Forschungsdatenökosystem am Universitätsspital Zürich

Vortrag im Rahmen des MIRACUM-DIFUTURE Kolloquiums

Matthias Joos & Patrick Hirschi, Team Research Data Products, Universitätsspital Zürich

"Der Biber ordnet die Daten, die Schmetterlinge verteilen sie wild, die Eule guckt groß und der Rest summt vor sich hin – so entsteht aus dem Chaos der Forschungsdaten ein harmonischer Datendschungel."

Deepseek R1, 2025

Matthias Joos

✉ **matthias.joos@usz.ch**

in **matthias-joos-099a51260**
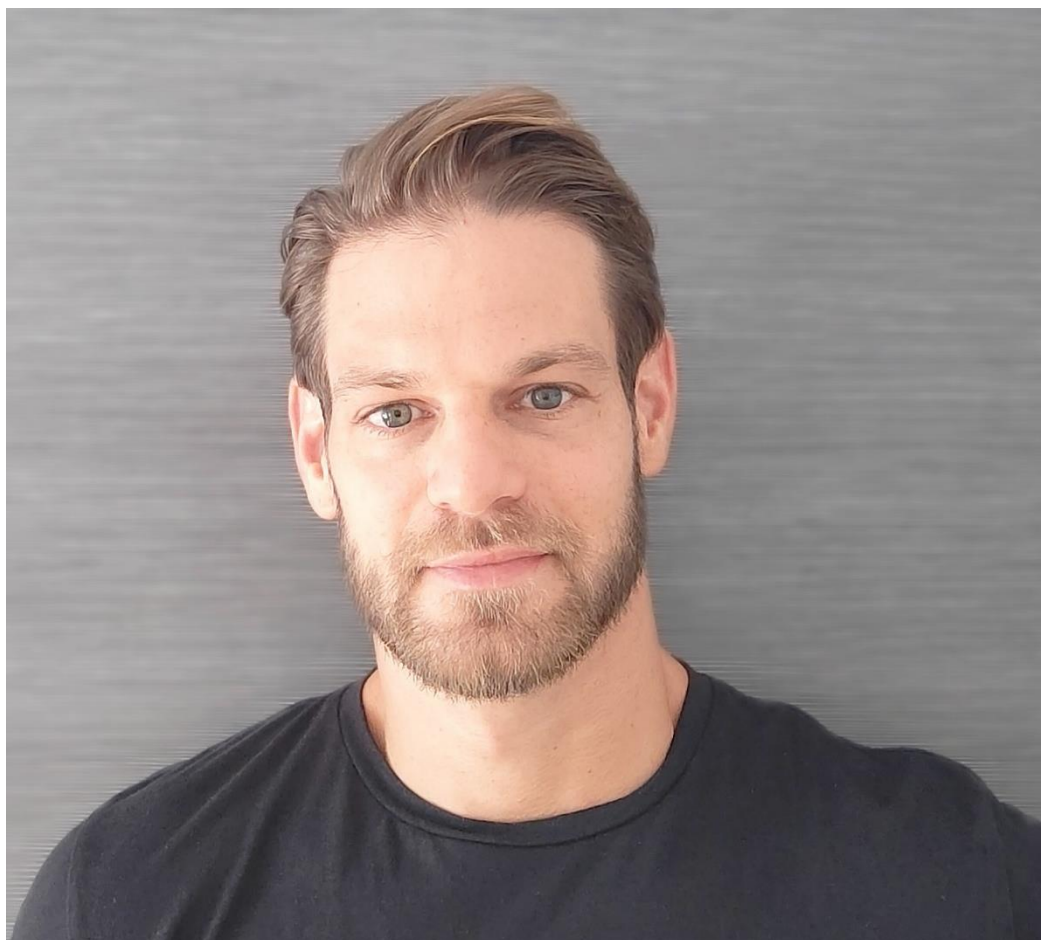
since 2022

Data Solution Engineer
Research Data Products

▶ *Development of USZ data products (data deliveries, data marts, ETL pipelines)*
▶ *Development of SPHN data pipelines*
▶ *Development of data platform tools*

2018 - 2022

Business Service Verantwortlicher
Medizinbereich Innere Medizin - Onkologie

▶ *Responsible for IT project portfolio of medical department*

Patrick Hirschi

mail@patrick-hirschi.ch

@patrickhirschi_

https://patrick-hirschi.ch

patrick.hirschi.5

since 2016

USZ
Universitäts
Spital Zürich

2015 - 2016

emineo

2020 **Lead Engineer**
Clinical Data Platform for Research

▶ *Team Lead for 10 data engineers and 3-5 associated virtual employees*
▶ *Responsibility and decision-making competence regarding methodology, technology, and architecture of the Clinical Data Platform*
▶ *Responsible for the technical implementation of interfaces from USZ to the Swiss Personalized Health Network (SPHN) and The LOOP Zurich Biomedical Informatics Platform (BMIP) to allow for translational data-driven research across swiss hospitals*

2018 **Data Engineer**
Clinical Data Platform for Research

▶ *Architecture, development, maintenance and administration of the Clinical Data Platform and its underlying data lake, databases, and data marts*

2016 **Application Manager / Developer**
Clinical Data Management

▶ *Data solution and product development on the Clinical Data Platform*

2015 **Healthcare Consultant**
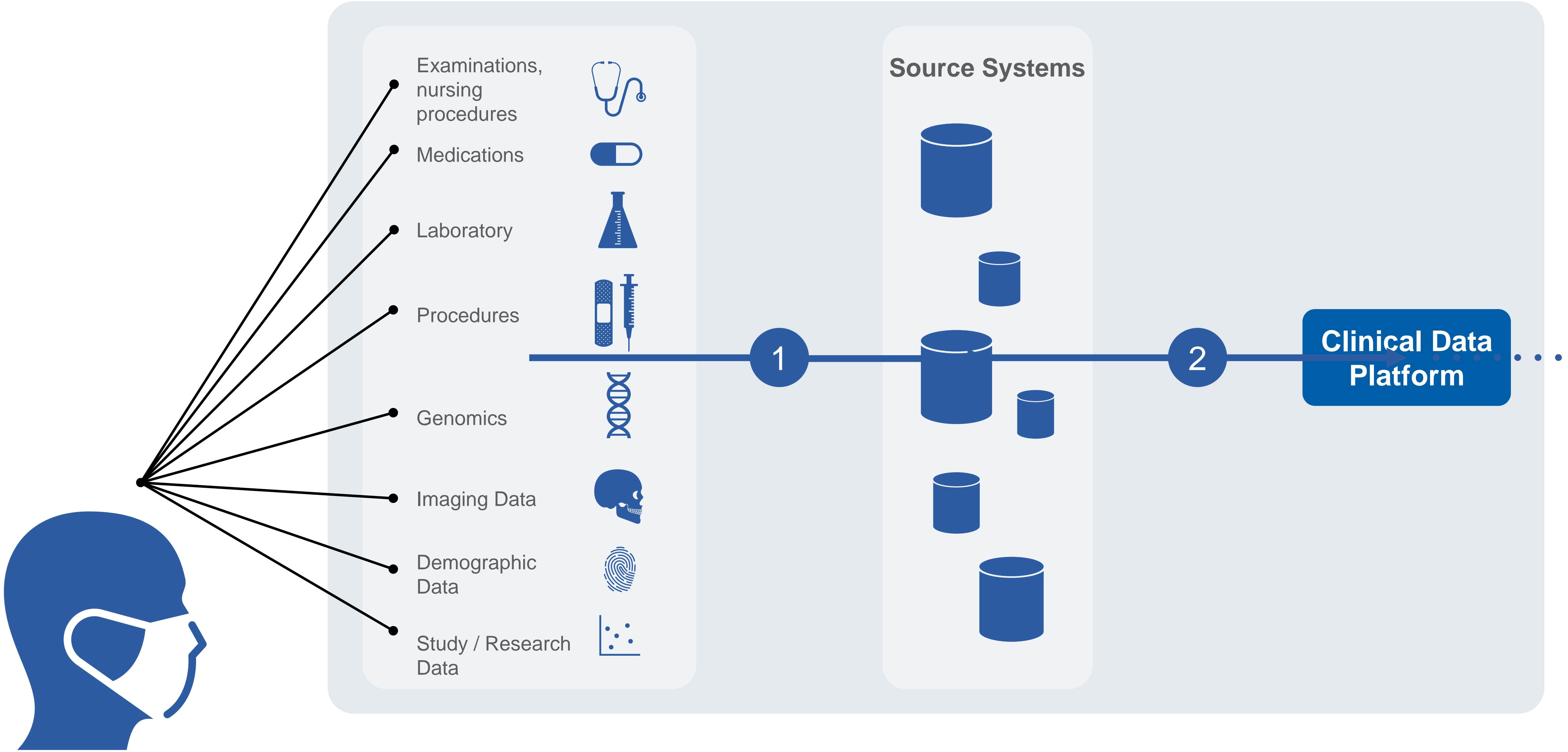*emineo AG (Switzerland, Zurich)*

