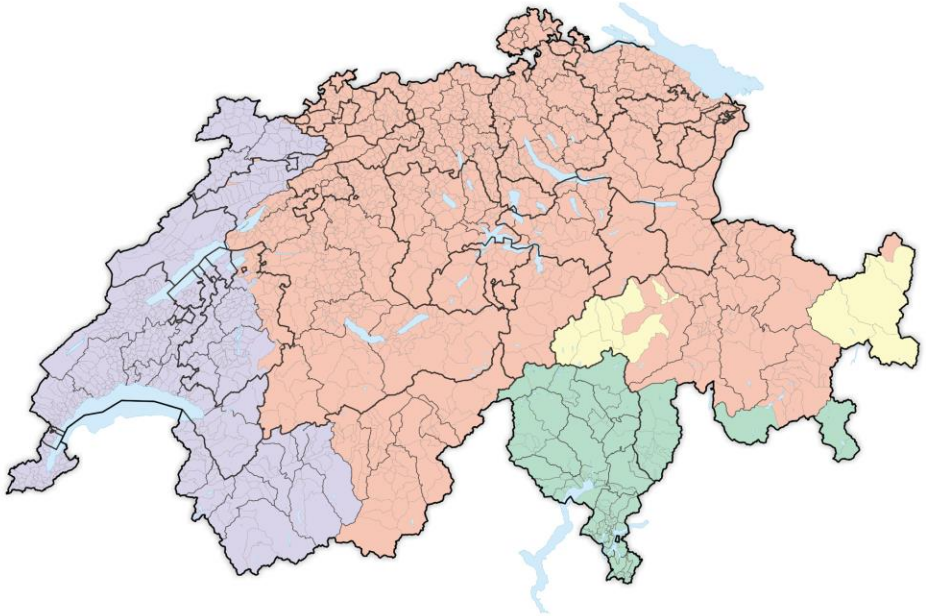
Swiss Personalized Health Network Federated Clinical Routine Dataset

Dr. sc ETH Sabine Österle
Swiss Personalized Health Network (SPHN), SIB Swiss Institute of Bioinformatics

A project of

Switzerland at a Glance



The Swiss Confederation divided into cantons and linguistic regions (French in purple, German in orange, Italian in green, and Romansh in yellow).

Country overview

- Population: ~9 million.

Federal and multilingual system

- 26 cantons.
- 4 national languages.

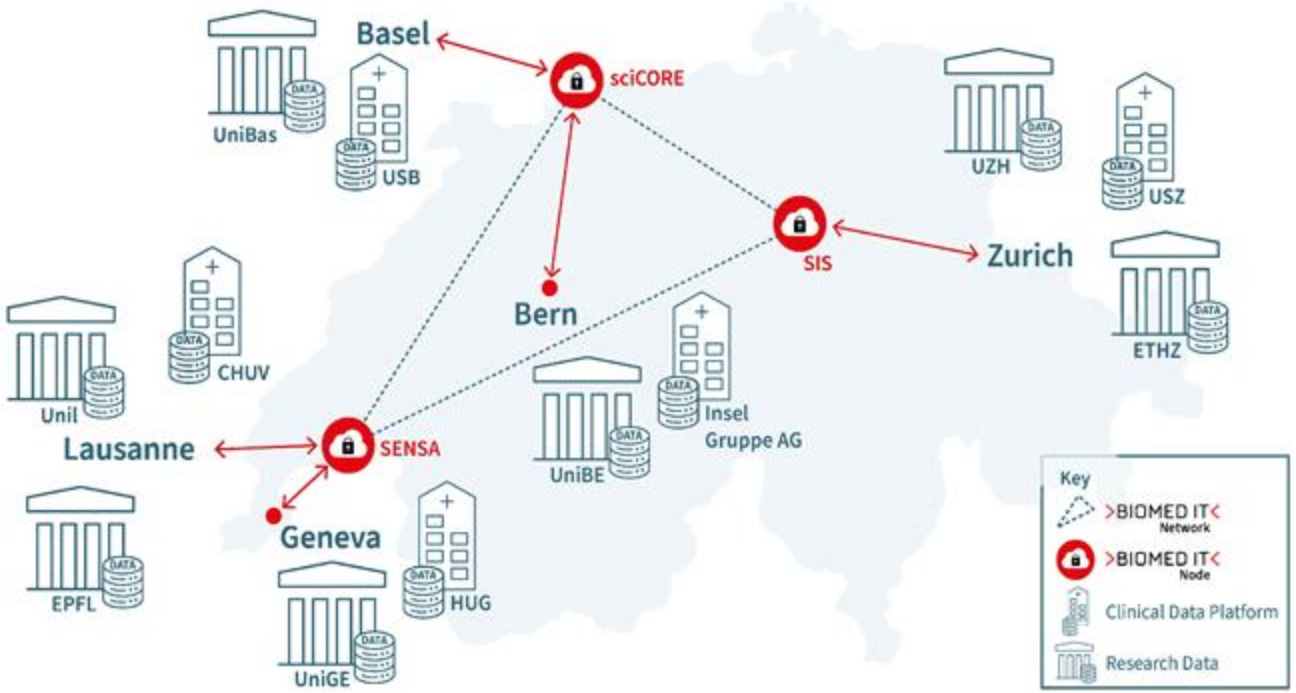
Healthcare landscape

- 5 University Hospitals.
- 40 major hospitals nationwide.

Data protection

- Strict regulations at federal and cantonal levels.
- Strong controls on use and sharing of health data.

