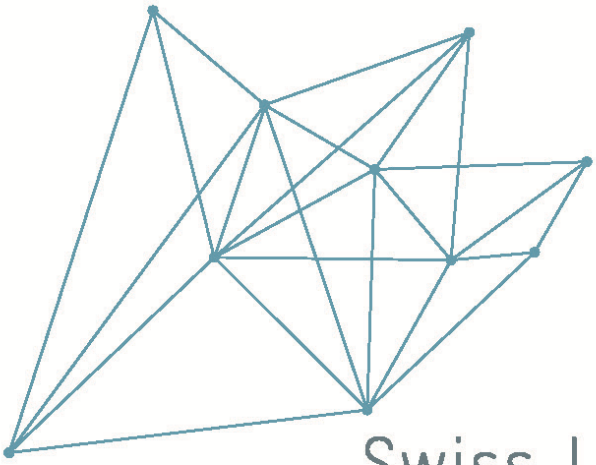


Swiss Learning Health System

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

Eliane Maalouf, Alessio De Santo, Paul Cotofrei, Kilian Stoffel

31.01.2020

www.slhs.ch

Keywords

Data, Swiss health information system, (de)centralized, metadata, security

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

Paul Cotofrei, MSc, PhD – Information management institute, University of Neuchâtel, Switzerland

Kilian Stoffel, MSc, PhD, Professor – 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., Cotofrei, P., & Stoffel, K. (2020). Design principles of a central metadata repository as a key element of an integrated health information system. Swiss Learning Health System. https://www.slhs.ch/images/learning-cycles/topics/2020-Maalouf/PB_final_IMI_2.pdf

Table of Contents

List of abbreviations

FOPH	Federal Office of Public Health
FOS	Federal Office of Statitstics
HIS	Health Information System
OBSAN	Swiss Health Observatory
SHIS	Swiss Health Information System
SHS	Swiss Health System
SOA	Service-oriented Architecture
SPHN	Swiss Personalized Health Network
URI	Universal Resource Identifier
URL	Universal Resource Location
WS	Web Services
WSDL	Web Services Description Languages
W3C	World Wide Web Consortium

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

Executive Summary

Background and Context

The Swiss Health System (SHS) is a distributed and complex system with various interacting elements, including different groups of actors: (1.) population, (2.) health providers, (3.) payers, (4.) human resources and (5.) governance of the system [1]. For each of these groups, data is collected and analyzed for distinct purposes using different methods (e.g. for billing, research, regular statistics, legal requirements). For example, the Swiss Federal Office of Statistics (FOS) collects data about the health status of the population and about the costs and financing of the health system; the Swiss Federal Office of Public Health (FOPH) collects data about the mandatory health insurance and on the general governance of the system; and the Swiss Medical Association collects data about practicing physicians, medical graduates, postgraduate training or hospital beds. Overall, there is large heterogeneity and dispersion of available data, and information that is relevant to the health system and to health services in Switzerland in general.

Integrated Health Information Systems (HIS) aim to generate information to improve health management decisions at all levels of the system through data collection, data processing and reporting [2]. The development of a 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 that can help managing the heterogeneity and dispersion of Swiss health-related data.

According to the literature, a data integration infrastructure relevant in the context of a SHIS can take two different approaches: a *centralized approach* (Data Warehouse and/or Data Lake - for accessing data from a single location) and a *decentralized approach* (Federated Databases using Web services – for accessing data at their original locations). The choice of the optimal approach must not only consider technical aspects but also legal (e.g., data protection), cultural (e.g., privacy) and political (e.g., federalization) aspects relevant to the Swiss political and administrative system. For these reasons, this brief will describe a hybrid infrastructure to support the management of health-related data. The design of this solution is based on the following potential scenarios that would have to be considered by a governing body of a SHIS:

- Recurrent data analysis or reporting, using data sources with a permanent right to be copied locally. The best approach is a central data warehouse, allowing authorized organizations easier access to data (aka “bridge access mode”).
- Recurrent data analysis or reporting, using data sources without the replication right. The best approach is a federated database using Web services, based on permanent agreements with data providers (aka “ferryboat access mode”).
- Specific data analysis or reporting, using data sources with restricted access rights. The best approach is a temporal local database, based on a one-time access mode (aka “crane access mode”).

With a hybrid data management infrastructure, with no single central data warehouse but several types of access modes, the challenge is to identify the information about the available datasets. Therefore, a metadata management infrastructure becomes crucial in order to guarantee that the necessary data are retrieved and accessed from the right databases while respecting for each dataset its specificities. The main component of such infrastructure is a

central metadata repository designed as the main access point for the identification, description and location of health-related resources. This structure would not only allow searching and finding specific data resources, but also establishing relationships between various resources.

Metadata Management

Metadata can be envisioned as a kind of map that allows representing in a simpler form the complexity of a resource (datasets, documents, etc.). Formally metadata is defined as a structured set of information used to describe a resource at different levels of granularity. The consistency of the resource description (represented as a three-part statement: *subject – relation – object*) is ensured by a set of models denoted by standards. On one hand, a metadata schema is a model, which controls the structure of the statements defining a resource (e.g., what are the allowed relations). On the other hand, an information model controls the meaning (i.e., the semantics) of the description terms according to a standard vocabulary.

Metadata management is defined as the process of using metadata standards (i.e., metadata schemes and information models) in order to structure metadata items related to resources. Metadata standards are required to establish a common understanding of the meaning of a particular type of information in a specific domain. The inner structure of metadata allows searching and identifying particular relations (e.g. semantic links) between various resources.

In the context of a metadata repository for a SHIS, the policy brief proposes to follow the recommendations established by W3C and the best practices employed by similar health-related metadata repositories. Consequently, an appropriate choice for a metadata schema would be the Dublin Core, a descriptive schema widely adopted, designed to describe digital resources. The core set of the schema contains 15 elements related to resource description, as *Creator, Format, Date, Rights, Subject, Identifier, Language*, etc. An important recommendation of this schema is the use of Universal Resource Identifiers (URI) as a formal unique identifier for a resource.

Conceptually, an information model corresponds to a linguistic ontology, i.e. a terminology agreement inside a domain, focusing on terms and their relationships. The knowledge encapsulated by an information model is represented under different forms or structures. The simplest form is a *nomenclature* (a list of preferred lexical terms from a particular field), whereas a *terminology* is a nomenclature enriched with synonyms and definitions. The forms *taxonomy* and *classification* are hierarchical structures, constructed using super-/sub-concepts, and based on the principles of class, exhaustiveness and mutual exclusiveness. The most complex structure is the logically defined model denoted *formal ontology*, allowing reasoning through the manipulation of logical formulas, and using concepts (categories, classes), properties (attributes), relationships (semantic links) and instances (objects) as building blocks.

In the context of a SHIS, the metadata repository must consider appropriate metadata schemas and information models for health data resources. 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). It should be noted that many health standards, in particular, classification systems, have been developed for a specific purpose and used by a specific set of stakeholders. The

increasing number of health standards and their inherent competition poses the question of the optimal information model for a metadata repository.

Legal Requirements

Besides considering the conceptual and technical aspects of metadata management, it is necessary to also consider legal and privacy aspects in the design process of a metadata repository for health-related resources. The key aspects of data access and protection from a privacy-preserving point of view (e.g., the Swiss Federal Act on Data Protection) must be defined while focusing on specificities of health data. The privacy and security requirements (e.g., security level) related to data access are to be included in the repository as a metadata element. Several specialized security standards for health data are relevant to data providers, such as the "HL7 Health Care Privacy and Security Classification System", or the guidelines on privacy and security published by the "Public Health Data Standards Consortium".

Aside from the health data protection requirements, the data about the users of the metadata repository need also to be assessed and protected. This brief recommends using the "European General Protection Regulation" as guiding principles and best practices, in order to guarantee privacy by design and to make sure that the right procedures are in place.

Usability Principles

The problem of interaction between an information system and its users is analysed from the usability perspective, defined as the degree to which a software can be used by specified consumers to achieve quantified objectives with effectiveness, efficiency, and satisfaction. There is a rich literature describing methods to identify potential user profiles of the information system and to understand how and why they might want to interact with this system. In the context of a metadata repository for a SHIS, given the heterogeneity of the users and of their interests, the recommended approach is to follow the principles of human-centered design. On the one hand, these principles allow taking into account the different needs of potential users. On the other hand, this iterative design process updates the system if new users' profiles are connected in the future or if requirements should change. These principles also allow for larger flexibility during the development process and during the future maintenance phases as well.

The collection of necessary information for the design process is realized by qualitative research methods to gather useful data on users' behaviors. Several activities can be carried out, each one being relevant to collect certain types of information: stakeholder interviews, expert interviews or semi-structured interviews with the users.

Conclusions and Implementation Considerations

As previously argued, the particular specificities of a Swiss Health System (related to a federated political system, a decentralized administration, the privacy culture, a diversity of data owners, and the heterogeneity of stakeholders) favour the design and implementation of a hybrid infrastructure for data integration. In this context, a central metadata repository, designed as the main element of a metadata management infrastructure, becomes essential for the identification/description of health data resources.

In Switzerland, the concept of a metadata repository has been actually considered by several projects related to the setting up of an institutional framework and infrastructure to facilitate

data access/data linking/data analysis. The “Linkhub.ch” project, for example, intends to provide several services, among which is a metadata service offering information on data sets and how they could be accessed and linked. The “Swiss Personalized Health Network” (SPHN) (a project aiming to make data from hospitals and research centers interoperable) supports the "Metadata catalogue and request portal requirements" [3], which is in the conception phase at the five university hospitals in Switzerland. On the other hand, the semantic aspects of metadata are addressed by “eHealth Suisse”, which coordinates the implementation of the electronic patient record (EPR) and the development of the Swiss e-Health Strategy 2.0.

The in-depth analysis performed in this brief on the principles that guide the design of a metadata repository for health data resources, according to different perspectives (metadata management, legal requirements and usability principles), highlights several important key elements to be considered. Yet, a potential implementation project of a metadata management infrastructure for health-related data raises also a number of practical questions whose answers may be identified 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 key influential 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 ownners /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.

