Clinical Data Interoperability based on Archetype Transformation

Catalina Martínez Costa, Marcos Menárguez-Tortosa, Jesualdo Tomás Fernández-Breis*

Facultad de Informática, Universidad de Murcia, CP 30100, Murcia, Spain

*Correspondence author: Jesualdo Tomás Fernández-Breis; phone:+34868884613, fax:+34868884151; email:jfernand@um.es

Email addresses: `cmartinezcosta@um.es` (Catalina Martínez Costa), `marcos@um.es` (Marcos Menárguez-Tortosa), `jfernand@um.es` (Jesualdo Tomás Fernández-Breis*)

Abstract -The semantic interoperability between health information systems is a major challenge to improve the quality of clinical practice and patient safety. In recent years many projects have faced this problem and provided solutions based on specific standards and technologies in order to satisfy the needs of a particular scenario. Most of such solutions cannot be easily adapted to new scenarios, thus more global solutions are needed.

In this work, we have focused on the semantic interoperability of Electronic Healthcare Records standards based on the dual model architecture and we have developed a solution that has been applied to ISO 13606 and openEHR. The technological infrastructure combines reference models, archetypes, ontologies and Model-Driven Engineering techniques. For this purpose, the interoperability infrastructure developed in previous work by our group has been reused and extended to cover the requirements of data transformation.

Keywords : electronic healthcare records, semantic interoperability, ISO 13606, openEHR, archetypes, ontology

1. Introduction

The present world is becoming more and more global. Citizens move nowadays more often from one country to another for work or leisure. In addition to this, the technological development has removed many barriers from the past and citizens can use most of their services from anywhere. As stated in [1], access to healthcare services during temporary stays abroad is facilitated by the European Health Insurance Card, but, cross-border healthcare services are not only limited but also account for only 1% of public healthcare expenditure in the European Union. As a matter of fact, the clinical information of a patient can usually be available only at the healthcare institution in which that piece of information was generated. Hence, the semantic interoperability of electronic healthcare records (EHR) systems is a major challenge in eHealth [2, 3]. That should allow healthcare professionals to manage the complete EHR of citizens, independently from the institutions that generated the clinical data. Recent recommendations of the European Commission [4, 3] have stated that such semantic interoperability is essential to improve the quality and safety of patient care, public health, clinical research, and health service management. In the present work, semantic interoperability of clinical information is defined as the ability of information systems to exchange and understand clinical information independently of the system in which it was created.

Standards for the representation and exchange of EHR play a crucial role in the achievement of semantic interoperability. In this work, we are interested in standards based on the dual model architecture [5], because the previously mentioned international recommendations consider them a promising solution for achieving semantic interoperability of clinical information and knowledge. Dual model architectures are based on the definition of information and knowledge separately by defining two conceptual levels by using the reference and archetype models. They provide a way of specifying this clinical information by means of archetypes. Archetypes represent the minimal information unit that clinical information systems can exchange, thus they are the basic semantic interoperability unit [6]. Moreover, archetypes also include the context for interpreting clinical information. ISO 13606 [7], openEHR [8] and HL7 CDA [9] are the most widespread dual-model standards.

In this article we propose a methodology for making the exchannge of clinical data between dual model-based systems based on different standards be able possible. In [10, 11] we described our initial efforts related to the archetypes transformation, here we will focus on our new research results, which conform the last part of the methodology, namely, data transformation. This methodology intends to be generic enough to be applicable to any system based on the use of archetypes, and we applied it to ISO 13606 and openEHR. The first one has gained importance in Europe in recent years and a sign of it can be considered that in 2008 was selected interoperability standard in Sweden [12]. On the other hand, openEHR has a large community of users and developers, and it is used in countries such as Australia or Holland. As a result, there are many open-source tools and archetypes available for it that are useful for validating the methodology.

Next, the structure of this work is described. First, Section 2 provide some background about EHR standards and some of the existing interoperability approaches. Then, in Section 3 the methodology proposed is presented. This section will summarise our previous work and describe the solution proposed for data transformation. Finally, some discussion and conclusions are put forward in Section 4.

2. Background

This section provides a brief introduction to EHR standards, focusing on those following the dual model architecture. Afterwards, a description about clinical data and archetypes will be provided. This will include how both are represented and used in this work. Finally, some recent EHR interoperability approaches will be presented.

2.1. EHR standards

In order to manage that interoperability of clinical information, the digitalization of clinical records has been a major research issue for the last decades. An increasing number of countries have been making efforts in the digitalization of clinical records since the GEHR project [13] began the design of an electronic health architecture in the late 80s. As a consequence of those efforts, several EHR standards were proposed, including HL7, openEHR and ISO 13606. The last two mentioned standards follow a dual model architecture. It is based on the definition of two different conceptual levels, namely, information and knowledge. Each level uses a different data model, thus, the information level uses the reference model and the knowledge level uses the archetype one. The reference model represents the information level and defines the set of entities that form the building blocks of the electronic healthcare record. The archetype model is used to define clinical concepts in the form of structured and constrained combinations of the entities contained in the reference model, so clinical knowledge is defined at this level. These clinical concepts are named archetypes and are usually defined using the Archetype Definition Language (ADL) [14]. This language provides an abstract syntax which can be used to express archetypes for any reference model in a standard way.

Therefore, both ISO 13606 and openEHR share the same archetype model but have different reference models. OpenEHR defines richer data structures and types than ISO 13606. Thus, openEHR allows defining clinical information with more detail. An extract of both reference models is shown in Figure 1 . Both models consist of a set of concepts that organize the EHR information hierarchically. In openEHR and ISO 13606 the clinical information of a patient is usually defined by using a *COMPOSITION* concept, which is similar to a clinical form, and several *COMPOSITION* can be grouped in a *FOLDER*. The headers of a form in which clinical information is classified are represented by means of *SECTION* entities and in each one, clinical concepts can be defined by using *ENTRY* concepts. In ISO 13606 there is one *ENTRY* type, but, in openEHR, there are some types of *ENTRY* according to the information they contain, namely, *ADMIN_ENTRY*, *OBSERVATION*, *INSTRUCTION*, *EVALUATION* or *ACTION*. In addition to this, openEHR also defines a *GENERIC_ENTRY* entity to facilitate interoperability with other standards such as ISO 13606. The information contained in an *ENTRY* is organized by using a table, list or tree by means of a *CLUSTER* in ISO 13606 or specific data structures such as *ITEM_TABLE*, *ITEM_TREE*, *ITEM_LIST* or *ITEM_SINGLE* in openEHR. Finally, the leaf node of the EHR hierarchy is the *ELEMENT* that will contain a data value.

