

Article

Not peer-reviewed version

Five Years of the SPHN RDF Journey: FAIR Enough?

Vasundra Touré , Deepak Unni , [Harald Witte](#) , Jan Armida , [Sabine Österle](#) *

Posted Date: 24 October 2025

doi: 10.20944/preprints202510.1783.v1

Keywords: FAIR; health data; knowledge graph; semantics; infrastructure; SPHN; retrospective



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Five Years of the SPHN RDF Journey: FAIR Enough?

Vasundra Touré, Deepak Unni, Harald Witte, Jan Armida and Sabine Österle *

Personalized Health Informatics, SIB Swiss Institute of Bioinformatics, Basel, Switzerland

* Correspondence: sabine.oesterle@sib.swiss

Abstract

Since 2020, the Swiss Personalized Health Network has adopted Semantic Web technologies to standardize health-related data for research in Switzerland. The SPHN Semantic Interoperability Framework promotes semantic interoperability, following the FAIR principles. Within this framework, the SPHN RDF Schema has evolved over five years to define more than 200 concepts across domains such as patient demographics, diagnoses, laboratory results, procedures, omics, and imaging metadata, enabling the representation of structured and machine-interpretable datasets. This study evaluates the evolution of schema versions from 2021 to 2025 and their adoption, examining structural and semantic changes, and analyzing quantitative metadata from projects in the SPHN Metadata Catalog. Results show consistent reuse of core concepts, especially demographics, diagnoses, and laboratory-related concepts, with 67% of SPHN concepts used in projects. The SPHN framework has proven to be a viable national standard for FAIR health data representation. Nonetheless, semantic modeling alone does not guarantee full interoperability. Future efforts must enhance data structuring and quality at the source, promote RDF adoption in research workflows, and develop user-friendly tools for querying and visualizing data.

Keywords: FAIR; health data; knowledge graph; semantics; infrastructure; SPHN; retrospective

1. Introduction

In 2020, the Swiss Personalized Health Network (SPHN) adopted Semantic Web technologies, particularly the Resource Description Framework (RDF) as a common standard format for representing and exchanging health-related data [1]. This strategic choice, supported by all stakeholders, promotes semantic interoperability across Switzerland's decentralized healthcare and research landscape. Over the past five years, the SPHN Semantic Interoperability Framework has evolved into a rich semantic schema, namely the SPHN RDF Schema [2], encompassing 209 concepts spanning diverse domains such as patient demographics, diagnoses, procedures, laboratory tests, molecular analyses, and imaging metadata. The framework is designed to harmonize data provisioning and reuse for research while ensuring compliance with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles.

To evaluate and iteratively improve this framework, SPHN has supported a range of projects, including four National Data Streams (NDS) and eleven Demonstrator (DEM) projects, that cover diverse research use cases. These projects also had the opportunity to extend the SPHN RDF Schema with project-specific concepts to address domain-specific needs, thereby contributing to the iterative evolution of the framework. This study marks a five-year milestone, evaluating both the progress achieved and the challenges that persist. Specifically, it addresses two key questions: i. How has the SPHN framework evolved over five years? ii. How widely is the SPHN framework adopted as a standard, and what data coverage do projects show?

To answer these questions, we analyzed the temporal evolution of the SPHN RDF Schema, assessed its use across SPHN projects and identified areas for improvement to further enhance the semantic interoperability and data reuse in Switzerland.

2. Methods

This study is based on a longitudinal analysis of the SPHN RDF Schema releases from the first official release in 2021 until 2025, accessible here: <https://git.dcc.sib.swiss/sphn-semantic-framework/sphn-schema/-/releases>. We analyzed structural and conceptual changes across versions to characterize their evolution. The analysis focused on quantifying the number and types of concepts (i.e. classes) introduced, tracking modifications in the organization of concepts, and assessing their alignment with external terminologies.

In parallel, we reviewed project-specific information available through the SPHN Metadata Catalog (<https://fdp.dcc.sib.swiss/>), a FAIR Data Point providing an overview of health-related datasets available in Switzerland. This platform mainly contains quantitative and qualitative metadata describing RDF data collected primarily from Swiss hospitals and submitted by NDS and DEM projects. As of now, metadata has been provided for six out of fifteen projects, and we analyzed the coverage of key semantic domains. The quantitative metadata was analyzed using a combination of SPARQL queries and custom R scripts to extract and aggregate the relevant statistics.

