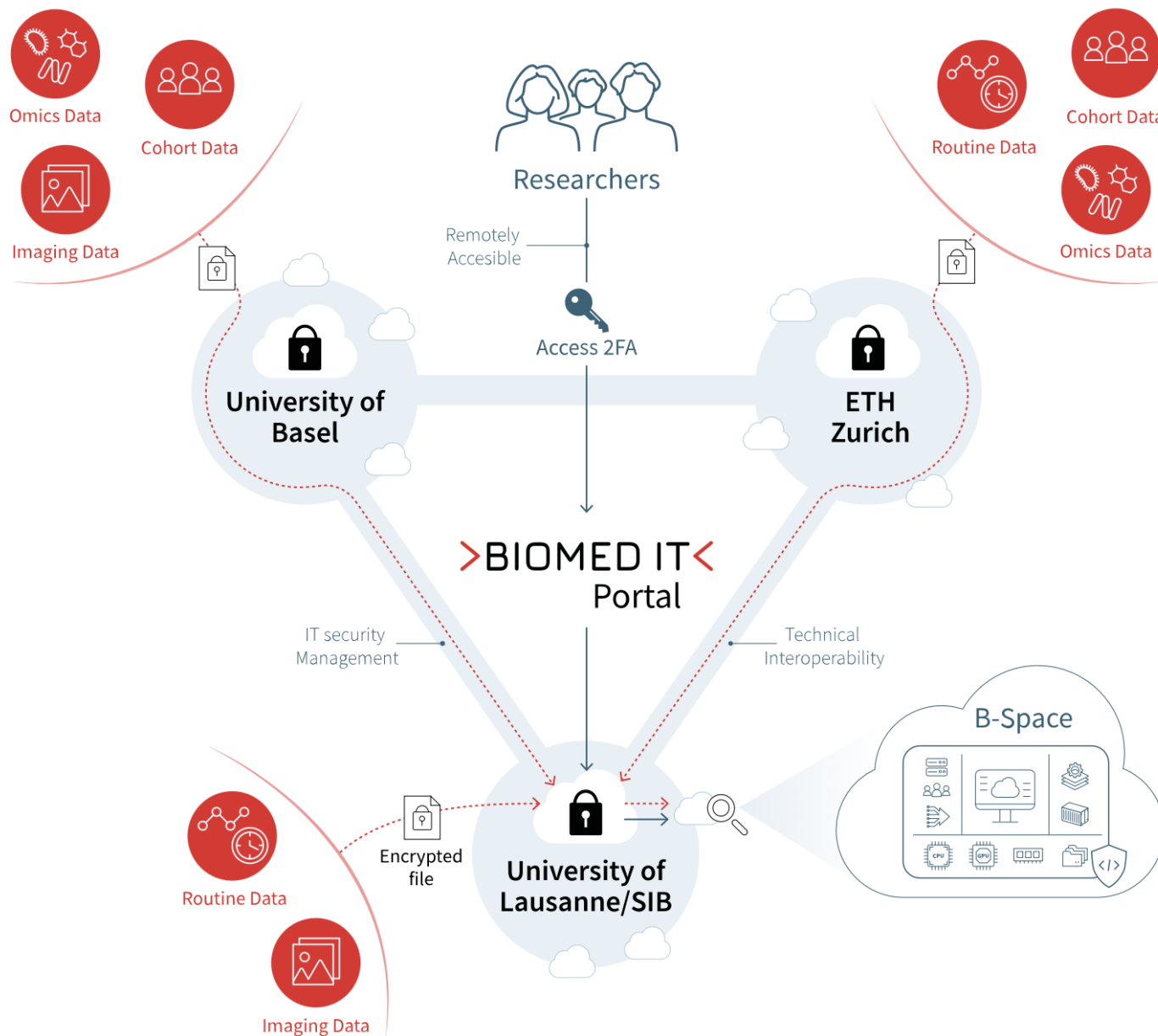
A secure way to perform digital research on highly sensitive data



