Reusing openEHR Archetypes for the Expression of Cerebral Palsy Electronic Medical Records

Afef Samet Ellouze¹, Rafik Bouaziz² and Nadia Bouassida³
1. Computer Science, Higher Institute of Technological Studies, Sfax 3001, Tunisia
2. Computer Science, Faculty of Economic Sciences and Management, Sfax University, Sfax 3013, Tunisia
3. Computer Science, Superior Institute of Multimedia Science, Sfax University, Sfax 3011, Tunisia

Corresponding author: Afef Samet Ellouze (samet_afef@yahoo.fr)

Abstract: Meta-modeling of healthcare records is necessary to capitalize and normalize the knowledge of domain experts. First, it enables knowledge interoperability and, it offers reuse of clinical concepts and provides a higher design quality of health information systems and electronic health record systems. However, to obtain these advantages, designers must be able to retrieve the suitable archetypes that cover the clinical concepts involved in their specific applications, and to overcome the difficulties of modeling an electronic health record system based on existing archetypes. In fact, designers often have difficulties to reuse and adapt archetypes provided by open repositories, when modeling new electronic health record systems. Thus, they need special tools assisting them in customizing the appropriate archetypes. This paper focuses on archetypes reuse for modeling the electronic health record of child affected by cerebral palsy. Also, this paper highlights the major difficulties encountered during this process. Then, it evaluates archetypes reuse and presents some perspectives to solve these difficulties.

Key words: Electronic health record, semantic interoperability, archetype reuse, cerebral palsy information management system, dual model approach.

1. Introduction

EHR (electronic health record) can be considered as a repository of information regarding health care in a computer processed form, stored and transmitted securely. It contains retrospective, current, and prospective information in order to support efficient and quality integrated health care [1]. One of the major advantages of an EHR is that the data, being recorded as part of healthcare delivery, present a valuable source of information that can be reused, shared and exchanged to ensure optimal clinical decision support and to coordinate distributed care processes among multiple specialists. However, EHR shared data must be understandable by different actors for all defined domain concepts. In fact, different EHR systems must be able to interoperate, so that one system can understand the context and meaning of information provided by another system, i.e., it is necessary to ensure semantic interoperability.

The advanced standards for EHR representation and communication [2, 3], such as ISO/EN 13606 “Health Informatics-EHR communication” [4] and openEHR [5], propose the use of an architecture based on the dual model approach, carried out the meta-modeling of healthcare records with two conceptual levels: RM (reference model), and archetype model. The first level models the semantics of data in terms of generic data types and structures in a stable form. It specifies (1) the global characteristics of medical record components, (2) how these components are aggregated/composed and (3) the information context necessary to meet the ethical, legal and provenance requirements. These components are defined in the RM as a set of classes that form the generic building
blocks of the EHR [4]. On the other hand, the second level, which is the archetype model, used to formalize clinical concepts and knowledge. It specifies the common features of entity types and, therefore, it defines valid information structures in terms of taxonomic (is a class of) and partonomic (is a part of) components of each clinical concept. Archetypes are usually defined by domain experts, in an attempt to standardize clinical practice.

According to the openEHR standard, in addition to demographic-archetypes class used to model demographic data, several other classes of archetypes are used to constrain the information building blocks defined in the RM:

- **Composition-archetypes**: These thematic archetypes are the containers of all clinical and administrative content of the EHR.
- **Section-archetypes**: These organizational archetypes correspond to document headings.
- **Entry**: The openEHR specification defines several ENTRY subtypes, which are OBSERVATION, EVALUATION, INSTRUCTION and ACTION.
- **Structure-archetypes**: Archetypes to model a structure (e.g., a tree or a list) of items.
- **Cluster-archetypes**: Archetypes to model a cluster of items and to embed this cluster in other archetypes.
- **Element-archetypes**: Archetypes to model a single item with the intention of its reuse.

