An archetype-based solution for the interoperability of computerised guidelines and electronic health records

Mar Marcos[1*], Jose A. Maldonado[2**], Begoña Martínez-Salvador[1], David Moner[2], Diego Boscá[2], and Montserrat Robles[2]

[1] Universitat Jaume I, Castellón, Spain,
[2] ITACA Institute, Universidad Politécnica de Valencia, Spain

Abstract. Clinical guidelines contain recommendations based on the best empirical evidence available at the moment. There is a wide consensus about the benefits of guidelines and about the fact that they should be deployed through clinical information systems, making them available during consultation time. However, one of the main obstacles to this integration is still the interaction with the electronic health record. In this paper we present an archetype-based approach to solve the interoperability problems of guideline systems, as well as to enable guideline sharing. We also describe the knowledge requirements for the development of archetype-enabled guideline systems, and then focus on the development of appropriate guideline archetypes and on the connection of these archetypes to the target electronic health record.

1 Introduction

Clinical guidelines contain recommendations about different aspects of clinical practice in relation to a specific clinical condition. These recommendations are based on the best empirical evidence available at the moment. For this reason, the use of guidelines has been promoted as a means to control variations in care, reduce inappropriate interventions and deliver more cost-effective care, among other things. Despite some discrepancies, there is a wide consensus on the benefits of guidelines. There is also consensus about the fact that guidelines should be deployed through clinical information systems, and that they should be made available to medical professionals during consultation time [1]. Current guideline systems include reminder systems as well as increasingly more complex systems representing significant parts of guideline procedural knowledge. Whatever is the case, there must be some interaction with the clinical information system in

* This research has been supported by Universitat Jaume I through project P11B2009-38, and by the Spanish Ministry of Education through grant PR2010-0279.
** This research has been supported by the Spanish Ministry of Science and Innovation under Grant TIN2010-21388-C02-01, and by the Health Institute Carlos III through the RETICS Combiomed, RD07/0067/2001.

general, and with the electronic health record (EHR) in particular, to obtain all the necessary information.

Over the last decade, clinical guidelines have been one focus of research in the areas of Artificial Intelligence and Medical Informatics. Significant contributions in these areas include a variety of languages for the representation of guidelines (see [2], [3] for a review). Recently the focus of attention is shifting from the representation of guidelines to the integration of guideline systems in realistic healthcare settings [4]. Despite these efforts, the interaction with the EHR remains as one of the main obstacles for the interoperability of guideline systems within clinical information systems [5]. An important problem is the lack of standardisation as regards data. There have been several initiatives, involving standardization bodies, to define a generic EHR architecture for the communication of health data. However, its use is not directly supported in the current guideline representation languages.

One of the main contributions of recent EHR architectures is the dual model methodology [6] 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 and stable properties of health record information (such as folder, document, section, and audit). The generality of the reference model is complemented by the particularity of archetypes. Archetypes are detailed, reusable and domain-specific definitions of clinical concepts (such as Apgar score, discharge report, and primary care EHR) in the form of structured and constrained combinations of the entities of the reference model. Their principal purpose 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 reference model.

In a previous paper we propose the utilisation of openEHR archetypes in the framework of guideline representation languages [7]. We view computerised guidelines (CGs) as objects with archetype-enabled fragments in specific points where interactions should occur, e.g. patient data queries and/or physician order generation. In this paper we take a further step and present an archetype-based approach to make the interoperability of CGs and EHRs possible. We also describe the knowledge requirements for the development of archetype-enabled CGs, and then focus on the development of a set of CG archetypes and on the connection of these archetypes to the EHR. The latter steps are illustrated with examples from a guideline for chronic heart failure.

2 Approach

We are concerned with the use of archetypes (openEHR or other) within CGs as a mechanism for the interaction with the EHR, and also as a way to make possible the shared use of CGs. In our view, shared use (and reuse) of CGs is a crucial issue because guideline recommendations are valid in a more or less wide scope (national or international) which implies that usually they have to be implemented in different healthcare institutions, possibly with some adaptations. Our interest in reuse is not limited to CG procedures but also covers the models of relevant clinical concepts in the CG.

We are also concerned with technical solutions to implement this approach. Technical implementation requires a platform for the access to the EHR data via archetypes, in the likely case the EHR system does not support archetypes natively. Additionally, a software for the execution of CGs (CG engine) supporting the use of archetypes is required. For the former, we plan to use the data integration engine of the LinkEHR Normalization Platform [8] (see section 4 for more details). With respect to the CG engine, the programming work to adapt an existing engine so that it supports data access via archetypes is underway.

In addition, the view of archetype-enabled CGs has several requirements as regards knowledge modelling:

1. it is necessary to design a collection of archetypes suitable for the decision tasks carried out in the guideline.
2. it is necessary to ensure that the guideline model (CG) is compliant with these archetypes.
3. it must be ensured that the connection with the target EHR (or clinical databases) via the designed archetypes is feasible.

With respect to (1.), to increase the chances of reuse it is important that the guideline archetypes are designed considering the available archetypes and standards. Requirement (2.) is also crucial since CGs are often modelled without regard to the interaction with the EHR, which hinders interoperability. Here again, CGs should be modelled taking into account EHR standards and available archetypes all along. Finally, requirement (3.) involves the definition of a series of mappings relating the archetyped concepts from (1.) to the corresponding data items in the target EHR. Because guidelines often operate on data abstracted from lower-level EHR data, these mappings may relate one archetype to several data items (or even archetypes), e.g. by means of logical abstractions. The rest of the paper is mainly devoted to the requirements (1.) and (3.).

Since the seminal work on the Arden syntax [9], different authors have sought the integration of EHR systems with decision support systems in general and with guideline systems in particular. The KDOM framework [10] and the MEIDA architecture [11] are remarkable examples of recent work dealing with this problem using standards. Overall, our approach is similar to the ideas of these latter platforms. A distinctive feature of our work lies in the utilisation of the full-fledged archetype framework instead of the simplified versions of HL7 RIM used in MEIDA and KDOM, for the definition of the virtual health record (VHR).

From a data model perspective, the utilisation of archetypes brings about several advantages over previous initiatives. First, the VHR/data models we work with are of a higher level (clinical concept level instead of reference model one) and can contain semantic descriptions (by means of terminology references). Second, archetypes allow dealing with VHRs based on different EHR architectures (e.g. CEN/ISO EN13606, openEHR or HL7 CDA), due to their deliberate independence of the reference model. Third, no matter what VHR is used, data access via archetypes operates in such a way that data instances of the VHR are actual instances of the underlying reference model. Last but not least, clinicians are the main actors in the development of archetype models, which ensures both the medical and the technical validity thereof.

From the perspective of integration architecture, the use of archetypes in theory makes possible the direct access to EHR data without any kind of wrapper mechanism, provided that the EHR system natively supports archetypes and uses archetype models compatible with the CG ones (which ensures e.g. that the value for systolic blood pressure will be always located in the same path within the blood pressure archetype, be it specialised or not).

3 Knowledge resources

A guideline for chronic heart failure We have worked with the guideline for the diagnosis and treatment of chronic heart failure (CHF) developed by the European Society of Cardiology (ESC) [12]. According to the ESC, there are at least 10 million patients with heart failure in the countries it represents. The prognosis of heart failure is poor, hence the importance of a correct patient management. The guideline had been previously modelled in the PROforma guideline representation language.

The ESC CHF guideline is a 26 page document containing recommendations for the diagnosis, assessment, and treatment of CHF, for use in clinical practice. An evidence-based approach has been applied in the elaboration of the guideline, except for the diagnosis part, which is based on consensus. The guideline has text format but nevertheless contains several explanatory figures and tables. A section on descriptive terms in heart failure as well as a few tables with definitions have been particularly useful for the purposes of our work.

The openEHR Clinical Knowledge Manager We have used as archetype source the openEHR Clinical Knowledge Manager[3] (CKM), which is a web-based repository allowing for archetype search, browse and download. Archetypes in the CKM have been created by independent domain experts, mainly clinicians and computer scientists, and then they have been released to the community as open source and freely available content. Before publication, archetypes undergo an iterative review process to ensure that they cover as many use-cases as possible and thus constitute a sensible maximal data set (with a high reuse potential).

