Knowledge-Based Systems 105 (2016) 175–189

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A semantic web based framework for the interoperability and exploitation of clinical models and EHR data

María del Carmen Legaz-García[a], Catalina Martínez-Costa[b], Marcos Menárguez-Tortosa[a], Jesualdo Tomás Fernández-Breis[a,*]

[a] Departamento de Informática y Sistemas, Universidad de Murcia, IMIB-Arrixaca, Spain
[b] Institute of Medical Informatics, Statistics, and Documentation, Medical University of Graz, Austria

ARTICLE INFO

Article history:
Received 14 June 2015
Revised 8 May 2016
Accepted 9 May 2016
Available online 12 May 2016

Keywords:
Semantic web
Medical informatics
Electronic health records
Ontology
Semantic interoperability

ABSTRACT

The advent of electronic healthcare records (EHR) systems has triggered the need for their semantic interoperability, which is reinforced by the opportunities for the secondary use of EHR data. The joint use of EHR standards and semantic resources has been identified as key for semantic interoperability. To date, existing tools focused on EHR standards permit to create, search, explore clinical models and to map data sources to clinical models, but do not provide an appropriate support and integration of semantic resources or permit the secondary use of EHR data. In this paper we describe an OWL-based framework that leverages EHR and Semantic Web technologies for the interoperability and exploitation of archetypes, EHR data and ontologies. It also enables the secondary use of clinical data. This framework has been implemented in the Archetype Management System (ArchMS). We also describe how ArchMS has been used in a real study in the colorectal cancer domain.

1. Introduction

The increasing use of electronic health records (EHRs) in our globalised world leads to a situation in which patients' health data are spread across different health systems. This situation demands the semantic interoperability of clinical information, that is, its meaningful communication across EHR systems. The lack of such interoperability has been internationally considered as a reason for inefficiencies within the healthcare system, contributing to the waste of billions of dollars in the United States annually [1].

In [2], the SemanticHEALTH project identified that EHR standards, ontologies and terminologies are key players to achieve the desired semantic interoperability. In the last decades, many efforts have addressed the development of EHR standards and specifications, including openEHR[1] or ISO 13606[2]. They are based on the dual model architecture, which distinguishes two modelling levels. On the one hand, the information model provides the generic building blocks to structure the EHR information (i.e. data types and data structures). On the other hand, clinical models are used to specify clinical recording scenarios by constraining the information model structures (i.e., what needs to be recorded about the measurement of blood pressure). In both openEHR and ISO 13606, clinical models are named archetypes, which are a promising way of sharing clinical data in a formal and scalable way [3]. The interest in archetypes is reinforced by the commitment of the Clinical Information Modeling Initiative (CIMI) to use them.[3] HL7 specifications[4] have also evolved to include artifacts similar to clinical models with the aim of facilitating sharing and interoperability. An example is the recent Fast Healthcare Interoperability Resources (FHIR) specification.[5]

The lack of appropriate tooling for applying and exploiting archetypes and archetype-based data in semantic interoperability environments is considered a barrier to the adoption of dual-model architectures by the majority of vendors. Therefore, the development of tools that permit to exploit the archetypes and the archetype-based data is needed [2].

Besides, the advent of EHR systems has also created new opportunities for the secondary use of data such as rapid cohort identification, quality of care assessment, comparative effectiveness research, data privacy and de-/re-identification research, phenotyping methodology and predictive modelling [4]. Some secondary

* Corresponding author. Fax.: +34868884151.
E-mail address: jfernand@um.es (J.T. Fernández-Breis).

[1] http://www.openehr.org/.
[2] http://www.en13606.org/.
[3] http://informatics.mayo.edu/CIMI/index.php/London_2011.
[4] http://www.hl7.org/.
[5] http://hl7.org/fhir.

http://dx.doi.org/10.1016/j.knosys.2016.05.016
0950-7051/

uses require combining data from different systems, which requires semantic interoperability between such systems, and in works like [5,6] the corresponding solutions are based on standards.

The achievement of the semantic interoperability will depend on the effective development and application of technologies able of supporting tasks such as detecting semantically equivalent archetypes and EHR data, or the joint exploitation of clinical models, clinical data, terminologies and ontologies for both primary and secondary uses.

The Semantic Web [7] is described as a new form of Web content meaningful to computers, and [8] proposed the Semantic Web as a natural space for the integration and exploitation of biomedical data. Semantic Web technologies are meant to enable the joint exploitation of heterogeneous, distributed content because machines are able to understand the meaning, which is provided in a precise way by means of ontologies. An ontology is defined in [9] as a common, shareable and reusable view of a particular application domain. Besides, Semantic Web technologies permit to infer new information by using automated reasoning, which can be very useful when working in semantic interoperability settings, in which discovering relations between content generated by different systems will be needed.