BioMedIT: Switzerland's Trusted Research Environment

- Secure cloud and HPC offerings
 - Central [IT Service] Management (SIB)
 - 3 physical nodes (at Swiss Universities) providing scientific IT support
 - BioMedIT Portal as entry point, central tools and services
 - Security by design: Common Information Security Policy (Swiss standard)
- Secure mobilization, processing and storage of sensitive research data
- End-to-end encrypted data transfers (via sett)

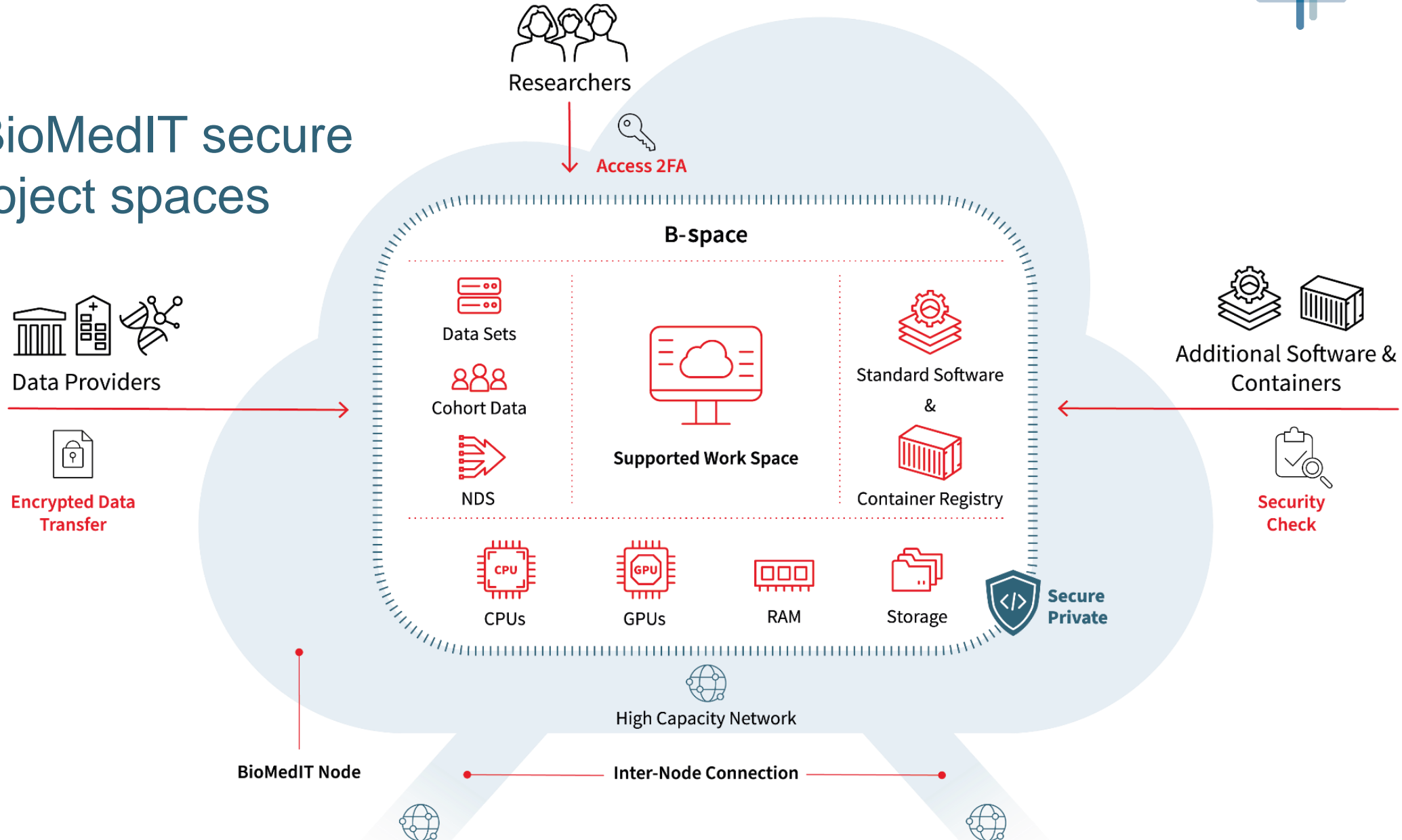




Data transfers from data providing institutions

- SFTP:
IP whitelisting
“snowflake architecture”
- HTTPs:
point-to-point

B-spaces: BioMedIT secure research project spaces





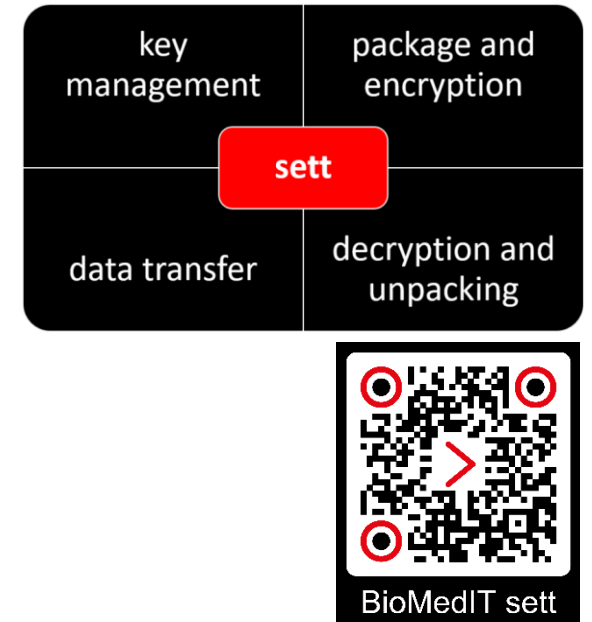
Parameter	sciCORE	SENSA	SIS
CPUs	2500	800	6000
GPUs	12	6	352
Ram (TB)	20	3	60
Storage (PB)	2	0.7	2.6
Cloud platform	openstack	openstack	openstack
HPC scheduler	slurm	slurm	slurm
File system	Ceph, NFS	Weka	Ceph/CephFS

- > 110 local, national and international projects
- > 30 data providers connected
- > 680 users



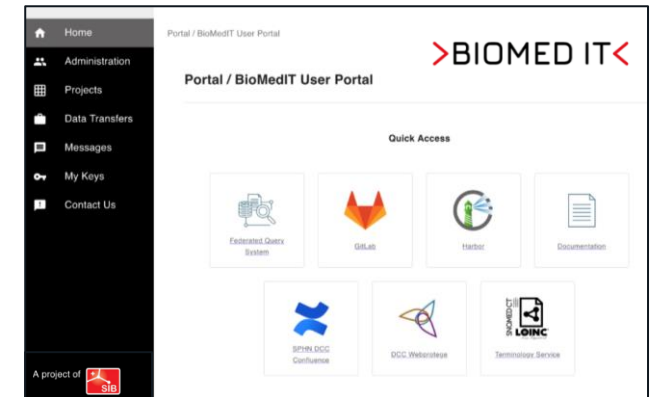
Secure Encryption & Transfer Tool (sett)

- Automation around standard packages
GPG, tar, SFTP, etc.
- Supports data encryption/decryption,
compression, check-summing
- Supports transfer via SFTP, LiquidFiles & HTTPs
- PGP key management: Key creation,
deletion, up-/download from key-server
- GUI / CLI versions available
- Open source and re-usable, contributions welcome



BioMedIT Portal: Enabling Secure User Access and facilitate usage

- Login (2FA) using SWITCH edu-ID/AAI
- Project and user management
- Data Transfer Requests and status overview
- View GPG keys on the BioMedIT key server
- Access project spaces (research data) on the nodes via Remote Desktop / SSH
- One-stop access to other central services
- Open source and re-usable, contributions welcome



Flavours of TRE



Oil pipeline

Serves a single purpose, but very effectively

If you want to go somewhere new, build a new pipe.

Once it is running, it runs well; high costs



Train system

Can be used in various ways, but with strong limitations

If you want to go somewhere new, build a new track.

Very safe and efficient but with many limitations; high costs



Car

Can be used for many purposes, in many ways.

Goes wherever there are roads, highly flexible.