Welcome to SPHN

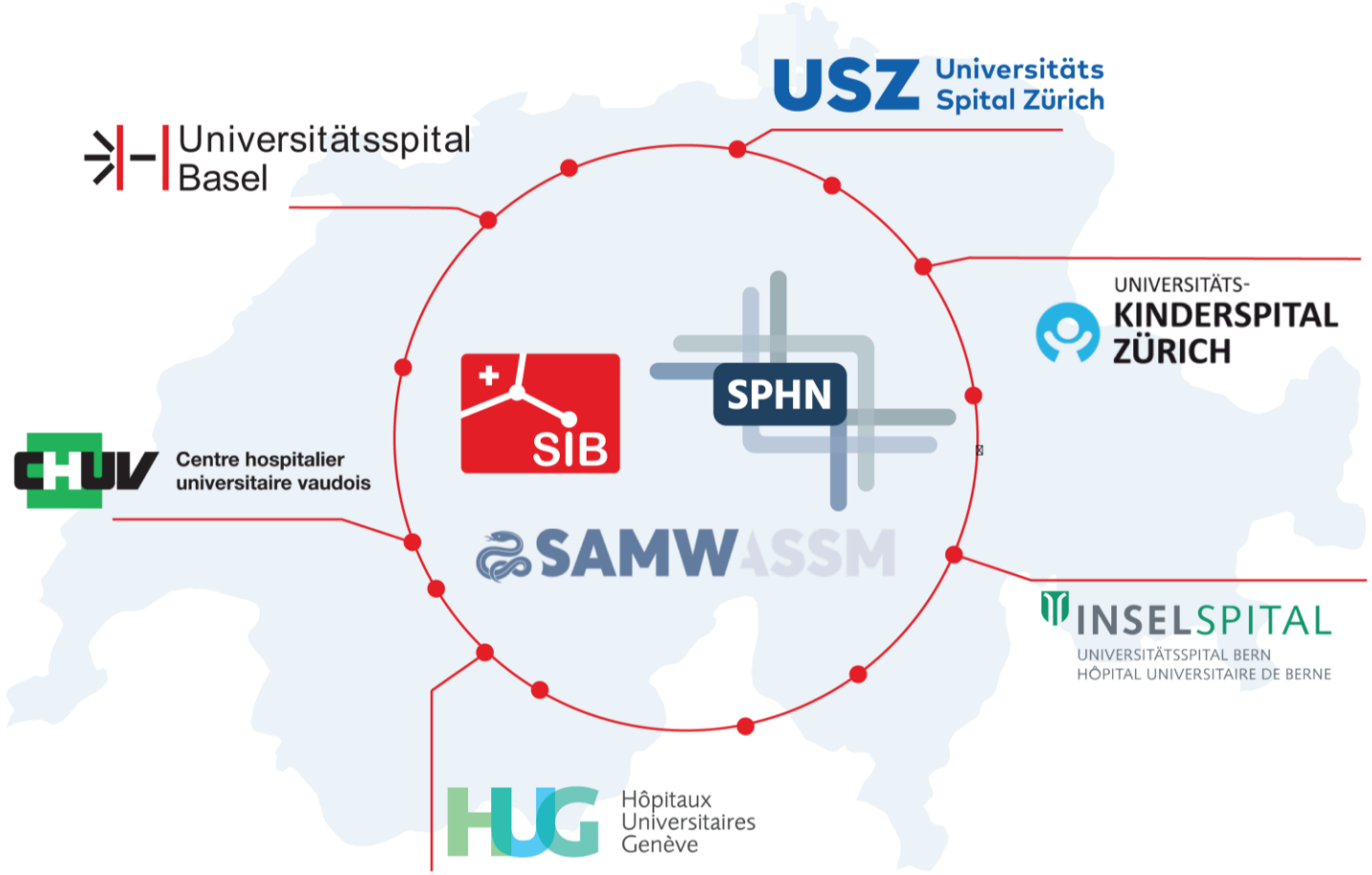


A national data infrastructure coordinated by

SAMW ASSM & SIB

Centre hospitalier universitaire vaudois · KSA · INSELSPITAL · UKBB · KSB · THE LOOP ZURICH · UNIVERSITÄTS-KINDERSPITAL ZÜRICH · ehealthsuisse · USZ Universitäts Spital Zürich · eoc · Personalized Health Alliance Basel-Zurich · HUG Hôpitaux Universitaires Genève · Kantonsspital St.Gallen · SWISS BIOBANKING PLATFORM · SCTO · SWISS CANCER INSTITUTE · FNSNF · University Hospital Basel · SSPH+ · luzerner kantonsspital · SDSC · life sciences cluster basel · health 2030 · swissuniversities · Universitäre Medizin Schweiz

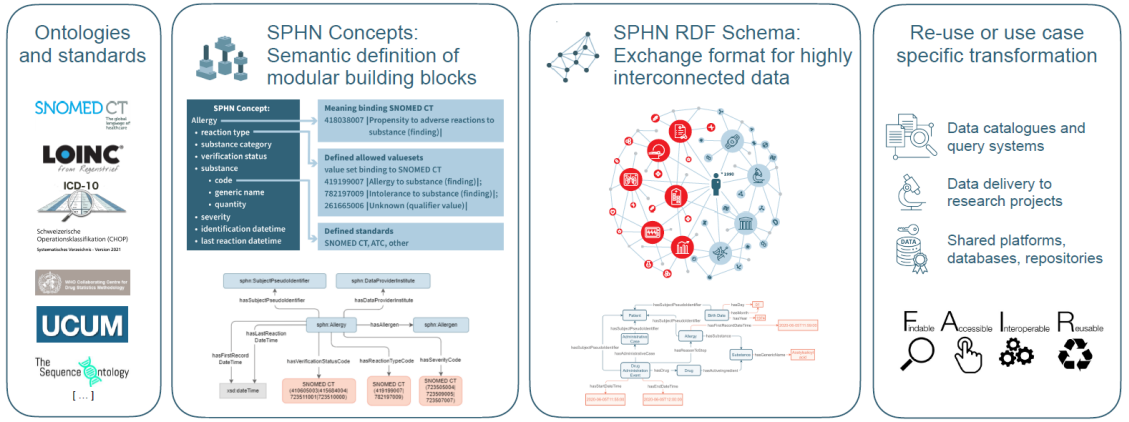
SPHN Federated Clinical Routine Dataset



SPHN Semantic Interoperability Framework

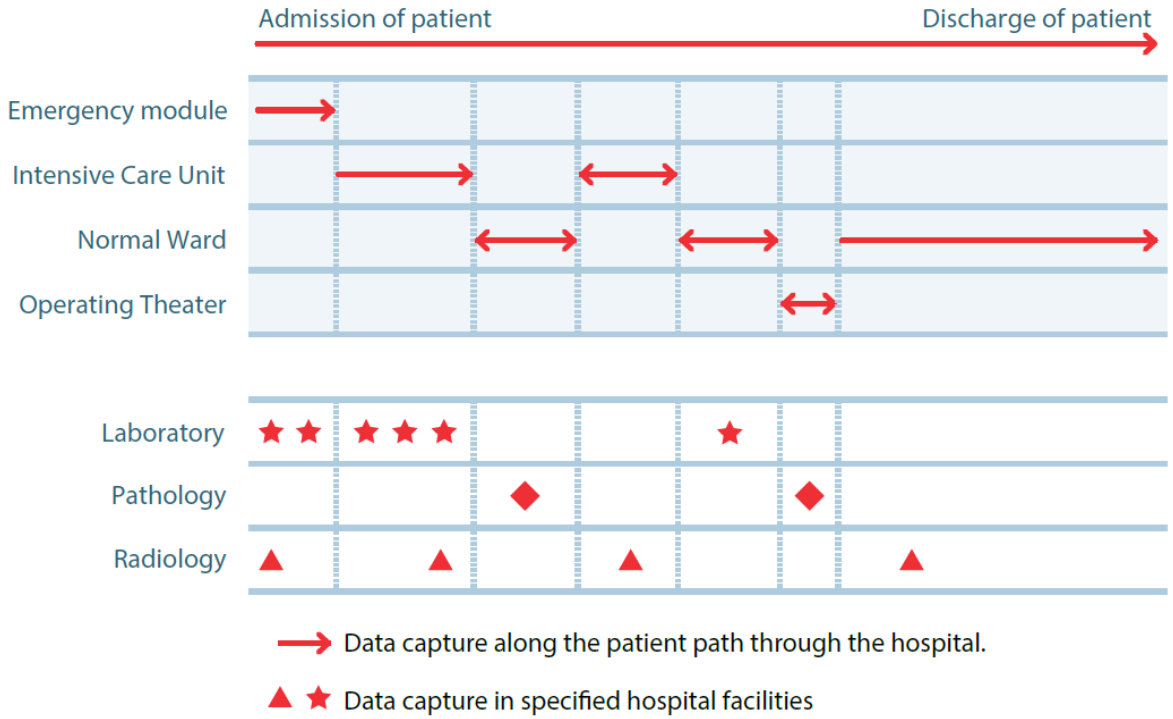
- **Real world data** is complex and diverse level of context of the data is available
- **Biological data** adds a layer of complicity
- W3C (RDF and OWL) and data (SNOMED CT and LOINC) standards allow to **speak a common language**
- **Semantic representation** of the data enables a better understanding of the data