Background and Context

The Swiss Health System (SHS) is a distributed and complex system with various interacting elements, including different groups of actors [1]. For each of these elements and groups, data is collected and analyzed for distinct purposes using different methods (e.g. for billing, research, regular statistics, legal requirements). For example, the Swiss Federal Office of Statistics (FOS) collects data about the health status of the population and about the costs and financing of the health system; the Swiss Federal Office of Public Health (FOPH) collects data about mandatory health insurance and on the general governance of the system; and the Swiss Medical Association collects data about practicing physicians, medical graduates, postgraduate training or hospital beds. Overall, there is large heterogeneity and dispersion of available data and information that is relevant to the health system and services in Switzerland. This makes it difficult to gather and collect knowledge in comprehensive, non-redundant ways and that would allow to address different levels of the health system. In this regard, the development and implementation of an integrated "Swiss Health Information System" (SHIS) would provide a useful starting point.

Data, information and *knowledge* are three distinct concepts in information systems [5, 6].

- Data is the (unstructured) unit of a fact/observation that contains the relevant measures (as characteristics of persons or objects), and, considered individually, have little inherent meaning or value.
- Information (structured data) is the output of a data process/analysis, usually represented as a pattern (data table, XML structures, graphs, etc.).
- Knowledge is created using a modeling approach on the information and, subsequently applied (under the form of recommendations, action rules or predictive models), to make decisions and change human behavior (also called wisdom)

Despite these distinct definitions, due to the extremely common usage of words *data* and *information* in various contexts, it is not easy to strictly respect their meanings. Therefore, in this document, the concept *data* will cover both data and information.

Integrated Health Information Systems

In general, integrated Health Information Systems (HIS) aim at generating information to improve health management decisions at all levels of the health system through data collection, data processing and reporting [2]. In order to provide this information, HIS have to ensure that reliable and timely health information is continuously available. However, this may prove to be a challenging task, as an integrated HIS would have to identify and access data from various heterogeneous and dispersed sources. Hence, it is necessary to identify a feasible data management solution that would allow managing the very same heterogeneity and dispersion of data that also characterises the current situation in Switzerland.

Data Management Solutions

Public health data management faces similar challenges as, for example, research in bioinformatics. Data needs to be identified and accessed from various heterogeneous sources and then needs to be integrated in order to provide relevant insights. Different data integration infrastructures have been described in the literature [7, 8]:

Centralized model - Data warehousing: aims collecting all relevant data resources about the system under study in a central location [9]. This approach requires setting up a uniform data model, which dictates the relevant information to extract, transform and load from the different sources [10]. This implies that the database developer must know in advance how the data is going to be used and what types of information are most relevant, in order to implement a data warehouse adapted to the specific users' needs. It has the advantage of providing one access point to the data and a local execution for all queries. However, in a context where many users might need the data for different types of applications, which are not known or definable in advance, the data warehouse cannot be adapted quickly to the users` needs. Furthermore, the maintenance of a data warehouse is very costly, as the data model cannot adapt automatically to potential modifications of the data structures. Consequently, the developer of the warehouse is forced to be constantly in the process of writing and updating data integration routines. Further, issues around data synchronization and privacy are likely to create additional problems.
More recently, with the big data era, the concept of the "data lake" has emerged [11]. A data lake is a central repository for data without imposing any structure. Intuitively, it can be seen as a large container where all types and formats of data can be deposited. This approach reduces the integration efforts needed during the collection phase, while providing a central interface from which the data can be retrieved. However, each query execution implies preparatory steps, including data cleaning and data transformation, which needs to be done prior to the analysis. Therefore, even though this approach could provide an added flexibility in the data collection stage against various user requirements, it would also add more complexity downstream when data would need to be integrated into analysis pipelines (e.g. queries).

Distributed model - Federated databases and Web Services: is a decentralized approach allowing users to access heterogeneous databases at their original location as if accessing a single database, while avoiding centralized integration [8, 12]. In a federated warehouse a query must first be translated against the different databases where the data resides. After the execution of distributed queries, results are aggregated by a central interface. Users always access an up-to-date version of the data resources without the need to replicate the data locally. The costs of data integration are lower than in the warehouse, as data is accessed on demand. However, distributed querying can be very complex and needs to adapt to the different schemas of the data resources.
The technological support for automatically accessing heterogeneous databases, as well as for managing and linking data properly is represented by Web Services (WS) [13, 14]. These services allow computer-to-computer communication and data access through a common language known as Web Services Description Language (WSDL). WS technology uses Service-oriented Architecture (SOA), an architectural model that decouples the service provider (data source) from the service consumer (user). Thus, the service consumer can choose any service from any provider no matter which query language is locally used and what database application is running. Web Application Programming Interfaces (web API) allow computers to request access to the remote data or to request a computation to be performed on the remote data.

The challenge however is that the providers need to open their data (or parts of it) and to provide a WS registry. Finally, one of the major problems of WSDL based Web services is the semantic heterogeneity, caused by disagreement about the meaning, interpretation, or intended use of the same or related data and services. Therefore, all providers must apply standardizations to data identities and nomenclature.

A hybrid infrastructure for a Swiss Health Information System (SHIS)

Besides the technical constraints, other aspects have to be considered when developing an integrated HIS and choosing an appropriate data management solution. Those aspects are mainly related to the acceptance (shaped by different legal, cultural and political implications of the implemented approach) by potential users. Additionally, the potential use of data would also have an impact on the choice of the data management model.

The following aspects need to be considered [1]:

- The Swiss political system and administration is federated and decentralized with decisions being made at federal, cantonal and communal levels. Each level of the administration has varying administrative authorities on public health issues.
- Decentralization and privacy (another important aspect of the Swiss culture) are tightly related to each other, as the latter is seen as a guarantee of the first.
- Owners of health-related data are diverse; they include, but are not limited to, the federal office of statistics, hospitals, as well as insurance companies.
- Potential users of data are also very heterogeneous (researchers, policy makers, patient associations, etc.) with different needs.

Additionally, one has also to consider the timeframes during which and the scope for which data resources are used, or can be useful. Some data resources might be relevant for one specific project: their utility is limited to a specific time period and for a specific goal. Some other data resources might be relevant to different successive projects: their utility is not limited in time or scope. Lastly, interests of ownership of data resources have to be considered when deciding to copy, store and reuse some sources.

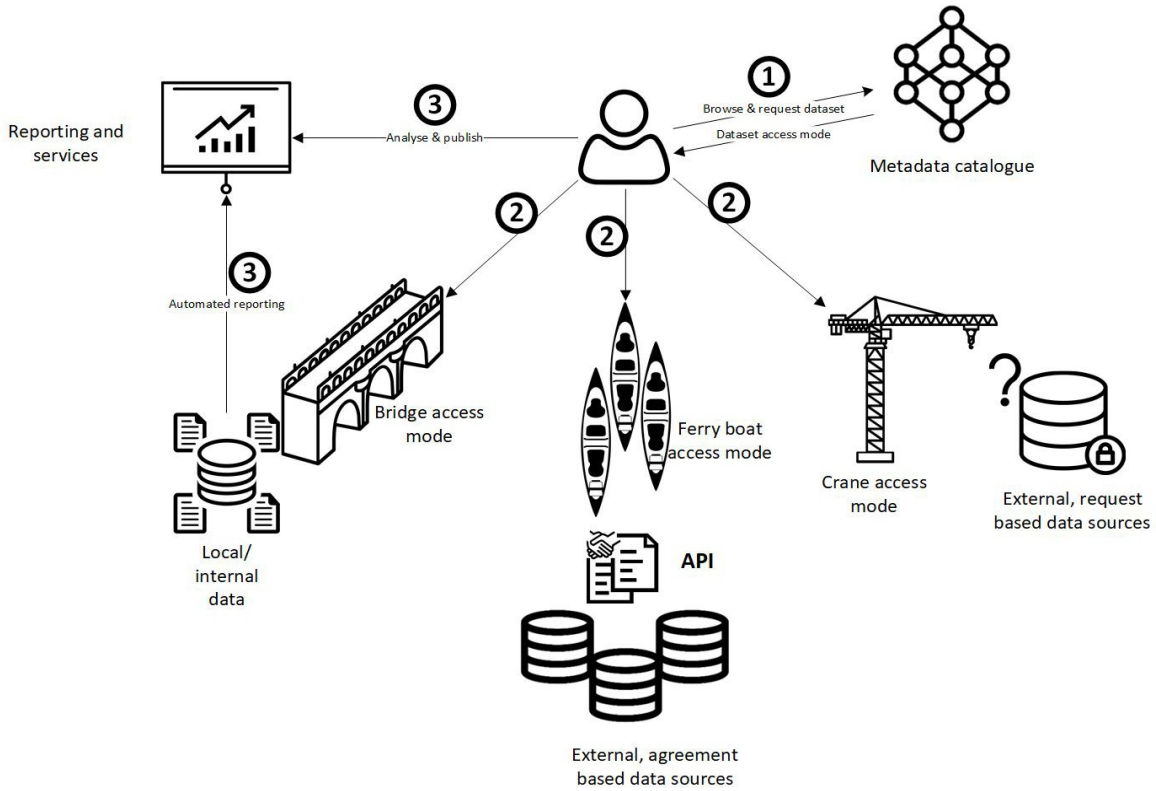
The above highlights potential constraints that may hinder the development of a purely centralized approach that would be useful for different users in Switzerland. Overall, collecting and integrating all the relevant information into a central location may become an unfeasible task. For one, information about all potentially useful data needs to be known in advance for an efficient integration. Moreover, the same effort and resources must be spent for a data resource that might be useful once and for a data resource that might be useful more often. Centralization might also lead to a duplication of databases (and, consequently, to a synchronization issue), as some organizations might continue to update only their local data resources. Furthermore, a centralized approach risks being rejected by potential (end) users as it does not reflect the decision making process in the Swiss health system and is not sensitive to the public's culture and privacy concerns. Also, ownership and data protection become very challenging in a centralized infrastructure. Given its setup as a single point of entry, it may bear some risks (including system failure), which may eventually affect a large number of individuals and organizations.

For these reasons it is argued that a **hybrid infrastructure** would be currently the most adequate design to support the data management of a SHIS. Three scenarios (illustrated in Figure

1) can be considered by a potential governing body of a SHIS that aims at implementing a hybrid infrastructure for data integration:

1. **Scenario - "The bridge access mode"**: recurrent scenarios of data analysis or reporting are defined, using data sources with a permanent right to be copied locally (so no synchronization issues). For these sources and such frequent usages, a central warehouse would be beneficial, allowing authorized organizations an easier access to the data.
2. **Scenario - "The ferry boat access mode"**: recurrent scenarios of data analysis or reporting are defined, using data sources without replication right. Agreements could be set up with the data providers to open parts of their data sources for frequent access. For these sources, federated databases using Web services would be more suitable.
3. **Scenario - "The crane access mode"**: particular, unique scenario of data analysis or reporting is defined, using a data source with restricted access rights. For this case, a one-time access mode based on a local, non-synchronized data copy, would be sufficient.

