SPHN: Building an interoperable environment for biomedical research in Switzerland

Thierry Sengstag
SPHN/BioMedIT node head, SIB - Swiss Institute of Bioinformatics and University of Basel

on behalf of, and with contributions from:
SPHN Data Coordination Center
BioMedIT nodes (University of Basel, ETH-Z, SIB)

e-IRG Workshop, CERN
21 May 2019

Swiss Personalized Health Network
SIB Swiss Institute of Bioinformatics

Bridging the gap between data science and health

Rationale

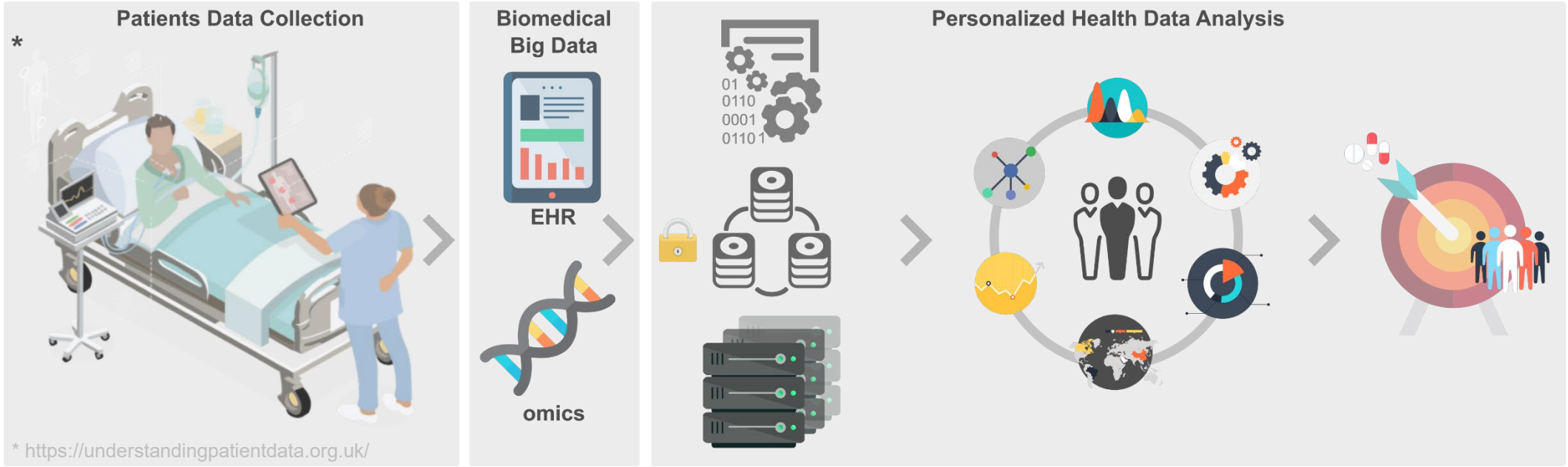
Homo sapiens is now the best characterized model organism for biomedical research

- More data collected in hospitals than in any biomedical research project
- Healthy citizen data increasingly collected with smart mobile technologies

Large, complex datasets require advanced data science techniques

- Computing infrastructures and methodologies

How to leverage these data ethically, legally and efficiently for the benefit of patients, citizens and the society?



Swiss Personalized Health Network SPHN

- A national initiative to promote research in Personalized Health
- Proposed in 2014, started in 2017
- Partners: university hospitals, schools of higher education, research institutes, and other organizations
- Mandated by the SERI (SBFI), co-mandated by the FOPH (BAG)

SAMW ASSM

SIB
Swiss Institute of Bioinformatics

BioMedIT

Create secure environment to work with human data in Swiss academic institutions

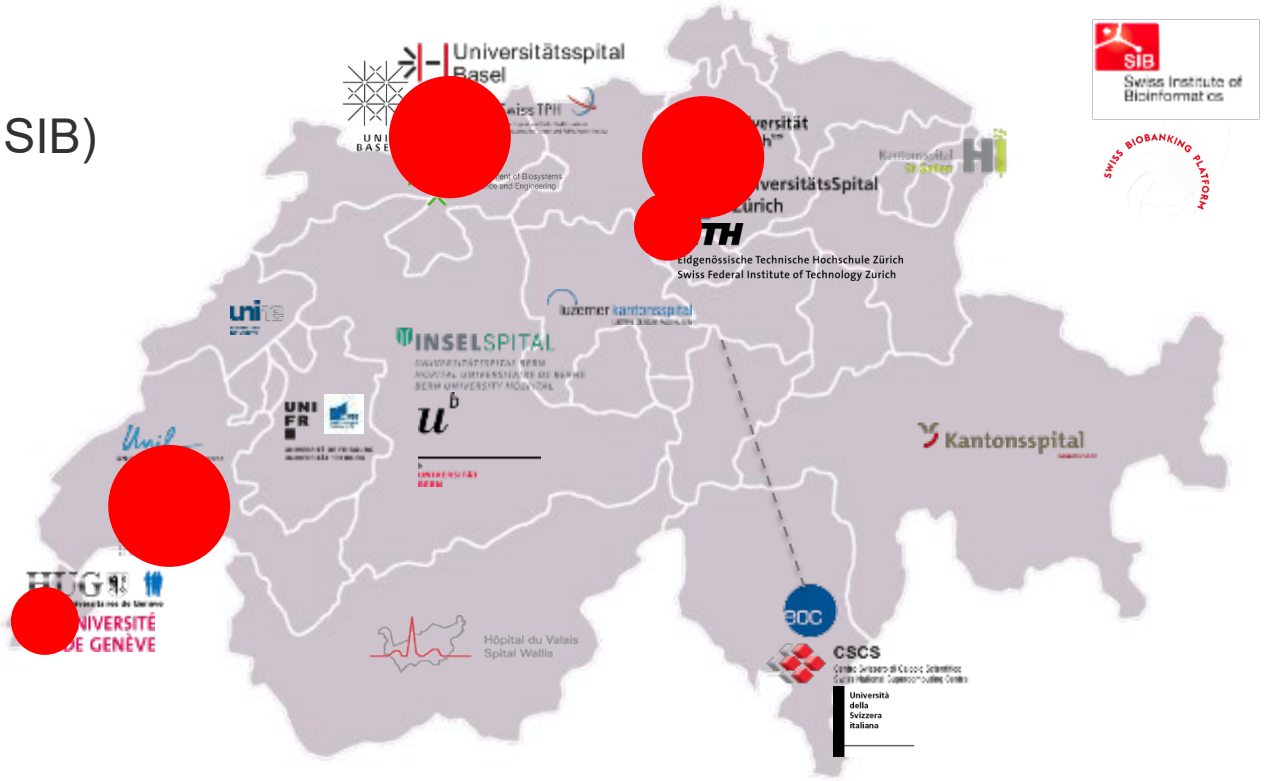
BioMedIT network
coordinated by SPHN DCC (mandate to SIB)

IT infrastructure and project support
- **Core-IT** (SIB, Lausanne)
- **sciCORE** (University Basel)
- **SIS** (ETHZ)

Data infrastructure
- **HES-SO** (Geneva)

Project support
- **S3IT** (University Zurich)

Aligned with SPHN



SPHN, BioMedIT, PHRT

Biomedical research
Healthcare-oriented

SPHN: a SAMW/SIB project with the aim to **promote the development of personalized medicine/health in Switzerland**

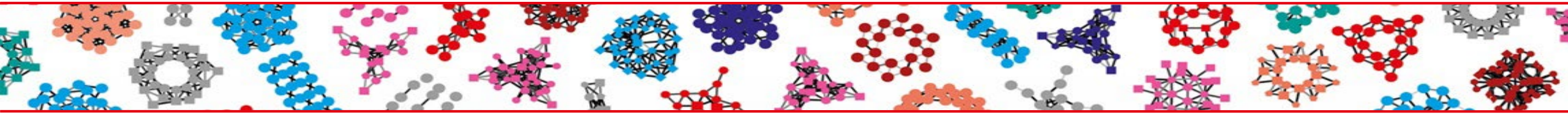
Swiss Institute of Bioinformatics

Bioinformatics IT
Research-oriented

BioMedIT: an SIB project with the aim to **create an IT environment for research with biomedical data in academia**

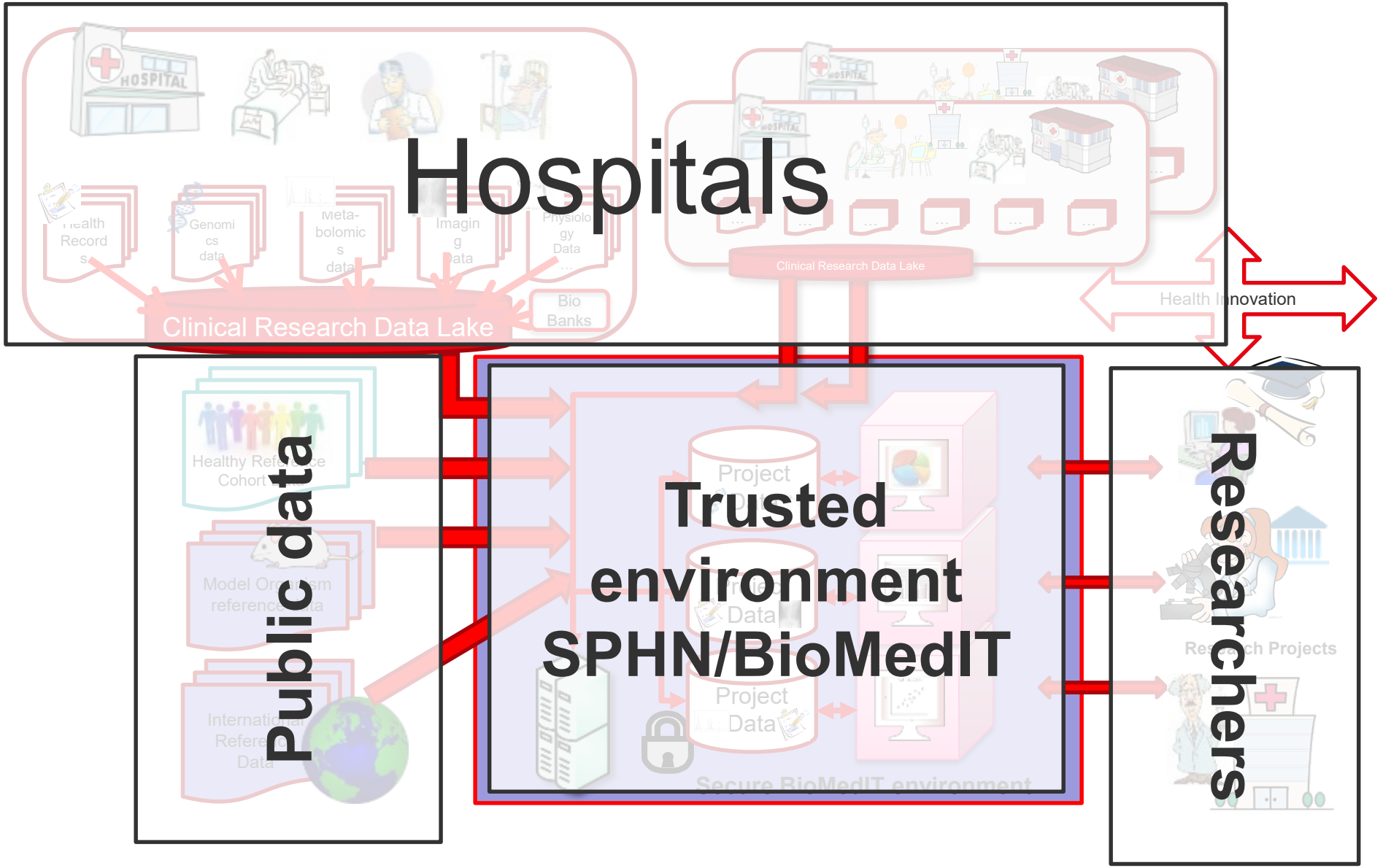
Healthcare technologies
Applications-oriented

PHRT: an ETH project with the aim to **develop new technologies in support of personalized health**



BioMedIT focus: infrastructure

SPHN/BioMedIT – integrating hospital data for research



Anchoring developments in actual research needs

To develop and implement the required infrastructure, SPHN funds the the following project types:

Project type	Description	Funding scheme
Infrastructure implementation projects	Projects that are devoted to build a progressive shareable data system enabling nationwide interoperability of molecular and clinical patient data.	Collaboration agreements
Infrastructure development projects	Projects that thrive to develop and test new technologies, methods and infrastructures for personalized health related research in connection with infrastructure implementation.	Call for proposals*
Driver projects	These projects are based in a concrete research field (e.g. cancer research/oncology) and will push the development of clinical data management systems in all University Hospitals by testing data interoperability & data sharing principles within the whole network.	Call for proposals*

*Coordinated call for proposals with ETH- Domain SFA-PHRT: Joint proposals between ETH researchers and non ETH researchers can be submitted for both type of projects.

