Journal of Biomedical Informatics 46 (2013) 676–689

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

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

Interoperability of clinical decision-support systems and electronic health records using archetypes: A case study in clinical trial eligibility

Mar Marcos [a,*], Jose A. Maldonado [b], Begoña Martínez-Salvador [a], Diego Boscá [b], Montserrat Robles [b]

[a] Dept. of Computer Engineering and Science, Universitat Jaume I, Av. de Vicent Sos Baynat s/n, 12071 Castellón, Spain
[b] Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Received 11 January 2013
Accepted 14 May 2013
Available online 22 May 2013

Keywords:
Clinical decision support systems
Electronic health records
Systems integration
Clinical trials
Terminology
SNOMED CT

ABSTRACT

Clinical decision-support systems (CDSSs) comprise systems as diverse as sophisticated platforms to store and manage clinical data, tools to alert clinicians of problematic situations, or decision-making tools to assist clinicians. Irrespective of the kind of decision-support task CDSSs should be smoothly integrated within the clinical information system, interacting with other components, in particular with the *electronic health record* (EHR). However, despite decades of developments, most CDSSs lack interoperability features.

We deal with the interoperability problem of CDSSs and EHRs by exploiting the *dual-model methodology*. This methodology distinguishes a reference model and archetypes. A reference model is represented by a stable and small object-oriented model that describes the generic properties of health record information. For their part, archetypes are reusable and domain-specific definitions of clinical concepts in the form of structured and constrained combinations of the entities of the reference model. We rely on archetypes to make the CDSS compatible with EHRs from different institutions. Concretely, we use archetypes for modelling the clinical concepts that the CDSS requires, in conjunction with a series of knowledge-intensive mappings relating the archetypes to the data sources (EHR and/or other archetypes) they depend on.

We introduce a comprehensive approach, including a set of tools as well as methodological guidelines, to deal with the interoperability of CDSSs and EHRs based on archetypes. Archetypes are used to build a conceptual layer of the kind of a *virtual health record* (VHR) over the EHR whose contents need to be integrated and used in the CDSS, associating them with structural and terminology-based semantics. Subsequently, the archetypes are mapped to the EHR by means of an expressive mapping language and specific-purpose tools. We also describe a case study where the tools and methodology have been employed in a CDSS to support patient recruitment in the framework of a clinical trial for colorectal cancer screening.

The utilisation of archetypes not only has proved satisfactory to achieve interoperability between CDSSs and EHRs but also offers various advantages, in particular from a data model perspective. First, the VHR/data models we work with are of a high level of abstraction and can incorporate semantic descriptions. Second, archetypes can potentially deal with different EHR architectures, due to their deliberate independence of the reference model. Third, the archetype instances we obtain are valid instances of the underlying reference model, which would enable e.g. feeding back the EHR with data derived by abstraction mechanisms. Lastly, the medical and technical validity of archetype models would be assured, since in principle clinicians should be the main actors in their development.

1. Introduction

A *clinical decision-support system* (CDSS) can be defined as "any computer program designed to help health professionals make clinical decision" [1]. This definition encompasses systems as diverse as sophisticated platforms to store and manage clinical data, tools to alert clinicians of problematic situations (e.g. drug-drug interactions), or decision-making tools to assist clinicians by providing patient-specific recommendations. In a broader sense, other systems which use clinical data to support decisions not directly related to patient care can also be considered to be CDSSs. Systems to support patient recruitment for clinical research trials are a representative example of such CDSSs.

Irrespective of the kind of decision-support task, ideally CDSSs should be smoothly integrated into the computer tools that are

* Corresponding author. Fax: +34 964 72 84 86.
E-mail address: Mar.Marcos@uji.es (M. Marcos).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.05.004

routinely used by clinicians, and more importantly they should be able to operate without the manual entry of data already entered using some other system [1]. This implies some interaction with other components of the clinical information system, in particular with the *electronic health record* (EHR) to access all the clinical data required. However, after more than 3 decades of developments most of CDSSs have been either stand-alone systems or small components embedded within EHR or physician order entry systems [1,2].

An important problem is the heterogeneity of clinical data sources, which may differ in the data models, schemas, naming conventions, and degree of detail used to represent similar data [3]. On the other hand, CDSSs very often require data at a level of abstraction higher than raw clinical data, a problem which has been referred to as the "impedance mismatch" between the CDSS and the EHR [4,5]. There have been several initiatives, involving standardisation bodies, to define generic EHR architectures for the communication of health data, such as CEN/ISO EN13606 [6], openEHR [7], HL7 CDA [8], or CDISC ODM [9]. However, their use is not widespread in current CDSSs.