Figure 1: SHIS Data Access Modes



With a hybrid infrastructure for data management, a new challenge arises: If not all data are local and if the sources can be accessed with different modes, how can information about available datasets in the system be found? To solve this challenge, we need to consider the concept of **metadata** (data about data): “Metadata is the information and documentation which makes data understandable and shareable for users over time. Data remain useable, shareable,

and understandable as long as the metadata remain accessible." (ISO/IEC 11179-11[2]). In the case of a hybrid data infrastructure, a **metadata management infrastructure** becomes crucial in order to guarantee that the needed data are retrieved and accessed from the right data source, respecting each dataset's specificities. A so-called **metadata repository** would be the main element of such a metadata management infrastructure.

[2] https://www.iso.org/standard/61932.html

The following sections will provide an in-depth analysis of the principles guiding the design of such a metadata repository taking into account three different perspectives (technical, legal and usability perspectives). This brief will present the main technical aspects of metadata management, specifically in the context of health-related data. Further, the brief will elaborate on the key legal aspects that mainly relate to privacy concerns in the field of health(-related) data, as well as describing key elements to be considered that address the usability of such a metadata management infrastructure. For all three perspectives, challenges and opportunities for implementation will be highlighted.

Metadata Management

Metadata repositories can provide a structured point of entry to existing data sources, allowing users to quickly locate and gather data of interest in order to conduct analyses or reporting. A central metadata repository is a flexible structure that is able to handle a wide variety of unknown in advance and ever-changing data. This structure would not only permit to search and find specific data sources, but also to identify specific relations (e.g., semantic links) between these digital resources.

Tagged with metadata, any data source can be associated with other relevant elements in a metadata repository. Capturing metadata in a centralized, easily accessible repository could facilitate the discovery of relevant information and the retrieval of data sources.

In the following the basic elements of a metadata repository, as well as a detailed description of best practices for building and maintaining such a repository are discussed. Further, an analysis of existing metadata schemas and information models for the **public health domain** will be presented.

- METADATA CATALOG – location storing metadata definitions
- METADATA REPOSITORY - database storing metadata and metadata relationships

Metadata catalogs (or registries) and metadata repositories have already been implemented in various health-related contexts. Examples include the “Cancer Data Standards Registry and Repository” (caDSR) or the “United States Health Information Knowledgebase” (USHIK), the “Australian Metadata Online Registry” (METeOR), the UK cancer-grid project or the “Medical Data Models Portal”. Also in Switzerland metadata cataloges have been explicitly considered by other projects. The “Linkhub.ch” project (https://linkhub.ch), for example, plans to provide a metadata service with information on datasets including information on how to access and link those datasets. The “Swiss Personalized Health Network” (SPHN) (www.sphn.ch) also aims to establish a national infrastructure, consisting of various modules/components, enabling/facilitating a FAIR (findable, accessible, interoperable, re-usable) use of health-related data for research. A main component of the data-flow is the "Metadata catalogue and request portal requirements" [3], which is currenly in the conception phase. Semantic aspects of metadata are addressed by “eHealth Suisse”, which coordinates the implementation of the electronic patient record (EPR) and the development of the Swiss e-Health Strategy

Metadata: Basic elements and principles

Beyond the limited and imprecise definition “data about data”, the term metadata is a broad concept that covers many facets of data description and discovery. Although the term "metadata" is still relatively new, the idea of metadata goes back to the first libraries and the catalogs they kept. Printed catalogs helped users to find materials in the libraries’ collections. At that time, physical documents were cataloged by genre, title, author’s name, total number of pages, as well as some other attributes, facilitating the identification of the location of the needed items. Nowadays, it is still useful to have some kind of map to help in the description and discovery of resources. Metadata serves as this kind of map. It allows representing in a simpler form the complexity of a resource [15]. A metadata catalog documents the content,

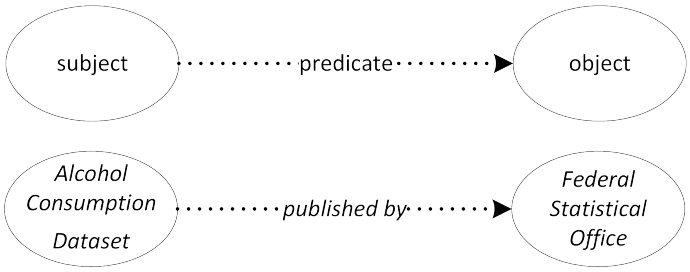
the form and the access or acquisition methods of relevant resources (e.g. data sources or other generic digital documents).

Similarly to a catalog card describing a book, metadata elements can describe resources at different levels of granularity. Three main types of metadata elements exist: descriptive, structural and administrative [16, 17]. These types are essential to establish an accurate representation of the nature of the resource (content, context and structure).

- Descriptive metadata elements allow the identification and retrieval of a resource. These elements provide usually information on the title, the subject or the creator.
- Structural metadata elements provide information about the internal structure (e.g. page, section, indices, etc.) of resources and they also describe relationships between resources
- Administrative metadata elements provide information about the origin of data resources, their type and access rights. It helps to manage and preserve resources in a collection.

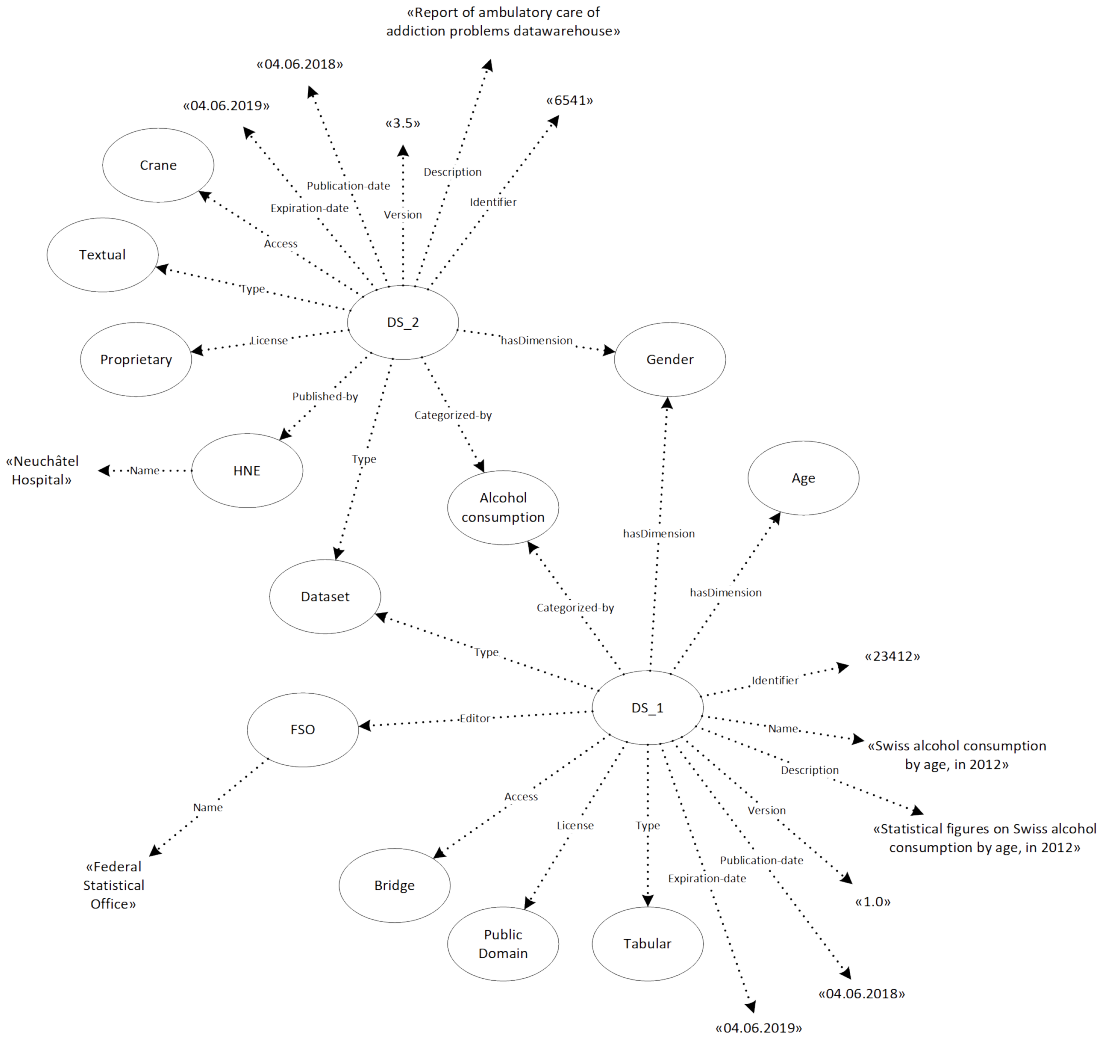
In general, metadata can be considered to be a statement about a potential resource [15]. To make a statement on a resource, firstly it is necessary to identify the resource, and secondly to have something to say about it. A statement can be decomposed into three parts: the *subject*, the *object* and the relation between the two, usually called a *predicate*. This three-part statement is known as a triple and is a fundamental building block of semantic representation [18]. Each triple represents a statement about a subject and an object (see **Error! Reference source not found.**).

Figure 2: Metadata triple representing a statement on a resource



Tied together, multiple triples form a so-called directed graph. Directed graphs are well-known data structures in computer science and mathematics. This kind of graph allows representing unlimited statements interconnecting infinite resources [19]. Figure 3 shows a diagram of some triples structured as a graph, with two subjects (nodes representing two resources categorized by alcohol consumption: DS_1 and DS_2), several predicates represented as directed arcs (e.g. *Published by*, *Access*, *Type*, etc.) and a number of objects represented as nodes (e.g. *FSO*, *Bridge, Proprietary, Textual*).

Figure 3: Metadata directed graph



Although very flexible, directed graphs can become rapidly an uncontrolled network of information of unsystematic statements if a minimum consistency is not considered. **Metadata schemas**, which are sets of rules about what kind of statements are allowed to be constructed, can help providing such consistency [20]. For example, the Dublin Core[3], which is a widely used metadata schema, has been designed to enable the description of any resource. It defines a restricted set of allowed predicates to be used by statements describing a digital resource (e.g. “Title”, “Creator” or “Date” [21]).

