

An Ontology-driven Proposal for Semantic Interaction among Heterogeneous Health Information Systems

Idoia Berges, Jesús Bermúdez and Arantza Illarramendi

University of the Basque Country, Paseo Manuel de Lardizabal, 1
20018 Donostia-San Sebastián, Spain

Abstract. The adoption of Electronic Health Records (EHRs) has brought multiple advantages to the healthcare area. However, the goal of achieving semantic interoperability of EHR information between heterogeneous Health Information Systems has not been accomplished yet. In such scenario, the purpose of this paper is twofold: On the one hand, the presentation of our ontology-based approach to the problem of EHR interoperability (restricted to the case of medical observations), which goes one step further with respect to other approaches for the same goal, and on the other hand, the presentation of two additional features that complement our approach: path mappings for transforming individuals that represent EHR information and rules for medical knowledge sharing.

1 Introduction

In 2009 the European Community presented a longer-term research and deployment roadmap that provides the key steps for achieving semantic interoperability in the area of healthcare[1]. The motivation for that is that nowadays the idea of one person receiving health assistance from the same medical institution throughout all his life is no longer realistic. Thus, medical institutions must be prepared to receive patients from other regions or countries without the quality of service being affected. The incorporation some years ago of Electronic Health Records to the institutions may be seen as the first step towards the goal, since, apart from local advantages over manual records such as avoiding legibility problems due to poor handwriting which may lead to misunderstandings, they favour a fast exchange of clinical data between different organizations. However, the fact that most institutions have developed their health information systems in an autonomous way has resulted in a proliferation of heterogeneous health information systems, each one with its own proprietary models for representing and storing EHR information, which difficults the task of interoperating with each other.

In many areas, the adoption of knowledge representation standards stands out as the most usual approach to solve interoperability problems. This happens also in the healthcare area, where some standards such as openEHR¹, CEN-13606² and HL7-CDA³ are

¹ www.openehr.org

² www.en13606.org

³ www.hl7.org

under development for this purpose. All those three follow a dual model-based methodology for representing the information that may appear in an EHR: On the one hand, the Reference Model is a stable model which defines the basic structures for representing EHR information (such as List, Table, etc.). On the other hand, the Archetype Model defines specific knowledge elements (such as Respiration Rate) by using and constraining the elements of the Reference Model. Although the idea of using a standard may seem suitable for the goal, we think that interoperability does not mean to have a unique representation but a semantically acknowledgeable equivalent one. This would relieve medical institutions from being forced to use one standard in the representation of their knowledge and moreover, since several standards are being developed for the same purpose, the interoperability problem will remain unsolved unless these standards merge into a single one.

In this paper we present a proposal to move towards the notion of full semantic interoperability of EHRs, which states that when one particular system receives some EHR information from another institution, the received information can be seamlessly integrated into its underlying repository because the differences in the language, in the representation of the information and in the storing systems do not cause any misunderstanding[1]. Our solution is based on the use of semantic technologies, and more precisely on OWL2[2] ontologies and corresponding reasoners.

In the area of EHRs semantic interoperability a certain number of related works can be found at present. The works mentioned next also rely on semantic technologies that facilitate semantic interoperation between heterogeneous information systems as opposed to other formats for interchanging data such as XML which do not deal with the semantics of the exchanged data[3]. [4] provides a solution to achieve semantic interoperability between systems that have been developed under the HL7 reference model. However, this proposal requires that the source system has some prior knowledge about the target system and moreover, it does not tackle the communication between systems that use proprietary EHR specifications. In [5] ontology mappings are proposed between pairs of archetype-based models. Finally, in [6] a model-driven engineering approach that transforms archetypes of the CEN-13606 standard into OWL models is presented.