2.2. Archetypes

As it has been aforementioned archetypes can be used to define clinical concepts such as the heart rate, a laboratory test, a blood pressure measurement, etc. An archetype can be defined as a specialization of another one, can include another archetype fragment in it, can be used in combination with others by means of templates, and so on. They constitute a standardized way of capturing clinical data according to a knowledge model. This

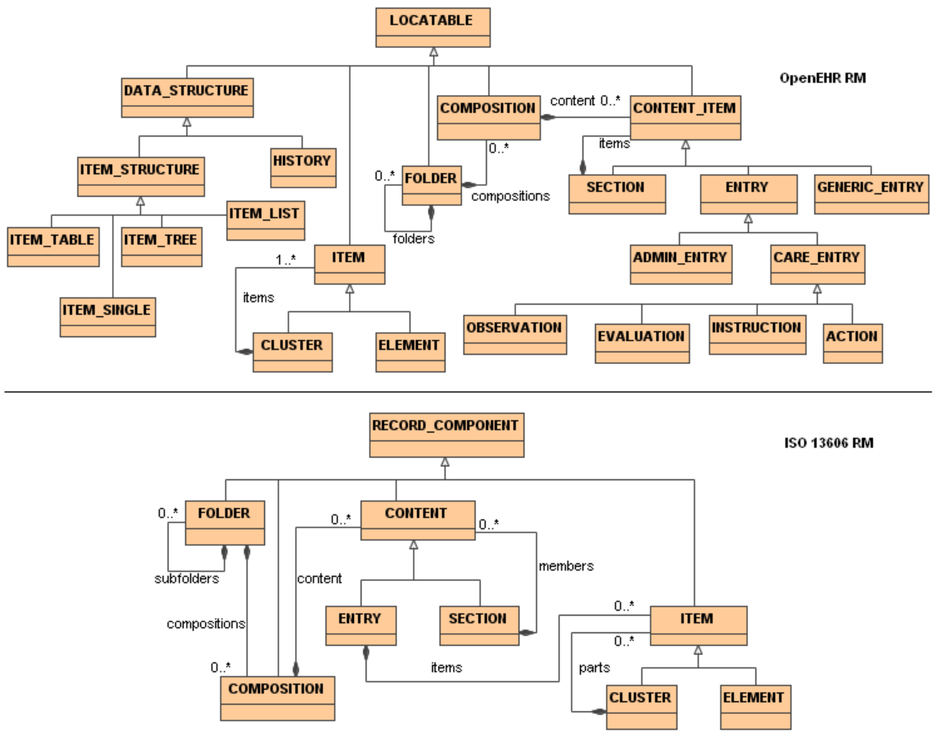

Figure 1: OpenEHR and ISO 13606 main data structures

model is the archetype itself, which provides the context for interpreting the clinical information.

They are usually defined in ADL and the following four main sections can be distinguished in it: header, description, definition and ontology. The first two ones give information about the name of the archetype, the language it is written in, original author, purpose, etc. The definition section is the most important one because it contains the structures and constraints associated with the clinical concept defined by the archetype. The ontology section provides the textual description for each concept from the definition section and bindings to other terminologies.

Figure 2 shows an excerpt of the definition part of the ISO 13606 ADL archetype for defining regular drugs. The root node of this archetype is the composition (*COMPOSITION[at0000]*) which comprises the *ENTRY[at0001]*. This *ENTRY* contains the following *ELEMENTs*:

- The *name* (*ELEMENT[at0008]*), *route* (*ELEMENT[at0006]*) and *side effects* (*ELEMENT[at0002]*) of the drug are defined by using a text type (*SIMPLE_TEXT[at0018]*, *SIMPLE_TEXT[at0016]* and *SIMPLE_TEXT [at0014]*). The *name* and *side effects* are free text, but the *route* is defined by a list of options (*Auricular*, *Buccal*, *Oral*, etc).

- The period of time (*ELEMENT[at0009]*) in which the drug is taken, is defined by means of a boolean type (*BL[at0019]*) that says whether it is during a long term.

- The *dosage* of the drug (*ELEMENT[at0003]*) by using a physical quantity type (*PQ[at0024]*) that constrains the *unit* of measure to (*mg*) and the *value* to be a positive quantity.

```
definition
   COMPOSITION[at0000] occurrences matches {1..1} matches {   -- Regular_Drugs
      content cardinality matches{0..*; unordered; unique} matches {
         ENTRY[at0001] occurrences matches {0..*} matches {   -- Regular Drug
            items cardinality matches{0..*; unordered; unique} matches {
               ELEMENT[at0002] occurrences matches {0..*} matches {   -- Side effects
                  value matches {0..1} matches {
                     SIMPLE_TEXT[at0014] occurrences matches {0..1} matches {
                        originalText existence matches {0..1} matches {*}
                     }}}
               ELEMENT[at0003] occurrences matches {1..1} matches {   -- Dosage
                  value matches {
                     PQ[at0024] occurrences matches {0..1} matches {
                        units matches {
                           CS[at0025] occurrences matches {0..1} matches {
                              codeValue matches{"mg"}
                        }}
                        value matches {|>=0.0|}
                     }}}
               ELEMENT[at0006] occurrences matches {1..1} matches {   -- Route
                  value matches {
                     SIMPLE_TEXT[at0016] occurrences matches {0..1} matches {
                        originalText matches {
                          "Auricular","Conjuctival","Buccal","Cutaneous","Oral"}
                     }}}
               ELEMENT[at0008] occurrences matches {1..1} matches {   -- Name
                  value existence matches {0..1} matches {
                     SIMPLE_TEXT[at0018] occurrences matches {0..1} matches {
                        originalText matches {*}
                     }}}
               ELEMENT[at0009] occurrences matches {0..1} matches {   -- Is long term
                  value matches {
                     BL[at0019] matches {
                        value matches {True, False}
                     }}...}}}}
```

Figure 2: Excerpt of the ISO 13606 regular drugs archetype

2.3. Data Extracts

In dual model-based EHR standards, data is usually represented in XML according to the schemas defined for each standard, although some standards do not have an official one to date. In this work, we will use the non-official XML Schema defined for ISO 13606 that is publicly available at [15], and the official one for openEHR [8].

Figure 3 shows an example of data extract for the ISO 13606 regular drugs archetype. This extract contains the definition of the *name* of the drug (*Ranitidine*), its *dosage* (*75mg*) and its *route* (*Oral*). Moreover, it also mentions that the data has been captured using the regular drugs archetype. The complete extract can be found in [16].

```
<EXTRACTS>
   <archetype_ID>CEN-EN13606-COMPOSITION.Regular_Drugs.v1</archetype_ID>
   <extract>
      <id_pat>0012</id_pat>
      <data type="COMPOSITION">
         <name type="SIMPLE_TEXT">
            <originalText>Regular Drugs</originalText>
         </name>
         <archetype_id>at0000</archetype_id>
         <content type="ENTRY">
            <name type="SIMPLE_TEXT">
               <originalText>Regular Drug</originalText>
            </name>
            <archetype_id>at0001</archetype_id>
            <items type="ELEMENT">
               <name type="SIMPLE_TEXT">
                  <originalText>Dosage</originalText>
               </name>
               <archetype_id>at0003</archetype_id>
               <value type="PQ">
                  <archetype_id>at0024</archetype_id>
                  <value>75</value>
                  <units type="CS">
                     <archetype_id>at0025</archetype_id>
                     <codeValue>mg</codeValue>
                  </units>
               </value>
            </items>
            <items xsi:type="ELEMENT">
               <name xsi:type="SIMPLE_TEXT">
                  <originalText>Route</originalText>
               </name>
               <archetype_id>at0006</archetype_id>
               <value xsi:type="SIMPLE_TEXT">
                  <archetype_id>at0016</archetype_id>
                  <originalText>ORAL</originalText>
               </value>
            </items>
            <items xsi:type="ELEMENT">
               <name xsi:type="SIMPLE_TEXT">
                  <originalText>Name</originalText>
               </name>
               <archetype_id>at0008</archetype_id>
               <value xsi:type="SIMPLE_TEXT">
                  <archetype_id>at0018</archetype_id>
                  <originalText>Ranitidine</originalText>
               </value>
            </items>
            ...
         </content>
      </data>
   </extract>
</EXTRACTS>
```


Figure 3: ISO 13606 Regular Drugs XML data extract

2.4. EHR interoperability approaches

In the last years many projects have pursued the interoperability of EHR information systems. The different approaches have proposed solutions based on specific standards and technologies in order to satisfy the needs of a particular scenario, but no global interoperability frameworks have been provided so far, as we describe next.

The first group of approaches uses XML technologies for achieving interoperability. An example is the interoperability experience between a propietary system and HL7 CDA that has been run in Taiwan [17]. In this solution, both representation models are represented in XML and their mappings are defined by using XSLT rules. The second group of approaches makes use of XML and OWL, although with different purposes. In [18], the interoperability among different health care systems conforming to different EHR standards was pursued by annotating the Web Service messages through archetypes defined in OWL. In that approach mappings between these archetypes are defined manually by using a graphical tool. Then, these mappings are applied to XML clinical data instances. The same research group presented in [19] an approach based on archetypes, ontologies and semantic techniques for the interoperability between HL7 CDA and ISO 13606 systems. This work shares with their previous results that archetypes are represented in OWL, but they approach the problem by using HL7 RIM as basic and common information model. Consequently, this information model supports the semiautomatic definition of the mappings between the standards.The mappings are defined between archetypes by comparing whether the structures used in both standards are specializations of the same class of the HL7 RIM by using reasoning mechanisms. The resulting mappings are applied to XML clinical data, which are transformed to OWL for the reasoning and then transformed back to XML. Another real interoperability experience is the one between openEHR and the Cambio COSMIC system [20]. In this work mappings between both information models were defined. This solution, as the one run in Taiwan, is oriented to solve the specific interoperability problem of these two systems and therefore difficult to adapt to its use with other clinical models or standards.

Given the importance of the semantic interoperability, the main clinical standardization organizations, ISO, HL7 and openEHR, are making an effort for harmonizing their specifications. An evidence of this is the part three of ISO 13606 named ”Reference archetypes and term lists” that provides an informative guide in order to represent clinical information codified according to HL7 and openEHR by using ISO 13606 structures. Moreover, that specification states that, in order to achieve full semantic interoperability of systems based on these norms, it is necessary to harmonize the vocabulary and the data types used. In order to harmonize data types, the ISO 21090 standard is being developed. In the near future the current EHR standards are propable going to support that norm. In order to harmonize the vocabulary used, clinical terminologies will play an important role. On the other side, the openEHR Foundation is defining the mappings between ISO 13606 and openEHR data structures and between openEHR and ISO 21090 data types. In addition to this, Detailed Clinical Models (DCM) [21] have been recently defined as an initiative for defining clinical information independently of an specific clinical standard but with the aim of offering the possibility of being transformed into other standards.

In summary, each particular solution help to provide access to the patient clinical information to some clinical organizations. One of the disadvantages of those approaches is that they do not propose frameworks and methods easy to apply to different standards. On the one hand, XML technologies provide a limited semantic model which does not seem the best solution for achieving the semantic interoperability objective since, for instance, the integration of clinical data with terminologies would be difficult. In particular, the usage of XSLT for defining the mapping rules is no longer the best available option given the availability of mapping and transformation languages created by the Model-driven Engineering community which have better properties in terms of maintenance, reusability and support to software development processes. On the other hand, the use of OWL and ontologies for supporting such interoperability processes is a good practice that is recommended by the Semantic Health report, so global solutions should make use of such technology.

3. Method

In this section the method for transforming clinical data instances between dual model systems is presented. This method reuses some pieces developed by us in previous works, which will be briefly described in this section because they are fundamental for understanding the method for data transformation. Then, the implementation and validation of the method will be reported.

3.1. Overview

The method proposed here intends to enable systems based on different dual model standards to exchange clinical data. Our scenario comprises two EHR systems, namely, A and B, that have been used for capturing the EHR of a particular citizen, and A needs to retrieve some information stored in B. In an ideal situation, both systems would share a representation standard and semantic interoperability would be granted, but this situation is unrealistic nowadays. If we assume that both systems use different standards based on the dual-model architecture, a solution for the achievement of the semantic interoperability could be the one shown in Figure 4. There, two different steps are identified:

- Archetype transformation: the archetypes used in B for capturing the data to be exchanged are transformed into archetypes for A.
- Data transformation: the data captured in B are transformed into data for A.

Our solution for the interoperability of data will provide methods for both archetype and data transformation. This work constitutes then our third step towards the achievement of the semantic interoperability between ISO 13606 and openEHR. Due to this, we will be able to reuse two major previous results: (1) a method for obtaining the ontological representation of archetypes [10]; and (2) a method for transforming archetypes between ISO 13606 and openEHR standards [11].

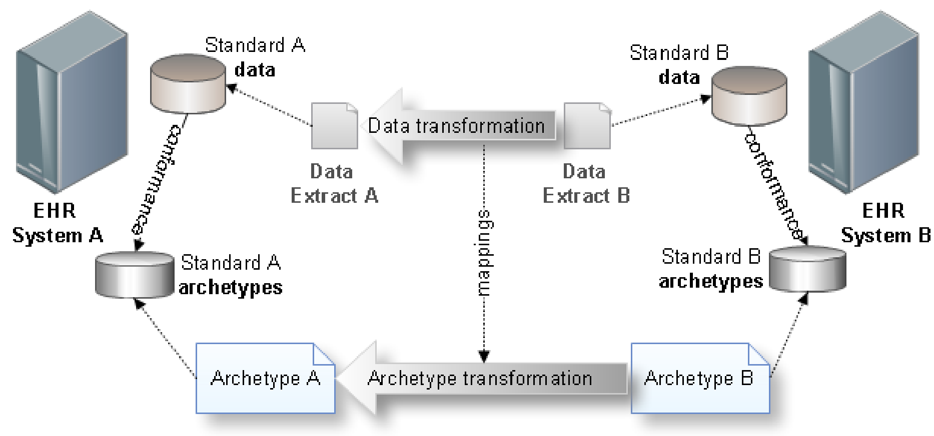

Figure 4: Data Interoperability between two dual model-based EHR systems

3.2. Ontology infrastructure

In dual model architectures, the reference model, the archetype model and terminological definitions are each other related by using string properties, which is not an optimal decision for pursuing semantic goals. In order to implement the interoperability approach above presented it would be better to provide a more semantically precise definition of those relations. For this purpose, we designed an ontology infrastructure for the consecution of interoperability [22], which is shown in Figure5 for ISO 13606 and openEHR.