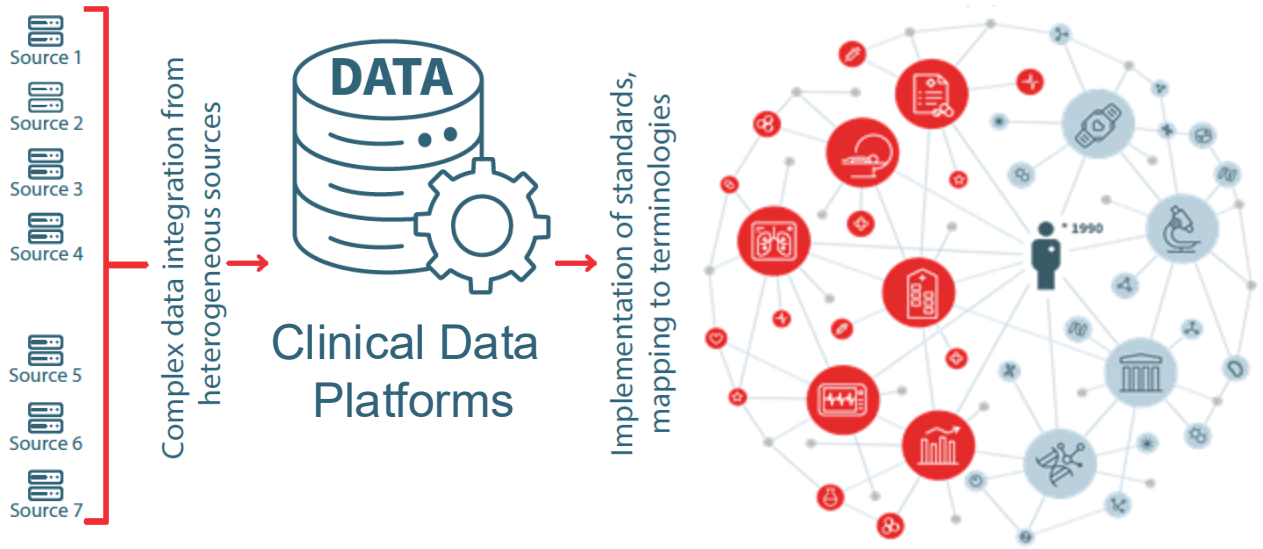
Multi-layered **data integration** into a knowledge graph enables a comprehensive view of the patient



https://sphn-semantic-framework.readthedocs.io/

Core infrastructures at Swiss University Hospitals





- → Human and machine-readable meaning of the data
- → Context - provenance and metadata
- → Linking data from different sources

The SPHN Schema to generate knowledge graphs

210 SPHN concepts (2026.1)

– Nationally aligned

– Meaning defined once

Domains covered

– Administrative items (administrative case, consent)

– Demographics, medications, procedures, diagnosis, lab, measurements (vital signs)

– Medical Imaging, Sample and biobanks

– Special concept for oncology, intensive care, microbiology

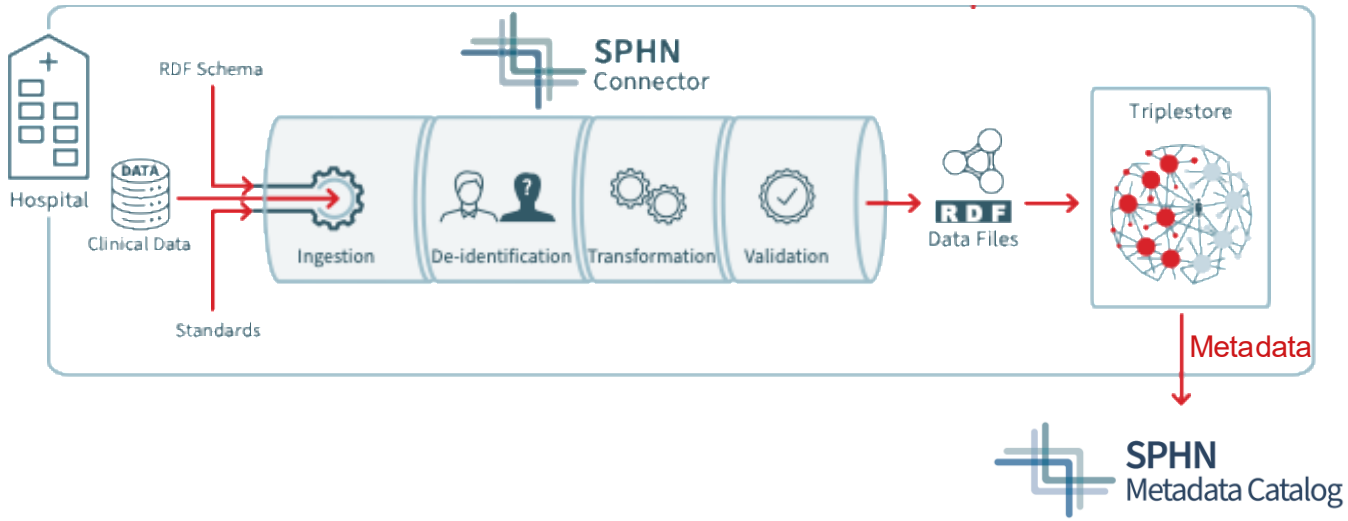
– Omics results and processing

– Data provenance

SPHN Federated Clinical Routine Dataset

Data stays in the hospital

- Hospitals are in control of the data
- Ownership of the platform
- Security and data privacy
- Easier processes for
  - Data updates
  - Consent revocation



SPHN Federated Clinical Routine Dataset

Data context

- 6 hospitals
- Cases from 2018 to current

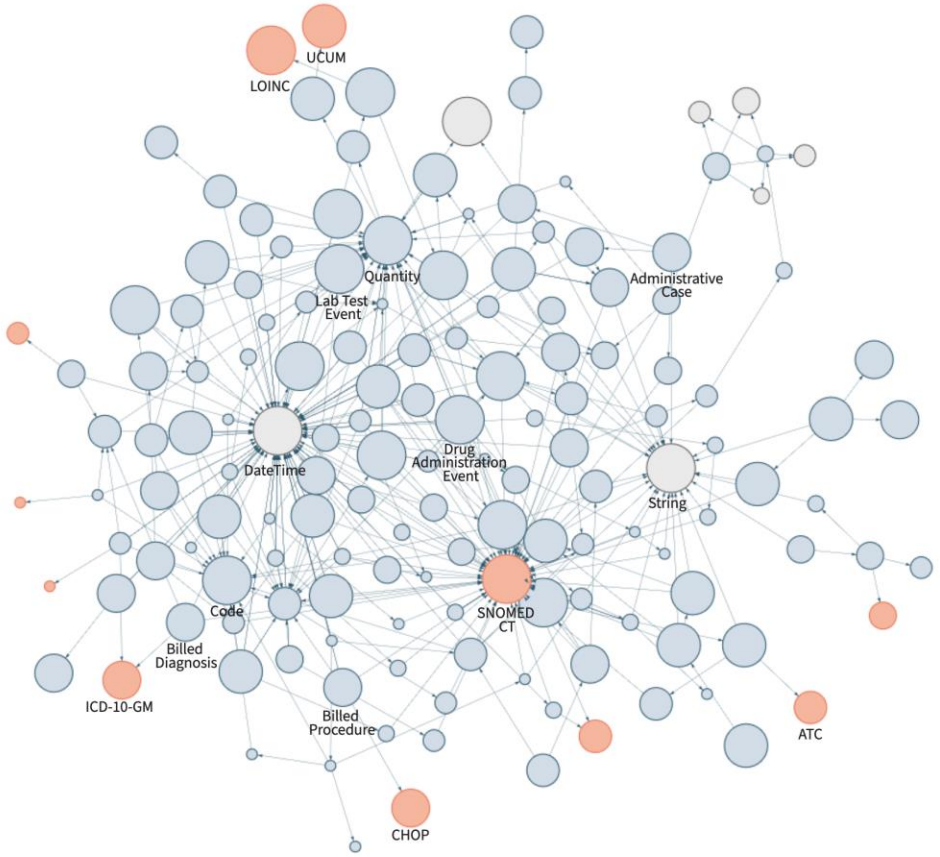
Data numbers

> 800 000 patients (broad consent)

