

# Dealing with the Archetypes Development Process for a Regional EHR System

M.R. Santos<sup>1</sup>; M.P. Bax<sup>1</sup>; D. Kalra<sup>2</sup>

<sup>1</sup>Universidade Federal de Minas Gerais, Brazil; <sup>2</sup>University College London, United Kingdom.

## Keywords

Health information systems, electronic health record, archetype modelling, two-level modelling, archetypes, ISO 13606; openEHR

## Summary

**Objective:** This paper aims to present the archetype modelling process used for the Health Department of Minas Gerais State, Brazil (SES/MG), to support building its regional EHR system, and the lessons learned during this process.

**Methods:** This study was undertaken within the Minas Gerais project. The EHR system architecture was built assuming the reference model from the ISO 13606 norm. The whole archetype development process took about ten months, coordinated by a clinical team co-ordinated by three health professionals and one systems analyst from the SES/MG. They were supported by around 30 health professionals from the internal SES/MG areas, and 5 systems analysts from the PRODEMGE. Based on a bottom-up approach, the project team used technical interviews and brainstorming sessions to conduct the modelling process.

**Results:** The main steps of the archetype modelling process were identified and described, and 20 archetypes were created.

Lessons learned:

- The set of principles established during the selection of PCS elements helped the clinical team to keep the focus in their objectives;
- The initial focus on the archetype structural organization aspects was important;
- The data elements identified were subjected to a rigorous analysis aimed at determining the most suitable clinical domain;
- Levelling the concepts to accommodate them within the hierarchical levels in the reference model was definitely no easy task, and the use of a mind mapping tool facilitated the modelling process;
- Part of the difficulty experienced by the clinical team was related to a view focused on the original forms previously used;
- The use of worksheets facilitated the modelling process by health professionals;
- It was important to have a health professional that knew about the domain tables and health classifications from the Brazilian Federal Government as member in the clinical team.

**Conclusion:** The archetypes (referencing terminology, domain tables and term lists) provided a favorable condition for the use of a controlled vocabulary between the central repository and the EMR systems and, probably, will increase the chances of preserving the semantics from the knowledge domain. Finally, the reference model from the ISO 13606 norm, along with the archetypes, proved sufficient to meet the specificities for the creation of an EHR system for basic healthcare in a Brazilian state.

**Correspondence to:**

Marcelo Rodrigues dos Santos  
Universidade Federal de Minas Gerais  
Brazil  
E-mail: marrsantos@terra.com.br

**Appl Clin Inf 2012; 3: 258–275**

doi:10.4338/ACI-2011-12-RA-0074

received: December 28, 2011

accepted: May 15, 2012

published: July 4, 2012

**Citation:** Santos MR; Bax MP; Kalra D. Dealing with the archetypes development process for a regional EHR system. Appl Clin Inf 2012; 3: 258–275  
<http://dx.doi.org/10.4338/ACI-2011-12-RA-0074>

## 1. Introduction

The development of Information Systems (IS) that can satisfy the fundamental requirements of the clinical data electronic recording process has attracted the efforts of the healthcare information technology scientific community [1]. These types of IS, generically known as the Electronic Health Record (EHR) systems, must, on the one hand, have the flexibility to represent specific medical knowledge in a way that allows it to assimilate its complexity and rapid evolution (of clinical concepts) and, on the other hand, satisfy the need for interoperability with other EHR systems. Such systems are built based on different data models, terminologies, concepts etc. As they exchange clinical data, there is a risk that mistaken interpretations may lower the accuracy of the communicated record and negatively impact on the quality of medical care.

Interoperability between EHR systems implies an agreement on the structure and meaning of the data being communicated. One proposal to handle this complexity (dynamic nature and sharing of the concepts) is two-level modelling. In this approach, two models are used: a generic information model (reference model) and a knowledge domain model, closer to the domain specialist (not the IT specialist). Suitable for the modelling of complex systems that involve a large number of domain specialists, the approach has strongly influenced the design of interoperability standards in the EHR systems field.

An example of an interoperability standard that uses this approach is the ISO 13606 norm [2], which specifies a reference model aimed at EHR system interoperability. This ISO norm proposes a model that represents either the complete EHR or selected parts of an EHR, and it is a simplification of the approach proposed by the openEHR Foundation<sup>1</sup>. In this way, data can be extracted from an EHR system for the purpose of communicating with another system or responding to a requesting system. [3–5]. This norm was used in the case of the Health Department of Minas Gerais State (SES/MG)<sup>2</sup> [6].

This article aims to present the results of the archetype modelling process used by SES/MG to support the establishment of a regional EHR system – based on a central repository – and the lessons learned from this activity.

## 2. Background

### 2.1. Two-level modelling

The idea of separation between a model of information and a model of knowledge has its origins in previous research within the Information Science field [7–12]. These concepts were applied in the health informatics field by certain European projects [3, 13–15] and subsequently expanded by the openEHR Foundation [16].

According to openEHR approach, the two-level modelling (or multi-level modelling), distinguishes between a reference model (information level) that is used to represent the generic properties of health record data, and an archetype model (knowledge level) used to represent the specific information that needs to be documented within health records by each clinical domain. The advantage of this split is the possibility to construct a domain specialist model (knowledge model) that is more independent from the developers' activities, which remain focused on the reference model [17, 18].

The reference model represents the global traits of the components of an EHR, the manner in which they are aggregated and the data context required to meet legal and ethical requirements, and the provenance (origin) of the data. Following this approach, it defines a set of classes that make up the structural blocks of an EHR, reflecting its stable traits. In order for this model to be stable and able to handle any particular clinical domain, it should be complemented by metadata structures reflect-

<sup>1</sup> The openEHR approach is a comprehensive open specifications for EHR systems originally based on the results of the European Union's GEHR-Project in the early 1990s – <http://www.openehr.org>.