In addition to agreeing on the structure of statements, one must also agree on the meaning of the terms used in the statements. **Information models** provide consistency over the significance of the terms used. There are two types of schemas related to information models for resource description, one to specify the syntax, and one to specify the vocabulary. The **syntax-encoding schema** is a set of rules that provides standards to follow on how to format and

[3] http://dublincore.org/

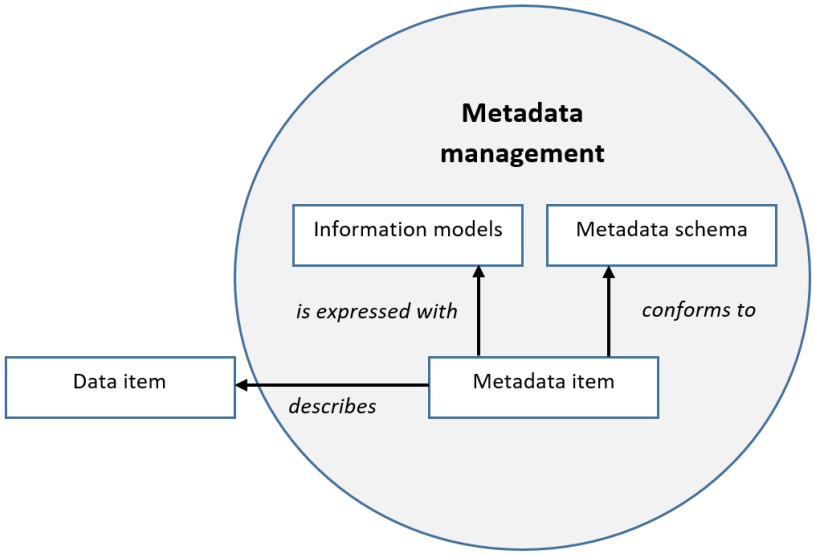
encode a specific type of data. For example, ISO 8601 is a standard for the representation of dates and times. A date and time encoded in ISO 8601, will look like the following: 2018-07-05T10:10:00. These kinds of encoding schemas are used to provide a standard for how various types of data are represented in metadata. The **vocabulary-encoding schema** on the other hand, provides a set of finite strings (called words) that may be used to make statements. In that way, subjects and objects of statements can come only from limited and controlled vocabularies, uniquely understood and approved by a domain community, e.g. the medical domain.

Managing Metadata

Metadata management refers to the process of using metadata schemas and information models in order to structure metadata items (or information units) related to data items (or data units) (see Figure 4).

A **data item** represents anything stored in an information system that describes some real world entity. Data items are described by the previously described three-part statements. Those statements are also refererred to as **metadata items**.

Figure 4: Major concepts and relationships related to metadata



The previously mentioned metadata schemas and information models specify the names and the semantics of a set of statements in a specific domain or for a particular type of information resource. Every metadata schema/information model is expressed in a formal language, also called **metadata language**. Across disciplines, many organizations are developing and maintaining many different metadata schemas and information models. These so called **metadata standards** are required to establish a common understanding of the meaning or semantics of the data. Yet, the increasing number of metadata standards due to a growing number of

metadata languages has led to the problem of competing schemas among the same domains and the dilemma of which standards to adopt if no agreement exists[4].

- METADATA: statement about a potential information resource
- METADATA ITEM: three-part statement used to describe any particular data item
- STATEMENT: can be separated in three parts - the subject, the object and the relation (predicate) between them
- METADATA SCHEMA: set of rules about what kind of statements are allowed to be constructed, e.g. it defines a set of allowed predicates
- INFORMATION MODEL: a vocabulary providing consistency over the significance of terms used in the subject and object parts of a statement
- METADATA STANDARDS: pairs of metadata schemas and information models developed and maintained inside a domain
- METADATA MANAGEMENT: process of using metadata schemas and information models in order to structure metadata items

[4] "Standards are like toothbrushes, every one agrees they're a good idea, but nobody wants to use anyone else's" - Attributed to Murtha Baca, Getty Research Institute

Metadata Schema

One commonly used metadata schema is the Dublin Core, which is a descriptive metadata schema that was designed to describe any kind of resource. The Dublin Core schema emerged in 1995 from a workshop in Dublin, Ohio: *"to advance the state of the art in the development of resource description (or metadata) records for networked electronic information objects"* [24]. Because of its simplicity and ease of use, it has been widely adopted and its use has become good practice.

The Dublin Core has 15 elements (see Table 1) and each represents a predicate of a statement that can be made about a resource. There is no limitation on the number of statements made by element; therefore, a resource (subject) could perfectly be related to several objects by the same element (see the example of element *Type* in Figure 3).

Table 1: Dublin Core metadata elements

Element	Definition
Contributor	An entity responsible for making contributions to the resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Creator	An entity primarily responsible for making the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Description	An account of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Language	A language of the resource.
Publisher	An entity responsible for making the resource available.
Relation	A related resource.
Rights	Information about rights held in and over the resource.
Source	A related resource from which the described resource is derived.
Subject	The topic of the resource.
Title	A name given to the resource.
Type	The nature or genre of the resource.

Like many metadata schemas, the Dublin Core includes rules and recommendations for the construction of statements. An example of a recommendation is, for instance, to use a syntax-encoding schema such as the already mentioned ISO 8601 for the *Date* element. An important

recommendation of the Dublin Core concerns the *Identifier* element (see Table 1): it is recommended to use a value conforming to a formal unique identifier system, and in general the Dublin Core and the W3C consortium support the adoption of the Universal Resource Identifier (URI). By design, anything one can talk about can be assigned to an URI. For example, a digital resource accessible on the Web is identified by its URL (Universal Resource Locator), which is a particular type of URI.

Information models

The semantics[5] of the terms used by Dublin Core elements are provided by information models, which are specific for each domain of interest related to the corresponding element (e.g., legal access rights for *Rights* and *Coverage*, data model for *Format* and *Type*, or healthcare for *Description* and *Subject*).

The information model (an abstract description of knowledge representing concepts, relationships, constraints, rules and operations) expresses the meaning of items for a specific domain. The knowledge representation encapsulated by an information model varies from simple, flat structures to complex hierarchical structures [25]. A non-exhaustive list of such structures includes:

- *Nomenclature* - a set of rules used to form the names/terms in a particular field, e.g. in the field of medicine. In its most basic sense nomenclature is about naming things.
- *Classification* - a coding system based on the principle of class (set of similar concepts considered as equivalent). The classification process must satisfy the principles of exhaustiveness (complete coverage) and mutual exclusiveness (no overlaps).
- *Terminology* - vocabulary of terms, including preferred lexical terms, synonyms and their definitions; may be represented as a graph.
- *Taxonomy* - representation constructed using super-concepts (e.g. genus) and sub-concepts (e.g. species) related to each other by subsumption relationships.
- *Ontology* - domain representation, based on a logically defined formalism, and using concepts (categories, classes, sets), properties (attributes), relationships (semantic links), instances (objects) or axioms (definitions) as building blocks.

[5] I.e, an agreement about the meaning

Ontology

As the most complex structure for knowledge representation, the concept of ontology has been defined in many different ways, one of the most widely accepted definitions being "a **formal**,

- **Formal**: domain specifications are expressed in a mathematical or machine-understandable form.
- **Explicit**: specifications of the concepts, relations and constraints of the domain are made in a complete and direct manner.
- **Shared**: must be agreed by the domain community.

explicit specification of a **shared** conceptualization" [27].

Ontologies, as generic knowledge representation structures, can be classified according to different criteria [28]. One of these criteria is the expressivity of an ontology, reflected by the types of components that can be defined. According to this criterion an ontology may be classified as:

- *Information Ontology* - used by humans, it focuses on concepts, instances and their relationships. This kind of ontology is usually defined by a visual language, such as MindMap for exmaple.
- *Linguistic/Terminological Ontology* - considered as a result of a terminology agreement inside a domain, it focuses on terms and their relationships. This kind of ontology is usually represented by dictionaries, nomenclatures, terminologies, taxonomies, folksonomies or lexical databases. An example of a defining language is the Resource Description Framework (RDF), recommended by W3C for creating metadata structures that define data on the Web.
- *Software Ontology* - provides conceptual schemas focusing on data storage and data manipulation. An example of a defining language is the graphical language Unified Modeling Language (UML).
- *Formal Ontology* - use formal logic to define the meaning of concepts and of relationships and focus on the reasoning through the manipulation of logical formulas. Examples of defining languages are Description Logics (DL), Conceptual Graphs (CG), First Order Logic (FOL) or Web Ontology Language (OWL) (the last being the standard recommended by W3C).

Another criterion considered for classification is the scope of the objects described. According to this criterion an ontology can be classified as:

- *Local Ontology* - represents the particular model of a domain according to a single perspective of a user (no consensus or knowledge sharing)
- *Domain Ontology* - represents a specific model of a domain according to the perspective of a group of users (knowledge sharing).
- *Core Reference Ontology* - represents a standard model of a domain according to different perspectives related to specific groups of users (integration of multiple domain ontologies).
- *Fundational/Top Level Ontology* - represents general knowledge (not related to a particular domain) defining basic notions like objects, relations, events, processes and so on.

As previously stated, metadata standards may include several information models, each one incapsulating a different knowledge representation (e.g., nomenclatures, terminologies, ontologies). In the context of a metadata repository for health-related data resources, the information model providing the semantics for the Dublin Core elements *Description* and *Subject* should be expressed as a Linguistic/Terminological, Core Reference Ontology related to health domain.

Health Ontologies

In the context of a SHIS, health ontologies would serve as a standardized representation of health-related data (i.e., healthcare standards), allowing the exchange, consolidation and interpretation of data. This standardization step is crucial in improving interoperability between the systems (e.g. computer-to-computer) and the general understandability of the data (i.e. parties agree by default on the definitions of the standard concepts). It should also be noted that many healthcare standards, in particular classification systems, have been developed for a specific purpose and used by specific sets of stakeholders.

Hereafter a non-exhaustive list of **healthcare standards** is represented.

ICD[6] - The "International Classification of Diseases" is a standard classification of diseases, disorders, injuries and other related health conditions. It currently exists in version 11. Current stakeholders include physicians, nurses, health workers, researchers, health information managers, policy makers, insurers and national health program managers, with applicability in the domains of medico economics, post-market surveillance and epidemiology. It is worth to note that ICD11 was constructed in a hierarchical mode following ontology creation standards and based on the Web Ontology Language OWL.

