Swiss Personalized Health Network – A national, graph-based framework for the semantic representation of FAIR health data

Dr. sc. ETH Sabine Österle
Team Lead Data Interoperability, Personalized Health Informatics Group
SIB Swiss Institute of Bioinformatics

A project of

Dr. sc. ETH Sabine Österle

- Team Lead Data Interoperability at the SPHN Data coordination center (SPHN-DCC), SIB Swiss Institute of Bioinformatics
- Doctorate, ETH Zurich
- BSc and MSc Interdisciplinary Science ETH Zurich

Relationships with a for-profit and/or a not-for-profit organization to disclose:

- Lecturer at ETH Zurich
- Member of the GA4GH Standards Steering Committee

Objectives

1) To understand the Swiss Personalized Health Network (SPHN) framework for the semantic representation of health data in a knowledge graph.
2) To enable researchers to build medical knowledge graphs in a simplified way: a tool stack for the easy design, generation, and validation of RDF graph data from multiple sources, and to explore the potential to apply them to your project.
3) To explore the challenges and limitations associated with using clinical routine data for research and knowledge graph development, including issues related to data quality, privacy concerns, data silos, and interoperability hurdles.
4) To learn about the limitations of existing medical standards and SPHN's approaches to making such vocabularies more FAIR (Findable, Accessible, Interoperable, Reusable principles) and more usable.

Switzerland

8.6 Mio people

Four national languages

26 cantons

- Smallest one 37 km^2

Healthcare is in the responsibility of the cantons

SPHN

Government funded initiative

135 Mio CHF (210 Mio CAD) for 2016-2024

Secure infrastructure network for FAIR health-related data for research

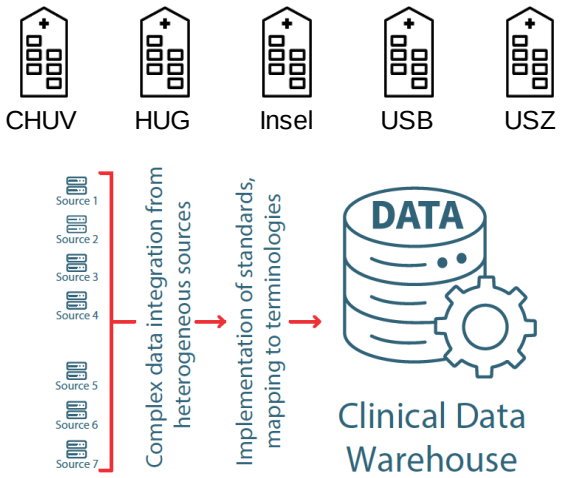
Data alone is not enough

For a functioning data ecosystem and broad use of health data in research and innovation, there is a need for:

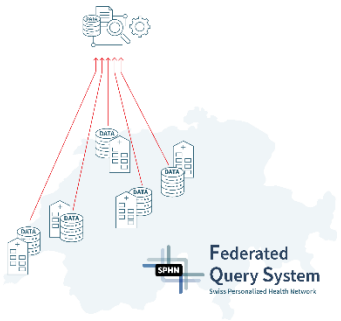
- Scalable and sustainable core infrastructures
- A smart governance framework
- Public acceptability
- A close coordination at national and international level
- A change in research culture: to foster collaboration, transparency, and reproducibility and to ensure responsible data practices

SPHN infrastructures, services and support

UH Clinical Data Platforms



Data discoverability and reuse



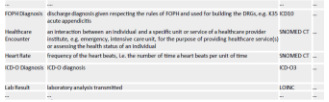
Swiss cohort findability

mælstrom

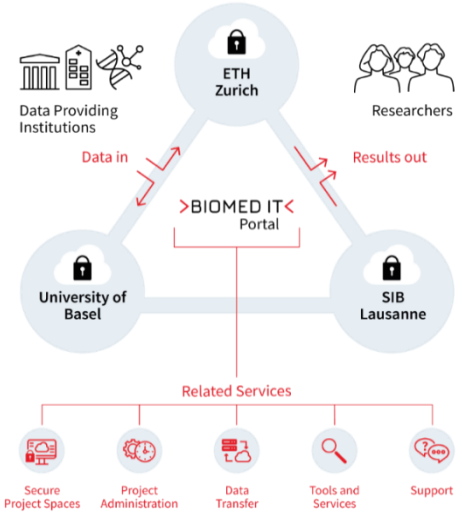
Federated data exploration

TUNE INSIGHT

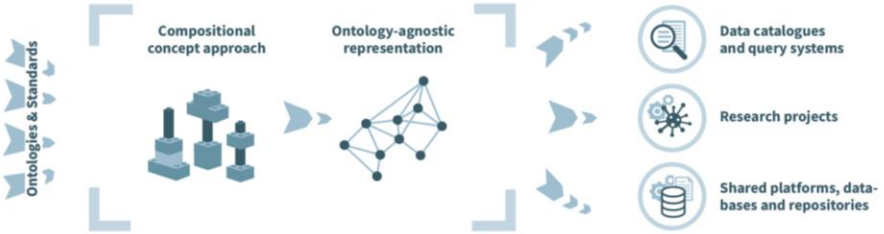
SPHN Metadata catalogue



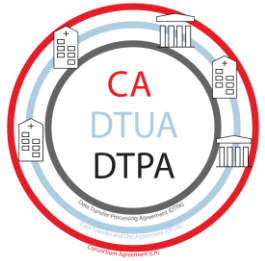
BioMedIT Network



SPHN Semantic Interoperability Framework



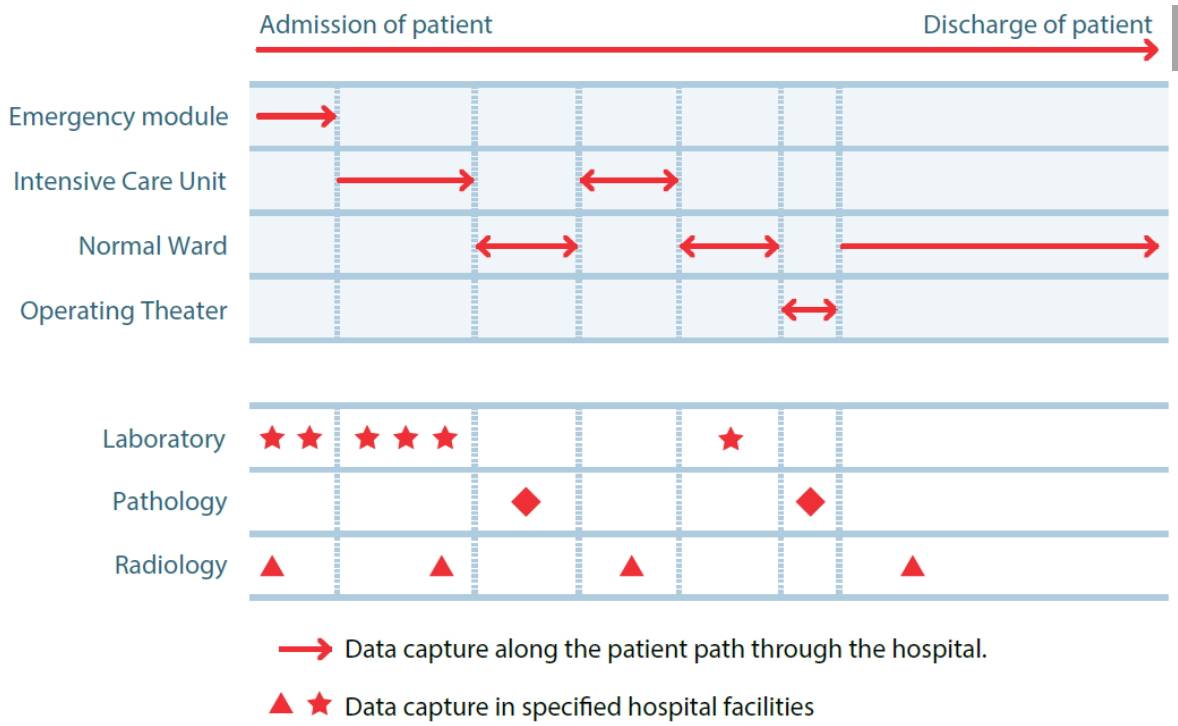
Legal Framework



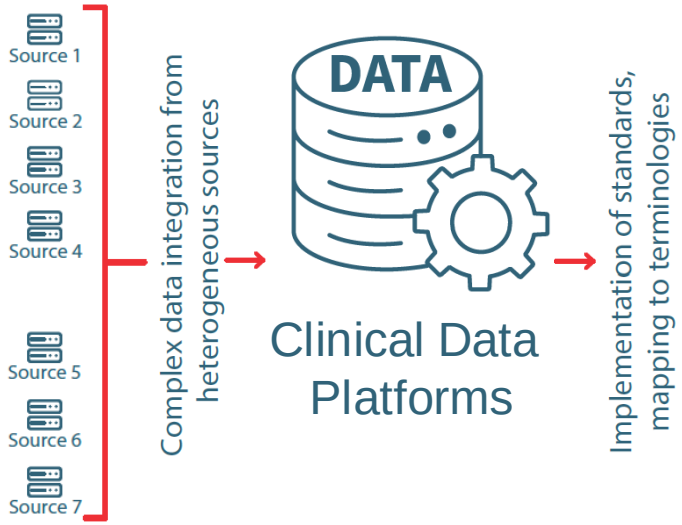
Outreach and Training

- Responsible use of health-related data for research
- FAIR Health data for research
- Use of the Graph data and the SPHN Semantic Tool stack
- Information Security and privacy
- BioMedIT: Research with sensitive data

Core infrastructures at Swiss University Hospitals



[primary use]



[primary use]
- Clinical care
- Tumorboards
- Quality control

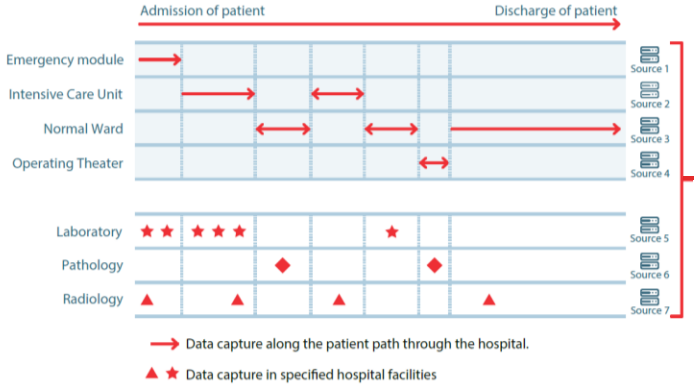
[secondary use]
- Research (in-house and external)
- National initiatives
- Data transfers to registries, authorities
- etc.

