SPHN Data Coordination Centre
Clinical Data Semantic Interoperability Working Group
Strategy paper

1 Introduction

This document has been developed within the Clinical Data Semantic Interoperability Working Group (CDS-IOP WG) of the Data Coordination Center (DCC) built within the Swiss Personalized Health Network (SPHN) initiative. The aim of this document is to describe the overall strategy of building a clinical data semantic interoperability framework in the SPHN initiative.

There have been numerous initiatives worldwide to build clinical data semantic interoperability in a distributed and heterogeneous clinical data acquisition landscape. Some of them are used in the work of the CDS-IOP such as:

- The OMOP Common Data Model, that allows for the systematic analysis of disparate observational databases (1);
- The i2b2 data model, that is a cohort selection tool aggregating inpatient data providers (2);
- The HL7 Reference Information Model (3) and the HL7 FHIR variables set (4);
- Several European projects, such as FP7 DebugIT (5); FP7 SemanticHealthNet (6); IMI EHR4CR (7);
- CDISC, an organisation developing data standards to promote clinical research and interoperability in healthcare (8)

Some specific country-based initiatives have been developped, such as:

- Germany: the Medical Data Model (MDM) portal at the university of Münster (9); The German Medical Informatics Initiative (10);
- The Netherlands: Health and Care Information Models for the standardisation of patient information for multiple usage from the National competence centre eHealth in the Netherlands (11)

The aim is that a pragmatic, meaning feasible, but sustainable way is the main guiding principle of the methodology. The approach is pragmatic in that the selection of the variables is based directly on the needs of the SPHN needs; it is sustainable, it tries to build stepwise using existing standards and state-of-the art knowledge representation.

Guiding principle
- Pragmatism
- Feasibility
- Sustainability

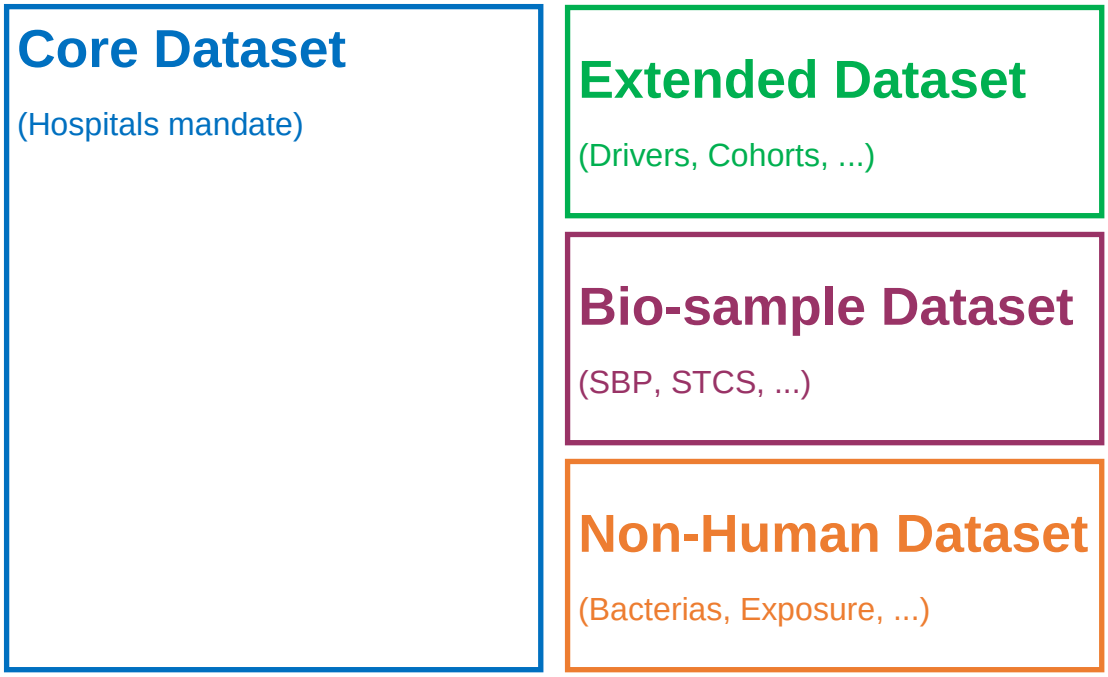
2 Datasets

The pragmatic approach leads to start with datasets requirements instead of designing or choosing top-down data models.

The datasets have lifecycle, that is versioning, and will progressively evolved and be enriched. The succinct description below is only for information and pertains to the first release.