▶ *Project Management and Consulting of Hospitals*
▶ *Analysis of clinical source systems and interfaces*
▶ *Technical implementation of interfaces, reports, applications / solutions for various clinics (traumatology, infectiology, emergency department)*
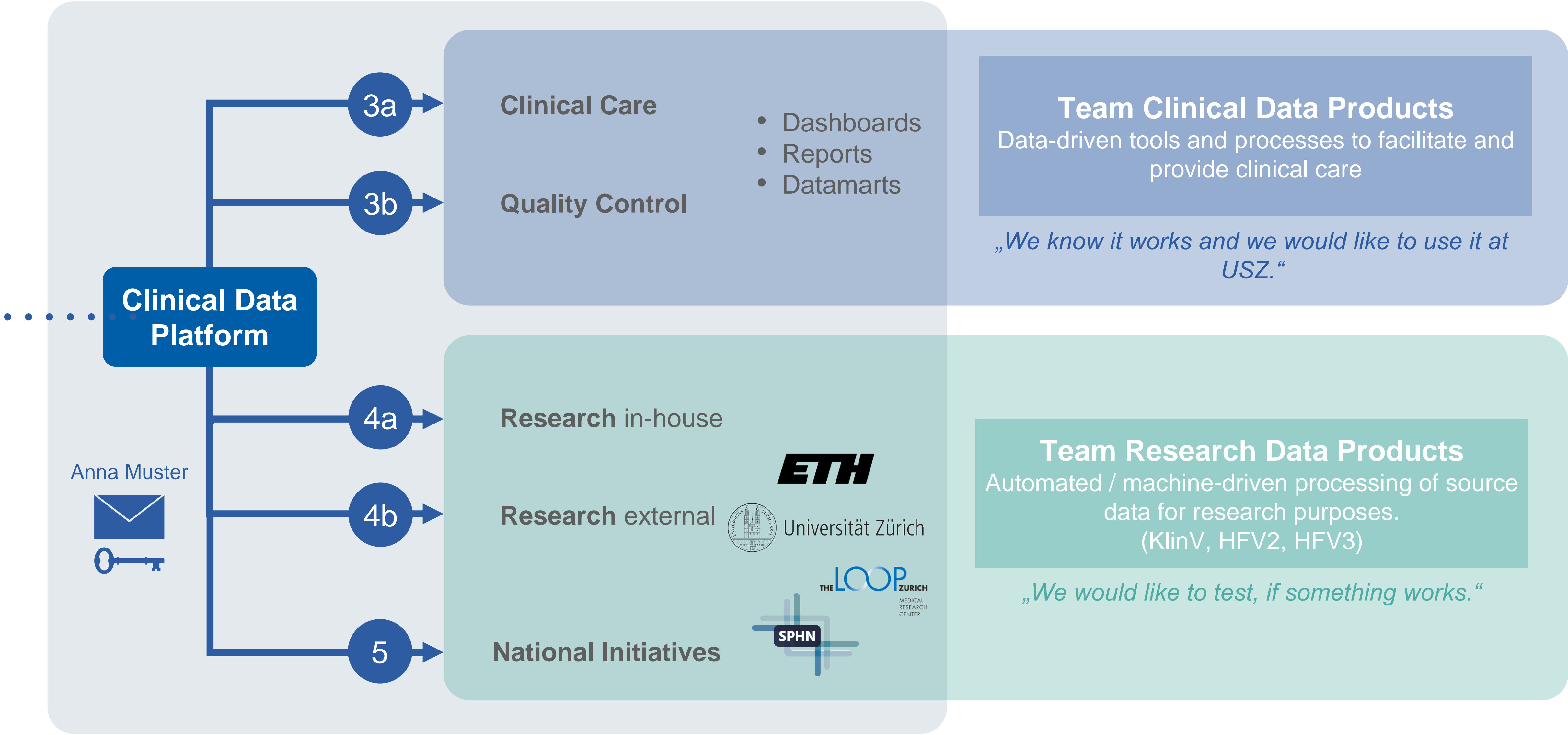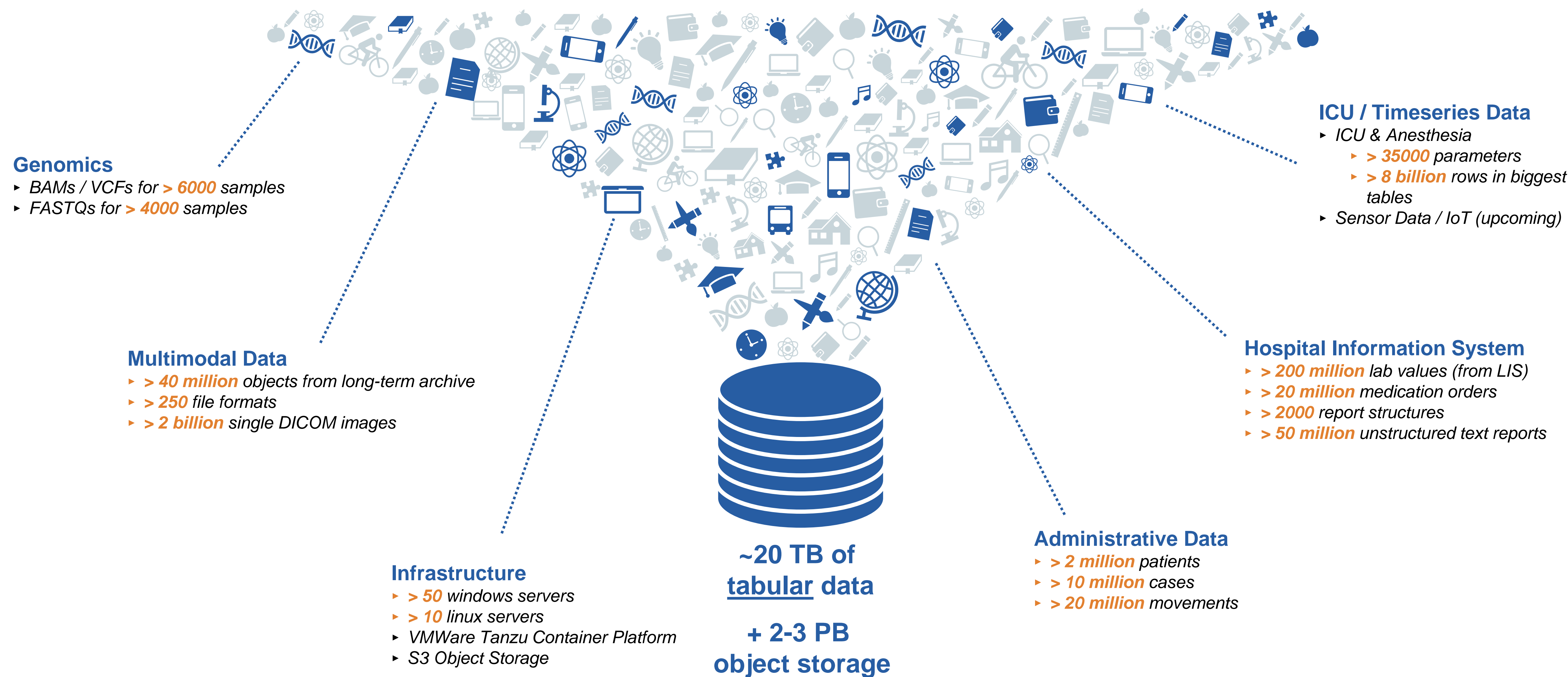
Clinical Data Platform

USZ Universitäts Spital Zürich

Genomics
- *BAMs / VCFs for > 6000 samples*
- *FASTQs for > 4000 samples*

ICU / Timeseries Data
- *ICU & Anesthesia*
  - *> 35000 parameters*
  - *> 8 billion rows in biggest tables*
- *Sensor Data / IoT (upcoming)*

Multimodal Data
- *> 40 million objects from long-term archive*
- *> 250 file formats*
- *> 2 billion single DICOM images*

Hospital Information System
- *> 200 million lab values (from LIS)*
- *> 20 million medication orders*
- *> 2000 report structures*
- *> 50 million unstructured text reports*