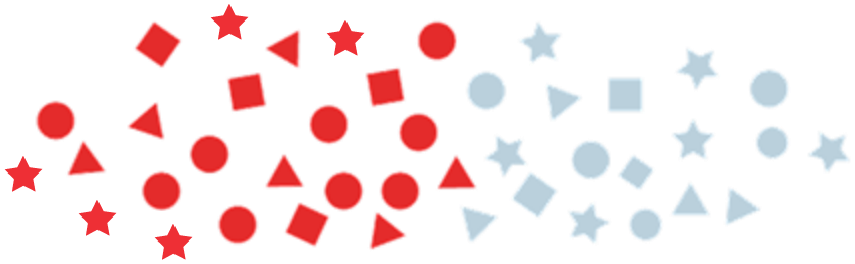
[standards, codes]

[data specification]

The need for data interoperability [at the source]

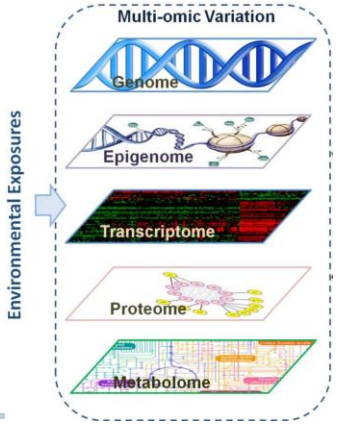
Clinical routine data





Other health-related data

- Omics Data
- Clinical Trial Data
- Registry Data
- Patient Reported Outcomes
- Sensor Data
- Medical Imaging
- etc.



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5742494/

→ Understanding the meaning of the data (by humans & machines)

→ Easy linkage of data from different sources

The SPHN Interoperability Framework: Making sense of data and adding value

Data semantics

Use of controlled vocabulary for concepts, valuesets and data

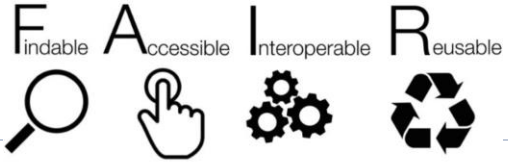
- Data standards: ATC, SNOMED CT, ICD-10-GM, CHOP, UCUM, LOINC, NANDA, ICD-O, GENO, SO...
- Multi-level coding and mappings

Language for knowledge representation

- Semantic Web (RDF Schema)
- Controlled vocabulary (rdfs, owl, skos)

Quality control framework

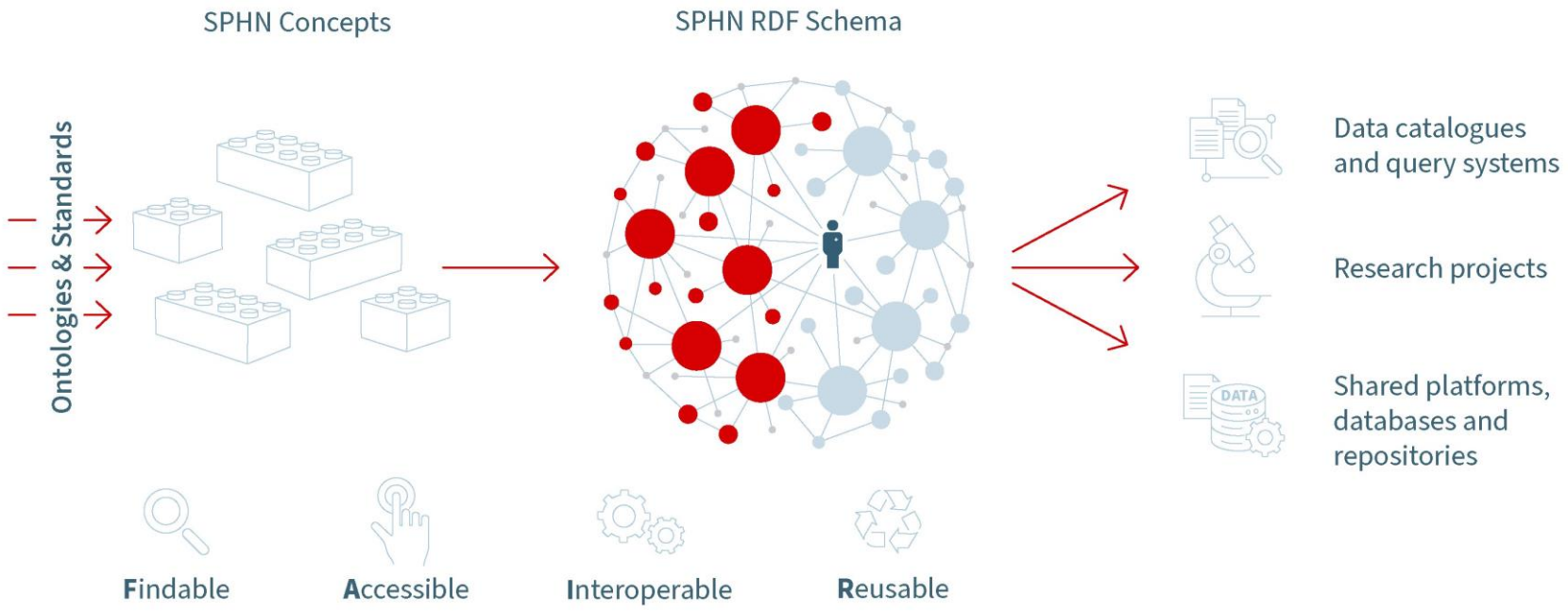
- Semantic Web (SHACL, SPARQL)



Why a knowledge graph?

- **Semantic representation of the data** rather than harmonization to fit in one data model
- **Flexibility**: Allows representation of complex, interrelated data and the integration of multimodal data from different sources
- **Scalability**: KGs can handle large amounts of data and are designed to be scalable
- **Intuitive and visual** presentation of relationships and connections between data elements, allowing researchers to easily explore relationships, and to identify new signals
- **Linked data:** seamless use of data and semantic standards
- **Knowledge inference**
- **FAIR principles** implementation

SPHN Semantic Interoperability Framework



SPHN Concepts

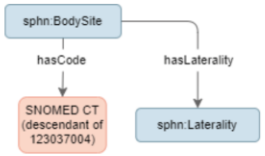
Body Site [C]

Classes

URI https://biomedit.ch/rdf/sphn-ontology/sphn#BodySite

Description any anatomical structure, any nonspecific and anatomical site, as well as morphologic abnormalities

Schema representation



Meaning binding (Equivalent-classes)
SNOMED 123037004 | Body structure (body structure) | [C]
LOINC 39111-0 | Body site | [C]

Parents SPHN Concept [C]

Property (In the domain of)		Cardinality	Class or Datatype	Restriction
	has laterality [op]	0 .. 1	Laterality [C]	
	has code [op]	1 .. 1		Yes

Restrictions
▾ has code [op] : descendants of the given classes are allowed

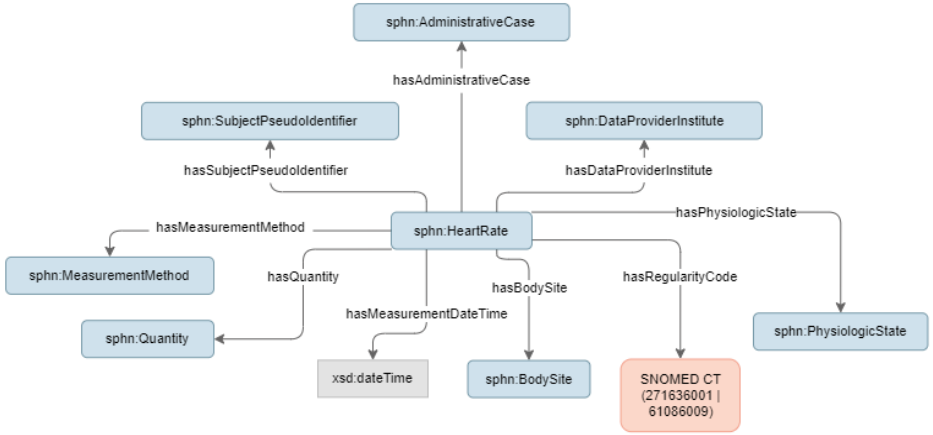
SNOMED 123037004 | Body structure (body structure) | [C]

Heart Rate [C]

URI https://biomedit.ch/rdf/sphn-ontology/sphn#HeartRate

Description frequency of the heart beats, i.e. the number of time a heart beats per unit of time