> 12 billion triples

> 6 500 SNOMED CT codes

> 3 600 LOINC codes



Armida et al. (2025); https://doi.org/10.21203/rs.3.rs-8250886/v1

SPHN Federated Clinical Routine Dataset

Data context

- 6 hospitals
- Cases from 2018 to current

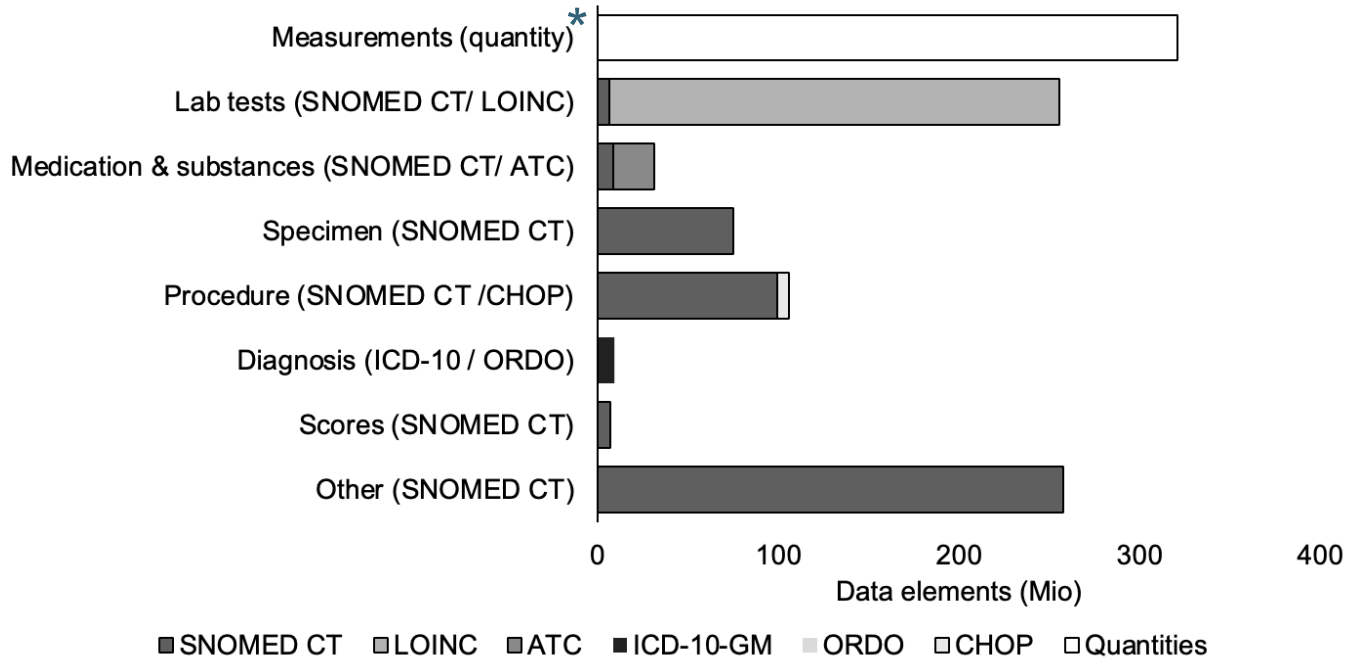
Data numbers

> 800 000 patients (broad consent)

> 12 billion triples

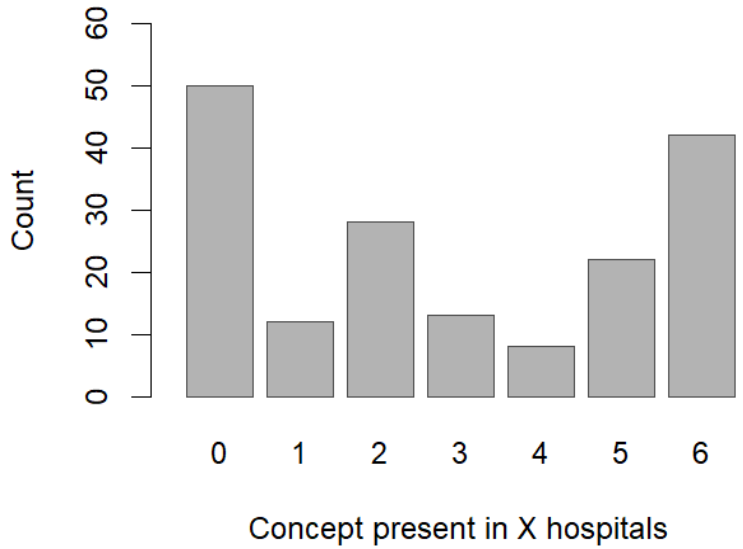
> 6 500 SNOMED CT codes

> 3 600 LOINC codes



* excluding timeseries (PDMS)

Challenge 1: Data Completeness



6. Administrative, Medication, Diagnosis, Procedure, Lab, Sample, Oxygen Administration, Scores, etc.

5. Allergy, BMI, Body Weight, Civil Status, Imaging procedure, Vital signs ...

4. Circumference, Country, Isolate, Nationality, Pharmaceutical Dose Form …

3. Access Device, Insurance Status, Oncology Diagnosis, Organ Support, …

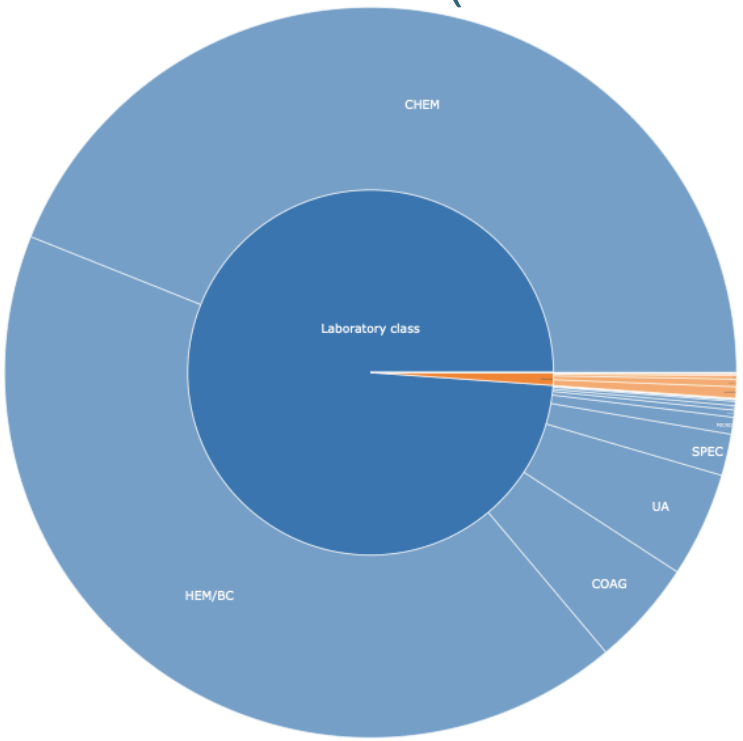
2. Biobanksample, Body Surface Area, Cardiac Index, Cardiac Output, Microbiology, Resuscitation Directive, Tumor grade, Tumor stage, …