~20 TB of tabular data

+ 2-3 PB object storage

Infrastructure
- *> 50 windows servers*
- *> 10 linux servers*
- *VMWare Tanzu Container Platform*
- *S3 Object Storage*

Administrative Data
- *> 2 million patients*
- *> 10 million cases*
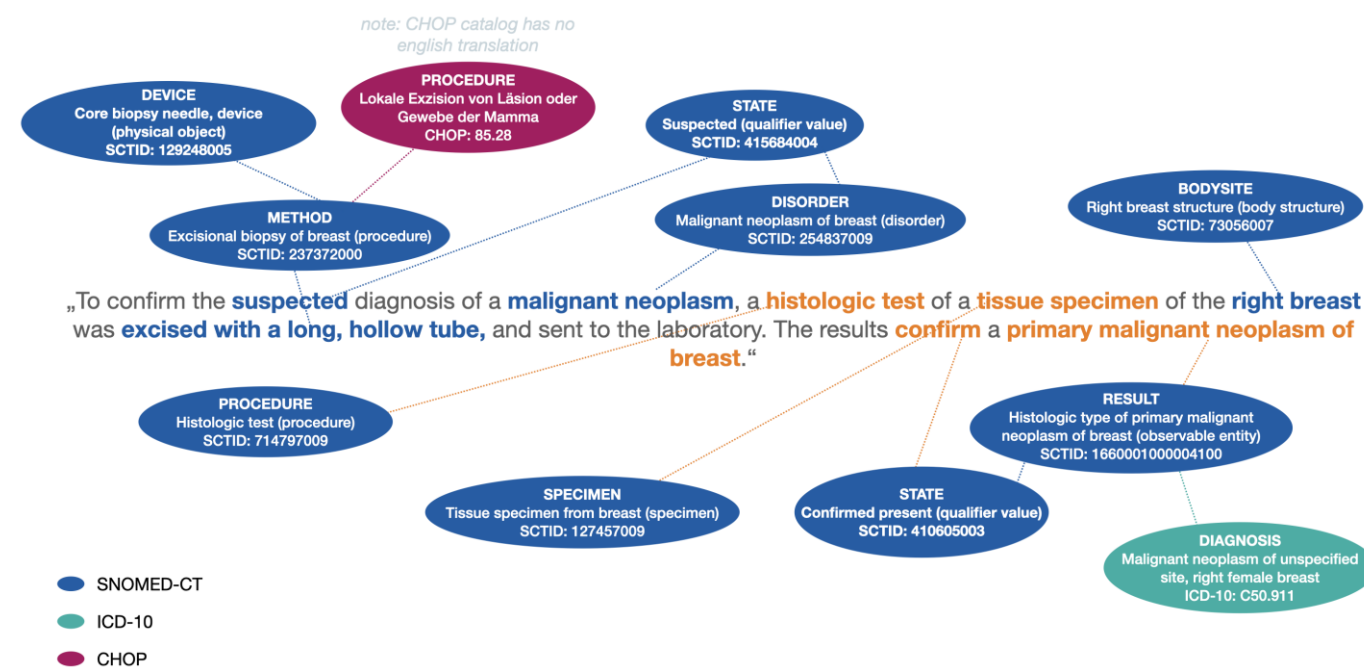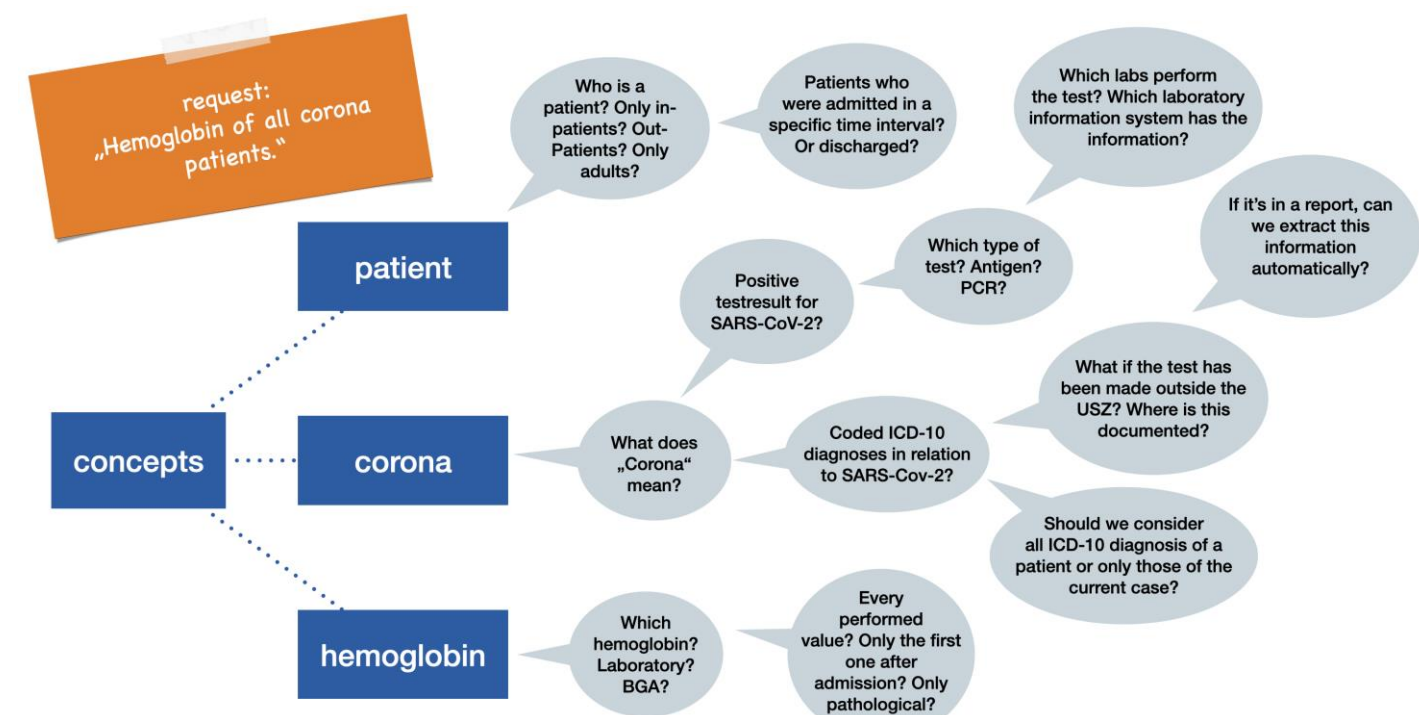- *> 20 million movements*

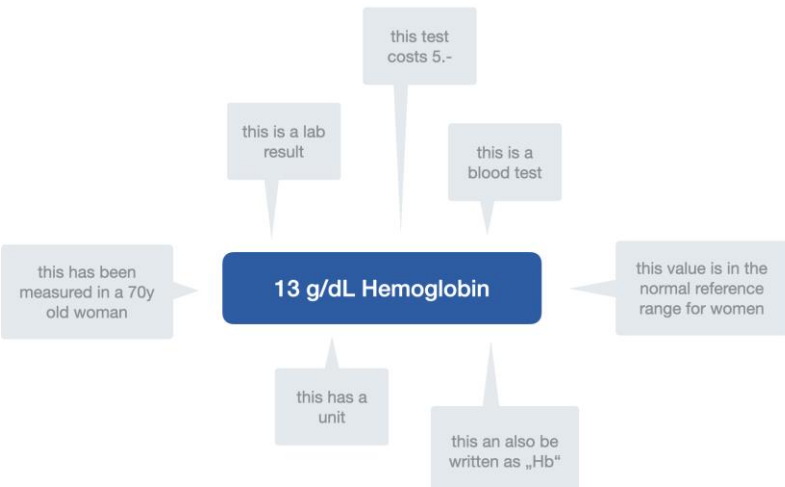**all numbers from Mar 2024**

Research Data Products

Data Specification
What is needed to answer the research question?

