Strategy paper
Clinical Data Semantic Interoperability

1 Introduction

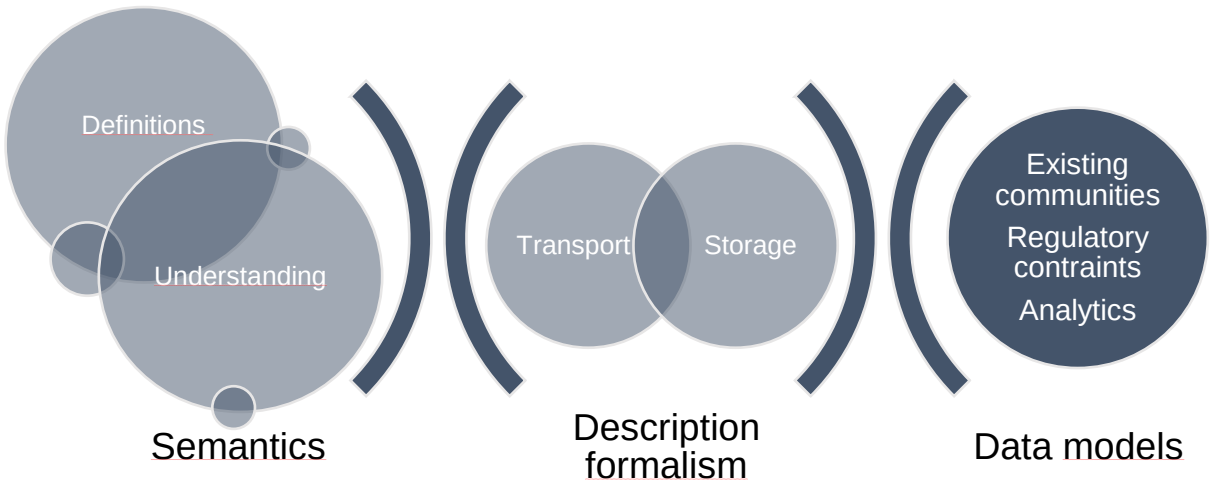
The aim of this document is to describe the overall strategy of building a Swiss wide clinical data semantic interoperability framework for the SPHN initiative. It is based on existing work from standardizing initiatives and organizations [1] and countries initiatives [2].

2 Vision and mission statement

The Clinical Data Semantic Interoperability (CSI) Working Group (WG) develops a semantically driven framework to enable the use of health related data for research purposes. This framework will empower a data driven approach in the field of Personalized Health research.

To achieve these goals, the CSI WG's work is based on a "three pillar" strategy:

1. Build a strong semantic layer that leverages interpretability of health related data by supporting data/object composition;
2. Achieve a highest possible degree of flexibility for data use by implementing a data model-agnostic transport and storage method, such as applying a formal descriptive language, e.g. Resource Description Framework (RDF);
3. Based on the needs of the research communities and requirements of research projects different data models (such as OMOP, CDISC, i2b2, etc.) can be used.



Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

The Personalized Health Informatics Group manages
the SPHN Data Coordination Centre and the BioMedIT project

3 Background

3.1 Swiss Federal Shared Patient Record - landscape

Since 2005, the Swiss Federal Council has started an eHealth Strategy that required more than 10 year's work to be finalized in a new federal law about the Swiss shared patient record, and the associated regulation, that is enforced since Spring 2017. In this work, a strong involvement of the Swiss Society of Medical Informatics has led to the adoption of international standards to insure appropriate technical processes, under the Integrating Healthcare Enterprise (IHE) framework [3,4]. Standards supporting distributed systems have been selected to comply with the Swiss regulatory environment. In addition, the use of HL7 Clinical Document Architecture (CDA) documents allows a progressive adoption of structured information, starting with low structure and PDF documents, up to highly granular structured information [5]. All documents, source code of the eHealth Swiss connectors, test bench, etc. are freely available on the national e-health Suisse portal [6,7]. Since early 2017, there is also a national license to use SNOMED-CT and all its tools in Switzerland, as it is one of the major systems that has been selected to improve clinical data interoperability.

