Swiss Personalized Health Network: Making health data FAIR in Switzerland

Deepak Unni, Sabine Österle, Katrin Crameri

on behalf of the Data Interoperability Team of the Personalized Health Informatics Group

SIB Swiss Institute of Bioinformatics, Basel, Switzerland

The Swiss Personalized Health Network (SPHN) is a national initiative responsible for the development, implementation, and validation of coordinated data infrastructures that make health-related data FAIR (Findable, Accessible, Interoperable, Reusable) and available for research in Switzerland in a legally and ethically compliant manner. SPHN brings together stakeholders from various university hospitals and research institutions across Switzerland to enable the secondary use of health-related data, including but not limited to clinical routine data, omics and cohort data for personalized health research. To that end, the SPHN initiative consists of several key elements. Firstly, SPHN provides an Semantic Interoperability Framework for the definition and harmonization of health data semantics. Secondly, SPHN provides an ecosystem that supports the generation, quality check, dissemination, and analysis of health data. The system enables the translation of health data into RDF, provides a terminology service for access to external terminologies in RDF, and a schema template to support projects to create and work with their individual subset of concepts and attributes. Lastly, SPHN also provides services that improve the discoverability of data, and of metadata, enabling researchers to explore and identify cohorts of interest and request access to the data. Since 2017, SPHN has contributed to the establishment of clinical data management platforms at the five Swiss University Hospitals to make health data efficiently available for research, has supported researchers in the discovery and analysis of data, and built the required infrastructures to bridge the gap between healthcare and data-driven research.

Data

Data - such as routine clinical data - are provided by the University Hospitals. Data from several systems and sources are harmonized, coded using terminologies, and integrated into a clinical data warehouse (CDW) at each participating hospital. The CDW is a centralized repository of clinical data that provides clinicians and researchers with a single, comprehensive view of patient information.

Semantics

The semantics are defined as concepts and attributes that can be used as building blocks to define more complex concepts. The **SPHN Dataset** is the collection of concepts, attributes, meaning binding, value sets and subsets, and any additional metadata that enriches the semantics.

Oxygen Saturation	fraction of oxygen present in the blood	SNOMED:103228002
quantity	value and unit of the concept	
measurement datetime	datetime of measurement	
measurement method	method of measurement of the concept	
body site	body site where the concept was measured, performed or collected	

Formal Representation of Semantics

The collection of concepts are then modeled using RDFS, SKOS and OWL, represented in RDF, and available as the **SPHN RDF Schema**. The translation of semantics to RDF is done via the **Dataset2RDF**.

```
sphn:OxygenSaturation a owl:Class ;
    rdfs:label "Oxygen Saturation" ;
    skos:definition "fraction of oxygen present in the blood" ;
    rdfs:subClassOf sphn:Measurement ;
    owl:equivalentClass snomed:103228002 ;
```

Oxygen Saturation

|	
URI	https://biomedit.ch/rdf/sphn-ontology/sphn#OxygenSaturation
Description	fraction of oxygen present in the blood
Schema representation	
Meaning binding (Equivalent-classes)	SNOMED 103228002	Hemoglobin saturation with oxygen (observable entity)
Parents	Measurement

Data Reuse

Data with richly defined semantics and metadata are made available by the participating university hospitals for the reuse by the researchers in compliance with institutional governance and Swiss legal and ethical regulations.

The SPHN contractual framework with its legal templates regulate data disclosure and principles of collaboration.

Data Preparation and Quality Control

SPHN provides various tools to support querying, validation, and exploration of data.

SHACLER - a tool to generate SHACL shapes from the SPHN RDF Schema, for data validation.

SPARQLER - a tool to generate SPARQL queries and statistical queries from the SPHN RDF Schema, for querying and analyzing RDF data.

SPHN Connector - a tool to generate and validate SPHN-compliant RDF data.

Interoperability

The SPHN strives to ensure semantic interoperability of health-related data by adopting a three pillar strategy.

