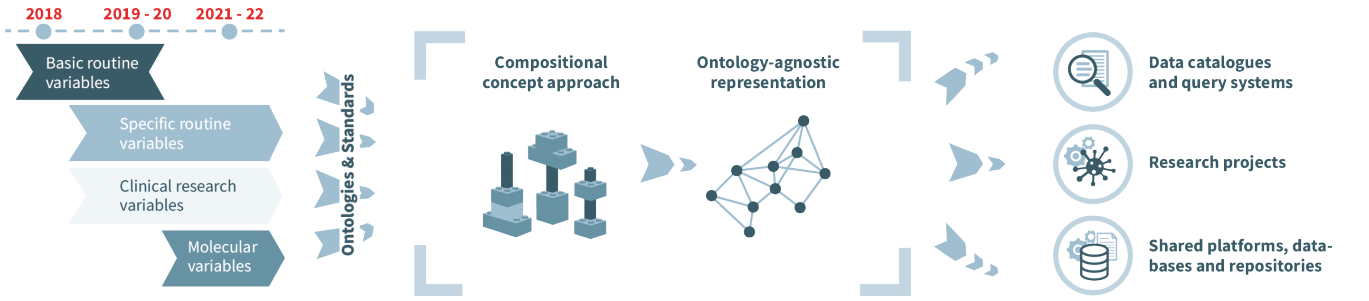
2021 Fact sheet
The SPHN Semantic Interoperability Framework

1 Summary

Health-related data, ranging from hospital routine data to biomedical research data, are quite diverse, stored in different databases and formats, and often coded in local standards. This fragmentation and diversity make it very time-consuming to combine data from different sources for doing research on a particular topic. It is generally difficult to understand data and the intended meaning due to the lack of common standards, metadata, or a common data dictionary. Following the FAIR (findable, accessible, interoperable and reusable) principles, SPHN builds an infrastructure to overcome these hurdles and enables collaborative research by making the meaning of health-related data understandable to humans and machines. This allows for an easy combination of the data from different sources, thus simplifying the use and exploration of data across Switzerland. Experience shows that the diverse needs and requirements of the different use cases make it impossible to agree on one single data model. Not only the required data elements differ between projects, but also the level of information granularity required for a single datapoint can vary significantly. Therefore, SPHN developed a framework based on a strong semantic layer of information (pillar 1), and graph technologies for the exchange layer, which can be extended by the individual projects to fit their purposes (pillar 2). Thus, a universal exchange language for healthcare is established, using the "words" from various international standard vocabularies (such as SNOMED CT or LOINC), a simple "grammar" (subject-predicate-object; expressed in RDF), and additional SPHN guidelines and rules to establish good practices for FAIR data.

The SPHN Semantic Interoperability Framework

A collaborative, national endeavour coordinated by the SPHN Data Coordination Centre, with the expert know-how and contribution of the SPHN community, the Clinical Data Semantic Interoperability- and the Hospital IT working group.



Project needs

Variables and associated information needs are collected from the SPHN Projects. In a close dialogue with all stakeholders a global understanding for a data-driven representation is developed. Individual research needs are aligned across projects and translated into a common semantic representation.

01

Semantic representation

Strong semantic definitions are concept-ualized in building blocks, so-called "concepts". Concept definitions take into account existing medical ontologies and standards. In a modular way, concepts can be reused and combined into more comprehensive concepts.

02

Data transport and storage

A flexible data exchange format, able to carry the meaning of data, is recom-mended for data transport and storage.

This format is capable to support various data-driven research scenarios in personalized health.

03

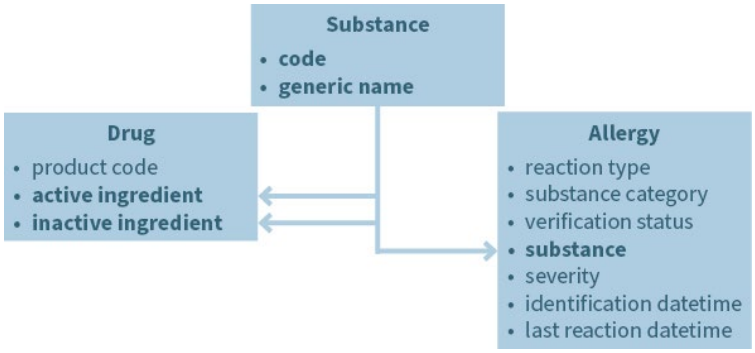
Use cases

Diverse use cases and applications in healthcare, biomedical research and data science can be served with a flexible, multi-dimensional semantic data architecture, which makes it easy to transform data into desired, use case specific data models.