ICF[7] - The "International Classification of Functioning, Disability and Health" provides a standard language (classifications and codes) and conceptual basis for the definition and measurement of disability. ICF, which is complementary to ICD, provides a multi-perspective of functioning: biological, individual and social perspectives. Current stakeholderers are healthcare workers, researchers and policy makers, with applicability in the domains of clinical practice, population statistics or health policy.

ICHI[8] - The "International Classification of Health Interventions" is a classification under construction providing a tool for reporting and analyzing health interventions. It covers interventions across the full scope of the health system: acute care, primary care, rehabilitation, assistance with functioning, prevention and public health. The classification is built around three axes: Target (the entity on which the Action is carried out), Action (a deed done by an actor to a target) and Means (the processes and methods by which the Action is carried out).

HL7[9] - The "Health Level Seven" is a group of standards for the exchange, integration, sharing, and retrieval of electronic health information. The HL7 v3 suite[10] based on the Reference Information Model[11] (HL7-RIM) is a normative group of standards. The goal of this suite is to

[6] https://www.who.int/classifications/icd/en/

[7] https://www.who.int/classifications/icf/en/

[8] https://www.who.int/classifications/ichi/en/

[9] http://www.hl7.org/implement/standards/

[10] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=186

[11] http://www.hl7.org/implement/standards/rim.cfm

ensure efficient exchange of health information by producing messages and electronic documents expressed in XML syntax and by focusing on semantic interoperability. It defines health information transactions, health information documents and other data and knowledge related concepts.

LOINC[12] - The "Logical Observation Identifiers Names and Codes" (LOINC) is a vocabulary to identify health measurements, observations and documents. It codes the observation but not its value. The value could be taken from other structured vocabularies. LOINC has two categories: Laboratory and Clinical. It allows coding individual observations or a collection of observations. Current stakeholders are pharmaceutical industries, lab vendors, and tumour registries.

SNOMED CT[13] - The "Systematized Nomenclature of Human and Veterinary Medicine - Clinical Terms" (SNOMED CT) provides a standardized way to represent clinical phrases captured by the clinician and enables automatic interpretation of these. This supports the development of comprehensive high-quality clinical content in health records and meaning- based retrieval. In Switzerland, current stakeholders are physicians, with applicability as a decision-support tool for specialists in infectious diseases.

Other related vocabularies include:

- International Classification of Primary Care (ICPC-2)
- International Classification for Nursing Practice (ICNP)
- CDISC Standards in the Clinical Research Process
- Disease Ontology
- Medical Subject Headings (MeSH)
- Unified Medical Language System (UMLS)

[12] https://loinc.org/

[13] http://www.snomed.org/snomed-ct/five-step-briefing

Legal requirements

A central metadata repository, as any software system that directly interacts with users, must consider a number of legal aspects when dealing with users' related data. Moreover, according to one of the Dublin Core elements (*Rights*), information about rights (including legal ones) held in and over the resource (and particularly in the case of health-related data) has to be included. Therefore, it is mandatory to tackle legal and privacy aspects [29] in the design process of a metadata repository.

Data Protection in Switzerland

The Swiss Federal Act on Data Protection[14] (FADP) aims to protect the privacy and the fundamental rights of persons when their data is processed. Article 3 of the FADP defines health data as particularly sensitive [30]. Depending on the context, almost all personal data can be considered sensitive. In this context, two important perspectives have to be considered:

- on the one hand, data represents a great economic interest, as it can provide insights into people`s habits and preferences;
- on the other hand, the right to dispose of our personal data is an important element of our social order.

According to these perspectives everyone should be able to determine, as far as possible, what own personal information can be transmitted, to whom, at what time and in what context. Data protection must ensure that the proportionality principle is respected, meaning that collection and processing of data will involve as little personal data as possible, and never more than what is strictly necessary. It must also ensure that the person has the opportunity to check as far as possible the processing of his/her own data so that he/she can, if needed, oppose.

As the metadata repository envisioned in this policy brief represents only a platform that enables references to existing data resources and does not directly collect nor store these data resources, legal and privacy questions should already have been tackled at the data collection level by the relevant organizations. Nevertheless, the privacy and security requirements to access the data resources are to be included as a form of metadata in the repository. Each data resource will require an explicit security level related to the sensitivity of its content. Furthermore, the reuse of data and communication about it requires also an authorisation in the form of licenses in order to protect any attached copyrights.

Based on ISO/IEC 27001[15], information classification contains three major security levels:

1. *Public or open information*: may be broadly distributed and reproduced without restrictions. It can be simply deleted when not needed anymore.
2. *Internal or proprietary information*: remains inside the organization and would require authorization to disclose to the public. Reproduction is limited to authorized persons.
3. *Confidential or restricted*: highly sensitive and valuable information. Access and reuse of this information requires the highest level of control. Might need to apply encryption to

[14] https://www.admin.ch/opc/fr/classified-compilation/19920153/index.html
[15] http://www.iso27001security.com/ISO27k_Model_policy_on_information_classification

store it. Disposal should also follow strict and secure measures to allow for correct deletion of data.

Health Data Security Standards

There are some specialized security standards for health data that are relevant to data providers. The aforementioned heathcare standard "HL7 Health Care Privacy and Security Classification System"[16], for example, describes rules for automated labeling and segmentation of protected healthcare information by access control systems to enforce privacy and security policies. Another example is the "Public Health Data Standards Consortium", which provides guidelines for privacy and security standards[17]. Some relevant documents on security classification can also be found on the NHS Digital website[18] and the IHE website[19]. In theory there should be no adjunction of new sensitive data at the metadata repository level, as only descriptive data of already existing data resources are created. However, a metadata repository providing a single point of entry to potentially sensitive information stored somewhere in the SHIS network could attract malicious hackers. In order to ensure the security of the metadata information, some strategic and political questions have to be addressed:

- Who is authorised to navigate through the metadata repository? How is authorization acquired?
- Who is authorised to add datasets to the metadata repository? How is authorization acquired?
- Are the datasets of SHIS partners sufficiently described on the levels of sensitivity, privacy and access control?
- How can the organization maintaining the metadata repository be legally protected against security threats directed towards data that belongs to entities in the network, or from the non-observation of the relevant data protection principles by these entities?
- What is the legal framework for defining the classification levels of datasets and access management/authorization to these datasets?
- What is the legally binding documentation that needs to be generated in order to create and maintain the metadata repository (e.g. terms of use, discharge of responsibility, etc.)?

Confidentiality

Aside from the requirements reagarding the protection of the system's data resources, data about the users of the metadata repository also need to be assessed and protected. The European General Protection Regulation[20] (GDPR) provides the first detailed regulation in the world to ensure the protection of users' data assets. Even though it is not directly binding for Switzerland and there is no obligation for a SHIS to follow it, its principles could still be used as

[16] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=345
[17] http://www.phdsc.org/standards/health-information/PS_Standards.asp
[18] https://digital.nhs.uk/services/data-security-centre
[19] https://www.ihe.net/resources/technical_frameworks/
[20] https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32016R0679

guiding principles and best practices. On one hand, this would ensure high privacy standards for SHIS users and, on the other hand, it would set the stage for any future compliance verification that might be requested by a national or a foreign regulator. GDPR promotes the concept of "protection by design and by default" which means that personal data must be stored using pseudonymization or full anonymization. The highest possible privacy settings need to be applied by default, so that the data is not available publicly without explicit, informed consent, and cannot be used to identify a subject without additional information stored separately. The subject has the right to revoke consent at any time. No personal data may be processed unless it is done under a lawful basis or an unambiguous and individualized affirmation of consent is received from the data subject. Based on the "12 steps to GDPR compliance" web page[21], the most relevant principals are listed below:

- Make sure that the importance of users' privacy is understood by the stakeholders of the project.
- Document the personal data that the system will need to hold about its users, where it came from and who it is shared with (e.g. dataflows).
- Make sure that procedures are in place to accommodate the rights of individuals to be provided with their personal data in a commonly used format, and that this data can be deleted upon the data subject's request.
- Identify the lawful basis under which the system or other related entities can process the users' data. Make sure the privacy notices communicated to the users are clear enough to explain the extent of the users' data collection and its processing.
- Make sure that procedures are in place to seek, record and manage consent.
- Make sure that procedures are in place to detect, report, and investigate a data breach.
- Designate a person or a function to take responsibility for data protection compliance.

In conclusion, the legal requirements that need to be observed in the context of creating and maintaining a metadata repository for a SHIS mainly concern the protection of metadata from malicious users that might aim to attack the data owners after harvesting information from the catalog. The security and protection of the data resources themselves would remain under the responsibility of the data owners/data providers. Nevertheless, descriptions of the sensitivity, privacy and access control for each registered data resource in the repository would need to be added. Concerning the protection of users' personal data, the GDPR recommendations could be used as guiding principles to guarantee privacy by design and to make sure that the right procedures are in place.

[21] https://www.esecurityplanet.com/network-security/how-to-comply-with-gdpr.html#compliance_steps

Usability principles

Information technologies allow an increase of performances in almost all areas of activity. However, performance gains are often slowed down by users' hesitance to accept and use available information systems [31]. Because of the persistence and importance of this problem, users' acceptance of information technology represents a wide and long-standing issue in the area of research on information systems. A better understanding of the determinants of information systems usage would be of great theoretical and practical value.

Usability of an information system refers to the degree to which a software can be used by specified consumers to achieve quantified objectives with effectiveness, efficiency, and satisfaction in a quantified context of use as defined by ISO 9241-11[22]. The ISO 9241 part 210[23] standard on the ergonomics of human-system interaction recommends following a human-centered design approach for interactive systems, in order to achieve the highest levels of usability.

User-centered design

User centeredness is defined as the "approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques" [32].

