Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Metadata-based Assessment of SNOMED CT and LOINC Adoption Across 800 000 Swiss Patients

Vasundra Touré
SIB Swiss Institute of Bioinformatics
Jan Armida
SIB Swiss Institute of Bioinformatics
Harald Witte
SIB Swiss Institute of Bioinformatics
Deepak Unni
SIB Swiss Institute of Bioinformatics
Katie Kalt
University Hospital Zurich
Olga Endrich
Inselspital University Hospital Bern
Karen Triep
Inselspital University Hospital Bern
Mathias Gassner
University Children's Hospital Zurich
Seraphina Kissling
University Children's Hospital Zurich
Gaëlle Vuaridel-Thurre
University Hospital Lausanne
Pero Grgic
eHealth Suisse
Amanda Ramirez Ramos
University Hospital Zurich
Bram Stieltjes
University Hospital Basel
Marouan Borja
Geneva University Hospitals
Sabine Österle
sabine.oesterle@sib.swiss

SIB Swiss Institute of Bioinformatics

Research Article

Keywords: LOINC, SNOMED CT, Semantic interoperability, Clinical terminologies, Real world data, Metadata analysis

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10377450/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Metadata-based Assessment of SNOMED CT and LOINC Adoption Across 800 000 Swiss Patients

Vasundra Touré[1], Jan Armida[1], Harald Witte[1], Deepak Unni[1], Katie Kalt[2], Olga Endrich[3], Karen Triep[3], Mathias Gassner[4], Seraphina Kissling[4], Gaëlle Vuaridel-Thurre[5], Pero Grgic[6], Amanda Ramirez Ramos[2], Bram Stieltjes[7], Marouan Borja[8], Sabine Österle[1 *]

[1]*Swiss Personalized Health Network, SIB Swiss Institute of Bioinformatics, Basel, Switzerland*
[2]*University Hospital Zurich, Zurich, Switzerland*
[3]*Inselspital University Hospital Bern, Bern, Switzerland*
[4]*University Children's Hospital Zurich, Zurich, Switzerland*
[5]*University Hospital Lausanne, Lausanne, Switzerland*
[6]*eHealth Suisse, Bern, Switzerland*
[7]*University Hospital Basel, Basel, Switzerland*
[8] *Geneva University Hospitals, Geneva, Switzerland*

**Corresponding author: Sabine Österle, sabine.oesterle@sib.swiss*

Abstract

Background: Semantic interoperability is essential for enabling the secondary use of clinical data in research and healthcare. International standards such as SNOMED CT and LOINC are key enablers, yet empirical evidence on their real-world adoption in Switzerland remains limited.

Objective: This study presents a nationwide analysis of terminology usage based on metadata collected from hospitals contributing to the SPHN Federated Clinical Routine Dataset. The dataset includes information of approximately 800 000 patients, with a minimum data coverage spanning 2018 to 2025 across six Swiss University Hospitals. The goal is to quantify the adoption of semantic standards, identify strength and gaps in their usage, and highlight areas requiring further harmonization to support data quality and interoperability.

Methods: We analyze the use of SNOMED CT and LOINC international standard terminologies across participating hospitals by querying aggregated metadata from the SPHN Metadata Catalog via the SPARQL endpoint.

Results: A total of 6 557 SNOMED CT and 3 693 LOINC distinct codes were identified, representing approximately 717 million data instances. SNOMED CT and LOINC accounted for 60% and 32% of coded data instances, respectively. The remaining code instances originated from other coding systems (e.g., ATC, ICD-10-GM, CHOP). Despite this adoption, substantial heterogeneity was observed across hospitals, with most codes being institution-specific. A greater diversity of distinct codes was observed in domains such as “Substance” and “Organism”, whereas frequently instantiated concepts were concentrated in “Clinical Finding”, “Procedure”, and “Qualifier Value” hierarchies. LOINC code usage was dominated by routine laboratory tests across all institutions but with little overlap in the coding practice across institutions.

Conclusions:

While SNOMED CT and LOINC are widely used within the SPHN framework, significant variability in code usage and limited cross-institutional overlap highlight ongoing challenges for semantic harmonization. This study provides the first nationwide empirical baseline of terminology adoption in SPHN and demonstrates the value of metadata-driven approaches for monitoring interoperability readiness. The findings support targeted efforts to improve standardization and enhance the reuse of clinical data for research.

Keywords

LOINC, SNOMED CT, Semantic interoperability, Clinical terminologies, Real world data, Metadata analysis

Introduction

Interoperability of health data is a cornerstone of modern, data-driven healthcare and research. The consistent use of standardized clinical terminologies enables meaningful exchange, aggregation, and analysis of health data across institutions and national borders. Additionally, it provides a critical foundation for developing and validating robust and scalable artificial intelligence and machine learning models across clinical data from multiple sources [1].

Among the most widely adopted standards are the 10th revision of the International Classification of Diseases, specifically in Switzerland the German modification (ICD-10-GM [2]), a medical classification system used, among other things, for disease classification, cause of death certificates and billing; the Anatomical Therapeutic Chemical Classification System (ATC [3]), which classifies the active ingredients of a drug; the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT [4]); and the Logical Observation Identifiers Names and Codes (LOINC [5,6]) standard, which harmonizes the representation of laboratory tests and clinical measurements.

SNOMED CT is a multi-hierarchical terminology system that evolved since 2002 into the world's most comprehensive multilingual clinical terminology [4,7,8]. It is the result of the merger of SNOMED Reference Terminology (SNOMED RT) and the United Kingdom's Clinical Terms Version 3 (CTV3, creating a unified terminology for representing clinical information in electronic health records [9]. Over the past two decades, it has expanded significantly in scope, supporting the standardized representation of clinical findings, diagnoses, procedures, body structures, medications, devices, and many other healthcare concepts [10].

LOINC was developed in 1994 by the Regenstrief Institute to address the lack of standardized identifiers for laboratory tests and clinical observations exchanged between healthcare institutions. Today, LOINC is maintained as a freely available international standard that is widely used to support the interoperable exchange of laboratory and clinical observation data

across electronic health record systems, public health reporting, research networks, and exchange of health information [5,6].

Together, these standards form the semantic foundation for interoperable electronic health records (EHRs), clinical decision support systems, electronic patient dossiers, and large-scale health data research infrastructures.

European and national policy context

At the European level, the European Health Data Space (EHDS), adopted as Regulation (EU) 2025/327 and in force since March 2025, aims to establish a common framework for the secure (re)use and exchange of health information across Member States [11]. The EHDS promotes the adoption and use of shared semantic and technical standards, including SNOMED CT, LOINC, and Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR [12]), to ensure cross-border interoperability and to facilitate both primary care delivery and secondary use of health data for research. Although Switzerland is not a member of the European Union, EHDS requirements are expected to influence Swiss health information ecosystems due to the country's participation in European research collaborations, cross-border patient mobility, and alignment efforts in digital health policy.

At a national level, Switzerland made substantial progress toward establishing standardized health data exchange through the Electronic Patient Record, which mandates the use of internationally recognized terminologies for standardized data exchange [13]. In this context, Switzerland became a national member of SNOMED International in 2016, and a Swiss SNOMED CT extension is maintained by eHealth Suisse [14], the national coordination body for digital networking in the healthcare system. Complementing these initiatives, the Swiss Personalized Health Network (SPHN) developed a comprehensive semantic interoperability framework [15] that promotes the use of SNOMED CT as main terminology and LOINC for coding laboratory and measurement data, alongside other common biomedical standards. These efforts are closely aligned with the FAIR (Findable, Accessible, Interoperable,

Reusable) data principles [16], an internationally adopted framework for facilitating data sharing and reuse.

Building on these foundations, the DigiSanté program [17] and its Swiss Health Data Space represent a significant step forward in Switzerland's digital health transformation.
DigiSanté was launched in 2025 as a joint federal initiative by the Federal Office of Public Health (FOPH) and the Federal Statistical Office (FSO). DigiSanté aims to digitize the national health information infrastructure and promote the standardization and interoperability of health data across the healthcare ecosystem. International standards such as SNOMED CT, LOINC, and HL7 FHIR are identified as key enablers for semantic and technical interoperability. By establishing consistent data standards and governance frameworks, DigiSanté aims to improve data quality, facilitate secondary data use, and enhance the overall efficiency of Swiss healthcare services.

International adoption

Large-scale adoption of SNOMED CT in national healthcare initiatives began in the early 2000s and has expanded globally since then [18–21]. As of May 2026, SNOMED International comprises 53 member countries [22], including Switzerland, and the International Edition of SNOMED CT contains more than 360 000 concepts. The terminology is used in more than 80 countries to support semantic interoperability, clinical decision support, research, and health information exchange [22]. It has become a foundational terminology in numerous national eHealth programs [23–25], research networks, and interoperability initiatives worldwide [26]. Several countries have integrated SNOMED CT into their national health infrastructures; for instance, National Health Service England mandates the use of SNOMED CT as the standard clinical terminology across EHR systems in both primary and secondary care. Since its phased implementation, SNOMED CT is widely implemented in EHR platforms where it is used to structure clinical documentation, including diagnoses, symptoms, and procedures, enabling

consistent data capture and interoperability across care settings [18]. The United States also adopted SNOMED CT as a core component of its national interoperability strategy. It is designated as a required standard within federal health information technology certification criteria and is widely implemented in certified EHR systems to support semantic interoperability across providers, health information exchanges, and research infrastructures [27]. More recently, South Korea has incorporated SNOMED CT into its national digital health strategy as part of broader efforts to standardize clinical data exchange across institutions [21,28]. Beyond national adoption, SNOMED CT has also been employed extensively in clinical research, health information exchange, clinical decision support, and large-scale data harmonization projects, reflecting its central role in enabling semantic interoperability across healthcare systems [29–31].