Driver projects (call 2017)

	**Swiss Frailty Network and Repository** Heike Bischoff-Ferrari
	**Population-wide screens of the human immune repertoire: a reverse personalized-medicine approach (CCIP)** Adriano Aguzzi
	**Swiss Molecular Pathology Breakthrough Platform (SOCIBP)** Mark Rubin co-funded with Strategic Focus Area Personalized Health and Related Technologies
	**Swiss Personalized Oncology (SPO)** Olivier Michiélin and Mohamed Bentires-Alj
	**PRECISE: Identification of biomarkers and therapeutic targets in inflammatory disease immunotherapy by high-dimensional single cell analysis and cluster proteomics** Manfred Claassen co-funded with Strategic Focus Area Personalized Health and Related Technologies
	**PSSS: Personalized Swiss Sepsis Study**: Detection and modelling of sepsis using machine learning to analyse continuous ICU monitoring, laboratory, microbiology, and -omics data for personalized sepsis management Adrian Egli co-funded with Strategic Focus Area Personalized Health and Related Technologies

+ 8 infrastructure development projects

Domains: patient-data governance, e-consent management,
data de-identification, NLP processing of patient records, variants database (oncology)

Driver projects (call 2018)

	**SACR: The Swiss Ageing Citizen Reference** Nicole Probst
	**CREATE PRIMA: Clinical Research from multi-modality big data sources without proprietary interfaces in a multicenter approach** Joerg Leuppi
	**IMAGINE: Radiomics for comprehensive patient and disease phenotyping in personalized health** Matthias Guckenberger
	**SOIN: Swiss Ophthalmic Imaging Network** Thomas Wolffensberger
	**SHFN: SWISSHEART Failure Network** Christian Matter, Johachim Buhmann

co-funded with Strategic Focus Area Personalized Health and Related Technologies

+ 5 infrastructure development projects

Domains: variants database (non-oncology), reference datasets, QC metrics in radiology, secure distributed computing algorithms

Technical mandates in a nutshell

Hospitals (and supporting platforms)

- Make patient data discoverable (aggregated queries, semantics)
- Use harmonized processes (contracts) for data exchange

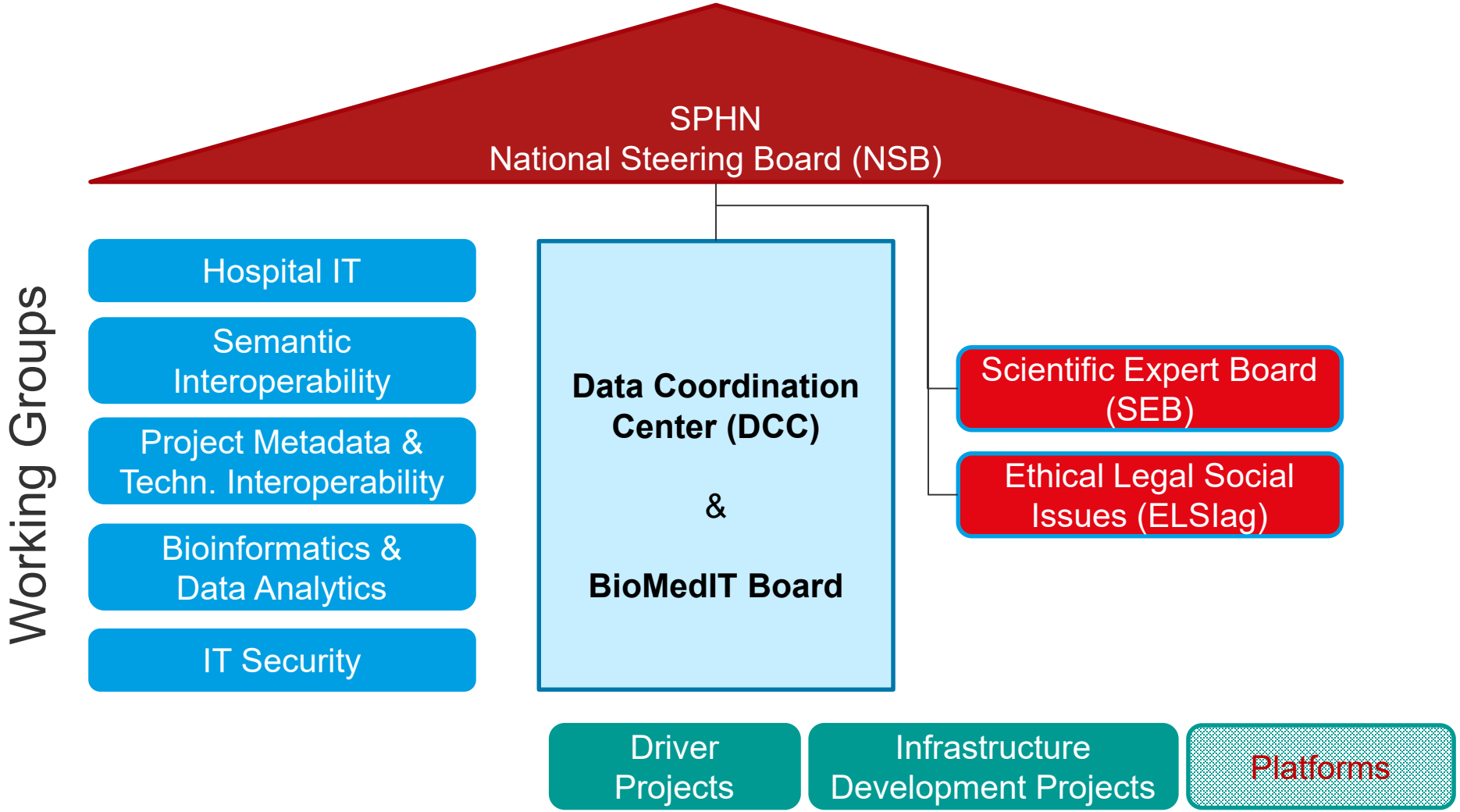
Research organizations

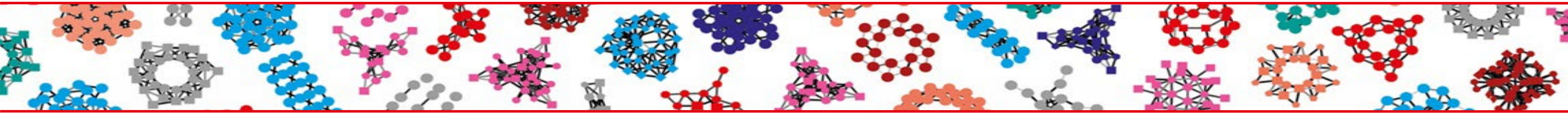
- Establish IT infrastructure adhering to common security standards
- Establish interoperable services (data management and compute) in support of biomedical research
- Make IT infrastructure available for any research project in Switzerland

Researchers (with SPHN funding)

- Help building the infrastructure:
  - by providing real-life research use cases
  - by committing to use the infrastructure

BioMedIT project organization





IT security consensus

Biomedical data are sensitive

Subject to data protection laws and strong ethical requirements
Consensus required between healthcare providers and research organizations

Regulations applicable to biomedical research

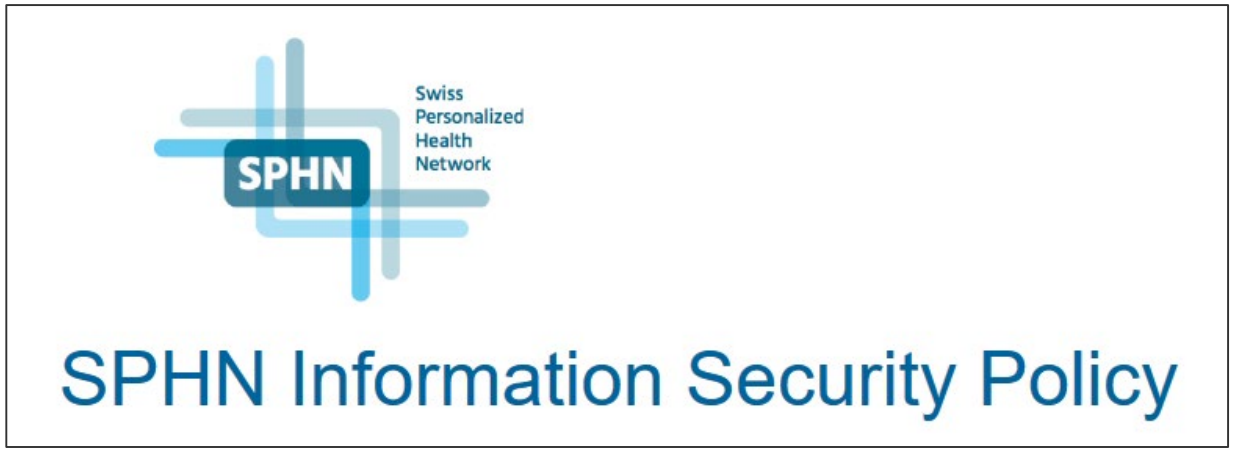
Switzerland:
- Federal Act on Data Protection (SR 235.1)
- Federal Data Protection Ordinance (SR 235.11)
- Human Research Act (SR 810.30)
- Human Research Ordinance (SR 810.301)
- Swiss Penal Code (SR 311.0)

State:
- Data access conditions
- Scope of usage context (**informed consent**)
- General processes of data management
- Consequences in case of misconduct

European Union:
- General Data Protection Regulation (Regulation (EU) 2016/679)

Impact on Swiss regulations anticipated

SPHN information security framework



Initial workshop
12-13 December 2017

Stakeholders:
- Hospitals IT security officers
- Industry consultants
- Academic research IT

