Swiss Personalized Health Network Schema Forge

Demonstration

Dr. Sabine Österle

Lead Semantic Interoperability Strategy and FAIR Data Team, SPHN Data Coordination Center, SIB Swiss Institute of Bioinformatics

A project of

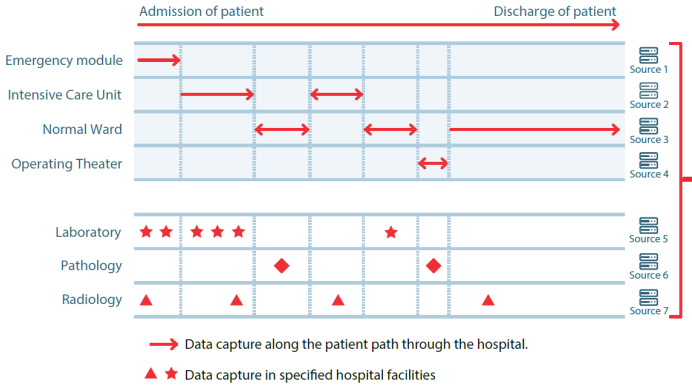
SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences

SIB Swiss Institute of Bioinformatics

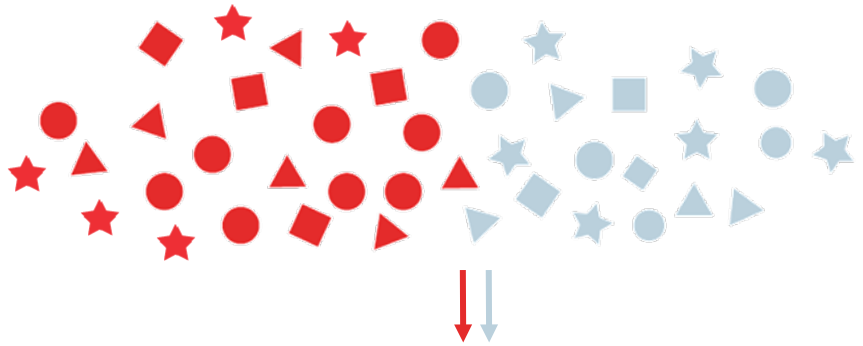
23.05.2023

The need for data interoperability

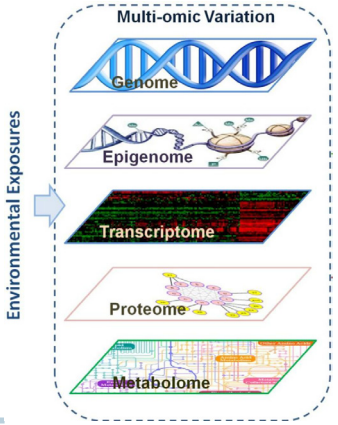
Clinical routine data



Other health-related data



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

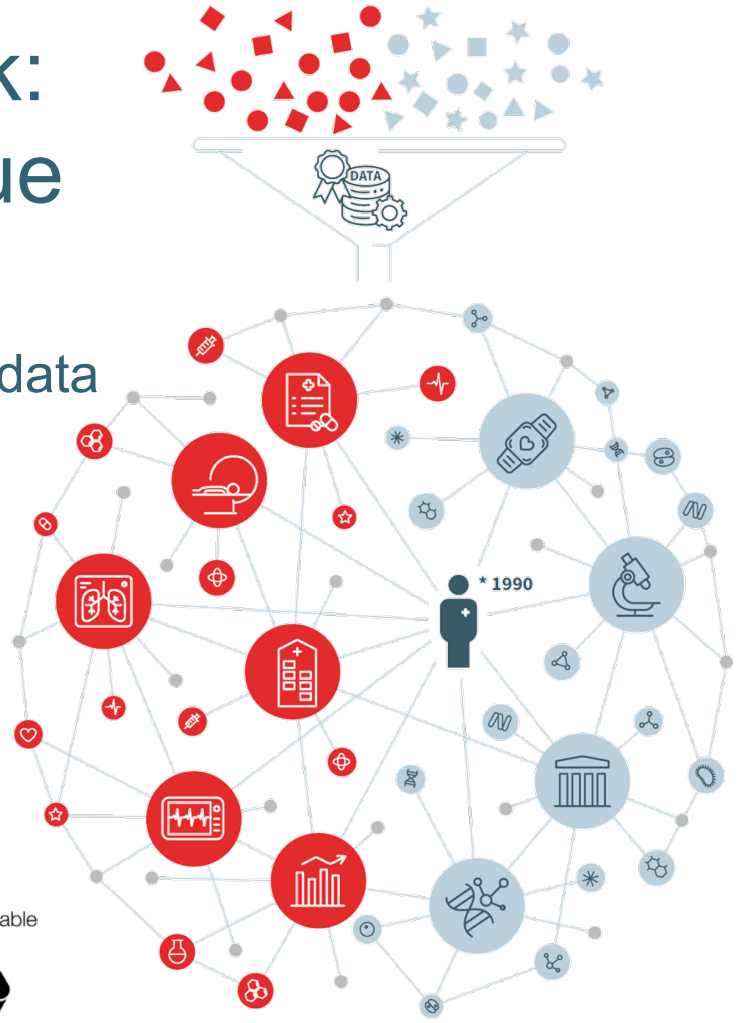


https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5742494/

→ 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

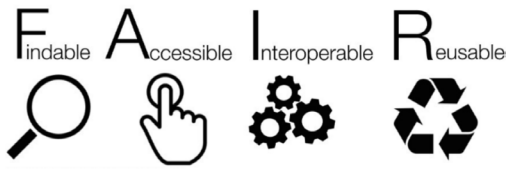


Data semantics
- Use of controlled vocabulary for concepts, valuesets and data
- Data standards: ATC, SNOMED CT, ICD-10-GM, CHOP, UCUM, LOINC, NANDA, ICD-O, GENO, SO...
- Multi-level coding and mappings

Language for knowledge representation
- Semantic Web (RDF Schema)

Quality control framework
- Semantic Web (SHACL, SPARQL)