Similarly, LOINC is widely adopted as a standard for laboratory and clinical observations with over 100 000 codes defined as of 2026 [5]. In the United States, LOINC is a required standard within federal health information technology certification criteria for the electronic exchange of laboratory results and clinical observations, forming a core component of national interoperability infrastructure [27]. It is widely implemented in EHR and laboratory information systems to ensure semantic consistency in the reporting and exchange of clinical test results across healthcare providers and public health agencies [32].

Together, these terminologies are increasingly promoted through national eHealth strategies, interoperability frameworks, and research infrastructures to support the secondary use and exchange of health data.

Challenges in Switzerland

In Switzerland, clinical routine data are not necessarily captured using international semantic standards at the point of care. Instead, SNOMED CT and LOINC have been introduced mainly

through research-driven initiatives, most notably SPHN, which required the use of standardized terminologies to represent interoperable data according to the SPHN specifications for secondary reuse [33]. As a result, terminology coding is often performed retrospectively or specifically within the SPHN context, rather than being natively embedded in routine clinical documentation workflows.

Therefore, relying solely on the availability of national terminology infrastructures may lead to an overly optimistic assessment of interoperability readiness. In Switzerland, systematic analyses of the usage of SNOMED CT and LOINC across hospitals are not yet available. Consequently, approaches that quantify interoperability without relying on patient-level data, such as evaluating code usage, frequency, and cross-institutional consistency at the institutional or dataset level, help in assessing progress and guiding further implementation efforts.

Objectives of this study

Within SPHN, Swiss hospitals built the SPHN Federated Clinical Routine Dataset (FedData), one of the largest Swiss real-world clinical datasets designed to achieve semantic interoperability in accordance with the SPHN framework [34]. The dataset contains de-identified data from all patients who provided broad (i.e., general) consent and had at least one clinical case since 2018. Data has been collected across six Swiss university hospitals: the University Children's Hospital Zurich (KISPI), University Hospital Basel (USB), University Hospital Bern (INSEL), University Hospital Geneva (HUG), University Hospital Lausanne (CHUV), and University Hospital Zurich (USZ).

The present study is based on a snapshot of the SPHN FedData reflecting the state of the datasets as of May 2026, comprising 817 856 patients. While the SPHN FedData subsets remain stored locally at the participating hospitals, our analysis relies on the metadata collected across these datasets as published in the SPHN Metadata Catalog [35]. Metadata-

based analysis provides a scalable and data privacy-preserving alternative to patient-level data analysis, making it particularly suitable for federated infrastructures.

The study aims to quantify and characterize the use of SNOMED CT and LOINC codes within the SPHN FedData, thereby providing an empirical baseline for assessing semantic standard adoption in the Swiss clinical data landscape. Such evidence is essential for guiding harmonization efforts and informing national strategies for interoperable, research-ready health data infrastructures.

Methods

Study Design and Objectives

This study used a cross-sectional, observational approach to assess the overall readiness of Swiss hospitals to adopt SNOMED CT and LOINC terminologies, in the areas covered by SPHN. Using aggregated metadata from the SPHN Metadata Catalog [35], we quantified the presence and coverage of standardized codes in the SPHN FedData [34] across hospitals without accessing any patient-level data (i.e., all analyses exclusively relied on aggregated metadata). The primary objective was to evaluate interoperability readiness by identifying which SNOMED CT and LOINC codes are used in this SPHN-aligned dataset, providing a baseline for monitoring terminology adoption and guiding future harmonization efforts across hospitals.

Data Source

In this study, we analyzed metadata associated with the SPHN FedData, which includes semantically harmonized, de-identified clinical data generated locally at six Swiss university hospitals: CHUV, HUG, INSEL, KISPI, USB, and USZ. The minimum dataset requirement comprises records from patients who provided broad consent (mandatory condition) and had at least one inpatient or outpatient encounter in the hospital (corresponding to the SPHN

concept “Administrative Case”) between 2018 and 2025. All data elements in the SPHN FedData are mapped to the SPHN Resource Description Framework (RDF) Schema 2025.1 [36], a semantic model based on semantic web technologies for representing healthcare concepts [33]. In that schema version, data elements conform to the December 2024 SNOMED CT International Edition and Swiss Edition, and to LOINC version 2.78. Details on the generation of the SPHN FedData, its metadata extraction and processing are described by *Armida et al.* in reference [34]. It is important to note that instance counts of codes below ten are reported as zero in the metadata itself, to indicate code availability without disclosing the exact number.

The SPHN FedData metadata is available in the SPHN Metadata Catalog [37], capturing the structural and semantic characteristics of the datasets, including clinical terminologies and codes used. Our study is based on the metadata published on the SPHN Metadata Catalog, as of May 2026.

SNOMED CT and LOINC analysis

We programmatically retrieved relevant metadata for this study via SPARQL queries run against the SPARQL endpoint [38] of the SPHN Metadata Catalog, targeting the ‘code frequency’ metadata of the SPHN FedData. The complete analysis workflow is available as an R Markdown document on Gitlab [39]. All metrics were computed on aggregated metadata for the selected snapshot available in Zenodo [40].

The analysis focused on the following metrics:

i) descriptive statistics on delivered codes, at the level of distinct codes and coded instances, where ‘coded instance’ refers to a reported occurrence count of a terminology identifier, and ‘distinct code’ refers to unique terminology identifier independent of their frequency of occurrence;

ii) distribution of delivered codes across SNOMED CT top-level categories and LOINC laboratory classes to characterize thematic coverage of each terminology;

iii) overlap of distinct codes across hospitals to assess cross-institutional semantic alignment, used as an indicative measure of completeness;
iv) identification of the ten most frequently used instantiated codes to characterize commonly shared semantic representations;
v) analysis of SNOMED CT code depth across categories and hospitals as an exploratory proxy for coding granularity.

Coding Granularity Depth Analysis

For the coding granularity analysis, SNOMED CT code depth was calculated as the hierarchical distance between an instantiated code and its corresponding top-level category, defined as the direct child concepts of the SNOMED CT root concept 138875005 ‘SNOMED CT Concept’. With SNOMED CT being polyhierarchical and individual codes possibly having multiple parent paths of differing lengths, depth was defined as the length of the shortest hierarchical path connecting a concept to its assigned top-level category. This method yields a deterministic value for each code ensuring comparability of coding granularity across hospitals. Mean depth, standard deviation, and minimum–maximum depth range were computed for each category and hospital. These measures were interpreted descriptively as indicators of coding granularity, not as measures of semantic correctness or coding quality.

Results

Across the over 800 000 patients and the 125 SPHN concepts delivered, the SPHN FedData contains a total of 6 557 distinct SNOMED CT codes, approximately 2% of the total SNOMED CT International release, and 3 693 distinct LOINC codes, approximately 4% of the total LOINC standard (see Table 1). These codes account for approximately 717 million instances of data (464 million SNOMED CT and 253 million LOINC). Substantial variation is observed in the number of distinct terminology codes provided across hospitals. The number of distinct

SNOMED CT codes per hospital ranges from 669 to 3 154, while the number of distinct LOINC codes ranges from 220 to 1 426.

Table 1. Distribution of patients, SPHN Concepts and distinct codes across the Swiss hospitals. H1-H6 corresponds to the participating Swiss hospitals included in the study. The 'SPHN Concepts' row refers to the number of SPHN semantic concepts mapped in each hospital's dataset. The 'Combined (unique)' column indicates for each row the union of values across all six hospitals in the FedData. **Note:** This table extends data previously reported in reference [34]. Patient counts and SPHN concept counts for some participating hospitals were reported previously; additional variables and updated counts are presented here.

	H1	H2	H3	H4	H5	H6	**Combined (unique)**
Patients	102116	113194	230242	106075	145366	120863	**817856**
SPHN concepts	82	92	89	73	86	101	**125**
SNOMED CT value sets	44	33	31	24	20	47	**88**
Distinct SNOMED CT codes	3154	2119	669	871	1603	2683	**6557**
Distinct LOINC codes	1113	1359	1426	220	699	890	**3693**

As illustrated in Figure 1, coded data instances in the SPHN FedData rely on international standard terminologies. SNOMED CT represents the largest share with 60% of all code instances, followed by LOINC with 32%. Other terminologies, ATC, CHOP, and ICD-10-GM collectively represent 8% of instances. This distribution shows the abundance of SNOMED CT and LOINC compared to the other terminologies in the SPHN FedData.

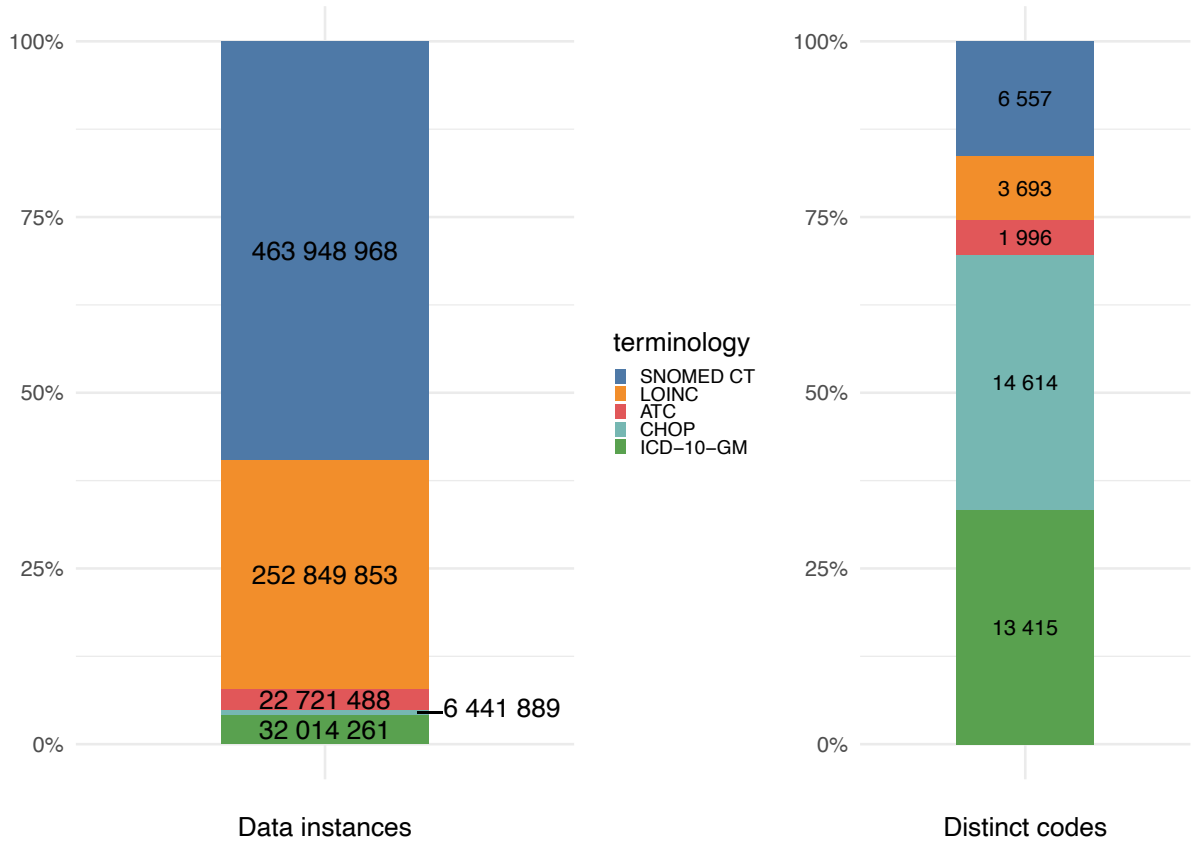


Figure 1. Distribution of coded data instances (left) and distinct terminology codes (right), with count overlaid. The left panel shows the percentage of all code instances of SNOMED CT (blue), LOINC (orange), ATC (red), ICD-10-GM (green), and CHOP (turquoise) terminologies. The right panel shows the percentage of distinct codes, showing the proportion of unique codes sourced from the previously cited terminologies. In both panels, the exact counts of either instances or distinct codes are displayed on the bars. This visualization highlights the difference between the frequency of code usage (instances) and the diversity of used codes (distinct codes) across the relevant terminologies.

SNOMED CT

When looking at the SNOMED CT value sets defined in the underlying SPHN RDF Schema, a total of 88 value sets across the 125 delivered concepts (Table 1) are possibly implemented

across the contributing hospitals. A value set is a collection of codes drawn from the relevant terminologies that define the permitted values for a given SPHN concept within the schema. Of the 88 value sets, 32 value sets are defined in the schema as explicit lists of distinct codes, ranging from a single code up to 11 distinct codes, and 46 are defined using hierarchical SNOMED CT constraints, allowing descendants of one or more parent concepts, that range between four and 124 028 allowed codes (i.e., due to enabling ‘Clinical finding’ concepts for the annotation of Allergy Episode). The final 10 value sets permit the use of the full SNOMED CT without restriction. Across all hospitals, the number of potential value sets based on the concepts delivered varies considerably, ranging from 20 to 47 per hospital (see Additional File 1 for the list of value sets defined in the SPHN RDF Schema).

The distribution of distinct SNOMED CT codes across semantic domains was uneven. Most distinct codes delivered are concentrated in the ‘Substance’, ‘Organism’, and ‘Body structure’ domains (see Figure 2B), although their coverage in terms of instance data is rather limited. On the other hand, ‘Clinical Finding‘, ‘Procedure‘, ‘Specimen‘ and ‘Qualifier value’ dominate in instance counts (Figure 2A), contrasting with their modest representation in the distinct code counts (Figure 2B). Concepts from the ‘Procedure’ and ‘Staging and scales’ are present but sparsely coded. This divergence reveals that a small set of procedure and qualifier codes drives the majority of SNOMED CT usage in SPHN, while semantic diversity is concentrated in substance, organism, and body structure domains.

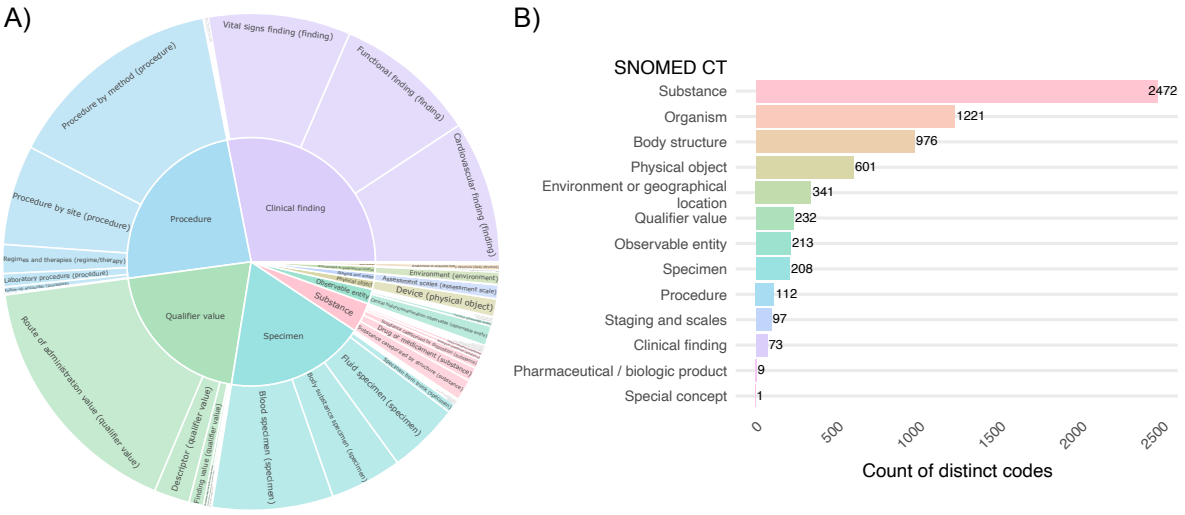


Figure 2. SNOMED CT usage in SPHN FedData.

A) Sunburst visualization of SNOMED CT code usage across hospitals, at instance level. Each segment size represents the number of code instances, and colors indicate SNOMED CT code categories. The inner ring shows the top-level categories and the outer ring reflects the second-level categories of coded data instances in the SPHN FedData. An interactive version of the plot is available in reference [41], including exact instantiated codes with their corresponding counts.

B) Bar plot showing the count of distinct SNOMED CT codes available in the SPHN FedData, aggregated by their top-level category.

This pattern is further illustrated in Figure 3A, where the count of distinct codes per SNOMED CT category is displayed on a logarithmic scale alongside the number of hospitals sharing each code. 'Substance', 'Organism', and 'Body structure' exhibit the largest absolute counts of distinct codes, yet most of these codes are used by only one, two or three hospitals, underscoring their institution-specific character. In contrast, categories such as ‘Qualifier value’ or ‘Clinical finding’, while contributing fewer distinct codes overall, show a comparatively higher degree of sharing across four to more hospitals.

The upper bound for cross-hospital code sharing observed in Figure 3A is constrained by two factors: the number of hospitals delivering such codes and the hospital contributing to the

fewest codes. This is illustrated in Figure 3B, which shows for each category, the total count of distinct codes delivered per hospital. Some categories are represented across all six hospitals (e.g., 'Substance', 'Staging and scales') while others are delivered by only a subset (e.g., 'Pharmaceutical / biologic product' by three hospitals, 'Specimen' by five). An example for the latter is the fact that one hospital contributed only four distinct codes to the 'Substance' category, inherently limiting the observed cross-hospital sharing for this category in Figure 3A, where all other hospitals delivered at least 291 codes.