The national eHealth Suisse Coordination organ is hosting the "Swiss Competence and Coordination Centre of the Confederation and the Cantons - eHealth Suisse Exchange Formats and Semantics" in charge, amongst others, to coordinate all relations with SNOMED INTERNATIONAL.

3.2 Swiss Personalized Health Network (SPHN) - landscape

SPHN is an initiative of the Swiss federal government, namely the State Secretariat for Education, Research and Innovation (SERI) and the Federal Office of Public Health (FOPH). The goal is to collaboratively advance Personalized Health research and innovation for the benefit of society. The Swiss Academy of Medical Sciences (SAMS) and the Swiss Institute of Bioinformatics (SIB) are responsible for the implementation of the mandate.

A total of CHF 68 million was allocated to the initiative for the period 2017-2020 to support the development and implementation of coordinated infrastructures in several ways:

1. Compatible data management systems are developed in the five University Hospitals through collaboration agreements.
2. Infrastructure development projects develop and test new technologies, methods and infrastructures at single or joint sites, to be made available to other institutions after proof of concept phase. Infrastructure development projects most relevant to semantic interoperability are NLP-powered mapping of clinical reports onto SNOMED-CT concepts for tumor classification (NLPforTC) and LOINC for Swiss laboratories (L4CHLAB).
3. Driver projects guide the development of SPHN by "test driving" the infrastructures and interoperability for multi-site research in a specific area or pathology. Each driver project typically involves multiple data providers as well as teams of data recipients who analyze the data.

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

The Personalized Health Informatics (PHI) group at SIB is responsible for the coordination and implementation of several key elements of SPHN, such as the BioMedIT project and the SPHN Data Coordination Centre (DCC). In addition, the PHI group coordinates the DCC working groups – among them the CSI WG – and facilitates information flow between the working groups, SPHN projects and SPHN partners and collaborators. A second funding period of SPHN (2021-2024) is planned.

4 Guiding principles

The CSI WG focuses on the semantic definition of health related concepts.

4.1 Concepts

Semantics approaches span from extremely conceptual ontology-driven knowledge representations up to data-type centered definitions. The goal is to find the right level between abstraction and granularity to optimize the power of expression. The CSI is developing a generally applicable framework that can be effectively used to exchange data in Switzerland.

The approach is based on conceptual definitions. A property of this approach is compositionality. A concept can be defined based on other concepts. In this case we speak about composed concepts. Usability is preferred over conceptualization.

4.2 Interrelations

Semantic definition of a concept is independent from context, transport, model.

This can be illustrated by medication and the concept “DRUG”.

There are many different possible understandings of a DRUG. In most models, drugs are seen as active ingredients that are given to patients to treat some conditions. Most models do have this understanding, such as the HL7 RIM model. However, DRUG can also be seen as a product, that as a chemical/biological composition. It can be seen as a logistical product, that must be ordered, stored, that generates costs, profits. It can also be seen as an allergy, as a contra-indication, etc. If one looks at all the possibilities, it is clear that DRUG must be usable in all these contexts. Thus, it is important to have a way to identify clearly a DRUG, and, separately, each context of use. For example, the definition of DRUG should be independent from the definition of the prescription context. The prescription context should refer to DRUG and add all elements required to characterize a prescription, such as case, indication, date and time, dosage, route, etc.

The concepts required to describe a context requires usually also a semantical definition in order to ensure interoperability.

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

4.3 Semantic representation must be generalizable

This can be illustrated by laboratory.

