SPHN Working Group on Semantics

Version and date: Version 1.0, 30 August 2022

Starting Position

The Swiss Personalized Health Network (SPHN) promotes infrastructures that enable the interoperable use of health data for multi-site research projects in Switzerland. In the first funding phase, the SPHN Data Coordination Center (DCC), the SPHN Clinical data semantic interoperability Working Group and the SPHN partners developed the "Three pillar strategy" for SPHN[1,2]:

1. Strong semantic layer that leverages interpretability of health-related data by supporting data/object composition;
2. A data model-agnostic transport and storage method to achieve a highest possible degree of flexibility for data use, the Resource Description Framework (RDF);
3. Data models according to the researchers' needs (such as OMOP, CDISC, i2b2).

To date, the DCC and its partners have developed and organized the following for the implementation of the first pillar:
- The SPHN Dataset[3] (yearly released) containing semantic definition of (core) data elements:
  - 63 SPHN concepts (general building blocks):
    - Basic concepts, which can be reused in different contexts, e.g., Quantity, Body site or Code;
    - Concepts of general interest, e.g., Administrative gender, Heart rate, Diagnosis or Procedure;
    - Specialized concepts from, e.g., intensive care or oncology.
  - including 16 different nomenclatures, classifications and ontologies used in SPHN concepts;
  - 47 SPHN concepts have a meaning binding to internationally controlled vocabulary (e.g., SNOMED CT or LOINC);
  - 37 of the total 53 value sets are coded in SNOMED CT;
- A documentation and user guide[4]
- An SPHN training on semantic standards[5]
- Two SPHN webinars: LOINC in Italy and Introduction to ICD-11

The SPHN Dataset forms the semantic basis for the technical data specifications of SPHN projects and services, such as the SPHN data exchange format specification, namely the SPHN RDF schema, and the SPHN Federated Query System (FQS).

In order to support the new National Data Streams (NDS) in the use of the SPHN concepts, the following challenges and gaps have been identified and need to be tackled:
- Data representation: Gap between the desired, data-driven representation of research data and what is captured in the source system or software solution during healthcare processes
- Missing concepts needed by the NDS and other SPHN projects:
  - Concepts for representing genomics, microbiology, cost information, PROMS as well as further ICU and oncology data
  - Meta-data concepts for multi-modal data enabling linking raw data files to clinical data

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

SIB | Swiss Institute of Bioinformatics
PHI | Personalized Health Informatics Group
www.sphn.ch | dcc@sib.swiss

– Data provenance and quality indicators: when integrating data from different sources, information about where the data comes from and how it was generated is very important
– Missing expertise:
– Standards, such as SNOMED CT, allow different ways of annotating data and the most suitable design of concepts and value sets can only be developed jointly by terminology and domain experts.

Vision and Mission

The ultimate goal of SPHN and its interoperability framework is to make health data accessible in a standardized and coded format and usable for data-driven, personalized health research. The Semantic Working Group therefore aims to facilitate the development of SPHN concepts in line with the recommendations of the SPHN Semantics Strategy, according to the following main guiding principles:
– data elements should be separated in their meanings
– semantic standards should be used defining concepts and (meta)data (FAIR principle I2);
– rich metadata should be linked (FAIR principle F2);
– aggregated and calculated values should be avoided or presented together with their source values.
SPHN data should be flexible to assemble and aggregate for data analysis, e.g., to derive features as input for machine learning algorithms. Such a detailed data representation is especially needed for future research projects to allow the full flexibility in data analysis.

Tasks and Timeline of the Working Group

The term of the WG is planned until the end of the second funding period of SPHN (2021-2024).

Tasks:

[1] Support the NDS and other SPHN projects on the use of SPHN concepts and the design of their project specific concepts by organizing workshops and trainings for concept development;

[2] Identify standardization needs and take responsibility for developing appropriate new concepts;

[3] Organize domain expert reviews for concepts and change requests as needed to ensure that concepts published in the SPHN Dataset are well structured and include appropriated semantic standards. Concepts should always be developed together with subject matter experts (SME) or reviewed by SME;

[4] Review new concepts and change requests along the SPHN Semantic Framework to ensure that the concepts comply with the guiding principles for concept design.

Deliverable 1: Yearly release of the SPHN Dataset

The content of each release will be discussed and prioritized by the working group. For the 2023 release, the focus lies on concepts needed by the NDS and those already in development and/or used in Driver Projects, and come from the following specialties: genomics, microbiology, imaging, cardiology, intensive care, oncology. New concepts from NDSs can be included when finalized.

Deliverable 2: Documentation, user guide and training

New concepts and change requests need to be documented to ensure traceability and understandability. The working group supports the DCC and project partners in the development and review these documents. Further, a comprehensive user guide and trainings are offered to support researchers in the design of new concepts and the use of semantic standards in general.

Deliverable 3: Workshop for concept development
To facilitate concept development and the exchange with SME, especially from the NDS, several "Concept development" workshop will be organized. This also includes the exchange with international experts as needed. The working group supports the DCC in the content preparation and prioritization as well as moderation of the workshops.

Composition of the Working Group

The University Hospitals nominated the following representatives:

- Insel: Alexander Leichtle
- HUG: Deniz Geres
- CHUV: Pierre Chodanowski, Daniel Damian
- USZ: Katie Kalt
- USB: Amanda Herbrand, Benjamin Kasenda

Kristin Gnodtke, Personalized Health Informatics (PHI) group, SIB Swiss Institute of Bioinformatics chairs the working group. The Semantic WG will work closely together with the RDF WG and the Data Standards and Data Quality (DASAQ) WG. The Semantic WG representatives are responsible for discussing and coordinating the feedback for new concepts and change request within their university hospital. Further, they are the internal contact persons with regards to semantic standards and the SPHN Dataset within their university hospital.

---

[1] https://sphn.ch/document/csi_wg_strategy/

[2] Gaudet-Blavignac, C.; Raisaro, J.L.; Touré, V.; Österle, S.; Crameri, K.; Christian Lovis, L. A National, Semantic-Driven, Three-Pillar Strategy to Enable Health Data Secondary Usage Interoperability for Research Within the Swiss Personalized Health Network: Methodological Study (doi:10.2196/27591).

[3] https://sphn.ch/document/sphn-dataset/

[4] https://sphn-semantic-framework.readthedocs.io/en/latest/sphn_framework/sphndataset.html

[5] https://sphn.ch/training/semantic-standards-training/