MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-827

A Process for the Representation of openEHR ADL Archetypes in OWL Ontologies

Alex Mateus Porn[a], Leticia Mara Peres[a], Marcos Didonet Del Fabro[a]

[a] *Informatics department, Federal University of Paraná, Brazil*

Abstract

ADL is a formal language to express archetypes, independent of standards or domain. However, its specification is not precise enough in relation to the specialization and semantic of archetypes, presenting difficulties in implementation and a few available tools. Archetypes may be implemented using other languages such as XML or OWL, increasing integration with Semantic Web tools. Exchanging and transforming data can be better implemented with semantics oriented models, for example using OWL which is a language to define and instantiate Web ontologies defined by W3C. OWL permits defining significant, detailed, precise and consistent distinctions among classes, properties and relations by the user, ensuring the consistency of knowledge than using ADL techniques. This paper presents a process of an openEHR ADL archetypes representation in OWL ontologies. This process consists of ADL archetypes conversion in OWL ontologies and validation of OWL resultant ontologies using the mutation test.

Keywords:

Archetypes, ADL, Ontologies, OWL.

Introduction

An archetype is a formal expression of a distinct concept in the domain level, expressed as constraints on data whose instances are in accordance with the reference model. The development process of an archetype consists mainly of translating a clinical concept for Reference Model entities, defining structure and data representation in an Electronic Health Record (EHR) [28]. Information sources for its development are varied: expert knowledge, specialized publications, data entry screens and system reports, term lists, clinical system data models, messages and regulatory forms, and forms used in clinical appointments or medical records [7, 8].

Clinical archetypes developed according to the openEHR [16] model use the Archetype Definition Language (ADL) as a standard [26]. This language presents an abstract syntax that can be used to express archetypes based on any information model. However, it presents difficulties in implementation and a few available tools to manage contents [3].

Although ADL language has its own syntax, it can be described using other languages such as eXtensible Markup Language (XML) [27] and Web Ontology Language (OWL) [10], increasing the integration with other Semantic Web tools and decreasing interoperability problems between health information systems [3, 4]. Therefore, processes for exchanging and transformating data can be better implemented with semantics oriented models, such as ontologies using OWL [1].

Ontologies define terms to describe and represent a knowledge domain. They are descriptions of concepts and relationships that can be used by people or software agents to share information in this domain. So, they can be used as a unifying structure for semantic representation of information [17, 18].

Ontologies have been frequently used to represent biomedical knowledge in recent years [12, 13]. They provide strict structures for knowledge management tasks related to archetypes and EHR systems representation [3].

We define in this article a process for the representation of ADL archetypes in ontologies described in OWL (or OWL ontologies) which are validated using mutation test. This process has two steps: the conversion of ADL archetypes in OWL ontologies using the algorithm of Elkin et al. [14] and the validation of OWL ontologies using the mutation test of Porn et al. [19, 20].

Archetype Definition Language (ADL)

ADL is a formal language to express archetypes, where the structure is independent of standard or domain. In general, it is not a language for clinical domains. It can be used to define any type of archetype for different reference models therefore it is possible to represent the same syntactic structure among diverse models [1].

ADL archetype is basically composed of three sections [1, 9]: (1) the section *header,* which contains the name of the archetype, a unique code that identifies the clinical concept and information of specialization and the language in which it was written; (2) the *definition* section, which specifies the structure and constraints associated with the clinical concept, restricting the cardinality and the content of the instances of the information model; (3) and the *ontology* section, which represents terminological definition to each clinical concept. In this paper only the d*efinition* section of archetypes is used to the conversion process to OWL.

ADL has some difficulties in implementation. When performing semantic processing two elements are required: one ADL parser and one parser of the reference model to ensure clinical accuracy of content [1].

The ADL parser produces a set of objects without explicit semantic relations among them. The semantics are unknown to the parser and only the association among the elements of *Definition* and *Ontology* section can be defined [1]. Thus, ADL reasoning possibilities are currently very limited [3], as well as the availability of tools to manage ADL contents is reduced. Consequently, the formalization of exchanging data and data transformation processes among systems becomes more difficult than using semantics oriented models, such as OWL ontologies.