According to openEHR the main categories for the description of clinical concepts are observation, evaluation, instruction and action. This categorisation is related to the way in which information is created during the care process: an *observation* is created by an act of observation, measurement, or testing; an *evaluation* is obtained by inference from observations, using personal experience and/or published knowledge; an *instruction* is an evaluation-based instruction to

[3] See `http://www.openehr.org/knowledge/`.

be performed by healthcare agents; and an *action* is a record of the interventions that have occurred, instruction-related or not. The number and specificity of available archetypes differs significantly among and within categories. E.g. within observation, there exist very specific archetypes such as *Apgar score* and *body weight*, while other archetypes like *imaging test* are rather generic.

4 Software tools and methods

The LinkEHR platform The LinkEHR platform The LinkEHR Normalization Platform[4] is a set of tools and modules that allow: i) the creation of an archetype-based customizable view over a set of heterogeneous and distributed EHR data sources [8]; ii) the editing of archetypes based on different reference models (standards) as long as an XML Schema is available [13] (several reference models have been tested successfully: EN13606, openEHR, HL7 CDA, CDISC and CCR); and iii) the specification of declarative mappings between archetypes and data sources, and from these mappings, the automatic generation of XQuery scripts which translate source XML data into XML documents that are archetype compliant.

LinkEHR employs archetypes for both the semantic description of legacy EHRs and the publication of existing clinical information in the form of standardized EHR extracts. Since health data reside in the underlying EHR systems, it is necessary to define some kind of mapping information that links entities in the archetype to data elements in data repositories (e.g. elements and attributes in the case of XML sources). Basically, these mappings specify how to create archetype instances from the content of the data sources.

Mapping methods in LinkEHR At the schema level, the above mentioned mappings require an explicit representation of how the source schema (legacy EHR data schema) and target schema (archetype) are related to each other. The effort required to create and manage such mappings is considerable. The common case is to write intricate and non-reusable software to perform the required transformations. This is even more complex with archetypes, since they are used to model generic concepts without any consideration regarding the internal architecture of the EHR. LinkEHR allows the definition high-level non-procedural mapping specifications that consist in a set of correspondences between entities of archetypes and source schemas. Two types of correspondences are supported: value and structural correspondences. The former specify how to calculate atomic values, whereas the latter may be used to control the generation and grouping of elements in the target.

Value correspondences are defined by a set of pairs, consisting of a transformation function that specifies how to calculate a value in the target from a set of source values, and a filter indicating the conditions that source data must satisfy to be used in the transformation function. The simplest kind of transformation function is the identity function which copies a source value into a target value.

[4] See `http://www.linkehr.com`.

But quite often it is necessary to specify arbitrary complex functions. For this purpose, the tool comes with a variety of functions such as type conversion, and mathematical, logical, string, date and time functions. The example in Table 1 illustrates a simple correspondence transforming gender codes. It transforms the local gender code in the path `/patient/gender` of an XML EHR extract (source data) into a normalized code (to be stored in the target data). Note that the order is relevant and only the first applicable function is used.

Table 1. A mapping transforming the gender codes from an XML source.

Filter	Function
/patient/gender='M' OR /patient/gender='m'	0
/patient/gender='W' OR /patient/gender='w'	1
/patient/gender=0 OR /patient/gender=1	/patient/gender
true	9

5 Archetypes for guideline interoperability

Identification of clinical concepts from the guideline As a first step, we have identified the clinical concepts necessary to support CHF management and hence requiring archetypes. We have reviewed the ESC guideline with the help of medical experts, who had previously highlighted in the text all the relevant terms, including clinical concepts, tests, interventions, etc. Then we have analysed the identified concepts, using a mind map as a tool, making explicit the relationships among them.

The guideline sometimes refers to rather high-level/abstract concepts. This has been depicted in the mind map by means of abstraction relationships linking the high-level concept to the lower-level EHR ones from which it can be obtained. An example is the concept *ACEI intolerant*[5], which is based on the concepts *cough*, *hypotension*, *renal insufficiency*, *hyperkalaemia*, *syncope*, and *angioedema*. In a few cases there is no definition for the abstract concept in the guideline. Here we have sought additional information to specify the abstraction relationship. As a result the mind map not only includes the concepts in the guideline but also related concepts necessary for the definition thereof. A list of the concepts used in the ESC guideline, together with indications of the most suitable CKM archetypes to store the necessary information, can be found in a previous paper [7].