This infrastructure comprises a series of ontologies that were the result of the semantic interpretation of both reference and archetype models of the standards. Both ISO 13606 and openEHR ontologies combine concepts from both reference and archetype models, because they are reusing the specific ontologies developed for the reference models (ISO 13606-RM and openEHR-RM) and the archetype model (common to both standards). The archetype structure is expressed in these ontologies in a more processing-friendly and

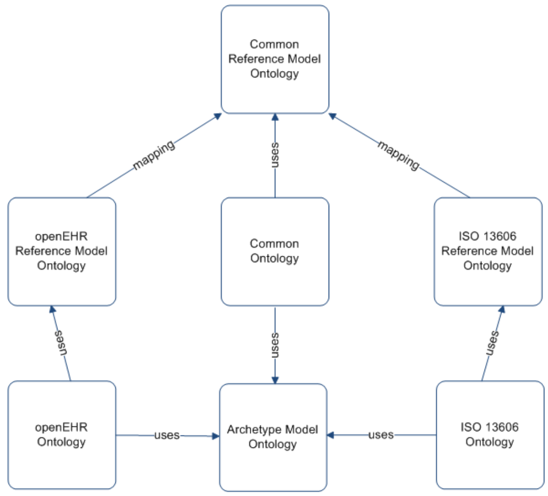

Figure 5: Ontologies relations map

comprehensible way. For instance, concepts as archetype, archetype description, archetype description item, occurrences, clinical datatype or archetype term exist in the archetype model, but more standard-specific ones such as folder, composition, section, entry or element belong to the reference model.

In order to define the transformations between both standards, the definition of these ontologies might be enough. However, we aim to develop a generic and extensible architecture, capable of dealing in the future with other standards such as HL7 CDA. This self-imposed requirement led us to develop a common ontology for EHR standards. This ontology covers the global aspects of archetypes in dual model approaches and offers a common representation for them. This collection of ontologies can be retrieved from [23].

3.3. Archetype Transformation

The archetype transformation can be considered an initial step towards the transformation of data instances. In [11] we presented an approach for transforming archetypes between ISO 13606 and openEHR standards. There, both reference models were analyzed by identifying their common and disjoint entities. As a result, a set of conceptual mappings at concept and property level for defining how a particular component of an archetype can be transformed into a component of the other standard was provided [24].

As it has been mentioned in Section 2.1 both standards use the same archetype model but different reference models. Every ISO 13606 entity has a similar one defined in openEHR, but the opposite does not happen because openEHR provides richer data structures and data types. For instance, the openEHR *ENTRY* is an abstract concept with a number of specializations but the ISO 13606 *ENTRY* is a concrete entity without specializations. Thus, all openEHR *ENTRYs* have to be transformed into the same concept in ISO 13606. This transformation requires adapting their internal structure by means of *CLUSTER* and *ELEMENT* entities in ISO 13606. Moreover, the openEHR reference model contains more entities than ISO 13606. For instance, the openEHR *ITEM_STRUCTURE* hierarchy has no direct correspondence in ISO 13606. Consequently, these entities are also represented in ISO 13606 as a *CLUSTER/ELEMENT* composition.

Therefore, mappings from openEHR to ISO 13606 might be interpreted as a generalization. In order to keep the meaning of the clinical data, each ISO 13606 data structure preserves the name of the openEHR property and type in the *meaning* property, which is defined for every ISO 13606 data structure. This information could be used in the target system to interpret more accurately the transformed archetype.

Otherwise, every ISO 13606 concept has a corresponding one in openEHR, although their properties might also differ. In this case, the property can be mapped into a *CLUSTER* or *ELEMENT*, and the name of the ISO 13606 property is preserved in the *uid* property of openEHR data structures. This attribute has been used because it is not currently needed in this standard.

With regard to data types, each standard defines its own set. The number of data types in ISO 13606 is lower than in openEHR, and some common ones are defined differently. Moreover, a *CLUSTER/ELEMENT* composition plus using the *meaning* property is not applicable for data types. Therefore, lossless mappings cannot be granted for some of them. For instance, the property *codingSchemeName* of the ISO 13606 data type Physical Quantity (*PQ*), does not have a similar property in the openEHR *DV_QUANTITY* type. Thus, this information could not be transformed. According to our experience, nearly all situations in which some information cannot be transformed is related to some particular data types. However, those spurious losses do not prevent from getting the archetypes transformed. Moreover, the method manages and reports such losses.

The architecture proposed for transforming ISO 13606 archetypes into openEHR ones is shown in Figure 6. There, two layers can be distinguished: ontology and metamodel.

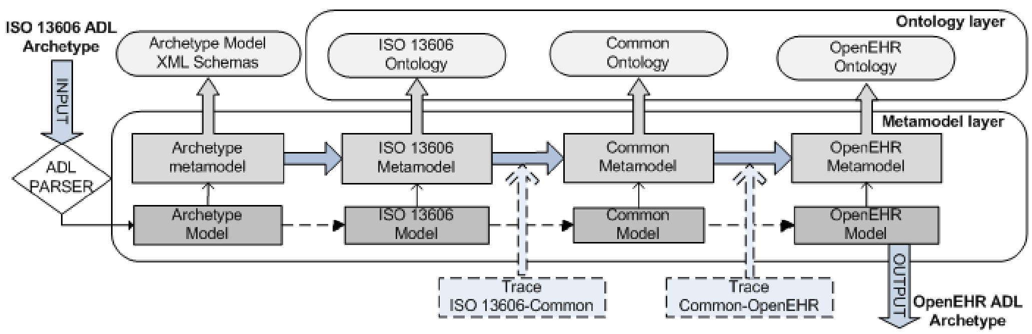

Figure 6: Archetypes transformation architecture

The ontology layer provides the formal semantics of our domain, and is composed of the ontologies previously described. The metamodel layer contains the metamodels corresponding to the archetype model and to the representation defined in the ontology layer, that is metamodels for the ISO 13606, openEHR and Common ontologies. The transformation mappings are formalized and the transformation of archetypes is done in the metamodel layer due to the availability and maturity of tools provided by Model-driven engineering (MDE). The conceptual mappings are defined between the particular standard, ISO 13606 or openEHR and the Common metamodel at concept and property levels and are implemented by using a model transformation language. The archetype transformation method requires these four phases:

1. The ISO13606 ADL archetype is transformed into a model according to the ontological representation of ISO 13606.
2. The ontological ISO 13606 model is transformed into a model conforming to the Common representation.

3. The Common model is transformed into a model conforming to the ontological representation of openEHR.
4. The openEHR model is transformed into ADL.

The process from openEHR to ISO 13606 would be similar. If this transformation method is applied to the ISO 13606 regular drugs archetype presented in Section 2.2 the ISO 13606 *COMPOSITION* will be transformed into an openEHR *COMPOSITION* concept. In ISO 13606, it organizes its content by using an *ENTRY* but in openEHR an *ENTRY* is an abstract concept that is specialized in other ones more specific such as *OBSERVATION*, *ACTION*, *INSTRUCTION*, etc. Here, it will be represented as a *GENERIC_ENTRY*. The previous ISO 13606 *ELEMENTs* will be now openEHR ones that will be grouped by a tree structure (*ITEM_TREE*) and each ISO 13606 data type will be transformed into its corresponding openEHR representation.