Ontology Web Language (OWL)

An alternative to problems related to ADL language is the archetypes representation through OWL ontologies. OWL allows definitions that are meaningful, detailed and accurate with consistent distinctions among classes, properties and relations defined by the user [1]. Thus, the construction of archetypes using OWL can ensure the consistency of knowledge in a more general way than when using ADL techniques due to the lack of tools for managing these models.

The OWL language is used to define and instantiate Web ontologies and is defined by the W3C (World Wide Web Consortium) [10]. OWL ontology specifies how to derive logical consequences through the formal semantic OWL, clarifying facts that are not presented in the ontology but are bounded by the semantics [11].

Ontologies impose strict structures to knowledge management related to archetypes and EHR systems [3], so that several methodologies and tools have been proposed to compare different ontologies, mixing and identifying inconsistencies. Thus, activities such as comparison, selection, classification and consistency checking can be performed in the OWL in an easier and more efficient way than in the ADL content. As a matter of fact, OWL is a knowledge representation language.

The use of ontologies to represent biomedical knowledge is not a new process, as ontologies have been used in biomedical domains in recent years in different purposes [12, 13]. Thus, OWL becomes appropriate to represent clinical archetypes and information about EHR, so that constraints on archetypes can be determined by OWL or by defining appropriate elements.

Materials and Methods

For the representation of ADL archetypes in OWL ontologies two steps were established:

1. Convert openEHR archetypes developed in ADL language to OWL ontologies, based on an algorithm for mapping ADL objects to OWL, adapted from a model proposed by Elkin et al. [14].
2. Validate OWL ontologies obtained after conversion. For this, OWL ontologies are tested using mutation tests [19, 20].

Figure 1 shows an example of *definition* section of an openEHR archetype implemented in ADL language, that will be converted to an OWL ontology.

```
definition
    EVALUATION[at0000] matches {     -- A health oriented check list
        data matches {
            ITEM_TREE[at0001] matches { -- Tree
                items cardinality matches
                {1..*; unordered} matches {
                    CLUSTER[at0004] occurrences matches
                    {1..*} matches {     -- Question group
                        items cardinality matches
                        {1..*; unordered} matches {
                            CLUSTER[at0002] occurrences matches
                            {1..*} matches {     -- Question
                                items cardinality matches
                                {1..2; unordered} matches {
                                    ELEMENT[at0003] matches {*}
                                    ELEMENT[at0005] occurrences matches
                                    {0..1} matches {     -- A comment on the answer
                                        value matches {
                                            DV_TEXT matches {*}
                                        }}}}}}
                    ELEMENT[at0006] occurrences matches
                    {0..1} matches {     -- Summary
                        value matches {
                            DV_TEXT matches {*}
                        }}}}}}
```

Figure 1 – Definition section of an ADL archetype

To represent archetypes in OWL, it is necessary to define in OWL the classes from the Reference Model used by the archetype. Thus, the following algorithm is used to the first step of the conversion process:

1. Reference Model classes used by the archetype are represented as OWL classes;
2. Each archetype node defined as a clinical concept is represented as a subclass of one of the classes of the Reference Model used;
3. Object properties that define associations among nodes and archetype components are defined as OWL Object Properties. The domain declaration of this properties is indicated for the archetype class itself, and the range declaration to the set of objects that should be associated to the domain of this property;
4. Properties that represent data structures in ADL are defined as OWL Data Type Properties, being the domain declaration indicated as the cardinality of objects associated to this property, and the range declaration defined as the data types of all objects found in the archetype with respect to this attribute;
5. Properties constraints and cardinalities defined in ADL archetype are converted in constraints and axioms of OWL language.

Basically, an archetype has a hierarchical structure and constraints started with a root class. Thus, for an ADL archetype that presents several archetype nodes of the Reference Model of the same type, as ELEMENTs, in OWL just one class is defined to represent this node as a root class of the hierarchy and the clinical concepts of the archetype are represented as subclasses of these classes of the Reference Model.