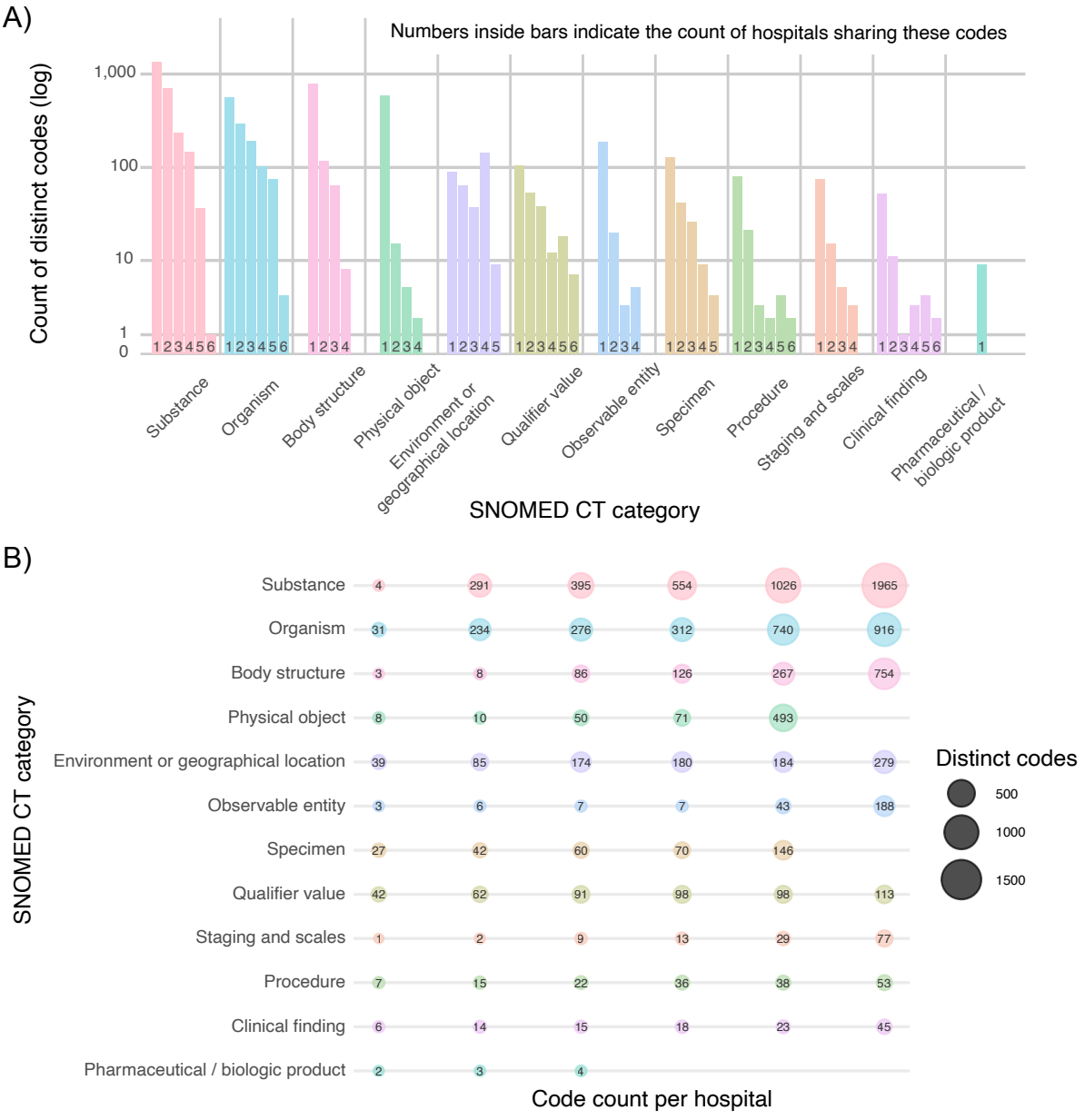


Figure 3. Overlap of SNOMED CT codes across hospitals. A) Log-scale distribution of

distinct SNOMED CT codes shared across hospitals, grouped by top-level category and stratified by the number of hospitals in which each code occurs. The number of hospitals delivering each code is indicated by the number at the bottom of each bar. For example, only one code was delivered in Substance category by all six hospitals while there were more than 1 000 distinct codes delivered only by one of the hospitals. B) Bubble plot of distinct SNOMED CT codes delivered per hospital across the top-level categories. Each row represents one SNOMED CT top-level category; and within a row, each bubble represents one hospital, ordered from left to right by increasing code count. The number of bubbles in a row indicates how many hospitals contributed at least one code in that class, while bubble size reflects the exact count of distinct codes contributed.

An analysis of the ten most frequently instantiated SNOMED CT codes across all hospitals (Table 2) reveals the predominance of generic laboratory- and procedure-related qualifier values (e.g., 'Intravenous route', 'Subcutaneous route', 'Intramuscular route'), alongside basic demographic and administrative codes predefined in the SPHN RDF Schema, including 'Female', 'Male', 'Accepted' and 'Generalized', with the latter two used to represent patient consent status and type, respectively. The consistent use of these codes across institutions reflects a strong harmonization for predefined value sets and highlights the stabilizing effect of semantic constraints defined at the data schema level. These top ten codes account for 11,6% (54 million of 464 million) of the SNOMED CT data instances delivered in the SPHN FedData.

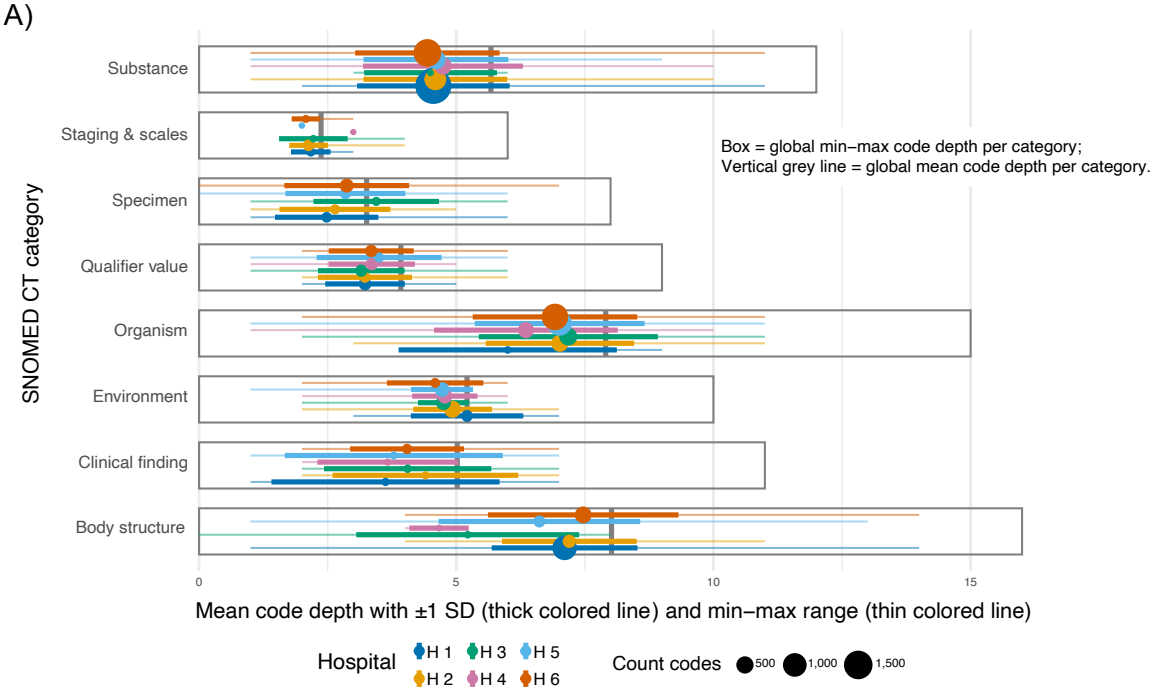
Table 2. Top ten SNOMED CT codes used across all six hospitals. The SNOMED CT code, its associated label and instance count, representing codes used by every participating institution, are shown.

