Hygehos Ontology for Electronic Health Records

Naiara Muro[1,2], Eider Sanchez[1,3], Manuel Graña[2], Eduardo Carrasco[1,3], Fran Manzano[4], Jose María Susperregi[5], Agustin Agirre[5] and Jesús Gómez[5]

[1] Vicomtech-IK4, San Sebastian, Spain
(nmuro@vicomtech.org)
[2] Computational Intelligence Group, University of The Basque Country UPV/EHU, San Sebastian, Spain
[3] Biodonostia Health Research Institute, San Sebastian, Spain
[4] Igarle, San Sebastian, Spain
[5] La Asunción Clinic, Tolosa, Spain

Abstract. During the last years a high effort on standardization of Electronic Health Records has been made. Standards ISO EN 13606 and OpenEHR with their dual approach have promoted semantic interoperability in the real clinical practice. Recently, the focus has been set on the extraction of knowledge from the clinical information stored in EHR, but current approaches based on archetypes do not provide a complete solution regarding content structuration limitation. In this paper we propose an ontology for Hygehos Electronic Health Records (EHR), that we call the Hygehos Ontology. The introduction of such ontology on the EHR system will facilitate the development of reasoning and knowledge extraction tools over the stored clinical information. In our approach, we first align the Hygehos EHR to the dual model of OpenEHR and generate the corresponding archetypes for every part of the system. Secondly, we formalize a methodology for structuring the clinical contents of Hygehos EHR into the Hygehos Ontology.

Keywords: EHR, ADL archetype, OWL ontology, Hygehos

1 Introduction

Interoperability and clinical data processing play an important role in biomedical research through a variety of applications linked to hospitals' databases. Different approaches have been applied in the search of interoperability between heterogeneous information systems [1].

Standardized Electronic Health Records have allowed the communication and interpretation of this data within a wide variety of medical centers [2]. Its dual structure composed of a reference information model and an archetype model has allowed separation between clinical content and its structuration, in order to promote this interoperability [3]. However, the archetypes responsible for the definition of clinical concepts are not able to support complex reasoning and knowledge discovery requirements [4]. A more complete structuration of the information is needed to overcome this lack in order to be able to process semantic queries.

Ontologies are used primarily as a source of vocabulary standardization and integration, and many applications can be benefited of using them for computable knowledge extraction. Transforming the current archetypes into OWL (Web Ontology Language) classes and extending them into an ontology would overcome the current lack of ADL (Archetype Definition Language) archetypes when introducing the obtained reasoning conclusions from the content in their own definition [5, 6].

In this paper we propose an ontology for Electronic Health Records (EHR), that we call the Hygehos Ontology. The introduction of such ontology on the EHR system will allow the application of reasoning and knowledge extraction tools over the stored clinical information. We based our approach on the Hygehos EHR [7]. We first align the Hygehos EHR with the dual model of the OpenEHR standard and generated the corresponding archetypes for every part of the system. Then, we formalize the methodology that we have followed for structuring the clinical contents of the Hygehos EHR into the Hygehos Ontology.

This paper is structured as follows: in Section 2 we present a brief state of the art of relevant concepts and approaches; in Section 3 we describe the methodology followed by the Hygehos system; in Section 4 such methodology is formalized into the Hygehos Ontology; in Section 5 we present the mapping between the Hygehos Ontology with the dual modelling of Hygehos to OpenEHR, and finally in Section 6 we summarize conclusions and future work.

2 Background concepts

This section describes some relevant technologies of our approach, such as ontologies and EHR. Also, we briefly introduce the Hygehos EHR, over which our work has been developed.