<sup>2</sup> Health Department of Minas Gerais State – <http://www.saude.mg.gov.br>.

ing domain-specific traits, restricted by the general classes from the reference model. The formalism for the domain-specific traits is known as archetypes [3].

Archetypes are formal expressions of a domain content model in the form of declarations of structured constraints based on a Reference Model [1]. They specify a hierarchy of sub-classes that are components of a record, defining or constraining the names of each sub-class and other relevant attribute values, such as optionality and multiplicity at any point of the hierarchy, the types of data and value intervals the data elements may assume, and dependence constraints [3]. Generally designed for broad use, they can, however, be specialized to include local particularities and can accommodate any number of idioms and terminologies [18]. Archetype instances conform to a formal model known as archetype model [19].

## 2.2. ISO 13606 Norm

The ISO 13606 standard was developed by the Technical Committee ISO/TC 215, Health Informatics, and conceived from practical experience obtained during the implementation of the European precursor pre-standard, ENV 13606. This norm has focus on the communication between electronic health record systems. Its reference model is a simplification of the architecture put forward by the openEHR Foundation, and it is also based on the two-level modelling approach. In this alignment, the main classes of its reference model aims to establish an information model to communicate part or the entire electronic health record of a given patient (EHR extracts): preserving the original clinical meaning intended by the author; and reflecting the confidentiality of each data as intended by both author and patient [2].

Its reference model defines the following hierarchical levels for EHR data organization (containers): EHR\_Extract – the electronic health record for one person; Folder – high-level organization of the EHR; Composition – Clinical care session, encounter or document; Section – clinical headings reflecting the workflow and consultation process; Entry – Clinical “statements” about observations, evaluations, etc; Cluster – Nested multi-part data structures; Element – Leaf nodes with single data values. See these hierarchical levels in Figure 1.

The domain knowledge in the ISO 13606 norm also can be represented by a set of archetypes. According to the part 2 of the norm, an archetype represents a set of constraints placed upon the structure of the data present in the reference model. Usually individual archetypes are represented as serialized objects through a formalism named Archetype Definition Language (ADL).

## 2.3 Archetype Modelling

Archetypes can be developed through similar principles to the ones used in the creation of guidelines or protocols, based on best clinical practices and through development by multidisciplinary teams [20]. However, the clinical knowledge represented by an archetype goes far beyond that represented by a guideline. The archetype requires evidence about a knowledge object containing the details on the aspects of an intervention that has been undertaken and documented. Furthermore, it requires evidence associated with the concepts such as: rules, measurement intervals, data types, presentation formats, data representation conditions (codes, terminologies), etc. The data contained in an archetype should be sufficient to allow its interpretation on its own and be as complete as possible so that it may meet the needs of multiple sectors, purposes and priorities. However, they may be used in specific contexts, reflecting the data elements typical of a more focused environment [18, 20–21].

The openEHR Data Modelling Approach (ODMA) is a known process for the development of archetypes [22] that suggests the following steps should be taken:

1. search in all media (records in paper, guidelines, specialized magazines, etc.) for possible data elements;
2. group the elements into coherent concepts;
3. search for the identified concepts within a repository of archetypes and verify the existence of compatible archetypes;
4. develop the new archetypes that are needed by using a specific editor;
5. design templates to adapt archetypes to local circumstances or in order to combine different archetypes to represent a higher-level clinical documentation workflow.

## 2.4. Regional EHR System of Minas Gerais State

The SES/MG has begun a project to develop an EHR for all the cities in the state, which includes building a state-wide EHR repository, aimed at consolidating demographic data and producing a Patient's Clinical Summary (PCS) of its citizens for the Family Healthcare Programs (FHP).

The regional EHR system proposed for this project was based on the division of the State into macro-regions and on the implementation of vendor-based Electronic Medical Record (EMR) system to handle each one of them. Each macro-region, composed of several healthcare units, should use an EMR system and have a local datacenter. Therefore, the EHR system architecture was based on a central repository and on the creation of a messaging infrastructure, based on the ISO 13606 norm, aimed at ensuring the interoperability of the contracted EMR systems – the decision by use this norm was made mainly because its focus on EHR/EMR systems interoperability. The repository was designed to contain demographic data and the PCS necessary to meet the demands of the FHP. The architecture prescribes the use of archetypes to represent knowledge and includes four basic services: clinical data services, demographic data services, archetype services and terminology services [23]. See Figure 2.

The decision to use archetypes in the central repository was based on the flexibility and ease these artefacts can offer for medical knowledge reuse. Additionally, the archetype-based approach enabled the SES/MG clinical team to take direct responsibility for any evolutionary changes in the concepts specific to the health domain and for any new clinical data representations, while PRODEMGE would bear the “more technological” responsibility of maintaining the archetype processing infrastructure and managing the servers that host the central repository. This decision was made with the goal of obtaining the advantages expected with the two-level modelling approach [3].

The EHR system architecture was built on a Service Oriented Architecture (SOA) approach, the clinical data services were specifically designed to allow the exchange of patients' clinical data packages between systems and the central repository. These packages are called EHR extracts, which were developed based on the ISO 13606 norm reference model (Section 2.2) and complemented by a set of archetypes that define the central repository concepts. Thus, two services were proposed: one for sending data to the central repository and another to permit requests for data from the central repository – see Figure 2. In order to send data to the central repository, every day the EMR systems operating in the macro-regions will have to dispatch extracts composed of clinical data sets (within compositions) generated for each patient at the healthcare units and update the central repository.

At the time of preparing this manuscript, the selection process for the EMR systems that would participate in the State's EHR solution is in process, being conducted by the PRODEMGE<sup>3</sup>. Interoperability tests with the central repository are scheduled to begin in mid-2012 and, by 2013, the pilot period and the activation of the system in some regions/cities of the State are expected to have begun.