Our hypothesis is that Semantic Web technologies provide an appropriate support for performing the previously described tasks in the area of semantic interoperability. By Semantic Web technologies we mean the formalisms and languages that permit the semantic representation, query and exploitation of information and knowledge. Hence, in this paper we propose a Semantic Web-based framework for the joint exploitation of clinical data, archetypes, ontologies and terminologies for semantic interoperability environments. This framework has been implemented in the Archetype Management System (ArchMS), whose technological infrastructure permits to manage archetypes and EHR data from different standards using Semantic Web technologies.

The contributions of this work are (1) the exploitation of patient data, archetypes and classification rules using Semantic Web formalisms; (2) the reuse of content from existing archetypes and ontologies for the management and exploitation of clinical models and EHR data; and (3) enabling reuse of the ArchMS content by third parties because of the application of Semantic Web representation principles. As a prototypical tool, we think that ArchMS represents a good example of how Semantic Web technologies can contribute to semantic interoperability environments.

The structure of the rest of the paper is described next. In Section 2, some background and description of the state of the art in archetypes and Semantic Web technologies are presented. Our Semantic Web framework is described in Section 3. The validation of the platform in a real scenario is described in Section 4. Finally, some discussion and conclusions are put forward in Sections 5 and 6.

2. Background

2.1. Archetypes technologies

Archetypes are used to specify clinical recording scenarios such as a laboratory test, a blood pressure measurement, a medication order, etc. An archetype can be defined as a specialization of another one, can include other archetypes through the slots mechanism, and can be used in combination with others by means of templates. Archetypes are expressed in the Archetype Definition Language[6] (ADL), which structures the content in four main sections: header, description, definition and ontology. Header and description give general information about the archetype, such as name, language, author or purpose. The definition section contains the structures and constraints associated with the clinical recording scenario defined by the archetype. The ontology section provides textual descriptions for each element from the definition section and bindings to other terminologies. It should be noted that the ontology section is called terminology in the most recent version of ADL. For example, the openEHR blood pressure archetype records specific data related to the blood pressure measurement, such as systolic and diastolic blood pressure values; the protocol followed, method used, device, state of the patient in the moment of the recording, etc.

In the last years, a series of tools have been developed by the archetype community. LinkEHR[7] and the tools developed by the openEHR community, like the Archetype Editor[8] (AE), ADL Workbench[9] (AW) and the Clinical Knowledge Manager[10] (CKM) are likely to be the most widely used ones. LinkEHR permits the edition of archetypes, the representation of legacy data using archetypes as described in [10], and the view of EHR extracts. AE permits the edition of archetypes, AW permits to create archetypes and templates for ISO 13606, openEHR and CIMI and to perform management tasks related to archetypes and ADL technologies, and CKM provides a repository for managing sets of archetypes. The state of the art on tooling in this community shows the following limitations from a semantic interoperability perspective: (1) data, archetypes and terminologies are not represented and exploited in the same formalism, what limits the effectiveness of these approaches; (2) they are based on ADL technologies, whose limitations to perform activities as detecting equivalent archetypes have been shown in works like [11], since it is not easy to perform or support automated reasoning on ADL-based content.

2.2. Semantic web technologies

The Web Ontology Language (OWL)[11] is the *de facto* standard for ontology implementation, and it enables the precise description of data meaning. The subset of OWL based on Description Logics (DL), namely, OWL DL, permits the use of DL reasoning, which in this context enables performing inference tasks over the clinical models and the clinical data. In recent years, different works based on Semantic Web technologies have provided preliminary results of the feasibility of our research hypothesis:

1. OWL representations of clinical information and clinical models from different EHR standards such as ISO 13606, openEHR, HL7 or Clinical Element Models (CEM)[12] have been proposed [11–13].
2. OWL representations of clinical information and clinical models have supported the transformation of clinical models and clinical data between different EHR standards in [14,15].
3. OWL reasoning has been used for validating and checking the consistency of clinical models in works like [16] (for checking the correctness of terminological bindings) or [17] (for checking the correctness of specialization relations between archetypes).
4. OWL reasoning has been used to support the transformation of clinical models between specifications [18].
5. OWL reasoning has been used for the detection of isosemantic content in heterogeneous EHR systems, that is content with the same meaning but structurally different [19].

[6] http://www.openehr.org/releases/trunk/architecture/am/adl2.pdf.

[7] http://www.linkehr.com/.

[8] http://www.openehr.org/downloads/archetypeeditor/home.

[9] http://www.openehr.org/downloads/ADLworkbench/home.

[10] http://www.openehr.org/ckm.

[11] http://www.w3.org/TR/owl2-overview/.

[12] http://www.clinicalelement.com.