→ FAIR research data



SPHN Tool Stack

SPHN Schema Forge

– Automatic generation of semantic artefacts in less than 5 min

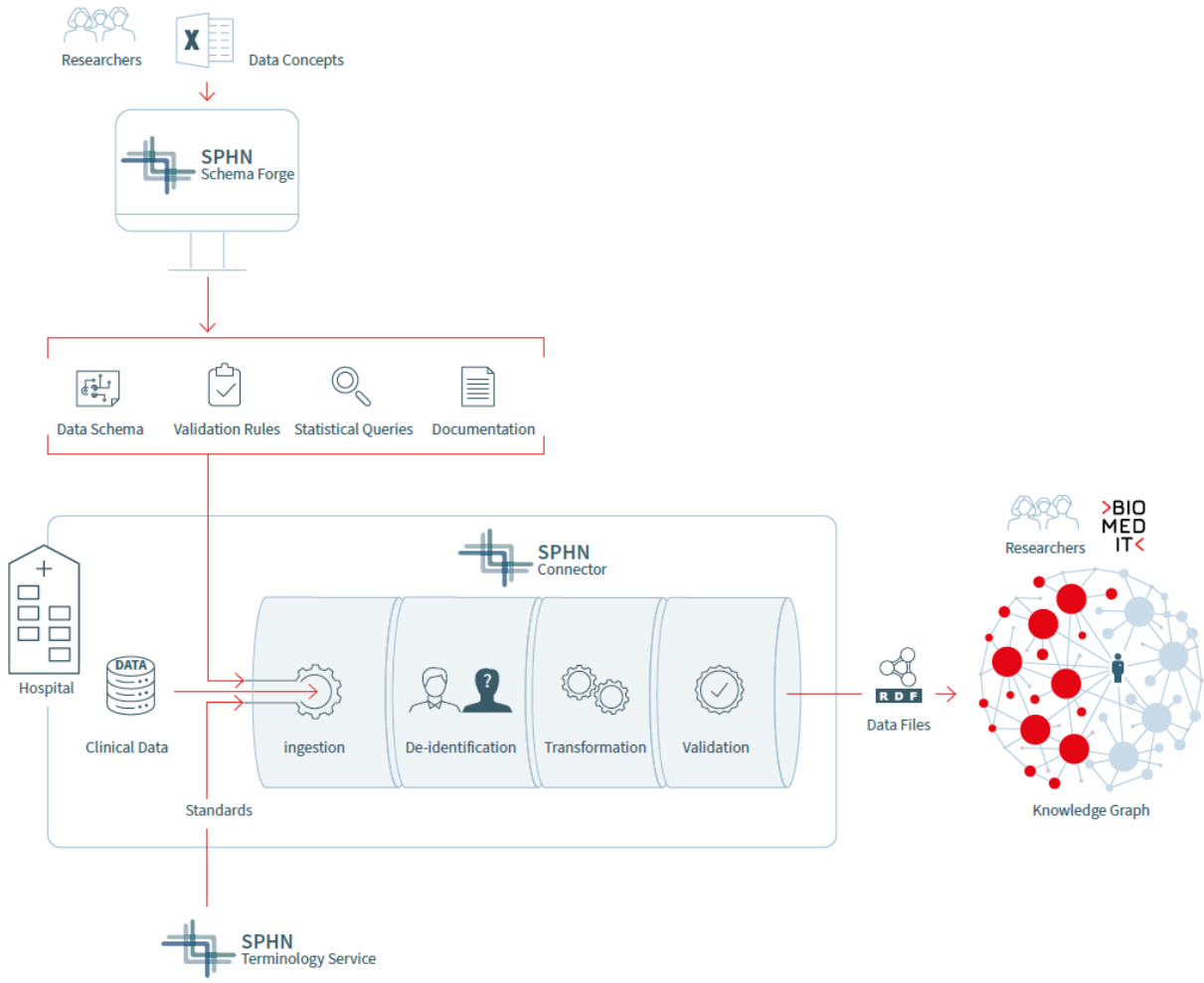
SPHN Connector

– Standardized component for data generation pipeline

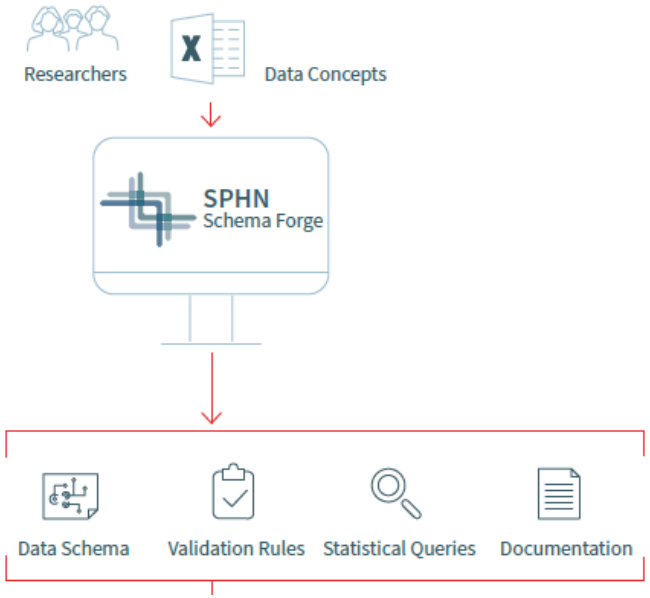
DCC Terminology Service