Schema representation

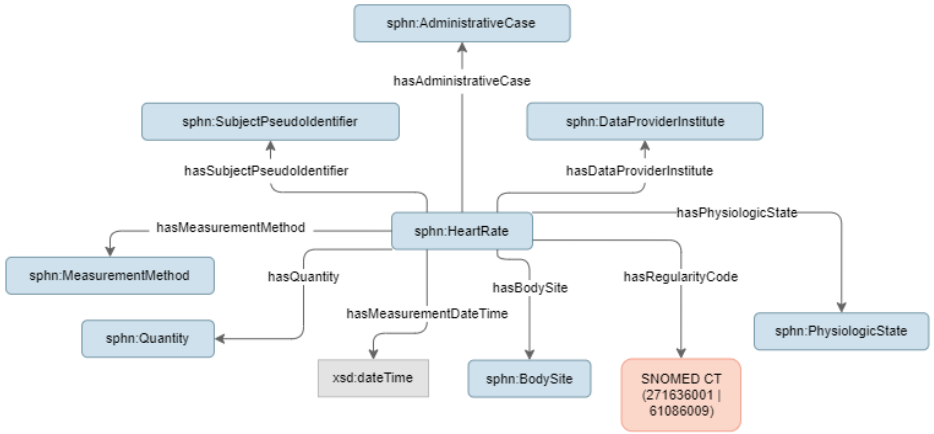


Meaning binding (Equivalent-classes)
LOINC 8867-4 | Heart rate | [C]
SNOMED 364075005 | Heart rate (observable entity) | [C]

https://www.biomedit.ch/rdf/sphn-ontology/sphn/2023/2

Meaning binding (e.g. SNOMED CT)

Heart Rate [c]

URI	https://biomedit.ch/rdf/sphn-ontology/sphn#HeartRate
Description	frequency of the heart beats, i.e. the number of time a heart beats per unit of time
Schema representation	
Meaning binding (Equivalent-classes)	LOINC 8867-4 | Heart rate | [c] SNOMED 364075005 | Heart rate (observable entity) | [c]

Defining SPHN concepts

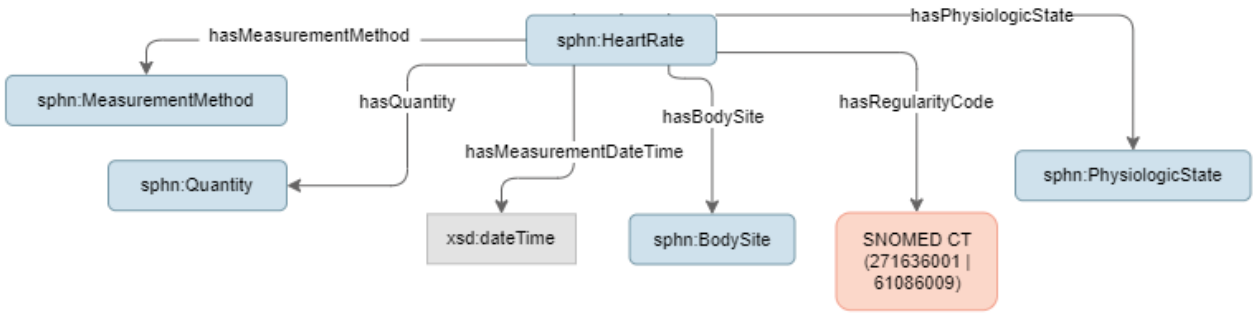
e.g., SPHN Concept Heart Rate – bound to SNOMED CT concept 364075005 |Heart rate (observable entity)|

→ Implemented via owl:equivalentClass

→ 64/87 of SPHN concepts have a meaning binding

→ Used standards: SNOMED CT and LOINC

Value set binding (e.g. SNOMED CT)



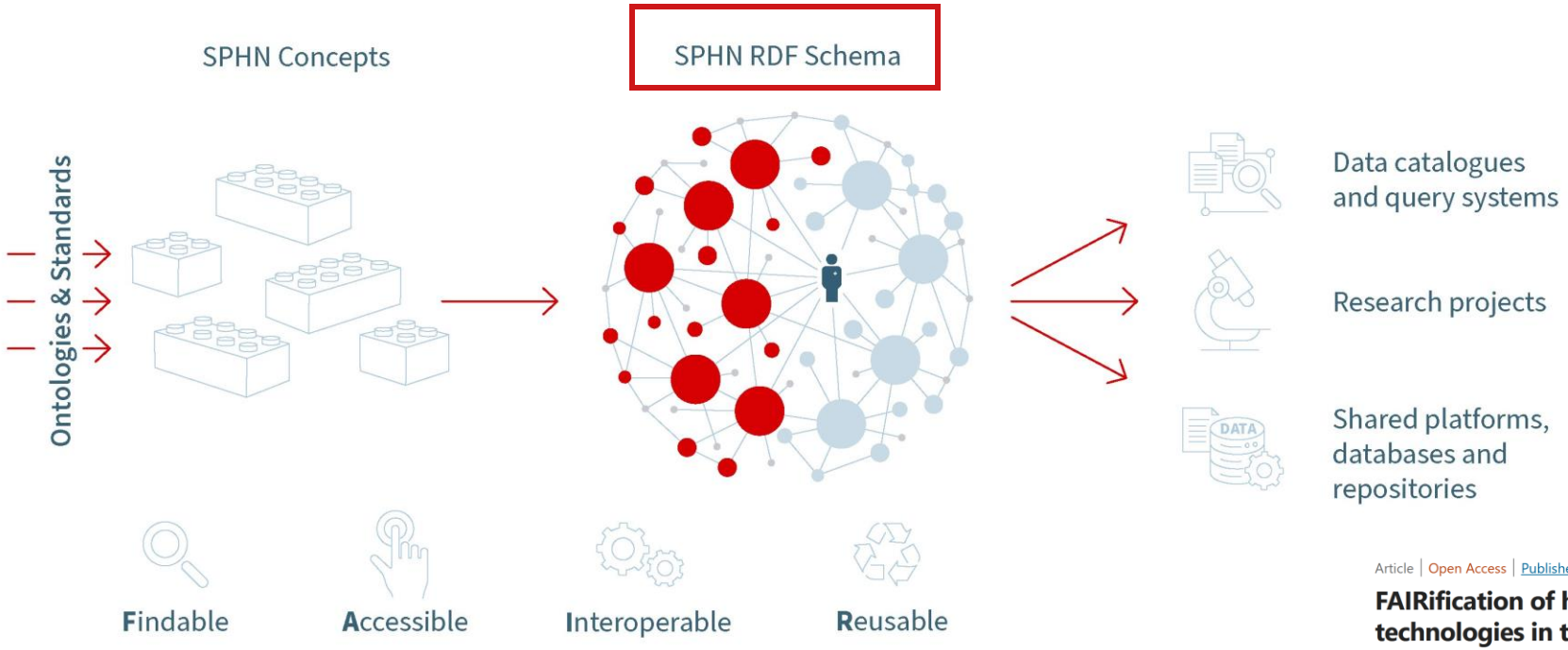
Restrictions

▼ has regularity code [op] : descendants of the given classes are not allowed

SNOMED 271636001 | Pulse regular (finding) | [c] SNOMED 61086009 | Pulse irregular (finding) | [c]

- Specific value set: Distinct and selected numbers of SNOMED CT codes allowed
- General value sets: Entire (or multiple) subgraph(s) as possible values allowed
- 37/52 SPHN value sets have SNOMED CT, GENO or SO value set binding

SPHN Semantic Interoperability Framework



Article | Open Access | Published: 10 March 2023

FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network

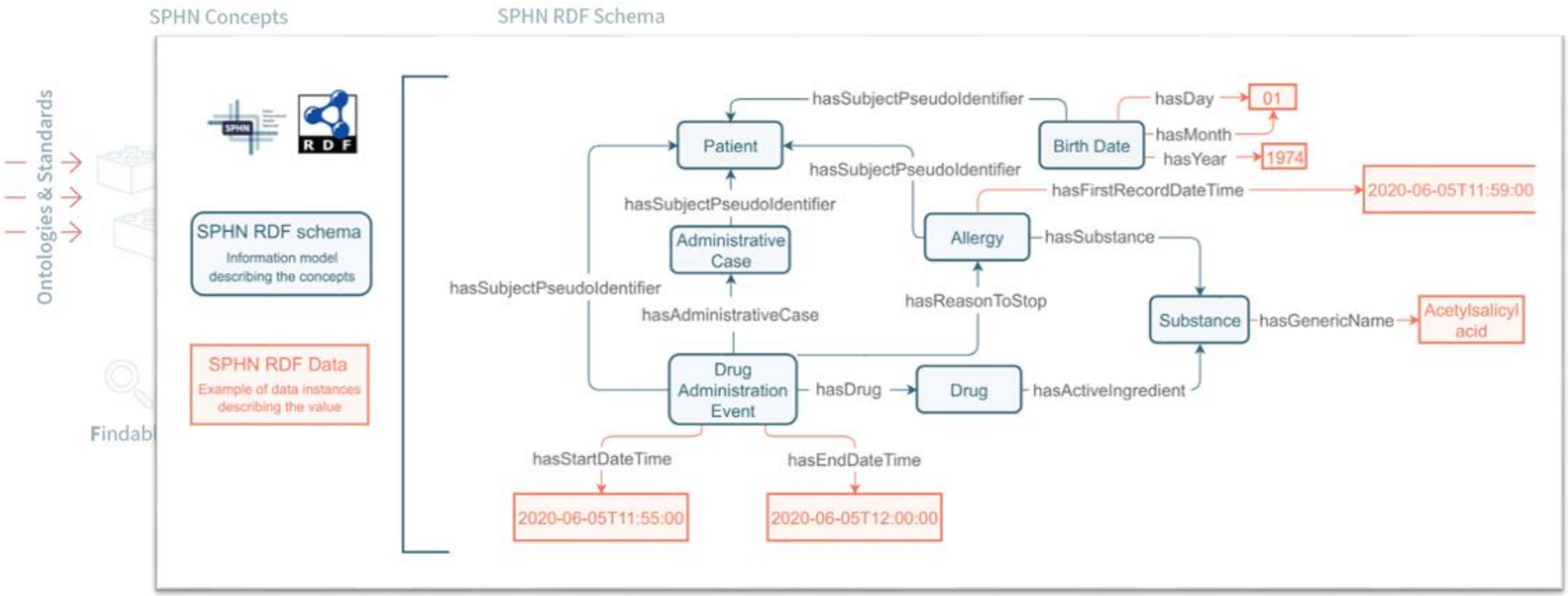
Vasundra Touré, Philip Krauss, Kristin Gnodtke, Jascha Buchhorn, Deepak Unni, Petar Horki, Jean Louis Raisaro, Katie Kalt, Daniel Teixeira, Katrin Crameri & Sabine Österle