Published: 23 August 2018

Swiss regulations
- Federal Act on Data Protection (SR 235.1)
- Federal Data Protection Ordinance (SR 235.11)
- Freedom of Information Act (SR 152.3)
- Human Research Act (SR 810.30)
- Human Research Ordinance (SR 810.301)
- Swiss Penal Code (notably Art. 321bis, SR 311.0)
- SPHN Ethical Framework

EU regulations
- General Data Protection Regulation (Regulation (EU) 2016/679) (or the Directive 95/46/EC until May 24, 2018)

IT security management standards
- ISO/IEC 27001:2013 Information Technology – Security Techniques – Information security – management systems – requirements
- ISO/IEC 27002:2013 Information technology — Security techniques — Code of practice for information security controls
- NIST Cybersecurity Framework

Security framework (highlights)

SPHN Information Security Policy

Secure authentication
- Reference researcher identity provider with institutional control (e.g. Swiss EDU-ID)
- Two-factor authentication

Secure data transfers
- Encryption specifications

Secure storage
- Unix group-based access control
- Physical infrastructure access control
- Encryption of inactive data (e.g. backups)

Auditable processes and incident reporting
- Data and infrastructure access logging
- Mandatory reporting of incidents (processes)

User awareness
- Data Use Agreement
- Training

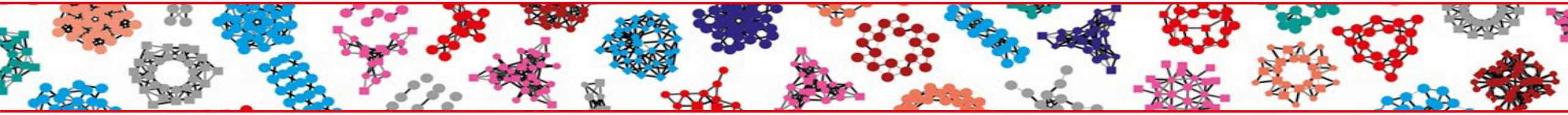
Data classification

SPHN Information Security Policy

Public: information shared and available outside of SPHN.

Internal: information that may be shared within SPHN. This information is not intended to be shared outside SPHN and respective projects/initiatives. The impact of the information leaking outside of SPHN would be minor.

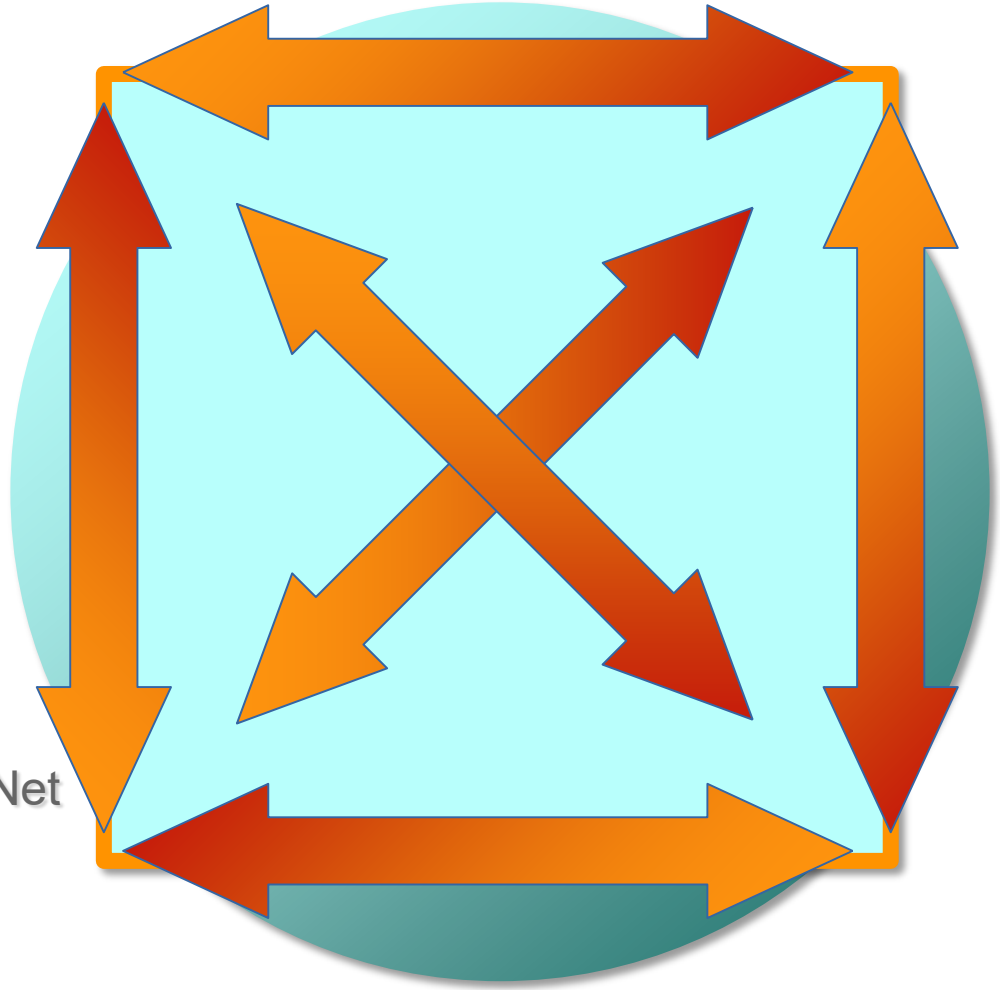
Confidential: the access to this information must be restricted only to those parties within SPHN that have a legitimate need to have access to it. **All personal data (either identifying data or pseudonymized) are confidential.** The impact of leaking this information to parties with no legitimate use or to the public may cause major harm to the person from whom the data originate, to the original Data Provider (Controller), to the research organization, or to SPHN.



IT infrastructures for biomedical research

BioMedIT infrastructures goals

Regulations
- Legal
- Ethical
- Best Practices
- CH, USA, EU



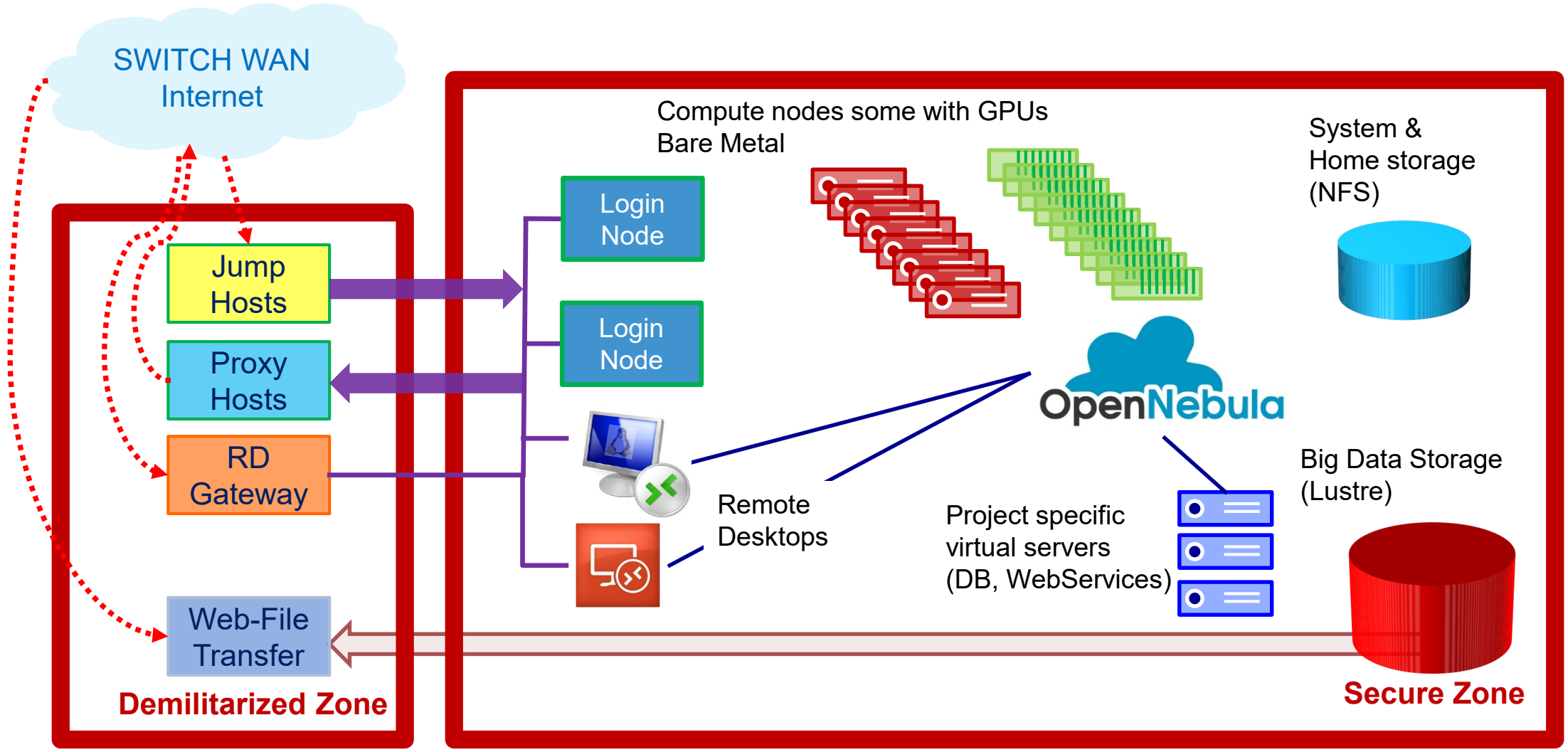
High Performance
- Fast Network
- GPUs
- Parallel Filesystems

Easy to use
- As on the notebook
- No security hassles
- Free access to the Net
- Interactive

