SPHN Working Group on Data Standards and Quality Control (DASAQ WG)

Version and date: Approved Version (V4), 3 February 2022

Starting position

The Swiss Personalized Health Network (SPHN) promotes the development, implementation, and validation of coordinated infrastructures to make health data interoperable and shareable for research in Switzerland. As part of the SPHN initiative, various efforts are currently underway to harmonize and define data standards in order to ensure the interoperability of health-related data, with a specific focus on data from the five Swiss University Hospitals (UH). In the first funding phase, the SPHN Data Coordination Center (DCC) and the SPHN partners started the development of several infrastructure elements, among them:

- Hospital specific infrastructures to provide clinical data;
- Semantic definition of (core) data elements, introducing meaning binding of concepts and value sets[1] to internationally acknowledged controlled vocabulary;
- Introduction of RDF as data exchange format with related SPHN schema definition, which brings numerous advantages regarding the FAIRification of data, e.g. the machine-readability of controlled vocabulary through standard-specific unique resource identifiers;
- Data specification and harmonization of generally consented data (from 2015 onwards) for the setup of a Federated Query System (FQS), currently containing over 75 million data entries of 5 UH.

From the currently running SPHN Driver Projects and the work regarding the FQS as well as the SPHN semantic interoperability framework[2], SPHN has identified a series of critical issues, among them the need for more standardization (i.e. introduction of controlled vocabulary at the sources) and quality assurance on the side of the data providing institutions.

In particular, the following challenges and gaps have been identified and need to be tackled:

- Data coded in local codes are often not usable without significant efforts for researchers and numerous callbacks to the data providers;
- LOINC and UCUM have not been introduced at the source (i.e. the main labs in the hospitals), but mapped in the CDWs – a process that is error prone and potentially leading to false codes as soon as changes in the laboratories are introduced without communication to the CDWs;
- There is a lack of resources and only minimal benefits for the clinical routine process (i.e. the cost-benefit ratio not good enough to bind clinical resources) at the UH to systematically introduce SNOMED

[1] "A meaning binding is a terminology binding that represents the clinical meaning of a data item or collection of data items." https://confluence.ihtsdotools.org/display/DOCGLOSS/terminology+binding

[2] https://sphn.ch/network/data-coordination-center/the-sphn-semantic-interoperability-framework/

CT codes, so coding is done in a project specific manner and cannot be re-used for other projects/purposes;
- The design and efficiency of quality control pipelines is different in every UH, resulting in frequent feedback due to data quality deficiencies from the research side; also, quality assurance processes should link back to the CDWs and other hospital IT system, in order to impact also the quality of data used for future projects.

Vision and mission

The envisioned processes to introduced standards and to a basic quality assessment should not be geared exclusively to the needs of SPHN and related research projects, but should also cover internal needs of the UH (e.g. own research) and local research initiatives (such as The Loop in Zurich), as well as additional cases where hospitals are obliged to provide data to external parties (e.g. the feeding of registers, public health data collections, data deliveries to authorities, etc.).

Therefore, with regards to the standardization and basic quality control processes, the "once-only" principle should be the most important premise. Standardization and coding of data should happen as early as possible in the processing chain in order to be universally usable (for research and other purposes).

Data scope

Core concepts (according to SPHN dataset[3]):

- Demographics: Data Provider Institute, Administrative Case, Healthcare Encounter, Subject Pseudo Identifier, Administrative Gender, Birth Date, Death Status, Consent
- Clinical Measurements: Heart Rate, Systemic Arterial Blood Pressure, Body Temperature, Oxygen Saturation, Central Venous Pressure, Circumference Measure (e.g. head, hip, waist or neck), Height, and Weight.
- Diagnoses: Problem Condition (clinical condition, problem, diagnosis, or other event, situation, issue, or clinical concept that has risen to a level of concern in coded form), FOPH Diagnosis, Nursing Diagnosis, ICD-O Diagnosis, and Oncology Treatment Assessment.
- Procedures: FOPH Procedure, Radiotherapy Procedure, and Catheter.
- Medications: Drug Administration Event, Drug Prescription, and Allergy.
- Adverse events
- Lab Results: Clinical Chemistry, Hematology, Immunology, Microbiology, and Virology.
- Pathology: TNM Classification, Tumor Specimen, Tumor Stage and Tumor Grade.
- Biobanking: Sample availability information "yes/no", Biosample.

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

The DASAQ WG will discuss the time period, the prioritization, and how far back to include the data and make a proposal to the HIT-STAG for adoption.