The datasets have provisory been divided in four major groups:

1- The core dataset is the one that is considered by the mandate to the Swiss university Hospitals and is devoted to support distributed queries for feasibility studies. It is mainly covering demographics and administrative data such as encounters, basic clinical elements such as weight, simple information about medication, routine and emergency laboratory, procedure and diagnosis has encoded at discharge.
2- The extended dataset is the one that is starting to cover the needs of specific projects, and primarily the SPHN drivers' projects. It is currently made by the five SPHN drivers projects of the 2017 SPHN call. It contains more extensive information, such as scores.
3- The bio-sample dataset will be devoted to cover the needs of SBP. It is currently being analyzed starting with liquids and tissues.
4- The non-human dataset is devoted to non-human elements, such as pathogens, environments, exposition factors, etc.



There is no overlap between the datasets, meaning that the core dataset must be used when either the extended or the bio-sample datasets are used. Of course, any primary data producer can produce the core dataset, which is not restricted to the hospitals mandate, this has only been used as a definition purpose.

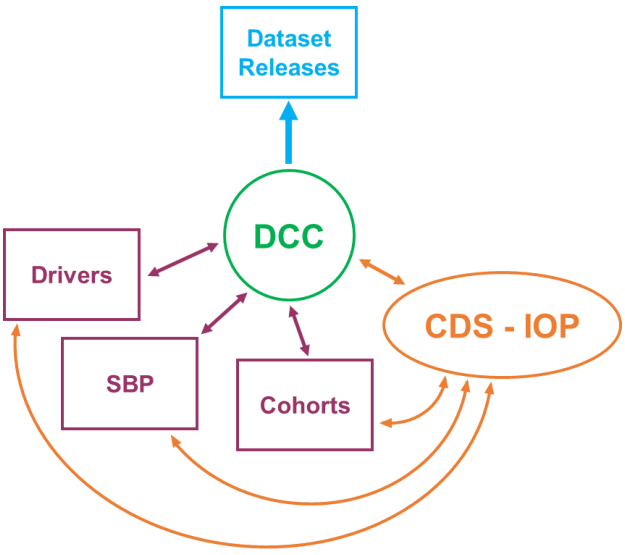
This is an operational way of looking at datasets, and allows other simultaneous organizations, such as by projects, pathologies, datatypes, etc. In these organizations, part of the datasets as described might be involved.

2.1 Building the datasets

The DCC is leading the process of building the datasets. All requirements are sent to DCC which manages the decision process about the needs and prioritization of including the new variables into a dataset release.

The DCC works in tight collaboration with the projects requiring extensions of the datasets. Each proposal will be asked to consider the global organisation of the datasets in their demands. It will also be required to consider ranking all variables according three levels:

The A Level	The A level is reserved for variables that have a “global” impact. In a SPHN Driver project, that would mean that this variable would be used for all patients and all participating centers at least, preferably be of use for other projects.
The B Level	The B level is reserved for variables that have a “regional” impact. In a driver project, such a variable would either be are specific to a center, or a group of patients, or strictly limited to that driver project.
The C level	The C level are for variables that have a “local” impact. In a driver project, that would mean that such a variable is specific for a sub project of that driver, or that would fit only a limited scope within the overall consideration of datasets.



All variables will be considered. The ranking in A, B and C level is only used in this early phase of the building of a converging interoperability framework and while the whole community is building capacity. It allows to have same metrics to allocate the sparse competent resources in the field.

The DCC work in tight collaboration with the CDS-IOP WG which is ultimately building the coherent clinical semantic interoperability framework. For that aim, the CDS-IOP is also working closely, in conjunction with the DCC, with the projects requiring new variables in order to clarify the context and the semantics pertaining to the new needs.

Variables have unique ID's. A variable is changed when any of its determinants is changed, including the range of answers (the valueset of that variable). When a variable is changed, a new variable is created linked to the previous one, in a chain of changes. A new unique ID is produced. All changes are have a track change including date, authorship, etc.

A variable is never deleted from a dataset. If a variable shall no more be used, it will stay in the dataset with a dated status set to "inactive".

Building principle
- DCC coordination
- Strong collaboration with CDS-IOP
- Strong collaboration with projects
- Prioritization of the variables
- Strong historization and unique identification of changes

2.2 Dataset releases

The DCC is sole competent for the release of datasets. Each dataset is uniquely identified and insures backwards compatibility. There is at most 4 releases per year.

2.3 Errors reporting

Errors ae reported to DCC following the appropriate procedure.

3 Convergence Healthcare system and research

In Switzerland, the Healthcare System is weighing about 80 billion CHF per year, which is about 12% of the NGP. This is massively more than what is available for the total research funded in Switzerland. Aware of the massive impact of the healthcare system, major players, such as Google, Apple, IBM are entering the field. Remarkably, they are adopting the Healthcare system standards, such as HL7 FHIR, rather than creating new standards and trying to impose them.