2.1 Ontologies

Following Gruber's definition [8] ontologies are formal specifications of a conceptualization. Ontologies allow the descriptive representation of rich and complex knowledge about concepts and their relations. They determine semantic identifiers and formal descriptions representing the classes of entities of a specific domain [9]. Ontologies have been applied in the literature to the clinical domain in different approaches and applications, such as (i) the agreement about the knowledge model of a domain [10], (ii) clinical decision support systems [11] and recently also for EHR enhancement [12, 13]. In particular the work of Beale and Heard [12] describe an interoperable and safely computable clinical information model based on an ontological analysis of the process of clinical care delivery, seen as a scientific problem-solving process. The works of [13] address the semantic interoperability of two EHR standards (OpenEHR and ISO EN 13606) by a solution capable of transforming

OpenEHR archetypes into ISO EN 13606 and vice versa that combines Semantic Web and Model-driven Engineering technologies.

2.2 EHR

EHR are defined as structured clinical data repositories. Provided that certain minimum requirements are met regarding ubiquity, they allow accessibility from anywhere at any time [14]. Their main features are: (i) to ensure readable, interpretable and persistent information record, (ii) to provide a unequivocal and unique identification for each patient, (iii) to allow interoperability between different healthcare centers as well as between different departments within the same center, (iv) to apply a standardization of the recorded data in the EHR to make interoperability possible, and (v) to facilitate the use and access to all patients' EHR and aid with the visualization and processing of the data. All this must be applied taking into account the security and privacy policy imposed by law to clinical data, and ensuring the authenticity of all documents filed in the EHR, through a signature of the responsible [15]. The multiple features described before have made its implementation in medical centers very favorable and widespread.

Over the last decade different EHR standards had been developed for EHR. The most extended ones are (i) HL7, (ii) ISO EN13606 and (iii) OpenEHR. Although each standard brings a particular feature compared to others, all agree on a dual model structure, consisting on a reference model (RM) and an Archetype Model (AM). The RM supports information within a structure, based on well-established concepts independent from knowledge. It represents the characteristics of the general components and their organization. AM defines and models concepts of clinical knowledge following the structure and constraints imposed by the RM. The combination of both models in a single frame provides of structure and semantic interpretation to the content stored in the EHR.

Archetypes are plain text files written following a syntax called ADL. Their goal is to represent particular clinical concepts without taking into account their information representation structure, giving only their clinical definition. The dual model approach was implemented just for that reason: to separate the clinical concept from the information model that embodies it. In that way particular clinical concepts are represented as a set of constraints on the generic model of information [14].

As we reported in our previous work [16], some challenges concerning knowledge extraction processes from the EHR that follow from the dual model are still unsolved. For that reason, we proposed in [16] an extension of the dual model to a triple model approach. This model extension proffers a new structure, which instead of relying directly on the timescale defined by the facts (clinical documents), was actually based on a scale defined by time dependent Decisional Events (DE). In this work, we present an ontology that can serve as a mapper between the document-based approach to a DE-based approach, allowing application of reasoning and knowledge extraction tools into EHR.

3 Hygehos Methodology

Hygehos is a proprietary Electronic Health Record developed during the last 15 years by the Spanish IT companies Igarle and STT and the clinical team of the La Asunción clinic[1]. The system covers hospital-side information system, but also contains a module, called Hygehos Home, for remote monitoring of patients and patient-doctor telecommunication services [17].

Hygehos follows a generic approach for the acquisition of the clinical information of patients. This fact allows the direct implantation of such system into almost every medical center and hospital, independently to the specialization level or type. Currently it is running in more than 15 hospitals and medical centers in Spain, each of them oriented to a different specialization, for instance a monographic center in oncology and a primary care center. Such generic approach, the vision and the methodology for EHR followed by Hygehos are presented next.

Hygehos collects all the information concerning the status of the patient and classifies it into three types: (i) Permanent Information, (ii) Episodic Information and (iii) Evolution Information. The Permanent Information contains relevant patient information of general interest for any health professional in charge of that patient as clinical background, vaccinations, living will and social history, amongst others. It has to be always available for query submissions and it lasts over time, increasing its contents. It will be applied in any clinical procedure regardless of its origin source.