Based on the conversion algorithm, it is possible observe at Figure 2 the definition of the archetype nodes EVALUATION, ITEM_TREE, CLUSTER and ELEMENT of the Figure 1, represented as OWL classes.

```
<Declaration>
    <Class IRI="#CLUSTER"/>
</Declaration>
<Declaration>
    <Class IRI="#ELEMENT"/>
</Declaration>
<Declaration>
    <Class IRI="#EVALUATION"/>
</Declaration>
<Declaration>
    <Class IRI="#ITEM_TREE"/>
</Declaration>
```

Figure 2 – openEHR Reference Model classes in OWL

In the hierarchy of an ADL archetype, properties that define associations among archetype nodes, refer to Information Reference Model classes, therefore they associate individuals of classes from a lower level, to individuals of classes from a higher level, defining classes and subclasses from the archetype [44]. Figures 3 to 7 show excerpts of one OWL ontology representing each structure of ADL archetype of Figure 1. Figure 3 presents OWL Object Properties which represent the properties that define associations among nodes and components of the ADL archetype.

```
<Declaration>
    <ObjectProperty IRI="#data_matches"/>
</Declaration>
<Declaration>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
</Declaration>
        .
        .
        .
<Declaration>
    <ObjectProperty IRI="#value_matches"/>
</Declaration>
```

Figure 3 – ADL Operators as OWL Object Properties

In Figure 4 it is possible to observe the domain declarations defined to the OWL Object Properties presented in Figure 3.

```
<ObjectPropertyDomain>
    <ObjectProperty IRI="#data_matches"/>
    <Class IRI="#A_health_oriented_check_list"/>
</ObjectPropertyDomain>
<ObjectPropertyDomain>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
    <Class IRI="#Tree"/>
</ObjectPropertyDomain>
            .
            .
            .
<ObjectPropertyDomain>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#ELEMENT"/>
</ObjectPropertyDomain>
```

Figure 4 – Domain definitions of OWL Object Properties

Figure 5 presents range declarations defined for OWL Object Properties presented in Figure 3.

```
<ObjectPropertyRange>
    <ObjectProperty IRI="#data_matches"/>
    <Class IRI="#ITEM_TREE"/>
</ObjectPropertyRange>
<ObjectPropertyRange>
    <ObjectProperty IRI="#item_cardinality_matches_summary"/>
    <Class IRI="#CLUSTER"/>
</ObjectPropertyRange>
            .
            .
            .
<ObjectPropertyRange>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#Answer"/>
</ObjectPropertyRange>
<ObjectPropertyRange>
    <ObjectProperty IRI="#value_matches"/>
    <Class IRI="#Summary"/>
</ObjectPropertyRange>
```

Figure 5 – Range Definitions of OWL Object Properties

It is possible to observe in Figure 6 the definition in OWL of the properties that represent data structures in ADL from Figure 1, with domain and range declarations of these properties defined as shown for OWL Object Properties.

```
<Declaration>
    <DataProperty IRI="#value_matches_A_comment_on_the_answer"/>
</Declaration>
<Declaration>
    <DataProperty IRI="#value_matches_Summary"/>
</Declaration>
<Declaration>
    <DataProperty IRI="#value_matches_answer"/>
</Declaration>
```

Figure 6 – ADL Operators as OWL Data Type Properties

As ADL data properties are represented as OWL Data Type Properties according to the conversion algorithm, associations can be made through OWL constructors *hasValue*, *someValuesFrom* and *allValuesFrom*.

So, according to the last step of the conversion algorithm, ADL property constraints and cardinalities are defined as OWL constraints, as shown in Figure 7 the constraints concerning the ELEMENT and ITEM_TREE classes of the archetype from Figure 1.

```
<EquivalentClasses>
    <Class IRI="#ELEMENT"/>
    <ObjectIntersectionOf>
        <ObjectSomeValuesFrom>
            <ObjectProperty IRI="#items_cardinality_matches_question"/>
            <Class IRI="#Question"/>
        </ObjectSomeValuesFrom>
        <ObjectSomeValuesFrom>
            <ObjectProperty IRI="#items_cardinality_matches_tree"/>
            <Class IRI="#Tree"/>
        </ObjectSomeValuesFrom>
    </ObjectIntersectionOf>
</EquivalentClasses>
```

Figure 7 – ADL constraints as OWL constraints

The conversion process and the resulting representation of ADL archetypes to OWL were assisted by a tool for creating and editing of ontologies called Protégé [15]. The following steps were defined:

1. Manual selection of openEHR ADL archetypes in the online repository CKM [16]. Seven archetypes in the review state were selected. Three archetypes are OBSERVATION type, respectively the "Apgar score", "Autopsy examination" and "Fetal heart rate" archetypes. Two are EVALUATION type, the "Alert" and "A health oriented check list" archetypes. One is INSTRUCTION type and one is an ACTION type, respectively "Informed consent request" and "Follow up action" archetypes.
2. Manual conversion of ADL archetypes to OWL based on the conversion algorithm, being the process realized with the aid of Protégé tool;

To validate archetypes and reveal possible defects inserted in OWL ontologies (proposed in step 2 of the process), we applied the mutation test technique to OWL Ontologies [19, 20].

Results

OWL ontologies obtained after the conversion process are readable for humans and computers, facilitating the interpretation of clinical concepts as well as the development of queries and content management.

Test methods which validate OWL models help to ensure adequacy and quality in ontologies. Tests allow the representation of knowledge domain with security and accuracy. Written queries in SPARQL [25] or DL Query [21] can be used to validate the correction of syntactic and semantic structures in tests of instantiation, recovery, achievement, satisfaction and classification [22].

Published studies have shown that mutation test is the most effective in revealing defects [23, 29]. Mutation test is a defect-based technique proposed in [19, 20, 24] to simulate defects on ontologies, generating mutant ontologies, according to predefined mutation operators known. Mutant ontologies must reveal defects or be considered equivalent (non defective) to the original ontology, according to the DL Query test data used.

After seven openEHR ADL archetypes have been selected in the online repository CKM and converted to OWL, it was possible to generate a total of 2000 mutants ontologies and 187 defects in all OWL ontologies [20].

Detailing in the "Apgar score", "Autopsy examination" and "Fetal heart rate" archetypes were found respectively 41, 19 and 44 defects. In the "Alert" and "A health oriented check list" archetypes were found 16 and 21 defects and, in the "Informed consent request" and "Follow up action" archetypes revealed 31 and 15 defects. 139 mutant ontologies were defined as equivalent to the original ontology and other mutants were killed with test data generated.

Discussion

Archetypes are used to guide clinical practice, requiring the exploration, comparison, classification and integration of information originating from different heterogeneous systems. The OWL language presents excellent mechanisms for these activities, enabling the representation of ADL archetypes in OWL ontologies and providing an excellent semantic representation of the clinical concepts addressed.

The representation of archetypes in OWL requires the semantic interpretation of clinical archetype, so that the ADL structures defined in the archetype specialize in openEHR Reference Model classes [9]. Therefore, defects can be created by the developer during the definition of the OWL archetype constraints, and these defects can cause failures. This was possible to observe with the reached results after the application of the mutation test technique in the archetypes was represented in OWL.

It is also possible to observe with the reached results, the large number of mutants generated, determining a high possibility of occurrence of defects in the development. However, the existence of defects in obtained models does not characterize problems in the conversion process. This is because OWL ontologies are based on open world assumption, thus the representation of archetypes in OWL requires the semantic interpretation of clinical archetype.

The purpose of applying the mutation test technique was to validate the obtained results after the conversion process, in this case OWL ontologies, and not the conversion process proposed. Consistently selected archetypes in the online repository CKM were in their review state. Defects were found after the conversion process in OWL ontologies; however, this did not invalidate the proposed method of conversion ADL archetypes to OWL ontologies.