Scientific Data **10**, Article number: 127 (2023) | Cite this article

709 Accesses | **9** Altmetric | Metrics

https://www.nature.com/articles/s41597-023-02028-y

SPHN Semantic Interoperability Framework



National Data Streams

IICU

Personalized, data-driven prediction and assessment of infection-related outcomes in Swiss ICUs (IICU)

SPO-NDS

Swiss Personalized Oncology National Data Stream (SPO-NDS)

SwissPed Health

Pediatric personalized research network Switzerland (SwissPedHealth) – a Joint Pediatric National Data Stream

LUCID

LUCID, Low Value of Care in Hospitalized Patients, a National Data Stream on Quality of Care in Swiss university hospitals

SPHN Tool Stack

SPHN Schema Forge

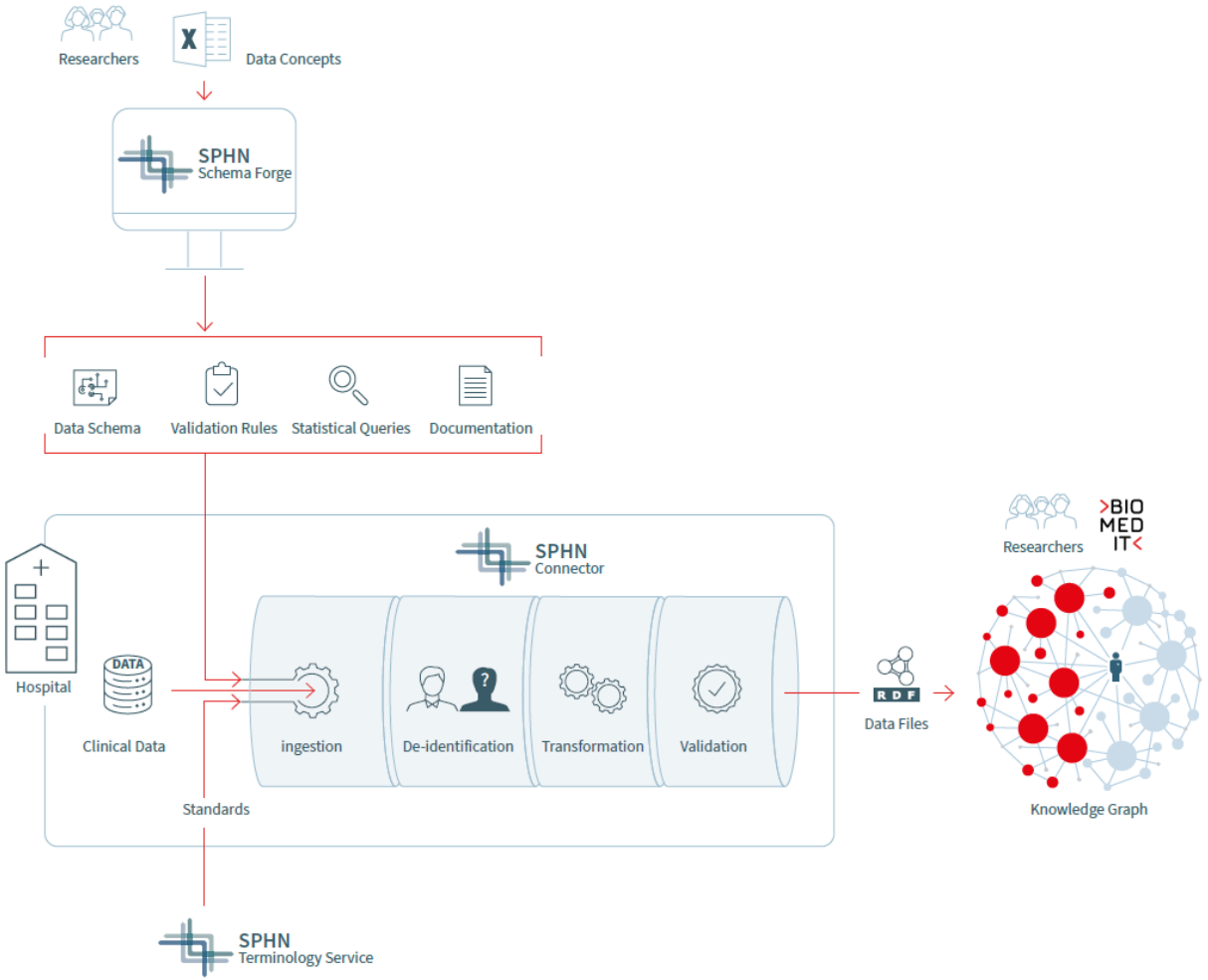
- Automatic generation of semantic artefacts in less than 5 min

SPHN Connector

- Standardized component for data generation pipeline

DCC Terminology Service

- FAIRifyed and historized
- SPHN compliant RDF format



SPHN Tool Stack

SPHN Schema Forge

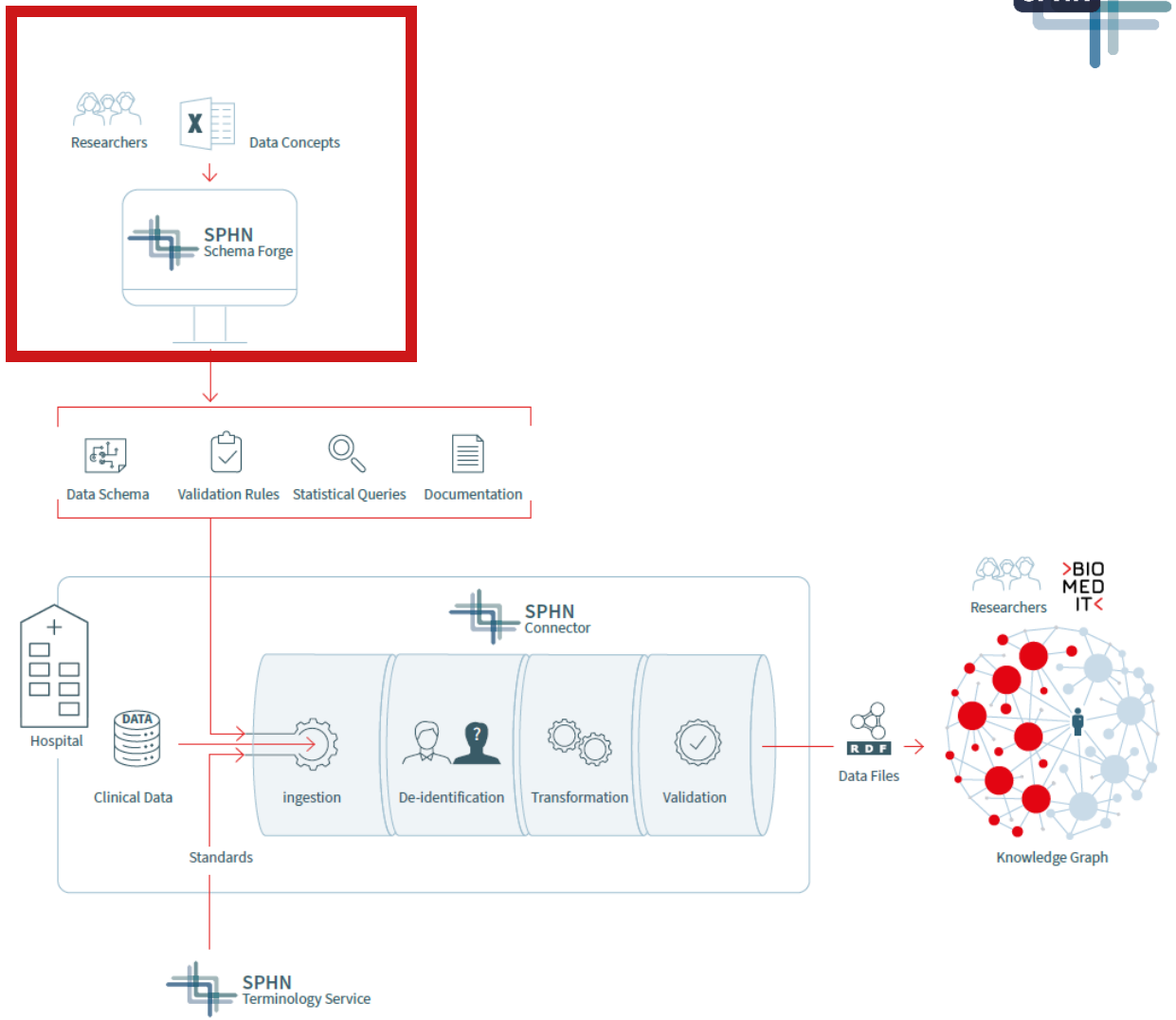
- Automatic generation of semantic artefacts in less than 5 min