3. Results

3.1. SPHN RDF Schema Evolution over the Years

The first official release of the SPHN RDF Schema was published in June 2021 and included 64 concepts that primarily focused on core clinical domains such as encounters, medical devices, and allergies. In parallel, key standard terminologies (e.g. ATC, LOINC) were also provided in RDF via the SPHN DCC Terminology Service [3] to facilitate the integration of codes into the data and their subsequent analysis.

Typically, one major release occurs each year in Q1, occasionally followed by a minor mid-year release update. Over time, the schema expanded to cover additional domains (see Figure 1): laboratory tests (2022), genomics (2023), provenance, microbiology, assessments (2024), genomic variants and imaging (2025). The latest release (2025.2) includes 209 concepts. With each release, relevant standard terminologies were included or updated to support the increasing diversity and granularity of SPHN data, bringing the total to 18 included in release 2025.2.

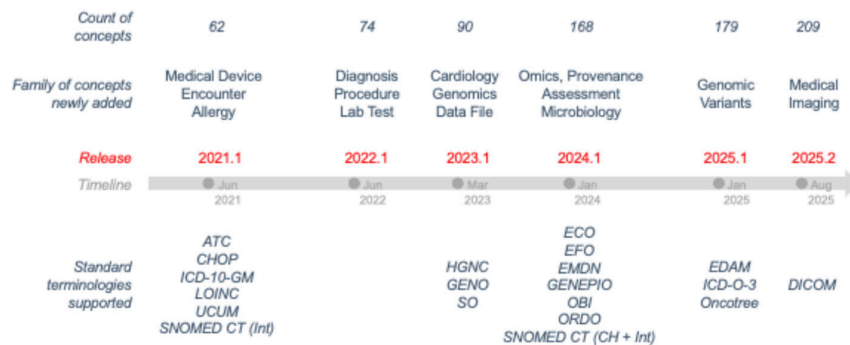


Figure 1. Timeline of the SPHN RDF Schema evolution, following the decision to adopt RDF as the standard knowledge representation in 2020. Each timestamp corresponds to a release of the SPHN Semantic Interoperability Framework, showing families of concepts developed and standard terminologies incorporated. The release year is embedded in the version number. The count of concepts indicates the total number of concepts defined in a release. Only major releases are shown; intermediary ones were discarded for readability.

3.2. Project Use of the SPHN RDF Schema

Most SPHN projects received RDF data from Swiss hospitals conforming to version 2024.1 of the SPHN RDF schema. These data were generated via the SPHN Connector, a tool that enforces schema-driven transformation and validation. Across projects, we observed that between 26% and 50% of the SPHN concepts were reused (see Figure 2). In total, 113 of the 168 concepts defined in version 2024.1 were used, showing the diversity of data needs across projects. Additionally, 71 project-specific concepts were introduced across these projects, demonstrating the flexibility of the SPHN schema to accommodate diverse project-specific requirements. General concepts, such as demographics, diagnoses, and laboratory values, are typically available in a structured and coded form within the clinical data platforms and are consistently used across projects. In contrast, more specific content (e.g., genomic variants or oncology-related diagnoses or assessments) is often only documented in free-text form and requires project-specific extraction and coding efforts.

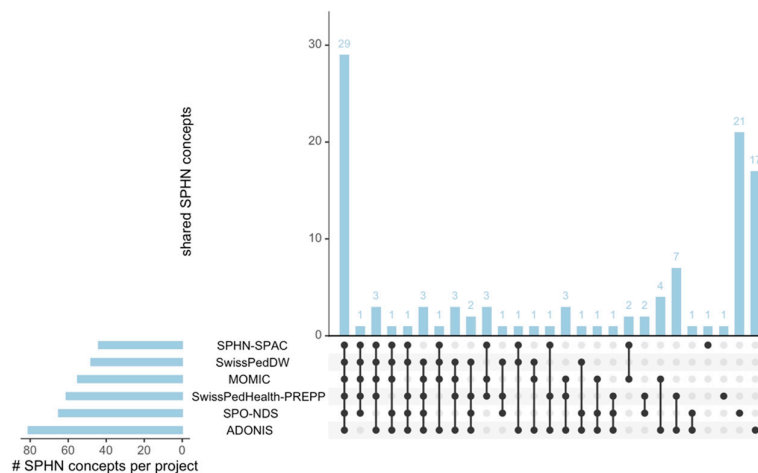


Figure 2. Overlap of concepts across SPHN DEMs and NDS projects. Vertical bars represent the number of SPHN concepts shared among specific projects, ordered from those shared by the most projects to those shared by fewer. Horizontal bars indicate the total number of SPHN concepts used in each project. The connected dots below the vertical bars denote which projects share each set of concepts. Full project metadata is available at <https://fdp.dcc.sib.swiss/>.

