Journal of Biomedical Informatics 55 (2015) 143–152

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Automatic generation of computable implementation guides from clinical information models

Diego Boscá [a,*], José Alberto Maldonado [a,b], David Moner [a], Montserrat Robles [a]

[a] Instituto Universitario de Aplicaciones de las Tecnologías de la Información y Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Valencia, Spain
[b] VeraTech for Health SL, Valencia, Spain

ARTICLE INFO

Article history:
Received 13 November 2014
Revised 3 February 2015
Accepted 7 April 2015
Available online 21 April 2015

Keywords:
Archetype
Natural Rule Language
Implementation guide
Data validation
Clinical information model

ABSTRACT

Clinical information models are increasingly used to describe the contents of Electronic Health Records. Implementation guides are a common specification mechanism used to define such models. They contain, among other reference materials, all the constraints and rules that clinical information must obey. However, these implementation guides typically are oriented to human-readability, and thus cannot be processed by computers. As a consequence, they must be reinterpreted and transformed manually into an executable language such as Schematron or Object Constraint Language (OCL). This task can be difficult and error prone due to the big gap between both representations. The challenge is to develop a methodology for the specification of implementation guides in such a way that humans can read and understand easily and at the same time can be processed by computers. In this paper, we propose and describe a novel methodology that uses archetypes as basis for generation of implementation guides. We use archetypes to generate formal rules expressed in Natural Rule Language (NRL) and other reference materials usually included in implementation guides such as sample XML instances. We also generate Schematron rules from NRL rules to be used for the validation of data instances. We have implemented these methods in LinkEHR, an archetype editing platform, and exemplify our approach by generating NRL rules and implementation guides from EN ISO 13606, openEHR, and HL7 CDA archetypes.

1. Introduction

Capturing requirements in the clinical domain is a difficult task [1]. Traditional requirements capture methodologies fail due to the continuous evolution of clinical knowledge, the different vocabularies of clinicians and implementers, and the implicit definition of domain concepts [2]. Typically clinicians rely on non-formal approaches (such as spreadsheet or word processor files) to document their domain requirements. This kind of approach is not suitable for cooperative and long term use as it is prone to errors and version control problems. In order to solve these problems several methodologies have been proposed.

Templates are the mechanism used by HL7 CDA [3] for the specification of clinical information models. In spite of not being computable, CDA Implementation guides are the most common way for the specification of such templates in an understandable way. They usually include an introductory section describing purpose, scope, intended audience, conventions used in the guide, and separated sections for each kind of CDA components (mainly document, section, and clinical statement templates). Each one of these sections contains all the relevant templates for a given clinical model. For each template, a template identifier, a description, a set of constraints over the attributes of a given CDA component, and an XML example are provided. The implementation guide is usually completed with terminological value sets and bibliographic references. Implementation guides play a central role in HL7 world. As an example, they have been adopted for the definition of the Consolidated CDA (C-CDA) Templates [4], which are being used to help providers to meet the applicable Meaningful Use objectives [5]. However, the interpretation of the constraints in an implementation guide may differ from person to person [6], therefore limiting semantic interoperability.

Another type of resource for the specification of clinical information models are archetypes. Archetypes are a key part of the dual model approach on which the EN ISO 13606 norm [7] and the openEHR specification [8] are based. The dual model approach is a recent paradigm for the specification of EHR Architectures

* Corresponding author at: Instituto ITACA, Universitat Politècnica de València, Camino de vera s/n, Edificio 8G, Spain. Tel.: +34 963870000x75277.
E-mail address: diebosto@upv.es (D. Boscá).

http://dx.doi.org/10.1016/j.jbi.2015.04.002
1532-0464/

(EHRA). It distinguishes two models: the Reference Model (RM) and Archetype Model. In a broad sense, a RM is an abstract representation of the generic and stable entities and relationships of a given domain. It is designed to provide a basis for the development of more concrete models and implementations. In the domain of Electronic Health Records (EHR), a RM defines the framework for describing all EHR entries or clinical statements, the way how they are aggregated, and the context information needed to meet ethical, legal and provenance requirements. The generality of the RM is completed by the particularity of archetypes. Archetypes are detailed and domain-specific definitions of clinical concepts in the form of structured and constrained combinations of the entities of the reference model. Archetypes may logically include other archetypes, and can be specialized to better fit the specific requirements of each use case. They can be bound to clinical terminologies and ontologies to semantically describe the elements of information. What is important here is that for each domain concept, a definition can be developed in terms of constraints on the RM entities. Each domain concept is also given an archetype node identifier (following the 'atNNNN' pattern where N stands for a digit) and a textual label. ADL (Archetype Definition Language) [9] is a formal language developed by openEHR for expressing archetypes that has been adopted by EN ISO 13606 standard. Even if archetypes are based on a formal language (ADL) understandable by computers, users still need specific tools and knowledge of the underlying reference model to define and understand the clinical models completely.