Design of archetypes from the clinical concepts From the clinical concepts identified in the previous step we have developed a set of archetypes for use in a CG for CHF management. We have proceeded in a bottom-up way, starting with the concepts directly related to EHR data. As an illustration, in the rest of the

[5] ACEI stands for angiotensin-converting enzyme (ACE) inhibitors.

section we focus on the archetypes for the concept *cough* and for the more abstract one *ACEI intolerant*. These required the specialisation of generic CKM archetypes, concretely `openEHR-EHR-CLUSTER.symptom.v1` in the case of *cough* and `openEHR-EHR-EVALUATION.adverse.v1` in the case of *ACEI intolerant*.

Due to the generality of CKM archetypes (they are designed to be used in a wide range of scenarios), many entities (attributes or types) may not be relevant to all the potential usages. When defining our set of archetypes the principal task has been the selection of relevant entities and associated terminologies. For instance, in the case of `openEHR-EHR-CLUSTER.symptom-cough` (a specialisation of `openEHR-EHR-CLUSTER.symptom.v1`), we have only retained the entities to hold the information about the type, character, duration and severity of cough. For coding the type of cough, we have chosen to use SNOMED-CT. For coding the cough severity, we have considered the use of a local terminology as defined in the original CKM archetype. Figure 1 shows the overall structure of the archetype `openEHR-EHR-CLUSTER.symptom-cough`.

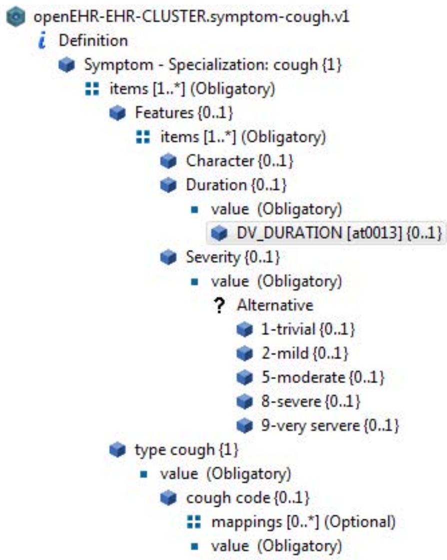

Fig. 1. Overall structure of the archetype `openEHR-EHR-CLUSTER.symptom-cough`.

Mapping of archetypes to a clinical database Unlike archetype definition, mapping specifications are particular to each EHR setting. In the mapping examples we describe next, we assume particular database features. The first mapping examples are related to the cough severity as defined in the archetype `openEHR-EHR-CLUSTER.symptom-cough` (see Severity in Figure 2). The severity is modelled by a `DV_ORDINAL`, which is a data type used to model finite scores where there is an implied ordering. This type contains two attributes: value, that represents the ordinal (position) in the enumeration of values, and symbol, that is the textual representation of the value. See the left hand side of Figure 2 for the definition of a DV_CODED_TEXT. According to the generic archetype, our archetype defines five levels of severity, namely: trivial (value=1), mild (value=2), moderate (value=5), severe (value=8) and very severe (value=9). Let us suppose that the source database offers its data as XML documents, and that it uses a severity scale ranging from 1 to 9 (source path is `/root/severity`) that needs to be mapped to the five-level scale used in the archetype. At least two mappings need to be defined for this scale mapping: one for the attribute holding the textual representation of the local code (mapping **a** in Figure 2), and one for the attribute holding the ordinal (mapping **b** in Figure 2).