SPHN Connector

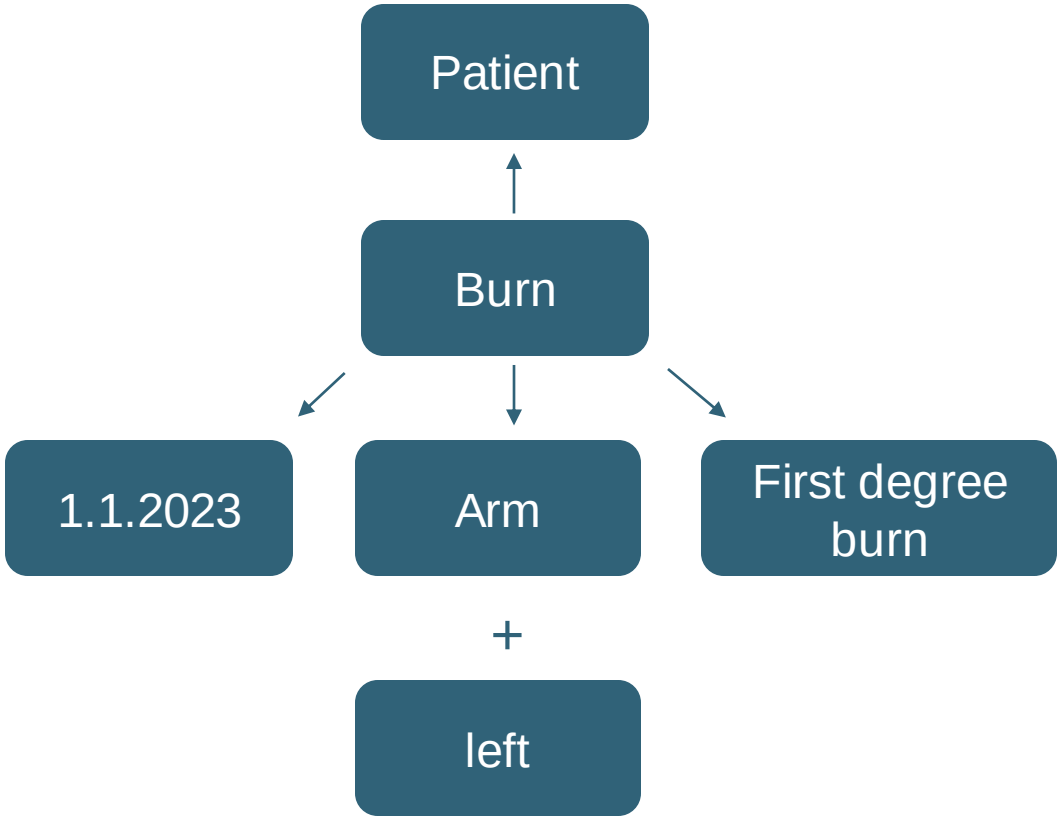
- Standardized component for data generation pipeline

DCC Terminology Service

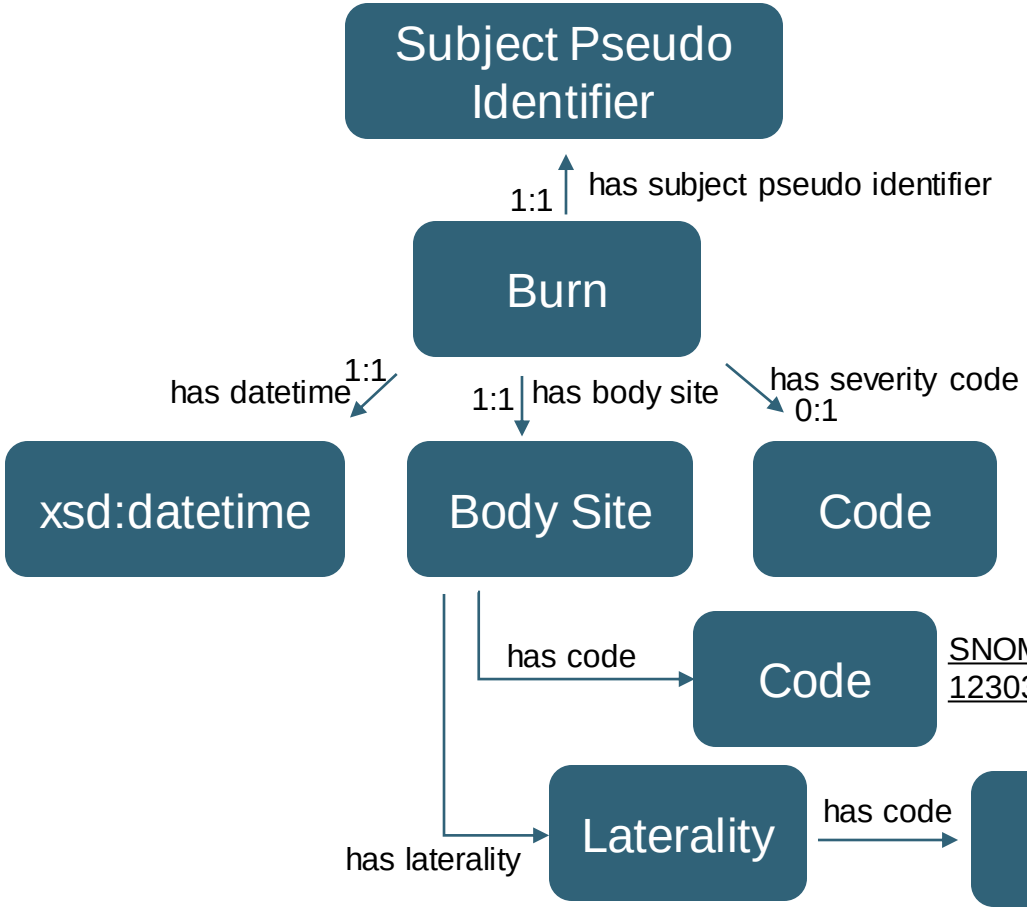
- FAIRifyed and historized
- SPHN compliant RDF format



An Example: Burn event



Build a Concept

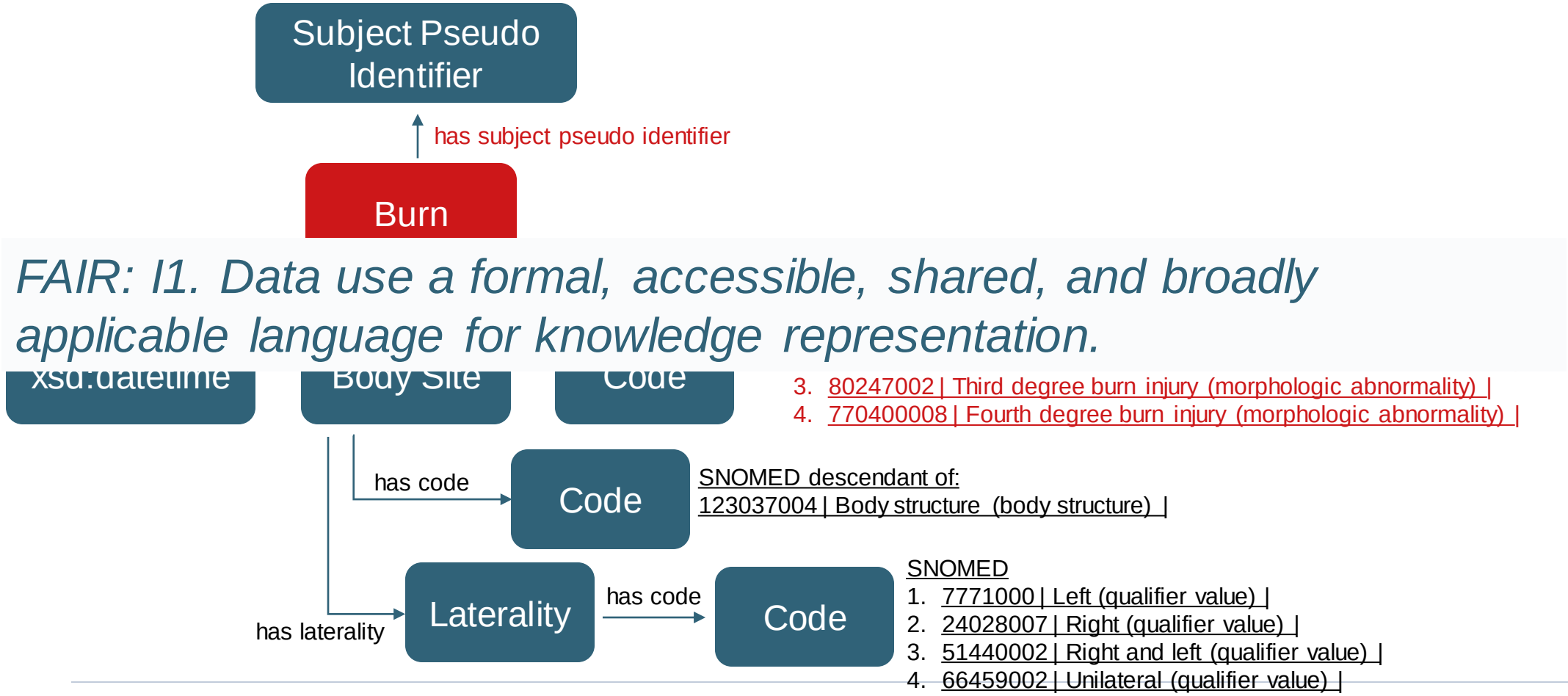


SNOMED descendant of:
123037004 | Body structure (body structure) |

SNOMED
1. 77140003 | First degree burn injury (morphologic abnormality)
2. 46541008 | Second degree burn injury (morphologic abnormality) |
3. 80247002 | Third degree burn injury (morphologic abnormality) |
4. 770400008 | Fourth degree burn injury (morphologic abnormality) |

