Semantic Repositories in the Interoperability of Health Information Systems

Jallal Manar, Bouhaji Mouna, Ait Moudden Naima, Housbâne Samy, Serhier Zineb and Bennani Othmani Mohammed

Medical Informatics Laboratory at the Casablanca Faculty of Medicine, BP 96353 Casablanca Nassim, Morocco

Abstract: Information systems in the health field should be communicating to foster cooperation of professionals in the health process centered on the patient, and to assist in medical decision. Therefore, communication through computer tools and the creation of computerized patient records require the use of semantic repositories whose control is required. Indeed, the sharing of patient data principle is needed particularly because of the development of medical knowledge, which helps to segment the expertise, skills and roles of players. The application of this principle allows greater coordination between professionals and requires appropriate and scalable information systems. The necessary data sharing and the development of information systems strongly emphasizes the issue of semantic interoperability of health systems and management repositories. However, if, in the context of the interoperability of information systems, choice, maintenance and the use of semantic standards are necessary conditions, the behavior of players and practices of evaluation procedures are nevertheless essential component for proper functioning.

Key words: Semantic interoperability, system of health information, reference terminology.

1. Introduction

In the area of health, medical language is characterized by an extremely rich vocabulary that is difficult to interpret. The terms used are often very inaccurate and are rarely subject to rigorous definitions. Indeed, there are many ways of expressing the same thing, as well as several possible interpretations for similar terms. This does not prevent the medical staff from communicating, but significantly complicates the automation of these communications [1]. Similarly to the natural language that allows humans to communicate based on a common vocabulary and a common syntax, the language used for data interchange between information systems is based on information structures comprising syntax and constrained vocabularies that provide the words to place in these structures to compose phrases [2]. Therefore, to treat medical information with a machine, a formal model has to be provided [3]. This model consists of all terms of language and relationships that connect the general concepts to more specific concepts. Several models exist and the key to the medical field are terminology and ontology. The combined use of these data-formalized and validated knowledge bases allows to assist health care professionals in their decisions. Moreover, clinical data must be standardized and interoperable health facilities to be used by players other than their producers and in different contexts [4]. The growing scope of these systems, initially limited to the same health care facility, is now for health networks that extend to the medical community and involve all health professionals in an illness or in any given situation. This creates new demands for the implementation of these health information systems especially that one of their primary goals is to ensure the interoperability of the different players of network [5]. This interoperability whose principle responds to the scheme according to which the information is unique in the system, and omnipresent and shared in professional work environments [6].
ability of independent heterogeneous systems to work with each other harmoniously, to share or make available to the user, a workable way of information without the necessity of special adaptations between systems and developments [6]. It also allows independence of geographic scope or data processing context [4]. In this respect, productive systems and data utilizing systems must use a common language to share information effectively, and for reliable treatments from the information exchanged. This is why the information structures must be standardized vocabularies and forced to be derived from reference dictionaries, repositories of meaning of words and semantic relationships that exist between these words. These dictionaries are called “reference terminologies” [2]. Finally, it is noted that the mastery of semantic interoperability in an information system or between different systems conditioned by the mastery of the dictionary data and the semantic reference [6], which is a normalized representation of the medical information, by concepts whose meaning is fixed and unalterable, and whose goal is bearing some disadvantages of natural language [7].

The aim of our work is to highlight the use of reference terminologies in the medical field and describe their role in the interoperability of health information systems.

2. Approaches and Terminology

2.1. Semiotic Triangle

The formalization of meaning refers to a worldview characterized by the three points of a semiotics triangle [3].

- We suppose that we can identify concrete or abstract objects in the world (for example, “heart”, “reason”). It is about these objects that we want to convey information or express knowledge.
- We identify objects by forming an idea, idealizing it as concepts.
- We talk about concepts or objects using statements in a language. We assume that a concept or an object can be conveyed by an expression, a word of that language.

Accordingly, this concept creates the connection between the object and the designation. This is represented by the semiotic triangle of Ogden-Richard (Fig. 1).

