Data Integration for Clinical Decision Support Based on openEHR Archetypes and HL7 Virtual Medical Record

Arturo González-Ferrer[1], Mor Peleg[1], Bert Verhees[2], Jan-Marc Verlinden[2], Carlos Marcos[3]

[1]Department of Information Systems, University of Haifa, Israel, 31905
[2]ZORGGemak BV, Papelaan 85F, 2252 EG Voorschoten, The Netherlands
[3]ATOS Spain, C/Albarracín, Madrid, Spain, 28037

Abstract. Clinical Decision Support Systems (CDSS) have gained relevance due to their potential to support patient-centric care, but their deployment still has to overcome barriers to become successful. One of these barriers is the integration of patient data with the CDSS engine, a tough challenge given the need to address interoperability with many different existing systems and medical devices. The MobiGuide project aims to build such a CDSS, providing guideline-based clinical decision support through a Personal Health Record (PHR). This PHR is the main component through which the CDSS could access patient data originating from hospital EMRs and wearable sensors, but it also contains the log of the recommendations provided by the CDSS. Using a case study, we compare data-representation standards through which the PHR could be developed, while considering expressiveness and usability requirements. We propose to develop the PHR by combining openEHR archetypes and the HL7 Virtual Medical Record standard, supported by a service oriented framework for data exchange. This proposal aims to close the gap between the HL7 and the ISO/CEN 13606 by using an openEHR-based approach.

Keywords: CDSS, PHR, interoperability, archetypes, openEHR, vMR

1 Introduction

Recent work in the area of Medical Informatics [1] suggests that the development and deployment of CDSS in Healthcare Organizations will improve patient-centric care, while providing the possibility of carrying out shared decision making processes between patients and physicians. On the other hand, it is agreed that this will not be feasible without overcoming traditional barriers for the integration of different patient data [2] that can be found scattered throughout different Information Systems, like Electronic Medical Records (EMR), or more dynamically generated from patient-worn mobile sensors connected to Body Area Networks (BANs).

Besides the traditional terminology standardization issues mentioned in [2], where different coding specifications can be used to assign an agreed code to a specific clinical concept, further technical and semantic aspects should be considered when developing a CDSS, where other complex interactions between different system compo-

nents can usually be needed. Concretely, different standards for both the representation and exchange of clinical data between different systems have been developed in the last two decades. Initially, these standards were designed considering the technical and computational issues of clinical data management (e.g., the HL7 v2.x message standard). While this has been a first big challenge to overcome for the adoption of IT systems in healthcare, these standards are not straightforwardly usable by humans, thus they are not optimal for data representation. Therefore, new standards following a higher abstraction level were developed recently, given the need of stakeholders to manage and interact with data. Some examples are standards based on the HL7 RIM (like HL7 CDA [3] or HL7 vMR), or detailed clinical models like archetypes [4].

Different stakeholders are involved in the process of developing a CDSS. Consider the case study of a knowledge engineer who is in charge of modeling a computer-interpretable guideline (CIG) for a concrete disease. This CIG will constitute the knowledge base for a guideline-based CDSS. The knowledge modeling step entails the representation of decision criteria relating to clinical abstractions (e.g., 'if the patient is taking an oral anti-diabetic medication and has blood pressure higher than the goal level, take step A, otherwise take step B'). When a CIG is enacted by a CIG engine, data needs to be acquired from the PHR, corresponding to the lower-level CIG concepts so that the CIG's decision criteria are evaluated. The PHR data can be much more specific than the one needed by the guideline; for example, the CIG may evaluate if the patient is taking an anti-diabetic medication, whereas the PHR may hold different codes for specific medications with particular dosage.

Mapping CIG knowledge to raw PHR data involves more than one-to-one mapping of CIG concepts to PHR data codes. The mapping [5] is a knowledge-data integration problem, where high-level concepts (e.g. 'high blood pressure') need to be linked to low-level data (e.g., systolic and diastolic BP values), whose evaluation will determine if the pressure is high or not. Such knowledge that defines abstract concepts in terms of more concrete concepts can be defined by the CIG modeler as part of the CIG Knowledge Base (KB) or as part of the mapping KB.

Furthermore, the potential of having the patient information scattered throughout several information systems or devices makes it beneficial to use a PHR which stores not only clinical data but also recommendations output by the CDSS. The type of PHR to be developed is known as *integrated* or *interconnected* PHR [6], since the data imported may be generated in different hospitals, medical devices, etc., and where the patient and the physicians (and possibly other roles likes nurses or patient relatives supporting the care process) are allowed to enter information into selected areas of the record. The data stored in the PHR should later be viewed, searched, and analyzed (e.g., for compliance, for finding patterns) by clinical staff, patients, and researchers. Therefore, the data should be provided in a way that is understandable for these stakeholders. In such scenario, not only the representation of data is relevant but also the interfaces provided for external systems to access and exchange data. Even for the mapping task and the exchange of data, a comprehensible and intuitive data model is needed in order to help the guideline modeler, the database administrator, and the clinical expert work together to define correct mappings from knowledge to data.

This paper aims to address the selection of clinical data standards for the design of such a PHR in order to 1) integrate and represent patient information from different sources, considering not only relevant literature, but also real market needs, and 2) facilitate the integration of patient data with a guideline-based CDSS and also the representation of CDSS output information, considering the different stakeholders involved in the process of designing and setting up the system. The project where this study is framed is described next.

The MobiGuide project (FP7-287811, www.mobiguide-project.eu) aims to build a guideline-based CDSS supported by the Asbru language and tools [7], initially covering the domains of Gestational Diabetes Mellitus (GDM) and Atrial Fibrillation (AF), but aimed to be portable to other domains in the future. One of the project's challenges is the integrated representation of different sources of patient-related information. By integrating patient data into a PHR, MobiGuide aims to have access to more dynamic information than the hospitals' EMRs usually include, thus being closer to provide patient-centric decision support. This patient data can be related to several aspects of the evaluation of the patient condition, considering both inputs and outputs of the MobiGuide system: the clinical history of the patient (e.g., previous diseases or conditions), his/her socio-demographic aspects (e.g., environment, habits or family support), the information coming from different medical sensors in order to monitor and evaluate the actual patient condition (e.g., blood pressure, physical activity monitoring), specific knowledge abstractions derived from inference processes made by the system components, or guideline-based recommendations and instructions provided as output by the system.

For this aim, data integration is a critical issue, and needs to be addressed according to the variety of data and information, but also for the purpose of providing decision support, given the nature of the MobiGuide project. This differs substantially from other projects where the effort is directed to share patient summaries among organizations, like the epSOS initiative (www.epsos.eu) which aims to provide cross-border services (like e-prescription) for citizens travelling across Europe.

The rest of the paper is structured as follows. Section 2 describes the standards selected for evaluation and the experiments carried out. Section 3 includes the evaluation of the standards, Section 4 describes our proposal, Section 5 shows related work and Section 6 includes a discussion and presents our conclusions.

2　　Materials and Methods

In this section we describe the different standards that we have evaluated for data representation in the PHR, the ISO/CEN 13606 European Norm for Healthcare IT systems, and finally the experiments carried out in order to ground our decision.

2.1　　Possible Standards for the PHR

In Section 4, we evaluate different clinical standards available regarding the representation of patient data, without losing sight of the influence they could have on the interoperability of the system components. Given that the MobiGuide system is de-

signed following a distributed architecture, we intend to support data exchange by using standard service-oriented interfaces (e.g. SOAP or RESTful web services).

HL7 RIM and HL7 vMR (www.hl7.org). The HL7 Reference Information Model (RIM) is the cornerstone of the HL7 v3.x development process. It is a model shared among all clinical domains. The RIM is an ANSI (American National Standards Institute) approved standard and it is also adopted by ISO (International Organization of Standardization), concretely ISO/HL7 21731:2006. With the RIM, it is very simple to express any fragment of patient data, as can be observed in **Fig. 1(a)**.

HL7 Virtual Medical Record (vMR) is a recent standard (derived from the RIM) especially designed for the purpose of integrating patient data with CDSS. The group in charge of this standard conducted a multi-institutional analysis of CDSS data needs [8] encompassing twenty CDSSs from four nations, which included both large-scale home-grown CDSSs and a number of commercial CDSSs.

Fig. 1. (a) An example of an HL7 RIM frame representing a heart rate measurement, and (b) an example of an HL7 vMR frame representing a CDSS recommendation for the administration of the substance 'Celestone' twice a day for 2 days (using Protégé, http://protege.standford.edu)

HL7 CDA version 2. HL7 CDA (the Clinical Document Architecture [3], see example in **Fig. 2**) is an XML standard that specifies the structure and semantics of clinical documents for the purpose of exchange. CDA version 2 is focused on structured content, so that it enables the formal representation of clinical statements by means of the CDA 'entry' element, which also conforms to the RIM. Only the narrative blocks required are not based on the RIM, as they store unstructured content. CDA was originally intended as a standardized way of communicating clinical notes, but the user

community has utilized it as a persistence model as well. It can also be enhanced by HL7 templates, used to refine these existing models with a focused scope or domain.

{} title	Vital Signs	
▲ entry		
▲ observation		
═ classCode	OBS	
═ moodCode	EVN	
▲ code		
═ code	364075005	
═ codeSystem	2.16.840.1.113883.6.96	
═ codeSystemNa...	Snomed-CT	
═ displayName	Heart rate	
▾ statusCode code=completed		
▲ effectiveTime		
═ value	20111219	
▲ value		
═ xsi:type	RTO_PQ_PQ	
▾ numerator value=60		
▾ denominator value=1 unit=min		

Fig. 2. An example of an HL7 CDA representation of a heart rate measurement (using the XML grid view of XMLSpy, http://www.altova.com/xmlspy.html)

OpenEHR Archetypes. The most distinctive feature of the openEHR standard (www.openehr.org) is the *archetype*. Via archetypes [9], a separation between clinical concerns and the technical design of data storage is made possible using **two-level modeling**. While the first level model takes care of the *technical* concerns and deals with the information structure and data types using an underlying Reference Model (RM), the second level handles the concerns of the clinical domains, which are about how to represent and communicate the *semantics* of the clinical content. Archetypes can be designed from scratch, or adapted from preexisting ones. Furthermore, different archetypes can be aggregated into one by means of archetypes templates, which also support semi-automatic derivation of user interfaces.

2.2   The ISO/CEN 13606 Norm

This multi-part standard [10], include terminology, security and interface considerations for the standardized exchange of Electronic Health Records, and concerning information modeling, it propose to use a dual modeling approach [4, 9] without specifying the format (either the openEHR model or the CDA combined with templates are expected to be possibilities in this sense [11]. However, EN13606 does not provide all the requirements to create EHR systems. Instead, it is more directed to communication of EHR extracts between components, and it defines a detailed and flexible authorization-mechanism, usable in almost any legal situation worldwide. CEN acknowledges the fact that standards like openEHR or HL7 RIM can provide the semantic level for representing patient data, so their effort has been directed to align the 13606 standard with both initiatives, instead of trying to develop yet another data standard. As described in Section 9.5 of [12], CEN signed a Memorandum of Under-

standing (MoU) with HL7 for aligning the CEN information model and the RIM, another MoU with openEHR to adopt the archetype concept. On the basis of these agreements, the RIM has been also influencing both openEHR and CEN reference models. For these reasons, we didn't evaluate the ISO/CEN 13606 capability for the semantic representation of patient-related information (an issue it does not address properly [11]). Instead we checked if the guidelines that this norm propose regarding 1) the use of a two-level modeling approach and 2) security concerns, could be adequately followed by using the data standards evaluated in Section 2.1 (e.g., security is not addressed by HL7 or openEHR, which delegate it to be solved during system deployment). Thus, it is also our intention that the solution proposed will comply with the requirements and standards of the enlarged European market.

2.3 Experiments

In order to support a decision about the standards to be used in MobiGuide, we used the following examples to represent several concrete patient data items with the different standards described previously in Section 2.1:

- **EMR data (quantitative)**: Heart rate result: 60 bpm measured on 19/12/2011.
- **EMR data (qualitative)**: Brother of patient X has diagnosis of "Myocardial Infarction", recorded on 19/12/2011.
- **BAN data**: Heart rate waveform: heart rate results recorded every second for 5 minutes starting at 8 a.m. on 19/12/2011.
- **Abstraction**: Tachycardia (e.g., heart_rate > 115 bpm) during the interval of 8:00-8:30 on 19/12/2011.
- **Decision-support**: the next recommendations were given at 8am on 19/12/2011: (1) measure serum urea every 3 days; (2) hospitalize patient now; (3) perform echo umbilical 2-3 weekly; (4) give celestone 12mg 2 times every 24hr for 2 days.

Fig. 1(a) and **Fig. 2** show how the first observation can be represented using the HL7 RIM and CDA respectively. HL7 RIM uses the *Act_code* attribute to refer to the subject of the observation (heart rate and its code from a controlled vocabulary), the *mood code* to refer to a recorded event (of heart rate), the *value* to report 60 bpm, and the *critical_time* to refer to the observation's date. CDA has a section for vital signs. In it there are entries about observations. As done in HL7 RIM, CDA specifies the observation using similar attributes. Archetypes have a more flexible structure than HL7 RIM-based models thus different modeling styles may be used to create them. While the vMR is available as an XML schema, we represented the vMR classes[1] as archetypes, since our proposal is linked to this technical solution, as further described in Section 4. We created 22 high-level archetypes using LinkEHR [13] corresponding to the HL7 vMR classes (e.g. encounters, observations, problems, procedures, substance administrations, etc.), which we utilized to represent the examples. In

---

[1] See the vMR model at http://wiki.hl7.org/index.php?title=Virtual_Medical_Record_(vMR)

Fig.3 (left) we see the openEHR composition approach. The different information about observations is grouped by using a "composition", where the "archetype slots" are linked to other (4 in our case) fine-grained lower level archetypes, based on other vMR classes. (a) "Observation Order" archetype is used to represent an order to conduct an observation, such as a laboratory test. (b) "Observation Proposal" can be used to represent recommendations e.g., by a CDSS, for an observation to take place. (c) "unconducted observation" can be used to indicate that an observation was not made (e.g., 'alcohol addiction' was not assessed). In the example selected, we represented the result of an observation (heart rate 60 bpm) by means of (d) "Observation Result archetype", shown in **Fig.3** (right). The figure points to values of the attributes observationValue (60bpm), observationEventTime (19/12/2011), and observationFocus (which holds the vocabulary code for Heart Rate, 138875005, from SNOMED). Further information can be specified, like the interpretation value (e.g., is the measurement value considered high or low?), the body part where the measurement was taken (linking to the "Target Body Site" archetype), or the observationMethod (e.g., was a direct measurement or an indirect calculation).
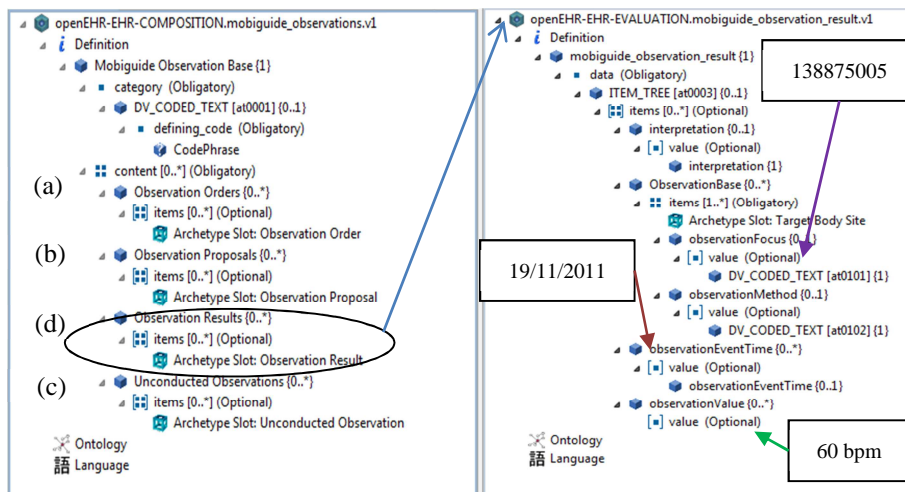
Fig.3. Composition representing encounter-related information through different Observation Archetypes (left). Observation Result Archetype, representing a heart rate measurement (right)

Fig. 1(b) represents the underlined recommendation: "give celestone 12mg 2 times every 24h for the next 2 days". HL7 vMR includes classes for representing recommendations proposed by a CDSS, e.g. 'SubstanceAdministrationProposal' class, shown in **Fig. 1(b)**.

It is important to highlight that although the set of examples selected might not encompass all the possible data types that need to be represented when developing a generic CDSS like MobiGuide, we considered it representative enough for our aims of evaluating the aspects later commented in Section 3. This derives from the fact that we are not only evaluating expressiveness of a concrete standard to represent the examples, but also factors which are related to how the standard facilitates building a

CDSS, how the information is structured and how this affects the knowledge-data mapping process, or what mechanisms it provides for semantic data integration. Furthermore, we will show that our solution is based on a standard that already encompassed a more extensive study about expressiveness requirements during its development, which makes us trust in the robustness of our decision.

3   Evaluation

In this section we evaluate the standards using criteria relating to **back-end** (where the data is actually represented and stored) and **front-end** (1. the data exchange interfaces between the EMRs and medical devices and the PHR, and 2. the conceptual link of the DSS component with the PHR, where knowledge-data mappings will be defined and used for the CIG interpretation) interfaces. As we already commented, not only the representation of data in the back-end will be relevant in our study, but also how it affects the development of the CDSS. That is the reason to also include evaluation of the front-end interfaces. We consider openEHR and HL7 CDA for both front-end and back-end, but focus on HL7 RIM for the back-end, and HL7 vMR (which is actually a subset of the RIM) for the front-end, since the latter is specifically designed for data integration with CDSS, but can also be used for integration of data from the EMRs to the PHR, as will be shown.

