Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Using semantic technologies to promote interoperability between electronic healthcare records' information models

Jesualdo Tomás Fernández-Breis, Pedro José Vivancos-Vicente, Marcos Menárguez-Tortosa, David Moner, José Alberto Maldonado, Rafael Valencia-García, Teddy Gonzalo Miranda-Mena

Abstract—**There are currently different standards for representing electronic healthcare records (EHR). Each standard defines its own information models, so that, in order to promote the interoperability among standard-compliant information systems, the different information models must be semantically integrated. In this work, we present an ontological approach to promote interoperability among CEN- and OpenEHR- compliant information systems.**

I. Introduction

Nowadays, the advantages of flexibility and openness offered by the Internet to promote connectivity between computers is not in line with the connectivity between applications. The main techniques that have traditionally been applied to obtain integration and interoperability at application level are adaptors and exchange formats, whose success has not been very significant to date. Therefore, alternative approaches are currently making use of semantic technologies to facilitate integration and interoperability [6]. In particular, we are interested in how semantic technologies may support and promote interoperability among electronic healthcare records systems.

The healthcare record of a patient is the non-redundant, ordered and complete set of information that is generated during the relation between the patient and the healthcare system. In the last years, different working groups have been actively working in the definition of architectures and information models for electronic healthcare records. Hence, we face the problem of interoperability between different electronic healthcare record models; each model implies a working environment in which the meaning of data varies. This requisite is fulfilled by semantic technologies, because they offer a working environment that allows for connecting distributed data and describe their content according to the particular contexts. Therefore, the use of semantic technologies makes it possible the description of the nature and logical context of the information to exchange, whereas it allows each system to remain independent.

An advantage of using semantic approaches is the fact that they do not require to replace current integration technologies, databases and applications, but they add a new layer that take advantage of the already existing infrastructure [3]. In this work, integration and interoperability are carried out at data level. Data integration requires real-time transformations of the information that flows between systems. The transformations must take into account the semantic differences between the applications. The most important factors that make it difficult to integrate and obtain interoperability between systems are the semantic and structural heterogeneity, as well as the different meanings information has in different systems.

Finally, the structure of this paper is the following. In section 2, models for representing electronic healthcare records are discussed. Then, the semantic technology used to facilitate interoperability in section 3. In Section 4, the semantic infrastructure developed to obtain the interoperability will be presented. Finally, some conclusions will be put forward.

II. Electronic Healthcare Records Representation

Recently, the dual model architecture [1] is gaining relevance to develop electronic healthcare records systems. This architecture is based on the meta-modelling of healthcare records, and takes into account the dynamic nature of the healthcare environment, so scalable and adaptable models are needed. Dual modelling follows two main principles. The first one is the separation of concepts in two levels, one defining the reference model and another formed by archetypes, which are formal models of domain concepts. Each archetype defines a different clinical concept. The second principle is that computing systems are based on the reference model, and valid healthcare records extracts are instances of such reference model. This methodology is currently being used by the major standards for representing electronic healthcare records, such as OpenEHR, CEN, and the last version of HL7.

The reference model represents the global features of the

Manuscript received April 3, 2006. This work has been possible thanks to the Spanish Ministry for Science and Education through the projects TSI2004-06475-C02; the Seneca Foundation through the Project 00670/PI/04; FUNDESOCO through project FDS-2004-001-01; and Spanish Ministry of Health under Grant PI050815

Jesualdo Tomás Fernández-Breis*, Pedro José Vivancos-Vicente, Marcos Menárguez Tortosa and Rafael Valencia-García are with the Departamento de Informática y Sistemas, Universidad de Murcia CP 30100 Murcia, Spain (*corresponding author; phone: +34968364613; fax: +34968364151; e-mail: jfernand@um.es).

Teddy Gonzalo Miranda-Mena is with Intelligent Methodologies (IMET) (e-mail: tegomime@yahoo.com).

David Moner and José Alberto Maldonado are with the BET Group, Technical University of Valencia (e-mail: damoca@gmail.com; jamaldo@upvnet.upv.es).

annotation in the healthcare record, how they are aggregated, and the contextual information required to meet the ethical and legal requirements. This model defines the set of classes that constitute the building blocks and it shows the non-volatile features of an electronic healthcare record. On the other hand, the reference model must be complemented by a formal method for defining, communicating and comparing aggregates of valid data, defining datatypes, value ranges and other restrictions to ensure interoperability, consistency and quality of data. Archetypes are this complement. An archetype is a knowledge model which specifies the common features of entity types and, therefore, defines valid domain structures. Archetypes limit the business objects defined in a reference model. They bridge the generality of the business concepts defined in the reference model and the variability of clinical practice. They are formally adjusted to a formal model known as archetype model, which is formally related to the reference model. After introducing how electronic healthcare records can be represented at conceptual level, let us describe the two standards we are focusing on in this research work: CEN and OpenEHR.

The CEN/TC251 (http://www.centc251.org), Technical Committee 251 of the Normalization European Committee, is in charge of developing standards in the field of medical informatics. The activity of one of its working groups is devoted to the standardization of the architecture and information models of electronic healthcare records. This working group is currently working in the final version, which was initially planned for 2004. This model has five parts: reference model, archetype interchange specification, reference archetypes and term lists, security features, and exchange models. The OpenEHR Foundation (http://www.openehr.org) is a non-profit organization whose objectives are the development of open specifications to make the interoperable, life-long electronic healthcare record a reality and to improve healthcare in the information society. The specification of its EHR model consists of the following components: reference model, service model, and archetype model.

III. THE SEMANTIC TECHNOLOGY

Amongst the different available semantic technologies, the Ontology Web Language (OWL, http://www.w3.org/TR/owl-ref/) is currently the most used to represent semantic content. OWL is used to code a widespread knowledge representation technology, the ontology. Ontologies are used to give meaning to information structures that are exchanged by information systems [4]. An ontology can be seen as a semantic model containing concepts, their properties, interconceptual relations, and axioms related to the previous elements. Ontologies have become a key technology due to its advantages [2]. On the one hand, ontologies are reusable, that is, the same ontology can be reused in different applications, either individually or in combination with other ontologies. On the other hand, ontologies are shareable, that is, their knowledge allows for being shared by a particular community. In the context of integration, they facilitate the human understanding of the information besides the access based on information and the integration of information of very different information systems. For our purpose, the information model semantics is formalized by means of ontologies.

However, the creation of an ontology is not a simple task but time consuming. In integration and interoperability tasks, ontologies can simplify the understanding of the domain by introducing generalizations, that is, ignoring the details to highlight general ideas. It is obvious that the existence of semantic heterogeneity and divergence makes generalization more difficult, because the common aspects of two entities can be semantically represented in different ways. Due to this fact, an ontological integrative approach facilitates the achievement of generalization [5]. Ontologies allow for differentiating among resources, and this is especially useful when there are resources with redundant data. Thus, they help to fully understand the meaning and context of information. This is important for our objective of achieving semantic interoperability among electronic healthcare record systems built on top of different information models.

IV. THE SEMANTIC INFRASTRUCTURE

In this section, the ontological infrastructure developed to facilitate the interoperability of CEN and OpenEHR electronic healthcare records information models is described. This would facilitate healthcare information systems to use extracts of electronic healthcare records coming from health institution using CEN or OpenEHR information systems. In this work, the following releases of the CEN and OpenEHR specifications have been used: CEN (09/2004) and OpenEHR (09/2005). For this purpose, the following steps have been followed. First, each information model was analyzed. Then, each model was represented using OWL to obtain its ontological representation. At this point, the ontological information and archetype models were compared in order to find similarities and differences among the CEN and OpenEHR representations. It is more appropriate to perform this comparison at ontological level due to different reasons. First, ontologies are formal models so that formal reasoning can be performed on both models. Second, representing both models using the same formalism provides a common representation framework for the comparison process. Third, if we want to come to an integrated model, it is more appropriate to have the components represented with the same formalism and at the same granularity level.

Hence, let us discuss the conclusions drawn from the analysis of the information and archetype models. First, the representation of the information models is more oriented to

the transmission via a communication network rather than to represent semantically the contents. In fact, the different model diagrams provided in the documentation of both standards has little semantic information, they are similar to UML class diagrams. For instance, there are some references to elements of different classes modeled by string attributes. This representation may provoke consistency problems. Therefore, it would be more appropriate to model this reference through a relation between the corresponding classes. In our ontological approach, referential semantics is modeled through semantic relations between the concepts. Moreover, the UML representation is not suitable for performing formal reasoning at conceptual level, so that, better use of the information contained in the model might be made. However, the UML-like representation is not very appropriate for human understanding, the ontological one being more suitable for both human and machine understanding.