Data Interoperability, Standards & Quality
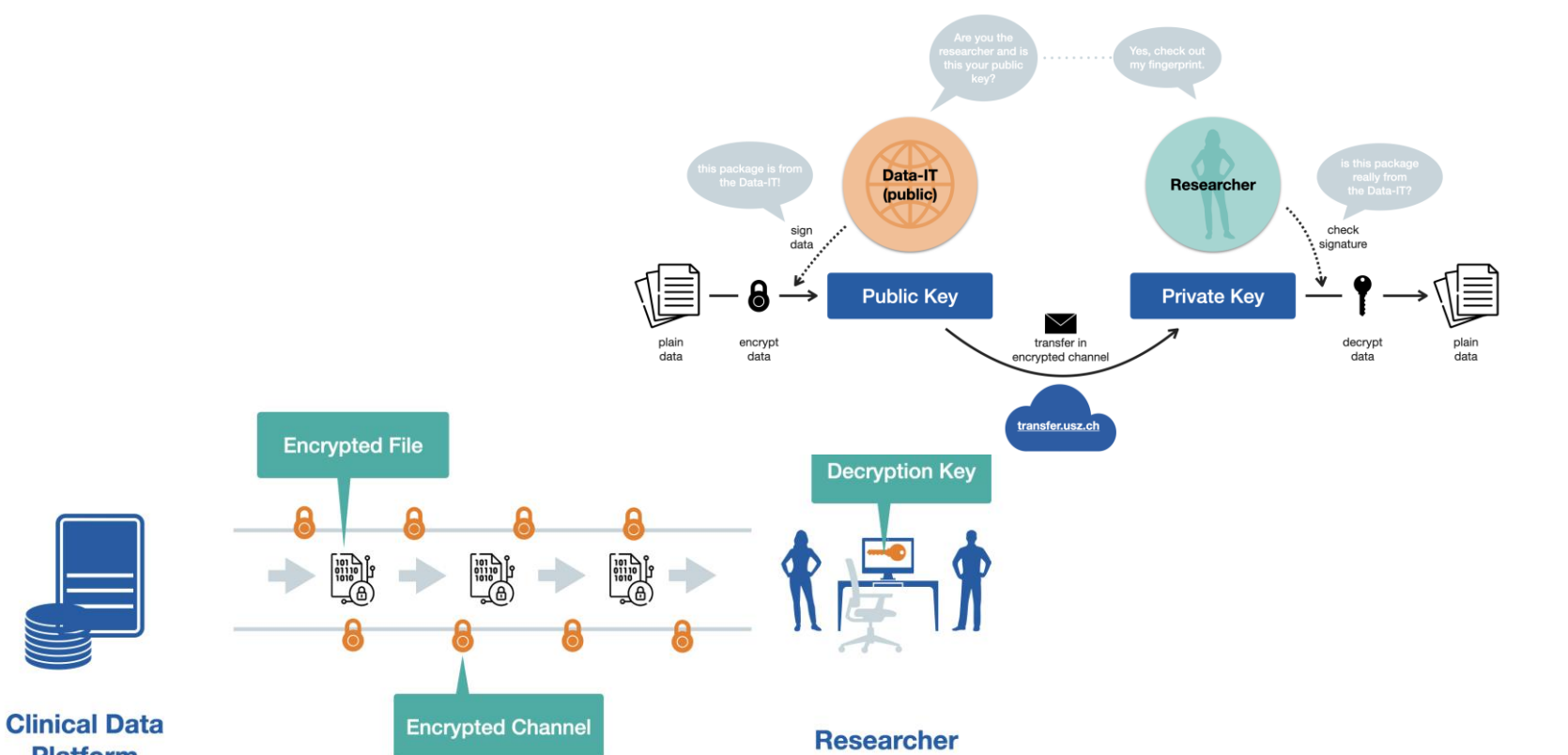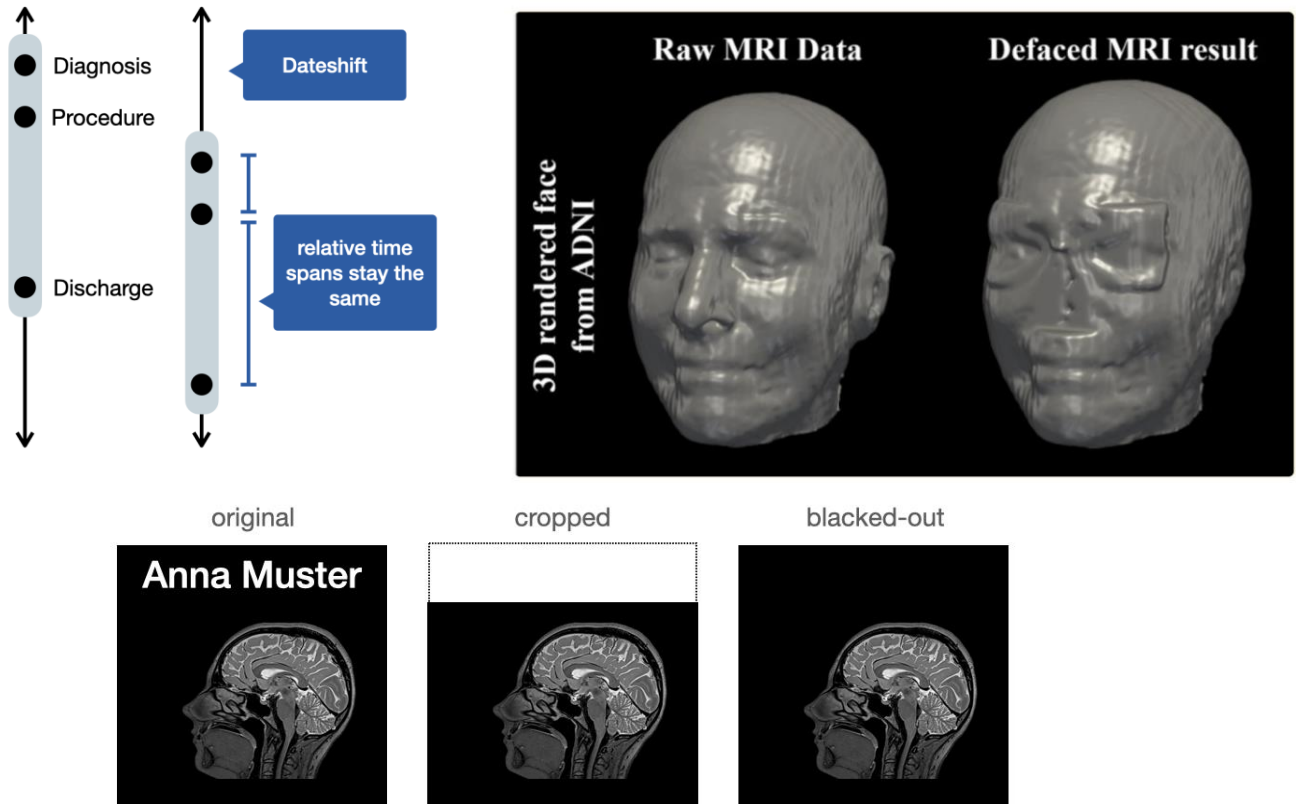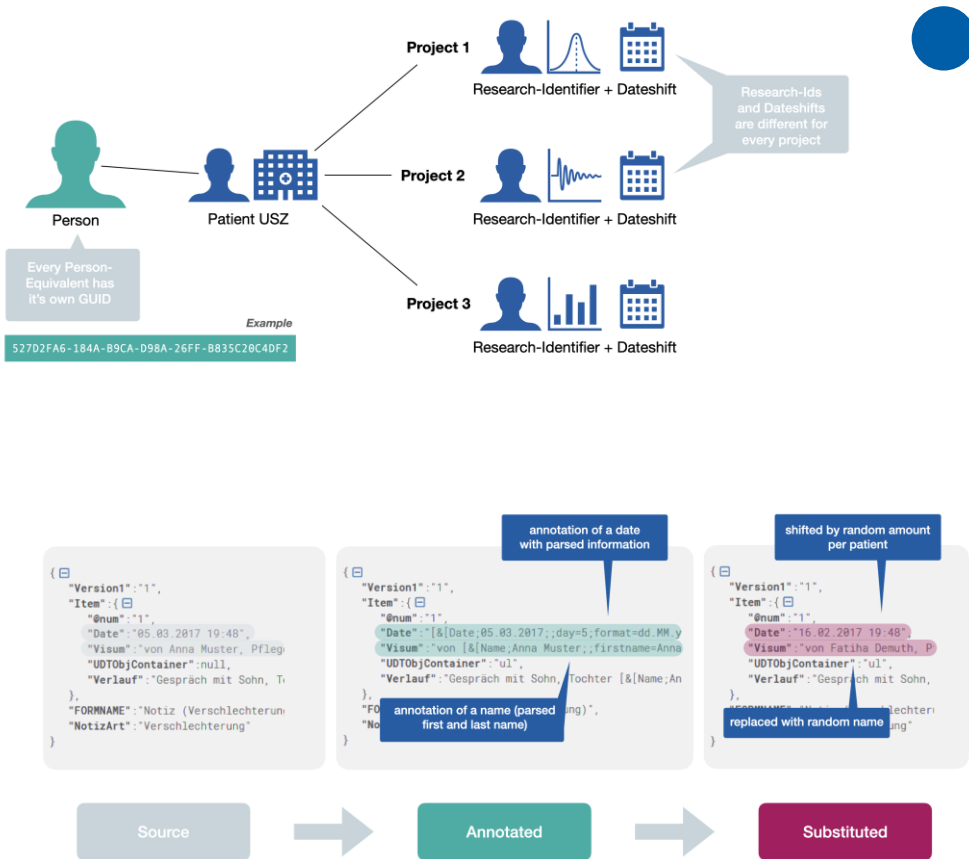Terminologies, Ontologies, Standards