Based on these archetypes, templates for local use can be defined. Templates combine several archetypes, according to a local context, into larger structures like a screen form, document, report or message. They may add further local constraints on archetypes, including removing or mandating optional sections, and may define default values. Archetypes and templates are used at runtime by the system. Their primary purpose is to validate data during their capture, and to ensure storing data according to the archetype definition [6, 7].

As a formal definition of clinical concepts, archetypes need to be of high technical and clinical quality. Thus, they are the most promising pathways to semantic interoperability of health information systems. For this, the openEHR foundation [5] has established the archetype editorial group whose members collaborate openly on the web using the openEHR CKM (clinical knowledge manager) (http://www.openehr.org/knowledge), created by Ocean Informatics (http://www.oceaninformatics.com) for the openEHR foundation. The web-based review and validation process offered by CKM for archetypes enables international involvement and participates in the construction of archetypes of high clinical and technical quality. However, to design archetypes of such quality, archetype construction is also related to other issues such as versioning, specialization, and composition. Versioning is necessary since the medical domain is a dynamic environment, which has evolving clinical results. Thus, archetype construction approaches must take into account this fact and allow defining and managing multi-version archetypes. For each archetype, the schema versioning must allow creating new schema versions and preserve old ones. For legal reasons, CKM supports discovery of all versions, with only one valid version and some erroneous ones for each archetype [4]. But it does not support archetype versioning management. So, we need new tools to (1) make CKM users able to treat archetype schema changes as a versioning process and not as a simple non destructive updating of archetypes, (2) be able to maintain many valid versions for each archetype, stamped by their valid times, and (3) formulate queries with temporal specifications.

On another hand, archetypes can be defined as extensions or specializations of existing ones [8, 9]. Thus, archetypes management approaches must take into account archetypes reuse and have the ability to capitalize on prior expertise. Finally, some archetypes are defined as parts of the structure of other ones. So mechanisms for the management of these patronymic structures must also be provided.
In this study, we focus on the problem of modeling cerebral palsy EHR, which is a multidisciplinary disease. The relevant information has to be made available at the right time and place in an adequate form to all involved healthcare providers. All shared concepts must be expressed and structured in a reusable form for all the involved specialties to ensure semantic interoperability. Neurology Department in Hedi-Chaker-Sfax-Tunisia University Hospital uses a paper-based system as a current medical record system. This record system consists in a list of documents relative to the involved therapies. Doctors use these documents to record observations about their patients. Doctors involved in cerebral palsy are specialists in
• Neonatology to diagnose child history and circumstances of birth;
• Pediatric neurology to provide primary care and diagnose neurological symptoms of cerebral palsy.
• Orthopedic to predict diagnoses and treat associated muscle, tendon, and bone problems;
• Physical therapy to design and supervise special exercise programs to improve movement and strength;
• Ergothrapy to evaluate and improve infant’s functional independence;
• Speech and language therapy to diagnose and treat communication problems.

We find several EHR implementations related to child affected by cerebral palsy. However, due to the heterogeneity of systems and underlying databases, medical information cannot be easily shared between care professionals, thus limiting the scope and the scale of scientific research that can be carried out and effectiveness of clinical decisions making. But we can find EHR development solutions [10-13] which aim is to make possible shared information, due to the overlapping domains, to the similarity of problems, and to the requirements they need to fulfill. Indeed, these solutions intend to store and communicate medical information efficiently and reliably to ensure faithful interoperability between healthcare providers.