The purpose of this paper is twofold: On the one hand, the presentation of our ontology-based approach to the problem of EHR interoperability (restricted to the case of medical observations) and on the other hand, the discussion of some features that complement the current approach -which may be also relevant to other ontology-based interoperability solutions. More specifically, we would like to stress two of them: first, the usefulness of defining a new category of mappings between the elements of two ontologies -called *path mappings*- which indicate some kind of relationship between two property paths in the ontologies and facilitate the transmission of the information about individuals between two ontologies. Secondly, the convenience of incorporating SWRL[7] rules to the ontologies to define and share medical knowledge among institutions.

The rest of the paper is divided as follows: first, a general overview of our ontology-based approach for semantic interoperability of EHRs is presented in section 2. Section

3 tackles the complementary features pinpointed above. Finally conclusions are discussed in section 4.

2 Overview of the Framework

In general, an EHR includes clinical statements such as observations, laboratory tests, diagnostic imaging reports, treatments, therapies, administered drugs and allergies. In this paper we focus on the exchangeability of medical observation statements, which are used to record all notionally objective observations of phenomena and patient-reported phenomena, such as physical examinations, laboratory results or basic information about the patient (weight, sex,...). In Fig.1 the architecture of our proposal is shown.

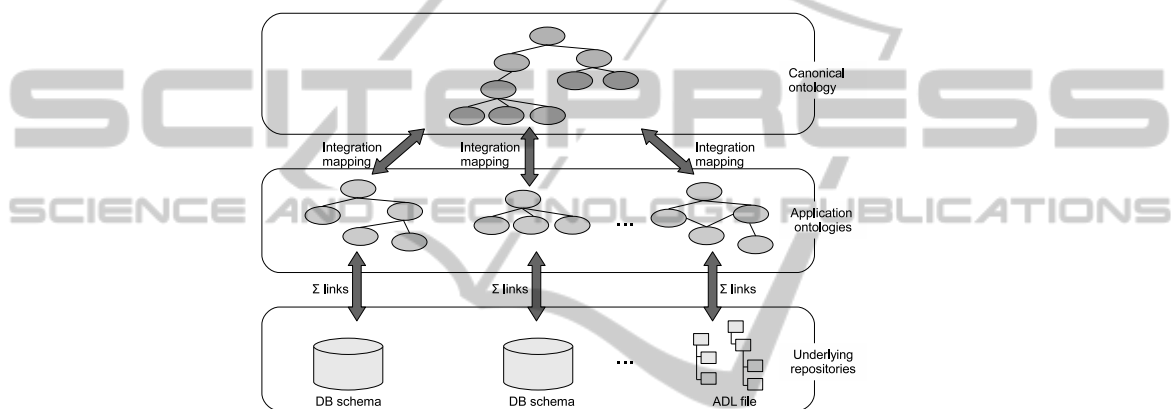


Fig. 1. Architecture of the solution.

This proposal is sustained in one of the approaches for interoperability among systems described in [8]: Using a canonical model to which the particular systems are linked. More precisely, we deal with a Canonical ontology that represents medical observations in a canonical way, that is using a general representation that is independent from the different conceptualizations of them that can exist. In that ontology we propose a subdivision of medical observations into two groups: simple observations and composite observations. Simple observations have a single value and unit of measurement. Additionally, we have also identified three properties that may be relevant at the time of characterizing an observation: *the protocol*, which records information about how the observation process was carried out, either by indicating a particular clinical protocol (e.g. the Balke protocol for treadmill graded exercise testing) or the medical device used for taking the measurement (e.g. a stethoscope); *the anatomical site*, to indicate the specific body location in which the observation was taken; and *the state of the patient*, which is intended to record the state of the subject of the observation during the observation process. In order to represent the information about the protocol in a controlled way, we advocate for using the terms of an ontology that comprises classes from the Device and Procedure categories of SNOMED-CT[9]. Moreover, in order to represent anatomical information, the terms of the Foundational Model of Anatomy

ontology[10] are suggested. Finally, we have developed one ontology for describing information about the state of the patient, which contains 121 classes divided into 28 categories to represent states such as the level of exertion (low, medium, high intensity) or the position of the patient (standing, sitting,...).

On the other hand, composite observations are composed of two or more observations, either simple or composite. They are intended to represent observations of phenomena such as the Blood Pressure, which is composed of the systolic and diastolic blood pressures. Below, we present some OWL2⁴ axioms that represent the general terms for representing medical observations⁵:

$$\begin{aligned} c:\text{Observation} &\equiv c:\text{Simple_Obs} \sqcup c:\text{Comp_Obs} \\ c:\text{Simple_Obs} &\equiv =0 \ c:\text{comp} \\ c:\text{Simple_Obs} &\sqsubseteq =1 \ c:\text{value} \sqcap \leq 1 \ c:\text{unit} \sqcap \leq 1 \ c:\text{protocol}.c:\text{Protocol} \sqcap \\ &\quad \forall c:\text{state}.c:\text{State} \sqcap =1 \ c:\text{site}.c:\text{AnatomicalSite} \\ c:\text{Comp_Obs} &\equiv \geq 2 \ c:\text{comp}.c:\text{Observation} \end{aligned}$$

Specific observations are described as specializations of these general terms.

Other main components of the proposal are Application ontologies, which represent the observations as they are understood in one particular health information system. When such a system wants to join the framework, the following steps must be followed: first its Application ontology has to be defined on top of its underlying data repository. One module named *Internal2OntoModule* has been developed for that task. In some cases the module will receive as input a database schema and after applying a set of rules founded on schema features (tables, keys, inclusion, exclusion and functional dependencies, null values and semantic integrity constraints), it constructs the corresponding ontology components (classes, properties, relations and restrictions). More details about the nature of the rules are described in a previous paper of our research group[12]. In other cases, the input will be an archetype description (e.g. of a EHR standard) written in Archetype Description Language[13] which is transformed to OWL. Moreover, this module is responsible for creating the Σ links (Fig.1) that regulate the information flow between the underlying repositories and the Application ontology, following the guidelines in [14]. Then, the particular system must import the above described Canonical ontology and create the integration mapping that relates the terms of its Application ontology with the terms in the Canonical ontology. A *MappingModule* has been developed for this purpose.

Once a particular system A has joined the framework it is prepared to send information about observations stored in its underlying repository to another system B in the framework. Thanks to the Σ links between the underlying repository and the Application ontology of system A, the information to be sent is converted into instances (individuals) of the classes of that Application ontology. Then, all the implicit knowledge (regarding the individuals) that can be inferred from the Applications layer is made explicit with the help of a reasoner. At this point, the integration mapping that

⁴ For the sake of conciseness we use the Description Logic (DL)[11] representation of axioms.

⁵ Throughout the paper, the namespace 'c' will be used for referring to terms in the Canonical ontology, and namespaces 'a' and 'b' will be used for referring to the ontologies of some specific systems A and B

has been defined between the Application ontology of system A and the Canonical ontology comes into play and as a reasoning result, the individuals are also classified as instances of the concepts of the Canonical ontology. All the inferred knowledge about the individuals is then sent to system B, which asserts it into its ontology. Since system B has also imported the Canonical ontology, this is a straightforward process. Moreover, thanks to the integration mapping between the Application ontology of system B and the Canonical ontology, the individuals are then recognized as instances of the specific terms of system B. Finally, the Σ links between the Application ontology of system B and its underlying repository allow to assert the knowledge into the latter. The whole process described above is directed by a reasoner.

To sum up, the main features of the framework presented in this section are the following:

- It is extensible to any model, either standard or proprietary.
- It is not based on peer-to-peer transformations but on the semantic acknowledgement of one instance of a class in the source ontology as instance of another class in the target ontology.
- The features of any specific system remain unknown to the other systems in the framework. Acknowledging and using the Canonical ontology as a shared model is enough.
- Reasoning plays a major role in several parts of the framework.