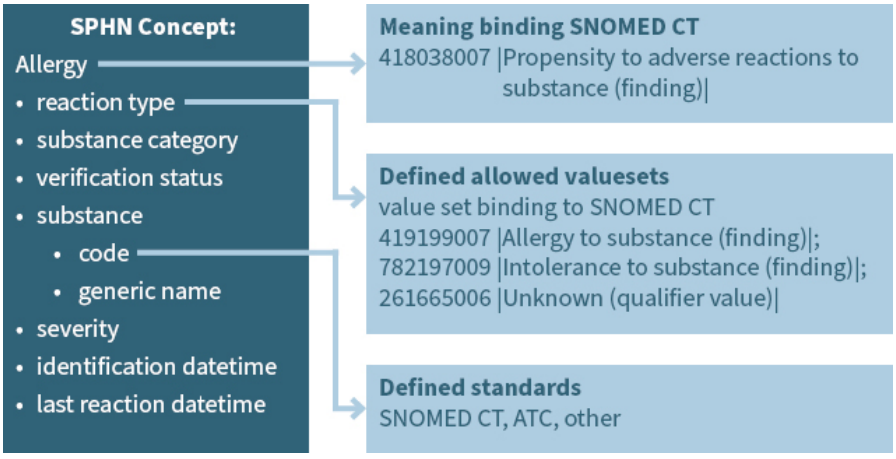
2 How it is implemented

Semantic representation (Pillar 1)

SPHN concepts are generalizable building blocks, which can be used in different contexts. Each concept contains all information necessary to understand it, and concepts can be combined to composed concepts, which again can be combined to more complex compositions. It is important to find the right level between abstraction and granularity to optimize the power of expression. The approach can be illustrated with the example of “substance”. A substance can be an active or inactive ingredient of a drug or it can be the substance someone is allergic to. Therefore, we can abstract “substance” as a concept on its own. The concept of substance is composed of two concepts: “code” and “generic name”. These concepts describe a substance no matter if it is the active ingredient of a drug (that is defined as the type “substance“) or the substance to which someone is allergic to.



To make the SPHN concepts comparable nationally and internationally, we express their meaning using existing semantic standards (controlled vocabularies), by creating a meaning binding wherever possible to SNOMED CT and/or LOINC. The data element of a concept can be expressed using one or several recommended standards (e.g. LOINC, SNOMED-CT, ICD-10, ICD-O-3, CHOP, ATC). For example, the instance of substance code under the concept “allergy” can be an ATC code, a SNOMED CT code, or a code form another semantic standard. If needed value sets are defined and if possible, a value set binding to SNOMED CT is added. Descriptions for concepts and value sets as well as standards are, whenever possible, aligned with national and international sources.