## 3. Methods

This study was undertaken in SES/MG. The whole archetype development process took about ten months. This process was based on the following steps: selecting data elements from the patient's clinical summary; identifying concepts that could become archetypes; modelling and validating entry archetypes; identifying domain tables, terminology and rules of permanence; searching for external sources of archetypes that could be used within this project; and codifying archetypes in ADL. The modelling process was coordinated by a clinical team formed by 3 health professionals and one systems analyst from the SES/MG, supported by the first author of this study. Additionally, the activities of data elements selection and concept definition involved around 30 health professionals from the internal SES/MG areas, and 5 systems analysts from the PRODEMGE. The clinical team used technical interviews and brainstorming sessions to conduct the modelling process.

<sup>3</sup> Information technology company of Minas Gerais State – <http://www.prodemge.gov.br>

The result validation process was mostly accomplished through the formalization of the data in spreadsheets that, once generated, were at each step sent to the participants by e-mail. The idea was to give each reviewer a means of collaborating on the basic clinical information structure without a special tool, and without having to face the full detail required to define each archetype. After validation, the recording of the approval was also done by e-mail. Some of the results were obtained during meetings held for this purpose and recorded in the minutes of the meetings.

## 4. Archetype Modelling Process in Minas Gerais Project

The archetypes were used exclusively to model clinical data that were to be kept in the central repository. The demographic data were represented in accordance with the model published by the Federal Government to support the national health card<sup>5</sup>. The aforementioned steps and the manner in which the clinical team conducted the archetype conception are detailed next.

### 4.1. Selection of Elements for the Patient's Clinical Summary

An essential part of the EHR scenario, the PCS, represents a set of clinical data elements resulting from the care provided at the state healthcare units. Initially, a set of principles guided the selection of the PCS data elements. These data elements should:

1. be sufficient to allow for patient mobility between regions of the State without compromising care;
2. include data for the monitoring of basic healthcare management indicators established by the SES/MG;
3. be the result of a consensus among the teams of various clinical specialties, who are in complete agreement with the comprehensive plan and contemplate the life cycle stages mentioned in the guidelines;
4. be structured (not free text) and, whenever possible, be compatible with the national healthcare terminology used by the Federal Government and with the domain tables used by the SES/MG.

Grounded on these principles, the clinical team used the FHP documents and the guidelines as a starting point for the identification of the mandatory data elements for the EMR systems that would be used. These documents were used in the various workshops carried out with the internal areas of the SES/MG. As a result, the clinical team identified several necessary data elements. Excluding similar or repeated data across the life cycle stages, 206 data elements were identified for the central repository: 84 relating to public health management and 122 directed at patient mobility – only those data elements related with patient mobility (clinical data) were subject of the archetype development. The demographic data elements were not included in this study (Table 1).

Thus, these 122 data elements were subjected to a rigorous analysis aimed at determining the most suitable clinical domain. After several meetings with the SES/MG internal areas, the clinical team achieved positive results. Only 10% of the 122 data elements would be free text (non-structured), the remainder were converted into internal term lists and domain tables or aligned with the national/international classifications. Once the data elements were identified, the candidate concept identification and the archetype modelling process was initiated.

### 4.2 Candidate Concept Identification and Archetype Modelling

Identification of the candidate concepts to be kept in the central repository, from the 122 data elements chosen, were targeted for archetype modelling. The modelling process began by considering the structural organization of the archetypes. At first, it was decided that a thematic archetype (composition) called “Clinical Summary” would be created. Also, it was agreed to use small Entry archetypes based on specific concepts, thus allowing greater reutilization flexibility. The later grouping of Entries into Sections proved to be a high complexity task, considering the still incipient modelling archetype knowledge that the team possessed. In order for them to be properly modelled, a well consolidated view of the use of these structures was needed. Therefore, postponing this modelling until

future archetype versions was expected increase the chances of their correct use.

Once the archetype organizational structure had been delineated, an analysis process aimed at surveying the concepts behind the identified data element groups began. To this end, the clinical team conducted brainstorming sessions in which the possible concepts were identified and the pertinence, adequacy and coherence of the data elements for the creation of these concepts were analyzed. Several candidate concept versions were created and a substantial modelling effort was put into this step of the process.

Levelling the concepts to accommodate them within the hierarchical levels in the reference model was definitely no easy task. The use of a mind mapping tool facilitated the modelling process, allowing the visualization of the relationship between concepts and data elements and the existing overlaps. The structures of the concepts resulting from the archetype-oriented modelling process were quite different (in terms of data grouping) from those employed in the original forms from which most data elements were extracted [5].

After an adjustment and correction process, the spreadsheet was finally validated by the Project Clinical Team. It should be noted that the team decided to use worksheets to facilitate the modelling process by health professionals. Many of them had difficulties in using archetypes editors directly because of the technical skills that were required. This worksheet was created according the ISO 13606 reference model classes and data type definitions. Also, the health professionals could use this spreadsheet to inform terminology codes and domain tables to be used. See a worksheet example in Table 2.

### 4.3 Terminology Analysis

The next step concerns the identification of the domain tables and terminology codes to be used to validate each data element. Control over the employed vocabulary is required in order to ensure the interoperability of the EMR systems with the central repository. The terminology server proposed within the EHR system architecture addresses this need (Section 3). This will aggregate the codes from a descriptive terminology<sup>4</sup>, term lists and domain table from SES/MG, and national/international classifications to be used. The clinical team analyzed the domain for each data element, specifying the appropriate term lists and/or domain tables to validate data elements. Some data elements needed to be validated against Federal Government classifications (e.g. clinical procedures and vaccines).

The decision regarding which descriptive terminology to use was, at the time of writing, still awaiting a decision from the Federal Government. The Committee for Health Informatics and Information of the Ministry of Health (CIINFO/MS) was intensifying efforts towards the adoption of a descriptive terminology in Brazil. It is important to note that whenever the SES/MG does require the EMR systems to start sending coded extracts using a descriptive terminology, the reference model of ISO 13606 already covers this type of use.