3 Additional Features for Complementing the Proposal

In this section two additional features that complement our approach will be discussed: path mappings for transforming information about individuals between two ontologies and rules for knowledge sharing. The usefulness of these features may be also relevant to other ontology-based interoperability solutions. Prior to that, subsection 3.1 is intended to present the definitions of the elements that will appear in the examples of the following subsections.

3.1 Scenario for the Examples

The Revised Trauma Score (RTS)[15] is a physiological scoring system for predicting death taking into account three measures: the Glasgow Coma Scale value, the Systolic Blood Pressure and the Respirations Rate. Moreover, the Glasgow Coma Scale(GCS) is a neurological scale that aims to give a reliable and objective way of recording the conscious state of a patient[16]. It is calculated from the result of three tests: the eye, motor and verbal responses.

Subset S_C of the Canonical Ontology. It contains the definitions of the observations Revised Trauma Score and Glasgow Coma Scale.

$$\begin{aligned} c:RTS &\equiv c:Comp_Obs \sqcap \exists c:comp.c:GCS \sqcap \exists c:comp.c:SysBP \\ &\quad \sqcap \exists c:comp.c:RespRate \\ c:GCS &\equiv c:Comp_Obs \sqcap \exists c:comp.c:EyeR \sqcap \exists c:comp.c:VerbalR \\ &\quad \sqcap \exists c:comp.c:MotorR \\ c:value &\in owl:DatatypeProperty \end{aligned}$$

Subset S_A of the Application Ontology of a Specific System A. In this subset only the observations related to the Revised Trauma Score are considered.

$$\begin{aligned} a:RTS &\equiv \exists a:hasEyeResp . a:EyeResp \\ &\quad \sqcap \exists a:hasMotorResp . a:MotorResp \\ &\quad \sqcap \exists a:hasVerbalResp . a:VerbalResp \\ &\quad \sqcap \exists a:hasSysBP . a:SysBP \sqcap \exists a:hasRespRate . a:RespRate \\ a:hasValue &\in owl:DatatypeProperty \end{aligned}$$

Notice the difference in the representation of the $a:RTS$ class with regard to the $c:RTS$ in the Canonical ontology. While in the latter the class $c:GCS$ is used in the definition, in the former the five values that ultimately are necessary to calculate the RTS score are indicated directly.

Subset S_B of the Ontology of a Specific System B. In this subset only the observations related to the Glasgow Coma Scale are considered.

$$\begin{aligned} b:GCS &\equiv \sqcap \exists b:hasEyeResponse . b:EyeResponse \\ &\quad \sqcap \exists b:hasVerbalResponse . b:VerbalResponse \\ &\quad \sqcap \exists c:hasMotorResponse . c:MotorResponse \\ b:hasValue &\in owl:DatatypeProperty \end{aligned}$$

Integration Mappings.⁶ Finally, let us imagine that the following integration mappings have been established by the MappingModule:

$$\begin{aligned} \mathcal{I}_{AC} &= \langle S_A, S_C, \\ &\quad \{a:RTS \equiv c:RTS, a:EyeResp \equiv c:EyeR, a:RespRate \equiv c:RespRate, a:hasEyeResp \sqsubseteq c:comp, \\ &\quad a:hasSysBP \sqsubseteq c:comp, a:hasRespRate \sqsubseteq c:comp, a:hasValue \equiv c:value\} \rangle \\ \mathcal{I}_{BC} &= \langle S_B, S_C, \\ &\quad \{b:GCS \equiv c:GCS, b:EyeResponse \equiv c:EyeR, b:hasEyeResponse \sqsubseteq c:comp, \\ &\quad b:hasValue \equiv c:value\} \rangle \end{aligned}$$

3.2 Path Mappings

As stated in the previous section, the MappingModule is in charge of creating the integration mapping between the Canonical Layer and the Applications Layer. Thanks to this integration mapping instances that initially belong to the Application Layer can be recognized as instances of the Canonical Layer (and viceversa). For example, if the aforementioned integration mapping \mathcal{I}_{BC} is considered, given the triples

$$\begin{aligned} (b:indGCS \text{ rdf:type } b:GCS) & \quad (b:indGCS \text{ b:hasEyeResponse } b:indER) \\ (b:indER \text{ rdf:type } b:EyeResponse) & \quad (b:indER \text{ b:hasValue } 4) \end{aligned}$$

the reasoner will infer the following statements, which classify all the information about the individuals $b:indGCS$ and $b:indER$ in the Canonical layer:

⁶ For the sake of visual clarity, in this integration mapping we indicate only the axioms related to the eye response component of the Glasgow Coma Scale. Please assume that the other two components are treated accordingly.

```
(b:indGCS rdf:type c:GCS)          (b:indGCS c:comp b:indER)
(b:indER  rdf:type c:EyeR)         (b:indER  c:value 4)
```

This is a quite straightforward process since the representation of the concept GCS is similar in both the Canonical ontology and the Application ontology of system B (i.e. in both cases the class GCS is directly related to each of its three components via an object property). The problem arises when the representation of the same concept in the source and target ontology is more heterogeneous than just different names for classes or properties. Let us compare the definitions of classes $a:RTS$ and $c:RTS$ in section 3.1. Looking at the description of $a:RTS$, it can be seen (Fig.2) that any individual belonging to that class will be directly related to an individual of the class $a:EyeResp$ via the role $a:hasEyeResp$ (assume the same intuition for the case of the motor and verbal responses). However, in the case of the descriptions in the Canonical ontology, it turns out that classes $c:RTS$ and $c:EyeR$ are not directly related, but indirectly: first $c:RTS$ is related to the class $c:GCS$ via the role $c:comp$ and then the class $c:GCS$ is related to the class $c:EyeR$ again via the role $c:comp$. Then it could be stated that there is a simple path between classes $a:RTS$ and $a:EyeResp$ (Fig.2a) and a composite path between classes $c:RTS$ and $c:EyeR$ (Fig.2b).

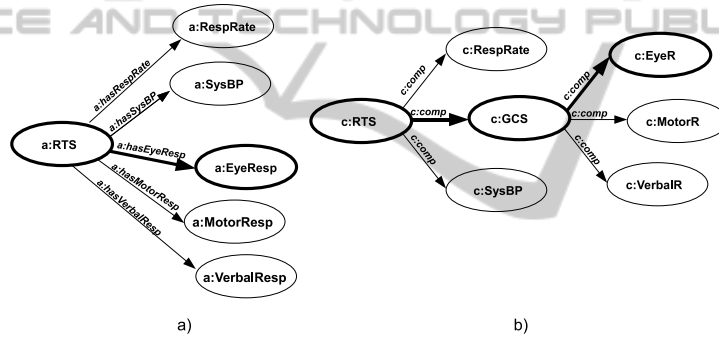


Fig. 2. Structurally different but semantically equivalent ontology paths.

Intuitively, those two paths could be regarded as equivalent, since their only difference is from the structural point of view caused by the heterogeneous origin of the ontologies, not from a semantic point of view. Let us denote those equivalences with the following statements where the expressions on both sides of the \equiv_p symbol represent a path. Each expression begins with a class name that is followed by (one or more) pairs *propertyName[className]*:

$$\begin{aligned}
 a:RTS.a:hasEyeResp[a:EyeResp] &\equiv_p c:RTS.c:comp[c:GCS].c:comp[c:EyeR] \\
 a:RTS.a:hasMotorResp[a:MotorResp] &\equiv_p c:RTS.c:comp[c:GCS].c:comp[c:MotorR] \\
 a:RTS.a:hasVerbalResp[a:VerbalResp] &\equiv_p c:RTS.c:comp[c:GCS].c:comp[c:VerbalR]
 \end{aligned}$$