Note that, while there is an agreement of the research community on the fact that the use of the archetype approach helps in supporting a better delivery of care and allows a high degree of semantic interoperability, none of the existing researches is interested in proposing cerebral palsy EHR implementation with the archetype approach. Within this context, this paper proposes to cover all concepts related to cerebral palsy through the reuse of archetypes. For this, we use three openEHR-specific software tools, i.e., openEHR CKM, Ocean Archetype Editor (http://www.oceaninformatics.com/Solutions/ocean-products/Clinical-Modelling/ocean-archetype-editor.html) and Ocean Template Designer (http://www.oceaninformatics.com/Solutions/ocean-products/Clinical-Modelling/ocean-template-designer.html). Moreover, this paper highlights major problems encountered during the modeling process and recommends some perspectives to solve these problems. In fact, EHR designers often don’t know how they can reuse and customize archetypes that are available in open repositories. Issues related to multi-version management of archetypes are not sufficiently supported by CKM. In fact, according to the openEHR specification, CKM ensures the maintaining of historic erroneous versions of archetypes for legal purposes. Each archetype version is provided in a snapshot form and associated to audit information (according to the reference model), without any temporal stamps. Indeed, CKM does not consider archetypes as logical entities evolving through time with different valid versions. Thus, we think that it is advantageous to provide a temporal management process, considering that each version is valid according to a time interval, and tools supporting temporal queries.

On another hand, if common archetype editors, as well as the openEHR ADL Workbench\(^1\), provide tools to specialize and to extend archetypes, they do not assist

the designer: which parts could be specialized, extended or enhanced. They just allow changing archetypes. We think that it could be interesting to provide an interactive assistance, giving the user suggestions when reusing and adapting archetypes, during the various steps of the modeling process.

2. Methods

Modeling cerebral palsy EHR with the dual-model approach consists in identifying and constraining the EHR components with archetypes. For this purpose, the information elements, i.e., the items to be documented, are captured and analyzed according to their structure and context. For this, items are collected from the paper-based patient record available system. This first step has required a rigorous observation of the progress of each child affected, with cerebral palsy clinical exams. Interviews with the involved therapists clarified the meaning of each item and helped us to determine sharable concepts between several cerebral palsy therapies. In fact, a deep understanding of the clinical terminology and data collected from the several palsy specialties is necessary to define the clinical concepts that will drive the design of the openEHR archetypes.

Documented item categories and their value ranges have to be ascertained from existing paper-based documentation system. Problems of synonymies, i.e., different item names assigned to the same data in the involved specialties, and polysemies, i.e., the same name identifying different items, may exist and have to be resolved. With regard to their content, the current documentation system includes a set of items used by several cerebral palsy therapies. These items correspond to:

- Information detailing the course of pregnancy and the circumstances of birth;
- Clinically detailing family history information, i.e., past and present family medical history;
- Information about patient identification.

The paper-based patient record system contains also items that can be shared by some specialties. For example, the set of information related to speech therapy is used both by the neurology and the speech therapies.

The second step has consisted in structuring and combining the identified items under distinct and non-overlapping concepts. The quality of archetypes to identify depends strongly on the quality of this structuring process. In fact, this step requires an harmonization process that gathers items into concepts. Thus, concepts and their items are structured hierarchically and the structure should be visualized, for example, in a mind map. During this process, multiple modeling of items needed in several archetypes can be avoided. The harmonization process helps to have an idea of how archetypes are generally designed, in an attempt to make each one cover a distinct concept in a generic way so that the archetype can be reused in different contexts [14].

As part of this second step, concepts involved in several therapies are identified to discover shareable and reused archetypes between these therapies. For example, a generic archetype modeling pregnancy can be reused in all implicated cerebral palsy therapies. Thus, it is necessary to identify therapists’ viewpoint for each concept in order to determine subsequently the behavior of the related archetype for each specialty. It is also imperative to determine interactions and relationships between concepts to facilitate the selection of the suitable archetypes. Fig. 1 illustrates items to be documented arranged into concepts belonging to the different involved by cerebral palsy; each concept regroups a list of items that can be modeled by one archetype class. The purple box illustrates concepts sharable by all specialties. The green box illustrates neuropediatric and neonatology sharable concepts. The blue box illustrates neuropediatric and ergotherapy sharable concepts. The yellow box illustrates neuropediatric and speech therapy sharable concepts. The pink box illustrates ergotherapy and othopedics sharable concepts.
The third step is reserved for mapping the identified concepts into archetypes. The adopted strategy is to reuse as much as possible existing archetypes and to eventually develop new ones if none is available. This avoids the development of overlapping archetypes that jeopardize the cerebral palsy EHR interoperability. This step is based on the principle that a single archetype should represent a single clinical concept and that archetypes need to be designed in a generic and reusable way because they have to be used in different contexts to enable interoperability and thus the sharing of EHRs [15].

