Journal of Biomedical Informatics 46 (2013) 304–317

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

OWL-based reasoning methods for validating archetypes

Marcos Menárguez-Tortosa, Jesualdo Tomás Fernández-Breis *

Departamento de Informática y Sistemas, Facultad de Informática, Universidad de Murcia, CP 30100 Murcia, Spain

ARTICLE INFO

Article history:
Received 4 September 2012
Accepted 24 November 2012
Available online 14 December 2012

Keywords:
Electronic healthcare records
Archetypes
Ontologies
OWL

ABSTRACT

Some modern Electronic Healthcare Record (EHR) architectures and standards are based on the dual model-based architecture, which defines two conceptual levels: reference model and archetype model. Such architectures represent EHR domain knowledge by means of archetypes, which are considered by many researchers to play a fundamental role for the achievement of semantic interoperability in healthcare. Consequently, formal methods for validating archetypes are necessary. In recent years, there has been an increasing interest in exploring how semantic web technologies in general, and ontologies in particular, can facilitate the representation and management of archetypes, including binding to terminologies, but no solution based on such technologies has been provided to date to validate archetypes. Our approach represents archetypes by means of OWL ontologies. This permits to combine the two levels of the dual model-based architecture in one modeling framework which can also integrate terminologies available in OWL format. The validation method consists of reasoning on those ontologies to find modeling errors in archetypes: incorrect restrictions over the reference model, non-conformant archetype specializations and inconsistent terminological bindings. The archetypes available in the repositories supported by the openEHR Foundation and the NHS Connecting for Health Program, which are the two largest publicly available ones, have been analyzed with our validation method. For such purpose, we have implemented a software tool called *Archeck*. Our results show that around 1/5 of archetype specializations contain modeling errors, the most common mistakes being related to coded terms and terminological bindings. The analysis of each repository reveals that different patterns of errors are found in both repositories. This result reinforces the need for making serious efforts in improving archetype design processes.

1. Introduction

The lifelong health history of a patient recorded in electronic format represents her Electronic Health Record (EHR). Several technological approaches have been proposed for representing and exchanging EHRs and for capturing EHR data in the last decades. Many of these technologies are based in the dual-model architecture [1]. OpenEHR [2] and ISO EN 13606 [3] follow this architecture which defines two conceptual levels: (1) reference model; and (2) archetype model. The reference model defines the set of entities that form the generic building blocks of the EHR such as organizational data structures (e.g., folder, section) or datatypes (e.g., text, quantity). It contains the non-volatile features of the EHR, so clinical information is defined at this level. On the other hand, archetypes define how to represent clinical concepts in the form of structured and constrained combinations of the entities contained in the reference model, so knowledge in the EHR domain is defined at this level. They refer to clinical concepts, such as blood pressure or exam of the chest, and represent EHR knowledge in the healthcare domain. HL7 CDA [4] is also considered a dual-model architecture that specifies the structure and semantics of clinical documents by means of the so-called templates. CEM (Clinical Element Models) [5] pursues the definition of detailed clinical models that retain "computable meaning when data is exchanged between heterogeneous computer systems" in a dual-model approach.

In the last decades, many efforts have been put on the achievement of semantic interoperability in healthcare to promote patient safety and to increase the quality of care [6–8]. The final report of the SemanticHEALTH project [6] and many researchers consider that archetypes should play a fundamental role in the achievement of semantic interoperability in healthcare and that archetypes should be the clinical knowledge unit exchanged by clinical systems in order to process the clinical data of the patients. The interest in archetypes was reinforced by the decision of CIMI [9] of using them, so researching and improving this technology becomes fundamental.

In order for archetypes to be successful and useful, its quality has to be assured. Archetypes are usually expressed using the Archetype Definition Language (ADL)[10]. This is a generic, syntactic language which can be used for writing archetypes for different

* Corresponding author. Fax: +34 868884151.
E-mail addresses: marcos@um.es (M. Menárguez-Tortosa), jfernand@um.es (J.T. Fernández-Breis).

1532-0464/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2012.11.009

EHR standards. The requirement of formal methods for validating the design and content of archetypes has been identified in [11,12]. ADL parsers usually check the compliance of the archetype with respect to the ADL grammar. Archetype editing tools would be then expected to provide methods for guaranteeing the consistency of the archetypes. So far, very few archetype authoring tools implement such techniques. The most significant case is the LinkEHR editor [13], which defines and implement an algebraic formal framework for archetype validation. This required a great development effort since methods or tools from the ADL community were not available for reuse. Such effort also lacked of a representation of archetypes that could help to detect, for instance, inconsistencies in the terminological bindings associated with archetype terms. Addressing such issue in ADL settings would require another specific development.