SNOMED
1. 7771000 | Left (qualifier value) |
2. 24028007 | Right (qualifier value) |
3. 51440002 | Right and left (qualifier value) |
4. 66459002 | Unilateral (qualifier value) |

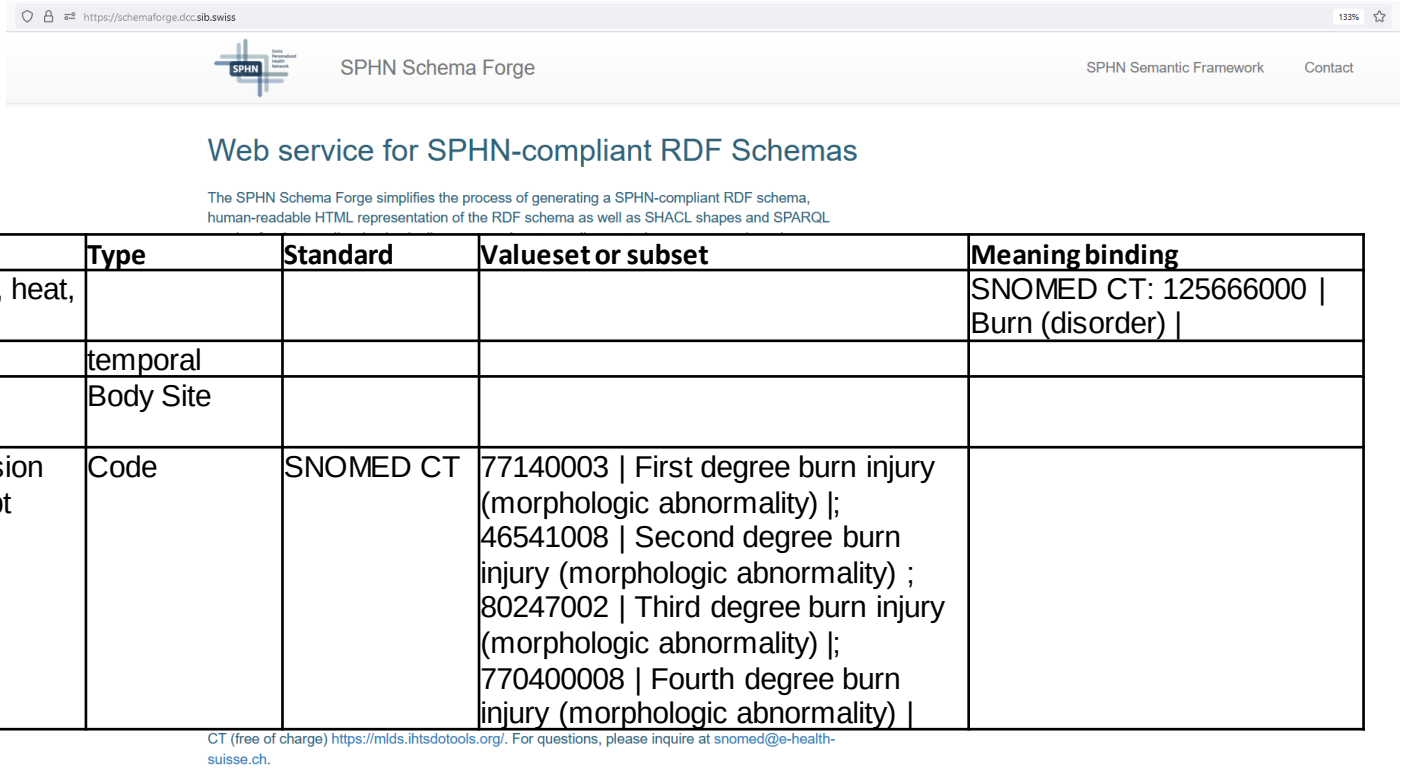
What can be re-used of SPHN



FAIR: I1. Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

SPHN Schema Forge

– SPHN Schema Forge



	Name	Description	Type	Standard	Valueset or subset	Meaning binding
concept	Burn	injure or damage by exposure to fire, heat, or radiation				SNOMED CT: 125666000 | Burn (disorder) |
composedOf	datetime	datetime of the concept	temporal			
composedOf	body site	body site where the concept was measured, performed or collected	Body Site			
composedOf	severity code	code, name, coding system and version describing the severity of the concept	Code	SNOMED CT	77140003 | First degree burn injury (morphologic abnormality) |; 46541008 | Second degree burn injury (morphologic abnormality) ; 80247002 | Third degree burn injury (morphologic abnormality) |; 770400008 | Fourth degree burn injury (morphologic abnormality) |	

RDF

- RDF is the model of how to represent data
- RDF Schema (RDFS) - a W3C standard - is the vocabulary/schema used to define conceptual models that capture knowledge
- **Used RDFS constructs**
  - rdfs:Resource
  - rdfs:Class
  - rdfs:Literal
  - rdfs:Datatype
  - rdfs:label
  - rdfs:comment
  - rdfs:domain
  - rdfs:range
  - rdfs:subClassOf
  - rdfs:subPropertyOf

OWL

- Web Ontology Language OWL
- Provides a catalogue of constructs for expressing relationships between concepts and properties
- **Used OWL constructs**
  - owl:Class
  - owl:EquivalentClass
  - owl:Restriction
  - owl:ObjectProperty
  - owl:DatatypeProperty
  - owl:NamedIndividual
  - owl:someValuesFrom
  - owl:allValuesFrom
  - owl:hasValue
  - owl:minCardinality
  - owl:maxCardinality

RDF snippet

```
sphn:Burn a owl:Class ;
        rdfs:label "Burn" ;
        rdfs:comment "injure or damage by exposure to fire, heat, or radiation" ;
        rdfs:subClassOf [ a owl:Restriction ;
                owl:onProperty sphn:hasSeverityCode ;
                owl:someValuesFrom [ a owl:Class ;
                        owl:unionOf ( snomed:77140003 snomed:46541008 snomed:80247002
                                      snomed:770400008 ) ] ],
            sphn:SPHNConcept ;
        owl:equivalentClass snomed:125666000 ;
        skos:definition "injure or damage by exposure to fire, heat, or radiation" ;
        skos:scopeNote “sphn:hasSeverityCode no subclasses allowed" .
```

*Not complete, excludes cardinalities

Visualization in HTML

Burn [c] # Classes

URI https://www.biomedit.ch/rdf/sphn-ontology/sphn/MIE#Burn

Description injure or damage by exposure to fire, heat, or radiation

Meaning binding (Equivalent-classes) SNOMED 125666000 | Burn (disorder) | [c]

Parents MIEConcept [c]

Property (In the domain of)		Cardinality	Class or Datatype	Restriction
	has data provider institute [op]	1 .. 1	SPHN | Data Provider Institute | [c]	
	has administrative case [op]	0 .. *	SPHN | Administrative Case | [c]	
	has severity code [op]	0 .. 1		Yes
	has datetime [dp]	0 .. 1	xsd:dateTime [c]	
	has subject pseudo identifier [op]	1 .. 1	SPHN | Subject Pseudo Identifier | [c]	
	has body site [op]	0 .. *	SPHN | Body Site | [c]	

Restrictions

▼ has severity code [op] : descendants of the given classes are not allowed

SNOMED 77140003 | First degree burn injury (morphologic abnormality) | [c] SNOMED 46541008 | Second degree burn injury (morphologic abnormality) | [c] SNOMED 80247002 | Third degree burn injury (morphologic abnormality) | [c] SNOMED 770400008 | Fourth degree burn injury (morphologic abnormality) | [c]

https://www.biomedit.ch/rdf/sphn-ontology/sphn/2023/2

Quality control rules SHACL

SHACLer automatic rule generation

Constraints checked

- Validity of classes and properties
- Timeframe validity
- Cardinality constraints
- Restricting on individuals/instances
- Restriction (sequence paths)
- Literal type constraints
- Code validity for versioned ontologies

Example:

```
constraints: sphn:Burn a sh:NodeShape ;
                sh:closed true ;
                sh:ignoredProperties ( rdf:type ) ;
                        [ sh:class sphn:SubjectPseudoIdentifier ;
                                sh:maxCount 1 ;
                                sh:minCount 1 ;
                                sh:path pistoia:hasSubjectPseudoIdentifier ],
                        [ sh:datatype xsd:dateTime ;
                                sh:maxCount 1 ;
                                sh:minCount 0 ;
                                sh:path pistoia:hasDateTime ],
                sh:targetClass sphn:Burn .
```

SPARQL Queries

SPARQLer automatic query generation

Statistics

- Counting instances per concept and predicates
- Minimum and maximum values/dates per predicate
- List and count of all used codes for hasCode

Conversion

- list of resources defined for a concept together with the direct property values

Example:

```
PREFIX sphn: <https://biomedit.ch/rdf/sphn-ontology/sphn#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT *
WHERE {
    {
    SELECT ?origin (MIN(?value) as ?min) (MAX(?value) as ?max)
     WHERE {
          ?resource a <https://www.biomedit.ch/rdf/sphn-
               ontology/sphn/SPHN#Burn> .
          BIND("sphn:hasDateTime" as ?origin)
          optional{ ?resource sphn:hasDateTime ?value . }
    }
   GROUP BY ?origin
}
```

SPHN Schema Forge

Web service to generate the following semantic artifacts within less then 5 min

- **RDF Schema by Dataset2RDF Tool:** Generates an SPHN compliant RDF Schema based on a Dataset input.
- **HTML Website by the Schema Visualization:** Generates a human-readable HTML document describing the project's RDF schema directly from the schema, based on pyLODE.
- **SHACL rules by SHACLer:** Extracts SHACL rules from an SPHN-compliant input ontology for facilitating data validation.
- **SPARQLs by the SPARQLer:** Extracts SPARQL queries for each concept from an SPHN compliant input RDF Schema.