According to the ISO 13606 norm, all term lists, domain tables, classifications and terminologies used in EHR extracts must be identified by OID numbers (Object Identifier). This is done in order to ensure that both the issuing and the receiving EMR systems correctly identify the terminology systems in use. As a final step, the spreadsheet used in the archetype modelling process was revised taking the identified term lists, domain tables and classifications into consideration.

At this point, the clinical team researched the openEHR Foundation Clinical Knowledge Management (CKM) portal in order to analyze the existing archetypes held there with the goal of comparing them with the modelled concepts. Despite the fact that there was still a small number of archetypes published on the CKM portal during the execution of this project, they served as research sources to guide the revision of some data elements.

<sup>4</sup> In this study we consider a descriptive terminology those destined to make the implicit knowledge in a data element explicit. Term lists and domain table from SES/MG should be the ones to determine the domain of one or various data elements.

## 4.4 Permanence Rules Analysis

As mentioned when describing the regional EHR system of Minas Gerais State (Section 3), the PCS database should be up-to-date with the clinical data for each citizen at all times. To this end, a set of rules, called permanence rules, was created to ensure that only up-to-date data elements remain in this database. Data elements have specific traits relating to the updating of the PCS. In some data elements, the last value to occur must always be maintained, while in others, the last  $n$  occurrences or all the occurrences, or even occurrences during a given time interval etc. are to be maintained. Therefore, in order for the PCS to be kept up-to-date, the appropriate permanence rule for each data element had to be identified by a specific analysis from the clinical team.

Once the EHR extract is semantically validated, it remains in the history database, preserving the composition history for each patient, and lastly, depending on the permanence rule for each data element, it may remain in the PCS database. For this reason, the platform applies permanence rules to the data elements present in the extract, keeping the PCS up-to-date. These permanence rules only apply to data elements (Elements) that have such rules defined for use during import into the repository. This is an especially important property for a summary.

In order to manage these permanence rules we decided to treat them similarly to terminology binding codes within the structure of the ADL files. In this way, all essential parameters necessary to define the knowledge structure for the central repository were held within the archetypes.

## 4.5 Encoding Archetypes into ADL

When the Entry validation process was completed by the clinical team, with discussions about term lists, domain tables, classifications, terminologies and permanence rules having been finalized, the archetype encoding process began. The goal of this step was the formalization of the constraints associated with each data element using ADL codes [1].

The ADL codes resulting from this activity would be used in tandem with a group of XML Schema from the reference model to validate the EHR extracts, defining the set of constraints to which the data instances should be subject in order to be considered valid by the central repository (Section 3). Finally, with the help of the linkEHR editor<sup>5</sup>, all of the concepts present in the spreadsheet approved by the clinical team were encoded in the ADL.

## 5. Results

The archetype modelling process carried out by the SES/MG can be summed up in the steps presented in Table 3 and yielded 20 archetypes as shown in Table 4.

Most of the EHR system implementation is still work in progress in the Minas Gerais EHR project, but steps have been taken towards confirming that the archetypes are suitable. Firstly, the archetypes were validated by the clinical staff of SES/MG who considered the artefacts created through the use of the editor LinkEHR well able to reflect the information requirements set out. Secondly, these archetypes (ADL codes) were also published by the government<sup>6</sup> in order to receive comments from the health community and vendors who participated in the bidding process for EMR systems – this bidding process was in progress at the time of this the study. Through these initiatives, we could confirm that all concepts and its constraints could be correctly formalized in ADL code. The clinical team now wishes to submit these archetypes to the openEHR CKM community in order to contribute with the international archetype development process.

<sup>5</sup> The LinkEHR editor was developed by Universidad Politécnica de València. At this time, the linkEHR was the unique editor that allows to create archetypes based on the ISO 13606 reference model. More information can be found in <http://www.linkehr.com>.

<sup>6</sup> The archetypes listed at the Table IV were published on the site: <http://sres.saude.mg.gov.br>.



## 6. Discussion

The two-level modelling approach is adherent to the requirements for the development of health information systems, particularly within architectures that aim for a high level of interoperability. In this sense, the archetypes are important semantic artefacts to achieve semantic interoperability between EMR/EHR systems [24–28]. However, few publications to date have specifically focused on the process of archetype development [20, 22, 29]. This paper contributes to our understanding of the necessary stages in the development of good quality archetypes that are based on the information model of ISO 13606 norm, and we present here the main lessons learned during this process.

In a traditional, top-down archetype modelling process, the candidate concepts are identified and discussed, taking many possible angles of analysis into consideration. In the present case, candidate concepts were chosen through a bottom-up approach; in other words, to represent the data elements previously chosen to constitute the PCS. As a result, all data elements in the central repository were mapped to archetypes. In this way, the constraints were mapped and made available, through ADL codes, to the EMR systems.

This EHR system architecture was designed to preserve semantics in a knowledge domain, and for this reason it includes a reference model, archetypes and terminology. These semantic artefacts are important to support the interoperability process. Thus, the architecture was designed to support interoperability between the EMR systems. However, full semantic interoperability will only be achieved when the SES/MG starts to use a descriptive terminology.