In some cases it was not possible to identify the defect applied by the mutation operator with any given test data because it was not considered in this analysis the instantiation of objects for each archetype classes represented in OWL, which might produce better results. The mutants of these cases are defined as equivalents.

For 139 mutants that have been defined as equivalent, it would be possible to obtain better results (that is, it is possible to define whether the ontology has or not defects) executing the mutant ontologies with same test data and instantiating individuals to ontology classes. Moreover, it can be concluded in these cases that the test data used were not efficient, because they do not produce distinct results from the original ontology and there is the possibility of generating new test data.

With the application of the mutation test on the archetypes in OWL, it is possible to validate the models obtained after the conversion process due to the identification of committed defects, correcting and ensuring the correction of the archetype.

Conclusion

Ontologies define concepts, classes, properties, relationships, constraints and axioms about a particular knowledge domain, which can represent the real and conceptual world through semantic identifiers. Ontologies have proven to be extremely useful to assist the development of computer systems, due to some of their own characteristics such as vocabulary for representing knowledge and the possibility for extending a generic model for a specific domain.

A clinical archetype represents a specific knowledge domain and can be modeled in ontologies in different ways. It was proposed in this paper to convert clinical archetypes of the openEHR standard implemented in ADL to OWL ontologies. In the conversion process, two steps were defined: (1) the conversion from ADL to OWL; (2) the validation of the model obtained.

The ontology developed tool Protégé was used to perform the first step of conversion process of openEHR archetypes to OWL ontologies as well as to perform the second step of process, where it was applied the mutation test in the OWL ontologies obtained. For each mutation performed, it generated a new ontology defined as mutant. For the generation and execution of test data, it used the DL Query language, a query language for ontologies available on the Protégé tool.

For this experiment, seven openEHR archetypes were used, generating an average of 285 mutants to each model. From the total of mutants ontologies obtained, 187 defects were identified and 139 mutants were defined as equivalent to the original ontologies. Each defect revealed in a mutant ontology corresponds to mutation performed by a mutation operator. The analysis and correction were made based on the characteristics of this known mutation operator.

The conversion process was experimented. A test method determined the correction of resulting OWL ontologies. A high number of mutants was generated and defects were revealed, consistent with the fact that they were under review. It is also possible to infer that this test method can be applied to other knowledge domains represented by OWL ontologies.

Acknowledgments

This research was partially funded by Capes.

References

[1] Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT, and Maldonado JA. A model-driven approach for representing clinical archetypes for Semantic Web environments. Journal of Biomedical Informatics, 42(1):150-164, feb. 2009.

[2] Kondo MNS. Mapeamento da base de conhecimento fundamentada em arquétipos: Contribuição à informática em saúde. Dissertação de Mestrado, Universidade de São Paulo, São Paulo, 2012.

[3] Martínez-Costa C, Menárguez-Tortosa M, and Fernández-Breis JT. An Approach for the Semantic Interoperability of ISO EN 13606 and OpenEHR Archetypes. Journal of Biomedical Informatics, 43(5):736-746, oct. 2010.

[4] Lezcano L, Sicilia MA, and Rodríguez-Solano C. Integrating reasoning and clinical archetypes using owl ontologies and swrl rules. Journal of Biomedical Informatics, 44(2):343-353, 2011.

[5] Beale T. Archetypes Constraint-based Domain Models for Futureproof Information Systems. OOPSLA 2002 workshop on behavioural semantics, 2002.

[6] Chen R, Garde S, Beale T, Nystrom M, Karlsson D, Klein GO, and Ahlfeldt H. An Archetype-based Testing Framework. Ehealth Beyond the Horizon: Get it there, volume 136 of Studies in health technology and informatics. p. 401-406, Amsterdam, 2008. IOS Press.

[7] Santos MR. Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais. Tese de Doutorado, Universidade Federal de Minas Gerais, Minas Gerais, 2011.