The openEHR CKM tool is used for the discovery of the appropriate archetypes. This tool provides excellent searching facilities to identify required archetypes based on a number of possible criteria, the most useful of which are to search by resource name or by clinical domain.

During this step, archetypes are combined, grouped and arranged using archetype slots. For example, the items of the concept “circumstance of birth” can be modeled within the Delivery cluster-archetype. Next, they can be included in the Pregnancy entry-evaluation-archetype, incorporated in the Diagnostic Report section-archetype and grouped in the Problem list composition-archetype.

Each archetype found should represent a complete clinical knowledge concept, thus covering the maximum possible data set [16]. Besides, when an archetype is found to cover only some fields of the regrouped concept, the repositories are explored again for further archetypes that covered these fields. However, if no other archetype could be found to model the information in these fields, the first one is extended to contain the missing items. If a found archetype covers a more general concept than one needed, this archetype is specialized to the more specific concept needed. Finally, if no appropriate archetype is discovered, a suitable structure for a new archetype has to be defined. This includes the decision if a clinical concept is an observation, an evaluation, an instruction or an action.

Therefore, it is possible that some archetypes must be extended, using an archetype editor, by adding data items of the appropriate type, binding data items to internal codes or terminologies like SNOMED CT [17]. Interviews with the involved therapists are required during this step to represent in a common way the new items. Once the set of archetypes that covers all the concepts is defined, a mapping is created between these concepts and the set of archetypes. Finally, archetypes are used to define templates in order to adapt these archetypes to the specific requirements of the cerebral palsy EHR system.

Special editing tools are available to create and maintain archetypes and templates. Ocean Informatics is a key openEHR services provider and partner in the openEHR Foundation (Ocean Informatics 2010). The services provide both an archetype editor and a template editor for this purpose. Editing tools also allow archetypes incorporating links or bindings to external clinical terminologies such as SNOMED which may be held in a separate terminology data store. This allows consistent use of clinical terminology to be applied to any developed archetypes and removes ambiguity in relation to use of clinical terms. Note that archetypes used to cover CP are
reviewed by all specialists involved by this disease, in Hedi-Chaker-Sfax-Tunisia University Hospital and there is a consensus on all items incorporated to these archetypes. Fig. 2 illustrates the process which models the cerebral palsy EHR.

3. Results and Discussion

Data collected from the several cerebral palsy therapies are modeled with archetype classes according to the openEHR and ISO/13606 standards. A list of 38 archetypes that fully specify all cerebral palsy therapies concepts are identified (Table 1). The hierarchical structure of the identified archetypes, related to the neurological folder class, is illustrated in Fig. 3. Six organizational FOLDER classes are defined to organize data within cerebral palsy EHR by means of the involved therapies. Three composition archetypes are needed to specify sharable concepts between all the involved therapies, which are HISTORY, REPORT and ENCOUNTER. The PROBLEM-LIST composition archetype is also used in the neonatology therapy. Table 1 illustrates all the archetypes used to specify the cerebral palsy concepts and therefore to model the related EHR. On another hand, all used archetypes are discovered using the openEHR CKM, thereby reinforcing archetypes contribution on semantic interoperability.