3.4. Data Transformation

In this section we describe the new developments of the framework to be able to transform data between dual-model EHR standards. First, we had to extend the archetype transformation method. This method is able now to manage and use the trace of the transformation process. As it can be observed in Figure 6, the transformation of archetypes is defined between the specific standards and the common representation. This transformation process creates two trace models for each archetype transformation, which reports the transformations that have been performed in the process.

Both trace models can be now processed for obtaining the corresponding semantic trace. A semantic trace contains semantic mappings, which associate the concepts and properties of the source and target archetype. They are named semantic, because they are obtained from the ontological representation of the archetype. Semantic mappings must be distinguished from the conceptual ones. The conceptual mappings are the transformations defined at reference model level and the semantic mappings are the transformations that have been executed for a particular archetype. Each semantic mapping has two semantic paths represented as a list of [concept]property/ elements and which provide the route of concepts and properties of the ontological representation of the archetype. Figure 7 depicts the semantic mappings obtained for the *units* property of the *dosage* from the regular drugs example.

```
[ARCHETYPE_CEN-EN13606-COMPOSITION.Regular_Drugs.v1]defines/
   [cen_COMPOSITION_at0000]content/
      [cen_ENTRY_at0001]items/
         [cen_ELEMENT_at0003]value/
            [cen_PQ_at0024]units/
               [cen_CS_at0025]codeValue/

[ARCHETYPE_openEHR-EHR-COMPOSITION.Regular_Drugs.v1]defines/
   [openehr_COMPOSITION_at0000]openehr_content/
      [openehr_GENERIC_ENTRY_at0001]openehr_generic_data/
         [openehr_ITEM_TREE]openehr_items/
            [openehr_ELEMENT_at0003]openehr_value_element/
               [openehr_DV_QUANTITY_at0024]openehr_quantity_units/
```

Figure 7: The semantic mapping for units

However, in order to transform clinical data instances, semantic mappings are not enough. They provide the conceptual mappings established between the ontological representation of archetypes but clinical data are defined as XML extracts according to ADL archetypes. Both ADL archetypes and data extracts are organized in a tree structure, so their elements can be accessed by using a path which accounts for the nodes from the root of the tree to the desired one. The path for accessing the archetype nodes is named ADL path [25]. It is formed from an alternation of segments made up of an attribute name and an optional object node identifier predicate delimited by brackets and separated by slash characters. The ADL path for the value of *units* in the ISO 13606 regular drugs archetype example is */content[at0001]/items[at0003]/value[at0024]/units[at0025]/codeValue*. It is built by the nodes from the root of the archetype to the *codeValue* property of *units*. The corresponding path in a sample data extract would be the same but adding the extract header, */extract[id]/data/content[at0001]/items[at0003] /value[at0024]/units[at0025]/codeValue*. In this work, these paths will be called syntactic paths because they are obtained from the XML representation of the extracts or the archetypes.

Thus, in order to allow the transformation of data extracts the alignment of each piece of data is needed, that is, the mapping from the source extract to the target extract. Each piece of data is in this case identified by a syntactic path. Such mappings at data level are named syntactic mappings and are obtained from the semantic mappings.

The general process for data transformation consists of two steps, namely, generation of the syntactic mappings and transformation of the data. This process for ISO 13606 and openEHR is depicted in Figure 8. There, it can be observed how the semantic trace obtained from the archetype transformation permits obtaining the syntactic mappings established between the ISO 13606 and openEHR ADL archetype representations, that is, the pairs of source and target syntactic paths. In the figure, an ISO 13606 data extract has to be transformed into openEHR. Thus, the source syntactic paths will be the ones obtained from the ADL ISO 13606 archetype representation, and the target syntactic ones will be those corresponding to the ADL openEHR archetype. Once all the syntactic mappings have been obtained, each source syntactic path will permit accessing data in the ISO 13606 extract, and each target syntactic one will provide the path for defining such data according to the openEHR representation.

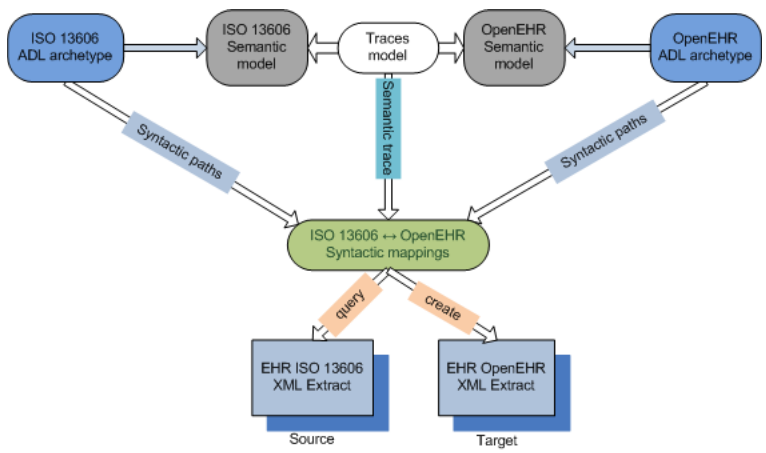

Figure 8: ISO 13606 and openEHR data transformation schema

Therefore, in order to transform the ISO 13606 regular drugs patient data presented in Section 2.3 to their openEHR representation, two main tasks are needed: (1) transformation of the ISO 13606 regular drugs archetype into its representation according to the openEHR standard; and (2) transformation of the ISO 13606 regular drugs data extract in an openEHR extract.

The first task has already been explained in Section 3.3. As a result of this task, the semantic trace between the concepts and properties of the regular drugs archetype for both standards is obtained. In order to carry out the second task, the semantic trace is used for generating the syntactic mappings. These last ones will provide the route of each data unit in the source data extract, namely, source syntactic path, and the corresponding one in the target data extract, namely, target syntactic path.

The transformation of the ISO 13606 extract generates the result published at [16], and whose excerpt is shown in Figure 9. There, it can be observed how the different data structures and types have been transformed into the corresponding openEHR ones like *GENERIC_ENTRY*, *DV_TEXT*, or *DV_QUANTITY*, etc. The structure of the content section of this transformed extract is described next. The type of data is *COMPOSITION*. Its content is a *GENERIC_ENTRY*, whose *id* is *at0001*. The data associated with this entry is an *ITEM_TREE*, whose items include an *ELEMENT*. The name of this element has the value *Dosage*, and the value of the element is a *DV_QUANTITY*, whose *id* is *at0024*. An openEHR *DV_QUANTITY* has the properties *magnitude* and *units*, whose values are, respectively, *75* and *mg*. Such values have been obtained by applying the corresponding semantic and syntactic mappings.

```
<EXTRACTS>
   <archetype_ID>
      openEHR-EHR-COMPOSITION.Regular_Drugs.v1
   </archetype_ID>
   <extract>
      <id_pat>0012</id_pat>
      <data type="COMPOSITION">
         <name type="DV_TEXT">
            <value>Regular Drugs</value>
         </name>
         <archetype_id>at0000</archetype_id>
         <content type="GENERIC_ENTRY">
            <name type="DV_TEXT">
               <value>Regular Drug</value>
          </name>
            <archetype_id>at0001</archetype_id>
            <data type="ITEM_TREE">
               <items type="ELEMENT">
                  <name type="DV_TEXT">
                     <value>Dosage</value>
                  </name>
                  <archetype_id>at0003</archetype_id>
                  <value type="DV_QUANTITY">
                     <archetype_id>at0024</archetype_id>
                     <magnitude>75</magnitude>
                     <units>mg</units>
                  </value>
               </items>
               ...
            </data>
         </content>
      </data>
   </extract>
</EXTRACTS>
```