## 7. Lessons Learned

The lessons learned during this case study can be summed up in the following points:

- The set of principles established during the selection of elements for the PCS guided the clinical team and helped them to keep focus on their objectives, even considering the participation of several other health professionals.
- The process assumed to develop the archetypes began considering the structural organization of the artefacts. It was very important because the clinical team could then understand how the concepts (information) should be organized.
- Each one of the data elements identified was subjected to a rigorous analysis aimed at determining the most suitable clinical domain. It forced the clinical team to undertake research into different sources of information resulting in more robust definitions for the concepts. Thus, only 10% of the data being free text (non-structured) constituted a gain in terms of knowledge representation and interoperability.
- Levelling the concepts to accommodate them within the hierarchical levels in the reference model was definitely no easy task. The use of a mind mapping tool facilitated the modelling process, allowing the visualization of the relationship between concepts and data elements and the existing overlaps.
- Part of the difficulty experienced by the clinical team in mapping the data elements into the candidate concepts was related to a view focused on the original forms previously used. The archetype modelling process should be oriented towards the formation of specialized concepts to allow their reutilization in different usage contexts.
- The grouping of Entries into Sections proved to be a high complexity task, considering the still incipient modelling archetype knowledge the team possessed. Therefore, postponing this modelling until future archetype versions should increase the chances of their correct use.
- The use of worksheets facilitated the modelling process by health professionals. Many of them had difficulties in using archetypes editors directly because of the technical skills that were required. These worksheets were created according to the ISO 13606 reference model classes and data type definitions.
- Considering the large number of domain table and classifications from the Federal Government in Brazil, it was very important to have a health professional that knew about this subject as member in the clinical team.

## 8. Conclusion

The archetype modelling process used by the SES/MG to support building the regional EHR system and the lessons learned are presented here. As a result, the steps in the archetype modelling process conducted by the SES/MG were identified and a spreadsheet of the generated archetypes was produced. From these results, one can see that the archetypes (referencing terminology, domain table and term lists) provided favourable conditions for the use of a controlled vocabulary between the central repository and the EMR systems, and probably, will increase the chances of preserving the semantics from the knowledge domain. It is worth stressing the degree of adequacy of the two-level approach that laid the foundation for the construction of the proposed architecture. Finally, the reference model from the ISO 13606 norm, along with the archetypes, proved sufficient to meet the specificities for the creation of a Regional EHR system for basic healthcare in a Brazilian state.

### Clinical Relevance Statement

The archetypes are important semantic artefacts to achieve semantic interoperability among EMR/EHR systems. However, few publications to date have specifically focused on the process of archetype development. This paper contributes to our understanding of the necessary stages in the development of good quality archetypes that are based on the information model of ISO 13606 norm, and we present here the main lessons learned during this process.

### Conflicts of interest

The authors declare that they have no conflicts of interest in the research.

### Human Subjects Protections

The authors declare that human and animal subjects were not included in this research.

### Acknowledgements

We acknowledge the support of the Health Secretary of State of Minas Gerais and Prodemege for the development of this research.