Given that archetypes have been considered a way of modeling clinical knowledge in the EHR domain, there has been an increasing interest in recent years in exploring how semantic web technologies in general, and ontologies in particular, can facilitate the representation and management of archetypes. This idea is reinforced by the relevance of the links with terminologies in archetypes and the ongoing efforts for providing a better link between terminologies and ontologies [14]. Exploring such relations has been a major research goal of our research group in the last years. First, we addressed the representation of archetypes based on the Web Ontology Language (OWL) [15], aiming to improve the management of clinical archetypes. The EHR domain of dual-model architectures was represented by OWL ontologies and archetypes were transformed into individuals of such ontologies. In this way, semantic operations, such as classifications and searches, could be efficiently performed over archetypes. Such representation handled archetypes as data objects and demonstrated its usefulness for supporting interoperability processes [16,17]. However, the development of archetype validation methods for such representation required an effort comparable to non-semantic approaches. The representation of archetypes using OWL has been approached in other ways. In [18,19] archetypes are expressed as OWL classes rather than individuals. This representation allows for processing EHR extracts as individuals of such OWL classes, but they did not support validation operations.

In this work, we use OWL-DL, which is the OWL sublanguage with a correspondence with Description Logics (DL) [20]. DL is a family of knowledge representation languages that permits integrating archetypes, information models (the reference model) and terminologies in one modeling framework. This is especially relevant because information or reference models deal with EHR data structures and terminologies deal with models of meaning [21]. On the one hand, OMG has developed the proposal ODM (Ontology Definition Metamodel) [22] that supports the bridges between software engineering modeling languages, such as UML, and knowledge representation languages, such as OWL. Reference models are usually defined as UML models. On the other hand, terminologies are models of meaning in the medical field that can be represented as ontologies or, at least, using ontology languages. In particular, SNOMED-CT is available as a DL ontology in OWL.

Archetypes have not been the unique EHR technology which has taken advantage of OWL and ontologies. OWL has been used for different purposes with HL7 technologies. An ontology architecture for HL7 V3 is discussed in [23] and an OWL-DL ontology for HL7 RIM is presented in [24]. Besides, OWL representations of SNOMED-CT and CDA have been used to validate CDA documents [25]. In such effort, CDA documents were represented as OWL individuals. On the other hand, an OWL metamodel for representing Clinical Element Models has been proposed in [26]. This approach attempts to be able to represent clinical data as OWL individuals. Ontologies have also been proposed for the alignment of HL7 V2 and V3 applications [27].

Hence, we believe that an ontology-based representation of archetypes capable of supporting validation would certainly be very useful for several reasons, including quality assurance. First, ontology models would be used for a proper representation of clinical knowledge, and this would facilitate the development of efficient knowledge management methods. Second, the combination of advanced ontology models, such as OWL, with reasoning techniques would certainly reduce the effort required for implementing quality assurance and validation methods.

In this work we present an OWL-based process for validating the consistency of archetypes and for supporting the implementation of archetype quality metrics. Such process will address the following requirements: (1) the inconsistencies of archetypes with respect to the reference model and base archetypes will be identified; (2) the development effort will be reduced by using state of the art OWL technologies; and (3) the consistency of the terminological bindings in archetype specializations will be evaluated. This method will use an OWL representation for archetypes which will represent terminological bindings, and the consistency checking method will exploit the possibilities of OWL reasoning, so reducing the development effort.

Next, the structure of the paper is described. In Section 2, background information about archetypes and OWL is provided. The methods for representing archetypes and validating archetypes using OWL will be presented in Section 3. Section 4 will describe the software tool that implements the methods and the results of the analysis of two publicly available archetype repositories. Finally some discussion and conclusions are provided in Section 5.

2. Background

2.1. Archetypes

Archetypes are detailed, domain-specific definitions of clinical concepts in the form of structured and constrained combinations of the entities of a reference model [1]. They refer to clinical concepts, such as blood pressure or exam of the chest, and represent EHR knowledge in the healthcare domain. As mentioned, the ISO EN 13606 and openEHR communities specify them using ADL, which can express archetypes for any reference model in a standard way. The openEHR community is more active in terms of archetype authoring and the most important repositories of archetypes are for such standard: the Clinical Knowledge Manager (CKM) of the openEHR Foundation [28]; and the repository supported by the National Health Service Connecting for Health Program [29].

An archetype can include other archetypes and can be used in combination to form templates [30]. These usually correspond to screen forms, printed reports and, in general, complete application-level information to be captured or sent. Moreover, archetypes are envisaged as guides for clinicians. An archetype is organized in three sections, namely, header, definition and ontology. The *header* section contains metadata about the archetype such as authorship or description. The *definition* section includes concept descriptions by means of constraints on entities of the reference model in a tree-like structure. Finally, the *ontology* section has the terminological definitions of the concepts included in the *definition* section. The terminological definition can be textual or it can be bound to terminologies such as SNOMED-CT [31] or LOINC [32].

The archetype formalism offers two constructs for reusing definitions, namely, internal reference and archetype slot. The former allows for including an archetyped concept in different