4. Discussion

The experience of the past five years demonstrates that adopting RDF as a foundation for semantic interoperability in health research is both feasible and broadly beneficial for data reuse. The SPHN schema offers a robust and extensible model, enabling alignment with international standards (e.g. SNOMED CT, LOINC), increasing interoperability. The framework is not only used by the SPHN projects but has also attracted interest from the international community, which has applied it for other use cases [4,5]. Local data heterogeneity, however, remains a challenge for semantic frameworks and may limit projects' insights.

At the start of SPHN, most hospitals had little to no SNOMED CT or LOINC coding in their clinical data platforms. Today, at least 9,000 distinct SNOMED CT and 4,000 LOINC code usages are recorded. Nevertheless, heterogeneity in local data coding still hinders full interoperability. These observations highlight that achieving truly interoperable health data requires not only well-defined semantic models but also harmonized data and coding practices across institutions, as well as supporting tools to ensure quality and usability for research. As potential next steps, feedback by projects to data providers and an adaptation of data collection at the source would be beneficial.

Balancing semantically precise and conceptually rich data representations with their practical usability remains an ongoing endeavor. In 2024, the SPHN RDF Schema underwent a major

restructuring to enhance consistency and clarity. For instance, body height was modeled as having a performer, which is semantically inaccurate. *Body height* represents an attribute defined by a value and a unit, whereas its *measurement* is a process that may involve a performer. This motivated a refactoring towards a **process-oriented modeling approach, as adopted by other initiatives [6]**, with the aim of clarifying the distinction between entities (e.g. Result, Code) and processes (e.g. Measurement, Medical Procedure, Assessment). The restructuring overall increased the number of concepts and initially raised concerns among implementers. However, we observed that concept patterns have since become more predictable, and no new patterns are required for newly developed concepts, suggesting that releases have reached structural stability.

This stability enables hospitals and other data providers to formalize their data according to the SPHN schema, guides data users in designing new concepts, and supports the development of tools for automated data transformation. As a result, the framework now offers improved semantic coherence and extensibility, which facilitates future developments and reduces modeling ambiguities.

5. Conclusion

The SPHN Semantic Interoperability Framework has established a national standard for health data representation, laying the groundwork for FAIR, reusable datasets in personalized health research. The schema's breadth and maturity provide a solid foundation for future data integration and cross-border collaboration. Several projects have successfully received data at scale, demonstrating the practical value of the approach. Moving forward, priorities must include enhancing data standardization and data quality at the source, strengthening adoption of RDF-based workflows into research environments, and developing intuitive tools for query building and data visualization.

References

1. Gaudet-Blavignac C, Raisaro JL, Touré V, Österle S, Crameri K, Lovis C. A National, Semantic-Driven, Three-Pillar Strategy to Enable Health Data Secondary Usage Interoperability for Research Within the Swiss Personalized Health Network: Methodological Study. *JMIR Med Inform* 2021;9:e27591. <https://doi.org/10.2196/27591>.
2. Touré V, Krauss P, Gnodtke K, Buchhorn J, Unni D, Horki P, et al. FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network. *Sci Data* 2023;10:127. <https://doi.org/10.1038/s41597-023-02028-y>.
3. Krauss P, Touré V, Gnodtke K, Crameri K, Österle S. DCC terminology service—an automated CI/CD pipeline for converting clinical and biomedical terminologies in graph format for the Swiss personalized health network. *Applied Sciences* 2021;11:11311. <https://doi.org/10.3390/app112311311>.
4. Jhee JH, Megina A, Constant Dit Beaufils P, Karakachoff M, Redon R, Gaignard A, et al. Predicting clinical outcomes from patient care pathways represented with temporal knowledge graphs. *European Semantic Web Conference, Springer; 2025*, p. 282–300. https://doi.org/10.1007/978-3-031-94575-5_16.
5. AIDAVA Reference Ontology n.d. <https://github.com/AIDAVA-DEV/AIDAVA-Reference-Ontology> (accessed October 17, 2025).
6. Kaliyaperumal R, Wilkinson MD, Moreno PA, Benis N, Cornet R, dos Santos Vieira B, et al. Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data. *J Biomed Semantics* 2022;13:9. <https://doi.org/10.1186/s13326-022-00264-6>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.