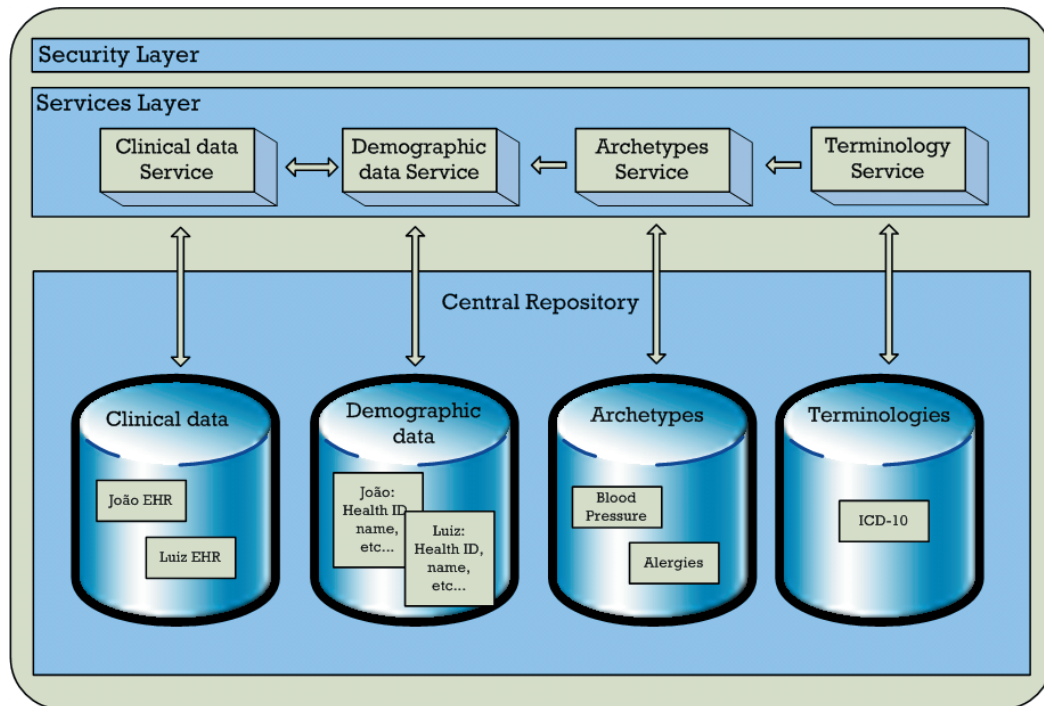


Fig. 1 Regional EHR System Architecture

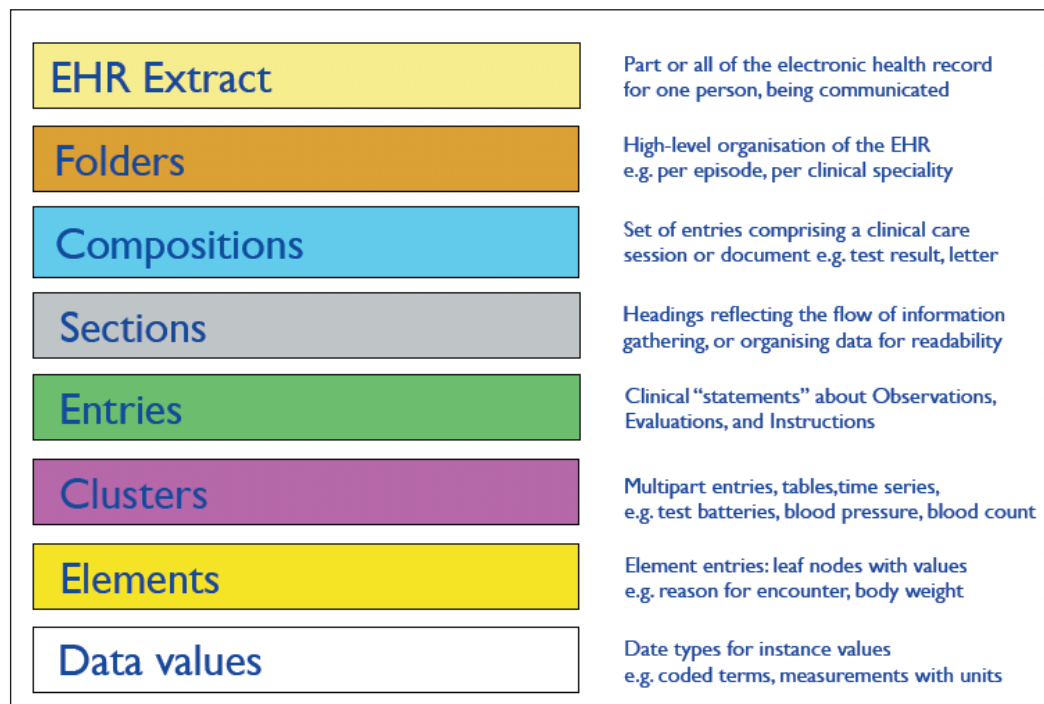


Fig. 2 ISO 13606 hierarchical levels [3]

**Tab. 1** 122 data elements identified for the PCS

#	Data Element	#	Data Element	#	Data Element
1	Periodontitis	29	Apgar score fifth minute	53	Posology of medication
2	Type of Delivery	30	History of dental traumatism / cause	54	Medication being used
3	Gestational Age	31	Other causes of dental traumatism	55	Date when use of medication began
4	Intercurrences during prenatal period	32	Classification of newborns	56	Medical prescription
5	Other Intercurrences during prenatal period	33	Skull Perimeter	57	Adverse effects of medication
6	VDRL	34	Intercurrences during neonatal period	58	Newborn Hearing Screening test
7	Toxoplasmosis	35	Other Intercurrences during neonatal period	59	Hearing Screening test alterations
8	Anti HIV	36	Measured value	60	Body mass index
9	HBsAg	37	Weight	61	Lesions affecting the mucosa of the oral cavity
10	Name of complementary examination	38	Orthoses and prostheses	62	Fluorosis
11	Specification of complementary examination	39	Presence of disability	63	Waist circumference
12	Complementary examination result date	40	Allergy	64	Cardiovascular risk associated with waist circumference
13	Complementary examination result	41	Description of Allergy	65	Systolic arterial pressure value
14	Glycated hemoglobin dosage result date	42	Name of immunobiologicals	66	Diastolic arterial pressure value
15	Glycated hemoglobin dosage result	43	Immunobiologicals dosage	67	Activities of Daily Living test – ADL – Self-care
16	Neonatal screening result: Hypothyroidism	44	Immunobiologicals lot number	68	Instrumental activities of daily living test – IADL
17	Neonatal screening result: Phenylketonuria	45	Expiration date of Immunobiologicals	69	Falls in the last year
18	Neonatal screening result: Cystic fibrosis	46	Immunobiologicals application date	70	Number of falls in the last year
19	Neonatal screening result: Sickle-cell disease	47	Immunobiologicals Adverse events	71	Fractures after falls
20	Indirect Coombs result date	48	Immunobiologicals are up-to-date	72	Activity limitation after falls
21	Indirect Coombs test result	49	Consumption of Alcohol	73	Incapacity after a fall
22	Hemoglobin test result	50	Tobaccoism	74	Compromised Mobility after a fall
23	Hematocrit test result	51	Use of other substances / licit or illicit	75	Cognitive disability
24	Fasting Glycemia test result	52	Name of the medication		
25	Routine urine test result				
26	Urinalysis test result				
27	Ultrasonography test result				
28	Apgar score first minute				

Tab. 1 Continued

#	Data Element	#	Data Element	#	Data Element
76	Postural instability	91	Number of gestations	106	Description of Edema
77	Immobility	92	Dental Caries Activity	107	Uterus
78	Incontinence	93	Gingivitis	108	Uterine Height
79	Iatrogenesis	94	Number of Abortions	109	Position of Fetus
80	Diagnostic impressions	95	Number of Deliveries	110	Condition of Fetus
81	Risk stratification for children	96	Number of Vaginal Births	111	Dorsal Portion of Fetus
82	Risk stratification for adolescents	97	Number of C-Section Births	112	Fetal Heart Rate
83	Risk stratification for the elderly	98	SISPRENATAL reference number	113	Fetal Movements
84	Cardiovascular risk stratification	99	Date of Last Menstrual Period	114	Cervix
85	Risk stratification – Diabetes	100	Probable Date of Delivery	115	Dilation Condition
86	Risk stratification for current gestation	101	Probable Date of Delivery after first Ultrasonography test	116	Dilation rate
87	Risk stratification – Mental Health	102	Result of First Ultrasonography test	117	Planned pregnancy
88	Risk stratification – Tuberculosis	103	Prenatal Consultation Reference Number	118	Care Plan
89	Risk stratification – Hansen’s disease	104	Gestational Age	119	Prenatal care site
90	Risk classification – oral health	105	Presence of Edema	120	Relation with maternity hospital of reference
				121	Scheduling Pregnant Women Group Meetings
				122	Observations on Current Gestation

**Tab. 2** Example of spreadsheet used during the modelling process based on ISO 13606 reference model

Entry	Cluster	Element	Optionality	Description	Format	Datatype CEN14796	Terminology	Element multiplicity
Allergy	Allergies list	Classification	Mandatory	Classification of allergies to food, medications and other substances.	Number – 2 characters	ORD	1. Allergy to foods 2. Allergy to animals 3. Allergy to cosmetics 4. Allergic to detergents 5. Allergy to drugs 6. Allergy to fungi 7. Allergy to ... ...	
Allergy	Allergies list	Observation	Optional	Observations regarding the allergies.	Alphanumeric – limited to 1000 characters	SIMPLE_TEXT		
Disabilities		Kind of disability	Optional	Kind of disability the patient has.	Number – 1 character	ORD	0 – no observed 1 – hearing 2 – Physical 3 – mental 4 – ostomy 5 – visual 6 – multiple	multivalued
Disabilities		Severity degree	Optional	Severity degree observed	Number – 1 character	ORD	1 – mild 2 – moderate 3 – severe	

Note: in order to fit the spreadsheet to this page just some columns were included and the concepts are not complete.

**Tab. 3** Detailing of the archetype modelling steps in the SES/MG

Step	Description	Methodology	Participants	Delivered artifact	Effort
Patient's Clinical Summary (PCS) element selection	Defining the data elements that would compose the PCS.	<ul style="list-style-type: none"> <li>✓ Workshops with the SES/MG internal areas;</li> <li>✓ Domain analysis for each identified element.</li> </ul>	Clinical Team; Representatives from the internal areas of the SES/MG.	122 PCS data elements identified.	60 days
Identification of the Candidate Concepts	Identifying the candidate concepts that properly accommodated the data elements chosen for the Summary.	<ul style="list-style-type: none"> <li>✓ Discussions to identify candidate concepts;</li> <li>✓ Development of Mind Maps.</li> </ul>	Clinical Team	Mind Maps including all concepts identified.	60 days
Archetype Modelling	Modelling the archetypes according to the hierarchical levels prescribed by the ISO 13606 norm.	<ul style="list-style-type: none"> <li>✓ Modelling the candidate concepts in accordance with the classes prescribed by the ISO 13606 norm;</li> <li>✓ Developing an MS Excel Spreadsheet based on the RM of ISO 13606.</li> </ul>	Clinical Team	MS Excel spreadsheet with the Entry modelling results.	90 days
Terminology Analysis and Identification	Validating term lists, domain table, classifications and terminology to be used.	<ul style="list-style-type: none"> <li>✓ Analyzing applicable descriptive terminology;</li> <li>✓ Analyzing term lists and domain tables for data element validation;</li> <li>✓ Complementing the archetype modelling with the inclusion of term lists, domain tables and terminology.</li> </ul>	Clinical Team	MS Excel spreadsheet with the Entry modelling results complemented by terminologies.	45 days
Analysis and Identification of Permanence Rules	Studying the permanence rules to be used.	<ul style="list-style-type: none"> <li>✓ Analyzing data element permanence rules in the PCS;</li> <li>✓ Complementing the archetype modelling with the inclusion of the summarization rules.</li> </ul>	Clinical Team	MS Excel spreadsheet with the Entry modelling results complemented by permanence rules.	It was started in parallel of the Terminology Analysis – 45 days.
Existing Archetype Research	Researching existing archetypes and revising modeled Entries.	Accessing the openEHR/CKM website and analyzing the published archetypes.	Clinical Team	Revised archetypes.	It was started in parallel of the Terminology Analysis – 45 days.
ADL Archetype Encoding	Archetype encoding in ADL	<ul style="list-style-type: none"> <li>✓ Studying the ADL language;</li> <li>✓ Archetype encoding with the use of the linkEHR RC2 editor.</li> </ul>	PRODEMGE technical team	Archetypes encoded in ADL	45 days

**Tab. 4** Summary of the Archetypes for the Entries Created

Archetype	Entry
ISO13606-SESMG-ENTRY. Alergias.v1	Alergias (Allergies)
ISO13606-SESMG-ENTRY. Alimentacao.v1	Alimentação (Diet)
ISO13606-SESMG-ENTRY. AntecedentesObstetricos.v1	Antecedentes Obstétricos (Obstetric History)
ISO13606-SESMG-ENTRY.AntecedentesOdontologicos.v1	Antecedentes Odontológicos (Dental History)
ISO13606-SESMG-ENTRY. AtividadesCotidianas.v1	Atividades Cotidianas (Daily Activities)
ISO13606-SESMG-ENTRY.AvaliacaoFuncional.v1	Avaliação Funcional (Functional Evaluation)
ISO13606-SESMG-ENTRY.Deficiencias.v1	Deficiências (Disabilities)
ISO13606-SESMG-ENTRY.EstratificacaoRisco.v1	Estratificação de Risco (Risk stratification)
ISO13606-SESMG-ENTRY.ExameClinicoOdontologico.v1	Exame Clínico Odontológico (Dental Clinical Examination)
ISO13606-SESMG-ENTRY.ExameFisico.v1	Exame Físico (Physical Examination)
ISO13606-SESMG-ENTRY.ExamesComplementares.v1	Exames Complementares (Complementary Exams)
ISO13606-SESMG-ENTRY.GestacaoAtual.v1	Gestação Atual (Current Gestation)
ISO13606-SESMG-ENTRY.GigantesGeriatra.v1	Gigantes da Geriatria (Giants of Geriatry)
ISO13606-SESMG-ENTRY.ImpressaoDiagnostica.v1	Impressão Diagnóstica (Diagnostic Impression)
ISO13606-SESMG-ENTRY. Imunobiologicos.v1	Imunobilógicos (Immunobiologicals)
ISO13606-SESMG-ENTRY. Medicamentos.v1	Medicamentos (Medication)
ISO13606-SESMG-ENTRY. Mobilidade.v1	Mobilidade (Mobility)
ISO13606-SESMG-ENTRY. Parto.v1	Parto (Childbirth / Delivery)
ISO13606-SESMG-ENTRY.Prenatal.v1	Pré-natal (Prenatal care)
ISO13606-SESMG-ENTRY. Puerperio.v1	Puerpério (Puerperium)