Flexible
- Fast changes
- Cutting edge software
- State full nodes
- DB servers

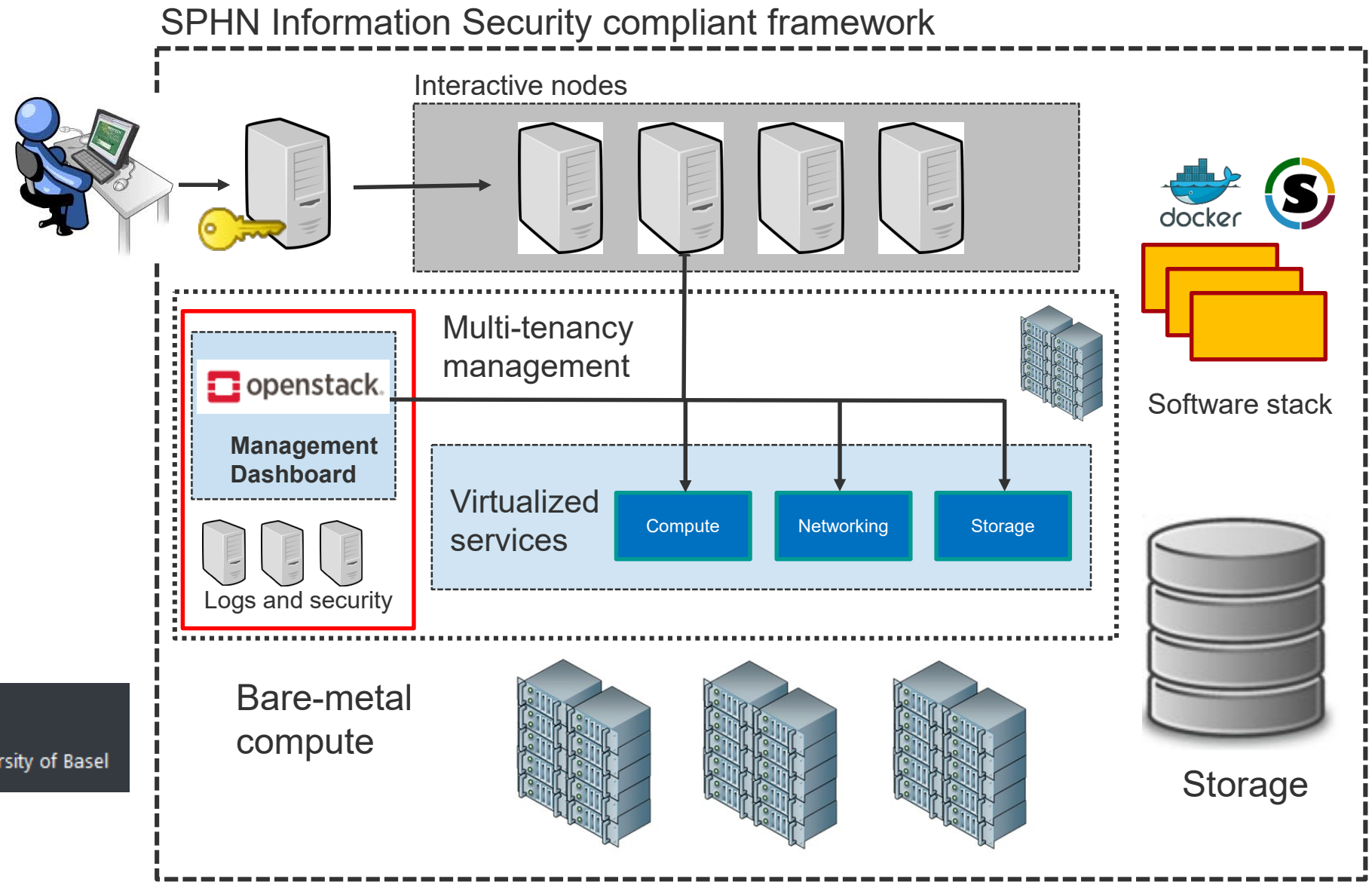
Slide credit: Christian Bolliger, SIS ETHZ

Leonhard Med > Security + Usability



Slide credit: Christian Bolliger, SIS ETHZ

sciCORE Med - BioMedIT architecture



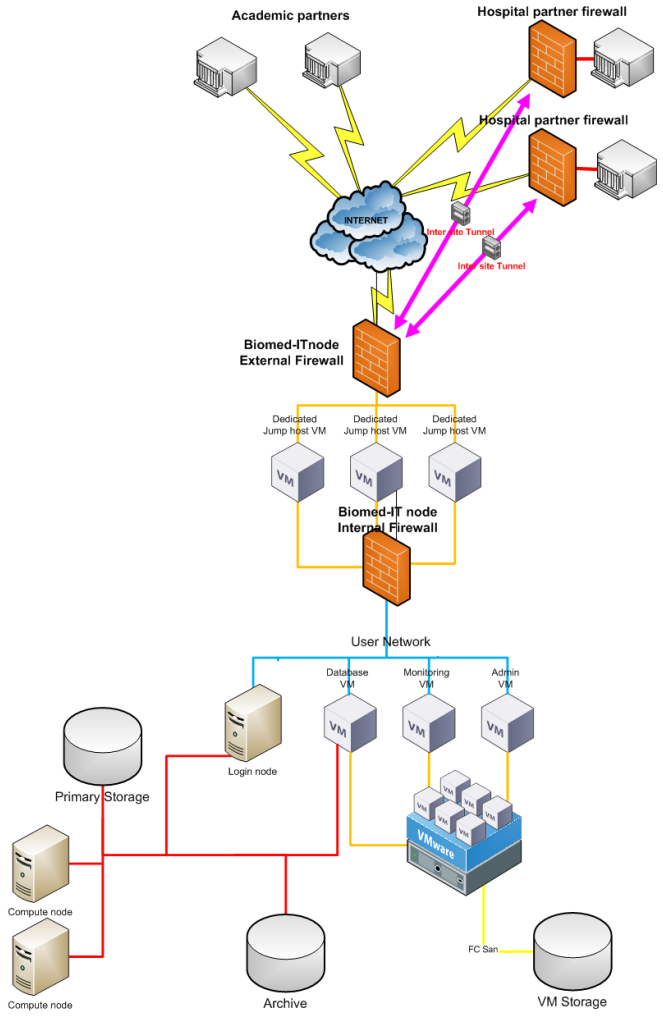
sciCORE
Center for scientific computing @ University of Basel