SNOMED CT code	SNOMED CT label	Instance count
http://snomed.info/id/47625008	Intravenous route (qualifier value)	41668857
http://snomed.info/id/34206005	Subcutaneous route (qualifier value)	6583232

http://snomed.info/id/63161005	Principal (qualifier value)	2131045
http://snomed.info/id/60132005	Generalized (qualifier value)	815091
http://snomed.info/id/385645004	Accepted (qualifier value)	815089
http://snomed.info/id/394656005	Inpatient care (regime/therapy)	533302
http://snomed.info/id/248152002	Female (finding)	413874
http://snomed.info/id/248153007	Male (finding)	401204
http://snomed.info/id/78421000	Intramuscular route (qualifier value)	163117
http://snomed.info/id/72607000	Intrathecal route (qualifier value)	14533

Beyond sharing patterns, coding granularity itself varied across domains and hospitals as illustrated in Figure 4A where the mean SNOMED CT code depth was calculated as the shortest hierarchical path between each code and its corresponding top-level category. Variation observed in mean SNOMED CT code depth across hospitals was primarily driven by differences between semantic domains. Categories such as ‘Body structure’, ‘Clinical Finding’, ‘Organism’, ‘Environment’, and ‘Substance’ show a greater mean code depth, indicating a fine-grained representation of these domains within SNOMED CT. In contrast, ‘Staging & scales’, ‘Specimen’ and ‘Qualifier Value’ showed lower mean depth. Overall, in each category, hospitals tended to use shallower codes than the global SNOMED CT mean. At the hospital level, coding granularity remained comparatively consistent for most categories, except ‘Staging & scales’ and ‘Body structure’, where greater variations across hospitals are observed. Particularly, in ‘Body structure’, multiple hospitals used distinct but hierarchically related concepts to represent anatomically comparable information. Figure 4B illustrates this pattern using the Body structure hierarchy. Starting from the top-level category ‘Body structure’ (depth 0), increasingly specific anatomical representations extend through ‘Ankle region structure’ (depth 6) and ‘Ankle joint structure’ (depth 7) to laterality-specific concepts such as ‘Left ankle joint structure’ and ‘Right ankle joint structure’ (depth 8). Hospitals differed in the depth at which such concepts were recorded, reflecting alternative representation choices within the same semantic hierarchy.

A)



B)

code depth	SNOMED CT code	hospital
0	▼ 123037004 | Body structure (body structure) |	
1	▼ 442083009 | Anatomical or acquired body structure (body structure) |	
2	▼ 91723000 | Anatomical structure (body structure) |	
3	▼ 52530000 | Body region structure (body structure) |	
4	▼ 785818007 | Structure of joint region (body structure) |	
5	▼ 787050004 | Structure of joint region of lower limb (body structure) |	
6	▼ 344001 | Ankle region structure (body structure) |	H1, H6
7	▼ 70258002 | Ankle joint structure (body structure) |	H1, H5
8	▼ 720533003 | Right ankle joint structure (body structure) | ▼ 720532008 | Left ankle joint structure (body structure) |	H1, H6 H1, H6
9	▬ 737464007 | Structure of articular cartilage of left ankle joint (body structure) |	

Figure 4. Depth analysis of SNOMED CT coding. A) Mean SNOMED CT code depth across top-level categories for the six hospitals. Each colored point represents the mean depth of codes used within the SNOMED CT category (e.g., ‘Body structure‘, ‘Qualifier value‘, ‘Substance’) for each hospital. Thick colored lines represent standard deviation (SD) and thin

bars indicate the minimum-maximum depth range observed within each hospital. Grey boxes represent the global range of depth (with minimum and maximum) for each category, while the vertical grey lines indicate the global mean depth. Differences across hospitals highlight variation in coding granularity between and within the data providers. The depth of a code was calculated as the shortest path between the code of interest and the top-level SNOMED CT category. B) Example of SNOMED CT code depth within the 'Body structure' category. Beginning with the top-level code 'Body structure' (depth 0), this figure shows the path through increasingly specific anatomical codes: 'Ankle region structure' (depth 6), 'Ankle joint structure' (depth 7), and the laterality-specific codes 'Left ankle joint structure' and 'Right ankle joint structure' (depth 8). Codes highlighted in red were used by three hospitals, whereas black nodes indicate concepts not instantiated in the analyzed SPHN FedData dataset. Inverted triangle means the code has children codes. Dash means the code is a leaf code.

LOINC

The distribution of distinct LOINC codes across laboratory classes and hospitals was similarly uneven (Figure 5). The semantic diversity of the delivered codes is concentrated within 'Chemistry – non-challenge' (dominating class), 'Allergy', 'Microbiology', 'Drug toxicology' and 'Hematology and Cell counts', which together account for 82% of the delivered distinct codes (Figure 5B). In contrast, panels and molecular pathologies subcategories contribute only marginally to the overall instance and distinct code count.

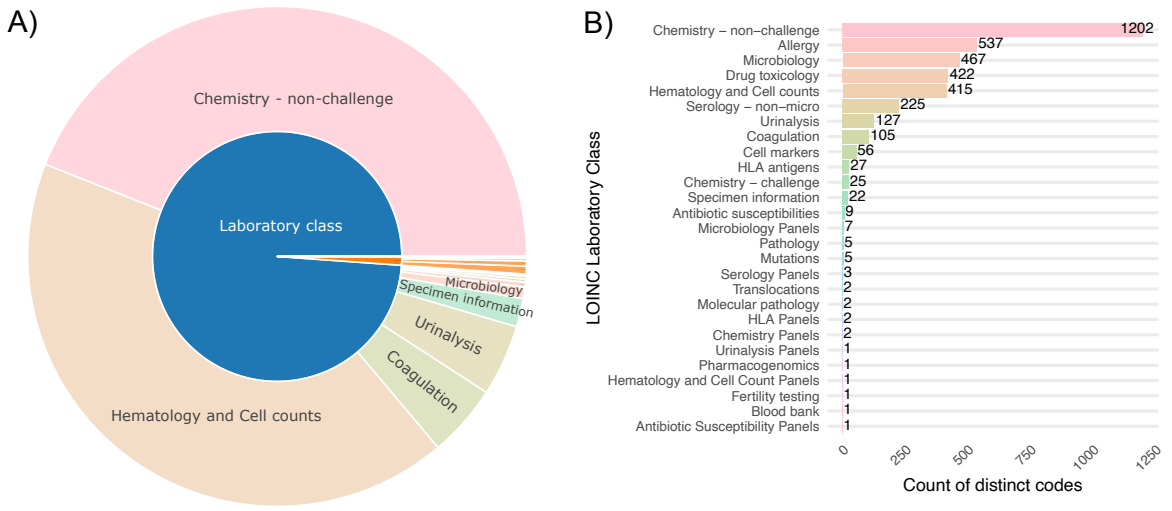


Figure 5. LOINC code usage in the SPHN FedData. A) Sunburst visualization of LOINC class types code usage at the instance level across hospitals. An interactive version of the plot is available at reference [41]. B) Count of distinct LOINC codes grouped by the 27 LOINC class types. Only the laboratory-related class types are shown, as the remaining four non-laboratory class type codes are limited in size and fall outside the scope of this analysis.

The ten most frequent LOINC codes across hospitals (Table 3) are entirely composed of routine hematology and basic metabolic tests, reflecting a universal set of core routine tests performed across all institutions and a consistent usage of LOINC codes for representing these tests. These ten codes collectively reflect 21% (54 million of 253 million) of the LOINC data instances delivered in the SPHN FedData.

Table 3. Top ten LOINC codes used across the hospitals. The LOINC code, its associated label, and instance count are shown. This table also reflects the top ten codes shared across all six participating hospitals.

LOINC code	LOINC label	Instance count
https://loinc.org/777-3	Platelets [#/volume] in Blood by Automated count	5579607
https://loinc.org/786-4	MCHC [Mass/volume] by Automated count	5469499
https://loinc.org/2823-3	Potassium [Moles/volume] in Serum or Plasma	5427323
https://loinc.org/6690-2	Leukocytes [#/volume] in Blood by Automated count	5412252
https://loinc.org/2951-2	Sodium [Moles/volume] in Serum or Plasma	5372283

https://loinc.org/4544-3	Hematocrit [Volume Fraction] of Blood by Automated count	5361108
https://loinc.org/787-2	MCV [Entitic volume] by Automated count	5332596
https://loinc.org/789-8	Erythrocytes [#/volume] in Blood by Automated count	5319672
https://loinc.org/785-6	MCH [Entitic mass] by Automated count	5269217
https://loinc.org/14682-9	Creatinine [Moles/volume] in Serum or Plasma	5257819

The codes listed in Table 3 belong to 'Hematology and Cell counts' and to 'Chemistry – non-challenge', the same classes driving the highest cross-hospital sharing observed in Figure 6. Figure 6 illustrates the distribution of LOINC laboratory class codes and degree of cross-hospital sharing. For example, for 'Microbiology', one distinct code is shared by exactly three hospitals, while for 'Chemistry – non-challenge' and 'Hematology and Cell count', there are more than ten codes shared by all six hospitals. The overarching observation is that the overlap of codes across hospitals is limited, rarely exceeding ten shared codes when checking across four or more hospitals. The notable exceptions are 'Allergy', 'Chemistry – non-challenge', 'Drug toxicology', and 'Hematology and Cell counts', which also represent the most extensively coded laboratory classes. At the same time, high cross-hospital sharing would not have been detectable with classes like 'Cell markers', 'Chemistry – challenge' or 'Specimen information' due to the limiting factor of certain hospitals contributing only a small number of codes (Figure 6B). This joins the constraint highlighted previously for SNOMED CT: The hospital contributing the lowest number of codes sets the boundaries for the degree of observable overlap across hospitals.

