Design principles of a central metadata repository as a key element of an integrated health information system

Key Messages

The Challenge

The Swiss Health System (SHS) is a distributed and complex system with various interacting elements, involving five major groups of actors (population, health providers, payers, human resources and governance) for which data is collected and analyzed for distinct purposes. An integrated national health information system (HIS) is designed with the objective of generating information to improve health management decisions at all levels, based on reliable and timely data. In order to reach the overarching objective of data integration, a national Swiss Health Information System (SHIS) is confronted with the challenge of identifying and accessing data from various heterogeneous sources in order to provide relevant insights. Therefore, it is necessary to identify a data management solution to help manage the heterogeneity and dispersion of Swiss health-related data.

In the context of a SHIS, two different approaches can be considered for an appropriate data integration infrastructure: a *centralized approach* (Data Warehouse and/or Data Lake) and a *decentralized approach* (Federated Databases using Web services). Identifying an optimal approach for Switzerland has to consider technical aspects, as well as legal (e.g., data protection), cultural (e.g., privacy) and political (e.g., federalization) aspects relevant to the Swiss context.

Options to address the challenge

In light of the current constraints, this brief proposes a hybrid infrastructure that supports the following data access modes (1) *bridge access mode*: for data stored in a central data-warehouse, (2) *ferryboat access mode*: for data stored in persistent, local repositories, and (3) *crane access mode*: for data stored in temporal, local databases.

The challenge raised by a hybrid data infrastructure, with no central data warehouse, is the capability to retrieve the information about the available datasets. Therefore, a metadata management infrastructure is crucial in order to guarantee that the necessary data are retrieved and accessed from the right databases, respecting for each dataset its specificities. The main component of such an infrastructure is a central metadata repository designed as the main access point for the identification, description and location of health-related data (resources). This structure would not only permit to search and find specific data (resources), but also to establish relations between them.

Implementation Considerations

Metadata is structured as a three-part statement (*subject - relation - object*) about a resource. The consistency of the resource description is ensured by a set of models denoted by standards. The metadata schema controls the structure of the statements that can be defined (e.g., the set of allowed relations), whereas the information model controls the meaning (i.e., the semantics) of the description terms according to a standard vocabulary.

Following the recommendations established by W3C[1], the policy brief proposes a Dublin Core metadata schema as the appropriate choice to describe health-related data. The core set of the schema contains 15 elements (relations) related to resource description, including *Creator, Format, Date, Rights, Subject, Identifier, Language*, etc.

Conceptually, an information model corresponds to a linguistic ontology, represented under different forms: *nomenclature*, *terminology*, *taxonomy*, *classification* and (the most complex one) *formal ontology*. In the context of a metadata repository for a SHIS, the choice of the appropriate information model (or healthcare standard) depends on a number of criteria to be fulfilled, such as the domain coverage or the availability in all Swiss official languages. A non-exhaustive list of healthcare standards includes SNOMED CT (ontology), LOINC (terminology), ICD11 (ontology), ICD10, ICF and ICHI (classifications), WHO-ATC (classification), medDRA and openEHR (terminologies).

Opprtunities and barriers

For this brief an in-depth analysis on the principles guiding the design of a metadata repository for health data resources was conducted. Taking different perspectives into account (technical, legal and usability), the analysis highlighted several key elements related to the topic. However, a potential project for implementing a metadata management infrastructure of health-related data resources raises a number of practical questions, which should be addressed through an exchange of ideas, opinions and knowledge between relevant stakeholders (data providers, developers, policy makers, health researches). The ability of the project to achieve its main objective depends on several important factors (opportunities and barriers).

Opportunities to implement a central metadata repository may include:

- potential exchange of experience and know-how with similar projects in Switzerland;
- potential interest of health system decision-makers (political or economical);
- a single entry point for searching/retrieving health-related data resources;
- possibility to identify semantic relations between data resources;
- an increased capacity of research groups to share/access/analyse data;
- the necessity for an increased transparency about health related available resources.

Barriers to implement a central metadata repository may include:

- weak or no contribution of stakeholders during the design process of the metadata repository;
- weak or no support of data owners /data providers;
- a lack of consensus about the most appropriate healthcare standard for a metadata repository;
- no organisational structure able to manage the implementation and the maintenance of the project's hardware/software infrastructure;
- insufficient resources (workforce and financial) to move from a prototype implementation to a full operational system.

[1] The World Wide Web Consortium (W3C) is an international community that develops open standards to ensure the long-term growth of the Web