Six principles of human centered design:

1. The design is based upon an explicit understanding of users, tasks and environments.
2. The users are involved throughout design and development.
3. The design is driven and refined by user-centered evaluation.
4. The process is iterative.
5. The design addresses the whole user experience.
6. The design team includes multidisciplinary skills and perspectives

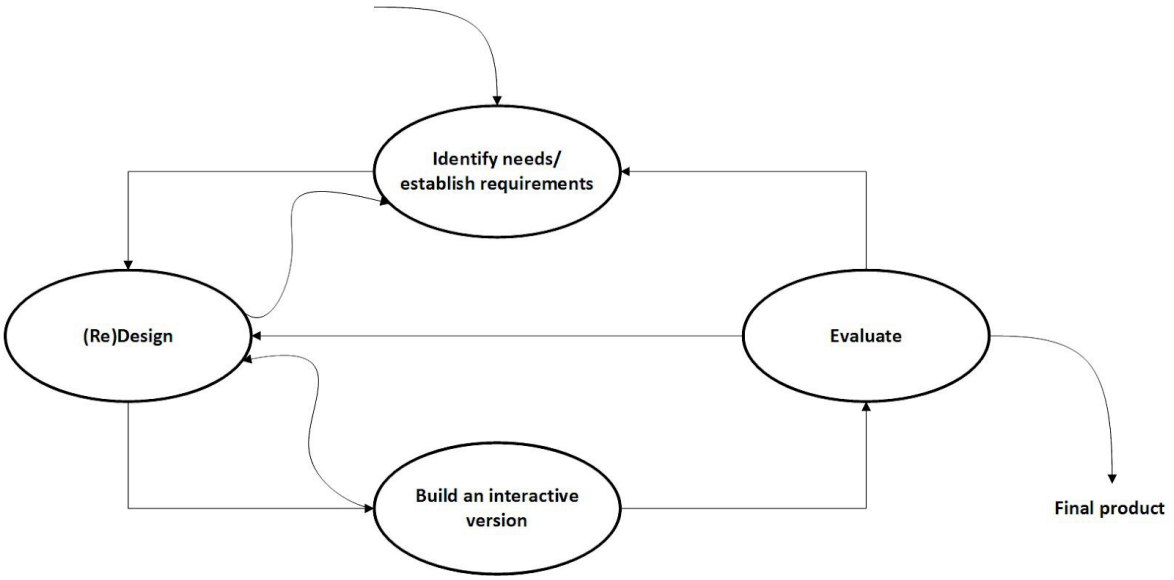
In order to achieve user centeredness, designers need to perform tasks in the following areas: understanding the context of use, understanding the users and the organizational requirements, devising the corresponding design solution and evaluation. Design activities from these different areas make up for a design cycle [33]. This cycle (see **Error! Reference source not found.**) is made up of processes: defining stakeholders' requirements, requirements analysis, implementation, verification and validation. Some of the activities involved in those processes are listed below, along with existing supportive standards [34]:

- *Defining stakeholder requirements*: this process allow to define the system's goals, to identify the system's users and their characteristics, to define the tasks to integrate in the system and to identify the environments in which the users will interact with the system. Supporting standards include "ISO 25063:2014 Context of use description" and "ISO 25064:2013 Needs report".

[22] https://www.iso.org/standard/63500.html
[23] https://www.iso.org/standard/52075.html

- *Requirements analysis and implementation*: these processes allow to define the context of use, to identify relevant user interfaces, and to define effectiveness, efficiency and satisfaction criteria. Additionally, during these steps the specifications of the user interaction are clarified through the designing of the workflow, the subtasks, the dialog model, the task specific usability objectives and the information architecture. These processes also include the identification and the design of the tasks and system objects, the user interface elements, the views, as well as the relationships and behaviors of the elements of the system. Supporting standards include "ISO 25065 User requirements specification".
- *Verification and validation*: this process aims at identifying usability problems, elaborating recommendations, and verifying conformity with user requirements. A formal summative evaluation is recommended and a baseline defined from which all systems improvements can be tracked and compared. Supporting standards include "ISO 25062 Usability test reports" and "ISO 25066 Evaluation report".

Figure 5: Design cycle



In the context of a SHIS, given the diversity of the users and their interests in the system, it is recommended to follow the principles of a human-centered design. On one hand, these principles allow to take into account the different needs and potential uses. On the other hand, the iterative process allows to adapt the system more dynamically if new users' profiles are emerging in the future or if requirements should change. These principles allow for a larger flexibility during the development process and during the future maintenance phases as well.

In general, most efforts that take into account the usability in the development of health technology are not methodical and do not follow a cohesive framework [35]. For this reason, this brief recommends, in the general context of the development of particular components of a SHIS, to follow a common standard like the ISO standards and also the recommendations from the human-computer interaction community [36].

It is important to highlight at this point that this chapter does not aim to provide final and actual usability requirements to develop a central metadata repository. These requirements would need to be identified during the implementation phases, through a structured study.

Instead, the objective here is to pinpoint to the necessary starting conditions from which a usability design and study could be initiated, as well as to present the process through which this study can be performed.

Design process

As stated previously, a user centered approach needs to focus on (1.) users, (2.) tasks and (3.) environments. In the context of a SHIS these are the three aspects that need to be firstly defined and then studied and evaluated during the implementation phases. For each of these aspects this brief identified a few initial questions to answer in order to guide the design process. Based on the literature [37, 38, 39, 40] and experience, groups of relevant questions were developed to answer to each aspect of the design of a SHIS information infrastructure.

Users

- **General background:** Who are the users that are going to interact with the system? Users could be for example: scientific researchers, political decision makers, decision makers in health organizations, health practitioners or others.; What are the main characteristics of each of the users' categories: education profiles, domain expertise, social context, etc.?; Are the users groups fixed or do they need to evolve during the lifetime of the information system?; Could the impact of the change of the users be previewed in advance?
- **Knowledge:** What is the average level of knowledge and acquaintance of each of the users' categories with information systems in general?
- **Expectations:** What are/would be each user's personal expectations, objectives and goals in using the information system being designed?; How do these users think about their activities?
- **Problems:** What are the users' (current) frustrations and problems with the situation prior to the existence of the information system being designed?
- **Patterns:** Can patterns of communication between the categories of users be identified and how do these patterns influence the use of the system?; Could the users support each other or could they enter in competition or conflict while using the information system?
- **Interaction:** Do all users directly interact with the system or are there some stakeholders that will interact indirectly through other users?; Can these indirect users be identified and their objectives from the interaction with the system mapped?

Tasks

- **Expectations:** What are the tasks that each user category expects to be able to perform by using the system being designed?
- **Users` objectives:** How would each of the tasks contribute to each of the users objectives/goals?; When, why and how is the system going to be used by each users category?; Are these tasks being currently executed without the information system under design? If yes, how?
- **Mapping:** What is the mapping between the tasks and the user categories that need to perform them?; What tasks need to be available to all users and what tasks need to be isolated to avoid cognitive load?; What is the mapping between tasks and environments?; Which tasks in the system need to be available in which contexts?

- **Organization:** Can the tasks be organized hierarchically from an overall goal of the system to more granular tasks' specific goals?; What are the preconditions /prerequisites or information needs of the users in each task?; What are the dependencies between the tasks (e.g. subtasks, workflows, further information needs)?
- **Exceptions:** Can exceptions be identified for the tasks?; Can be identified what could go wrong while the user performs the task in question?; Can corrective tasks or processes for these exceptions be deisgned?

Environments

- **Contexts:** In which contexts would the users interact with the information system that is being designed? Contexts could be for example: critical environments (e.g. a hospital), decision making environment (e.g. a boardroom), communication intensive environments (e.g. the parliament or in a dialogue setting), research environment (e.g.labs).
- **Access:** How would the information system be accessed in each of the environments?; What type of interfaces could be made available in each relevant context?
- **Mapping:** How would the environment affect the type of interaction possible with the system (e.g. voice interaction not possible due to noise in background)?; What is the mapping between environments and possible interactions?
- **Agenda:** Are there any organizational agendas related to the development of the project (e.g. political agenda, funding, organizational business goals)?; How does this agenda influence the design process and the staging of the implementation tasks?

Given the richness of information that needs to be collected in the design process, usability practitioners have developed/adapted qualitative research methods for gathering useful data on users' behaviors. Several qualitative research activities may be applied, each one being relevant to collect certain types of information [41]:

- *Stakeholder interviews* identify preliminary product vision, budget and schedule, technical constraints and opportunities, business drivers and the stakeholders' perceptions of the users.
- *Subject Matter Experts (SME) interviews* provide information on expert users, identify causative problems with the current situation and can provide very useful insights on complex or specialized domains.
- *User interviews* provide rich information on current uses and frustrations, goals in adopting the product being designed, the users roles in the implementation and maintenance of the system, identifying domain related issues and vocabulary, when, why and how the product is or will be used and what users need to know to do their jobs.

It is often noticed during such studies and interviews that the interviewed are not always able to verbally articulate their ideas and feelings. A useful method that is often used by usability practitioners is the observation of the users performing their tasks in the current environment or interacting with the prototypes of the proposed solution. This observation combined with verbal interviewing is thought to be largely effective and information-rich. Furthermore, an additional literature review activity would allow to collect information about existing systems, including relevant documentation from the technical literature, as well as publicly available publications and white papers. This step would allow to evaluate the existing systems from a

usability point of view and to discover discordant functionalities with the system being designed. Further, if a system is being designed from scratch and the users have not used a similar system before (i.e. similar in objectives), the existing literature could be used to develop the initial design prototypes in a form that can be tested and evaluated with the users during the interviews and the observation activities. Once the interviews are performed and the data gathered from them is analyzed, a functional list of requirements could be extracted and first steps for a functional prototype development could be initiated. This functional prototype could be evaluated with the users and the stakeholders during the evaluation step where the prototype is compared against the requirements in further rounds of interviews and observations.

In order to correctly perform all these activities, a qualitative study design needs to be rigorously prepared using a semi-structured interview process [42]. A fully structured interview with a fixed questionnaire could be very restrictive, especially when the relevant questions cannot all be identified prior to the interviews. A semi-structured format allows to guide the interview towards the goals of the interviewer while allowing for the possibility to discover or to go deeper into the topics with the interviewee. This method has the potential to collect a large set of data in the form of notes (and recordings) and has also the potential to identify issues or angles to the problem that the designers might have not accounted for at the beginning of the reflections [43]. This method of data collection is very relevant to a SHIS, since the concept of the system is not strictly defined, the users are very diverse with multiple levels of interaction with the system and all information about these users is not readily available at the beginning of the design process. The process can then be further supported with other methods once the vision gets clearer (e.g. tasks analysis, focus groups, usability testing) [44].

Data analysis has to follow a rigorous methodology, including information extraction, categorisation and reporting, to generate relevant data that is usable for designers to answer more specific questions (e.g. list requirements, list tasks, list goals, etc.) and to identify new issues that were not taken into account earlier. In a user-centered paradigm, the central piece of data organisation is the modeling of users as personas. Personas are composite archetypes based on behavior patterns of the observed users [41]. Personas play also an important role in reporting and documentation, as they provide a documented description of the designer's level of understanding of the users' needs. This documentation can evolve through time, given the updates and the feedback received from the users. This documentation also serves as the reference against which the implemented requirements will be evaluated. From the definition of these models, specific requirements can be extracted and transformed into system's functionalities to be implemented in a functional prototype.