The first criterion for evaluation is the *expressiveness* of the standards. We have checked five different examples, where a considerable set of typical patient data (and CDSS recommendations) attributes was represented. Concretely, we have represented up to 11 different data aspects: **time-specific observations** (e.g. heart rate), **periodic observations** (waveform heart rate) including **periodic temporal patterns** (every second for 5 minutes), **dates and times**, **abstractions** (e.g. tachycardia), **recommendations for periodic observations** (e.g. measure serum urea every 3 days), **time-specific instructions** (hospitalize patient), **instructions with several iterations** (perform echo umbilical 2-3 times a week), **substance administrations** specifying **substance** (celestone 12 mg), and the **periodic pattern for this recommendation** by means of frequency and duration (every 24h during the next 2 days). We considered good support when the standard was able to represent all these attributes.

The evaluated standards' RMs (that of openEHR and HL7 RIM) have been developed for more than a decade, addressing many different use cases, so they support all the data types we intended to represent. As commented previously in Section 2.2, the harmonization efforts between openEHR and HL7 to align their respective RMs are a fact (by means of the agreement CEN-HL7, since CEN proposes using a subset of the openEHR RM). Thus, not many differences can be found in the ability to represent all the constellation of data types with both standards. In fact, some work has been carried out in order to identify the equivalences between openEHR and HL7 data types[2], and specific agreements were signed recently (April 2012) between openEHR and HL7 New Zealand to progress on this and other issues.

[2] http://wiki.hl7.org/index.php?title=OpenEHR_datatypes_mapping

Focusing on aspects specifically related to the **front-end** interfaces, one of the most important criterions to evaluate is *user-friendliness*. First, this aspect should be interpreted as the suitability of the standard's conceptual model for representing clinical guideline data, which is relevant for the knowledge-mapping requirement. For this criterion, the vMR standard is the one that provides the best support. This derives from the fact that it has been designed for clinical decision support, and its conceptual model is very similar to what the physicians are used to (e.g. observations, problems, procedures, or clinical assessments and recommendations for care plans). Second, based on our experience, knowledge engineers and database administrators can understand it quite straightforwardly, since it encompasses a small set of classes with attributes clearly defined in HL7 documentation, and where all types of patient data are instances of these classes; this user-friendliness should enable hospitals to connect with and use our system. Note that this differs from using archetypes created from scratch for a specific purpose, for which there is no predefined structure, hence each data item could be defined differently. Considering HL7 CDA does not seems a good choice in this case, since this standard was created to store EHR extracts (documents) and no appropriate distinction is done in the standard for representing recommendations provided by the CDSS. A concrete data item would need to be found during the knowledge-mapping step within complex XML-based documents, which can be really challenging and inefficient. What's more, a vMR has been used in other projects sharing our same goal (see Section 5). Note that this criterion is related to the evaluation of "easiness to represent data in the back-end" described later in this section.

Finally, a third important criterion to evaluate is the *ability to link with the back-end*. This includes (1) data from EMRs and medical devices front-end needs to be stored in the back-end and (2) concepts from CIGs (DSS front-end) need to be linked to data represented in the back-end. In this sense, using openEHR archetypes in the backend for representing the vMR classes is the best solution we found since on one side, archetypes provide a high flexibility for possible adaptations of the vMR that might be needed (e.g., complex data like ECGs) and, on the other side, the conceptual linking between the guideline concepts and a vMR-based PHR is possible and more comprehensible than using either CDA or concept-specific archetypes (as the ones found traditionally in the openEHR Clinical Knowledge Manager). Furthermore, using an archetype-based representation of the vMR standard, we keep compliance with the ISO/CEN 13606 norm, sharing the two-level modeling approach view.

Regarding the **back-end**, we checked several functional and non-functional aspects. As functional aspects, we checked the *ease to represent data* and to *extend the standard*. For this aspect, the learning curve, the documentation, and how explicit was the representation of a specific data aspect in each standard (of the 11 aspects commented before), plus our personal experience while representing them were key issues to evaluate. Another functional aspect checked was the *provision of functionality for semantic integration* (i.e., querying interfaces and support for vocabularies). We don't consider HL7 CDA as the best suitable option, mainly because it has a high learning curve, and it uses a complex representation format (usually based on XML), nor it is trivial to extend it. Furthermore, we found that its provision of semantic integration is not as good as the one provided by openEHR, as explained below. On the other hand,

while the HL7 RIM provides ease to represent and extend the standard, openEHR outperforms it in terms of semantic integration functionalities; it provides a powerful mechanism to support vocabularies (by means of the archetypes' ontology section), and the querying of data can be enhanced by using Archetype Query Language (AQL) supported by a Service Oriented Architecture (SOA). Finally, openEHR can make the most of the two-level modeling approach, where the storage can be selected independently of the high-level archetypes representation, which makes it more flexible by separating both layers, which also improves the facility to represent data.

As non-functional aspects, the evaluation of *security and privacy* (authentication, authorization, and audit trail) was not possible, since the openEHR and HL7 standards don't include consideration of such aspects, deriving its support to concrete implementation. Given this lack of support and the relevance of considering these issues in our project, we considered beneficial using an openEHR-based middleware provided by a partner of our project that aims to support the security guidelines suggested by the ISO/CEN 13606 norm. Another aspect to check is *scalability*, in terms of the number of users and simultaneous access and the performance of the system. This aspect is not covered by the standards and also depends on implementation. Although we did not conduct experiments, the mentioned openEHR-based middleware can be deployed using load balancing techniques and it can be connected to a high-performance NoSQL database as low-level data infrastructure.
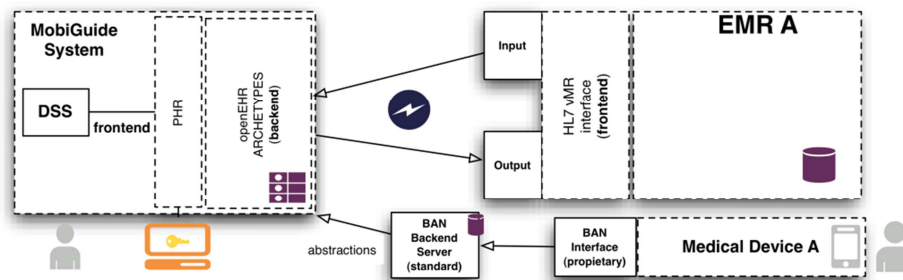
Fig. 4. A simplified view of frontend and backend interfaces for the PHR, EMRs and BAN

4   Selection of Standards Proposed

Based on the previous evaluation, our proposal is based on two main pillars. First, the HL7 vMR structure is ideal to address the different front-end implementation needs. Second, using openEHR archetypes designed following the structure of the HL7 vMR is ideal for addressing the back-end needs and the back-end/front-end communication. **Fig. 4** shows a schema of all the proposed interfaces.

If we first analyze the *front-end interface of the different EMRs with the system*, requiring the hospitals that would like to use MobiGuide to export their data following the HL7 vMR standard is technically affordable, beneficial for commercialization, simplicity and standardization. This occurs because of the similarities that the vMR shares with the models usually represented in most EMRs (i.e., hospital EMRs nor-

mally have sections for problem list, laboratory test results, medications, allergies, etc., as the vMR classes). This means that a data item found in a particular EMR can be easily mapped to a vMR class without effort. Second, the HL7 working group in charge of this standard is developing implementation guidelines for transforming HL7 v2.x messages to and from the vMR model (and also to/from CDA messages). This is very relevant, since it is a known fact that most hospitals are able to export messages in such a way. This would simplify tremendously the process of exporting patient data from new hospitals that want to use MobiGuide to the vMR service model. Furthermore, the vMR standard not only has defined input classes, but also output classes that can be used in order to represent and integrate recommendations given by the CDSS back into the hospitals' EMRs. What's more, the analysis of data needs undertaken by the HL7 vMR standard work group in [8] guarantees that it can cope with the versatility required for the representation of information needed for building such a system. Finally, the waveform data coming from medical devices, or abstractions derived from this data (e.g., AF episodes lasting 1 minute), could be represented using archetypes, which is important for the design of the PHR in the MobiGuide system.

Looking at the other front-end side, the *conceptual mapping between the CIGs Knowledge Base used by the DSS and the PHR*, we propose to develop openEHR archetypes, designed to comply with the structure of the HL7 vMR classes, specially designed for the goal of supporting CDSS. This provides a straightforward way of linking concepts represented in the CIGs to raw data represented through archetypes so that, by using the HL7 vMR class structure, we guarantee that the knowledge engineers in charge of linking raw data between both components will be familiar with the target structure, arranged in a very natural and comprehensible way. It is important to realize that the process of exporting data from the EMRs to the openEHR vMR–based PHR would be a realizable process, also highly reusable for new customers. First, as commented before, the mapping of data types between the openEHR and the HL7 RIM has been proved to be possible (see Section 3). Second, the export of data from HL7 messages to the vMR will be feasible, thanks to the guideline that is being developed by the HL7 working group in charge of the vMR standard. Third, the result of the previous export process is straightforwardly integrated into the corresponding openEHR archetypes, given that the mapping of data types is possible, as stated before, and the structure of the archetypes is the same (vMR classes).