1. Fluid Balance, Gene, Implant, Medical Device, Oncology surgery, Transplant, …

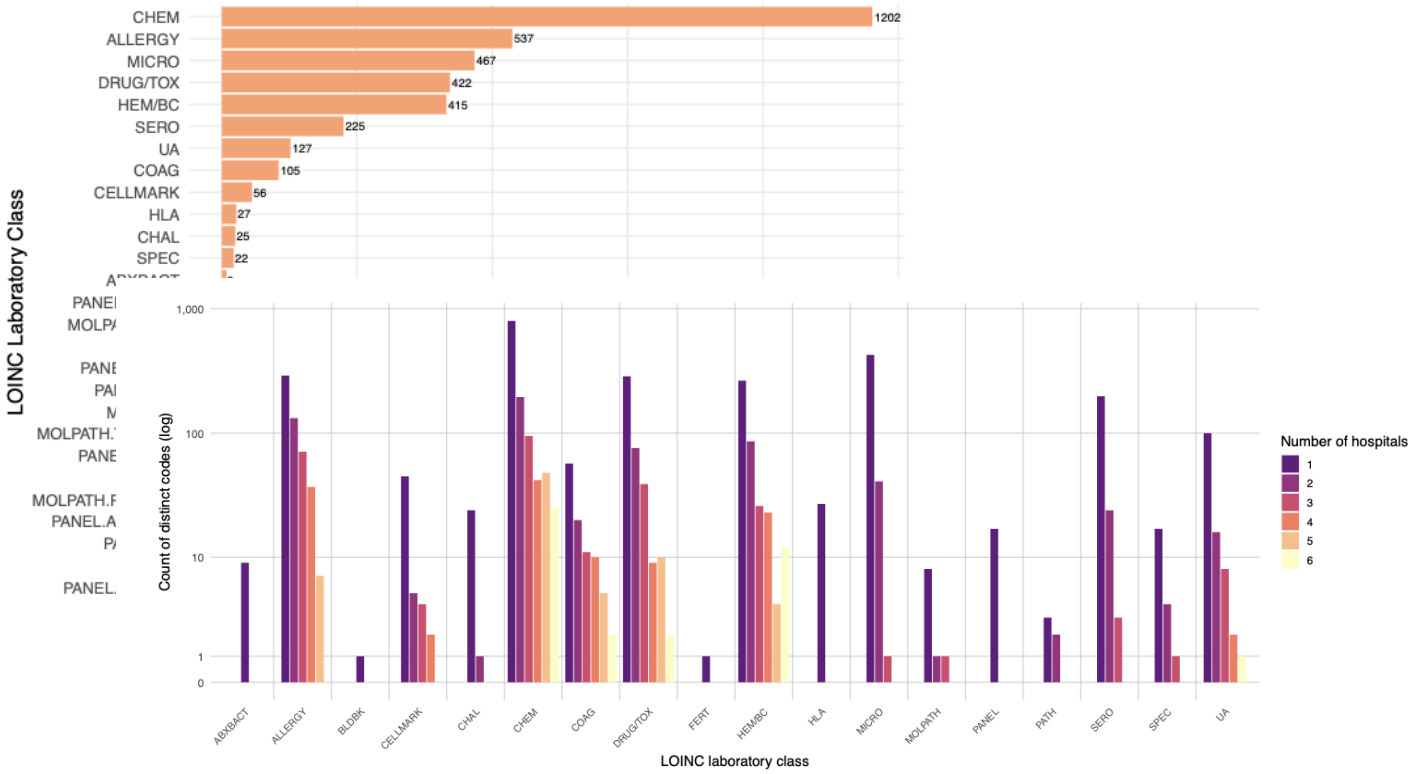
0. Non clinical routine concepts or data not structured at scale at the hospitals

Challenge 2: Data coding (LOINC)

Data Instances (12-122 Mio)