University of Basel

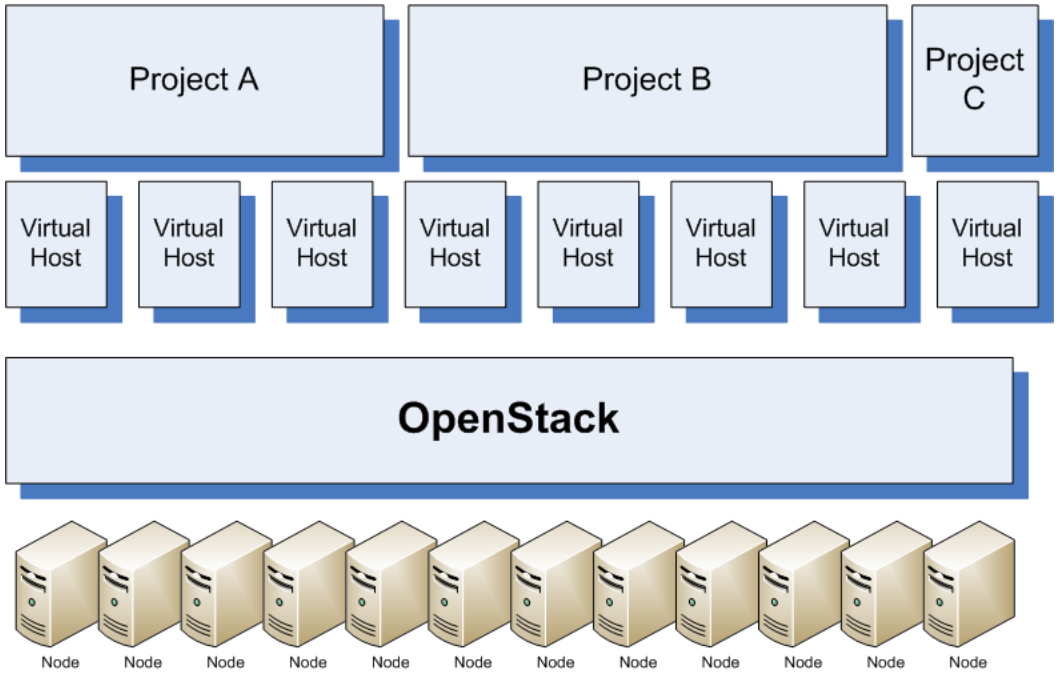
Core-IT BioMedIT Architecture

SIB Lausanne

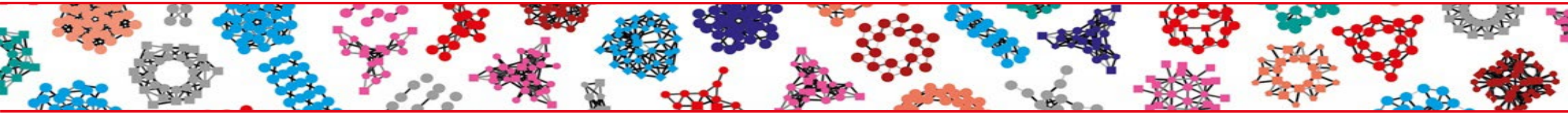
Physical layout



Logical layout



Slide credit: Roberto Fabbretti, Vital-IT

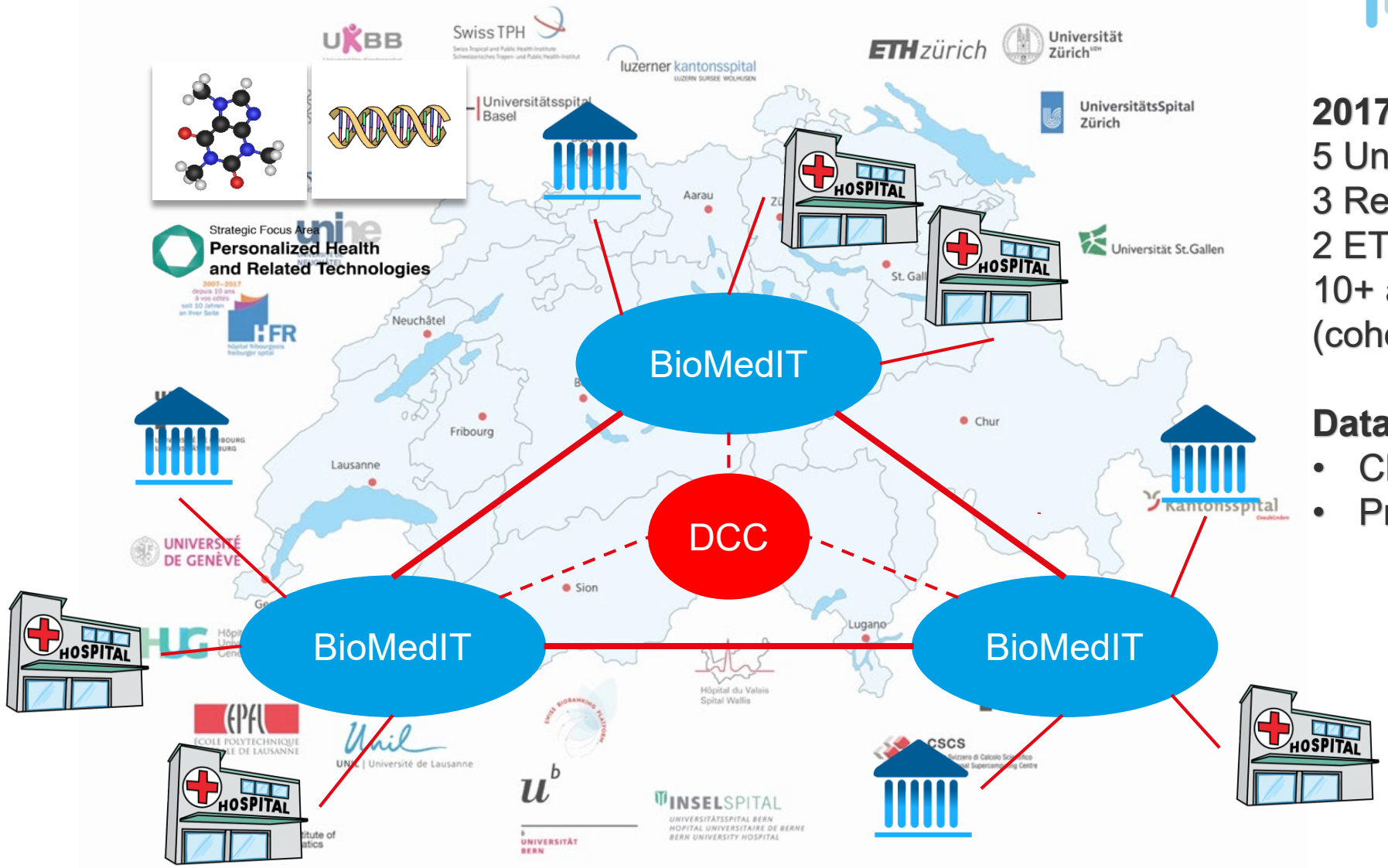


Interoperability

Network interoperability challenge

Swiss Personalized Health Network

Swiss Institute of Bioinformatics



2017-2020

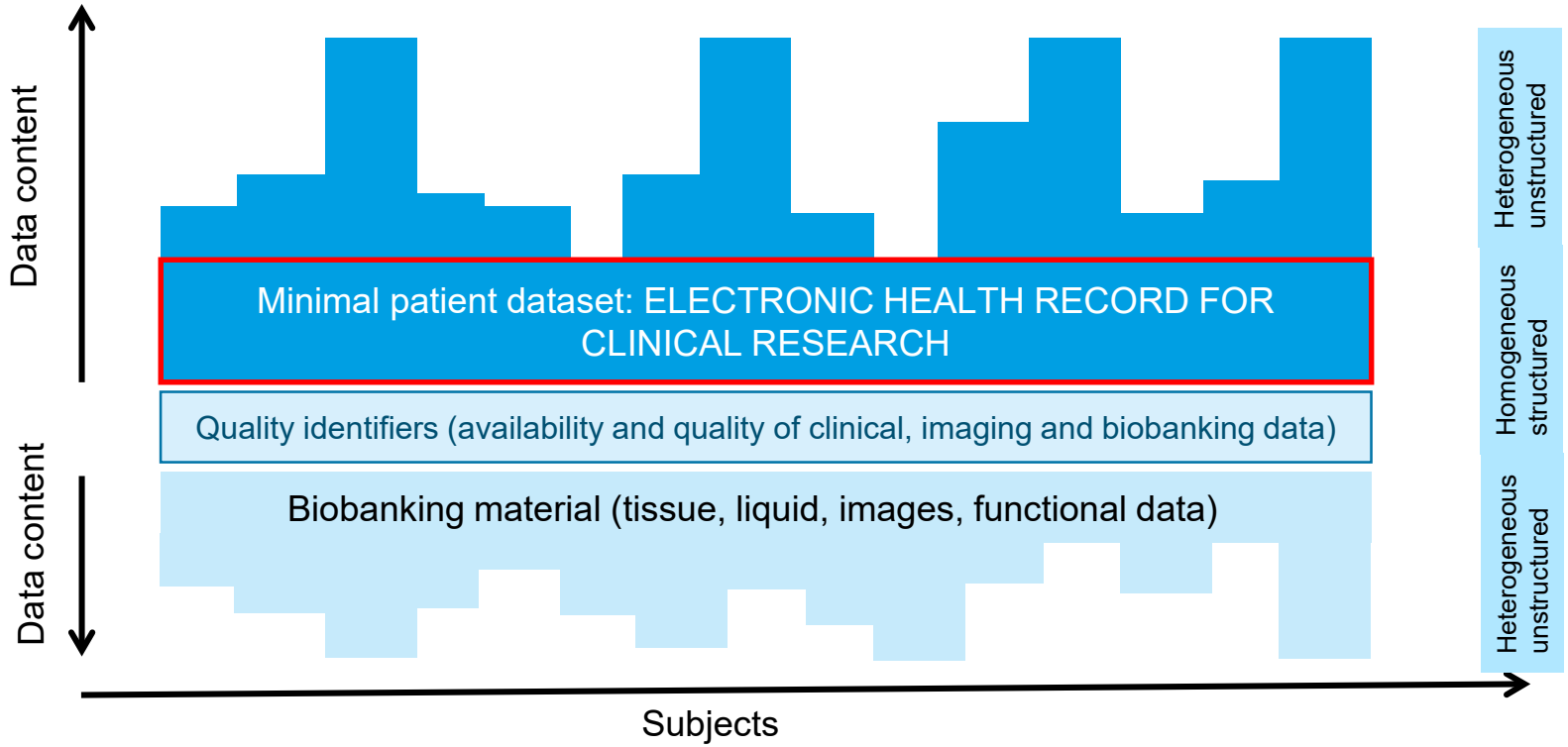
5 University hospitals
3 Research IT infrastructures
2 ETH platforms
10+ additional partners
(cohorts, regional hospitals, …)

Data transfers

- Channel encryption
- Project-specific encryption

Semantic challenge: Hospital data heterogeneity

Type of data: (low degree of data quality, high numbers, large heterogeneity)



Data semantics

Core terms defined through SPHN Driver projects

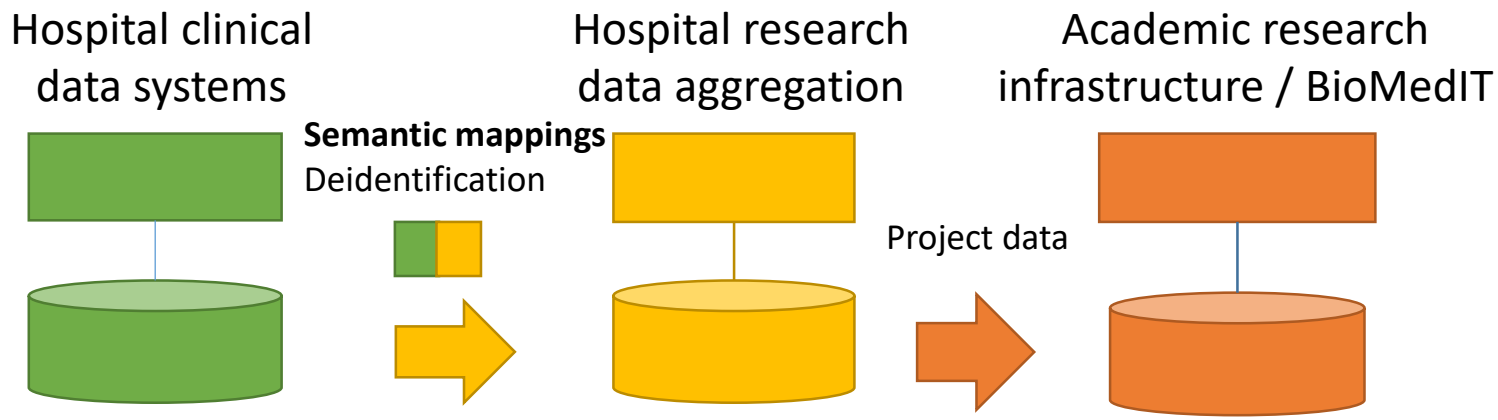
Call 2017

- Oncology
- Sepsis (intensive care)
- Immune repertoire
- Ageing

Swiss Personalized Health Network

SPHN semantic interoperability WG coordinates:

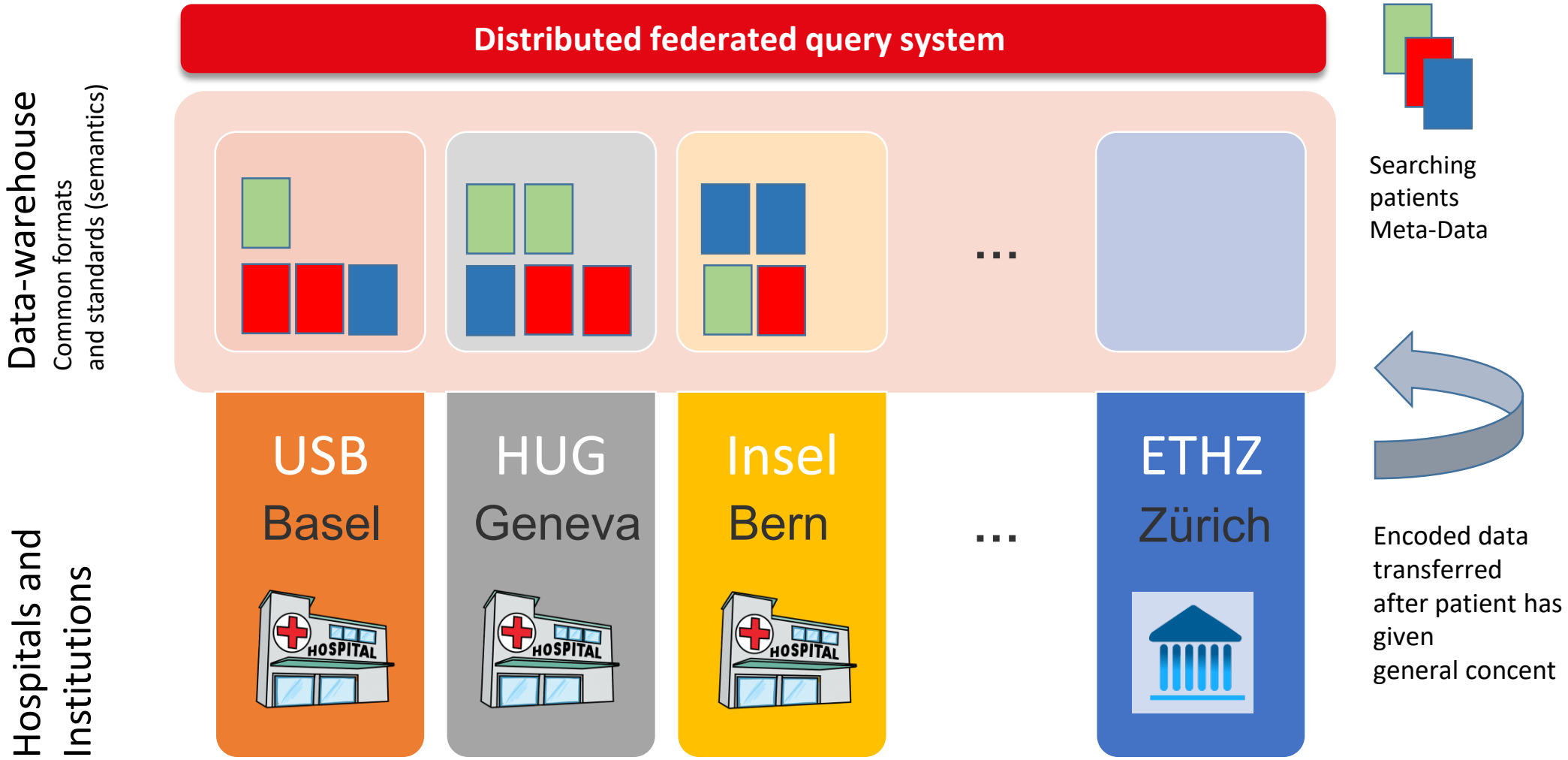
- Across hospitals
- Across domains



Core vocabularies

- SNOMED CT
- ICD-10, ICD-O
- LOINC

Hospitals data discoverability



Workflow interoperability

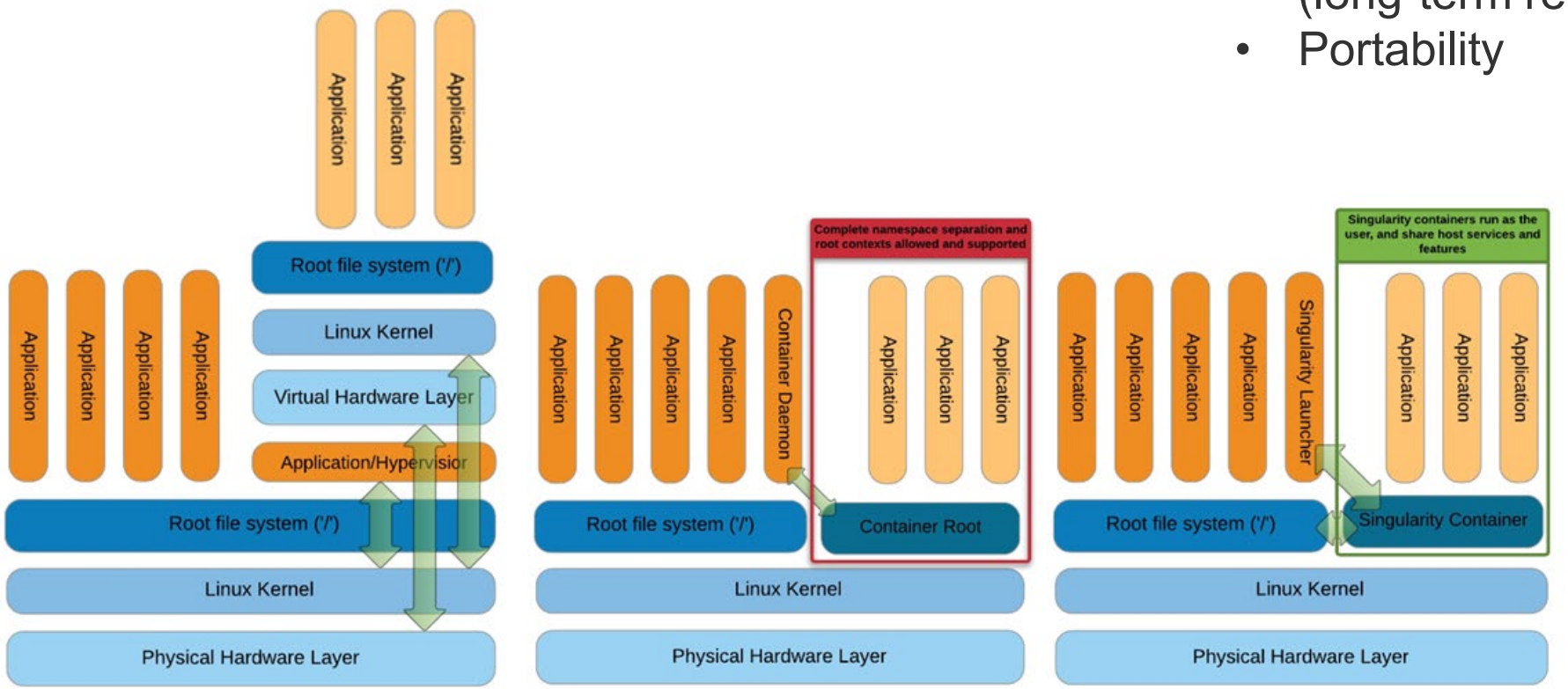
- Supporting reproducible research
  - Repeatable data analysis (same data)
  - Reusable in different context (different data)
- Workflows executable everywhere
  - Work with local data

Slide credit: Kevin Sayers, SPHN DCC

Virtualization technologies

Key for

- Dependencies encapsulation (long-term reproducibility)
- Portability



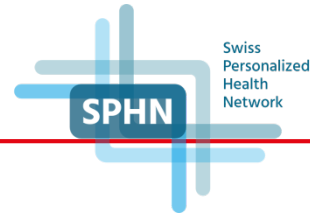
General VM
eg ESXi

General Container
eg Docker

HPC Container
Singularity

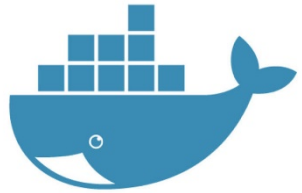
Source: Greg Kurtzer, Singularity Containers for Science

Container types



Swiss Institute of Bioinformatics

Docker



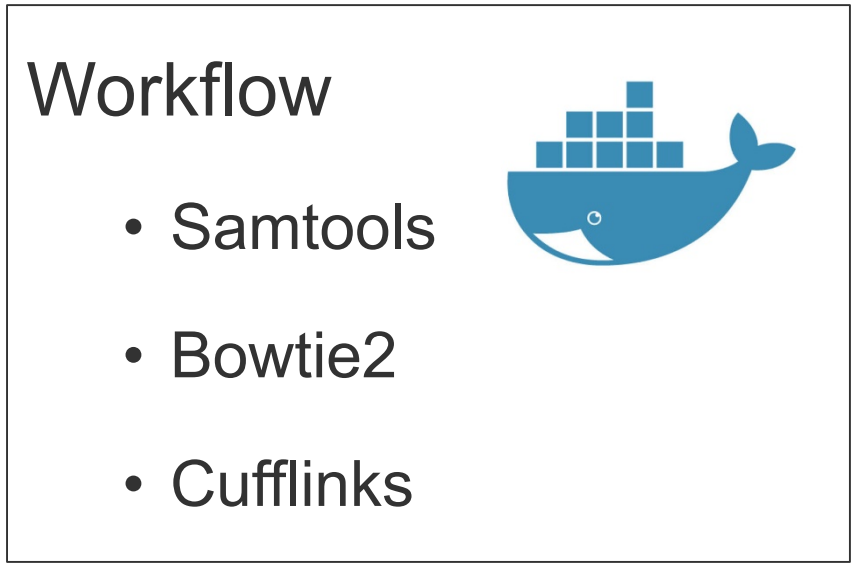
- Widely used
- Security issues (commonly unsupported in HPC environments)
- Federation and orchestration tools (Kubernetes, Swarm)
- Docker Hub repository
- Built-in support from cloud providers

Singularity

- More secure to run than Docker in multi-user environment
- Out of the box support for GPU's and other hardware
- Can pull Docker images and convert to Singularity
- Singularity Hub repository
- Convenient default data mounts

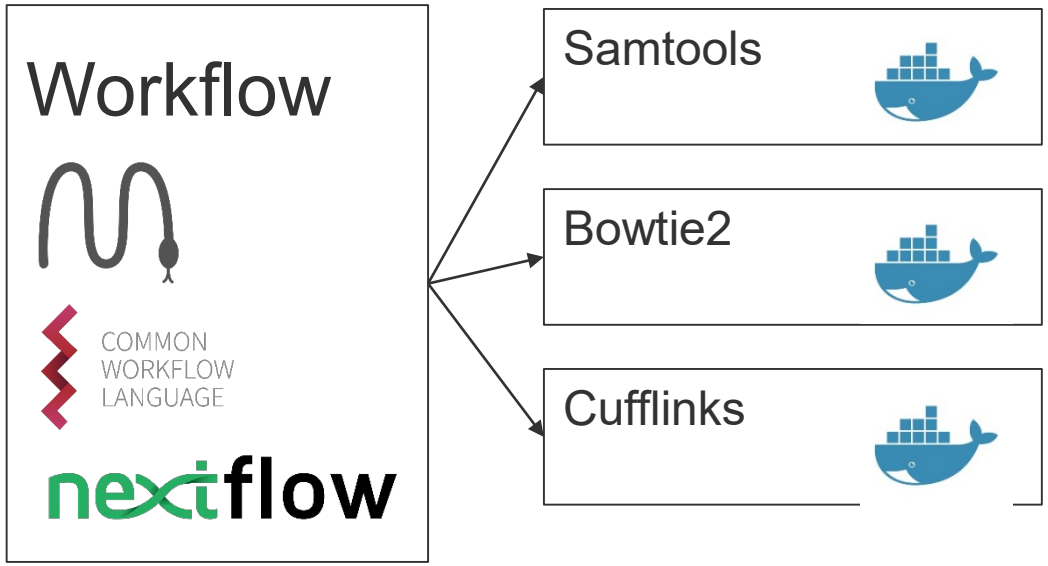
Slide credit: Kevin Sayers, SPHN DCC

Workflow strategies



One container with everything

- Pros:
  - One image
- Cons:
  - Potential incompatible dependencies
  - Large image size
  - Changing individual components tricky



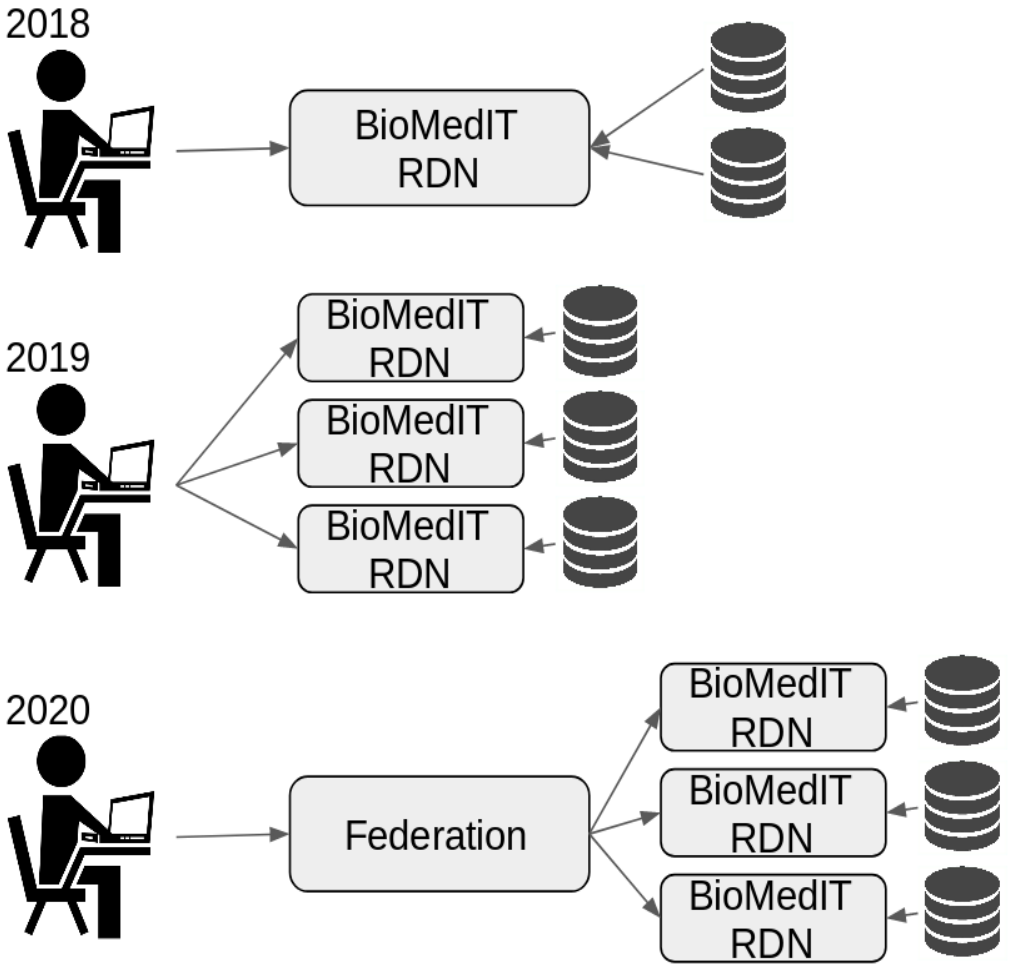
External workflow that orchestrate containers

- Pros:
  - Containers can be mixed and matched
- Cons:
  - Workflow system must be supported
  - Reproducibility concerns

Slide credit: Kevin Sayers, SPHN DCC

Enabling edge computing

"Bringing the algorithm to the data"



