MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-207

Case Study: Applying OpenEHR Archetypes to a Clinical Data Repository in a Chinese Hospital

Lingtong Min[a], Li Wang[a], Xudong Lu[a*] and Huilong Duan[a]

[a] *College of Biomedical Engineering & Instrument Science, Zhejiang University, China*

Abstract

openEHR is a flexible and scalable modeling methodology for clinical information and has been widely adopted in Europe and Australia. Due to the reasons of differences in clinical process and management, there are few research projects involving openEHR in China. To investigate the feasibility of openEHR methodology for clinical information modelling in China, this paper carries out a case study to apply openEHR archetypes to Clinical Data Repository (CDR) in a Chinese hospital. The results show that a set of 26 archetypes are found to cover all the concepts used in the CDR. Of all these, 9 (34.6%) are reused without change, 10 are modified and/or extended, and 7 are newly defined. The reasons for modification, extension and newly definition have been discussed, including granularity of archetype, metadata-level versus data-level modelling, and the representation of relationships between archetypes.

Keywords:

Archetype, openEHR, Clinical Data Repository, Chinese hospital.

Introduction

A Clinical data repository (CDR) is a repository that stores the clinical data integrated from various kinds of clinical information systems for analysis and research. An increasingly larger demand for clinical research and application has drawn international initiatives' attention and propels institutions such as the Mayo Clinic [1], Intermountain Healthcare [2], Stanford Medical Center [3], Massachusetts General Hospital [4] and Columbia University Medical Center [5], to build CDRs.

As healthcare is a highly complicated and rapidly developing domain, the flexibility of the data model in CDR is rather indispensable. Being an open architecture for Electronic Health Record, openEHR advocates a dual-level methodology to conduct data modelling for clinical information [6]. It also provides a flexible modeling methodology to adapt to the evolution of clinical concept and knowledge. The architecture of openEHR includes reference model and archetype to separate clinical knowledge from clinical information. The reference model defines the data type, data structure, basic framework of EHR and represents the global characteristics of health record entries [7]. The archetype is a conceptual model, which is built on the base of clinical knowledge by clinical experts and specifies constraints on the reference model.

Although openEHR methodology has been widely carried out and implemented in Europe and Australia [8-10], only a few researches studies utilizing openEHR have been conducted in China. Most of the existing publications in China were limited to the introduction or translation of the basic principles and concepts of openEHR [11-13]. Only few were related to openEHR implementation in the area of research [14], while none of them focuses on data modeling in clinical settings. The possible obstacles are the differences in clinical process of hospital and the healthcare management mechanisms between countries. To implement openEHR in China, an investigation of the feasibility of using openEHR methodology to model clinical data in China is a necessity. This paper introduces a case study of applying the openEHR archetypes to CDR in a Chinese hospital.

Methods

The selected tertiary class A hospital in the case study has deployed several information systems including HMIS (hospital management information system), CIS (clinical information system), PACS (picture archiving and communication system), RIS (radiology information system), and LIS (laboratory information system). The clinical data has been scattered in these silo systems, and the physicians have to access different patient data in corresponding system respectively. It is crucial to build a CDR to integrate all the clinical data from these heterogeneous information systems and provide real-time data access services for applications to browse the complete set of patient data in data viewing applications. To achieve this target, the CDR needs to contain all essential clinical data including domains of patient demographics, encounters, medication, imaging examination, and laboratory test. The authors have designed a 6-step method to model the data in CDR via openEHR.

1. Analyze MOH standards

The data requirements of the CDR are first analyzed with regard to several standards. Chinese MOH (Ministry of Health) has published a set of standards related to healthcare data sets to facilitate information collection, storage, and exchange, such as "WS 363-2011 Health data element dictionary" (WS 363-2011) and "WS 445-2014 Basic dataset of electronic medical record" (WS 445-2014). "WS 363-2011" defines the identification, naming, meaning, data type, representation, and value set for all the data elements in healthcare domain. The purpose of this standard is to provide the standardized definition of data elements for all applications in healthcare domain. "WS 445-2014" specifies the typical business model and clinical documentation of the electronic medical record (EMR) together with the data sets under this architecture. The data elements used in this

standard conform to "WS 363-2011", except for some that are further constrained to adapt to the EMR context.