– FAIRifyed and historized

– SPHN compliant RDF format



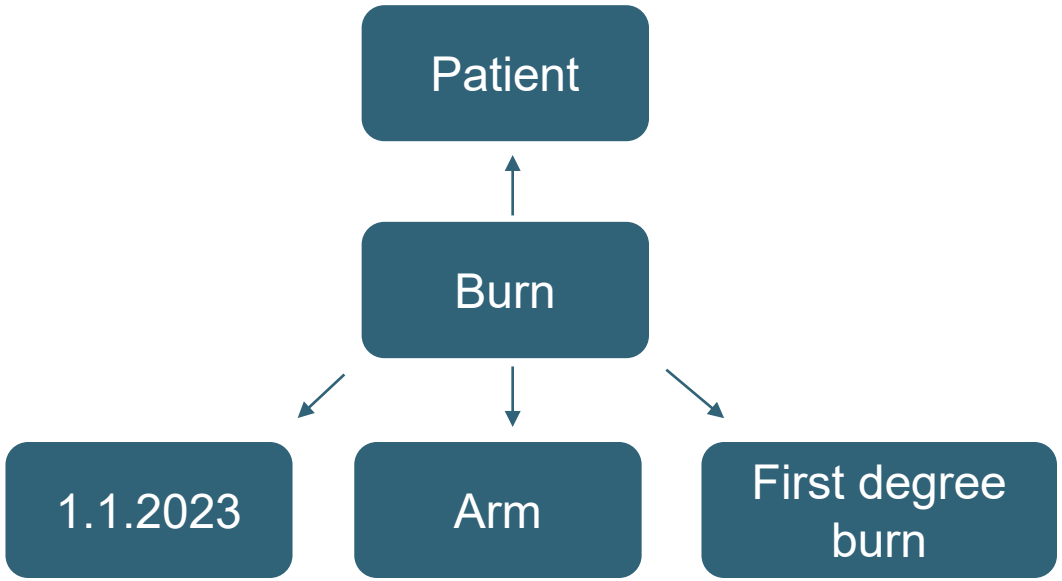
SPHN Schema Forge



Automatic Generation of Semantic Artefact

- RDF Schema
- SHACL rules for validation
- SPARQL queries for statistics
- HTML documentation

Build a Concept