The Episodic Information is defined as information limited in time, i.e. isolated facts. The measurement unit of the episodic information is an episode. Within it, we can distinguish three levels of information: (i) Emergencies (e.g. home-based, hospital-based), (ii) hospitalization (e.g. acute, subacute, home hospitalization) and (iii) outpatient (e.g. consultations, check-ups). For each episode different medical acts are made: (i) preventives (e.g. vaccinations, check-ups), (ii) diagnosis (e.g. radiography, colonoscopy, cytology, spirometry, blood analysis), (iii) therapies (e.g. surgery, rehabilitation), (iv) research (e.g. clinical trials) and (v) Aesthetic (e.g. Botulinum Toxin), amongst others. Depending on the type of each clinical procedure, some clinical documents are associated to it. For some of these clinical documents, there are some limits regarding minimum contents definition by law.

Hygehos makes a full scan of the treated person, not only a punctual study, so each episode is not understand as an isolated event but as different phases of clinical processes suffered by the patient. We can understand episodes as isolated or related to a process. As an example, pregnancy could be considered as a clinical process and different episodes will be associated to it (e.g. a gestational diabetes caused by hormonal change in pregnancy). For other cases in which the episode is related to a specific patient's condition or a particular clinical situation, such as being chronic heart failure patient, the different episodes relate in a particular way to the clinical processes. The

[1] Collaboration between La Asunción Clinic (http://www.clinicadelaasuncion.com/), Igarle (http://www.igarle.es/), and STT (http://www.stt-systems.com/)

definition of these relationships allows the extraction of global conclusions for personalized patient care.

The Evolutive Information is generated with a temporal discontinuity. It is composed by documents with previous annotations that are permanent and unalterable but that allow the adhesion of new comments or notes resulting from care processes. Combined together, they form a unit in content terms (clinical course, evolutionary notes, active treatments, etc.).

4 Hygehos Ontology

We have formalized the Hygehos Methodology into an ontology, which we call the Hygehos Ontology (depicted in Fig. 1). It represents the domain model of an Electronic Health Record, all the clinical information contained, as well as the rest of the elements that interact with it.

The main class **ElectronicHealthRecord** is related to class **ClinicalInformation** by the object type property *ehrContains*. **ClinicalInformation** represents every clinical data, which can be of different types, contained in the EHR. A clinical information will always correspond to a patient treated by clinical center's health personal. Class **Patient** is related to **ClinicalInformation** by the object type property *correspondsTo*. Class **Patient** has a data type property *patientId*. Class **HealthPersonal** is related to **ClinicalInformation** by the object type property *hasResponsible*. Class **HealthPersonal** has a data type property *healthPersonalId*. Each health personal is part of a medical unit of the hospital or center, according to their specialization. Class **ClinicalInformation** is related to class **MedicalUnit** by the object type property *isFrom*. Class **MedicalUnit** has a data type property *unitName*.

Class **ClinicalInformation** has three different disjoint subclasses: (i) **PermanentInformation**, (ii) **EvolutiveInformation** and (iii) **EpisodicInformation**. In turn, class **PermanentInformation** has six disjoint subclasses: **Allergies**, **Vaccination**, **WorkHistory**, **PersonalHistory**, **FamilyHistory** and **AdministrativeInformation**. Each of these classes contains different data type properties that will describe the different data stored for each case (e.g. class **Allergies** contains a data type property *allergyName*). In order to keep the information clear, Fig. 1 does not show such properties, and represents the set of data type properties corresponding to each of the last-level-subclasses as a unique empty block.

The set of data type properties corresponding to each of the last-level-subclasses of the Hygehos Ontology are mapped with the contents of archetypes that we implemented for the Hygehos system, as presented in Section 5. The same occurs with the rest of the last-level-subclasses of the classes **EvolutiveInformation** and **EpisodicInformation**.

Class **EvolutiveInformation** has two disjoint subclasses (last-level, and thus each of them has associated a set of data type properties): **ClinicalEvolution** and **ActiveMedicalOrders**, which includes the active treatments at the moment but also the ones applied before in order to evaluate the therapeutic indications taken into account.

These therapeutic indications covers widespread medical issues, not only pharmacological.