Laboratory medicine covers an extremely large range of potential data and information flows. Some are appropriately standardized; others, especially advanced technologies, are still missing standardized approaches. In all cases, laboratory medicine considers several important determinants of the quality and interpretability of the analysis results, such as pre, per and post analytical phases. In some cases, results can be well defined, such as the concentration of specific substances in a sample, in others, the results are less well defined, such as the measurement of “bacterial oro-fecal contaminations”. On one hand, the concepts to capture the measurement could be of the highest granularity, where each concept defines uniquely and precisely one unique type of measurement. On the other hand, and at the other extreme, concepts could be very generic, i.e. one concept is able to support any type of measurement, what requires to specify the measurement type itself as additional information. The latter is illustrated in the HL7 framework, which builds strong compatibility with generic concepts, but at the cost of interoperability. Technical compatibility does not provide semantic interoperability. In the CSI, a middle road is chosen. A generalization is made within a restricted domain that share a common semantics. For example, rather than building a concept for each lab analysis, a generic concept made of all what is required to define a specific lab analysis can be built, so that extending the list of lab supported doesn’t require to build new concepts, but only to extend the values supported by the generic concept, for example LOINC codes. In summary, it allows to define one concept in the SPHN semantic dataset, and then extend the source dictionary with all laboratory analyses that are needed and that are properly codable using LOINC in the Swiss context. Moreover, it allows to build the mapping and the consensus between lab providers, which is time costly, only for those lab analyses that are needed, and existing in LOINC.

4.4 Use of existing standards

Existing standards are used, except if otherwise required. International standards are preferred, such as SNOMED CT for clinical terms [8,9], MeSH for bibliographic terms [10], LOINC for laboratory [11,12], NANDA for nursing [13], ICD for diagnosis encoding, etc. In specific cases, national standards are used, such as CHOP surgical procedures.

4.5 Semantic sources and definition

Currently, the semantic sources and definition available from the DCC are built of three major elements:

1) Resources reference. Concepts can rely on resources such as coding systems, ontologies, classifications, or scores. Each resource is uniquely identified at the SPHN Xchange space.
2) Concepts. Each concept is uniquely identified and is persistent. Each concept which is tied to a resource will carry a link to this reference.
3) SPHN semantic datasets. The SPHN semantic datasets are periodically released and undergo a lifecycle management. Each SPHN semantic dataset release is uniquely identified.

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

5 Organization and processes

5.1 Responsibilities and organizational interfaces

The CSI WG is working on behalf of the DCC and has the mandate to specify the structure of the SPHN semantic datasets and to define the content by leveraging domain knowledge from the SPHN community, e.g. driver projects. The DCC is managing the SPHN semantic datasets. The SPHN semantic datasets are endorsed periodically by the national steering board (NSB). All involved stakeholders can provide feedback through their CSI WG representatives.

The concepts are prioritized based on the University Hospitals – SPHN collaboration agreement requirements, building a Swiss wide distributed query system to support feasibility studies, and on the requirements and needs of the SPHN driver projects.

5.2 Deliverables

- CSI Strategy paper
- List of concept definitions
- User Guide

5.3 Life cycle management

The list of concept definitions and the user guide underlie a version management. The list of concept definitions is published periodically and content modifications are documented in a change history.

5.4 Collaboration between CSI WG and driver projects

a) Definition of representatives of each SPHN Project

The SPHN projects identify at least some participants in charge of defining the semantics of concepts required by the project, the Project Representative (ProRep).

The CSI WG identifies a member of the CSI WG to be the CSI contact person for the project, the CSI Representative (CSIRep).

b) Building understanding.

The ProRep and the CSIRep get in contact in order to clarify mutual requirements and understandings.

c) Defining the required set of concepts

The ProRep defines the needs of the project in terms of concepts, by taking into account the existing SPHN semantic datasets. The ProRep and the CSIRep build the list according to CSI working rules and following the CSI Semantic dataset formalism. The ProRep delivers the CSIRep a first proposal with all concepts ranked according a 3-levels ranking. This 3-levels ranking is specific to each driver project and reflects the priorities of that driver. It will be used only to ease time and resource management of the CSI.