Safety determined by how it is built, AND very much how you use it; moderate costs

>BIO
MED
IT<

BioMedIT: A flexible approach to a TRE

Flexible and secure workspaces (B-spaces) where researchers can perform their analysis in ways they are used to

- Security awareness training for all users, analogous to ‘driving licence’
- Secure and controlled management of data import to B spaces
- Remote access to B spaces where researchers can run their application, workflows, containers (security checks required)
- Tight control of access to web services and data export

Nodes commit to a level of transparency, coherency and engagement that could not be expected or received when using a commercial cloud

BioMedIT nodes are operated by scientific IT providers with experience, technical and scientific knowhow and an understanding of the customers they support

The BioMedIT clients / use-cases

- Multi-site collaborative research projects (with data coming from different sources and a dataset accessed by researchers from various institutions)
 - Bilateral collaborative projects between national or international institutions involved
 - Institutional “outsourcing” of e.g., data management, ML/AI applications, compute needs, etc. by individual hospitals/institutions
 - Local projects of node-owning universities (own tenants, B-spaces, etc.)
- BioMedIT acts as processor on behalf of the data controllers (Datenverarbeitungsvereinbarungen, DTPAs)

BioMedIT Security Concept



Federated

Bring together multiple providers in order to deliver services. Internal transparency within federation.



Policy driven

Define central shared policies. Mandate required local policies. Focus on outcome not specific implementation.



Risk based

Assess risk levels of vulnerabilities, pseudonymization and information assets.



Agreement supported

All use based on standard documented agreements between data providers, research groups and BioMedIT nodes.



Security Awareness

Training and support to ensure that data providers, node staff and research teams understand their responsibilities.



Controlled access

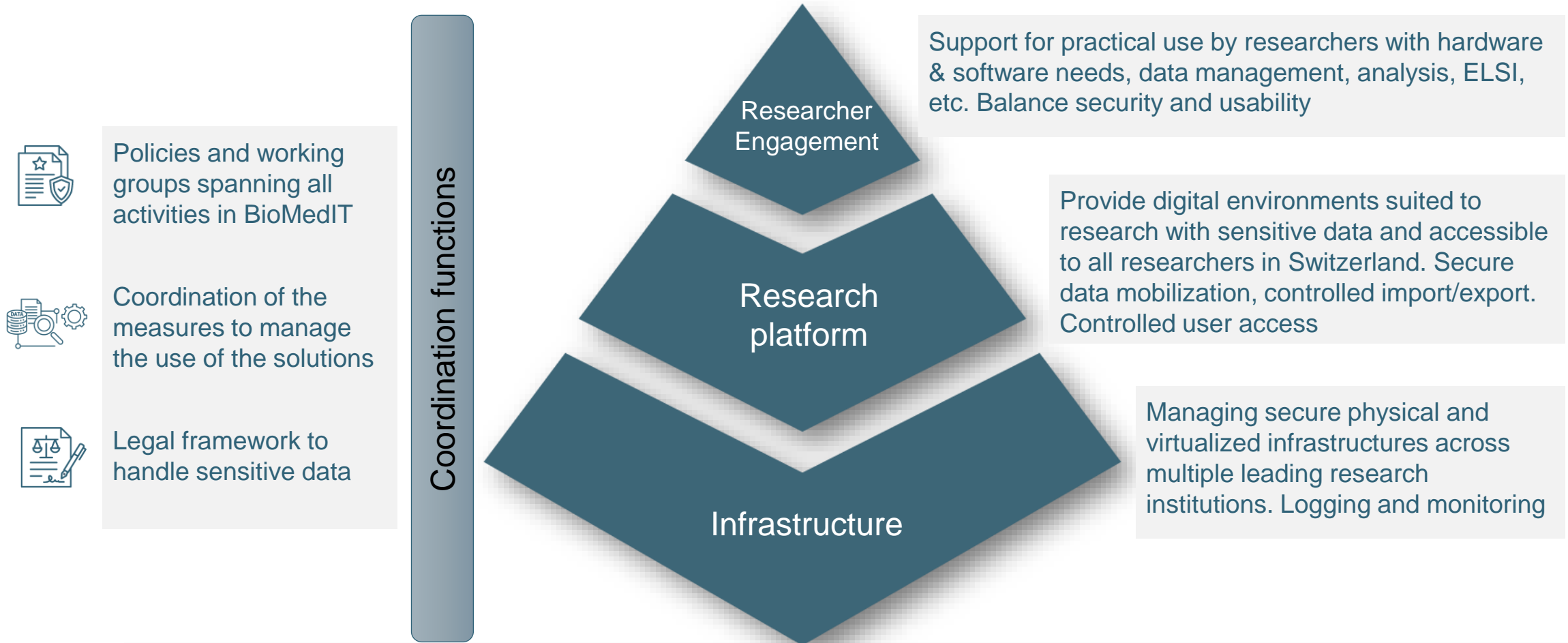
Central user management tied to national identity systems to manage users across projects and nodes.