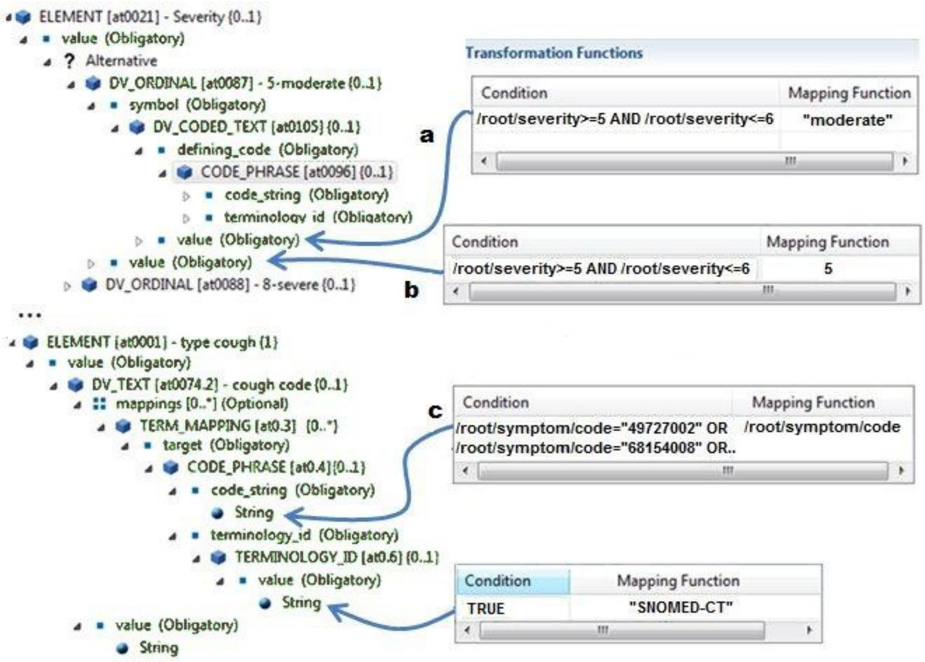

Fig. 2. Mappings for the archetype `openEHR-EHR-CLUSTER.symptom-cough`.

A different kind of mapping is related to the attribute holding the cough coded value. Let us assume that the source database uses a SNOMED-CT coding (source path is `/root/symptom/code`), and that the same coding is employed in the target data. Obviously, only those symptom codes related to cough must be used in the mapping, such as cough (SNOMED-CT::49727002), and chronic cough (SNOMED-CT:: 68154008). This terminological mapping appears in Figure 2 as mapping **c**.

Another possibility is using archetypes as sources, which allows the definition of more abstract concepts. A simple example is described next, which uses the archetype `openEHR-EHR-CLUSTER.symptom-cough` as source. Among other things, a patient that has either severe or very severe cough, including dry cough (SNOMED-CT::11833005), chronic cough (SNOMED-CT::68154008), or non-productive cough (SNOMED-CT::409596002), can be considered ACEI intolerant. The following expression uses the entities in `openEHR-EHR-CLUSTER.symptom-cough` to define a condition to be used as one of the filters in the mapping of `openEHR-EHR-EVALUATION.adverse-ACEI_intolerant`:

```
  --dry, chronic or non-productive cough
(/items[at0001]/value[at0074.2]/mappings[at0.3]/target[at0.4]/code_string=11833005 OR
 /items[at0001]/value[at0074.2]/mappings[at0.3]/target[at0.4]/code_string=68154008 OR
 /items[at0001]/value[at0074.2]/mappings[at0.3]/target[at0.4]/code_string=409596002) AND
  --severe or very severe cough
(/items[at0034]/items[at0021]/value[at0088]/value = 8 OR
/items[at0034]/items[at0021]/value[at0092]/value = 9)
```

6 Conclusions and future work

There is a general consensus about the fact that clinical guidelines should be deployed using some computer support and that this support should be integrated within the clinical information system, to take full advantage of their potential benefits. However, currently one of the main obstacles to this integration is the interaction with the EHR. With the aim of solving the interoperability problems of guideline systems, we propose the utilisation of archetypes as a canonical data representation of the concepts used in a CG. The utilisation of archetypes has several advantages with respect to previous initiatives, among others the possibility of dealing with VHRs described in terms of clinical concepts, possibly based on different EHR architectures.