Securely transfer data to a BioMedIT node for analysis

Provide single sign on (SSO) to access BioMedIT nodes where data is located

Federation that abstracts the underlying BioMedIT infrastructure

Slide credit: Kevin Sayers, SPHN DCC

BioMedIT Project status

Regulations
- SPHN Ethical framework (2017)
- SPHN Information Security Policy (2018)

IT infrastructure
- Infrastructure implementation ongoing on different sites (Zürich,
- Support to SPHN/PHRT projects ongoing (preparatory phase)
- Identity federation in development (Swiss EduID)

Workflow interoperability
- Proof-of-concept workflow interoperability completed (genomics)
- Workflow infrastructure (e.g. container repositories) and standards in development
- Workflow federation in development

Data interoperability
- Proof-of-concept node-to-node data transfers completed
- Technologies to link hospitals tested
- Key infrastructure, registries and automation in development
- Data catalogs, secondary data reuse in early development

BioMedIT environment usable for pilots…
… on track for production in 2020

Collaborations

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

SPHN : a driver project for GA4GH (February 2019)

Elixir collaborations on selected technology developments

SDSC

Outlook into 2021-2024

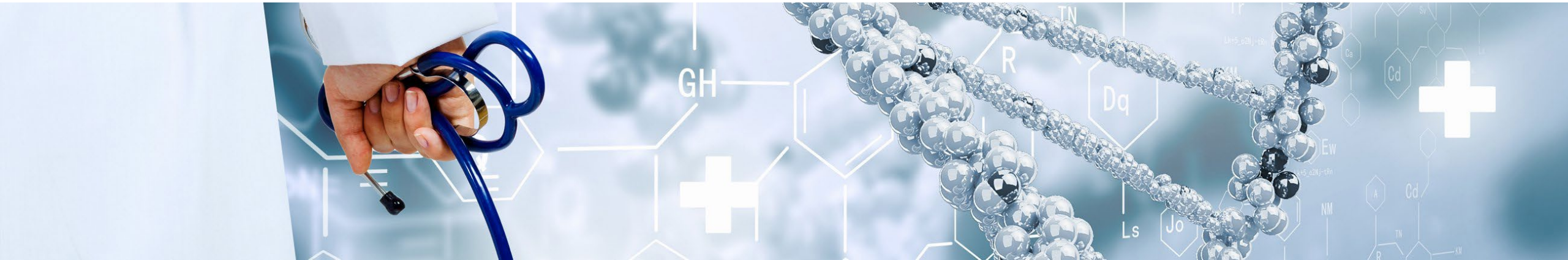
- Extend network to regional hospitals and private clinics
- Expand range of data collected (hospitals and platforms), consolidate semantics
- Integrate healthy citizen data
- Strengthen role of ELSIag and foster engagement of patients and citizens
- Establish PPPs (public-private partnerships) and international consortia
- Establish concepts for sustainable national center for PH research post 2024 (governance and technology)

More information

SPHN / BioMedIT

https://www.sphn.ch/
https://dcc.sib.swiss/

Documents
SPHN Glossary
SPHN Ethical Framework
SPHN Information Security Policy
SPHN Data Use Agreement Template



Acknowledgements

Key input for this presentation
Kevin Sayers (SIB, DCC)
Diana Coman Schmid (ETH)
Christian Bollinger (ETH)

SPHN/BioMedIT DCC
Katrin Crameri (chair), Torsten Schwede
Leila Alexander
Martin Fox
Sabine Österle
Kevin Sayers

BioMedIT interoperability team
Kevin Sayers (DCC)
Sergio Maffioletti (Uni Zürich)
Jaroslaw Surkont (Uni Basel)
Jorge Molina (Vital-IT)
Diana Coman Schmid (ETH)

BioMedIT nodes and board members
Bernd Rinn, Thomas Wüst, Leonhard Med team (ETH)
Heinz Stockinger, Roberto Fabbretti, Core-IT/Vital-IT teams (SIB)
Sofia Georgakopoulou (Uni Basel)

Christine Durinx (SIB)
Patrick Ruch, Douglas Teodoro (HES-SO)
Marcel Riedi, Sergio Maffioletti (Uni Zürich)

BioMedIT network collaborators
University hospitals, represented by:
- Christian Lovis (HUG, also chair semantics)
- Nicolas Rosat (CHUV)
- Alexander Leichtle (Insel)
- Markus Obreiter (USB)
- Roland Naef, Cornelia Kruschel-Weber (USZ)

All PIs and collaborators of SPHN projects

SPHN Swiss Personalized Health Network
SIB Swiss Institute of Bioinformatics
SAMW ASSM
Staatssekretariat für Bildung, Forschung und Innovation SBFI

Questions?

Supporting slides

Technical interoperability

“Cloud-of-services” federation

- Middleware-infrastructure agnostic (OpenStack, OpenNebula)
- Harmonization at the level of services

Project management (registry)

Identity and access control

Data access and data management

Workflow execution

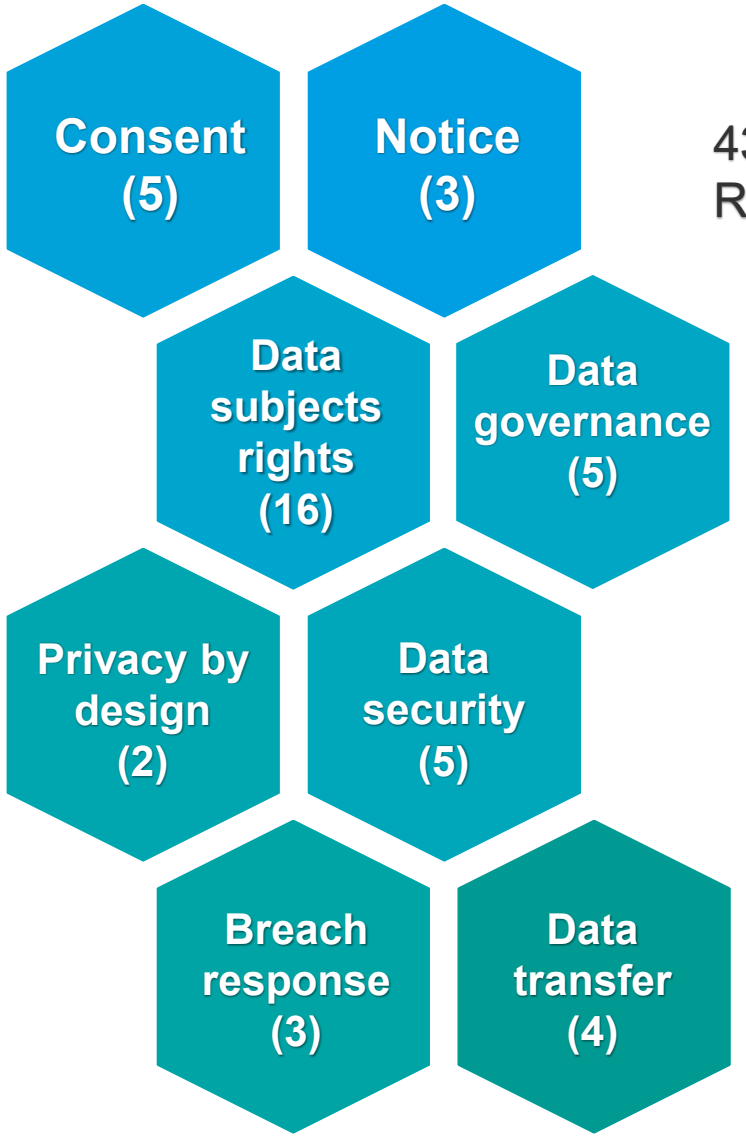
GDPR overview

European Council
Council of the European Union

The **General Data Protection Regulation (GDPR)** imposes new rules on organizations that offer goods and services to people in the European Union (EU), or that collect and analyze data tied to EU residents, no matter where they are located.

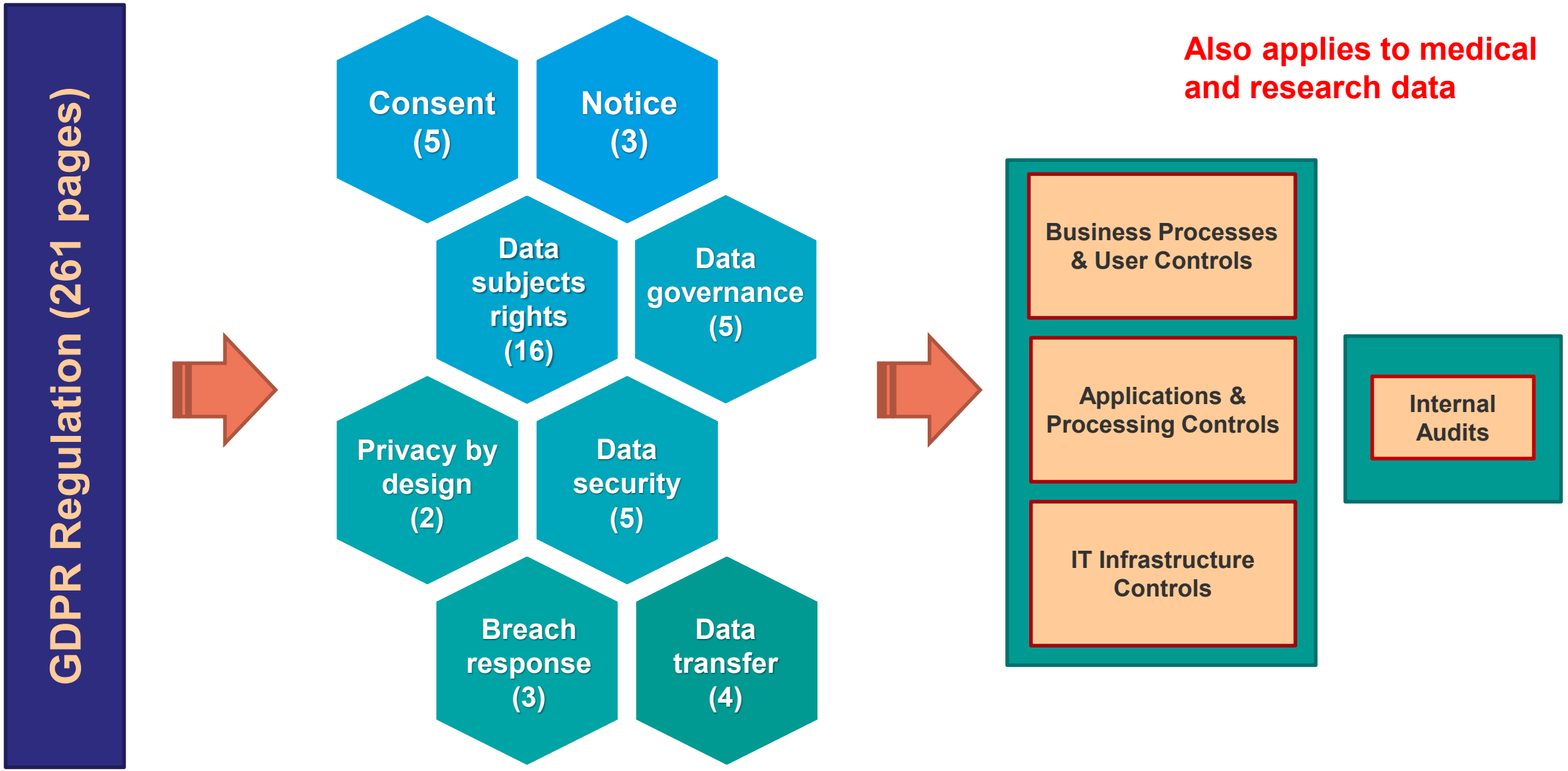
- **Enhanced** personal privacy rights
- **Increased** duty for protecting data
- **Mandatory** breach reporting
- **Significant** penalties for non-compliance

