Swiss Learning Health System

Summary of the Stakeholder-Dialog on the topic of:

Design principles of a central metadata repository under technical, legal and usability constraints, in the context of an integrated national health information system

Eliane Malouf and Alessio De Santo

Authors

Eliane Maalouf, MSc, PhD Student – Information management institute, University of Neuchâtel, Switzerland

Alessio De Santo, MSc, PhD Student – Information management institute, University of Neuchâtel, Switzerland

Address for correspondence

Eliane Maalouf
Information management institute University of Neuchâtel
A.L. Breguet 2
2000 Neuchâtel
E-Mail: eliane.maalouf@unine.ch

Suggested citation

Maalouf, E., & De Santo, A. (2020). Design principles of a central metadata repository under technical, legal and usability constraints, in the context of an integrated national health information system. Swiss Learning Health System.
https://www.slhs.ch/images/learning-cycles/topics/2020-Maalouf/SD_final_IMI.pdf

Content

Summary

This dialog discussed the topic of building a metadata repository as a central piece to enable Swiss-wide health information integration. It aimed to put in contact multiple stakeholders already involved in similar or related projects. The dialog also aimed to collect feedback on the authors' recommended design and initial proof of concept.

The discussion touched on seven key topics:

- Infrastructure & data sources
- Users & stakeholders of the metadata repository
- Standards' selection
- Metadata collection
- Legal requirements
- Governance
- Alternatives

The discussions could be summarized as follows:

1. **Infrastructure & data sources**: the idea of having a central metadata repository was challenged and it was deemed difficult to achieve for cataloging all health-related datasets in Switzerland. However, linking and searching different metadata repositories could be more feasible, especially that those repositories already exist or are being created (e.g. Linkhub, SWISSUbase, SPHN, FOS catalog).

   Nevertheless, if a metadata repository is to be built, a specific scope has to be defined for the type of data it can cover with clear inclusion/exclusion criteria on the datasets. This specification of scope impacts all the following questions: defining users and stakeholders, selecting relevant standards for the community in question and their data use, identifying the governing body that could enforce quality metadata collection and sustainability.

2. **Users & stakeholders of the metadata repository**: a distinction is to be made between the groups of "users" and "stakeholders", which could be very different from each other and ways of engaging with them are different. However, another distinction was made between "data users" and "metadata users", even though the overlap between these two groups is larger.

   In essence, stakeholders would include organizations that have a vested interest in the existence of a metadata repository or are furthering the interests of other groups. Stakeholders might interact directly with the data and the metadata resources, but they also might not interact at all. They include, but are not limited to: policy makers (decision makers, administrations and federal/cantonal offices); data and service providers (hospitals, practitioners, associations and NGOs), academic and research institutions.

   The users would be organizations or individuals having a need to actually access a data or a metadata resource. They include, but are not limited to: researchers, public health practitioners, hospitals, NGOs, patient organizations, standard development organizations (e.g. WHO, Elixir), media, healthcare industries, citizens.

   The scope of the metadata repository would further specify stakeholders and primary groups of users to be served.

3. **Standards selection:** in general, standards on data resources are usually selected for the purpose of use (e.g. billing, research). In the context of metadata management, standards are sought after, first, to structure the schema (e.g. Dublin Core, DACS, schema.org), and second, to structure the values of the metadata fields and align them to common domain terminologies and vocabularies (e.g. SNOMED-CT, ICD-10). However, stakeholders have cautioned that standards should be enforced at the source, and for the purpose of use of the data being collected, to avoid errors when mapping between standards at a later time. Standards intervene at different granularity levels (e.g. high-level reference models vs specific types of measurements). The granularity at which the metadata repository operates is a key element of standards selection.

There is currently no entity responsible for selecting and communicating about the use of unified standards in the health domain in Switzerland. However, some stakeholders form working groups to come up with proposals for standards and then those proposals are discussed for agreement.

4. **Metadata collection**: when metadata is collected manually, reducing the hurdles related to this task would be crucial to guarantee metadata quality and timely collection. Design choices in the input interface, incentives in forms of rewards or social status and (semi-)automating parts of the process could be important enablers. It was noticed that the entity that enters the metadata might not be the same as the one using it (e.g. inputter: physician in the field; user: researcher in a lab). Therefore, taking into account the environment (e.g. time, location, device) of the inputter is crucial for increasing her motivation to fill in the metadata.