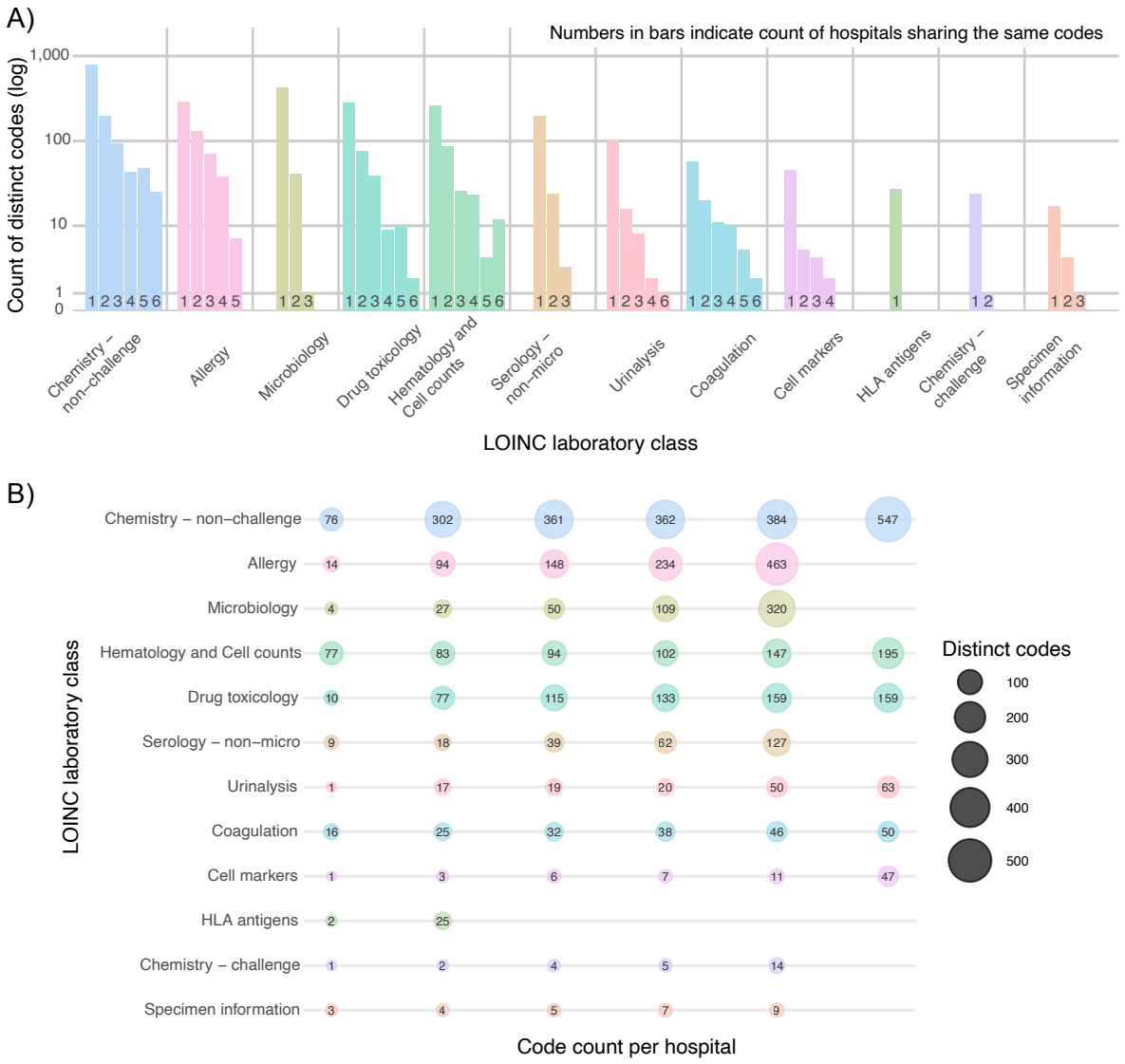


Figure 6. Distribution of LOINC codes across hospitals A) Log-scale distribution of distinct LOINC codes shared across hospitals, grouped by LOINC Laboratory classes. Numbers within bars indicate how many hospitals share the same set of codes. For example, for 'Microbiology', one code (count of distinct codes = 1) was delivered by exactly three hospitals, while for 'Urinalysis', there were approximately nine codes delivered by exactly three hospitals. B) Bubble plot of distinct LOINC codes delivered per hospital across LOINC Laboratory classes. Each row represents one LOINC class; and within a row, each bubble represents one hospital, ordered from left to right by increasing code count. The number of bubbles in a row indicates how many hospitals contributed at least one code in that class, while bubble size reflects the exact count of distinct codes contributed. Note that in both panels A and B, only

Laboratory classes are considered and categories with less than twenty codes shared have been omitted (these categories are visible in Figure 5B).

Discussion

This study provides the first nationwide, metadata-driven assessment of SNOMED CT and LOINC adoption in Swiss clinical routine data made available in the context of SPHN, leveraging metadata from the SPHN FedData across six university hospitals. By analyzing terminology usage patterns at scale without accessing patient-level data, it establishes an empirical baseline for monitoring semantic standard adoption in Switzerland and contributes evidence to ongoing national and European initiatives.

A major strength of this work lies in its national, multi-institutional scope. The analysis spans more than 800 000 patients, offering a comprehensive view of terminology usage patterns that would be difficult to obtain through institution-specific studies. The metadata-based approach enables large-scale analysis while avoiding privacy and governance constraints associated with patient-level data access. This makes the methodology reproducible, scalable, and well-aligned with federated data infrastructures, such as DigiSanté and the EHDS.

At the same time, the concepts and codes represented in the SPHN FedData are inherently constrained by the scope of the SPHN RDF Schema, as only schema-supported concepts can be represented and shared in a FAIR manner. Furthermore, the analyzed metadata may not reflect all terminology mappings available locally within hospitals, as the SPHN FedData is restricted to a defined patient population: individuals treated between 2018 and 2025, who provided general consent and, whose data were modeled according to the SPHN RDF Schema. These constraints should be considered when interpreting the observed terminology usage patterns.

Key Findings

The central finding of this study is that widespread terminology adoption does not necessarily translate into semantic harmonization. SNOMED CT and LOINC together accounted for 87% of all coded instances in the SPHN FedData, indicating adoption of international semantic standards across Swiss university hospitals. However, the limited overlap of distinct codes across institutions demonstrates that semantic harmonization remains incomplete (Figure 3 and 6). This heterogeneity reflects differences not only in terminology usage, but also in data availability, documentation practices, knowledge representations, technical pipelines and software, available resources and retrospective coding strategies.

Codes consistently used across all hospitals typically reflect universally applicable concepts that are routinely recorded for all patients. In some cases, the SPHN RDF Schema itself governs the degree of harmonization that can be achieved across institutions. For example, concepts such as ‘Administrative Sex’ or ‘Consent’ show no mapping variance, as they are constrained to predefined SNOMED CT values within the schema (e.g., ‘General’, ‘Accepted’, ‘Male’, and ‘Female’). While such constraints reduce heterogeneity, they could in some cases potentially limit the ability to capture local clinical nuances.

In contrast, concepts like ‘Body structure’ show greater variability, as local source systems may support multiple equally valid SNOMED CT representations, leading to divergent encodings across hospitals. Similarly, heterogeneity in terminology coverage across institutions (Table 1; Figure 3 and 6) can be due to variability in coding practices, including cases where codes were assigned retrospectively or manually outside of the routine clinical setting. It may also reflect technical variation between hospitals, such as differences in laboratory methods that lead to distinct LOINC codes. For example, two laboratory tests assessing the same analyte, but performed either manually or automatically, get assigned two different LOINC codes. The scope of terminologies used for coding can introduce further

heterogeneity. LOINC, for example, reflects six dimensions in its coding which results in individual codes for each combination of dimensions. Data availability and different data processing practices contribute further to heterogeneity, and all of these aspects together pose challenges for cross-institutional data comparability and federated analyses. The observed presence of hospital-specific codes may also partly reflect differences in clinical specialization and patient populations, with some hospitals emphasizing body structures, other substances, organisms, or other concept domains.

These findings highlight a key insight of the study: Terminology adoption is a necessary but insufficient condition for interoperability. Even within a shared semantic framework, local documentation practices, source-system constraints, data availability, and mapping strategies introduce substantial heterogeneity, posing challenges for data interoperability and comparability across hospital datasets.

Mean SNOMED CT depth and coding granularity

Mean SNOMED CT depth was used as an exploratory proxy to assess differences in coding granularity across terminology domains and hospitals. The observed variation suggests that semantic heterogeneity in the SPHN FedData is not limited to whether hospitals use codes from the same terminologies or value sets but also extends to the level of specificity at which clinical information is represented. Categories such as ‘Body structure’, ‘Clinical finding’, ‘Organism’, ‘Environment’, and ‘Substance’ showed greater mean depth which correlates with the depth of these categories in SNOMED CT (see Figure 3). In such categories, clinical data are highly contextualized and benefit from precise representation. In contrast, categories such as ‘Qualifier Value’, ‘Specimen’, and ‘Staging and scales’ showed lower mean depth, which may reflect schema-constrained value sets, or the more limited hierarchical depth of these SNOMED CT categories: these categories often provide fewer deeply nested codes.