For that reason, we have decided to incorporate a new kind of mappings to our framework: the so called *path mappings*, which establish equivalence or subsumption relations between two ontology paths. Path mappings are useful at the time of transforming individuals from one ontology so that they meet the requirements of the target

ontology. The implementation of path mappings is done by using SWRL rules. For example, the path mappings shown before would be implemented using the following rule:

```
a:RTS(?r) ∧ c:RTS(?r) ∧ a:hasEyeResp(?r,?e) ∧ a:hasMotorResp(?r,?m)
  ∧ a:hasVerbalResp(?r,?v) ∧ swrlx:createOWLThing(?g,?r)
  → c:comp(?r,?g) ∧ c:GCS(?g) ∧ c:comp(?g,?e) ∧ c:comp(?g,?m) ∧ c:comp(?g,?v)
```

Let us look at what happens when system A wants to send the following triples about a RTS reading to another system⁷:

```
(a:indRTS rdf:type a:RTS)           (a:indGCS a:hasEyeResp a:indER)
(a:indER rdf:type a:EyeResp)       (a:indER a:hasValue 4)
```

Following our proposal, thanks to the integration mapping \mathcal{I}_{AC} in the first place the individuals will be classified in the Canonical ontology. For example, individuals $a:indRTS$ and $a:indER$ will be recognized as instances of the classes $c:RTS$ and $c:EyeR$ respectively. In order to comply with the specification of the class $c:RTS$, there should be an individual of the class $c:GCS$ that acts as a connection between individuals $a:indRTS$ and $a:indER$. That individual is created by the SWRL rule above, which fires as soon as $a:indRTS$ is recognized as instance of the class $c:RTS$ (because the rest of the clauses in the body of the rule are also fulfilled).

3.3 Knowledge Rules

Up to now, we have presented a solution that allows the interoperability of data about medical observations between Health Information Systems. However we think that once this framework is set up, its potential could be enhanced to solve a more ambitious problem: the possibility of defining and sharing medical knowledge among those systems. It is widely known that EHRs hold great potential for clinical decision support, for example by translating practice guidelines into automated reminders and actionable recommendation[17] which can improve the quality and safety of healthcare as substantial evidence suggests[18]. Usually, medical experts are in charge of performing those translation tasks and of incorporating them into their systems, without sharing them outside their local context. However, it would be interesting to have the option to spread that knowledge from one ontology to another. For example, widely accepted knowledge directives could be integrated into the Canonical ontology, and due to the mappings between the Canonical ontology and the Application ontologies, spread to the diverse Application ontologies. This could incorporate valuable knowledge into the systems without too much effort on their part. Accordingly, a specific system could define knowledge directives in its Application ontology and spread them to the Canonical ontology and in consequence to other Application ontologies, although in this case some supervision should be carried out to avoid that one system infects other systems with knowledge that is relevant locally but not necessarily relevant for other systems.

An appropriate way of modelling knowledge directives related to diagnoses and treatment of illnesses is by using rules expressed in a declarative form, since they are

⁷ Accordingly for the remaining components of $a:RTS$.

suitable for obtaining conclusions from a set of data. More precisely, we have chosen again SWRL as language for representing these rules. For example, one of the rules that could be defined in the Canonical ontology is shown next. This rule is intended to calculate the Glasgow Coma Scale value of a patient as the sum of the values of each of the three components (Eye, motor and verbal response). The result is stored as the value of the `c:value` property of the corresponding `c:GCS` reading.

```
c:GCS(?g) ∧ c:comp(?g,?e) ∧ c:EyeR(?e) ∧ c:value(?e,?ev) ∧ c:comp(?g,?m)
  ∧ c:MotorR(?m) ∧ c:value(?m,?mv) ∧ c:comp(?g,?v) ∧ c:VerbalR(?v) ∧ c:value(?v,?vv)
  ∧ swrlb:add(?emv,?ev,?mv) ∧ swrlb:add(?emvv,?emv,?vv) → c:value(?g,?emvv)
```