To illustrate the idea of personas, we take the example of personas that were developed in the context of the CEDAR project[24]. CEDAR stands for "Center for Expanded Data Annotation and Retrieval". It aims to create information technologies to improve metadata authoring and reuse in the biomedical domain [45]. In this project five personas were defined for which the designers have identified a set of descriptive characteristics and a list of needs:

- *Template Creator* – needs to go through the metadata standards and prepare the templates that will be used by the other users.

[24] https://metadatacenter.org/purpose/user-personas

- *Metadata Creator* – needs to describe the datasets by identifying the relevant metadata and filling in the correct templates.
- *Dataset Curator* – is responsible for the quality and the completeness of the available metadata about the datasets and needs to follow up on any correction or modifications to ensure quality.
- *Biomedical Researcher* – needs to find one or more datasets that are relevant for her research.
- *Bioinformatics Developer* – needs to create new functionalities in the system as well as to maintain existing ones.

Even though these personas could not directly be reused in a SHIS, the CEDAR project does have some common objectives with a SHIS in terms of data management and metadata curation [46, 47, 48]. Also, many of the design methods used to develop CEDAR might also be relevant to the development of a SHIS.

A major phase of a human-centered design approach is the usability evaluation. We speak about two types of evaluations: formative and summative [49]. Formative evaluations are conducted during the design phase and are used to inform design decisions. They are qualitative in nature, part of the iterative process and are best conducted in the form of observing the users manipulating the system's prototypes. Summative evaluations on the other hand are conducted after the end of the development of the project or at the end of major releases of the product. Their purpose is mainly to identify pending problems to address in future development cycles. They are thoroughly documented and often conducted by third party professionals (e.g. quality assurance). Both of these evaluations are relevant to a SHIS. Formative evaluations are necessary to get quick feedback from the users in order to adapt the functionalities as early as possible to avoid incurring larger costs when these modifications are harder to implement in the running system. Summative evaluations will be necessary as part of a larger evaluation of the SHIS to first test for the usefulness of the system and second to evaluate if the current system is suitable to actually bring this usefulness to the users. This evaluation would allow for the design of future versions of the system and for major upgrades to the underlying objectives and mission.

This chapter provided an overview over the usability aspects of the SHIS. It identified the initial activities to perform in order to design the system with the users' needs in mind. Those steps consist of:

- getting preliminary answers to questions related to identifying relevant users, tasks and environments in relation to a SHIS;
- identification of common elements with existing platforms and the use of this information as reference for the analysis phase;
- designing a qualitative study to collect information from users;
- analysing data on users' behaviour and reporting of results in the form of personas and lists of requirements;
- priority setting for the functionalities and the users' needs;
- production of initial dummy and/or functional prototypes;
- iteration of the design process by performing formative evaluations with the users in order to update the requirements and to evaluate the designs.

When the proposed designs are validated, the actual implementation steps can begin. At the end of the primary functions implementation, a formal summative evaluation can be performed for the system and future developments could then be staged out. Formative evaluations can

then be reconducted in order to validate the designs of the following phases of the implementation. The actual implementation of the system will require the execution of the different steps of the process described earlier, as well as clear documentation and reporting about the results of this process.

Practical considerations

The implementation of a process supporting the generation and use of information in the context of a Swiss Health Information System (SHIS) would allow building a stronger integrated foundation addressing the entire Swiss health system. The steps of this process, including the identification of data resources, the setting up of data collection and transmission protocols, and the implementation of a data processing system, must be supported by an integrated data management infrastructure. As argued in the first part of this document, the particular specificities of the Swiss health system (related to a federated political system, a decentralized administration, the privacy culture, a diversity of data owners, and the heterogeneity of stakeholders) favour the design and implementation of a hybrid infrastructure for data integration instead of a centralized approach (e.g., a central data warehouse).

In the context of a decentralized approach for health data resources, a central metadata repository, designed as the main component of a metadata management infrastructure, becomes essential for the identification/description of specific data resources. The metadata repository would permit not only to search and find data resources, but also to establish relations between various resources. The consistency of the metadata representing a digital resource is provided by a metadata schema and a healthcare standard, specifying both the syntax rules and the semantics of the description of the resources.

The in-depth analysis that was provided in this policy brief highlighted some major key elements on the principles that could guide the design and development of a metadata managemenet infrastructure for a SHIS. Yet, the analysis only provides a starting point for a discussion concerning the possible development of such an infrastructure and the set of services provided with it. Stakeholders across the whole spectrum of the health system would have to be involved in the discussion and development of such a metadata management infrastructure, including data providers, developers, policy makers, and researchers.

The discussion may be structured along a number of practical questions that are relevant to the potential implementation of a metadata management infrastructure. In the following, based on their knowledge and past experience, the authors of this brief are expressing their opinion for some of these questions. These are by no means exhaustive and should solely serve as a basis for discussion.

Technical and Managerial aspects

- What is the governing/coordinating body that is going to manage the implementation and the maintenance of the hardware and the software infrastructure for a metadata repository project?
    - A potential candidate may be a public institution/organisation with a mandate defined by the Confederation that covers the domain of health-related data analysis and data exploration, and which is also an important data provider (e.g. OBSAN or FOS). On the other hand, if particular groups of stakeholders providing specific data resources (e.g., cohort data) or interested in specific topics (e.g., cancer studies) are considering the implementation of local, specialised metadata repositories, the management of such projects may be done by structures that are representative for those groups.
- What is the decision-making process for establishing/selecting the health standards for metadata items?

    - Under the hypothesis of a Swiss-wide central metadata repository, the ideal approach would be the emergence of a consensus between all the stakeholders on the choice of the optimal health information model. Because of the actual heterogeneity (interest and vision) of stakeholders, a more realistic approach could be a decision based on the recommendations for supporting semantic interoperability in the framework of the national e-Health strategy 2.0. However, if several projects for specialised metadata repositories are conducted, a decision-making process based on the consultation and agreement of the stakeholders related to these particular projects is more appropriate.
- What are the necessary resources (e.g. workforce, finances) needed to implement the project during its entire life cycle?
    - The complexity of a global project aiming at developing a metadata management infrastructure for health-related data resources makes an accurate evaluation of the necessary resources, at this point in time, quite difficult. More insights into the type and level of necessary resources may be obtained through a pre-project in the form of a prototype for a metadata repository (proof of concept to test proposed ideas). Moreover, the information/knowledge obtained during similar projects (finished, ongoing or in preparation) for metadata repositories, such as Linkhub.ch or SPHN, could also be used for resource evaluation.
- Who are the actors involved in the design and execution of the operational processes related to metadata creation (from data resources to metadata item): data identification, data collection, data access, data description?
    - The implication/contribution of stakeholders can be achieved only if these actors identify a real interest in the exploitation of a metadata repository and the use of the functionalities of such a system. As data providers (either institutional, private or academic) must normally deliver the highest level of involvement in the execution of the operational process, their attachment to the project is essential. Therefore, incentives should be provided to these actors.

Legal aspects

- What type of legal expertise is needed to be able to follow up on the legal questions related to the system and its data assets (e.g. licensing aspects)?
- Where is this expertise available? Is it accessible?
- What legal documentation needs to be produced to manage the operational aspects especially the ones related to data collection and data access (e.g. data contracts)?
    - As an overall opinion, the legal aspects of the project are expected to be addressed particularly by experts with specific knowledge in the field of law. If the acquisition of such expertise is proving to be difficult to obtain (e.g. insufficient resources), an alternative option would be to apply the related experience (or know-how) acquired by other similar projects. An example is the "Data Transfer and Use Agreement" template[25], developed by the SPHN and

[25] https://www.sphn.ch/dam/jcr:7cdb0e5b-79fc-4536-902e-f2d1ef91f4f9/Factsheet_20190403_DTA_template.pdf

Swiss Biobanking Platform. This document defines the rights, responsibilities and obligations of the involved parties regarding permitted use, ownership, publications, intellectual property and liability when health data are being transferred or accessed.

Usability aspects

- What is the process for identifying the stakeholders? How are they integrated in the project?
    - The development of a metadata repository would theroetically involve all stakeholders relevant to the health system, a heterogeneous group with different interests, roles and needs. However, for the first step of designing the system, only representative stakeholders (i.e., those with a proven interest for the project) would have to be identified. For this purpose we can apply a methodology known as "stakeholder analysis" [50] that can draw out the interests of stakeholders, asses their influence and importance and identify the relations which could enable “coalitions” of project participation, ownership and cooperation.
- What types of resources are needed to perform the usability study for the system?
- How is the usability study for the metadata repository project going to be performed in practice?
    - An appropriate evaluation model for assessing the usability for a metadata repository includes four interrelated metrics: satisfaction (consistency and standards), supportiveness (feedback), usefulness (information seeking, metadata creation) and effectiveness (navigation logic). The methodology for usability studies is based on focus group interviews and laboratory usability tests performed on an implemented prototype (as a proof of concept) for a metadata repository.

Opportunities and Barriers

Through an in-depth analysis of the expectations and constraints related to a SHIS, this policy brief has highlighted a number of theoretical arguments supporting the concept of a central metadata repository as the main part of metadata infrastructure for a SHIS. From a practical perspective, assessing the capacity of a project to achieve its objective (of building a functional metadata repository) can be done by identifying and evaluating the key influential factors. These factors are classified as internal, manipulable factors (Strengths & Weaknesses) and external, uncontrollable factors (Opportunities & Threats). The assessment and evaluation of these factors is done through a particular exploratory technique denoted SWOT analysis [51], which summarises the results in a 2 x 2 matrix.

Strengths • Main component of a metadata management infrastructure for health data • Single entry point for searching/retrieving available data • Consistent information due to use of standardized vocabularies (health ontologies) • Ontological support for identification of related resources	*Weaknesses* • No/weak participation of stakeholders during usability study • No/weak contribution of data providers to metadata repository • No consensus about the optimal information model (i.e., healthcare standard) • Insufficient support for legal aspects
Opportunities • Potential exchange of experience and know-how with similar projects in Switzerland (not necessarily related to health data) • Potential interest of health system decision makers (political or economical) • Potential debate about health-related standard vocabularies used in Switzerland • Increased capacity to share/access/ analyse data by research groups • Increased transparency about health related available resources	*Threats* • No organisational structure to manage the implementation and the maintenance of hardware/software project infrastructure • No/insufficient resources (workforce and financial) to pass from a prototype implementation (research project) to a full operational system (software development project) • Unexpected technological risks in software development*

*Concerning the last threat enumerated in the SWOT matrix, the specific risks related to the software development of a metadata repository (particularly, new technologies, functional requirements, system architecture and performance) will be evaluated through a proof-of-concept project for implementing a prototype of Swiss Health Ontological Supported MetaData Repository.