Structured & Unstructured Data (Text, Images, etc.)
Data Deidentification

Encrypted File Transfer
Data Provisioning

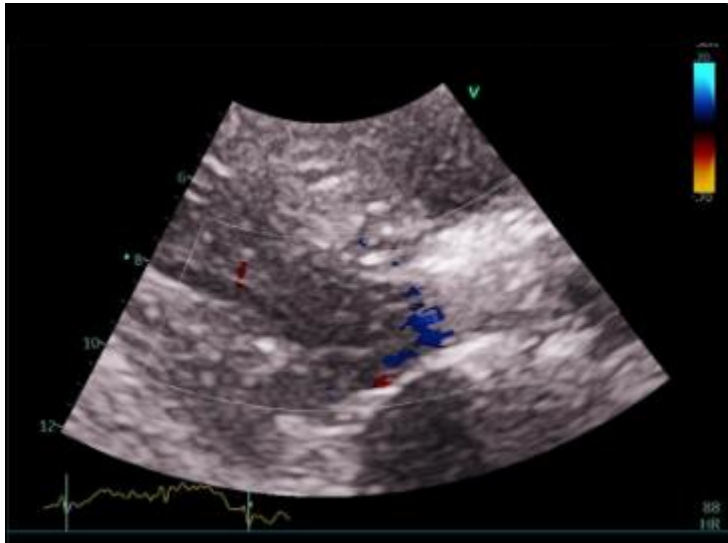## Self-service webportal for clinicians & researchers

DICOM DeId – Image de-identification for research studies
Device-specific removal of DICOM metadata and "burnt-in" annotations

ONTOCAT – Terminology Mapping
Intuitive user interface for facilitated mapping of clinical source data to international terminology

DocStore
Platform "library": Guidelines, manuals and further documentation

Impact of SPHN

Program Overview

Mandate   Research infrastructure initiative with the aim of making data from routine care (as well as human, biological data) better usable for research

Mandatary   State Secretariat for Education, Research and Innovation (SERI)

Funds   ERI Dispatch (SAMS and SIB); CHF 135 million (2017-2024)

Focus   Standardization of health data (semantics, coding, format, etc.) and infrastructure development for responsible and efficient use of (sensitive) data for research purposes

Scope   Primary and secondary use of data (also outside of research) with a view to the Swiss health data space and a digital public service
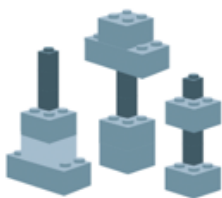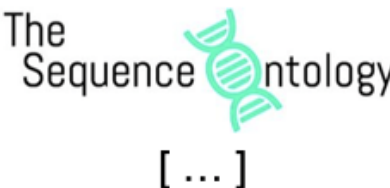
→ Enable data providing institutions to responsibly share interoperable health data
→ Enable researchers to access, integrate, analyze and re-use FAIR health data