[8] Tortosa MM, and Fernández-Breis JT. OWL-based reasoning methods for validating archetypes. Journal of Biomedical Informatics, 46(2):304-317, 2013.

[9] Kilic O, Bicer V, and Dogac A. Mapping Archetypes to OWL. Software Research and Development Center, 2005.

[10] OWL Web Ontology Language Overview. Available in: http://www.w3.org/TR/owl-features/. Access on: February of 2004.

[11] OWL Web Ontology Language Guide. Available in: http://www.w3.org/TR/owl-guide/. Access on: February of 2004.

[12] Schulz S, and Hahn U. Part-whole representation and reasoning in formal biomedical ontologies. Artificial Intelligence in Medicine, 34:179-200+, 2005.

[13] Smith B. From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. Journal of Biomedical Informatics, 39(3):288-298, jun. 2006.

[14] Elkin PL, Kernberg M, Beale T, Heard S, Shafarman M, Dolin R, and Nelson D. HL7 Template and Archetype Architecture. Template Special Interest Group, nov. 2003. http://www.hl7.org/documentcenter/public/wg/template/HL7_San_Diego_v2.3.zip. Accessed in: November 23rd, 2013.

[15] Protégé Ontology Editor. Available in: http://protege.stanford.edu/. Access on feb 12, 2014.

[16] openEHR. Clinical Knowledge Manager. Available in: http://www.openehr.org/ckm/. Access on: March of 2014.

[17] Davies J, Fensel D, and Harmelen F. Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons, Inc., New York, NY, USA, 2003.

[18] Gruber TR. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Intl J. of Human-Computer Studies - Special issue: the role of formal ontology in the information technology, 43(5-6):907-928, dec. 1995.

[19] Porn AM, and Peres LM. Mutation Test to OWL Ontologies. 13th Intl. Conf. on WWW/Internet, oct. 2014.

[20] Porn AM, Peres LM, and Fabro MD. Teste de Mutação para Arquétipos openEHR Representados por Ontologias OWL. XIV Congresso Brasileiro de Informática em Saúde, Santos, dez. 2014.

[21] DL Query. DL Query Plugin. Available in: http://protegewiki.stanford.edu/wiki/DLQueryTab. Access on: February of 2014.

[22] García-Ramos S, Otero A, and Fernández-López M. OntologyTest: A Tool Evaluate Ontologies through Tests Defined by the User. Springer, p. 91-98, 2009.

[23] Wong WE, Mathur AP, and Maldonado JC. Mutation Versus All-uses: An Empirical Evaluation of Cost, Strength and Efectiveness. Software Quality and Productivity: Theory, Practice and Training, p. 258-265, London, UK, 1995.

[24] Lee S, Bai X, and Chen Y. Automatic Mutation Testing and Simulation on OWL-S Specified Web Services. Simulation Symposium, Annual, 0:149-156, 2008.

[25] SPARQL. SPARQL Query Languagefor RDF. Available in: http://www.w3.org/TR/rdf-sparql-query. Access on: February of 2014.

[26] ADL. Archetype Definition Language. Available in: http://www.openehr.org/downloads/ADLworkbench/learning_about. Access on: February of 2014.

[27] XML. eXtensible Markup Language. Available in: http://www.w3.org/XML/. Access on: February of 2014.

[28] EHR. Electronic Health Record. Available in: http://www.healthit.gov/providers-professionals/faqs/what-electronic-health-record-ehr. Access on: February of 2014.

[29] Budd TA, DeMillo RA, Lipton RJ, and Sayward FG. Theoretical and Empirical Studies on Using Program Mutation to Test the Functional Correctness of Programs. Proceedings of the 7th ACM SIGPLANSIGACT Symposium on Principles of Programming Languages, GIT-ICS, p. 220-233, New York, NY, USA, 1980. Defense Technical Information Center.

Address for correspondence

Informatics department, Federal University of Paraná, Curitiba, Brazil. email: amporn@inf.ufpr.br, lmperes@inf.ufpr.br, didonet@inf.ufpr.br