A similar approach is being pursued in Switzerland.

Since 2005, The Swiss Federal Council has started a eHealth Strategy that required more than 10 years 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 **(12,13)**. The standards chosen do allow to comply the complex Swiss landscape, with a distributed environment, as well as to accommodate the slow adoption of the numerous stakeholders (IT, care providers, Cantonal legal framework, etc.) at proposing a system with various levels of semantic interoperability granularity using CDA documents **(14)**. All documents, source code of the eHealth Swiss connectors, test bench, etc. are freely available on the national e-health Swiss portal **(15,16)**. 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.

From a research perspective, since the very beginning of the SPHN initiative, mutual representative are part of the Semantic Working Group of both the Swiss healthcare framework, and the research initiative, with the intention to ensure as much as possible that the standards massively used in the care system are also used in the research initiative, and mutually.

Both for the research community and for the healthcare system, cross-borders research and cross-borders mobility have become "must have" properties.

As a result, there is a strong enforcement towards using international standards whenever possible.

Standards
- strong coordination healthcare - research
- Enforce the use of international standards

4 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 CDS-IOP WG to leverage the current drive and improve capacity in the field.

This can be done by disseminating guidelines, organizing specific education for the community and develop educational curricula in high education. There are several curriculum available in Europe and in the US of America that can be taken in example. In Switzerland, some High School, such as the Bern High School, are already offering introductory education in the field (MAS, CAS in Medical Informatics), or the University of Geneva (PhD in Life Science "Genomics and Digital Health"; Master MAS, CAS in Medical Informatics). This has to be extended.

Building capacity
- Increase competences within SPHN in medical informatics
- Increase competence in semantics around clinical information
- Promote pre/post graduated education curricula in Switzerland

5 References

Other SPHN documents:

I. SPHN Data Strategy V2 2017

II. SPHN L4CHLAB Laboratory strategy document

1. OMOP Common Data Model – OHDSI [Internet]. [cited 2018 Sep 8]. Available from: https://www.ohdsi.org/data-standardization/the-common-data-model/
2. Deshmukh VG, Meystre SM, Mitchell JA. Evaluating the informatics for integrating biology and the bedside system for clinical research. BMC Med Res Methodol. 2009 Oct 28;9:70.
3. Reference Information Model (RIM) Downloads [Internet]. [cited 2018 Sep 8]. Available from: http://www.hl7.org/implement/standards/rim.cfm
4. Overview - FHIR v3.0.1 [Internet]. [cited 2018 Sep 8]. Available from: https://www.hl7.org/fhir/overview.html
5. Teodoro D, Pasche E, Gobeill J, Emonet S, Ruch P, Lovis C. Building a Transnational Biosurveillance Network Using Semantic Web Technologies: Requirements, Design, and Preliminary Evaluation. Journal of Medical Internet Research. 2012;14(3):e73.
6. Home - SemanticHealthNet [Internet]. [cited 2018 Sep 8]. Available from: http://www.semantichealthnet.eu/
7. Electronic Health Records for Clinical Research - (EHR4CR) [Internet]. [cited 2018 Sep 8]. Available from: http://www.ehr4cr.eu/
8. CDISC [Internet]. CDISC. [cited 2018 Sep 8]. Available from: https://www.cdisc.org/
9. Dugas M. Portal of Medical Data Models (MDM-Portal) [Internet]. [cited 2018 Sep 8]. Available from: https://medical-data-models.org/
10. Shared data, shared benefits | Medical Informatics Initiative [Internet]. [cited 2018 Sep 8]. Available from: http://www.medizininformatik-initiative.de/en/start
11. National competence centre eHealth in the Netherlands [Internet]. Nictiz. [cited 2018 Sep 15]. Available from: https://www.nictiz.nl/english/
12. Integrating the Healthcare Enterprise (IHE) [Internet]. IHE International. [cited 2018 Sep 9]. Available from: https://www.ihe.net/
13. Holzer K, Gall W. Utilizing IHE-based Electronic Health Record systems for secondary use. Methods of information in medicine. 2011;50(4):319–25.
14. 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
15. Bern eHealth S 3003. eHealth Suisse - Deutsch [Internet]. [cited 2018 Sep 9]. Available from: https://www.e-health-suisse.ch/de/startseite.html
16. Bern eHealth S 3003. eHealth Suisse - français [Internet]. [cited 2018 Sep 9]. Available from: https://www.e-health-suisse.ch/fr/page-daccueil.html