Class **EpisodicInformation** is related to a class **Process** by the object type property *containsProcess*. Class **Process** is related to class **Episode** by the object type property *containsEpisode*. Class **Episode** is related to class **Procedure** by the object type property *containsProcedure*. Class **Procedure** is related to class **ClinicalDocument** by the object type property *containsClinicalDocument*. **ClinicalDocument** has seventeen disjoint subclasses (last-level, and thus each of them has associated a set of data type properties): **ClinicalStatisticalReport**, **EntryAuthorization**, **InformedConsent**, **AnamnesisAndPhysicalExploration**, **Evolution**, **MedicalOrder**, **ComplementaryExplorationReport**, **ReferralReport**, **SurgeryReport**, **Anesthesia**, **BirthReport**, **NursingCarePlan**, **PharmacologicalTreatment**, **VitalSignGraph**, **DischargeClinicalReport**, **NecropsyReport** and **Urgency**. This object will have as many subclasses as possible care events may occur in the present or in the future.

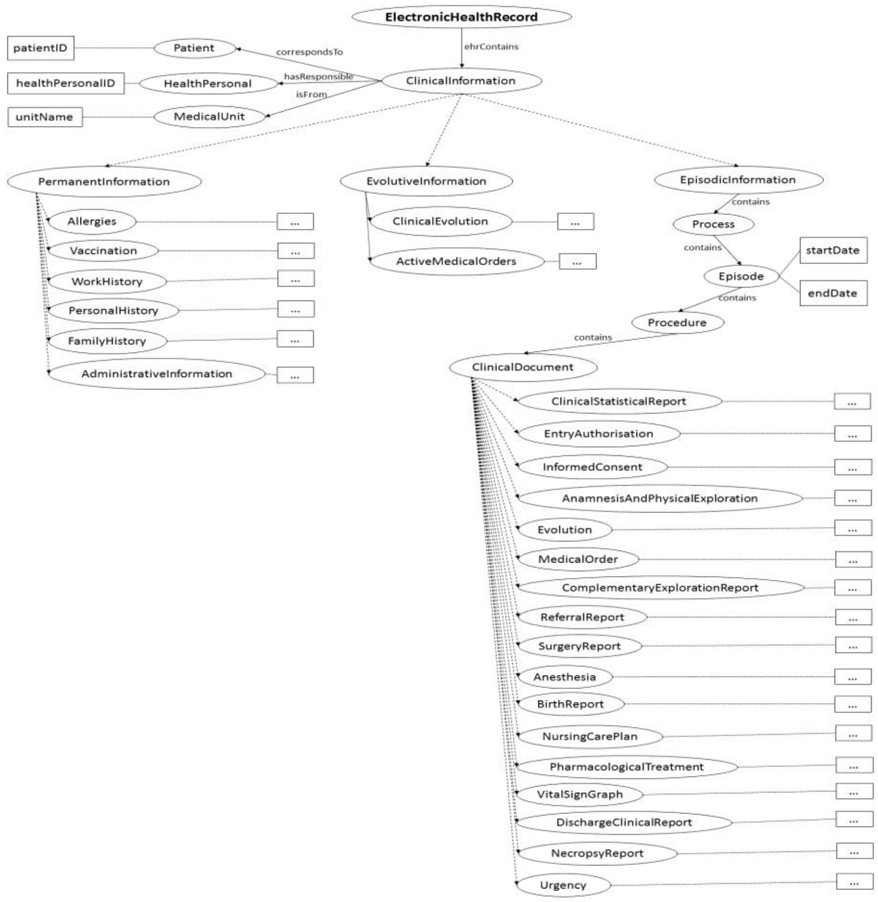

Fig. 1 Hygehos Ontology

5 Knowledge exploitation from Hygehos

The overall objective of this work is to include reasoning tools in the Hygehos System to extract more advanced and complex conclusions from the stored information into the clinical history. In order to put to work the generated Hygehos Ontology, first we have standardized the clinical history following the dual model of the openEHR standard. We created the different archetypes of the system, 25 in total. Each archetype is mapped to a class of the ontology, and each element of the archetype to a data type property (whose domain class is the one with which the archetype is mapped).

To illustrate the approach we present the archetype for the clinical document of anamnesis and physical exploration in Fig. 2. The elements defined are (i) registration date (data type DV_DATE_TIME); (ii) registration responsible professional (data type DV_TEXT) and corresponding medical unit (data type DV_TEXT); (iii) matter of the urgency (data type DV_TEXT); (iv) patient's background or medical history, with its type (data type DV_TEXT), date (data type DV_DATE_TIME) and description (data type DV_TEXT); (v) current illness for which medical care is requested (data type DV_TEXT); (vi) patient's biometric data, with the weight (data type CODED_TEXT), size (data type CODED_TEXT), body mass index (data type CODED_TEXT), and body surface (data type CODED_TEXT); (vii) description of the exploration (data type DV_TEXT); (viii) diagnosis (data type DV_TEXT), and (ix) the final recommendations (data type DV_TEXT). In Fig. 3 the graphical user interface corresponding to the archetype of Fig. 2 is shown. Fig. 4 depicts the part of the archetype where the link of each element with the corresponding class or properties of the Hygehos Ontology is shown.

```
definition
        OBSERVATION[at0001] matches {-- Anamnesis and Exploration Document
                ELEMENT[at0002] occurrences matches {0..1} matches { -- Date/Time
                        value matches {DV_DATE_TIME matches {value matches {yyyy-mm-dd}}}}
                CLUSTER[at0003] occurrences matches {0..*} matches { -- Responsible
                        items cardinality matches {0..*; unordered} matches {
                                ELEMENT[at0003.1] occurrences matches {0..1} matches {       -- Section
                                        value matches {DV_TEXT matches {*}}}
                                ELEMENT[at0003.2] occurrences matches {0..1} matches {       -- Practitioner
                                        value matches {DV_TEXT matches {*}}}}}
                ELEMENT[at0004] occurrences matches {0..1} matches { -- Matter Of Urgency
                        value matches {DV_TEXT matches {*}}}
                CLUSTER[at0005] occurrences matches {0..*} matches { -- Background/ Medical History
                        items cardinality matches {0..*; unordered} matches {
                                ELEMENT[at0005.1] occurrences matches {0..1} matches {       -- Type
                                        value matches {DV_TEXT matches {*}}}
                                ELEMENT[at0005.2] occurrences matches {0..1} matches {       -- Date
                                        value matches {DV_DATE_TIME matches {value matches {yyyy-mm-dd}}}}
                                ELEMENT[at0005.3] occurrences matches {0..1} matches {       -- Description
                                        value matches {DV_TEXT matches {*}}}}}
                ELEMENT[at0006] occurrences matches {0..1} matches { -- Current illness
                        value matches {DV_TEXT matches {*}}}
                CLUSTER[at0007] occurrences matches {0..*} matches { --  patient biometrical data
                        items cardinality matches {0..*; unordered} matches {
                                ELEMENT[at0007.1] matches {-- weight
                                        name matches {CODED_TEXT matches {code matches {[ac0007.1]} -- weight}
                                        value matches {QUANTITY matches {value matches {0..800}units matches {"kg"}}}}
                                ELEMENT[at0007.2] matches {-- size
                                        name matches {CODED_TEXT matches {code matches {[ac0007.2]} -- size}
                                        value matches {QUANTITY matches {value matches {0..5}units matches {"m"}}}}
                                ELEMENT[at0007.3] matches {-- BMI: Body Mass Index
                                        name matches {CODED_TEXT matches {code matches {[ac0007.3]} -- BMI}
                                        value matches {QUANTITY matches {value matches {0..100}}}}
                                ELEMENT[at0007.4] matches {-- body surface
                                        name matches {CODED_TEXT matches {code matches {[ac0007.4]} -- body surface}
                                        value matches {QUANTITY matches {value matches {0..3}units matches {"m^2"}}}}}}
                ELEMENT[at0008] occurrences matches {0..1} matches { -- Exploration
                        value matches {DV_TEXT matches {*}}}
                ELEMENT[at0009] occurrences matches {0..1} matches { -- Diagnosis
                        value matches {DV_TEXT matches {*}}}
                ELEMENT[at0010] occurrences matches {0..1} matches { -- Recommendations
                        value matches {DV_TEXT matches {*}}}}
```