2.2. Terminologies

Computer tools manipulate formal symbols and beyond their direct free texts as a human. To process information, a computer system needs a formal model. The IT system has to compare two representations and
determine if they are fully identical, partly identical or quite different then use its computing power. The need for consensus on the terms used to concepts creates the need of terminologies [9].

Building a terminology corresponds to a standardized inventory of the information that we want to handle: all relevant discussed concepts for the field and the relationships that connect the general concepts to more specific concepts or to build complex concepts from simple concepts [9]. Upon incorporation of terminology, polysemous words should be avoided or specified in other words, ensuring that each refers to a unique concept in the field. If inaccuracy is not deleted, it may then be well managed. Namely, the ability to prioritize concepts allows to explicitly link a generic term (indefinite) to more specific terms that can specify it [10]. Terminology models a system of standardized terms. Granting priority to concept, data of the field, conventional designator of the concept, the multilingual aspect is greatly simplified. Each concept may be designated by a specific term to each language. A multilingual terminology means that the concepts of a field are considered common to all languages. For a given field, there will be a unique system of concepts reflected in each language by an appropriate set of terms. Most terminologies have a standard setting. In fact, the use of standardized terms of reference terminology solves most exchange difficulties of information [3].

2.3 Use of Terminology as a Response to Coding Needs of Medical Information

At the current state of research, we do not yet have functionalities for automated processing features of natural languages that can inevitably unravel the ambiguities of these languages to fully assess all the nuances. The software in the medical field must therefore, in many situations capture, process, return to their users, and share them with qualitative data, including those used to describe the quantitative data. To perform these operations reliably, computers need to handle coded concepts, that is to say, accurate string characters; each associated with a specific and continuing significance and listed as such in a dictionary [2]. In this respect, coding a data consists in selecting from a vocabulary code that represents the meaning of this data for software, coupled with a natural language reference description that represents the same meaning for humans. Therefore, the coding of information ensures the integrity of its meaning over time, its automated processing in the local system and the fact that it is without data re-entering exportable to other systems, which in turn can use them in their treatments. To make these benefits accessible, coded qualitative data operation must comply with certain conditions, namely:

• Conservation of the link between the three elements constituting the encoded information when handled by the systems. These elements correspond to the carrier reference description of the meaning of information for people, the code for the string of characters representing the exact same meaning for software, and terminology corresponding to the computer dictionary that contains and connects the two.

• The Possibility of permanent access to the terminology by people and systems that may use the coded information to verify the relevance of the code, and find possible synonyms to the reference description (translation of this description in other language, or operating any semantic relations of this coded concept).

Table 1 Types of terminologies and their characteristics [11].

<table>
<thead>
<tr>
<th>Type of terminology</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controlled vocabulary</td>
<td>definition of terms</td>
</tr>
<tr>
<td>Classification</td>
<td>structurating named links between terms</td>
</tr>
<tr>
<td>Nomenclature</td>
<td>exhaustivity structuration</td>
</tr>
<tr>
<td>Thesaurus</td>
<td>Term standardization ambiguity reduction</td>
</tr>
</tbody>
</table>
• The Reference description associated with a code is permanent and unchangeable in the terminology; that is to say that the meaning of a coded concept is unalterable. Indeed, at the time of re-encoded information, the current version of the terminology accessed will always have the same code assigned to the same reference description.

When using codes to identify each word in different types of terminology, we may speak of a coding system. Designated as terminology system, a coding system is defined in ISO (2007) as a combination of a set of concepts, a set of codes and at least a mapping scheme between codes and concepts [1].

2.4. Classification of Terminologies

Evolution of terminologies has been described in terms of three generations [12]:
• First generation terminology “First generation”: they are characterized by a fixed organization (single hierarchy) and a simple representation as an indexed list in an alphabetical order. For example, the classification CIM10 and MeSH thesaurus;
• Second generation terminology “Second generation”: they are characterized by a dynamic organization (multiple hierarchy) with multiple indexing. For example, the medical dictionary of regulatory activities MedDRA or international SNOMED;
• Third generation terminology “Third generation”: they are based on a formal model with symbols that denote concepts and a set of rules allowing manipulating them. For example, SNOMED CT, GALEN.

Different treatment goals of medical information are obliged to create terminologies of different kinds [13]. However, few of these names have a stable definition and are universally applied in the health informatics field. We therefore reflect on their usual use rather than on the usual use that is conducted in the medical field: Ref. [1]

2.4.1 Controlled Vocabulary

Controlled vocabulary is a set of general terms without logical organization with their definitions. This definition encompasses the terms “terminological dictionaries”, “glossaries” and “vocabulary”.

2.4.2 Classification

Data collection is sometimes guided by a specific observation objective. This is, for example, the case of the collection of diagnoses for public health purposes or hospital activity evaluation. The system of concepts that we will set up to represent the possible answers to this question is directly influenced by this objective. To perform statistical calculations on the data collected, we will split up the space for answers in classes, preferably statistically balanced. These classes are a classification. The classification is a comprehensive set of mutually exclusive categories for the aggregation of data at a level of specific specialization. It can also be defined as a set of organized and hierarchical terms into classes and sub-classes. The structure of the classification and the granularity of classes depends on the objectives for which it was designed. Examples of medical classification hierarchy: CIM10 (International Classification of Diseases version10), ATC (Anatomical Therapeutic Classification), and the CCAM (Common Classification of Medical Acts).

2.4.3 Nomenclature

When the goal is to describe clinical information as accurately and authentically as possible, the classification as defined above, too oriented towards a specific goal, proves unsuited. In fact, it needs to provide a more diverse terminology and a specific range of medical concepts. Consequently, the nomenclature is intended to identify all the concepts of a domain, without restricting a priori for a particular purpose. It is a set of technical terms presented in a methodical classification, which aims to identify the terms of an exhaustive field. The terms of the nomenclature can be divided along several lines. This distribution enables to compose a complex concept by combining several concepts. One example is the SNOMED (nomenclature systematics of human and veterinary medicine).
2.4.4 Thesaurus

The search for information or documentary research aims to identify the documents containing information corresponding to an initial request. The best-known application of this technique is the literature search in databases of scientific articles. The latter uses a controlled vocabulary to index the documents: a thesaurus. Medline database is the most widely used in the biomedical field, using the MeSH (medical subject headings). The principle of indexing, done manually by professional indexers, is to describe an article by the main themes it addresses these subjects being chosen from those listed in thesaurus. The search for an article, for a user, then actually mentioning its themes of interest, and documents indexed by these concepts will be found. The concepts included in an information retrieval thesaurus are selected to cover the area with a degree of precision that depends on the effort to indexing. Finally, thesauri typically include relationships between concepts: specialization-generalization, are the most common and the general “related to” linking relationship, which allows for the search for documents dealing with related concepts.

2.5 Internal Semantic Relations Terminology

Among the relationships that may include terminology, there are relationships that allow prioritizing coded concepts from generic towards specific. This type of relationship between a specific concept to a more generic concept is noted “is a” in English, “is a kind of” in French. Some terminologies, such UCUM (unified code for units of measure), are devoid of that relationship. Other terminologies have a single chain of command. We refer to mono-axial terminology. This is the case of CIM-10, which has a single axis to classify diseases, with five deep levels at most. Other terminologies have several lines of authority. This is called multi-axial terminology. SNOMED CT is multi-axial: in its current version, it contains 19 lines of hierarchies. Diseases (disorders) are a sub-branch of one of its axes “findings” (Findings), with a number of deep levels that is not limited by conception [2].

2.6 Alignment between Terminologies to Enable Reuse of Data

In the medical field, every clinical case must be captured with precision and accuracy to allow the implementation of help features to medical decision, and re-use data in epidemiology and research work. Moreover, each case must be linked to a disease class to be classified in the right HPG (homogeneous patient group) and allow its medico-economic evaluation. Similarly, health monitoring of populations using aggregate data from data classes rather than fine-grained data [2].

Each activity has its own needs for information coding. On the one hand, documentation of care in a patient record specifically requires great descriptive accuracy. On the other, the medico-economic management of an institution requires indicators aggregating data care by broad categories.

One terminology is not enough to cover all needs. Hence, the importance of having semantic alignments between terminologies.

Therefore, to meet these needs, the use of data requires having details of a clinical case through the use of a descriptive terminology, and to store this case in a category using classification. To enable this double use from a single encoding, by the clinician, there must be a relationship between the terminologies used, namely here, the SNOMED CT and CIM-10, which allows to automatically connect with any clinical situation described fine granularity in SNOMED CT in disease class described in CIM-10. Semantic alignment is the relationship between the coded concepts of the two terminologies.

When semantics alignment is used to represent any concept of the coded terminology from a single encoded concept of arrival terminology, it is called transcoding.
3. Ontology

Ontology can be presented as the culmination of the formal definition of terminology [3]. In general, an ontology provides the means to express the concepts of a field by organizing them hierarchically and by defining their semantic properties in a formal language knowledge representation [14]. The hierarchical relationship “generalization-specialization” is unique, allowing to clearly define the subsumption between concepts. Several projects aim to developing an ontology for the representation of medical concepts. We may cite, as an example the GALEN ontologies (general architecture for language and BOMs) [15] and FMA (foundational model of anatomy) [16].

4. The Interoperability of Information Systems

4.1 The Different Levels of Interoperability

A successful exchange between players requires the consideration of various issues that can be divided into “Levels of Interoperability”.

The European Interoperability Framework for e-health (eHealth European Interoperability Framework—eHealth EIF) declined on four levels of interoperability between information systems.

The following diagram, taken from the model proposed in the EIF, has four levels of interoperability. A fifth level said syntactic or “Syntaxic interoperability” is also identified and uncouples the technical level, issues of exchange protocols, issues of exchange formats.

To each level, correspond standards and principles on which players should align to design and operate exchanges effectively.

- Policy Level:
  Shared visions, directions and converging strategies foster cooperation, communication and especially the exchanges between the different players, each with their level of activity.

- Legal Level:
  The exchanges have to comply with:
  - The legal framework upon which players (national and international law, intellectual property, privacy, etc.);
  - Contractual agreements between players (exchange arrangements, service levels, etc.).

- Organizational level:
  Organizational interoperability is related organizations and processes implemented notably to promote and operate the exchange. It also concerns the skills and knowledge associated with the operation of these organizations.

Fig. 2 The four levels of interoperability of the European Interoperability framework [2].
In terms of organization, it is, for example, to define the roles and responsibilities of those involved in the exchange within their entity. In terms of process, it is meant to define who releases data, when, following what event...but also how roles and responsibilities are shared between the different players.

- Semantics level:
  Semantics covers both the meaning of words, the relationship between the meaning of words (homonyms, synonyms, etc.) and also the life cycle of information, rules of aggregation or decomposition, etc. The meaning of words varies from organizations, businesses, players and contexts, both business and cultural. Any collaboration between entities requests communication in the sense of exchanging information. In this respect, these entities agree on the meaning of the data they exchange and the context of this exchange.

- Technology level: exchange and syntactic protocol
  Technical level deals with issues related to data exchange protocols, their formats and also the conditions and “storage” of these data. It is customary to divide this level into two parts. An “exchange protocol” part for everything related to the transport of data and therefore the “pipe” in which data flows and another part; “syntax” for everything concerning the technical format that can convey the data (their structure, coding...), regardless of their meaning that is treated at the semantic level.

  The syntactic interoperability provides first level of integration that can be called syntactic integration, in particular by defining the nature, type and format of the messages exchanged, it leads to the notion of open system to assume the heterogeneity components (interfaces, programming languages, etc.) [17]. This first level is insufficient: the formal coherence of messages does not guarantee, in itself, the consistency of meanings perceived by the different users of a system. A second level of integration, semantic integration based on semantic interoperability is necessary, which extends and complements the previous one.

  The field of health is characterized by a multitude of standards for the exchange of messages, each characterized by a syntax and a prime area of use. The heterogeneity of the syntax is however not an absolute obstacle to the interoperability of information systems. Indeed, it is always possible to design more generic templates and syntax translation tools in another translation[17]. Indeed, semantic specification messages are more important than the choice of the syntax used to support the message format [18].

---

![Diagram](image.png)

**Fig. 3** Interoperability between systems [7].
4.2. Semantic Interoperability between Information Systems

Semantic interoperability is a response to the semantic heterogeneity of data processed by different applications. It implies that the various users share consistent views about the underlying concepts of systems for various applications [17]. It is the ability of computer systems to exchange information with each other, ensuring the authenticity of the meaning of the information exchanged, preserving the sense in the long term, and exploiting this effect in their treatment [2]. The goal of semantic interoperability is how to ensure that the exchanges that take place between interconnected components retain their meaning, i.e., that the communicating parties share a common understanding of the meaning of data and services they exchange [19].

Semantic interoperability is based on the structuring of information and coding [2]. Accordingly, the meaning of the information exchanged is covered not only by words, but above all by the sentence that assembles and assigns a specific function to each of them, that is to say the information structure exchanged between the two systems and manipulated by them. Hence, the need to choose a model of interoperable information structure, known and interpreted in the same way by sharing partner systems. We must then ensure that the words inserted in the sentence (the information structure) have the same meaning for all systems involved in the exchange and this is achieved through the coding of qualitative data, building on reference terminologies. Each encodable element of a structure is associated with a set of values that contains the possible values for this information element. This set of values can represent a subset of a terminology reference or may optionally aggregate subsets of several terminologies reference.

4.3 The Essential Building Blocks of the Semantic Interoperability

Semantic interoperability operates essential blocks, listed below [20]:

- The Model of interoperable information structures on which systems need to grant prior to dematerialized exchanges or sharing health information.

Examples: medical summary part, e-prescribing, electronic care sheet, cross-border patient summary. Coded concepts that characterize meaningful words highlighted by the structural models and combine a carrier code meaning for computer systems, a term expressing the concept in verbal form for users, and the reference terminology in which this coded concept is defined.

![Fig. 4 Semantic interoperability between two systems](image_url)
Example: E11 | Diabetes mellitus type 2 | ICD-10-EN V2015

- Sets of values that list coded concepts allowing to enter a data element in a content conforming to a model of interoperable information structure. (The values of sets that link between information structure models and terminologies listing the coded concepts (code, description, terminology) employable to populate a specific coded field of an information structure).

Example: A set of values listing the health problems that can fill in the “object of investigation” field in medical biology exam application prescribed by a general practitioner.

- Terminologies that are the reference dictionaries assembling on the one hand represented by concepts encoded exploitable by those descriptions and, on the other hand by codes readable by computer processes (reference terminologies that provide formal definitions and univocal concepts coded for a specific area, their structure and their possible interrelationships).

Examples: CIM-10, CISP-2, SNOMED CT, LOINC, CCAM, MedDRA

- The Semantic alignments, establishing relationships between concepts of two coded terminologies allow transcoding of data for easy re-use.

Example: The International Classification of Primary Care ICPC-2 has a semantic alignment of its diagnoses and diseases with CIM-10. This alignment combines the coded concept T90 (diabetes type 2) of the CISP-2 to coded concepts E11, E12, E13, E14 and their specializations in CIM-10.

- Standardized services of access to terminology resources represented by sets of values, terminologies and semantic alignments posted. Software, health business players, use these services to control production or integration of interoperable information structures.