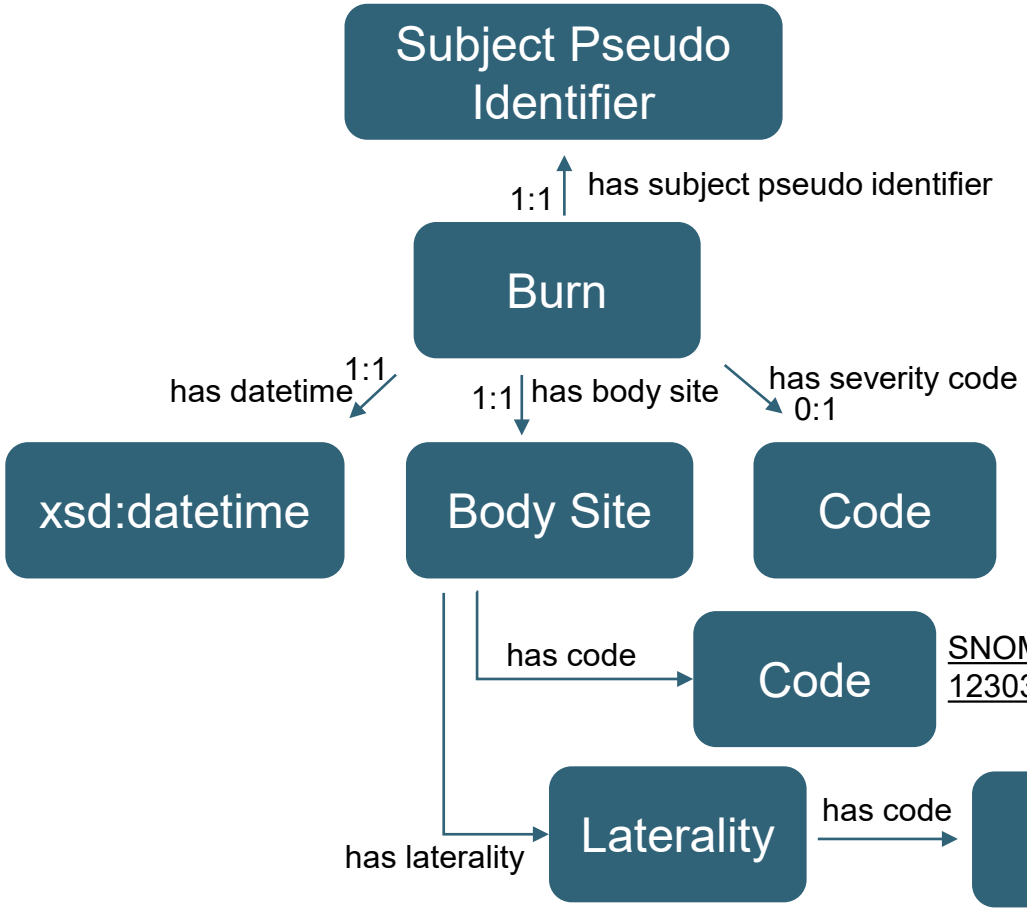
SNOMED
1. 77140003 | First degree burn injury (morphologic abnormality)
2. 46541008 | Second degree burn injury (morphologic abnormality) |
3. 80247002 | Third degree burn injury (morphologic abnormality) |
4. 770400008 | Fourth degree burn injury (morphologic abnormality) |