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

d) Aligning requirements with existing SPHN semantic datasets

The ProRep and the CSIRep align the needs with the existing SPHN semantic datasets and provide a gap analysis of the existing concepts not covering needs, or the need for new concepts.

6 Building capacity

A major challenge in the Swiss landscape is a a) severe lack of capacity in the field of semantic interoperability in health at the crossing between b) the healthcare system and the research community. It is thus of prime importance to improve that capacity.

While the movement around personalized medicine has initiated a positive drive of the public funding and superior education organizations in Switzerland, for most the reality of clinical semantic interoperability remains an uncovered field. Thus, there is a need for targeted and specific actions by the DCC and the CSI WG to leverage the current drive and improve capacity in the field.

Building capacity

- Increase competences within SPHN in medical informatics
- Increase competence in semantics around clinical information

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland

7 References

Online information in the SPHN DCC CSI WG Xchange space

https://xchange.dcc.sib.swiss/display/SPHNDCC/WG+I%3A+Clinical+Data+Semantic+Interoperability

Other SPHN documents:

I. SPHN clinical datasets

II. L4CHLAB dataset and white paper

III. DCC working group Hospital IT strategy

1. OMOP Common Data Model – OHDSI [Internet]. [cited 2018 Sep 8]. Available from: https://www.ohdsi.org/data-standardization/the-common-data-model/
2. Dugas M. Portal of Medical Data Models (MDM-Portal) [Internet]. [cited 2018 Sep 8]. Available from: https://medical-data-models.org/
3. Integrating the Healthcare Enterprise (IHE) [Internet]. IHE International. [cited 2018 Sep 9]. Available from: https://www.ihe.net/
4. Holzer K, Gall W. Utilizing IHE-based Electronic Health Record systems for secondary use. Methods of information in medicine 2011;50(4):319–325. [doi: 10.3414/ME10-01-0060]
5. HL7 Standards Product Brief - CDA® Release 2 [Internet]. [cited 2018 Sep 9]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7
6. Bern eHealth S 3003. eHealth Suisse - Deutsch [Internet]. [cited 2018 Sep 9]. Available from: https://www.e-health-suisse.ch/de/startseite.html
7. Bern eHealth S 3003. eHealth Suisse - français [Internet]. [cited 2018 Sep 9]. Available from: https://www.e-health-suisse.ch/fr/page-daccueil.html
8. SNOMED International [Internet]. [cited 2018 Jun 27]. Available from: https://www.snomed.org/snomed-ct/snomed-ct-worldwide/translations-of-snomed-ct
9. Randorff Hojen A, Rosenbeck Goeg K. Snomed CT implementation. Mapping guidelines facilitating reuse of data. Methods of information in medicine 2012;51(6):529–538. [doi: 10/gf2cfj]
10. MeSH Browser [Internet]. Bethesda (MD): National Library of Medicine (US). [cited 2015 Feb 23]. Available from: http://www.ncbi.nlm.nih.gov/mesh/68008490 MeSH Unique ID: D008490
11. LOINC — The freely available standard for identifying health measurements, observations, and documents. [Internet]. [cited 2018 Aug 1]. Available from: https://loinc.org/
12. Using LOINC with SNOMED CT - Using LOINC with SNOMED CT - SNOMED Confluence [Internet]. [cited 2018 Aug 1]. Available from: https://confluence.ihtsdotools.org/display/DOCLOINC/Using+LOINC+with+SNOMED+CT
13. Welcome to NANDA International Defining the Knowledge of Nursing | Just another WordPress site [Internet]. [cited 2019 Oct 12]. Available from: http://www.nanda.org/

Swiss Institute of Bioinformatics (SIB)
Personalized Health Informatics Group (PHI)
Elisabethenstrasse 43
CH-4063 Basel, Switzerland