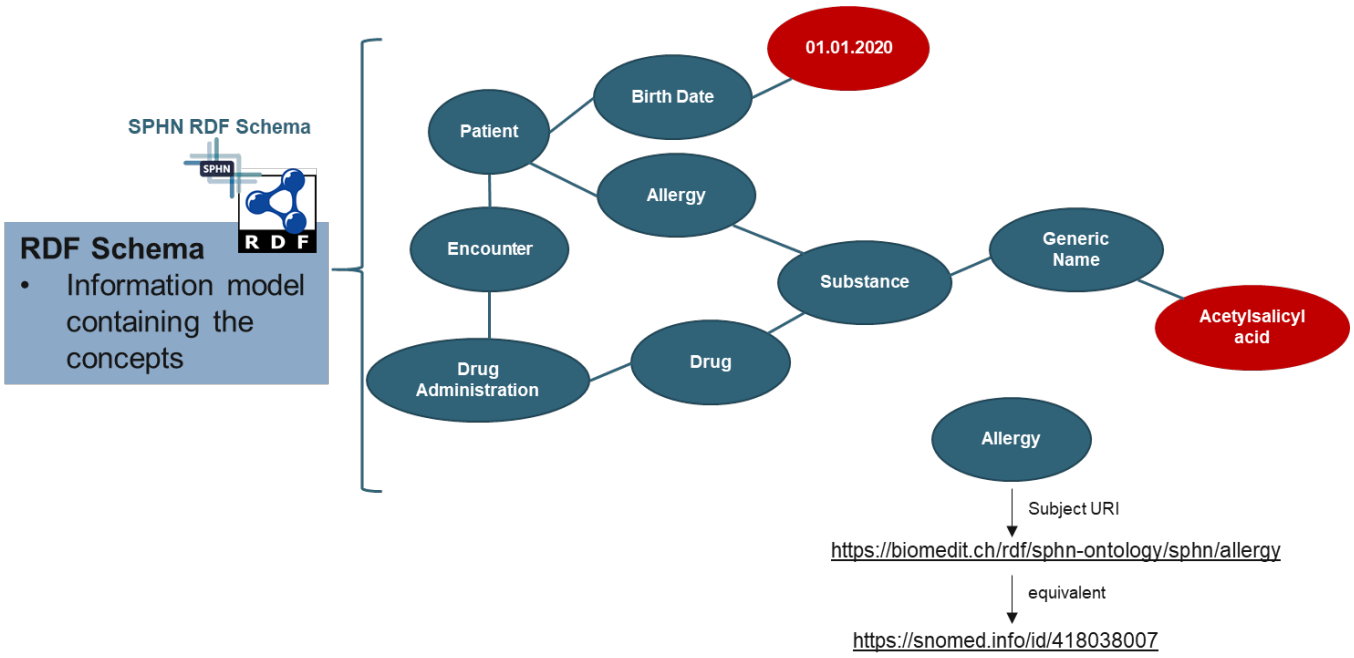
Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

Data transport and storage (Pillar 2)

SPHN concepts (blue) and their instances (red) can easily be mapped from/to other data representations or merged with other RDF data sets without losing their semantics.

The relations between different concepts and the data are expressed in the form of triples composed of a “subject”, a “predicate”, and an “object”. In the example, the RDF triple indicates that the “Patient” (subject) “has” (predicate) “Birth Date” (object), and “Birth Date” (now, subject) “has a datetime” (predicate) which is “01.01.2020” (object). Since RDF does not depend on a specific semantic standard, it allows the use of different semantic standards, and value sets as defined in the SPHN Dataset.

In RDF subjects, predicates and objects have a Unique Resource Identifier (URI) that enables the unique identification of these elements. In our example, the concept Allergy has the URI https://biomedit.ch/rdf/sphn-ontology/sphn/allergy, which uniquely and unambiguously identifies it in the context of SPHN. The SNOMED CT meaning binding is introduced by linking this URI to the corresponding URI in SNOMED CT in the SPHN RDF schema.



Use cases (Pillar 3)

Based on these needs, researchers can:

1. use the RDF files directly as input into their analysis software, e.g. rdflib in Python,
2. extract data into a flat file, e.g. Excel or CSV
3. load the data into other existing data models with adequate converters, e.g. i2b2, OMOP or a data management software

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

3 Selected SPHN Semantic Web Services

- **DCC WebProtégé** provides an easy-to-use, web-based ontology development environment to collaborate and share with others..
- **DCC Terminology Service provides** SPHN compatible, machine-readable versions of national (CHOP or ICD-10 GM) and international (SNOMED CT, LOINC, ATC, UCUM) terminologies and classifications in RDF format.
- **SPHN Quality Assurance Framework** contains a set of SHACL rules and statistical SPARQL queries to validate the compliance of the RDF data produced with the SPHN (and project-specific) ontologies.

4 FAIR principles

How does this strategy help to address the following FAIR criteria:

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
  --> in RDF, each concept, instance, and relation have a URI
- F2. Data are described with rich metadata
  --> each SPHN concept contains a set of information to describe the data element

Accessible

- A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
  --> SPARQL is the official RDF Query Language which is a standard protocol of the World Wide Web Consortium (W3C). It facilitates the exploration of data in RDF

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
  --> RDF is the standard language used for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
  --> SNOMED CT and LOINC along with other standards are used as vocabulary in the SPHN RDF Schema
- I3. (Meta)data include qualified references to other (meta)data
  --> when possible, the SPHN URIs of concepts and instances are linked to the URIs of the external resources like SNOMED CT, LOINC, ATC, ICD-10.

Find out more on: https://sphn.ch/network/data-coordination-center/the-sphn-semantic-interoperability-framework/

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland