INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 79 (2010) 585–597

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

Extraction of standardized archetyped data from Electronic Health Record systems based on the Entity-Attribute-Value Model

Georg Duftschmid, Thomas Wrba, Christoph Rinner*

Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria

ARTICLE INFO

Article history:
Received 4 June 2009
Received in revised form
29 April 2010
Accepted 30 April 2010

Keywords:
Medical Records
Medical Records Systems
Computerized
Reference Standards
Models
Theoretical

ABSTRACT

Objective: The ISO/EN 13606 Electronic Health Record architecture standard permits semantically interoperable exchange of Electronic Health Record data by using archetypes to define the structure and semantics of Electronic Health Record contents. Practical implementations of the ISO/EN 13606 standard have been scarcely reported on, and none of the publications describes in detail an efficient technique of archetype-compliant data extraction from an Electronic Health Record system. We address this research issue in the present report, and focus on a specific class of largely research-oriented Electronic Health Record systems which are internally based on the Entity-Attribute-Value Model.

Method: We propose an approach for extracting data described by archetypes from an Entity-Attribute-Value based Electronic Health Record system in an ISO/EN 13606-conformant manner. The approach is based on mapping from the structure of exported source documents to the archetype. It is implemented via standard XML technologies.

Results: We tested our approach on an Electronic Health Record system employed for clinical research at the Medical University of Vienna. Using test data defined by three different archetypes, source documents were successfully extracted as archetype-conformant ISO/EN 13606 Electronic Health Record extracts.

Conclusions: Electronic Health Record data may be effectively extracted from Entity-Attribute-Value based Electronic Health Record systems using the suggested approach. As a prerequisite for applying our approach, the internal data model of the Electronic Health Record system and the archetype must overlap in a way that a semantic mapping between them is possible. The system must further provide an XML interface which permits the export of the source documents in conventional format. The export must include data and metadata that are mandatorily postulated by the archetype and the ISO/EN 13606 Reference Model.

1. Introduction

Cross-organizational Electronic Health Record (EHR) communication, which enables authorized healthcare providers to access all relevant patient data regardless of where the data were created, will constitute a key component of future health care. The European Union acts as a political driving force behind this vision, and underlines its corresponding commitment by naming the "interoperability of EHRs" in its *eHealth*

* *Corresponding author*. Tel.: +43 1 40400 6696.
E-mail address: georg.duftschmid@meduniwien.ac.at (G. Duftschmid).
1386-5056/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2010.04.007

action plan as one of the goals to be strived for by member states [1]. As a consequence most European countries, Austria being one of them, are planning, or already working on, the implementation of a national EHR system [2,3]. The next goal – which is also pursued for a limited domain by the epSOS project [4] funded by the European Commission within the Seventh Research Framework Programme (FP7) – will be to achieve interoperability of the individual national solutions.According to a recent literature review of HL7's *EHR Interoperability Work Group*, a distinction is made between three levels of interoperability [5]:

- *Technical (syntactic) interoperability* focuses on the transference of data rather than their meaning. It aims to neutralize the effects of distance.
- *Semantic interoperability* communicates meaning and hereby aims to ensure that the communicated information is understood in exactly the same way by the sender and the recipient.
- *Process interoperability* primarily deals with methods for the optimal integration of computer systems into actual work settings. It aims to coordinate work processes.

The ISO/EN 13606 EHR architecture standard provides a framework for achieving interoperable EHRs [6,7]. It is based on the dual model approach, which combines two kinds of models – the Reference Model and archetypes – to represent EHR contents [8].

The *Reference Model (RM)* defines a set of generic classes from which any EHR content may be composed. It can be characterized as a communication model, i.e. a sending EHR system transforms the content to be exchanged from the system's internal data model to the RM and the receiving EHR system transforms it back to its own internal data model. A common RM, to which sender and receiver map their internal data models when exchanging data is essential to achieve technical (syntactic) interoperability [9].

Archetypes define the structure and semantics of EHR contents. For each EHR content, such as a blood pressure measurement or a pain symptom, a corresponding archetype specifies how the EHR content should be composed from instances of the RM. In other words the archetype acts as a "construction plan" (see Fig. 1). The idea is to keep the RM stable as a solid basis for EHR software developers and, at the same time, permit newly evolving medical knowledge to be easily integrated by means of the separate archetype layer.

By specifying the structure of individual EHR contents and providing an interface to medical terminologies within their *ontology section*, archetypes are an important means of achieving semantic interoperability between the different communicating EHR systems. To examine the practical applicability of archetypes and so-called "templates" (the latter representing groupings and refinements of archetypes for local applications), the British National Health Service (NHS) initiated several studies in the last three years, and drew positive conclusions in this regard [10–12].

Some existing systems – which we will report on in section 4.1 – utilize the dual model approach. However, very few of them actually implement the ISO/EN 13606 standard and none of the publications describes in detail an efficient technique of data extraction from an EHR system as a valid and archetype-compliant instance of an RM. In this report we present an approach to extract data from an EHR system as an ISO/EN 13606 RM instance, which satisfies the constraints prescribed by archetypes. As the ISO/EN 13606 EHR architecture is similar to the openEHR architecture, the latter could have served as an alternative basis for our work. We decided to rely on the ISO/EN 13606 EHR architecture because, in contrast to openEHR, it has been accorded the status of an official standard. Further, the openEHR specification[1] of the EHR extract information model is still under development.

According to ISO/EN 13606-1 and ISO/TR 20514 [13], an instance of the ISO/EN 13606 RM is named an *EHR extract*. In the present report we will use the term *archetyped EHR extract* for a set of EHR data that constitute a valid instance of the ISO/EN 13606 RM and additionally comply with an archetype.[2]

A simple way to create an archetyped EHR extract could be to instantiate the RM from the EHR system's data and then to derive a corresponding archetype from the resulting EHR extract "on the fly" (which could be done straightforwardly). This is not what we intend. Instead, we presume a scenario in which source EHR data have to be extracted and made compliant with *predefined* archetypes. We call an archetype *predefined* if it was developed and published in advance by an external party, ideally in standardized form by an organization that adheres to the principles of domain knowledge governance [14].

Our overall goal is to contribute to the cross-organizational semantic interoperable communication of EHR data based on archetyped EHR extracts. Within the communication network, each EHR system typically uses an individual internal data model. For the exchange of data, the data are transformed from their internal model to the model of the archetyped EHR extract and vice versa. This presupposes that the internal data models are generic enough to cover the model of the archetyped EHR extract. According to [15], the model of the archetyped EHR extract, which merges the prescriptions of the archetype and the RM, and the internal data models must possess a common "modeling safe zone". In this zone the internal data models of the communicating systems and the model of the archetyped EHR extract overlap in a way that a semantic mapping between them is possible. This also includes the terminologies used in the different models.

Our work focuses on the extraction of archetyped EHR extracts from EHR systems. We have not yet implemented the import of archetyped EHR extracts into an EHR system. As the source of data extraction, we will focus on a specific category of EHR systems, namely systems based on the so-called "Entity-Attribute-Value" (EAV) data model. This data model is the basis of several largely research-oriented EHR systems such as [16–18] because of the benefits of its generic design [19,20]. EAV and conventional database modeling were com-

[1] See http://www.openehr.org/svn/specification/TRUNK/publishing/architecture/rm/ehr_extract_im.pdf.

[2] Actually an archetyped EHR extract will frequently comply with more than one archetype. In the following we assume that there is always one "root" archetype, which may include other archetypes via slots.