We note that the archetypes discovered cover all concepts needed by the cerebral palsy disease. Thus, we do not need to define new archetypes, consolidating then the fact of semantic interoperability through the reuse of the available archetypes. Indeed, this indicates that the principle of reuse of archetypes is proved and that archetypes developed to represent clinical concepts in EHRs can generally be reused to represent clinical concepts in the cerebral palsy disease.

The CHECK_LIST entry evaluation archetype is specialized on three archetypes which are Check_list-Epilepsy, Check_list-Cerebral_Palsy and check_list-reflex.

During the step of mapping concepts into archetypes, these latter are combined, grouped and arranged using archetype slots. As illustrated in Fig. 3, each archetype composition regroups a slot of archetypes that can be modeled by one archetype class. The folder classes, which are in gray, group the composition classes which are in mauve, and the section classes, which are colored in green, are decomposed into entry classes.

Table 2 describes our interventions to adapt the archetypes existing in the open repositories, through extension. Many attempts to model EHRs with archetypes exist. Buck et al. [11] propose the modeling of an electronic patient record prototype for premature

![Fig. 2 Synthesis of archetype process modeling.](image-url)
Table 1  List of archetypes used to specify the cerebral palsy concepts.

<table>
<thead>
<tr>
<th>Class</th>
<th>Archetype</th>
<th>Type of uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition</td>
<td>Encounter</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Problem_list</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Report</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Diagnostic_report</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>SOAP</td>
<td>Extension</td>
</tr>
<tr>
<td>Section</td>
<td>Clinical_synopsis</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Pregnancy</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Reason_for_encounter</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>Risk_family_history</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>check_list</td>
<td>Specialization</td>
</tr>
<tr>
<td>Entry.Evaluation</td>
<td>Apgar</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>body_weight</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>Exam</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Story</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>fetal_movement</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>heart_rate</td>
<td>Reuse</td>
</tr>
<tr>
<td>Entry.Observation</td>
<td>Admission</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>Person_name</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Telecom_details</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Free_text</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Gait</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Move</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>Individual_professional</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Move-spine</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Move-joint</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Ambient_oxygen</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Balance</td>
<td>Reuse</td>
</tr>
<tr>
<td></td>
<td>Exam-generic</td>
<td>Extension</td>
</tr>
<tr>
<td></td>
<td>Dimensions</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Imaging</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Professional_role</td>
<td>Reuse</td>
</tr>
<tr>
<td>Cluster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEMOGRAPHIC-PERSON</td>
<td>Person</td>
<td>Specialization</td>
</tr>
<tr>
<td></td>
<td>Patient</td>
<td>Reuse</td>
</tr>
<tr>
<td>Demographic-ORGANIZATION</td>
<td>Organization</td>
<td>Reuse</td>
</tr>
</tbody>
</table>

Fig. 3  Hierarchical structure of archetypes related to the neurological folder class.
infants and newborns, established in Heidelberg university hospital, using openEHR archetypes. The involved concepts were analyzed and mapped into archetypes. A set of templates were designed according to the local context. However, in the model proposed we note two insufficiencies. Firstly, no steps were provided for extending existing archetypes, except by Buck [11]. Secondly, the case where the items of a certain concept might be distributed among two or more archetypes is ignored. If this case occurs, the respective concept should be split into two or more concepts and must be combined with other items.

Garde et al. [13] analyze the feasibility and usefulness of expressing clinical data sets (CDSs) as openEHR archetypes. They present an approach to transform the pediatric oncology CDS into archetypes and resume the problems identified with CDS and their corresponding solutions based on the openEHR approach. Indeed, they approve that archetypes offer a predefined structure for expressing clinical knowledge, thus providing CDS developers with the necessary structure to build their data sets in a uniform way adhering to basic design principles. Moreover, these authors note that to overcome overlapping and inconsistent concepts, systematic domain knowledge governance is required. Nevertheless, they report that insufficient support for hierarchical archetypes, which could flexibly reuse other existing archetypes, is the main problem encountered during the development of their archetypes. Indeed, no tooling support is available. Also binding archetypes to a terminology that supports post-coordination, like SNOMED-CT, can cause some challenges.

Atalag [18] describes an implementation of a gastrointestinal endoscopy health information system by using archetypes and multi-level modeling. He has developed a research prototype application to capture initial requirements and to get user feedback for better usability. However, its approach does not provide any support tool or method for the selection of the appropriate archetypes.

In Ref. [10], the archetype approach was applied to model the information of a prostate cancer biobank database. The authors showed that the majority of information in the relevant database was represented with existing archetypes. Moreover, some archetypes need to be extended and modified.

On another hand, we note that the versioning is an important aspect that must be handled by archetype approaches. In fact, the medical domain is an evolving domain where there are many versions of clinical data. Moreover, evolutions in time of schema of these records, as well as interpretation of temporal aspects in medical information, must be handled correctly. Thus,
for each archetype, there must be approaches for schema versioning that are able to create a new schema version and preserve old ones. This task is a challenging one since the archetype documents are characterized by a high degree of evolution in the time, not only in their data, but also in their structures.

In our study, we apply the same strategy that consists in reusing existing suitable archetypes and in adapting, through specialization and/or extension mechanisms, existing archetypes that do not conform to the specified concepts. The process of archetypes adaptation needs to be generic to ensure reusability of archetypes in different contexts enabling thus semantic interoperability and sharing of cerebral palsy EHR. However, replicated documentation of the same data and the problem of synonymies are the major obstacles during this first step.

Note that it would have been possible to model the CP data by other approaches like the HL7 clinical document architecture CDA [19]. However, the openEHR approach is chosen because of (1) the easy availability and useful of archetypes deposed in open repositories, (2) the complex maintenance of HL7 models.

Due to the lack of tools to adapt and reuse discovered archetypes, the open repositories are searched using only strings search function or through manual inspection of the ADL files. Other problems can be attributed to the presence of overlapping archetypes and to the lack of support for the selection of the appropriate archetypes. Indeed, the available selection tools do not contain semantic comparisons functions that ensure efficiency of the search process.

Note that, the archetype approach lacks modeling guidelines. Thus, the process of identifying and modeling suitable archetypes turned out to be manual and time-consuming, requiring automated tools to ensure coordinated archetype development and maintenance process. Thus, good tool support is imperative to promote participation of non-technical experts in the design of high-quality archetypes.

4. Conclusions

The widespread adoption of standard archetypes is a basis for capturing and reusing the best practices of clinical knowledge. In this work, we identified and analyzed data relative to the cerebral palsy disease. Then, open archetype repositories were explored, in order, to find the suitable archetypes. Note that, 21 existing archetypes were reused without change, 10 archetypes were extended and seven others were specialized. Through this application, we noticed a lack of assistance tools guiding the adaptation of the appropriate archetypes.

Our future works will focus on two research axes, firstly we will propose assistance and guidance for the development and reuse of archetypes, and secondly we will define approach for archetypes versioning. In fact, preserving complete therapeutic history of the patient and defining electronic medical records by archetypes represent a critical requirement for the modern medical information system. In addition to retrieval and browsing capabilities, the preservation of medical records history evolution also requires the ability of managing of schema versions related to such documents. Actually, these requirements are not sufficiently met and wait for the development of new enabling technology. Therefore we must define a novel approach to the management of EHR archives modeled by archetypes. This approach must be able to treat archetypes schema changes as a versioning process instead of a simple evolution of these archetypes. It must also ensure the consistency of the database since when a new archetype schema version is defined, it does not convert its previous schema versions applied in the EHR and does not revalidate the previous EHR which are valid to their archetypes schema versions.

Acknowledgments

The authors would like to thank Professor Chahinez Triki specialists in the Neurology Department in Hedi-Chaker-Sfax-Tunisia University Hospital for
heir valuable intervention in the process modeling of cerebral palsy EHR.

References