References

[1] C. De Pietro, P. Camenzind, I. Sturny, L. Crivelli, S. Edwards-Gravoglia, A. Sprangler, F. Wittenbecher, and W. Quentin, "Switzerland : Health System Review", Health Systems in Transition, vol. 17, no. 4, pp. 1–288, 2015.

[2] T. Lippeveld, R. Sauerborn, C. Bodart, WHO, et al., "Design and implementation of health information systems", 2000.

[3] K. Crameri, A. Lawrence, and L. Selter, "Sphn short report (2018)", Swiss Personalized Health Network, 2018.

[4] S. Stansfield, N. Orobaton, D. Lubinski, S. Uggowitzer, and H. Mwanyika, "The case for a national health information system architecture; a missing link to guiding national development and implementation", Making the eHealth Connection, Bellagio, 2008.

[5] S. K. Stansfield, J. Walsh, N. Prata, T. Evans, et al., "Information to improve decision making for health", Disease control priorities in developing countries, vol. 2, pp. 1017–1030, 2006.

[6] T. Mettler and V. Vimarlund, "Understanding business intelligence in the context of healthcare", Health Informatics Journal, vol. 15, no. 3, pp. 254–264, 2009.

[7] Z. Zhang, V. B. Bajic, J. Yu, K.-H. Cheung, and J. P. Townsend, "Data integration in bioinformatics: current efforts and challenges," in Bioinformatics-Trends and Methodologies, IntechOpen, 2011.

[8] P. K. Sreenivasaiah and D. H. Kim, "Current trends and new challenges of databases and web applications for systems driven biological research", Frontiers in physiology, vol. 1, p. 147, 2010.

[9] B. Devlin and L. D. Cote, "Data warehouse: from architecture to implementation". Addison-Wesley Longman Publishing Co., Inc., 1996.

[10] R. Kimball and J. Caserta, "The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data". John Wiley & Sons, 2011.

[11] H. Fang, "Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem", in 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), pp. 820–824, IEEE, 2015.

[12] M. N. Kamel and N. N. Kamel, "Federated database management system: Requirements issues and solutions", Computer Communications, vol. 15, no. 4, pp. 270–278, 1992.

[13] T. Malik, A. S. Szalay, T. Budavari, and A. R. Thakar, "Skyquery: A webservice approach to federate databases", arXiv preprint cs/0211023, 2002.

[14] X. Li and Y. Zhang, "Bioinformatics data distribution and integration via web services and xml", Genomics, proteomics & bioinformatics, vol. 1, no. 4, pp. 299–303, 2003.

[15] J. Pomerantz, "Metadata". MIT Press, 2015.

[16] B. Zhang and D. Gourley, "Metadata strategy", chapter in Creating Digital Collections (A. B. Zhang and D. Gourley, eds.), Chandos Information Professional Series, pp. 31–53, Chandos Publishing, 2009.

[17] S. J. Miller, "Metadata for digital collections: a how-to-do-it manual". Neal-Schuman Publishers New York, NY, 2011.

[18] M.-A. Sicilia, “Metadata, semantics, and ontology: providing meaning to information resources”, International Journal of Metadata, Semantics and Ontologies, vol. 1, no. 1, pp. 83–86, 2006.
[19] S. Cvetkovic, M. Stojanovic, and M. Stankovic, “An approach for extraction and visualization of scientific metadata”, ICT Innovations, pp. 161–170, 2010.
[20] E. Duval, W. Hodgins, S. Sutton, and S. L. Weibel, “Metadata principles and practicalities”, D-lib Magazine, vol. 8, no. 4, pp. 1082–9873, 2002.
[21] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, “Dublin core metadata for resource discovery”, Tech. Rep., Network Working Group, 1998.
[22] K. Feldman, L. Faust, X. Wu, C. Huang, and N. V. Chawla, “Beyond volume: The impact of complex healthcare data on the machine learning pipeline”, in Towards Integrative Machine Learning and Knowledge Extraction, pp. 150–169, Springer, 2017.
[23] J. Peckham and F. Maryanski, “Semantic data models”, ACM Computing Surveys (CSUR), vol. 20, no. 3, pp. 153–189, 1988.
[24] J. Pomerantz, “Descriptive metadata”, in Metadata, ch. 3, p. 67, MIT Press, 2015.
[25] F. Leiner, W. Gaus, R. Haux, and Knaup-Gregori, “Medical Data Management: A Practical Guide”. Springer-Verlag, 2003.
[26] R. Sojer, H. Müller, B. A. HES-SO, D. Aronsky, P. Ruch, and I. AG, “E-Health semantic and content for Switzerland”, 2011.
[27] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge engineering: principles and methods”, Data & knowledge engineering, vol. 25, no. 1-2, pp. 161–197, 1998.
[28] C. Roussey, F. Pinet, M. A. Kang, and O. Corcho, “An introduction to ontologies and ontology engineering”, in Ontologies in urban development projects, pp. 9–38, Springer, 2011.
[29] E. Vayena, T. Haeusermann, A. Adjekum, and A. Blasimme, “Digital health: meeting the ethical and policy challenges”, Swiss medical weekly, vol. 148, 2018.
[30] L. Morscher and K. Seidl-Nussbaumer, “Swiss Data Protection and Privacy”, in Data Protection & Privacy (M. Kuschewsky, ed.), Thomson Reuters, 2016.
[31] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology”, MIS quarterly, pp. 319–340, 1989.
[32] D. A. Norman and S. W. Draper, “User centered system design: New perspectives on human-computer interaction”. CRC Press, 1986.
[33] M. F. Theofanos and B. C. Stanton, “Usability standards across the development lifecycle”, in International Conference on Human Centered Design, pp. 130–137, Springer, 2011.
[34] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker, “What are User Requirements? Developing an ISO Standard”, in Human-Computer Interaction. Theories, Methods, and Human Issues, Part 1, pp. 4–13, Springer, 2018.
[35] P.-Y. Yen and S. Bakken, “Review of health information technology usability study methodologies”, Journal of the American Medical Informatics Association, vol. 19, no. 3, pp. 413–422, 2011.
[36] R. Mathew, A. Al Hajj, and A. Al Abri, “Human-computer interaction (hci): An overview”, in 2011 IEEE International Conference on Computer Science and Automation Engineering, vol. 1, pp. 99–100, IEEE, 2011.
[37] M. Johnson, T. R. Johnson, and J. Zhang, “A user-centered framework for redesigning health care interfaces”, Journal of biomedical informatics, vol. 38, no. 1, pp. 75–87, 2005.

[38] J. Zhang, “Human-centered computing in health information systems. part 1: analysis and design”, Journal of biomedical informatics, vol. 38, no. 1, p. 1, 2005.

[39] H. A. Taylor, D. Sullivan, C. Mullen, and C. M. Johnson, “Implementation of a usercentered framework in the development of a web-based health information databaseand call center”, Journal of Biomedical Informatics, vol. 44, no. 5, pp. 897–908, 2011.

[40] P.-Y. Yen, “Health information technology usability evaluation: methods, models, and measures”. Columbia University, 2010.

[41] Cooper, R. Reimann, and D. Cronin, “About face 3: the essentials of interaction design”. John Wiley & Sons, 2007.

[42] R. Dingwall, E. Murphy, P.Watson, D. Greatbatch, and S. Parker, “Catching goldfish: quality in qualitative research”, Journal of Health Services Research & Policy, vol. 3,no. 3, pp. 167–172, 1998.

[43] E. Murphy, “Qualitative methods and health policy research”. Routledge, 2017.

[44] C. Pope, N. Mays, S. Ziebland, A. le May, S. Williams, M. Coombs, A. Le May,D. Wicke, R. Coppin, P. Doorbar, et al., “Qualitative methods in health research”, in Methods, vol. 1, p. 2, 2000.

[45] M. A. Musen, C. A. Bean, K.-H. Cheung, M. Dumontier, K. A. Durante, O. Gevaert, A. Gonzalez-Beltran, P. Khatri, S. H. Kleinstein, M. J. O’connor, et al., “The center for expanded data annotation and retrieval”, Journal of the American Medical Informatics Association, vol. 22, no. 6, pp. 1148–1152, 2015.

[46] L. Egyedi, M. J. O’Connor, M. Martinez-Romero, D. Willrett, J. Hardi, J.Graybeal, and M. A Musen, “Using semantic technologies to enhance metadata submissions to public repositories in biomedicine”, in Proc. of Semantic Web Applications and Tools for Healthcare and Life Sciences, 2018.

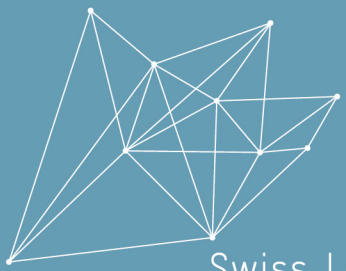
[47] M. Martinez-Romero, M. J. O’Connor, M. Dorf, J. Vendetti, D. Willrett, A. L. Egyedi, J. Graybeal, and M. A. Musen, “Supporting ontology-based standardization of biomedical metadata in the cedar workbench”, in Proc. of Int. Conf. on Biomedical Ontology (ICBO), in press, 2017.

[48] M. J. O’Connor, M. Martinez-Romero, A. L. Egyedi, D. Willrett, J. Graybeal, and M. A. Musen, “An open repository model for acquiring knowledge about scientific experiments”, in European Knowledge Acquisition Workshop, pp. 762–777, Springer, 2016.

[49] M. Taras, “Assessment–summative and formative–some theoretical reflections”, British Journal of Educational Studies, vol. 53, no. 4, pp. 466–478, 2005.

[50] R. F. Brugha, Z. Varvasovszky, “Stakeholder Analysis: A Review”, Healt Policy and Planning, vol. 15, nr. 3, 2000.

[51] M. M. Helms, R. Moore, and M. Ahmadi, “Information technology (IT) and the Healthcare industry: A SWOT analysis”, International Journal of Healthcare Information Systems and Informatics (IJHISI), vol. 3, no. 1, pp. 75–92, 2008.



Swiss Learning
Health System