To allow users unfamiliar with the archetype methodology or a particular reference model to understand clinical models without using specific tools, a formal document similar to the implementation guides is required. What we need is a formal document that has at least the same expressiveness than an archetype and at the same time is easily understandable even by non-technical users.

Our proposal, as described in Fig. 1, aims to achieve the automatic generation of computable implementation guides from archetypes. Our objectives are twofold:

(1) To generate implementation guides that can be used in the development of computer systems by IT technical staff. For this purpose, we use archetype texts, descriptions, and terminology bindings. We also include other automatically generated materials such as sample XML instances and validators.
(2) To document archetypes or templates in order to ease their understanding by health professionals without the need of specific tools. For this purpose, we transform the potentially complex archetype constraints into English-like rules. This is achieved by the use of Natural Rules Language (NRL) [10]. We also include additional reference materials in the implementation guide, such as a mindmaps, value sets and bibliographic references.

Our solution will improve the current implementation guides generation process in two different ways: Firstly, our implementation guides are based on a formal natural language that allows the direct application to end EHR systems and data, and secondly, implementation guides are generated automatically from the organization's information models, in our case archetypes.

We will exemplify our approach by generating implementation guides from an EN ISO 13606 archetype, openEHR archetypes from the Clinical Knowledge Manager (CKM) [11] and HL7 CDA archetypes from the Genetic Testing Report.

We will evaluate the correctness of our methodology from three different perspectives. First, we demonstrate formally that the generated data instances are valid with respect to the underlying information model. Second, we test if generated rules can correctly validate data instances. Third, we evaluate the quality of implementation guides and their usefulness for the development of EHR systems.

2. Background and related work

There exists a wide range of formal rule languages for the definition of constraints on data. One of the most known is the Object

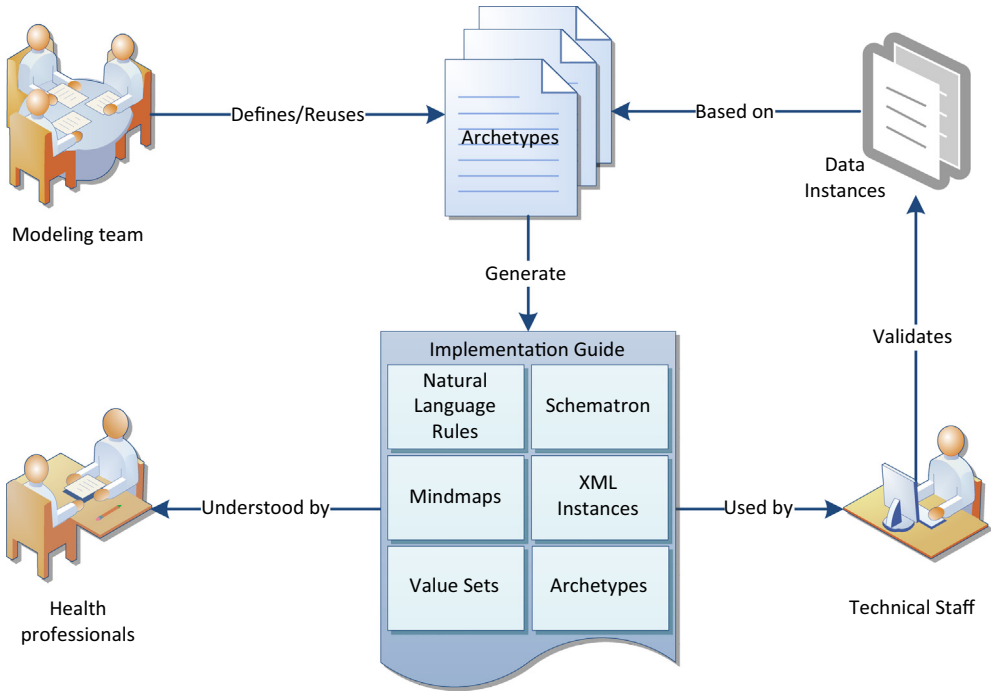

Fig. 1. Proposed architecture for the generation of implementation guides from archetypes.