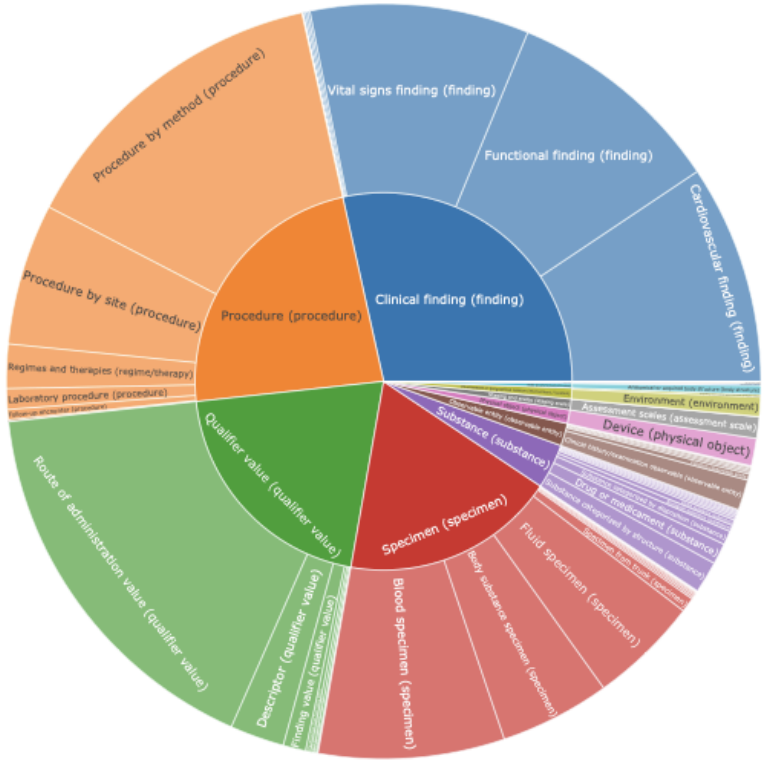


Codes (220 to 1 426 codes) per hospital

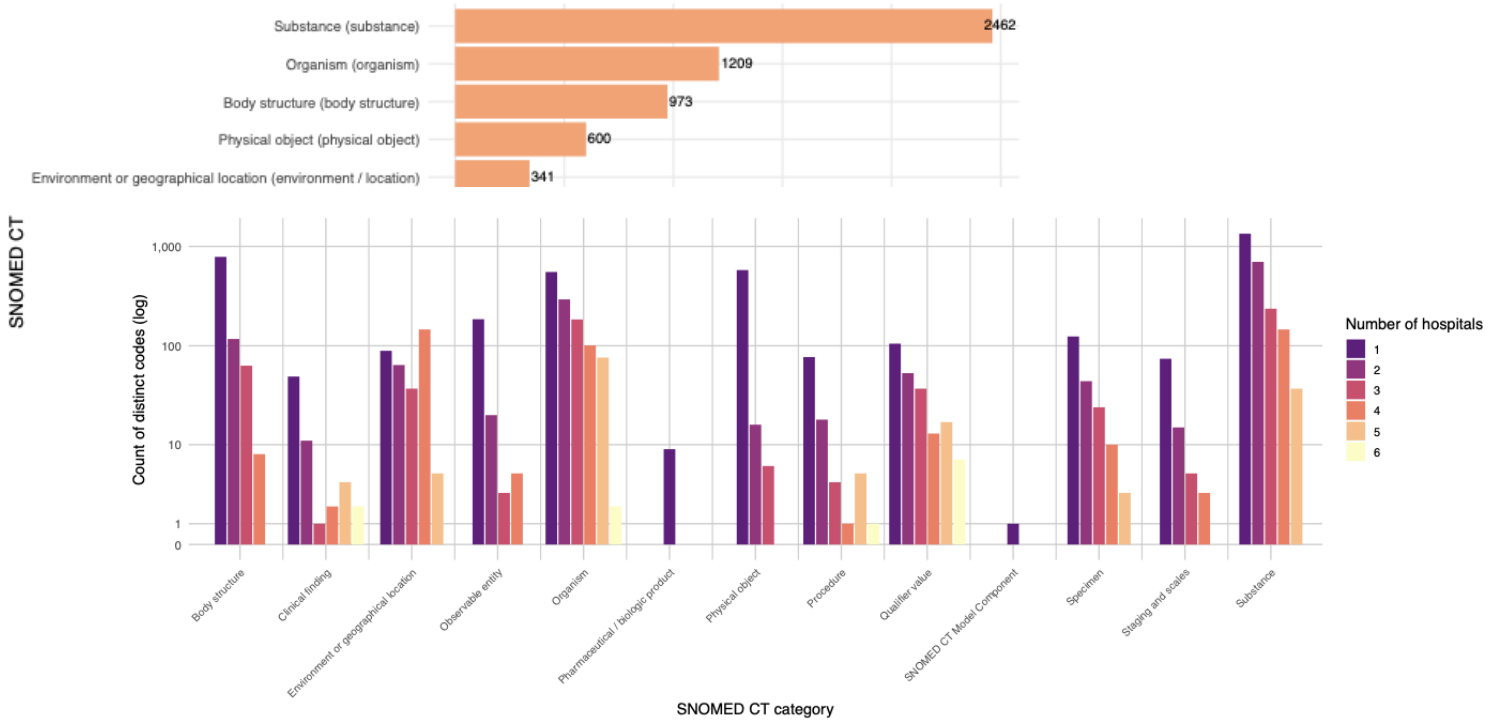


Challenge 2: Data coding (SNOMED CT)

Data Instances (17-56 Mio)



Codes (669 to 3 077 codes) per hospital



Challenge 3: Data quality

- Data completeness, not all concepts present per patient and per hospitals
- End datetime before start datetimes
- Extreme dates and values
- Code which does not exist in the official extension/local extension
- Extremely large patients “Patient with 2 Mio Drug prescriptions”
- Multiple time zones within one hospital
- There is still a lot of free text not yet coded
- Version information for coding's often missing

Challenge 3: Data quality

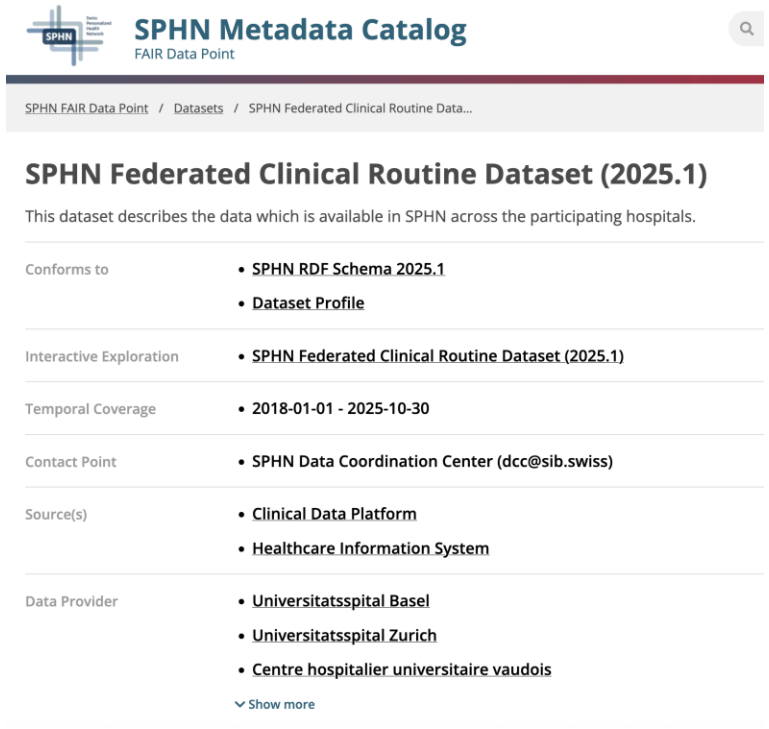
- Data completeness, not all concepts present per pati… …ospitals
- End datetime before start datetimes
- Extreme dates and values
- Code which does n… …ocal extension
- Extremely la… …mio Drug prescriptions”
- Multiple time… …e nospital
- There is still a… …or free text not yet coded
- Version information for coding's often missing

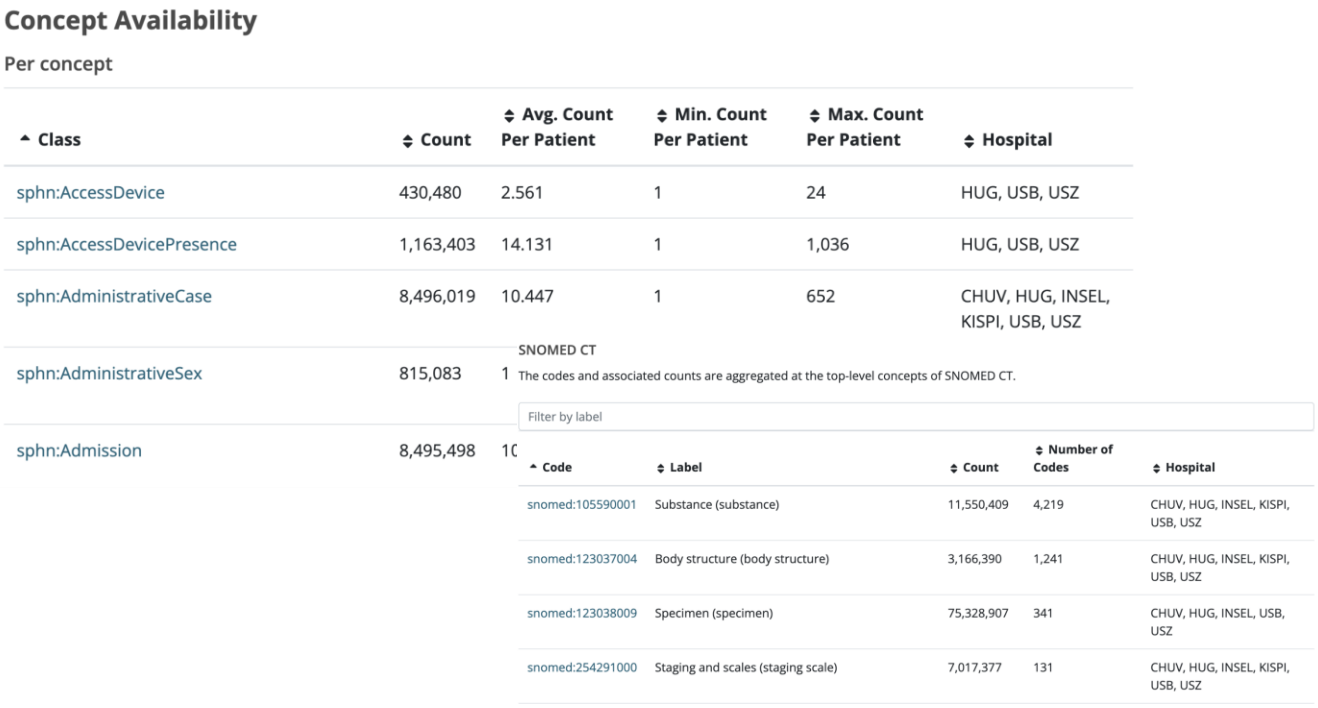
Very hard work!