Note: these archetypes can be found in <http://sres.saude.mg.gov.br>



## References

1. Beale T, Heard S. Archetype Definition Language 1.4: rev. 1.4.1 [Internet]. London: The OpenEHR Foundation, 2007. Available from: <http://www.openehr.org/svn/specification/TRUNK/publishing/architecture/am/adl1.4.pdf>.
2. International Organization For Standardization. ISO 13606–1:2008: Health informatics, Electronic health record communication, Part 1: Reference model, 2008 [Internet]. Available from: [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=40784](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40784).
3. Kalra D. Electronic Health Record Standards. *IMIA Yearbook of Medical Informatics*, 2006(1): 136–139.
4. Kalra D. Barriers, Approaches and research priorities for semantic interoperability in support of clinical care delivery. 2007. Available from: [http://www.semantichealth.org/DELIVERABLES/SemanticHealth\\_D4\\_1\\_final.pdf](http://www.semantichealth.org/DELIVERABLES/SemanticHealth_D4_1_final.pdf).
5. Garde S, P Knaup, Hovenga EJS, Heard S. Towards semantic interoperability for electronic health records: domain knowledge governance for openEHR archetypes. *Methods Inf Med* 2007; 46(3): 332–312.
6. Santos MR. Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais [Thesis]. Information Sciences Departament: Universidade Federal de Minas Gerais, 2011.
7. Newell A. The knowledge level [Internet]. *AI Magazine*. Summer 1982; 18(1).
8. Gruber TR. A translation approach to portable ontology specifications. *Knowledge Acquisition* 1993; 5(2): 199–220.
9. Guarino N. Formal Ontology, Conceptual analysis and knowledge representation. *Int J Human-Computer Studies* 1995; 43(5): 625–640.
10. Guarino N, Giaretta P. Ontologies and knowledge bases: towards a terminological clarification. In: Mars, NJI. *Towards very large knowledge bases: knowledge building and knowledge sharing*. Amsterdam: IOS Press, 1995. p. 25–32.
11. Johnson SB. Model formulation: generic data modelling for clinical repositories. *J Am Med Inform Assoc* 1996; 3: 328–339.
12. Chandrasekaran B, Josephson R, Benjamins VR. Ontology of tasks and methods. *Proceedings of the International Joint conference on artificial intelligence* 1999; 16: 31–43.
13. Kalra D. The Synapses project. *Proceedings of the Current Perspectives in healthcare computing conference*; 1997; Harrogate, England: BJHC, 1997. p. 405–406.
14. Grimson J, Grimson, W, Berry D, Stephens G, Felton E, Kalra D, et al. A CORBA – Based integration of distributed electronic healthcare records using the Synapses approach. *IEEE Trans Inf Technol Biomed* 1998; 2(3): 124–115.
15. Bird L, Goodchild A, Tun Z. Experiences with a two-level modelling approach to electronic health records. *J Res Prac Inf Tech* 2003; 35(2): 121–118.
16. OpenEHR [Internet]. OpenEHR future-proof and flexible; 2010. Available from: <http://www.openehr.org/about/origins.html>.
17. Michelsen L, Pedersen SS, Tilma HB, Andersen SK. Comparing different approaches to two-level modeling [Internet]. *Stud. Health Tech. Inform.* Amsterdam, 2005. Available from: <http://iospress.metapress.com/index/1hm0l4hd611u9dey.pdf>.
18. Beale T, Heard S. Archetype definitions and principles: rev. 1.0 [Internet]. London: The OpenEHR Foundation, 2007. Available from: [http://www.openehr.org/svn/specification/TRUNK/publishing/architecture/am/archetype\\_principles.pdf](http://www.openehr.org/svn/specification/TRUNK/publishing/architecture/am/archetype_principles.pdf).
19. Beale T, Heard S. Architecture overview: rev. 1.1 [Internet]. London: The OpenEHR Foundation, 2007. Available from: <http://www.openehr.org/svn/specification/TRUNK/publishing/architecture/overview.pdf>.
20. Hovenga EJS, Garde S, Carr T, Hullin CM. Innovative approaches and processes for capturing expert aged care knowledge for multiple purposes. *Electronic Journal of Health Informatics* 2007; 2(1).
21. Leslie H, Heard S. Archetypes 101. Health Informatics Society of Australia Ltd. 2006.
22. Buck J, Garde S, Kohl CD, Knaup P. Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach. *Int J Med Inf* 2009; 78(8): 521–511.
23. Santos MR, Bax MP, Kalra D. Building a logical EHR architecture based on ISO 13606 standard and Semantic Web Technologies. *Proceedings of the World congress on medical and health informatics* 13; 2010; South África, 2010.
24. Kalra D. Clinical foundations and information architecture for the implementation of a federated health record service [Thesis]. University College London, London, 2002.
25. Moner D, Maldonado JA, Bosca D, Fernández JT, et al. Archetype-based semantic integration and stan-

- standardization of clinical data. Proceedings of the EMBS IEEE Annual International Conference 2006; 28 New York: IEEE.
26. Atalag K. Archetype based domain modelling for health information systems [Thesis]. Department of Information Systems, The Graduate School of Informatics, Middle East Technical University, 2007.
  27. Munoz A, Somolinos R, Pascual M, Fragua JÁ, et al. Proof-of-concept design and development of an EN13606-based electronic health care record service. *J Am Med Inform Assoc* 2007; 14(1): 118–129.
  28. Wollersheim D, Sari A, Rahayu W. Archetype-based electronic health records: a literature review and evaluation of their applicability to health data interoperability and access. *HIM J* 2009; 38(2).
  29. Conde AM. Towards best practice in the archetype development process [Dissertation]. Trinity College of Dublin, 2010.