Fig. 2 Archetype for anamnesis and physical exploration

Having the archetype mapped to the ontology allows to reuse of the information contained to generate new conclusions and extract knowledge from the EHR. In our previous work [16] we proposed a new decisional model in which the information contained in the EHR would be represented based on each of the decisions made on a patient, rather than based on the documents and reports generated during each episode. Each decisional event in the new model describes the context in which such decision was made: (i) patient data, (ii) different criteria followed, such as clinical guidelines and protocols, (iii) the objective of the decision, such as the total recovery of the patient or the fastest partial recovery possible, (iv) the final decision value, for instance of the treatment prescribed to the patient, and (v) the result of the decision, in terms of success or failure achieved with respect to the objective. The analysis of such context allows the extraction of low-level conclusions about the patient, their symptoms, the treatments that could succeed and the ones that have had high failure rates in the past [18, 19].

In order to have patient information represented in such a decisional model, a mapping between the current approach (document-based) and the decisional model is needed. The Hygehos Ontology is the formalization of the current document-based model, which is a starting point for such mapping process. We are currently working on the mapping of the Hygehos Ontology into a decisional model.

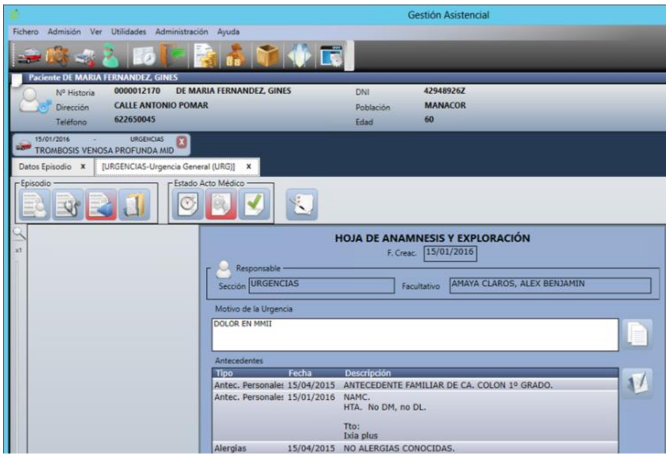

Fig. 3 User interface for anamnesis and physical exploration

6 Conclusions and Future Work

In this work we have built an ontology, called the Hygehos Ontology, which represents logically the generic methodology followed by the Hygehos EHR system. Our approach aims to provide knowledge extraction capabilities to EHR. The ontology we present in this paper formalizes the document-based approach followed by dual-model-EHR. In this sense, it can be used as an starting point to map such model into a decision based approach followed by the triple we proposed in our previous work [16] and for which we already proposed knowledge extraction and analysis tools [18, 19].

On the other side, we also mapped Hygehos to OpenEHR and provided the corresponding archetypes. In this way our work also allows to separate the graphical interfaces from the functionality and the contents, by allowing the generation of interfaces directly from archetypes. Such archetypes can be modeled directly by medical experts that do not have software technical expertise by using open archetype authoring tools provided by the OpenEHR community. In this way usability and user experience of Hygehos is improved, and personalized products can be provided to each customer (and their corporate image) with a small effort.

As future work, we will build the knowledge extraction tool based on the Hygehos Ontology. We will also build a tool that generates graphical interfaces directly from the definition of archetypes.

Acknowledgments. This research was partially funded by the Basque Business Development Agency (SPRI), dependent on the Basque Government, under the grant GAITEK2015-SemanHis.