Source: sphn.ch

Source: sphn.ch

Source: sphn.ch

USZ Universitäts Spital Zürich
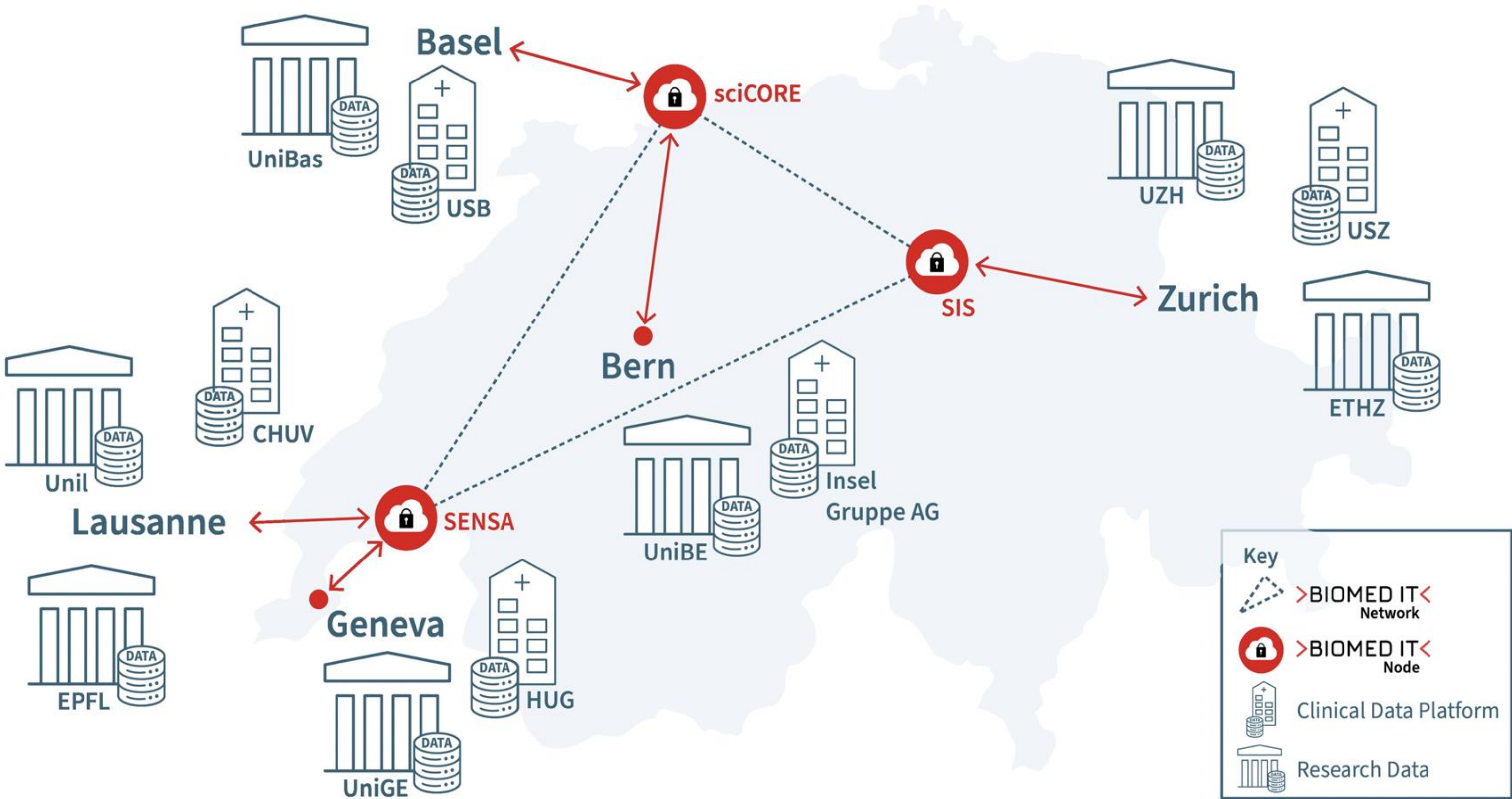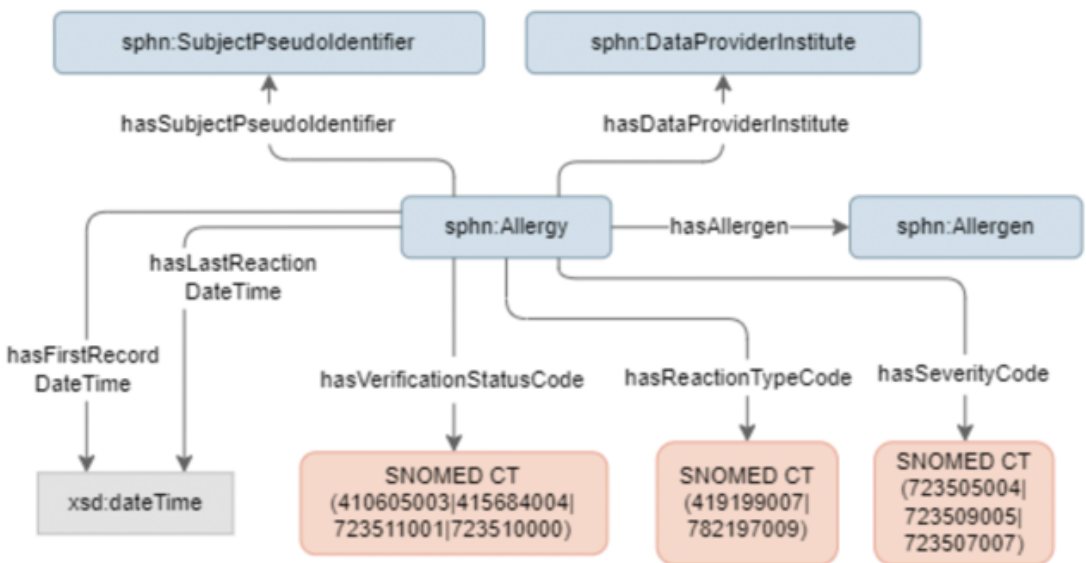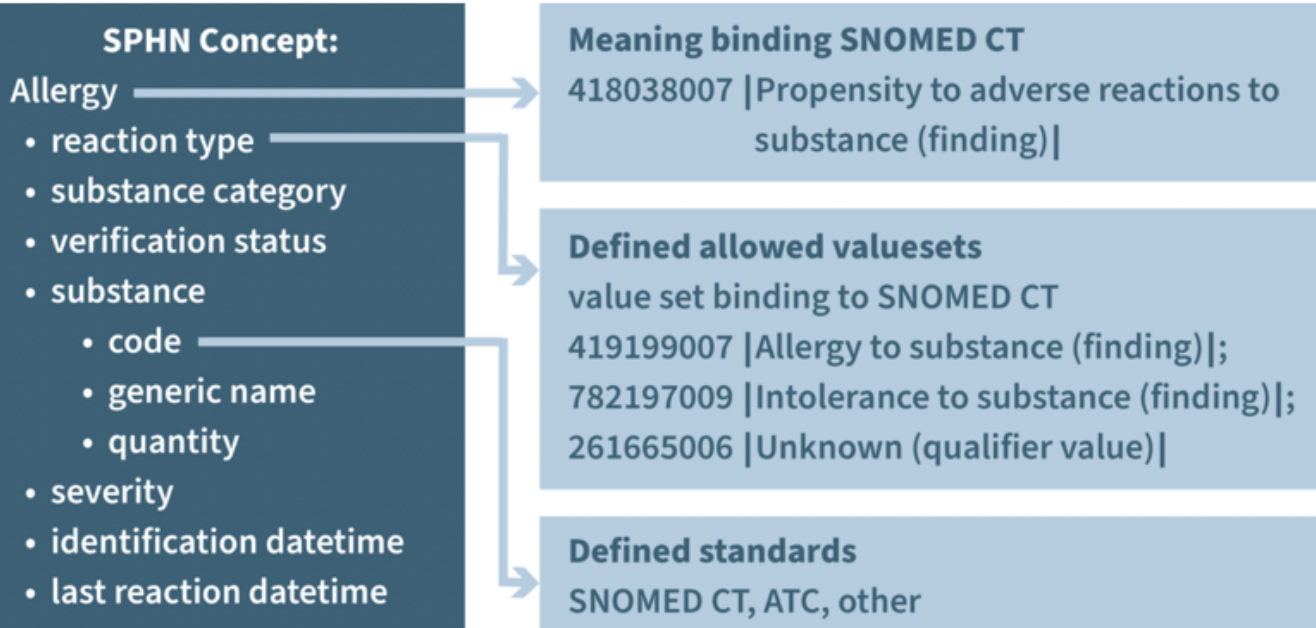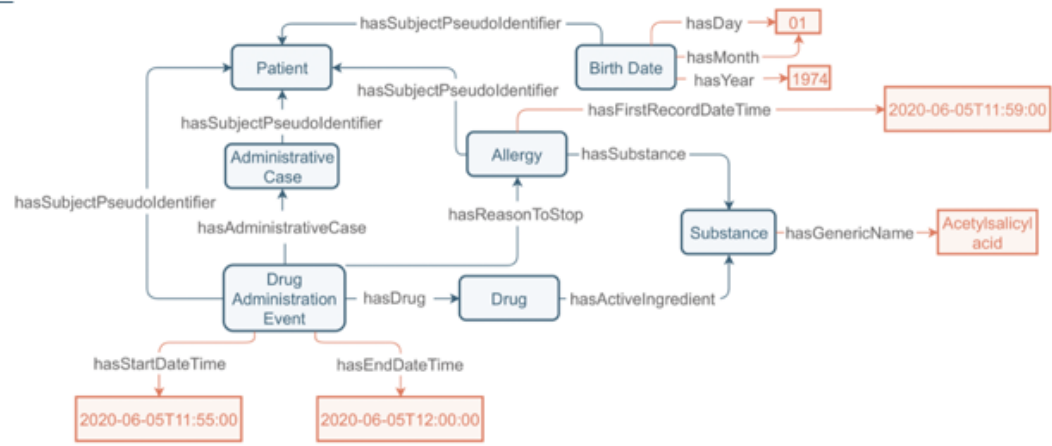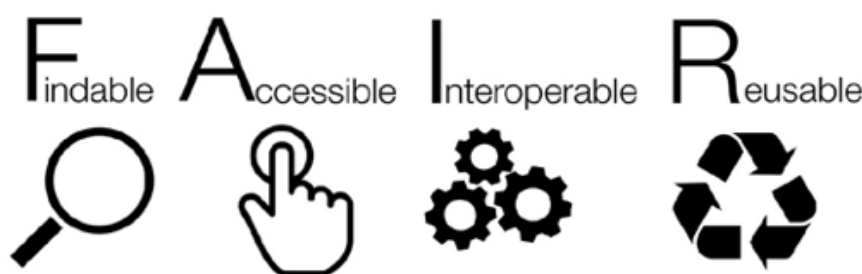
Clinical Data Platforms in Hospitals
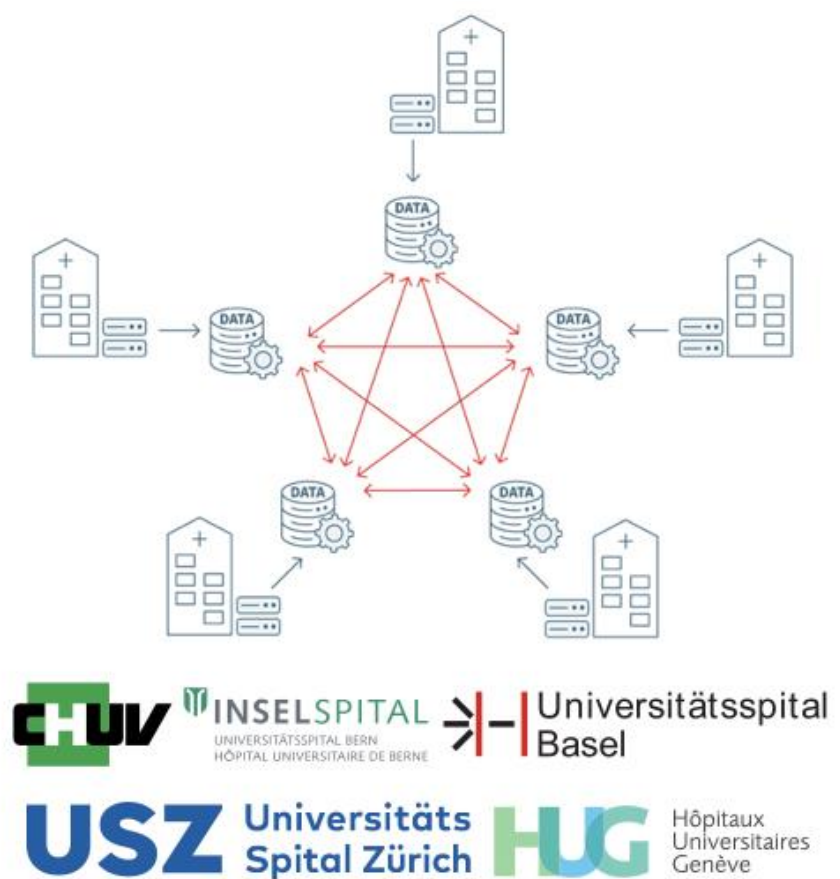
Production and Delivery of FAIR Data
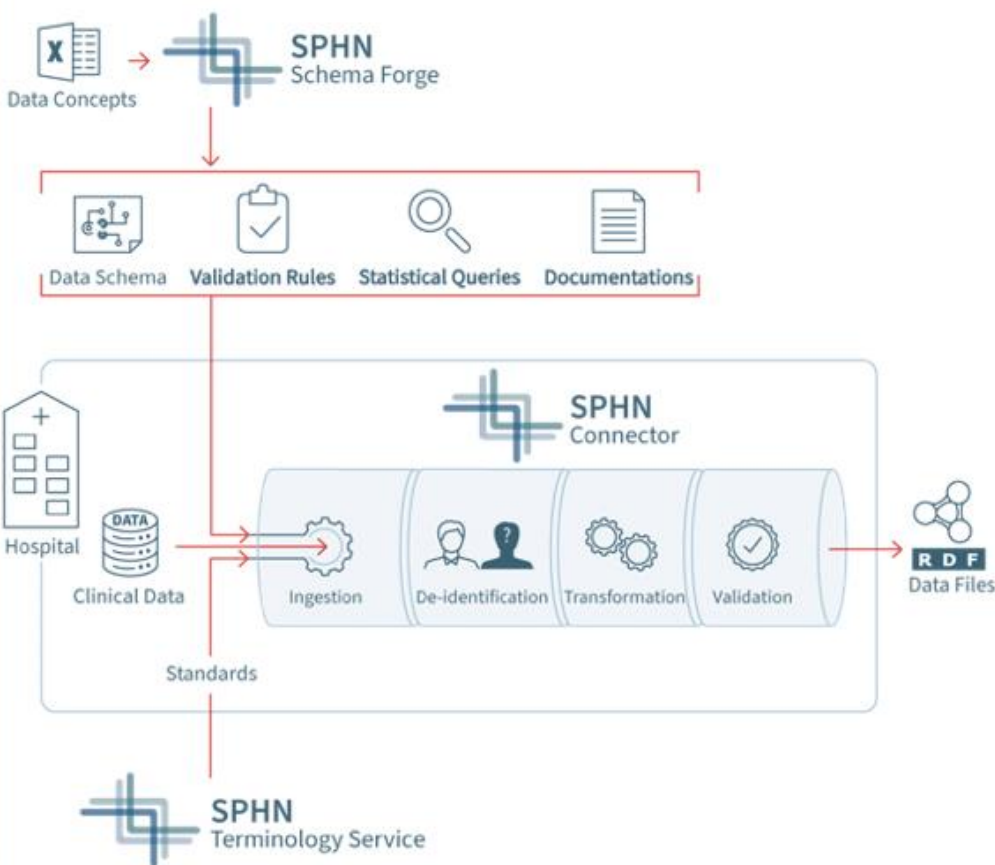
Health Data Discoverability and Reuse