An important aspect of guaranteeing quality and timely input of metadata is the governance of such a system that can oversee and enforce those aspects.

5. **Legal requirements**: defining clear ownership on the metadata could be a hard task. However, selecting an adequate license model overcomes the ownership question as it clarifies how the metadata can be used without the need to know its owner. Most licensing model on metadata are open (e.g. CC BY) to allow wide access to metadata catalogs by search engines and the public. Nevertheless, the data resources reference in the repository do not necessarily have the same license model and the metadata attached to them. Access to the actual data is often restricted and happens through a separate procedure (e.g. contracts).

Another legal question comes up when proposing (semi-)automatic metadata harvesting procedures that could require to access the actual content of the data resource in question. The metadata system will have to observe privacy restrictions, access rights and consent requests.

6. **Governance**: in trying to identify the right organization that could take up the responsibility to develop and maintain a centralized Swiss health metadata repository, it became clear that, as of today, there is no one organization that has the oversight on the whole health system and the data generated by it. However, multiple organizations were mentioned as possible contenders for such a role given their current mandate and their neutrality (e.g., FOPH, OBSAN, SSPH+, SNSF). Nevertheless, it is more reasonable to think that it would be easier to identify the right governing body depending on the scope of the repository (e.g. SNSF could be approached if the scope is health research data). This

might lead to multiple organizations taking the lead in operating different specialized metadata repositories. With any of those organizations, or others, it is important to clarify what is the added value they could gain from taking up the lead on a metadata management infrastructure. Funding agencies are thought to have a large influence on enforcing FAIR data practices as well as standardization.

7. **Alternatives**: two alternatives for the proposed design came up during the discussion. The first concerns replacing a central metadata repository by a tiered infrastructure linking multiple specialized metadata repositories. Each repository exposes their metadata with a common format. Any third-party engine could then query those repositories.

   The second alternative proposed to let go of the concept of metadata and to allow direct search inside the data. Even though this alternative is interesting, especially for unstructured data, it faces an important legal limitation related to privacy restrictions on systems to automatically access health data (e.g. if patients' data is involved, patients should have provided consent to the system to access it). Furthermore, having a health data catalog and the possibility to learn about the content of a data resource without actually accessing it, is still of importance for health researchers.

Key messages

Stakeholders generally agree that a landscape of Swiss health data is missing. However, heterogeneity of data sources and diversity of users and usages make it difficult to realize.

Currently, we have a limited knowledge of the actual end-users of health metadata repositories. A large scope metadata repository is hard to maintain and at some point, the usefulness of such a platform has still to be verified in confront of a multitude of specific metadata repositories.

There is no clear governance nor process in the standardization of health data. Each entity has its own code system raising major interoperability problems.

Selecting an open licensing model for the metadata would overcome the complexity of specifically defining its ownership.

Additionally, important elements are still missing to achieve sustainable health data integration in Switzerland, among them we can cite:

- Understanding of the complexities related to digitalization in health by decision makers;
- An initiative leader or a coordination of the various stakeholders in a national trust center for health data
- Identifiers allowing to uniquely identify resources and foster interoperability;
- Legal frameworks to allow linkage between data resources.

Recommendations

Stakeholders highlighted some recommendations to be considered for future work.

A tool to facilitate the discovery of health-related datasets in Switzerland is still relevant today, but questions as to the cross-cutting nature and scope of such a platform are still open. A first recommendation is to better define the scope of the metadata repository, this filters out end-users and their potential needs.

The usefulness and usability of one Swiss-wide health metadata repository is still to be clarified. One suggestion would be to build a community who could further investigate these prospects at a national level.

Important aspects of such a platform: metadata collection; metadata use. Both involve users' participation. The system has to be designed for the users' needs, to reduce the hurdles and leverage the enablers to participation. Seeking patients consent to allow (semi-)automatic metadata harvesting and practitioner's education are suggested as two vectors to improve metadata quality.

Finally, the lead to make such a project sustainable is yet to be found. Developing a large-scale proof of concept that shows the added value of a data catalog would be a first step towards reaching to relevant organizations and convince them to put the necessary resources for the putting such a system into production.