SPHN Tool Stack

SPHN Schema Forge

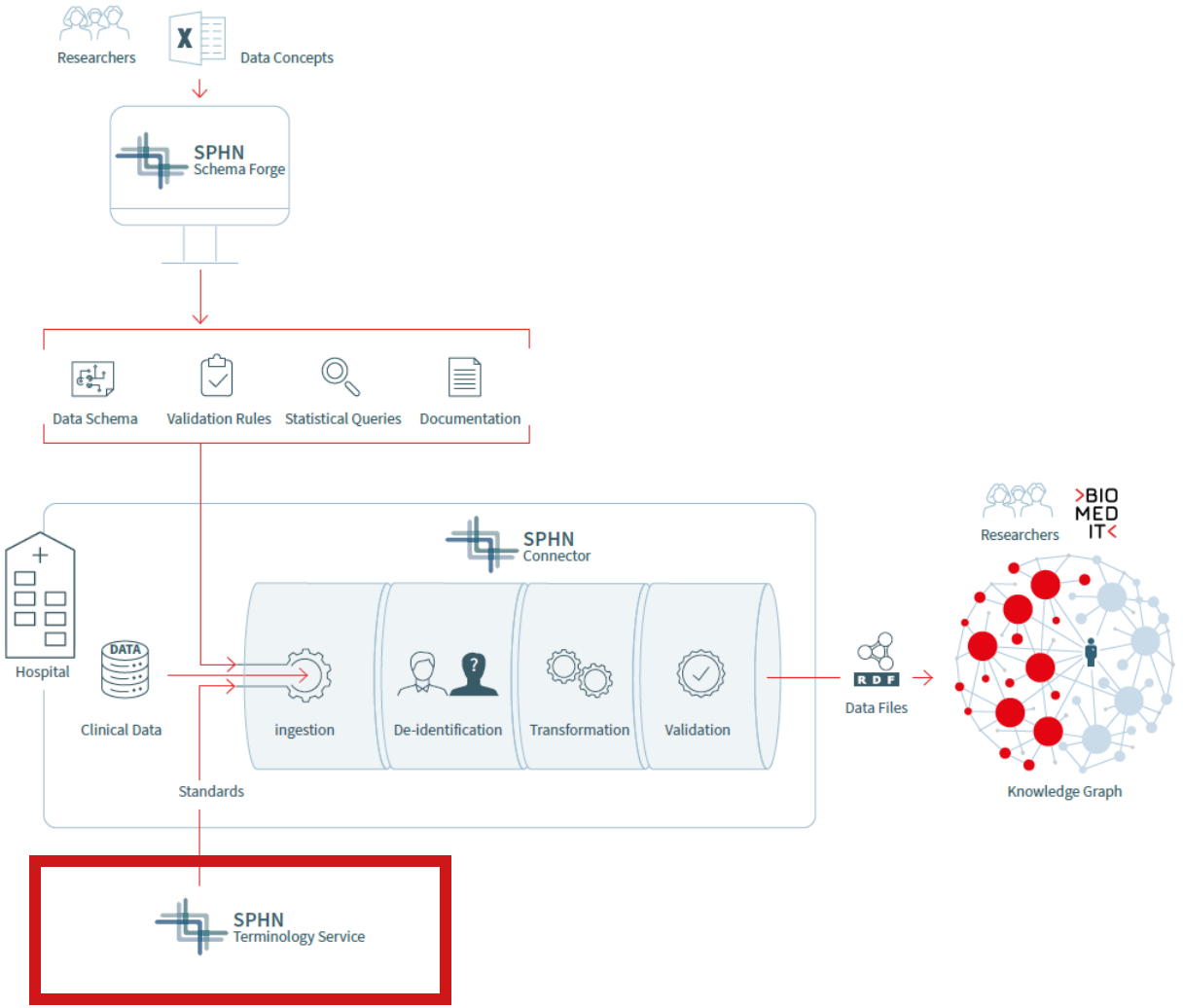
- Automatic generation of semantic artefacts in less than 5 min

DCC Terminology Service

- FAIRifyed and historized
- SPHN compliant RDF format

SPHN Connector

- Standardized component for data generation pipeline



Challenges working with Standards (Terminologies) in the setting of clinical routine data

- Often you can`t choose the standards, since they are used for example for billing or administration

→ Standards are not following the FAIR principles

→ Standards for healthcare and billing are often in local languages

→ Historical data (was) not of interest, so often the terminology version is missing

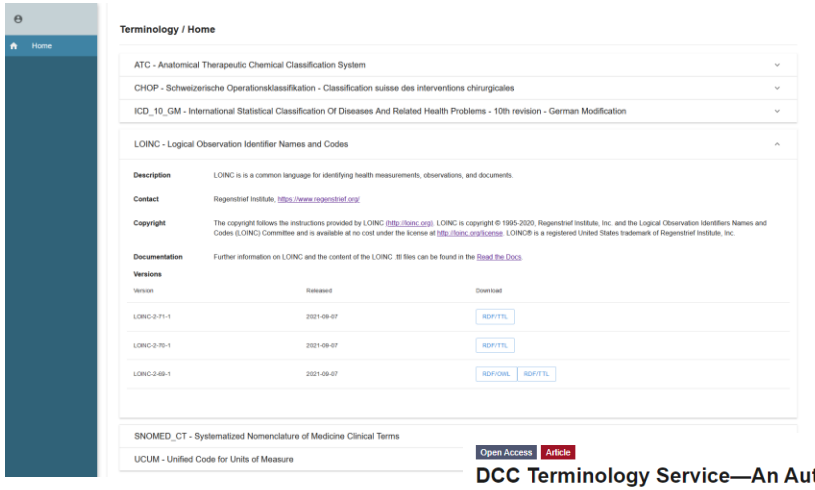
DCC Terminology service

Available Terminologies in SPHN compliant RDF

SNOMED CT, LOINC, ICD-10 GM, CHOP, ATC, UCUM, GENO, HGNC, SO

→ FAIRified

→ Historized (ATC, ICD-10 GM and CHOP)



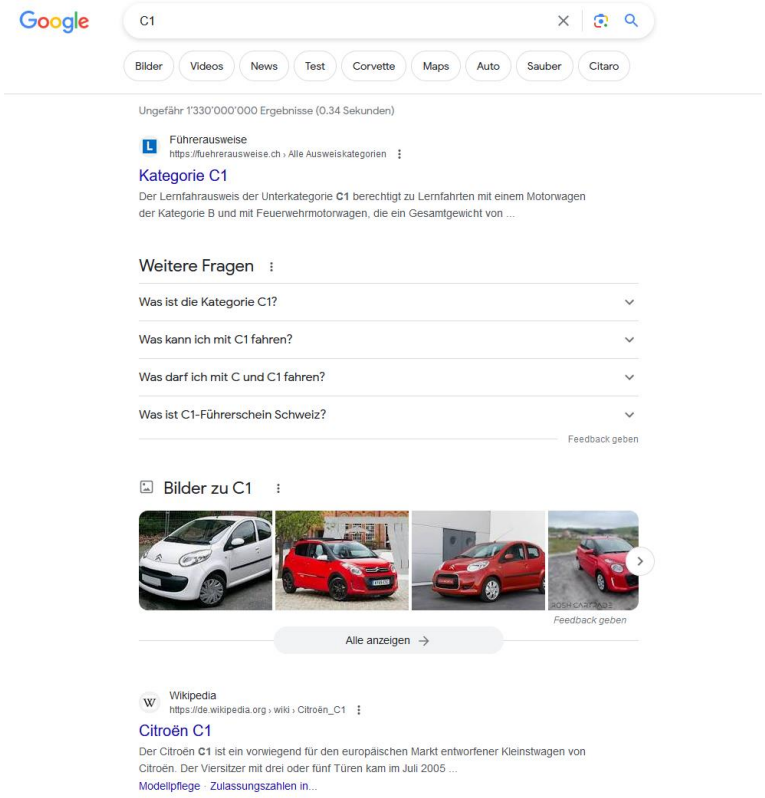
https://www.mdpi.com/2076-3417/11/23/11311

“FAIRification” of external terminologies

Level	Code	Label
1	C1	Operationen am Nervensystem (01–05)
2	Z01	Inzision und Exzision an Schädel, Gehirn und Hirnhäuten
3	Z01.0	Punktion am Schädel
4	Z01.01	Zisternenpunktion
5	Z01.11.0	Detail der Subkategorie 01.11
6	Z01.11.00	Geschlossene [perkutane] [Nadel-] Biopsie an den Hirnhäuten, n.n.bez.

- No unique identifier

Results Google for C1



“FAIRification” of external terminologies

Level	Code	Label
1	C1	Operationen am Nervensystem (01–05)
2	Z01	Inzision und Exzision an Schädel, Gehirn und Hirnhäuten
3	Z01.0	Punktion am Schädel
4	Z01.01	Zisternenpunktion
5	Z01.11.0	Detail der Subkategorie 01.11
6	Z01.11.00	Geschlossene [perkutane] [Nadel-] Biopsie an den Hirnhäuten, n.n.bez.