This process was performed in two steps: reference model, and archetype model. With respect to the archetype models, CEN and OpenEHR models are highly similar. It can be said that both models are identical in the way archetypes are built, containing the same type of information: translations, audit details descriptions, ontological section, and constraints. The ontological section and the constraint one can be considered the most important parts since they contain the definition of the different archetype terms. The archetype terms are instances of some specific type of concepts belonging to the respective (CEN/OpenEHR) reference models. However, the archetype model does not structure appropriately this information. It models the ontological part of the archetype as lists of term definitions, term bindings, constraint definitions, and constraint bindings. However, these elements are not formally modeled and related to the corresponding elements from the reference model, and the main differences appear when dealing with this issue, because they provide different types of archetype terms. In particular, the OpenEHR model is richer than the CEN one. So, what we have done is to remodel this archetype model by using as modeling focus the terms defined in the archetype. Our first goal was to develop an ontology that would represent an integrated view to the CEN and OpenEHR archetypes models, that is, the commonalities were identified, and the differences were kept to allow building archetypes for both reference models.

According to our approach, an archetype would be understood as follows. An archetype has general information as it is specified in the standard models. This general information encompasses the auditory details, the archetype description, assertions, translations to other languages, and a set of available terminologies. Furthermore, an archetype contains the definition of a concept (i.e., heart rate pulse), is a specialization of another archetype, and it contains a set of archetype terms. An archetype translation to a specific language is comprised of the set of translations of archetype terms to such language. Therefore, each archetype term has a set of translations associated. Each archetype term has also a definition, and a set of term bindings to the available terminologies.

Archetype terms can refer to restrictions and conceptual entities. Conceptual terms (called ontology terms) are divided into concepts (e.g.,, heart rate), complex terms (e.g., list, history), simple terms (e.g., position, device), or values (e.g., sitting, lying). A simple term has a set of values associated. Each complex term is comprised of a set of complex and simple terms. Values are of a particular datatype, which is given by the reference model (CEN / OpenEHR). In this case, both standards use the same basic datatypes so there is no need for making a distinction at this level. The situation is different for simple and complex terms, because both reference models offer different simple and complex modelling entities. Therefore, the type associated to a simple or complex term will depend on which type of archetype we are building.

As it was aforementioned, there are also constraint terms. This constraint terms are types of archetype terms and they are associated to simple terms. They define the linguistic expression associated to a particular simple term. A constraint term is also a type of constraint. There are other types of constraints accounting for the cardinality of terms having lists of values, the existence of a particular term, and the number of occurrence. Cardinality constraints are only compulsory for terms of types such as lists or sets, whereas every ontology term has an occurrence constraint associated. Each term associated to a reference model type, that is, complex and simple terms, have also an occurrence constraint associated, accounting for the occurrence of this type of node in the data under the owning term, that is, in the context of its parent archetype term. In order to show a graphical representation of what has been described, the following figures are provided. Figure 1 and Figure 2 contain the ontological representation of the archetype model. Figure 1 shows the vision at archetype level, whereas Figure 2 is the representation at archetype term level. Dotted lines means that the relation is optional, that is, the OWL restriction min_cardinality=0.

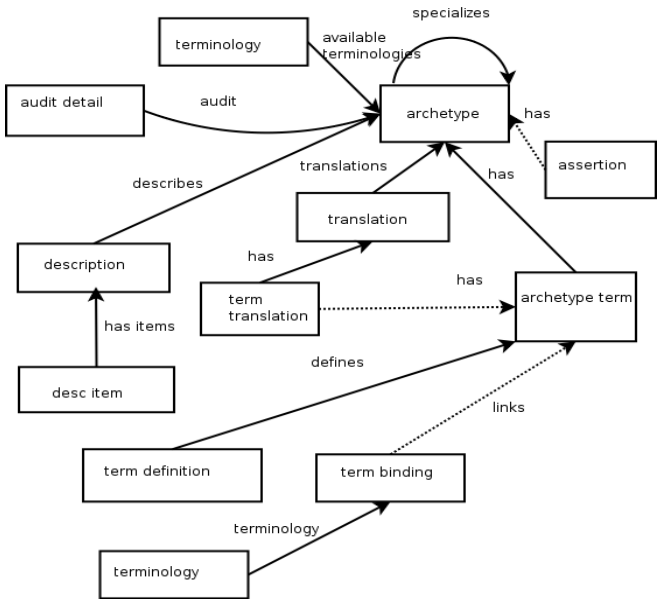

Fig.1. Partial representation of the archetype model, focusing on the archetype concept