References

1. Bellazi, R., Diomidous, M., Sarkar, I.N., Takabayashi, K., Ziegler, A., McCray, A.T.: Data Analysis and Data Mining: Current Issues in Biomedical Informatics. Methods Inf Med. 50:536-544 (2011)
2. Jha, AK., DesRoches CM., Campbell, EG., Donelan, K., Rao, S.R., Ferris, T.G, Shields, A., Rosenbaum, S., Blumenthal, D.: Use of electronic health records in U.S. hospitals. N Eng J Med. 360(16):1628-38 (2009)
3. Wollersheim, D., Anny, S. Wenny, R.: Archetype-based Electronic Health Records: A Literature Review and Evaluation of Their Applicability of Health Data Interoperability and Access. Health Information Management Journal. 38(2):7-17 (2009)
4. Iqbal, A.M., Shepherd, M., Abibi S.S.R.: An Ontology-Based Electronic Medical Record for Chronic Disease Management. 44th Hawaii International Conference on System Science (HICSS). (2011)
5. Pathak, J., Kiefer, R.C., Chute, C.G.: Using semantic web technologies for cohort identification from electronic health records for clinical research. AMIA JT Summits Transl Sci Proc. 10-19 (2012)
6. Lezcano, L., Sicilia, M.A., Rodríguez-Solano, C.: Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. Journal of Biomedical Informatics.44(2):343-353 (2010)

7. Carrasco, E., Sanchez, E., Artetxe, A., Toro, C., Graña, M., Guijarro, F., Susperregui, J.M., Aguirre, A.: Hygehos Home: an innovative remote follow-up system for chronic patients", in Studies in Health Technology and Informatics: Innovation in Medicine and Healthcare 2014, vol. 207, pp. 261-270, IOS Press, (2014)
8. Gruber, T.: "What is an Ontology." WWW Site http://www-ksl. stanford. edu/kst/whatis-an-ontology. html (accessed on 07-09-2004) (1993)
9. Martinez-Costa, C., Menarguez-Tortosa, M., Fernandez-Breis, J.T., Maldonado, J.A.: A model-driven approach for representing clinical archetypes for semantic web enviroments. Journal of Biomedical Informatics. 42(1): 150-164 (2009)
10. Gruber, T. R.: Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies, 43(5), 907-928 (1995)
11. Farion, K., Michalowski, W., Wilk, S., O'Sullivan, D. M., Rubin, S., & Weiss, D.: Clinical decision support system for point of care use: ontology driven design and software implementation. Methods of information in medicine, 48(4), 381-390 (2009)
12. Beale, T., Heard, S.: An ontology-based model of clinical information. In Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems (p. 760). IOS Press (2007)
13. Martínez-Costa, C., Menárguez-Tortosa, M., & Fernández-Breis, J. T.: An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. Journal of biomedical informatics, 43(5), 736-746 (2010)Gruber, T. R.: Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies, 43(5), 907-928 (1995)
14. Garde, S., Knaup, P., Hovenga, E.J.S., Heard, S.: Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes. Methods of Information in Medicine 46(3): 332–343 (2007)
15. Hayrinen, K., Saranto, K., Nykanen, P.: Definition, structure, content, use and impacts of electronic health records: A review of the research literature. International Journal of Medical Informatics. 77(5): 291-304. (2007)
16. Sanchez, E., Toro, C., Graña, M.: Integrating Electronic Health Records in Clinical Decision Support Systems, Innovation in Medicine and Healthcare, p.407--416 (2016)
17. Blobel, B., Kalra, D., Koehn, M., Lunn, M., Pharow, P., Ruotsalainen, P., Schulz, S., Smith, B.: The Role of Ontologies for Sustainable, Semantically Interoperable and Trustworthy EHR Solutions. European Federation for Medical Informatics (2009)
18. Sanchez, E., Toro, C., Artetxe A., Graña, M., Sanin, C., Szczerbicki, E., Carrasco, E., Guijarro, F.: Bridging challenges of clinical decision support systems with a semantic approach. A case study on breast cancer., Pattern Recognition Letters 34(14), 1758 – 1768 (2013).
19. Sanchez, E., Peng, W., Toro, C., Sanin, C., Graña, M., Szczerbicki, E., Carrasco, E., Guijarro, F., Brualla, L.: Decisional DNA for modeling and reuse of experiential clinical assessments in breast cancer diagnosis and treatment, Neurocomputing 146, 308 – 318 (2014).