Controlled transfer

Regulated input and output, requiring specific tools and procedures to ensure traceability and security.

BioMedIT: Architecture for a complete secure digital research solution



BioMedIT Funding (TRE only)

Network financing:

- Basic node funding (hardware, personnel): CHF 300K per node p.a.
- Additional funding (software, licences, consulting): CHF 100K per node p.a.
- Tools and Interoperability: 1 FTE per node p.a.
- Security: 1 FTE per node p.a.
- Data management and research support: 1 FTE per node p.a.
- Nodes provide 100% in-kind matching-funds

Central team financing:

- Personnel, hard- and software, projects, licenses, running costs: CHF 1'500K p.a.

BioMedIT: Lessons learned and remaining challenges

There are many challenges to face, most are not technical

- Tools and technologies (and money) help, but they only *support* solving problems
- Greater challenges are humans, cultural & organisational habits, particular interests, compliance, etc.

Data harmonization and interoperability is crucial

- Health data is too diverse, cannot automate the problem away
- Standardisation is a slow process, involves all stakeholders of care and research

Federated approaches are required, many groups to align

- Big data means multiple providers (hospitals, tech. platforms, other data sources)
- Modern science means multidisciplinary, collaborating research teams
- Trust is the most important currency in achieving harmonization and collaboration

BioMedIT: Lessons learned and remaining challenges

Legal requirements are strict but vague

- Many restrictions, instructions, requirements, interpretations
- No actual guidance and many opinions on *how* to achieve them

- Art. 7 Data security

¹ Personal data must be protected against unauthorised processing through adequate technical and organisational measures.

Security harmonization creates complex engagement between partners

- Technical alignment touches many internal parts of organizations, not just the surface
- Ensure compatibility of policies and day-to-day procedures without a top-down leverage
- Developing a shared understanding of risks, the means to control them, and the tolerance to accept them in certain cases

Summary and outlook

- Trusted Research Environments are crucial to gain trust of data providers and patients/citizens
- They help minimizing the risk when risk-based de-identification approaches are applied
- Depending on their architecture and use, TREs are complex and expensive to coordinate and operate because there is no one-size-fits all in biomedical data-driven research
- Everyone demands data protection and security, but long-term funding for secure infrastructures (as a national service) is hard to come by
- Reconciling high security requirements and user needs/expectations is difficult

Acknowledgements

The PHI Group:

Katrin Crameri, Deepak Unni, Harald Witte, Jan Armida, Julia Maurer, Kristin Gnodtke, Michael Müller-Breckenridge, Owen Appleton, Patricia Fernandez Pinilla, Petar Horki, Sabine Österle, Shubham Kapoor, Sergio Guarino, Simone Guzzi, Vasundra Touré

The **SPHN NSB** and **NAB**, **Task Forces** & **WGs**

The **BioMedIT Board** and **workforces** @ ETHZ, Unibas, Unil, **SIB Management**

The **Hospital workforces** @ USZ, USB, CHUV, Insel, HUG

The SPHN Management Office:

Thos Geiger, Liselotte Selter, Sarah Vermij, Christine Remund, Michaela Egli



@CrameriKatrin

@SPHN_ch



Katrin.Crameri@sib.swiss

dcc@sib.swiss | info@sphn.ch



www.sphn.ch | www.sib.swiss/phi

www.BioMedIT.ch