To model the reference models, the procedure was similar to the one followed for the archetypes model. They were analyzed in order to detect semantic representation flaws, and OWL models were developed. Then, both ontologies were semantically compared in order to look for mappings between both standards to develop an integrated model for the electronic healthcare records. The main difference is the richness for defining types of clinical data. The CEN model makes use of folders, sections, entries, items, clusters and elements, whereas the OpenEHR model uses a wider range of types, including some such as history, item list, item structure, and so on.

V. CONCLUSIONS

In this paper, a semantic approach to facilitate the integration and interoperability between CEN and OpenEHR compliant clinical information systems. The key semantic technology to achieve this goal is the ontology, which can be viewed as a conceptual model containing a set of interrelated elements whose existence is accepted by a particular community. Ontologies acquire more importance when they cover particular domains, because once achieved the semantic control of a domain, data integration or linking systems would be easier. Most of the scientific community agrees on the role and importance of the use of semantic technologies. In this paper, the effort has been put on generating ontological models of the CEN and OpenEHR reference models, as well as developing an integrated archetype model.

We are currently addressing the mapping between the ontologies corresponding to the CEN and OpenEHR reference models in order to obtain a global EHR model, so that, systems might work with data coming from both standards.

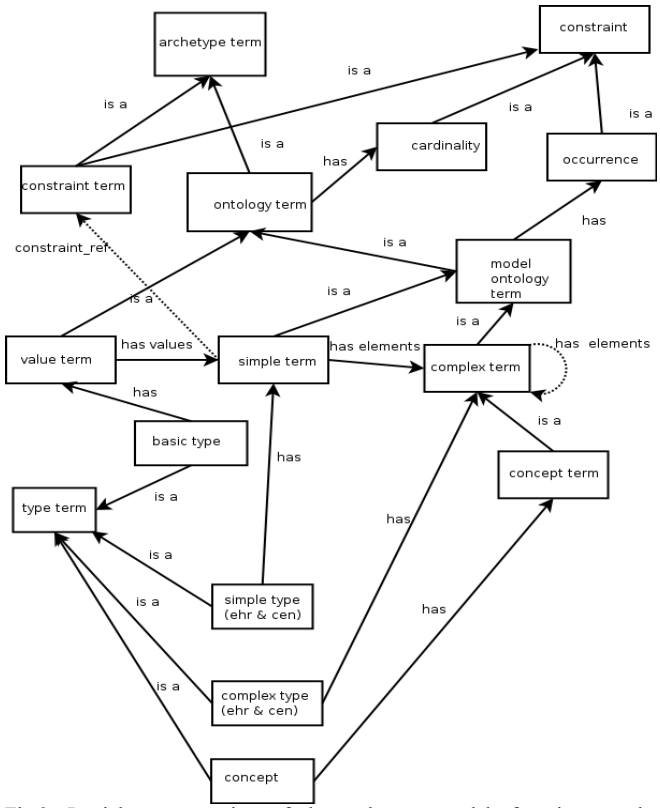

Fig.2. Partial representation of the archetype model, focusing on the archetype terms

REFERENCES

[1] T. Beale (2001). "Archetypes, Constraint-based Domain Models for Future-proof Information Systems". Available: http://www.deepthought.com.au/it/archetypes/archetypes.pdf

[2] Fernández-Breis, J.T., Martínez-Béjar,R. (2002) A Cooperative Framework for Integrating Ontologies. International Journal of Human-Computer Studies, 56(6):662-717

[3] Linthicum, D. (2004) Leveraging Ontologies: The Intersection of Data Integration and Business Intelligence. DMR Review Magazine, June-July.

[4] Missikoff, M. (2002) Harmonise: An Ontology-based Approach for Semantic Interoperability. ERCIM News 51.

[5] Partridge, C. (2002) The Role of Ontology in Semantic Integration, OOPSLA.

[6] Semantic Interoperability Community of Practice (2005). White Paper Series Module 1: Introducing Semantic Technologies and the Vision of the Semantic Web.