Importantly, mean depth should not be interpreted as a direct measure of coding quality. A higher depth may indicate greater specificity, but not necessarily better semantic alignment, while a lower depth may still represent clinically appropriate coding depending on the context and the structure of the terminology. Both institutional coding practices and ontology design, as drivers of the observed differences in granularity, are an important dimension of interoperability readiness, particularly for federated analyses where data is encoded at different levels of specificity.

Body Structure example: pre-coordination versus model-based post-coordination

Body structure-related concepts provide a particularly informative example of how differences in mean SNOMED CT depth can arise from alternative, but semantically comparable, representation strategies. In this context, pre-coordination refers to representing clinical information using a single, specialized SNOMED CT concept, whereas model-based post-coordination refers to distributing the same information across multiple linked elements in the SPHN RDF Schema, following definitions from SNOMED CT International in [42]. Accordingly, in some hospitals, anatomical information may be represented using more specific pre-coordinated SNOMED CT concepts, for example “left breast structure”. In other cases, the same information may be represented through the SPHN RDF Schema by combining a code for “breast structure” with a separate laterality value such as “left”, analogous to SNOMED CT post-coordination. While both approaches may preserve the relevant clinical information at a comparable level of granularity, they result in different SNOMED CT codes and depth profiles and may therefore appear heterogeneous in metadata-based analyses.

This illustrates that mean depth reflects not only terminology specificity, but also how information is distributed between the terminology and the data model. For users of federated data, such differences may be inconvenient because equivalent clinical information can

require different query patterns, joins, or normalization steps across hospitals. However, this form of heterogeneity is partly inherent to real-world clinical data, where source systems, documentation practices, and mapping workflows differ between institutions.

Consequently, mean SNOMED CT depth should be interpreted as a useful signal for identifying variation in representation patterns and potential user burden, rather than as a standalone measure of coding quality. This finding supports the development of clearer guidance, reusable query templates, and harmonization recommendations for domains such as 'Body structure', where pre-coordinated and model-based post-coordinated representations may coexist.

Data quality issues and semantic correctness

The observed heterogeneity suggests that harmonization cannot rely solely on the SPHN integration layer but must also be strengthened at the source within hospitals. While SPHN provides a common semantic framework for alignment and interoperability across institutions, different code selection and granularity limit the effectiveness of downstream harmonization. Improved guidance and standardization defined at the level of data generation, together with robust semantic mappings and validation mechanisms, are therefore essential.

A feedback loop with participating hospitals has been established, aiming to gradually improve data quality and data interoperability, enhance trust in the SPHN FedData, and maximize its reusability for research purposes. Engaging hospitals in reviewing discrepancies and implementing standardization protocols will systematically reduce errors while ensuring interoperability, wherever possible. Continuous validation therefore remains essential.

It is important to note that evaluating data quality using metadata alone has inherent limitations. Relationships between data elements and their clinical context can only be fully assessed within the actual data represented as a knowledge graph. Hence, the present analysis cannot assess semantic correctness, equivalence of representations across hospitals, or conformity with SNOMED CT editorial guidelines and the Machine-Readable Concept Model [43].

Nonetheless, metadata-based monitoring plays a critical complementary role by enabling scalable, privacy-preserving identification of trends, gaps, and priorities for improvement.

Challenges and needs for implementation

A further challenge lies in SNOMED CT itself. Despite its extensive scope, gaps remain, and data providers frequently encounter situations where no sufficiently precise code exists for the data element they need to represent. This may force users to select either a closely related but imprecise code, a hierarchically higher and more general concept, or no code at all, thereby reducing semantic precision. Although beyond the scope of this study, future work could assess the potential impact of the presence or absence of German, French, and Italian synonyms in the Swiss SNOMED CT extension on coding practices, particularly for semantically granular clinical concepts. This has been mentioned by [44] for issues related to expressions and semantic opposites between French and English. Regarding implementation, although SPHN uses the Swiss extension of SNOMED CT, no Swiss-specific SNOMED CT codes were identified in the delivered data. Additionally, all delivered SNOMED CT codes were valid codes with respect to the version used.

Addressing these challenges requires collaboration with terminology authorities, in the case of Switzerland, eHealth Suisse, to ensure that the evolving needs of clinical data providers are reflected in future releases of SNOMED CT. SPHN has contributed to SNOMED CT through concept requests: out of 43 new concepts submitted by the SPHN community, 33 were accepted and incorporated into SNOMED CT International in 2024, rather than introduced as extension to the Swiss Edition of SNOMED CT only. Improving semantic interoperability ultimately requires investment at the source, including education and training, appropriate tools, sufficient workforce, and a shared understanding of core semantic principles within hospitals. Consistent and high-quality mapping is inherently resource-intensive, often without

immediate payoff, demanding interdisciplinary expertise spanning clinical practice, terminology modeling, and biomedical informatics. Mid- to longer-term, however, improved data quality will result in benefits for both patients and researchers. Without sustained investment in these areas, heterogeneity in representation is likely to persist, even in the presence of shared standards and a common semantic framework.

Conclusion

This study presents the first nationwide, metadata-driven assessment of SNOMED CT and LOINC adoption in Swiss clinical routine data, drawing on metadata from approximately 800 000 patients collected across six university hospitals. Our findings confirm adoption of international semantic standards within the SPHN FedData, with SNOMED CT and LOINC together accounting for 87% of coded instances. However, limited code overlap across hospitals reveals substantial heterogeneity that restricts cross-institutional comparability. Although millions of coded instances are available, the comparatively smaller number of distinct coded values suggests that semantic coverage remains concentrated within a stable core set of concepts, with opportunities for further diversification and refinement.

These findings provide an evidence-based foundation for future harmonization efforts, including targeted feedback to participating hospitals and refinement of internal coding practices and guidance. Achieving true semantic interoperability will require improved standardization at the point of care, supported by robust semantic mapping and validation mechanisms. As Switzerland advances its digital health agenda through initiatives such as DigiSanté, the monitoring of terminology adoption will be critical to ensuring that clinical data are research-ready, interoperable, and fit for secondary use.

List of Abbreviations

ATC, Anatomical Therapeutic Chemical Classification System

CHOP, Swiss classification of surgical interventions

CHUV, Centre Hospitalier Universitaire Vaudois

DigiSanté, Swiss national digital health initiative

EHDS, European Health Data Space

EHR, Electronic Health Record

FAIR, Findable, Accessible, Interoperable, Reusable

FOPH, Federal Office of Public Health

FSO, Federal Statistical Office

HL7 FHIR, Health Level 7 Fast Healthcare Interoperability Resources

HUG, Hôpitaux Universitaires de Genève

ICD-10 GM, International Classification of Diseases, specifically the German modification

INSEL, Inselspital, Bern University Hospital

KISPI, Kinderspital Zürich (University Children's Hospital Zurich)

LOINC, Logical Observation Identifiers Names and Codes

RDF, Resource Description Framework

SPHN, Swiss Personalized Health Network

SPHN FedData, SPHN Federated Clinical Routine Dataset

SNOMED CT, Systematized Nomenclature of Medicine - Clinical Terms

UK, United Kingdom

USB, Universitätsspital Basel (University Hospital Basel)

USZ, Universitätsspital Zürich (University Hospital Zurich)

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the SPHN Metadata Catalog [45]. A snapshot of the code frequencies used for this analysis has been archived on Zenodo to ensure reproducibility [40].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the State Secretariat for Education, Research, and Innovation (SERI) via SPHN, a national data infrastructure initiative mandated by the Swiss Confederation and coordinated by the Swiss Academy of Medical Sciences (SAMS) and SIB Swiss Institute of Bioinformatics.

Authors' contributions

Study design: SOE, VT, Data Analysis: VT, JA, Methodology: VT, JA, SOE, Manuscript drafting: VT, JA, SOE. All authors reviewed and approved the final version.

Acknowledgments

We are grateful to the SPHN community for their collaborative work over the years on semantic harmonization, data standards use, and FAIR-aligned infrastructure, which made this analysis possible. In particular, we thank all university hospital teams that contributed to the coding of data, as well as those involved in the generation of the SPHN FedData. We also thank the BioMedIT team, the SPHN ELSI (Ethical, Legal, and Social Implications) team, the SPHN National Data Stream, and the SPHN Semantic, RDF, and RDF Support working groups for their respective contributions, including legal frameworks, IT infrastructure, user support, and semantic development. We also acknowledge the support of eHealth Suisse for promptly incorporating SNOMED CT codes requested in collaboration with hospitals and NDS projects. In addition, we thank the patients whose willingness to allow the use of their data for research enabled this work.

References

[1] Benson T, Grieve G. Why interoperability is hard. Principles of Health Interoperability: FHIR, HL7 and SNOMED CT, Springer; 2020, p. 21–40.