BioMedIT: Trusted Research Environment
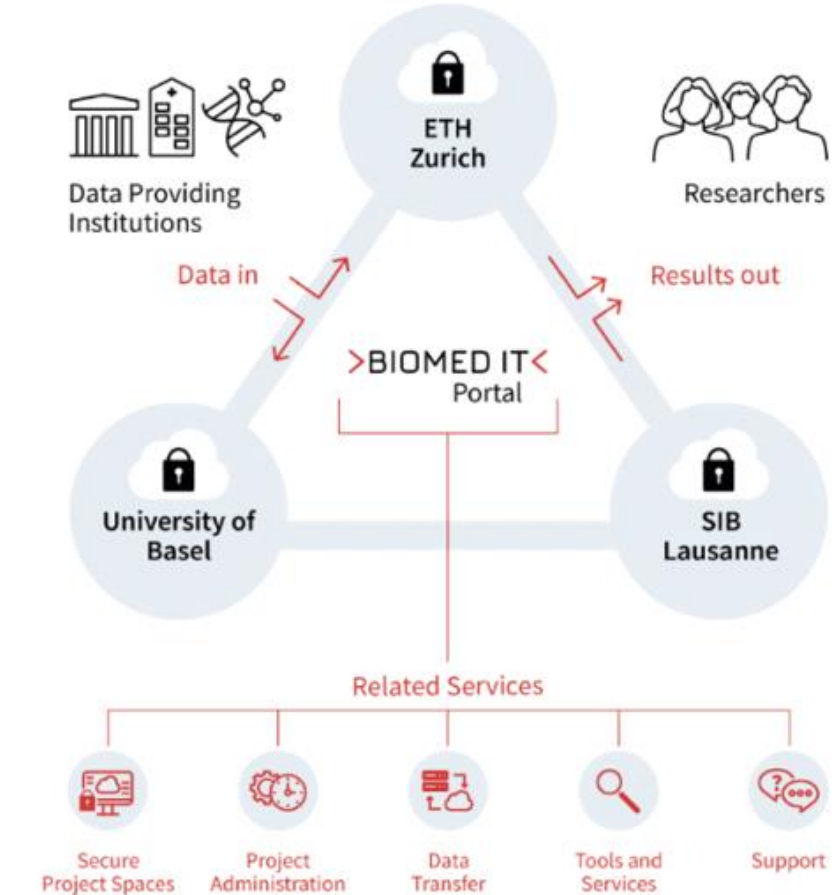
SPHN Interoperability Framework
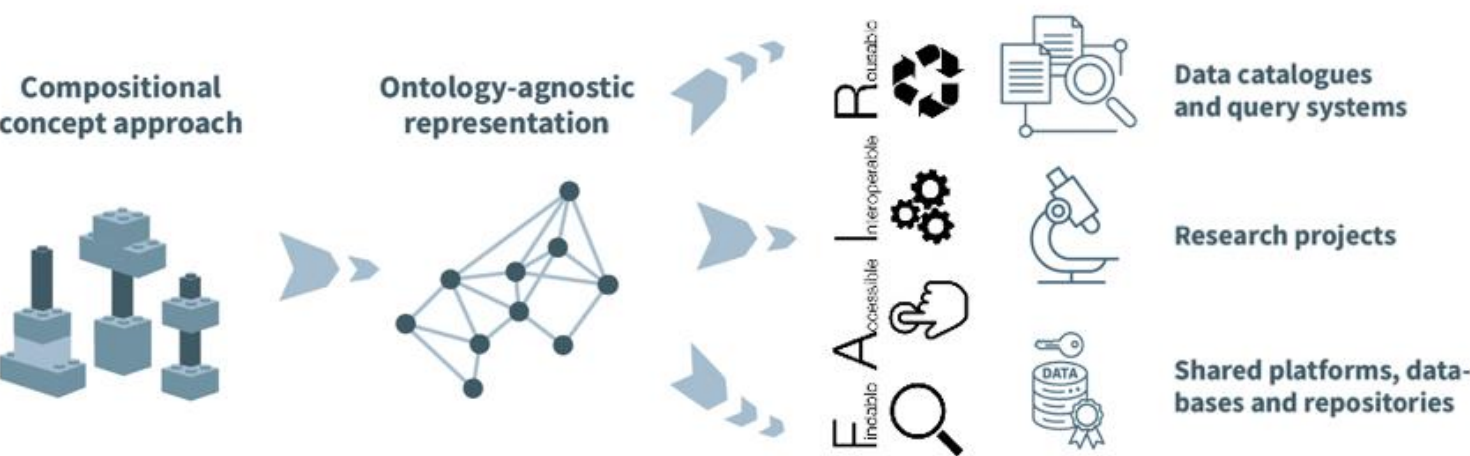
Legal Framework & Data Governance

Outreach and Training

- Responsible use of health-related data for research
- FAIR Health data for research
- Information security and privacy
- BioMedIT: Research with sensitive data

Source: sphn.ch

SPHN Concepts

- «Data catalog» for a specific data domain

- All hospitals must deliver the same set of concepts / data points

- Focus on the use of international terminologies (SNOMED, ICD, LOINC, etc.)