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

left

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

Build a Concept

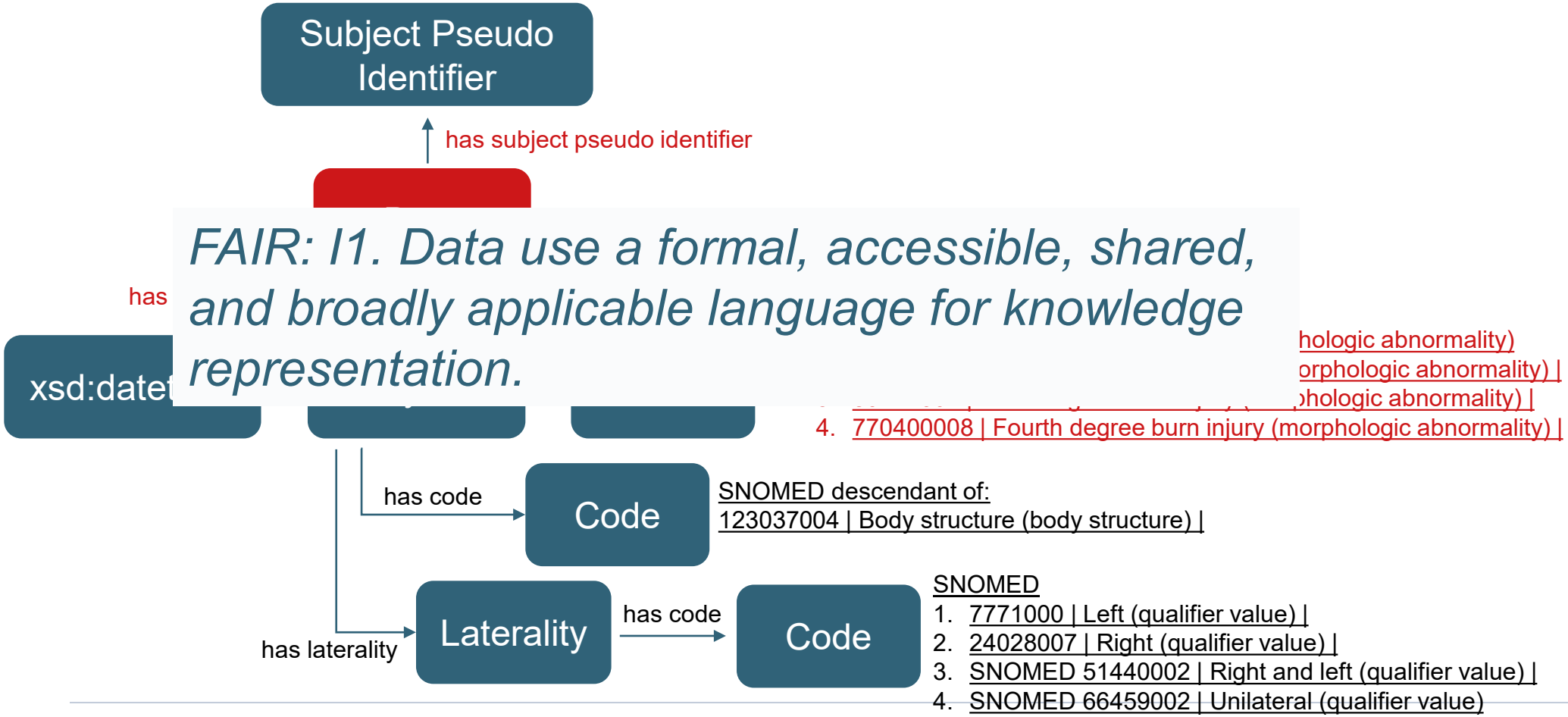


SNOMED descendant of:
123037004 | Body structure (body structure) |

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. SNOMED 51440002 | Right and left (qualifier value) |
4. SNOMED 66459002 | Unilateral (qualifier value)