Finally, analyzing the back-end requirements, using archetypes has many advantages. First, they can be easily connected to any medical vocabulary needed. Furthermore, an openEHR infrastructure can provide a very powerful query language (AQL), where data values can be retrieved to feed and support DSS in a smarter way than using XML-based query languages, making it easier to build more complex queries. From a pragmatic viewpoint for the MobiGuide project, the openEHR middleware available from one of the partners provides full audit trail, versioning and authorization following the ISO/CEN 13606 norm. It also provides SOAP and RESTful web services, very convenient in integration scenarios. What's more, for the case of future needs of interoperability of the PHR with external systems, export and import of openEHR data to/from other standards like HL7 messages is possible, and can be simplified by means of open source integration frameworks (e.g., Mirth Connect,

http://www.mirthcorp.com/products/mirth-connect). This improves the possibilities of integrating the DSS recommendations back into the hospitals EMRs. Finally, thinking on a world-wide view, the alignment of different norms like HL7, openEHR, and EN13606 is a very interesting objective of data integration research. We think that the proposal of using openEHR on the back-end, (using an ISO/CEN 13606 compliant component), and the HL7 vMR on the side of the EMRs' interfaces, could be very promising since it would demonstrate how all three initiatives could be integrated to pursue a common direction for world-wide interoperability for CDSS.

5   Related Work

Several projects have focused on the goal of data integration for supporting CDSS. One of the pioneering projects was the Virtual Medical Record developed originally in 2001 [14], which is the base of the HL7 vMR standard (see Subsection 2.1). It was aimed to ease the process of mapping guideline patient data items evaluated by CDSS, allowing decision, eligibility criteria and patient states to be defined in guideline models, by referencing to the vMR rather than to specific EMRs. Projects such as KDOM [5] or openCDS (www.opencds.org) also use a vMR model to achieve their goals. The purpose of the KDOM framework [5] is to allow specifying CIGs that refer to clinical abstractions, while using the framework to map the abstractions to the schema used by different EMRs. These mappings are defined as instances of an ontology of abstract mapping classes, so that SQL queries can be automatically generated from these instances. To reduce the effort of mapping a CIG to several, a vMR is defined and used as common data model for the EMRs. The purpose of the openCDS project (www.opencds.org) is to develop a SOA-based architecture for the integration of a CDSS with a common information model. This model is the HL7 vMR standard, and openCDS is actually its reference implementation.

Other projects have focused their interest in utilizing archetypes for the integration of patient data. LinkEHR [13] is a tool developed for editing archetypes, using different RMs. It is also able to automatically generate XQuery transformations from mapping functions, in order to link archetypes to existing EMR Schema. Marcos et al. [4] presented an archetype-based integration of CIGs and EMRs by using the LinkEHR tool and transformations, using existing openEHR archetypes that needed to be previously adapted to the concepts found in the CIGs. Chen et al. [15] presented a way to use archetypes and logic rules, expressed using CLIPS, for the definition of CIGs. They proposed to access the patient data by integrating AQL queries into the expressions evaluated by the CLIPS rules engine. Note that our proposal tries for the first time to integrate both approaches (archetypes on top of the HL7 vMR model), in order to take advantage of best features offered by both mechanisms.

6   Discussion and Conclusions

It is usually stressed that archetypes are a good mechanism for representing clinical concepts, and they have been usually reported in the literature as a means to specify

concepts in specific clinical domains. These archetypes can be later combined by using templates, and they have been utilized for designing new EMRs, improving the easy generation of user interfaces, or supporting the migration of proprietary EMRs, as shown in [16]. We agree that archetypes provide a powerful and flexible mechanism when representing clinical information, and in this paper we have presented other significant advantages of using the dual modeling approach and the ISO/CEN 13606 norm. However, we think that the way they are designed and used would need to be re-evaluated when thinking on developing a **domain-independent** guideline-based CDSS, as in the case of MobiGuide. For such a task, where specific raw data needs to be mapped between the CIG Knowledge Base and the PHR archetypes entries, we think that using a predefined set of static classes that are easy to understand (like the ones provided by the HL7 vMR standard) by the knowledge engineers would be the best solution. This way, the target structure used to link different guidelines concepts to raw data is always the same, and so the design of the PHR is not affected when porting the system to different clinical domains or customers' organizations.

The integration of patient data for the successful deployment of CDSS needs to advance in order to be done in a standardized, useful, user-friendly, and distributed manner, in order to simplify the data exchange and data representation between all the components involved in such a system. This is crucial when having different sources of patient information that should be integrated into an interconnected PHR [6].