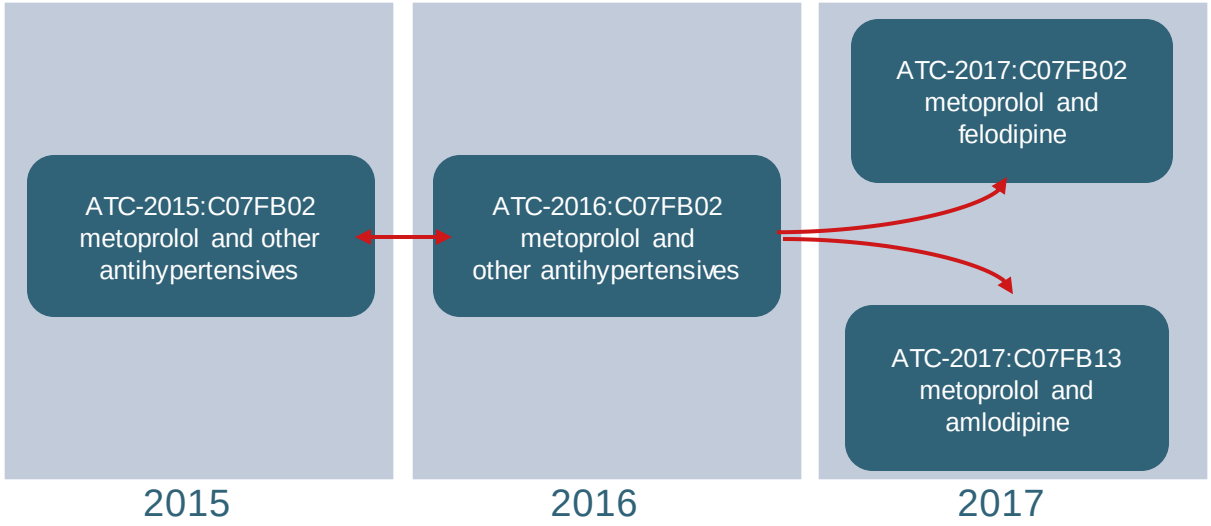
Implementation in SPHN

```
@prefix chop: <https://biomedit.ch/rdf/sphn-resource/chop/> .
chop:01 a rdfs:Class ;
      rdfs:label "Inzision und Exzision an Schädel, Gehirn und
      Hirnhäuten"@de,
      rdfs:subClassOf chop:C1 .
chop:01.0 a rdfs:Class ;
       rdfs:label "Punktion am Schädel"@de,
       rdfs:subClassOf chop:01 .

owl:versionIRI <https://biomedit.ch/rdf/sphn-resource/chop/2016/4>
```

- No unique identifier
- No language for knowledge representation
- Not machine-readable hierarchies
- No metadata in the file itself, version only in the file name

Historization



- Deprecated terms do not appear in the newer version
- Code splitting/reusage

Implementation in SPHN

- Keep all terms and mark as deprecated
- Terms which are the same between versions are marked as the same

→ Used in quality control

SPHN Tool Stack

SPHN Schema Forge

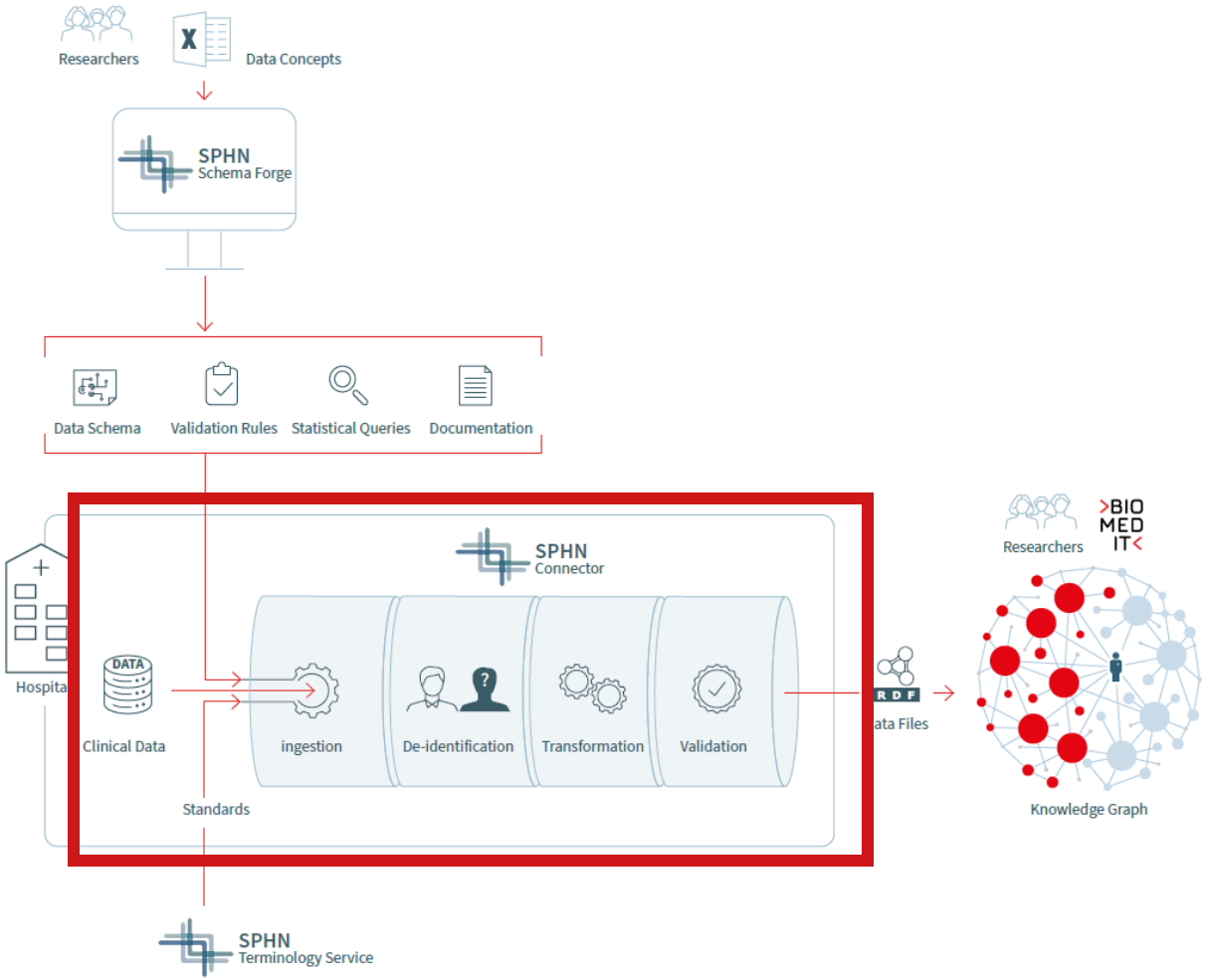
- Automatic generation of semantic artefacts in less than 5 min

SPHN Connector

- Standardized component for data generation pipeline

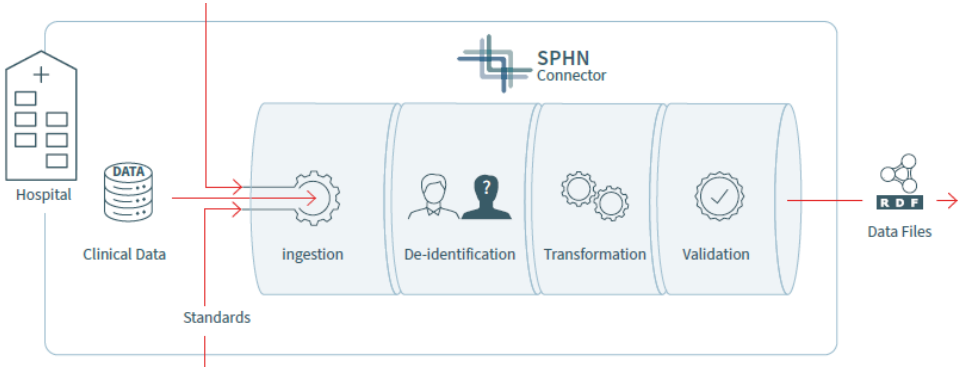
DCC Terminology Service

- FAIRifyed and historized
- SPHN compliant RDF format



SPHN Connector

- The SPHN Connector simplifies the process of connecting new hospitals to the SPHN ecosystem
- It facilitates the creation of valid and semantically rich RDF representations based on relational data in the hospital data platforms
- Automatically generates an ingestion interface from the RDF Schema (tables and json)
- Automatic format and quality checks as well as de-identification can be performed during the transformation process



The work will continue…

SPHN as an initiative will end in 2024, but the work to build and develop the Swiss health-data ecosystem will continue beyond 2025 under the leadership of the SPHN-DCC.

The future SPHN-DCC will be under the responsibility of the Swiss Academy of Medical Sciences.

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

SPHN Initiative
Project funding and administration

SPHN-DCC

SIB
Swiss Institute of Bioinformatics

SPHN Data Coordination Center
Researcher support, tools, services, BioMedIT

2017 - Implementation Phase 1

2021 - Implementation Phase 2

2025 + Consolidation

Take home message

- Everyone has to do his/her part to build a FAIR infrastructure
- Making data FAIR is not only helping others but also yourself
- Standards have to be introduced as early in the data processing chain as possible, best in the source
- Standardized data capturing at the source benefits healthcare and research

Acknowledgements

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

The SPHN NSB and NAB, Task Forces & WGs

The BioMedIT Board and workforces @ ETHZ, Unibas, Unil/SIB

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

The SPHN Management Office: Thos Geiger, Liselotte Selter, Sarah Vermij, Christine Remund

https://bit.ly/3Ja3cW8

https://bit.ly/3mU5fTl

https://bit.ly/31ng7RW

https://bit.ly/34mdNfh

go.nature.com/3Nq4KOe
https://bit.ly/43EEvJK

sabine.oesterle@sib.swiss

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