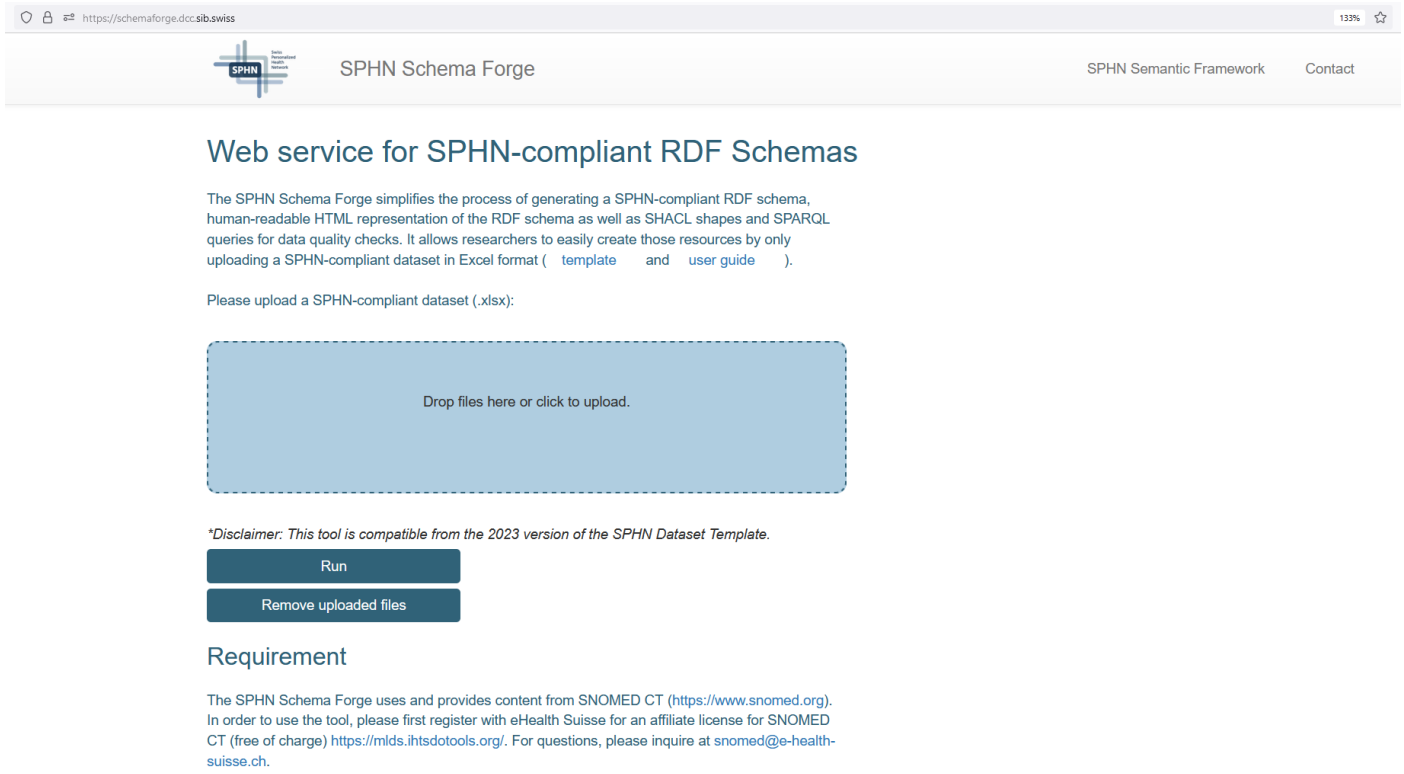
What can be re-used of SPHN



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

Hands on

- SPHN Schema Forge
- SPHN Dataset template
- User guide



Filling the template

	Name	Description	Type	Standard	Valueset or subset	Meaning binding
concept	MIE:Burn	injure or damage by exposure to fire, heat, or radiation				SNOMED CT: 125666000 | Burn (disorder) |
composedOf	MIE:datetime	datetime of the concept	temporal			
composedOf	MIE:body site	body site where the concept was measured, performed or collected	Body Site			
composedOf	MIE: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 on how to represent data
- RDF Schema (RDFS) - a W3C standard - is the vocabulary/schema use to define conceptual models that captures 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 relationship 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

Dataset2RDF

RDF snippet

https://bit.ly/3Odyhfa

```
mie: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 mie:hasSeverityCode ;
                owl:someValuesFrom [ a owl:Class ;
                        owl:unionOf ( snomed:77140003 snomed:46541008 snomed:80247002
                                      snomed:770400008 ) ] ],
                mie:MIEConcept ;
        owl:equivalentClass snomed:125666000 ;
        skos:definition "injure or damage by exposure to fire, heat, or radiation" ;
        skos:scopeNote "mie: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

SHACL

- Shape Constraint Language
- A W3C standard for validating RDF data against a set of constraints defined as shapes
- The constraints are generated based on semantics
- Supports complex rule-based validation