Figure 9: OpenEHR Regular Drugs XML data extract

3.5. Validation

This transformation method has been validated by using Poseacle Converter [26]. This tool provides functionality for both the transformation of an ADL archetype into OWL [27] and the transformation of ISO 13606 archetypes into openEHR ones and vice versa. The Poseacle Converter combines different technologies and languages included in our approach: Model-driven Engineering, Semantic Web and Java technologies. Apart from the online access to the tool a RESTful service for doing the transformation and an API to invoke it have been developed.

As a result of the current research work, an option for data transformation has been added to this tool. The option has been added as the tab *Data Transformation* (see Figure 10) whose interface is very similar to the one for archetype transformation. In order to get an extract transformed, the user has to perform the following actions:

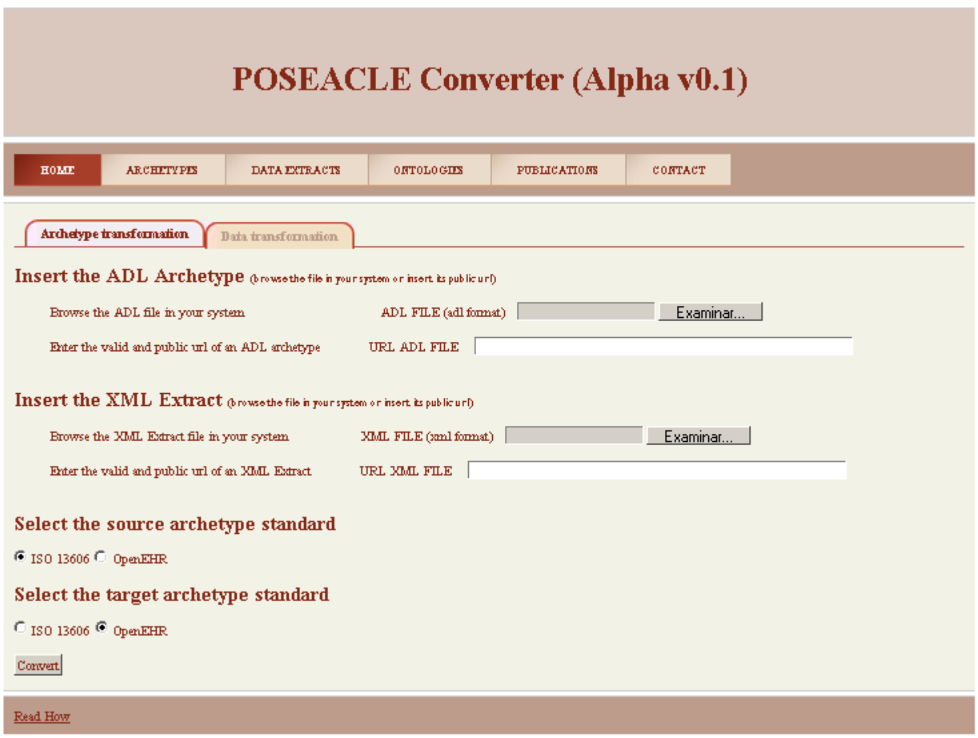

Figure 10: Poseacle Converter: ISO 13606/openEHR data transformation

- Input the ADL archetype by browsing the local file system or entering its URL.

- Input the XML data extract by browsing the local file system or entering its URL.
- Select the source and target EHR standard (ISO 13606/openEHR).

We have performed a technical validation of the process. This means that we have executed a series of data transformation processes and the results have been analyzed. The first set of tests were focused on the correct execution of the transformation rules. For this purpose, unit tests were executed and we checked the correctness by comparing the expected and the actual results.

Next, we tested the correct transformation of data extracts. For it, we manually created some sample extracts and we used some extracts facilitated by some colleagues. Those extracts were transformed both manually and by the Poseacle Converter and the correctness of the results was checked. The publicly available corresponding data extracts can be accessed in [16]. The results derived from the data transformation are in line with the ones derived from the archetype transformation. The data that are not transformed are those which corresponds to properties or concepts that have not been transformed at archetype level.

Finally, we analyzed the time performance of the process by using the Poseacle Converter. In order to evaluate this performance, it should be noted that the server has a 2.27GHz Core 2 Duo processor and 4 GB of memory, which is not fully dedicated to this tool. The process includes transformations at both archetype and data level. Consequently, both times will be considered independently. Our results show that,as expected, the time spent transformation of the archetypes and data are directly related to the size of the archetype and the extract. In general, the transformation at archetype level takes longer than the ones at data level, although both times are, in our opinion, satisfactory. The time spent in the archetype transformation ranged, in our experiments, between two and seven seconds, whereas data transformations took between one and two seconds. This would make the complete process to take between three and nine seconds. I should be pointed out that, in the context of a real deployment of the system, a library of transformations would avoid the need for the transformation at archetype level in most cases, so the global transformation time would be little more than the time of the transformation at data level.

4. Discussion and conclusions

The achievement of semantic interoperability of EHR systems is considered essential to improve the quality and safety of patient care, public health, clinical research, and health service management. The dual model architecture was developed with the aim of facilitating the semantic interoperability of clinical information systems. Unfortunately, a limited number of EHR systems make use of such standards for representing the clinical information and this makes the achievement of semantic interoperability harder.

In dual model architectures, archetypes are considered the basic interoperability unit. Nowadays, there are several standards based on this architecture. In order to enable the information and knowledge sharing between them, transformation methods have been developed by our research group in recent years. Hence, this works aims to offer mechanisms that help the adoption of dual model architectures, by providing mechanisms for the exchange of clinical information between different standards. More concretely, we have addressed in this work the transformation of archetyped-data between ISO 13606 and openEHR.

The present work reuses some of our previous results, such as the ontological infrastructure [28], the ADL to OWL transformation tool [10] and the archetype transformation methods [11]. They were focused on archetypes and their transformation between standards and have provided us the infrastructure for carrying out the work we present here, the transformation of clinical data instances. This infrastructure will be also the technological support for our future research and developments. The archetype transformation has been integrated into the LinkEHR [29] editor and also has been used in some tools developed in the research group which has contributed to better technical validating the different modules and tools developed. Since the beginning of the development of the methodology, the fact that it was based on previously developed methods and tools has allowed us to make them some improvements or extensions in order to facilitate the achievement of the next goals. For instance, among the improvements, the trace of the archetype transformation can be obtained as a result of this work. Apart from being useful for data transformation, it has allowed us to provide the user with an informative report of the process, that is, what has been transformed and what has not. The good results and the efficiency of the process makes us believe that our design decisions have been adequate in the last years.