Tasks and timeline of the Working Group

The timeline for the WG to complete the mandated tasks is 1 year (start of the WG: March 2022).

[1] The WG defines a concept how to measure the current data situation in the CDW of all data related topics per data domain within D1-D4 (e.g. in the routine lab data, the hospital has n Lab values per year, with n unique internal Lab codes. Out of these unique lab codes, n Lab values with n unique Lab codes are coded in LOINC, n Lab values are coded in L4CHLAB).

[2] The WG defines the target situation for each single data related topic per data domain within D1-D4 (e.g. Out of the top 300 used Lab codes, 90% are mapped to LOINC).

[3] The WG identifies the gap from current situation to the target situation and defines work packages within the hospital to close the gap.

[4] The WG defines the roadmap of the identified work packages above to be approved by the HIT-STAG.

[5] The WG supervises the implementation of the defined work packages within the hospital and reports the progress in the realm of the WG meetings. For the progress monitoring, the same KPI's as defined in [1] shall be used and compared to the target situation in [2] and be reported every 3 months to DCC as well. For quality assessment purposes, the WG determines a process for the delivery of sample data to BioMedIT.

[6] The WG defines and documents the process of maintenance of each data related topic per data domain within D1-D4

Deliverable 1: Standardized Lab tests and results in the CDW (LOINC/L4CHLAB and UCUM)

This should include [data related / process related]:

- Implementation of LOINC/L4CHLAB for lab test as defined in the CA 2021
- Definition of a process to introduce and maintain LOINC codes for lab analyses not covered in the L4CHLAB list as well as to extend the list for Clinical chemistry, Hematology, Immunology, Microbiology, and Virology (according to the HospFAIR program)
- Implementation of LOINC codes for lab tests of the Clinical chemistry, Hematology, Immunology, Microbiology, and Virology at the source or in the CDW
- Definition and implementation of a standardized coding (e.g LOINC Answer Lists) for qualitative lab results
- Definition of basic pre-processing needs for a correct LOINC mapping (e.g. combine information sampling time and lab result) and implement at the source or in the CDW
- Definition of basic pre-processing needs for units to adapt a Swiss-wide harmonized UCUM implementation and implement in the CDW

- Definitions of quality control measures for the laboratory test and results including outlier detection and unit of measure

Deliverable 2: Standardized medication details are available in the CDW

This should include [data related / process related]:

- Implementation of ATC as main standard for active ingredients of a drug as defined in the CA
- Definition of basic pre-processing needs (including transformation of a drug product dose [e.g. IRFEN Lactab 200mg 2 pills] to dose of active ingredient [ATC code: M01AE01, 400mg]) and implementation in the CDW[4]
- Definition of basic pre-processing needs for units to adapt a Swiss-wide harmonized UCUM implementation for medication doses and implementation in the CDW
- Definition of quality control measures for the medications including active ingredients doses and units

Deliverable 3: Value sets and values coded in SNOMED CT (according to SPHN Dataset) in the CDW

This should include [data related / process related]:

- Definition and implementation of a mapping for the value sets and values of the concepts defined in the HospFAIR program description (see data scope)
- Development and implementation of a cross-institutional peer-review process for the mapping

Deliverable 4: SPHN concepts (not covered in D1-D3) available according to SPHN standards in the CDW

- Definition and implementation of the mapping of local concepts data to the SPHN concepts/standards
- Implementation of the mapping from local value set to SPHN values sets
- Implementation of the mapping of local data to the SPHN recommended standard (not part of D1-D3, e.g., MedRA for adverse events, UCUM for units of vital signs, etc.)

[4] The necessity for pre-processing steps at the hospitals can be justified as follows: (i) the calculations need to be performed only once and not by each research project individually; (ii) the information needed for the calculation is available in the hospital, but often not to the researcher (licensing issue); (iii) substance amounts are need for the Federated Query System (next step) and potentially other use cases.

Composition of the Working Group

The HIT-STAG nominated the following representatives of the UH:

- Insel: Alex Leichtle (a.i.)
- HUG: Deniz Geres
- CHUV: Solange Zoergiebel
- USZ: Amanda Ramirez Ramos
- USB: Fabian Franzeck

The DCC will provide project management support and quality assurance (up to 0.3 FTE). After approval of the implementation roadmap by the HIT-STAG, the working group will provide a follow-up plan for the coordination of the implementation and maintenance processes. The WG will be chaired by the DCC; Stephan Meier (from IT-Logix) will be responsible for the project management and co-chairing the WG.