- Supports inference

SHACLer

SHACL generation

Constrains 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

SHACL snippet

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

SPARQL

- SPARQL Protocol And RDF Query Language
- A W3C standard for querying (and manipulating) data that follows the RDF specification
- Supports inference
- A database that supports storing and querying of RDF data is called Triplestore

SPARQLer

SPARQL 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

SAPRQL snippet

```
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/MIE#Burn> .
          BIND("sphn:hasDateTime" as ?origin)
          optional{ ?resource sphn:hasDateTime ?value . }
    }
   GROUP BY ?origin
}
```

Outlook

Release candidate for 2024.1 will be published for feedback in the coming days

Including additional:

- genomic concepts
- intensive care unit concepts
- provenance concepts
- demographics concepts

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

SIB
Swiss Institute of Bioinformatics

SPHN Initiative
Project funding and administration

SPHN-DCC

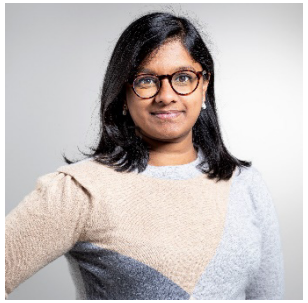
SPHN Data Coordination Center
Researcher support, tools, services, BioMedIT

2017 - Implementation Phase 1

2021 - Implementation Phase 2

2025 + Consolidation

Special thanks go to



The **Vasundra Touré and Deepak Unni** from the **Data Interoperability Team** of the SIB Personalized Health Informatics Group

And our collaborators from Trivadis Part of Accenture: **Philipp Krauss, Leon Hinderling, Jascha Buchhorn**

– SPHN Working Groups on Semantics and RDF with representatives of the five university hospitals

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