The backbone of our technological framework combines ontologies and metamodels. Both technological worlds are commonly accepted for the pursuit of semantic interoperability, even without being combined. The crucial role played by ontologies in interoperability settings is described in [30], and the recent effort presented in [31] uses a common ontology for facilitating the interoperability between ECG data standards. On the other hand, models drive the semantic services oriented architecture used in [32]. Hence, our technological infrastructure is based on the ontological representation of archetypes that will allow us to use in the future external terminological resources in order to improve the archetype and data transformation. Moreover, the representation of archetypes as ontological models has provided us with an interesting tool for a natural access to the archetype information, in which from each concept it is possible to access to all its related information: occurrences, cardinality, definition, bindings, etc.

In our approach, the transformation of data instances consists of two steps: (1) archetype transformation and (2) data transformation. The transformation of archetypes is executed as described in [11]. This process is based on the conceptual mappings between the reference models of ISO 13606 and openEHR defined by us, so they do not correspond to any community consensus. In fact, the openEHR Foundation is defining its own mappings [33], which are similar to ours. The archetype transformation process had to be extended to deal with the needs of the data transformation step. Thus, the new archetype transformation process generates the semantic trace which accounts for the set of transformations applied to the source archetype in order to get the target one. Such effective transformations are called semantic mappings; they consist of the paths of the archetype concept or property in the source and target ontological archetype.

The method proposed for transforming archetyped data instances makes use of that semantic trace and those semantic mappings to obtain the syntactic mappings. Each syntactic mapping includes two syntactic paths. Each syntactic path stands for the path of a particular archetype concept or property in the syntactic ADL archetype. The syntactic mappings are used for accessing the source clinical data and defining the target data extract in the target standard representation. The transformation of data instances might have been done by using XSLT transformations and, in such solutions, the semantic and syntactic mappings would not be required. However, as it has been previously mentioned, XML technologies have a limited semantics which make them not optimal for the semantic interoperability challenge. This does not mean that such technologies should not be used, since they are currently used by most standards for the representation of data extracts, but they should not constitute the technological kernel for defining and executing the mappings between the EHR standards. The Model-driven Engineering community has developed languages for defining and executing such mapping. Besides, such advanced Software Engineering approach also facilitates the combination of technological spaces different such as ontologies, archetypes and data extracts. Thus, a solution based on such mapping and transformation languages facilitate the development, reusability, extensibility and maintenance of our solution.

The transformation of data instances is not a lossless process. As discussed in [11], the transformation of archetypes may suffer from some loss of information, although this just happens in a reduced number of situations. Most information losses occurrs in the data types transformation and the ISO

21090 standard is being defined in order to provide a set of harmonized data types to be used in information exchanges. Both ISO 13606 and openEHR will likely use this new standard in the near future. Thus, our implementation will be also adapted to support it. In the situations in which some data can not be transformed, the Poseacle Converter does not only shows the data extract, but the user may also ask for the report of the transformation process, which includes the semantic trace of the process, and the data that have not been transformed. Consequently, the user is informed of what has been transformed and what has not, and may take the corresponding decisions and actions.

To the best of our knowledge, this is the first implementation of data interoperability between these standards, although different approaches for semantic interoperability have been proposed in recent years. Most of them are designed to give a solution to a very specific problem. In this way, mappings are usually defined directly betweem the EHR format representations or standards without using any common representation that would enable a more generic solution. Therefore, they resulting solutions are very difficult to mantain and very little adaptable. None of these approaches, including ours, formalize the mappings between the different stantards in an easy-to-reuse manner. Most approaches code the mappings into the software or, as we do, in a transformation language. Therefore, this makes it difficult for other researchers to reuse currently available solutions. Recently, the Ontology Management Working Group proposed a standard for Ontology Mapping [34], although it has not been adopted by the community. In addition to this, there are many mapping languages in Model-driven Engineering. However, the community has not developed a language good enough for defining the mappings between EHR standards.

Our solution is not perfect, since we have not used so far the terminological knowledge contained in the archetypes for optimizing the transformation processes. Thus, further research will be conducted for using SNOMED-CT as part of the transformation process. In fact, this offers a discussion about to what extent our solution is semantic. Our current solution incorporates a series of semantic components, although the transformation process is not semantic in a strict sense. The root of the approach is a set of ontologies for EHR standards written in OWL, which is a language whose construct have a formally defined meaning. Thus, archetypes can be expressed in a formal language such as OWL, and any semantic activity that could not be efficiently done with ADL can be perfomed on them. Hence, once archetypes are expressed in OWL, we can do semantic tasks as the ones included in our Archetype Management System [35]. This system allows for making automatic classification and search of archetypes by annotating them with external resources represented in OWL such as terminologies or other domain ontologies. It would be harder to perform such tasks with ADL archetypes. On the other hand, the mappings between the EHR standards are defined by using their ontological representations, and relations are defined between them, with formally defined meaning. This is therefore another sign of use of semantics in our approach. However, such semantics is not fully exploited in the archetype transformation process.

Our ontologies define the structural semantics of the archetypes and reference models, but terminological knowledge is not seamlessly integrated.

Thus, terminology can be associated with the terms included in the ontological archetypes, although as information units rather than knowledge units. This means that the terminological knowledge is not natively represented as part of the ontological knowledge, so additional processing would be required to combine the structural and terminological knowledge for supporting the transformation process. This does not prevent our data and archetype transformation process from obtaining correct results but, from a semantic point of view, the results could be more precise.

The usage and processing of the *meaning* property allows the method to be reversible, although this has not been implemented yet in our conversion tool. On the one hand, systems will mainly be interested in transforming the data to their own EHR standard. In this context, the reverse transformation is not expected to be used frequently. On the other hand, different transformation systems might interpret that *meaning* property in a different way, so this backward mechanism would make really sense in case the same converter is used for both transformations. Different interpretations of the *meaning* property would generate structurally different archetypes and data extracts.Thus, we did not assign a top priority to this functionality, although we will address this issue in short term.

In summary, we have developed methods for transforming data instances between two important EHR standards, namely, ISO 13606 and openEHR. The research results obtained in the last years have produced an interesting solution for the semantic interoperability of these standards, because we have developed automatic methods for the ontological representation of archetypes and for the transformation of both archetypes and data instances between

ISO 13606 and openEHR. Our results show that such exchange and sharing is possible, and we think that our approach could be applied to other dual model standards. All of them are going to share the same archetype model, so we are going to be able of obtaining models conforming to it without doing any change. Moreover, standards using a similar technological approach, like HL7 CDA, could also be integrated in our framework. For this purpose, we would need to build the ontology for this standard and to define the mappings with the common ontology. The mappings from/to an already integrated standard could be then automatically derived by composition of the mappings from/to the common ontology. The addition of a new standard might make the extension of the common ontology necessary. However, this would mean that the new concepts or properties are not considered in the already integrated standards, thus this new entities could not really be transformed in most cases.