Furthermore, in order to facilitate collaboration and interoperability with existing national and international healthcare communities, SPHN is actively investigating ways to translate its data into commonly used data models such as OMOP, i2b2, and FHIR. By utilizing these widely accepted data models, the SPHN aims to enhance the accessibility and compatibility of its data with external systems, promoting data exchange and facilitating comparative research across different datasets.
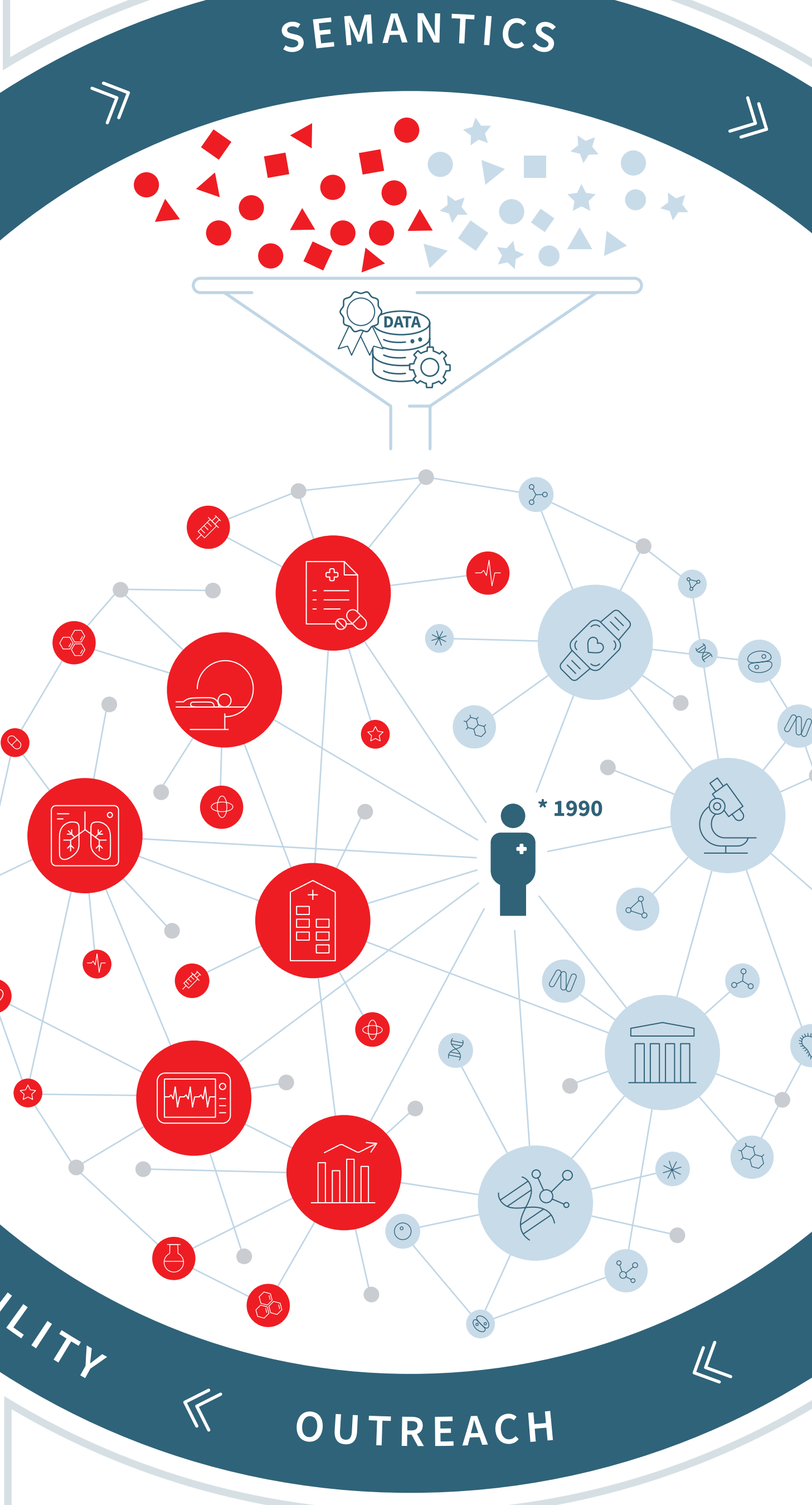
Outreach and Training

The SPHN provides outreach and training resources to generate awareness among the community on:

- Responsible Use of Health-Related Data for Research
- SPHN Ecosystem for FAIR Data
- FAIR Principles in Practice for Health Data
- Training on Federated Querying System

Discovery of (Meta)Data

The SPHN provides a Federated Query System (FQS) that combines consented data from five University Hospitals and enables researchers to query the clinical data to identify large number of patient records of interest and build cohorts.

Researchers can also browse existing Swiss cohort studies using the Maelstrom Catalogue.

SPHN Federated Query System