Therefore, we have presented an approach to cope with such complexity. After reviewing several standards that could be used for data representation and evaluating them against a set of relevant criteria, we selected a solution that meets all criteria examined. In this proposal, the same set of openEHR archetypes can be used as a back-end and front-end to the PHR, by conforming to the structure of the HL7 vMR classes, as described in the experiments of Section 2.3. By using this mechanism we aim to propose a design for PHRs sustainable through time, usable and portable to different domains, and which is supported on standards that have been properly designed for decision support [17]. At the same time we take advantage of powerful features provided by the two-level modeling approach, described through the paper.

The next stage of the project is committed to design the PHR considering more examples of real data needed in the scenarios mentioned (GDM and AF), further checking the viability of this proposal. We also plan to design a new interface for the KDOM tool [5] in order to follow a SOA approach, connecting to the PHR SOAP-based interfaces that the openEHR middleware provides.

References

1. Sim, I., Gorman, P., Greenes, R.A., Haynes, R.B., Kaplan, B., Lehmann, H., Tang, P.C.: Clinical decision support systems for the practice of evidence-based medicine. Journal of the American Medical Informatics Association 8, 527-534 (2001)

2. Ahmadian, L., van Engen-Verheul, M., Bakhshi-Raiez, F., Peek, N., Cornet, R., de Keizer, N.F.: The role of standardized data and terminological systems in computerized clinical decision support systems: Literature review and survey. International Journal of Medical Informatics 80, 81-93 (2011)

3. Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., Shvo, A.S.: HL7 clinical document architecture, release 2. Journal of the American Medical Informatics Association 13, 30-39 (2006)

4. Marcos, M., Maldonado, J., Martínez-Salvador, B., Moner, D., Boscá, D., Robles, M.: An archetype-based solution for the interoperability of computerised guidelines and electronic health records. Artificial Intelligence in Medicine, LNCS, vol. 6747, pp. 276-285. Springer (2011)

5. Peleg, M., Keren, S., Denekamp, Y.: Mapping computerized clinical guidelines to electronic medical records: Knowledge-data ontological mapper (KDOM). Journal of Biomedical Informatics 41, 180-201 (2008)

6. Detmer, D., Bloomrosen, M., Raymond, B., Tang, P.: Integrated personal health records: transformative tools for consumer-centric care. BMC Medical Informatics and Decision Making 8, 45 (2008)

7. Shahar, Y.: The "Human Cli-Knowme" Project: Building a Universal, Formal, Procedural and Declarative Clinical Knowledge Base, for the Automation of Therapy and Research. KR4HC Workshop, LNCS, vol. 6924, pp. 1-22, Springer (2012)

8. Kawamoto, K., Del Fiol, G., Strasberg, H.R., Hulse, N., Curtis, C., Cimino, J.J., Rocha, B.H., Maviglia, S., Fry, E., Scherpbier, H.J.: Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard. In: AMIA Symposium, pp. 377-381. (2010)

9. Garde, S., Knaup, P., Hovenga, E.J.S., Heard, S.: Towards Semantic Interoperability for Electronic Health Records--Domain Knowledge Governance for open EHR Archetypes. Methods of Information in Medicine 46, 332-343 (2007)

10. Muñoz, P., Trigo, J.D., Martínez, I., Muñoz, A., Escayola, J., García, J.: The ISO/EN 13606 Standard for the Interoperable Exchange of Electronic Health Records. Journal of Healthcare Engineering 2, 1-24 (2011)

11. Schloeffel, P., Beale, T., Hayworth, G., Heard, S., Leslie, H.: The relationship between CEN 13606, HL7, and openEHR. In: Proceedings of HIC pp. 1-4. HISA, (2006)

12. Eichelberg, M., Aden, T., Riesmeier, J., Dogac, A., Laleci, G.B.: A survey and analysis of Electronic Healthcare Record standards. ACM Computing Surveys 37, 277-315 (2005)

13. Maldonado, J.A., Moner, D., Boscá, D., Fernández-Breis, J.T., Angulo, C., Robles, M.: LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. International Journal of Medical Informatics 78, 559-570 (2009)

14. Johnson, P.D., Tu, S.W., Musen, M., Purves, I.: A virtual medical record for guideline-based decision support. In: Proceedings of AMIA Symposium, pp. 294-298. (2001)

15. Chen, R., Georgii-Hemming, P., Åhlfeldt, H.: Representing a chemotherapy guideline using openEHR and rules. Stud Health Tech Informat, vol. 150, 653-657, IOS (2009)

16. Chen, R., Klein, G., Sundvall, E., Karlsson, D., Åhlfeldt, H.: Archetype-based conversion of EHR content models: pilot experience with a regional EHR system. BMC Medical Informatics and Decision Making 9, 33 (2009)

17. Kawamoto, K., Del Fiol, G., Lobach, D.F., Jenders, R.A.: Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress. The Open Medical Informatics Journal 4, 235 (2010)