GDPR Regulation (261 pages)

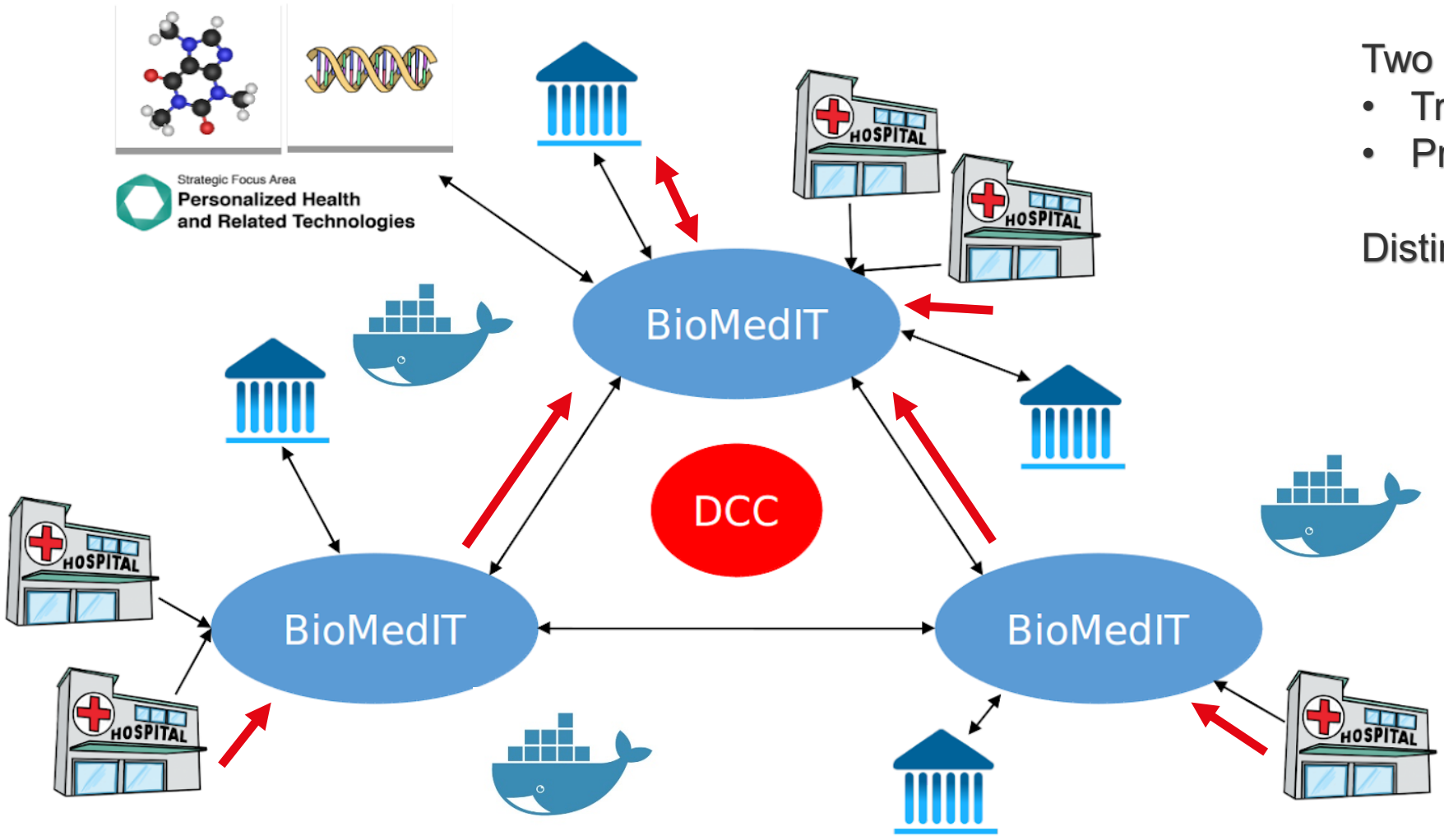


43 GDPR Requirements

GDPR overview



BioMedIT data transfer protocol

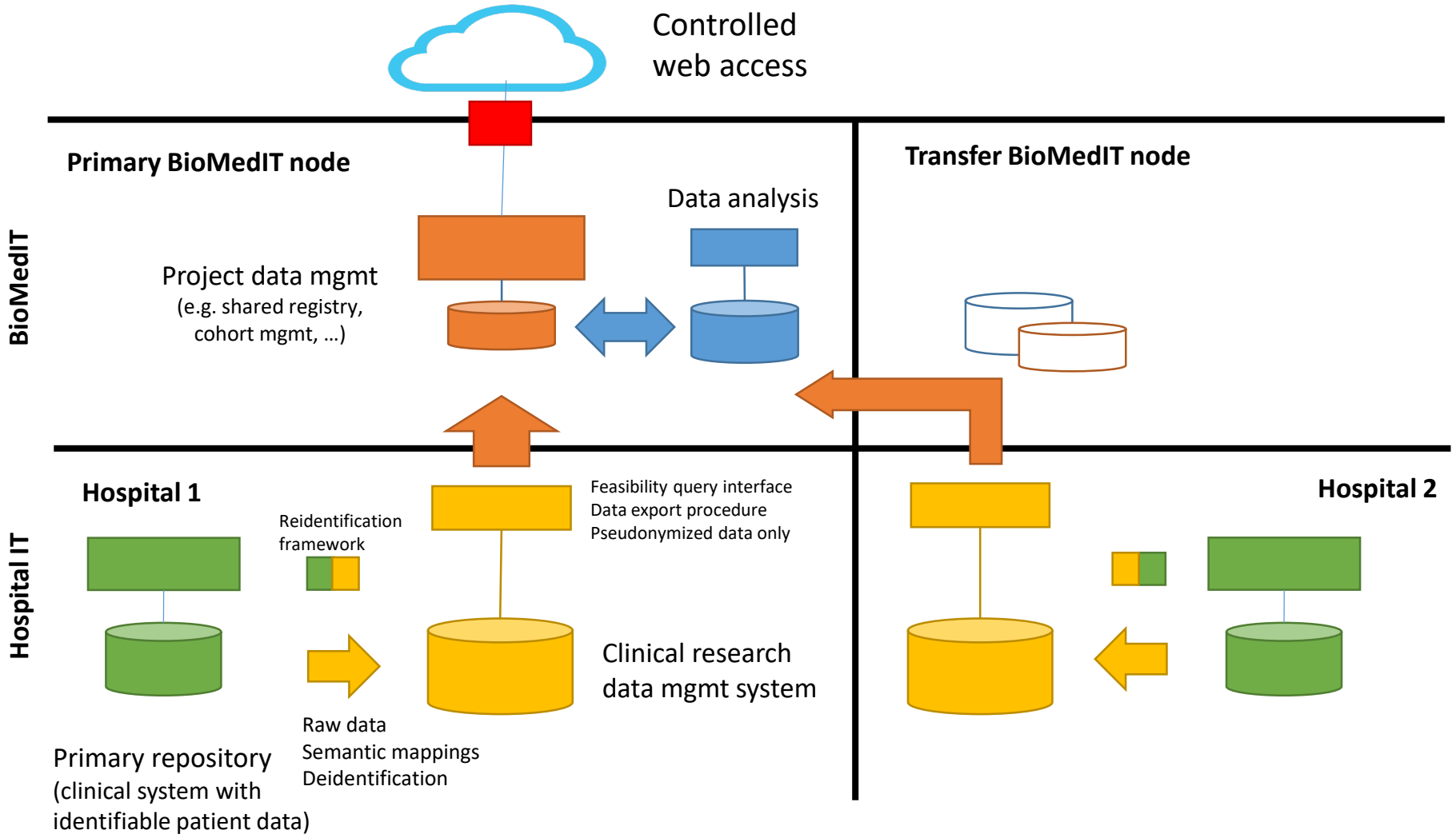


Two encryption layers:

- Transport encryption (SFTP)
- Project-level data encryption

Distinct asymmetric key pairs

SPHN/BioMedIT data flows



Generic description of data flows in SPHN

Container issues

Swiss Personalized Health Network

Swiss Institute of Bioinformatics

TECHNOLOGY

Hackers mined $90,000 worth of Monero with a simple Docker Hub trick

- Even if there is no mali
  functions as ex

container image data damned

Kromtech finds malicious code hiding in enterprise upstart's repositories of software

By Thomas Claburn in San Francisco 14 Jun 2018 at 07:01 3 SHARE

Slide credit: Kevin Sayers, SPHN DCC

Addressing the concerns

- Host a private SPHN repository
  - Local Gitlab can be used to build the images
- Utilize tools that scan repositories for image vulnerabilities
- Manual review of Dockerfile or Singularity files
- Container signing from developers

Slide credit: Kevin Sayers, SPHN DCC

EnhanceR Singularity tests

Swiss Personalized Health Network

Swiss Institute of Bioinformatics

EnhanceR | Enhancing Research through IT Expertise

EnhanceR is a Swissuniversities project on science IT coordination
ETHZ, Uni Basel, Uni Zürich, Uni Bern, SIB, EPFL, FMI, HES-SO

- Representative bioinformatics workflow in a Singularity container
- Demonstrated functionality in BioMedIT nodes
- Demonstrated analysis reproducibility
- White paper available

Collaboration: ETHZ, Uni Basel, Uni Zürich, Uni Bern, SIB

EnhanceR Singularity test matrix

Swiss Personalized Health Network

Swiss Institute of Bioinformatics

Milestone	UniBe	sciCORE UniBas	S3IT UZH	Vital-IT / SIB	SIS ETHZ
Singularity installed	✅ (v2.3)	✅	✅ v2.3-dist	✅ (v2.2.1 and v2.3)	✅ (v2.3)
"hello world" pipeline building	✅	✅	✅	✅ (from Dockerfile)	✅ (from Dockerfile)
"hello world" pipeline running	✅	✅	✅	✅ (from Singularity image)	✅
sciCORE pipeline1 & pipeline2 building	✅	✅	✅	✅	✅ (from Dockerfile)
sciCORE pipeline1 & pipeline2 running	✅	✅	-	✅ (only pipeline 2, pipeline1 is too big to run in home directory (quota issue))	✅
Docker -> Singularity conversion via Dockerfile	✅ (with"hello world" pipeline)	✅	✅	✅ (with SIB-publications pipeline)	✅ (with"hello world" pipeline)
Docker -> Singularity conversion via Dockerhub	✅ (with "whalesay" container)	✅ (with custom Dockerfiles for pipeline1 & pipeline2)	✅ poldracklab/fmriprep	not planned	✅ (with "whalesay" container and custom Dockerfiles)
Docker -> Singularity conversion via "private" Gitlab	✅ (with "hello world" pipeline, no auth)	✅	-	not planned	✅ (from private GitLab)
Runner script	✅ *(pipeline.py)*	✅ *(pipeline.py* from UniBe)	-	✅ (own runner script, tested with SIB-publications pipeline)	✅ *(pipeline.py* from UniBe)
Guideline document	✅	✅	✅	✅ all details are in: README.md and in final report	✅