- DCC / Semantic Working Group responsible for concept design

SPHN Core Concepts

- Age, Gender, Height, Weight, BMI

- Hospital Movements / Duration

- Diagnoses, Procedures

- Lab Values, Medication

National Data Streams (NDS)

Swiss Personalized Oncology

- Tumor Stage / Grade

- Variant Descriptor

- Allogeneic Transplantation

- Cancer Therapy Regimen

- Cancer Treatment Complex

- Metastasis Presence

- Molecular Test

- Radiological Oncology Disease Assessment

- Systemic Cancer Therapy

Demonstrator Projects (DEM)

ADONIS

Accelerating detection of neonatal sepsis (ADONIS): a machine learning-based approach

- Heart Rate, Respiratory Rate

- Environment Humidity Measurement

- Environment Temperature Measurement

- Oxygen Measurement

- Mean Inspiratory Airway Pressure Measurement

- Minute Ventilation Measurement

- Necrotizing Enterocolitis Diagnosis

Detailed concept list in the appendix

Milestones

- Development of extensive concept store

- Integration of complex data domains (e.g. NGS, ICU)

Challenges

- Mapping to other data standards (e.g. openEHR, FHIR) not yet available – under development

- Sustainability / long-term planning (infrastructure, contracts)

Vision Research Data Ecosystem

USZ Universitäts Spital Zürich

SPHN Terminology Service

SPHN Schema Forge

Data Concepts

Schema Input

Data Schema	Validation Rules	Statistical Queries	Documen- tation

Federated Data Exploration and Analytics System (DEAS) T∥NE INSIGHT

Federated Query **Federated Analytics**

Privacy-Preserving Federated Query and Analytics on:
- 100+ harmonized and semantically described concepts
- incl. image metadata
- incl. sample information

⬡ *kubernetes deployment possible*

VM *virtual machine deployment*

PM *physical machine deployment*

terminology download

FastAPI

(***) *optional, not part of MVP*
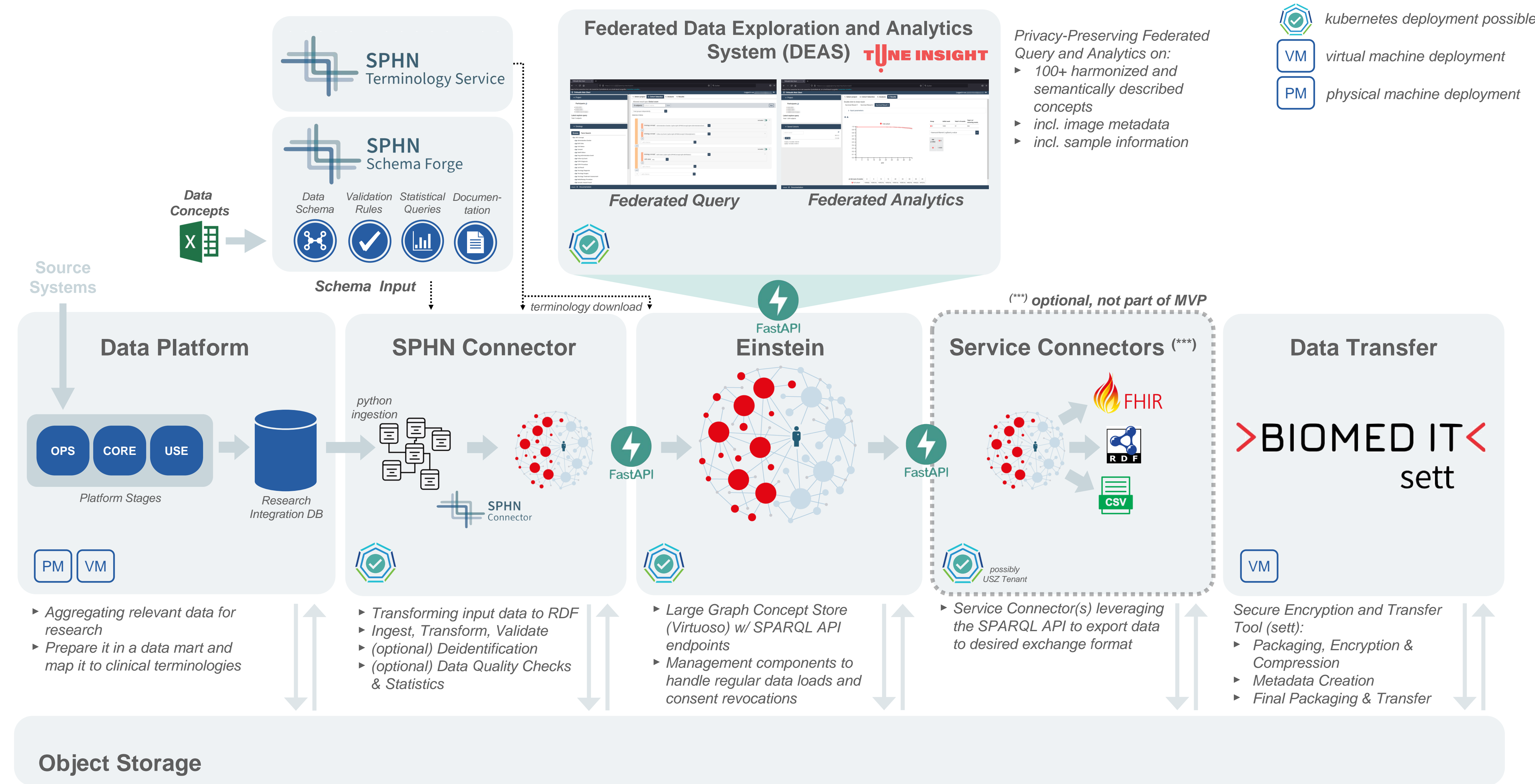
Source Systems

Data Platform

OPS CORE USE

Platform Stages

PM VM

SPHN Connector

python ingestion

SPHN Connector

FastAPI

Einstein

FastAPI

Service Connectors (***)

FHIR

RDF

CSV

possibly USZ Tenant

Data Transfer

>BIOMED IT< sett

VM

Research Integration DB

- Aggregating relevant data for research
- Prepare it in a data mart and map it to clinical terminologies

- Transforming input data to RDF
- Ingest, Transform, Validate
- (optional) Deidentification
- (optional) Data Quality Checks & Statistics

- Large Graph Concept Store (Virtuoso) w/ SPARQL API endpoints
- Management components to handle regular data loads and consent revocations

- Service Connector(s) leveraging the SPARQL API to export data to desired exchange format

Secure Encryption and Transfer Tool (sett):
- Packaging, Encryption & Compression
- Metadata Creation
- Final Packaging & Transfer

Object Storage

USZ Universitäts Spital Zürich

THE LOOP ZURICH

MEDICAL
RESEARCH
CENTER

BioMedical Informatics Platform (BMIP)

„LeonhardMed"
ETH SIS

KISPI

USZ

PUK

BALGRIST
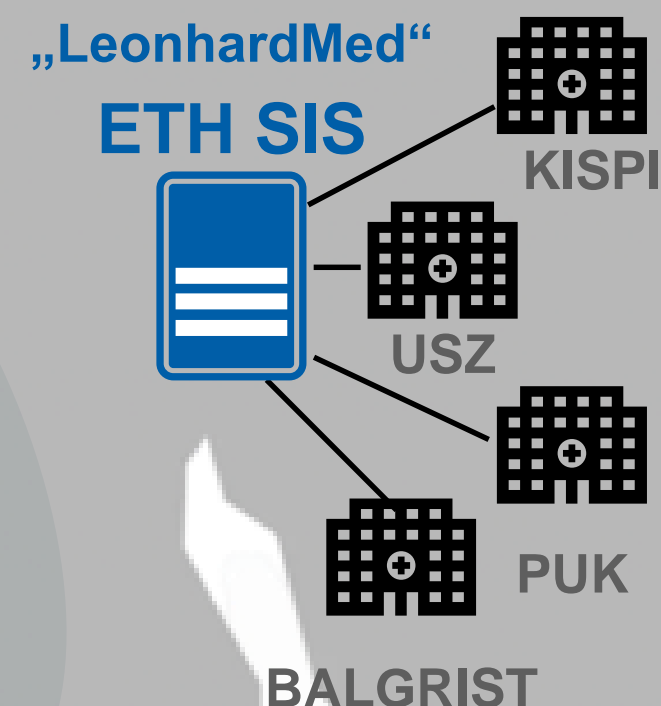
Zurich

„*The LOOP Zurich is establishing a central BioMedical Informatics Platform - The LOOP Zurich BioMedical Informatics Platform (The LOOP Zurich BMIP) - across four university hospitals in Zurich, i.e., the University Hospital Zurich (USZ), the University Children's Hospital Zurich (Kispi), the Balgrist University Hospital (Balgrist) and the University Hospital of Psychiatry Zurich (PUK) together with the two universities ETH Zurich (ETHZ) and the University of Zurich (UZH).*

The LOOP Zurich is seeking proposals from companies or individuals to assist in developing an IT research platform between the major Zurich hospitals and academic partners (University of Zurich and ETHZ) of strategic importance to the Zurich medical research ecosystem. The platform will allow for translational data-driven research involving routine clinical data. It is an integral part of The LOOP Zurich's precision medicine vision and aims to be compatible with national and international biomedical data sharing initiatives." (*Source*)

We look forward to future collaboration projects!