Often mistakes cannot be fixed retrospectively

SPHN Federated Clinical Routine Dataset

Explore online (HealthDCAT and VOID)



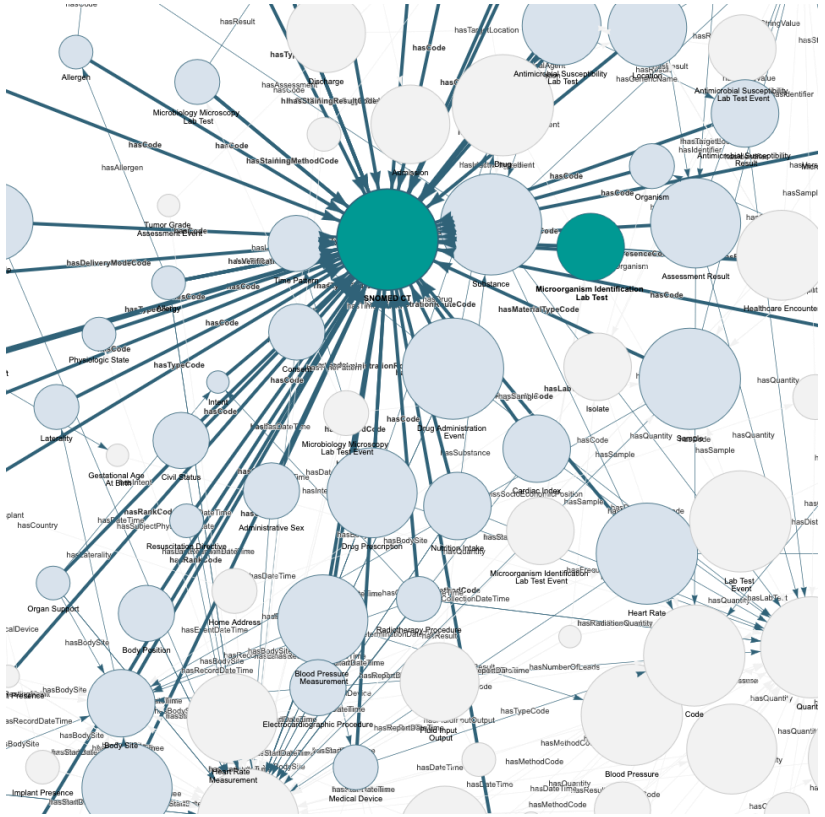


https://fdp.dcc.sib.swiss/catalog/fc6e936d-ad4b-5b21-b8a4-8c7826bbec99

Metadata Catalog- Browsing the graph

Explore Graphstructure online

- Concepts
  - Instance counts
  - Availability per hospital
- Codesystems
  - Instance counts
  - Number of different codes
  - Availability per hospital



Metadata Catalog – Querying Datasets

SPHN FDP SPARQL endpoint **https://fdp.dcc.sib.swiss/store/fdp/sparql**

```
SELECT DISTINCT ?class
WHERE {

  SERVICE <https://fdp.dcc.sib.swiss/store/fdp/sparql> {
   ?distribution dcterms:isPartOf fdp-dataset:58cd4839-
          5821-5dee-a859-72e4462c3659 .

   ?concepts dcterms:isPartOf ?distribution .
   ?concepts a void:Dataset .
   ?concepts void:classPartition/void:class ?class .

   ?class rdfs:subClassOf sphn:Measurement .
  }
}
```

Which SPHN measurements concepts are available in the FedData?

?measurements
sphn:BodyWeightMeasurement
sphn:HeartRateMeasurement
sphn:BodyTemperatureMeasurement
sphn:BodyHeightMeasurement
sphn:BloodPressureMeasurement
sphn:OxygenSaturationMeasurement
sphn:RespiratoryRateMeasurement
sphn:CircumferenceMeasurement
sphn:CardiacOutputMeasurement

Metadata Catalog – Querying Datasets

SPHN FDP SPARQL endpoint **https://fdp.dcc.sib.swiss/store/fdp/sparql**

```
SELECT ?class (COUNT(DISTINCT ?hosp) AS ?count_hospital)
WHERE {
          ?code_frequency a sphn-metacat:DetailedCodeFrequency .
          ?code_frequency void:classPartition ?classSet .
          ?classSet void:class ?class .

FILTER (?class = snomed:459231000124102)

?classSet sphn-metacat:hasHospital ?hosp .
?classSet void:entities ?instance .
          }
GROUP BY ?class
```

SOFA Score code

How many hospitals have the SOFA scores?

?hospital_count
4

Metadata Catalog – Querying Datasets

The Semantic Web approach enables federation...

Are there any pediatric-related datasets in SPHN and HealthRI?

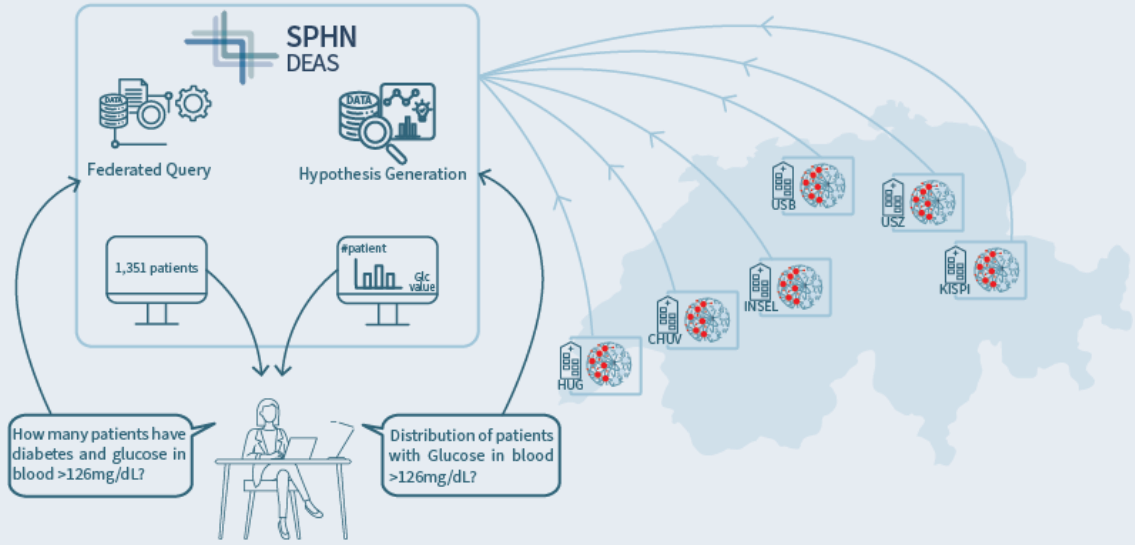
```
SELECT ?title ?source
WHERE {
  VALUES (?service ?source) {
    (<https://fdp.dcc.sib.swiss/store/fdp/sparql>
'SPHN')
    (<https://sparql.healthdata.nl/repositories/fdp>
'HealthRI') }

  SERVICE ?service {
    ?dataset a dcat:Dataset .
    ?dataset dcterms:title ?title .
    FILTER(CONTAINS(LCASE(STR(?desc)), 'pediatric'))
}
```