Example: The medical profession dealing software accesses the official semantic alignment CIM-10/CISP-2 to transcode in CISP-2 principal diagnosis coded in CIM-10 in the hospital report of its patient, knowing that help functions to the decision of this business software prefer coded data CISP-2.

5. Terminology Servers: Essential Tools for Standardizing Medical Vocabulary in Health Facilities

There exists in the field of health, as many terminologies as fields of application. In addition to the use of reference terminologies, other terminologies exist and are often local conceptions related to an institution or to a hospital, which raises another problem; that of regulation and standardization. The drive, which has emerged today, is to make these terminologies interoperable. Therefore, semantic interoperability requires a common format for representation of terms. As a result, LETRIM, CISMeF (Catalog and Index of Francophone medical sites- CHU Rouen) and Mondeca (industrialist), all specialized in the representation and processing of terminology and knowledge, have teamed up to design and implement a francophone “SMTS (health multi-terminologies server)”, to incorporate new terminology and make them interoperable. For terminologies to be interoperable, these systems must be based on common standards. This is to implement a common semantic repository for effective interaction with minimal loss of meaning. This is necessary when talking about medical record shared between different health professionals and patients, including various health terminologies. Terminology server is a tool for health professionals. It has access to Francophone health terminologies in real-time. Two steps were required for its design and its realization: the first on modeling and integration of terminologies and the second concerning the creation of a generic model capable of integrating each of these terminologies. The aim is to provide an open platform to specialized services developments by health players.
and industrial partners. SMTS has been operational since December 2007 and integrates seamlessly nine francophone health terminologies among the most used: CCAM, CIF, CIM-10, CISIP2, DRC, Medline, MeSH, CISMeF and SNOMED [21].

6. Dynamic Exploitation of Terminological Resources

Medical knowledge constantly evolves. Terminology resources (terminology references, sets of values, semantic alignments) accompany this development, and therefore, must evolve themselves. This evolution is reflected in the successive versions of these resources. Applications that are the players’ business in the medical sector are also changing in successive versions, according to their own logic, taking into account the regulatory changes and the needs of their users. A version of a business software
Semantic Repositories in the interoperability of health information systems

should be able to use multiple versions of a terminology or a set of values, or a semantic alignment, including future versions that even the software producer does not know about at the time of the release of his version of the business software. The business software should be able to explore and learn how to use new versions of terminology resources. The software achieves this capability by leveraging access to terminology resource services. Indeed, it is through these services that the software can assist a new version of terminology or a set of values, to structure and code a medical document for sharing or exchange [2].

7. Conclusion

Access and use of medical data have become major challenges for health professionals. In this context, several specialized medical terminologies have been created. These terminologies are mostly representational formats and different aims. The exchange of data between players in different health sector areas of activity, at the territorial, regional, national and international levels are conditioned by the development of semantic interoperability between computer systems, i.e. the ability for these systems to exchange information so that each system can interpret the meaning of the information received pursuant to the direction of the information produced, and use articulation with its own data [6]. Indeed, the use of multiple repositories by professionals is inevitable today. This leads to direct the development of systems and tools to facilitate multi-use terminology. It calls for the development of applied research to provide interfaces and tools tailored for the success of the process and to assist professionals in the use of the most appropriate ways, as transparently and efficiently as possible. In this respect, the conception of information systems should be based on an open architecture enabling the development of the scope of the system and the sharing of functions and tasks in the “business” process. It must also manage data basic standards and stable semantic repositories, perennial, recommended and updated at the national level and adopted by everybody, and finally, it must facilitate the work of professionals by integrating, as simply as possible, the collection and processing of information in the “business” process. The software involved in the progress of these processes must share data and knowledge, that is to say, they must be interoperable. This interoperability is not the only condition to bring the expected benefits in the quality of decision of patient care. Indeed, the behavior of players and procedures assessment practices are central components of functioning based on system interoperability. As a result, sensitization actions, training and incentives for players to accompany the governance of the actual interoperability and the proper use of repositories are essential.

References


Semantic Repositories In the interoperability of health information systems