The approach is built around the development of a collection of guideline-specific archetypes, taking into account the (medically valid) clinical models in existing archetype repositories. It implies the proper utilisation of CG archetypes within the guideline (guideline-archetype adjustment), as well as the specification of a series of mappings relating the archetypes to the corresponding data in the target clinical databases (archetype-database mapping). Despite the overload of developing the necessary archetypes and mappings, the resulting CG procedure and data models should have a high potential for reuse.

Another contribution of our work is the development of a set of archetypes for use in a real-world guideline, including sample archetype-database mappings. For these tasks we have used the LinkEHR data integration platform. Archetype editing with LinkEHR has been straightforward, after identifying the relevant CKM archetypes and determining the necessary adaptations (including links to terminologies). LinkEHR tools are also well suited for the definition of the archetype mappings in our sample guideline (terminological and abstraction ones).

As future work, we will proceed with the technical implementation of our approach, which requires the adaptation of an existing CG engine to support data access using archetypes. Also with the aim of validating the approach, we will tackle the development of archetypes for other clinical guidelines to determine if LinkEHR mapping methods are sufficient or, on the contrary, additional functionalities are required. At a different level, we intend to investigate ontological frameworks for the data model mappings we are dealing with, including conceptual models allowing for reasoning e.g. about how different data models relate.

References

1. Sonnenberg, F., Hagerty, C.: Computer-Interpretable Clinical Practice Guidelines. Where Are We and where Are We Going? IMIA Yearbook of Medical Informatics (2006) 145–158
2. Peleg, M., Miksch, S., Seyfang, A., Bury, J., Ciccarese, P., Fox, J., Greenes, R., Hall, R., Johnson, P., Jones, N., Kumar, A., Quaglini, S., Shortliffe, E., Stefanelli, M.: Comparing computer-interpretable guideline models: A case-study approach. J Am Med Inform Assn **10** (2003) 52–68
3. de Clercq, P., Blom, J., Korsten, H., Hasman, A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med **31** (2004) 1–27
4. Panzarasa, S., Quaglini, S., Cavallini, A., Micieli, G., Marcheselli, S., Stefanelli, M.: Technical Solutions for Integrating Clinical Practice Guidelines with Electronic Patient Records. In: Knowledge Representation for Health-Care. Data, Processes and Guidelines. Volume 5943 of LNAI. Springer-Verlag (February 2010) 141–154
5. Chen, R., Georgii-Hemming, P., Åhlfeldt, H.: Representing a Chemotherapy Guideline Using openEHR and Rules. In: Proc. of the 22nd Int. Conf. of the European Federation for Medical Informatics (MIE-09), IOS Press (2009) 653–657
6. Beale, T., Heard, S.: Architecture Overview. `http://www.openehr.org/releases/1.0.2/architecture/overview.pdf` (April 2007)
7. Marcos, M., Martínez-Salvador, B.: Towards the interoperability of computerised guidelines and electronic health records: an experiment with openEHR archetypes and a chronic heart failure guideline. In: Knowledge Representation for Health-Care. Volume 6512 of LNAI. Springer-Verlag (2011)
8. Angulo, C., Crespo, P., Maldonado, J., Moner, D., Pérez, D., Abad, I., Mandingorra, J., Robles, M.: Non-invasive lightweight integration engine for building EHR from autonomous distributed systems. Int J of Med Inform **76**(S3) (2007) 417–424
9. Hripcsak, G., Ludemann, P., Pryor, T.A., Wigertz, O.B., Clayton, P.D.: Rationale for the Arden Syntax. Comput Biomed Res **27** (1994) 291–324
10. Peleg, M., Keren, S., Denekamp, Y.: Mapping computerized clinical guidelines to electronic medical records: Knowledge-data ontological mapper (KDOM). J Biomed Inform **41**(1) (2008) 180–201
11. German, E., Leibowitz, A., Shahar, Y.: An architecture for linking medical decision-support applications to clinical databases and its evaluation. J Biomed Inform **42**(2) (2009) 203–218
12. The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the ESC: Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005). Eur Heart J **26** (2005) 1115–1140
13. Maldonado, J., Moner, D., Boscá, D., Fernández, J., Angulo, C., Robles, M.: LinkEHR-Ed: A multi-reference model archetype editor based on formal semantics. Int J of Med Inform **78**(8) (2009) 559–570