title	?source
SwissPedHealth Main Dataset	SPHN
SwissPedHealth Nested Project 1: Anthropometric routine data in Switzerland (SwissPedGrowth)	SPHN
Clinical features of COVID-19 in Pediatric Patients	HealthRI
Genetic Risk factors for Multi-system Inflammatory Syndrome in Children and Pediatric post COVID condition	HealthRI

From Querying Metadata to Data directly

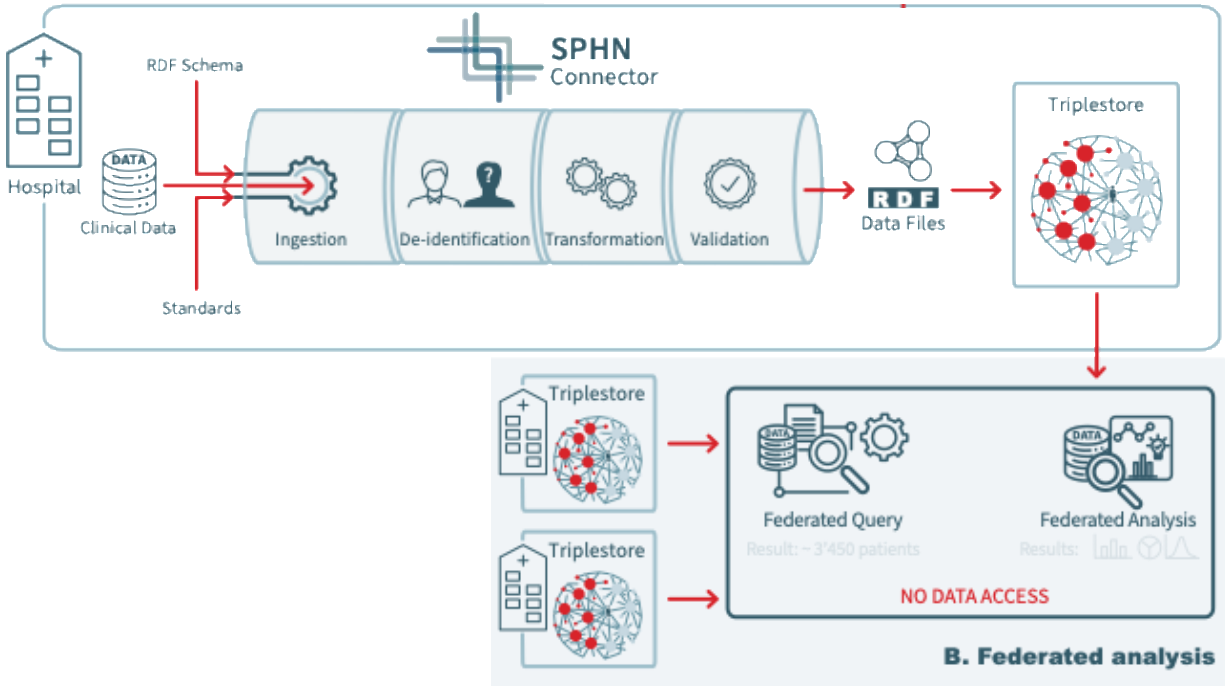
Is it possible to build an application layer on top of the Federated Clinical Routine Dataset?



What would be needed?

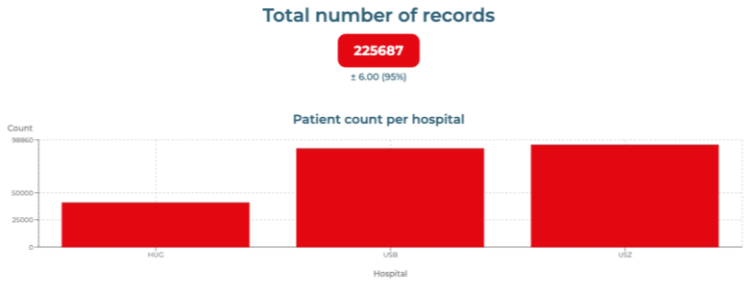
Application

- Way to control the queries (content and performance)
- Query and result federation

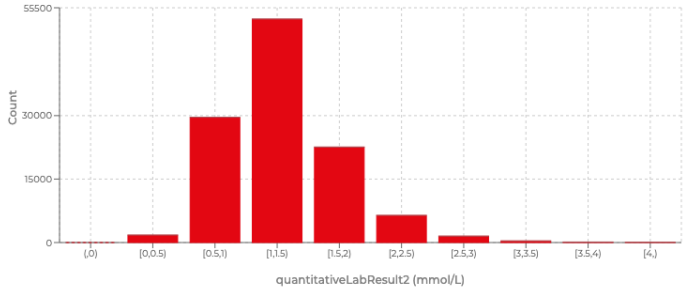


Use cases

Use case 1: Feasibility



User case 2: Value distribution



Use case 3: Means

→ User only see anonymized aggregated patient counts and statistics
→ Currently in the pilot phase
→ Go live planned mid 2026

In collaboration with TUNE INSIGHT

Next steps

Data

- Migrate to 2026.1 Schema
- Close the gaps (missing concepts and properties)
- Improve data quality
- Align SNOMED CT and LOINC coding

Query system

- Go live summer 2026 for university hospital users, then extend the users groups

Use cases

- Research projects using the FedData or subsets of it
- End to end process for data deliveries

Acknowledgments

The SPHN Implementation Teams:

Thomas Geiger, Christine Remund, Davide Chiarugi, **Deepak Unni**, Harald Witte, Jan Armida, Judit Kiss Blind, Julia Maurer, Manuela Paganini, Michaela Egli, Owen Appleton, **Andrea Brites Marto**, Patricia Fernandez Pinilla, Regan Geissmann, Sabine Österle, Sarah Vermij, Shubham Kapoor, Sergio Guarino, Simone Guzzi, **Vasundra Touré**, **Philip Krauss**

Local Project Teams at the hospitals (USB, USZ, INSEL; HUG; CHUV, KISPI):

Katie Kalt, Yves Jaggi, Jesus Fernandez, Xeni Deligianni, Maren Diepenbruck, Luca Leuenberger, Pedram Bürgin, Janshah Veettuvalappil Ikbal, Klaas Kreller, Helena Peic Tukuljac, Abdel Zalmate, Mathias Gassner