[2] ICD-10-GM n.d. https://www.bfarm.de/EN/Code-systems/Classifications/ICD/ICD-10-GM/_node.html.7 (accessed April 25, 2025).

[3] ATCDDD - Home n.d. https://atcddd.fhi.no (accessed April 25, 2025).

[4] Benson T, Grieve G. SNOMED CT. In: Benson T, Grieve G, editors. Principles of Health Interoperability: SNOMED CT, HL7 and FHIR, Cham: Springer International Publishing; 2016, p. 155–72. https://doi.org/10.1007/978-3-319-30370-3_9.

[5] LOINC n.d. https://loinc.org/ (accessed December 16, 2025).

[6] McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem 2003;49:624–33.

[7] Millar J. The need for a global language–SNOMED CT introduction. Nursing Informatics 2016 2016:683–5.

[8] SNOMED CT International n.d. https://www.snomed.org (accessed April 25, 2025).

[9] Cornet R, de Keizer N. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 2008;8:S2.

[10] Chang E, Mostafa J. The use of SNOMED CT, 2013-2020: a literature review. Journal of the American Medical Informatics Association 2021;28:2017–26.

[11] The European Health Data Space n.d. https://www.european-health-data-space.com/ (accessed May 22, 2026).

[12] Braunstein ML, Braunstein ML. Health Informatics on FHIR: How HL7's API is Transforming Healthcare. Springer; 2022. https://doi.org/10.1007/978-3-030-91563-6.

[13] Strategie eHealth Schweiz 2.0. 2018–2022 2018. https://www.bag.admin.ch/dam/de/sd-web/NFhHTfgiuzbB/181214_Strategie-eHealth-Suisse-2.0_d.pdf (accessed February 12, 2026).

[14] eHealth Suisse n.d. https://www.e-health-suisse.ch/en (accessed March 3, 2026).

[15] 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.

[16] Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.

[17] DigiSanté n.d. https://www.digisante.admin.ch/de (accessed March 3, 2026).

[18] Vuokko R, Vakkuri A, Palojoki S. Systematized nomenclature of medicine–clinical terminology (SNOMED CT) clinical use cases in the context of electronic health record systems: systematic literature review. JMIR Med Inform 2023;11:e43750.

[19] Agrawal A, He Z, Perl Y, Wei D, Halper M, Elhanan G, et al. The readiness of SNOMED problem list concepts for meaningful use of electronic health records. Artif Intell Med 2013;58:73–80. https://doi.org/https://doi.org/10.1016/j.artmed.2013.03.008.

[20] Lee DH, Lau FY. An exploratory analysis of SNOMED CT national editions. Journal of the American Medical Informatics Association 2026;33:383–93. https://doi.org/10.1093/jamia/ocaf184.

[21] Kim S, Shin S-Y, Hwang JE, Park H-A. Current status of SNOMED CT national extensions and terminology managements. MEDINFO 2023—The Future Is Accessible: Proceedings of the 19th World Congress on Medical and Health Informatics, SAGE Publications 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP; 2024, p. 1345–6.

[22] SNOMED CT - Members n.d. https://www.snomed.org/members (accessed June 5, 2026).

[23] Vorisek CN, Klopfenstein SAI, Löbe M, Schmidt CO, Mayer PJ, Golebiewski M, et al. Towards an Interoperability Landscape for a National Research Data Infrastructure for Personal Health Data. Sci Data 2024;11:772. https://doi.org/10.1038/s41597-024-03615-3.

[24] eHealth Suisse Strategie eHealth Schweiz 2.0 Schlussbericht zur Umsetzung – Mai 2025 n.d. https://www.gdk-cds.ch/fileadmin/docs/public/gdk/themen/ehealth/2025_Schlussbericht_Strategie_eHealth_2.0.pdf (accessed June 5, 2026).

[25] NHS - SNOMED CT n.d. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct (accessed June 5, 2026).

[26] Semler SC, Boeker M, Eils R, Krefting D, Loeffler M, Bussmann J, et al. Die Medizininformatik-Initiative im Überblick–Aufbau einer Gesundheitsforschungsdateninfrastruktur in Deutschland. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 2024;67:616–28.

[27] Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology n.d. https://www.federalregister.gov/documents/2010/07/28/2010-17210/health-information-technology-initial-set-of-standards-implementation-specifications-and (accessed June 10, 2026).

[28] Park H-A, Yu S-J, Jung H. Strategies for adopting and implementing SNOMED CT in Korea. Healthc Inform Res 2021;27:3–10.

[29] De Mello BH, Rigo SJ, Da Costa CA, da Rosa Righi R, Donida B, Bez MR, et al. Semantic interoperability in health records standards: a systematic literature review. Health Technol (Berl) 2022;12:255–72.

[30] Chu L, Kannan V, Basit MA, Schaeflein DJ, Ortuzar AR, Glorioso JF, et al. SNOMED CT concept hierarchies for computable clinical phenotypes from electronic health record data: comparison of intensional versus extensional value sets. JMIR Med Inform 2019;7:e11487.

[31] Willett DL, Kannan V, Chu L, Buchanan JR, Velasco FT, Clark JD, et al. SNOMED CT concept hierarchies for sharing definitions of clinical conditions using electronic health record data. Appl Clin Inform 2018;9:667–82.

[32] Vreeman DJ, McDonald CJ, Huff SM. LOINC®: a universal catalogue of individual clinical observations and uniform representation of enumerated collections. Int J Funct Inform Personal Med 2010;3:273–91.

[33] 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.

[34] Armida J, Touré V, Krauss P, Unni D, Witte H, Chiarugi D, et al. Semantic Interoperability at National Scale: The SPHN Federated Clinical Routine Dataset. Res Sq 2025. https://doi.org/10.21203/rs.3.rs-8250886/v1.

[35] Witte H, Unni DR, Krauss P, Touré V, Armida J, Österle S. The SPHN Metadata Catalog: A platform for health data discovery and exploration based on FAIR principles (Preprint). JMIR Preprints 2025. https://doi.org/10.2196/preprints.90146.

[36] The SPHN RDF Schema 2025.1 n.d. https://www.biomedit.ch/rdf/sphn-schema/sphn/2025/1 (accessed June 4, 2026).

[37] SPHN Metadata Catalog - FAIR Data Point n.d. https://fdp.dcc.sib.swiss/ (accessed June 4, 2026).

[38] SPHN Metadata Catalog - SPARQL Endpoint n.d. https://fdp.dcc.sib.swiss/store/fdp/sparql (accessed June 4, 2026).

[39] SNOMED CT - LOINC Analysis n.d. https://git.dcc.sib.swiss/sphn-semantic-framework/outreach/snomed-loinc-analysis (accessed June 12, 2026).

[40] Touré V, Armida J, Witte H, Unni D, Kalt K, Endrich O, et al. SPHN Federated Clinical Routine Dataset Metadata 2026. https://doi.org/10.5281/zenodo.20637770.

[41] SPHN FedData - ReadTheDocs n.d. https://sphn-semantic-framework.readthedocs.io/en/latest/sphn_framework/sphn_feddata.html (accessed May 26, 2026).

[42] SNOMED CT Expressions - Precoordination and Postcoordination n.d. https://docs.snomed.org/snomed-ct-practical-guides/snomed-ct-postcoordination-guide/snomed-ct-expressions/precoordination-and-postcoordination (accessed June 10, 2026).

[43] SNOMED CT Machine Readable Concept Model n.d. https://docs.snomed.org/snomed-ct-specifications/snomed-ct-machine-readable-concept-model (accessed May 20, 2026).

[44] Ehrsam J, Gaudet-Blavignac C, Mattei M, Baumann M, Lovis C. Semantics in action: a guide for representing clinical data elements with SNOMED CT. J Biomed Semantics 2025;16:7. https://doi.org/10.1186/s13326-025-00326-5.

[45] SPHN Federated Clinical Routine Dataset Catalog n.d. https://fdp.dcc.sib.swiss/catalog/fc6e936d-ad4b-5b21-b8a4-8c7826bbec99 (accessed June 12, 2026).

Additional Files

Additional File 1

File name: Additional file 1

File format: .XLS

Title: SNOMED CT value sets defined in the SPHN RDF Schema

Description: For each SPHN concept (column ‘SPHN Concept Name’) and concept property (column ‘Property’), this table lists the corresponding SNOMED CT value set specified in the SPHN RDF Schema (column ‘SNOMED CT value set’). These value sets define the SNOMED CT concepts allowed for annotating values of the respective concept properties. Value sets are specified either as explicit lists of SNOMED CT concepts, as hierarchical subsets using ‘descendant-of’ expressions, or as unrestricted SNOMED CT concepts sets (i.e., any SNOMED CT concept is permitted). Properties without SNOMED CT value set bindings are not included. Note that in column ‘Property’, an arrow (‘->’) indicates a nested property to which the value set

841 defined applies. For example, for the concept ‘Electrocardiographic Procedure’, the property
842 ‘hasCode’ nested within ‘hasSubjectPhysiologicalState’ has ‘resting state’ and ‘exercise state’
843 as allowed SNOMED CT values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.xlsx