Acknowledgements

This work has been possible thanks to the Spanish Ministry for Science and Education through grants TSI2007-66575-C02-02 and TIN2010-21388-C02-02. Special thanks to Daniel Karlsson, Marcelo Rodrigues dos Santos and Diego Boscá for providing us with data extracts for the evaluation of our method.

[1] European Commision, EU Citizenship Report 2010 : Dismantling the obstacles to EU citizens'rights (COM(2010) 603), Last accessed: April 2011.

[2] epSOS European eHealth project, http://www.epsos.eu/, Last accessed: April 2011.

[3] V. N. Stroetmann, D. Kalra, P. Lewalle, A. Rector, J. M. Rodrigues, K. A. Stroetmann, G. Surjan, B. Ustun, M. Virtanen, P. E. Zanstra, Semantic Interoperability for Better Health and Safer Healthcare. Deployment and Research Roadmap for Europe, ISBN-13 : 978-92-79-11139-6, 2009.

[4] V. Reding, Commission Recommendation of 2 July 2008 on cross-border interoperability of electronic health record systems (notified under document number (C(2008) 3282), Official Journal of the European Union, 2008.

[5] T. Beale, Archetypes: Constraint-based Domain Models for Future-proof Information Systems, in: Edited by Kenneth Baclawski and Haim Kilov. Northeastern University, Boston (Ed.), Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer, pp. 16–32.

[6] D. Kalra, T. Archana, The EHR and Clinical Archetypes: time for clinical engagement!, in: eHealth Planning and Management Symposium, Copenhagen, Denmark.

[7] ISO 13606 standard, http://www.iso.org/iso/home.htm, Last accessed: April 2011.

[8] OpenEHR specification, http://www.openehr.org/home.html, Last accessed: April 2011.

[9] Health Level Seven, http://www.hl7.org, Last accessed: April 2011.

[10] C. Martínez-Costa, M. Menárguez-Tortosa, J. T. Fernández-Breis, J. A. Maldonado, A model-driven approach for representing clinical archetypes for Semantic Web environments, Journal of Biomedical Informatics 42 (2009) 150 –64.

[11] C. Martínez-Costa, M. Menárguez-Tortosa, J. T. Fernández-Breis, An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes, Journal of Biomedical Informatics 43 (2010) 736 –46.

[12] A national decision in Sweden on healthcare informatic standard, http://www.e-recordservices.eu/attachments/019_Info_about_Swedish_decision.pdf, Last accessed: April 2011.

[13] Good European Health Record (GEHR) project, http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/index.htm, Last accessed: April 2011.

[14] OpenEHR Foundation. ADL language, Adl, http://www.openehr.org/releases/1.0.1/architecture/am/adl2.pdf, Last accessed: April 2011.

[15] CEN/ISO 13606 Information Site, http://www.en13606.eu/, Last accessed: April 2011.

[16] Examples of ISO13606/OpenEHR data extracts transformation, http://miuras.inf.um.es/~researchehr/websvn/, Last accessed: April 2011.

[17] W. Jian, C. Hsu, T. Hao, H. Wen, M. Hsu, Y. Lee, Y. Li, P. Chang, Building a portable data and information interoperability infrastructure - framework for a standard Taiwan Electronic Medical Record Template, Computer Methods and Programs in Biomedicine 88 (2007) 102–11.

[18] A. Dogac, G. Laleci, S. Kirbas, Y. Kabak, S. S. Sinir, A. Yildiz, Y. Gurcan, Artemis: Deploying semantically enriched Web services in the healthcare domain, Inf. Syst. 31 (2006) 321–39.

[19] O. Kilic, A. Dogac, Achieving Clinical Statement Interoperability Using R-MIM and Archetype-Based Semantic Transformations, IEEE Transactions on Information Technology in Biomedicine 13 (2006) 467–77.

[20] R. Chen, G. Klein, E. Sundval, D. Karlsson, H. Ahlfeldt, Archetype-based conversion of EHR content models: pilot experience with a regional EHR system, BMC Medical Informatics and Decision Making 9:33 (2009).

[21] Detailed Clinical Models, `http://www.detailedclinicalmodels.org`, Last accessed: April 2011.

[22] J. Fernández-Breis, P. Vivancos-Vicente, M. Menarguez-Tortosa, D. Moner, J. Maldonado, R. Valencia-García, T. Miranda-Mena, R. Martínez-Béjar, Towards the semantic interoperability of ehr information systems., in: Pacific Rim Knowledge Acquisition Workshop,Guilin, China.

[23] Ontology collection, http://miuras.inf.um.es/ researchehr/websvn/, Last accessed: April 2011.

[24] ISO 13606-OpenEHR Mappings, http://miuras.inf.um.es/~researchehr/ISO13606-OpenEHR-Mappings.html, Last accessed: April 2011.

[25] The openEHR Foundation, Archetype Definition Language. ADL 1.4, Last accessed: April 2011.

[26] Poseacle converter, http://miuras.inf.um.es:9080/PoseacleConverter, Last accessed: April 2011.

[27] OWL Web Ontology Language Reference, http://www.w3.org/tr/owl-ref/, Last accessed: April 2011.

[28] J. T. Fernández-Breis, M. Menárguez-Tortosa, D. Moner, R. Valencia-García, J. A. Maldonado, P. J. Vivancos-Vicente, T. Miranda-Mena, R. Martínez-Béjar, An Ontological Infrastructure for the Semantic Integration of Clinical Archetypes, Lecture Notes in Computer Science 4303 (2006) 156–67.

[29] J. A. Maldonado, D. Moner, D. Bosca, J. T. Fernández-Breis, C. Angulo, M. Robles, LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics, International Journal of Medical Informatics 78(8) (2009) 559–70.

[30] B. Blobel, C. González, F. Oemig, D. Lopéz, P. Nykänen, P. Ruotsalainen, The Role of Architecture and Ontology for Interoperability, Studies in Health Technology and Informatics 155 (2010) 33–9.

[31] B. Gonçalves, G. Guizzardia, J. G. Pereira Filhoa, Using an ECG reference ontology for semantic interoperability of ECG data, Journal of Biomedical Informatics In press (2011).

[32] G. A. Komatsoulis, D. B. Warzel, F. W. Hartel, K. Shanbhag, R. Chilukuri, G. Fragoso, S. Coronado, D. Reeves, J. B. Hadfield, C. Ludet, P. A. Covitz, caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability, Journal of Biomedical Informatics 41 (2008) 106–23.

[33] OpenEHR to ISO EN 13606-1 mapping, `http://www.openehr.org/wiki/display/stds/openEHR+to+ISO+13606-1,+ISO+21090+mapping`, Last accessed: April 2011.

[34] Ontology Management Working Group, OMWG D7.2: Ontology Mapping Language RDF/XML Syntax, `http://www.omwg.org/TR/d7/rdf-xml-syntax/`, Last accessed: April 2011.

[35] J. Fernández-Breis, M. Menárguez-Tortosa, C. Martínez-Costa, E. Fernández-Breis, J. Herrero-Sempere, D. M. Cano, J. Sánchez-Cuadrado, R. Valencia-García, M. Robles-Viejo, A Semantic Web-based System for Managing Clinical Archetypes, in: Conf Proc IEEE Eng Med Biol Soc., pp. 1482–5.