One of the main contributions of recent EHR architectures is the *dual-model methodology* [10] for the description of the structure and semantics of health data. The dual model methodology distinguishes a reference model and archetypes. A *reference model* is represented by a stable and small object-oriented model that describes the generic properties of health record information (such as folder, document, section, and audit). The generality of the reference model (RM) is complemented by the particularity of archetypes. An *archetype* is a detailed, reusable and domain-specific definition of a clinical concept (such as Apgar score, discharge report, and primary care EHR) in the form of a structured and constrained combination of the entities of the RM. The principal purpose of archetypes is to provide a powerful way of managing the description, creation, validation and querying of EHRs. From a data point of view, archetypes are a means for providing structural and terminology-based semantics to data instances that conform to some RM.

We deal with the interoperability problem of CDSSs and EHRs by exploiting dual-model EHR architectures. In previous articles we propose a solution that exploits openEHR archetypes for the interoperability of CDSSs based on clinical guidelines [11,12]. In this article we take a further step and describe the implementation of a prototype that demonstrates the feasibility of our proposal. The prototype is based on a case study dealing with the determination of patient eligibility in a clinical trial (CT) for colorectal cancer screening. Typically, both clinical guideline recommendations and CT eligibility criteria are intended to be shared across different institutions, at national or even at international level, and thus the standardised access to the EHR becomes a pressing need in CDSSs for these purposes.

2. Background

The advantages of integration with the EHR were already acknowledged in early CDSSs. Thus, different authors have sought such integration while pursuing the shared use of CDSSs, in particular in guideline-based CDSSs. One of the early approaches was to separate the site-specific data references from the logic rules. The best example of this approach is the Medical Logic Modules (MLM) of Arden syntax [13,14], currently a HL7 standard for representing clinical logic. In Arden Syntax MLMs, the site-specific mappings (queries) to EHR data are defined in a separated section, known as the data section. In this section, the specific details for retrieving a required data element from a data source, such as an EHR, are enclosed in a pair of curly braces. The problem of combining site-specificity with a standard syntax has been known as the "curly braces problem".

The problem of combining data residing at different sources and providing a unified view of these data, known as data integration [15], is not exclusive of the health-care domain. Among the different approaches to data integration, federated information systems are the most widely used. These systems leave data at the sources and provide querying access to the set of data sources through a virtual federated view (schema). The federation relies on schema mapping for the integration of data sources. The mediator/wrapper architecture [16] is one of the most commonly used approaches to achieve data federation. A mediator is a read-only virtual database which is introduced between the data sources and the client applications and is capable of answering queries about the underlying data [17].

Starting from the federated approach, other initiatives rely on the definition of a global virtual schema, known as Virtual Medical Record or Virtual Health Record (VHR), over a set of local EHR systems, and on a set of mappings from the VHR to the local EHR systems. The VHR includes an information model that defines generic concepts (such as Observation, and Instruction) for representing patient data, domains for attributes in the information model (e.g. terminologies), and a query language [18]. Queries for patient data in the CDSS are posed against the VHR schema. In order to answer them they are translated into an equivalent set of local subqueries that are executed against the local data sources, whose results are then combined. This approach alleviates the curly braces problem since it is only necessary to define the mappings between the VHR and the CDSS once. When a CDSS is to be bound to a new EHR system, only the mappings between the EHR system and the virtual view are needed, thus the CDSS remains unaltered and its portability is facilitated.

3. Approach

We are concerned with the use of archetypes within CDSSs as a standardised mechanism for the interaction with the EHR, in order to obtain CDSSs that can be shared across institutions without the need for modifications in the implementation. This problem is mentioned by Sujansky as one of the heterogeneous database integration challenges in Medical Informatics [3], and is usually solved by means of abstractions that make the CDSS compatible with clinical databases from different institutions. We propose to use archetypes to build a semantically-rich VHR for this purpose. More precisely, our proposal is to develop a series of archetypes for the data/concepts that the CDSS requires, and to include references to these archetypes in the parts of the CDSS knowledge base (KB) where interactions with the EHR should occur. It is important to note that our interest in shared use (and reuse) is not limited to the KB as a whole but also covers the archetypes modelling the necessary clinical data/concepts.

We are also concerned with technical solutions to implement our approach. Technical implementation requires on one hand a platform for the access to the EHR data via archetypes, in the likely case that the EHR does not support archetypes natively. On the other hand, an inference engine supporting the use of archetypes is required. For the former, we have used the data integration engine of the LinkEHR Normalization Platform [17] (see Section 4 for more details). With respect to the inference engine, in the absence of engines that support data access via archetypes, we have chosen to use an existing guideline execution engine in combination with a specific mediator module which allows taking input data from a variety of external data sources. Concretely, we have used the Tallis Engine, which is a non-commercial execution tool for the PROforma guideline representation language [19]. PROforma is partic-