4 Conclusions

In the first part of this paper we have presented a framework which has the following main features: First, it favors the notion of semantic interoperability among health information systems by using formal ontologies as canonical conceptual models, which allow to focus on semantic aspects that are independent of the languages or technologies used to describe EHRs. This reasoning-driven approach avoids the need of peer-to-peer transformations and in addition, the features of any specific system remain unknown to the other systems in the framework. Moreover, it favors the notion of extensibility to different models, since any medical institution can create its own Application ontology and relate it to the terms of the Canonical layer via an integration mapping. Finally it facilitates this seamless adaptation by providing of one module that facilitates the task of obtaining the definitions of the Application ontology from a particular underlying system and another module that facilitates the task of linking definitions of the Application ontology to definitions of the Canonical ontology. In the rest of the paper some features that complement the proposal have been discussed: the usefulness of path mappings in the transformation of ontology individuals and the convenience of using rules for knowledge representation and sharing.

Acknowledgements

This work is supported by the Spanish Ministry of Education and Science (TIN2007-68091-C02-01) and the Basque Government (IT-427-07). The work of Idoia Berges is supported by the Basque Government (Programa de Formación de Investigadores del Departamento de Educación, Universidades e Investigación).

References

1. Kalra, D., Lewalle, P., Rector, A., Rodrigues, J.M., Stroetmann, K.A., Surjan, G., Ustun, B., Virtanen, M., Zanstra, P.E.: Semantic Interoperability for Better Health and Safer Healthcare. Technical report, European Commission (2009)
2. OWL2 Web Ontology Language (2009) <http://www.w3.org/TR/owl2-overview/>.
3. Hefflin, J., Hendler, J.: Semantic Interoperability on the Web. In: Proceedings of Extreme Markup Languages 2000, Graphic Communications Association (2000) 111–120

4. Kilic, O., Dogac, A.: Achieving Clinical Statement Interoperability using R-MIM and Archetype-based Semantic Transformations. *IEEE Transactions on Information Technology in Biomedicine*, 13 (2009) 467–477
5. Bicer, V., Kilic, O., Dogac, A., Laleci, G.B.: Archetype-Based Semantic Interoperability of Web Service Messages in the Health Care Domain. *Int'l Journal on Semantic Web & Information Systems* 1 (2005) 1–22
6. Martínez-Costa, C., Menárguez-Tortosa, M., Valencia-García, R., Maldonado, J., Fernández-Breis, J.T.: A Model-driven Approach for Representing Clinical Archetypes for Semantic Web Environments. *Journal of Biomedical Informatics* 42 (2009) 150–164
7. Swrl (2009) Available at <http://www.w3.org/Submission/SWRL/>.
8. Kashyap, V., Sheth, A.P.: Semantic and schematic similarities between database objects: A context based approach. *The Very Large Databases Journal* 5 (1996) 276–304
9. SNOMED (2009) Available at <http://www.ihtsdo.org/snomed-ct/>.
10. Rosse, C., Mejino, J.L.V.: A Reference Ontology for Biomedical Informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics* 36 (2003) 478–500
11. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: *The Description Logic Handbook. Theory, Implementation and Applications*. Cambridge University Press (2003)
12. Blanco, J.M., Illarramendi, A., Goñi, A.: Building a Federated Relational Database System: An Approach Using a Knowledge-Based System. *International Journal on Cooperative Information Systems* 3 (1994) 415–456
13. The openEHR Foundation: *Archetype Definition Language* (2007)
14. Mena, E., Illarramendi, A., Kashyap, V., Sheth, A.: OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies. *International journal on Distributed And Parallel Databases* 8 (2000) 223–272
15. Champion, H.R.: A Revision of the Trauma Score. *The Journal of Trauma* 29 (1989) 623–629
16. Teasdale, G., Jennett, B.: Assessment of Coma and Impaired Consciousness. a Practical Scale. *Lancet* 2 (1974) 81–84
17. Hoffman, L.: Implementing Electronic Medical Records. *Communications of the ACM* 52 (2009) 18–20
18. Kawamoto, K., Houlihan, C.A., Balas, E.A., Lobach, D.F.: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 330 (2005) 765–768