Firstly, the standards of "WS 445-2014" and "WS 363-2011" were analyzed according to the requirements of the CDR scope, content, and data elements. The coverage of standards for the requirements is listed in Table 1 and Table 2.

To be more specific, for each CDR data requirement, we need to find the corresponding data sets of "WS 445-2014" and refer to "WS 363-2011" for collecting the necessary data items in CDR, "none" indicates that the CDR requirements are not included in standards. For instance, the imaging examination requirements in the CDR refers to three data sets of "outpatient and emergency medical record", "examination and laboratory test record" and "inpatient progress note" in "WS 445-2014". After that, the relevant imaging examination data items can be found within "identification", "assistant examination" and "healthcare organization" in "WS 363-2011".

Table 1 – CDR requirements and WS 445-2014

WS 445-2014	CDR requirements
1) medical record summary	patient demographics, encounters
2) outpatient and emergency medical record	imaging examination
3) outpatient and emergency prescription	medication
4) examination and laboratory test record	imaging examination, laboratory test
5) general therapy and treatment record	medication
6) delivery record of therapy and treatment	none
7) nursing operation records	none
8) nursing valuation and plan	none
9) informing information	none
10) home page of inpatient medical record	none
11) home page of inpatient medical record summary of TCM	none
12) admission record	encounters
13) inpatient progress note	imaging examination
14) inpatient order	medication
15) discharged brief	encounters
16) transfer record	encounters
17) medical institution information	none

Table 2 – CDR requirements and WS 363-2011

WS 363-2011	CDR requirements
1) identification	patient demographics, encounters, medication, imaging examination, laboratory test
2) demographics and social economics characteristics	patient demographics
3) health history	none
4) health risk factor	none
5) chief complaint and symptom	none
6) physical examination	none
7) assistant examination	imaging examination
8) laboratory examination	patient demographics, laboratory test
9) diagnosis	encounters
10) medical assessment	encounters
11) medical plan and intervention	encounters, medication
12) health expenditure	none
13) healthcare organization	patient demographics, encounters, medication, imaging examination, laboratory test
14) health personnel	none
15) drug and material	medication
16) health management	none

2. Analyze existing database schemas

The data items defined in the standards can only cover part of the CDR data requirements, and the schemas of the existing heterogeneous information systems should also be taken as a source for collecting the necessary data elements in CDR as shown in Table 3. For example, the data items of imaging examination can be found from Exam request, Exam item, Exam procedure, Image, and Report four tables in the existing database schemas.

Table 3 – CDR requirements and existing database schemas

CDR requirements	Information systems	Table schemas
patient demographics	HMIS	Patient
encounters	CIS	Visit, Inpatient admission
imaging examination	RIS	Exam request, Exam item, Exam procedure, Image, Report
laboratory test	LIS	Test request, Test item, Report, Sample
medication	CIS	Medication order

3. Merge data items

Data items collected from the above two steps are merged into the final data set for CDR. If the data type, value set, or coding for same data element is not compatible in the standards and the database of existing systems, the one from the standards was used. After merging, 217 data elements were finally defined in a structured format and ready to be used as shown in Table 4. The data items of imaging examination is composed of 38 items among which 13 items came from standards and 25 items came from information systems.

Table 4 – Number of data items collected from standards and information systems

CDR requirements	Standards	Information systems
patient demographics	18	21
encounters	16	45
imaging examination	13	25
laboratory test	10	17
medication	16	36

4. Organize data items into concepts

Since each archetype models only one distinct concept, the data items should be organized into concepts before modeling. Although the data items are classified into basic data sets in the standards, the classification is mainly clinical documentation oriented and the data sets often contain a number of distinct concepts. For example, the data sets in "part 4) examination and laboratory test record" of "WS 445-

2014" contain two sub-domains of laboratory test and examination, and there are concepts about request, result, report, and specimen in each sub-domain. The data elements in "WS 363-2011" are just listed in a row one by one. On the other hand, the data schemas from different information systems are heterogeneous and overlapped. To organize the data items from both standards and data schemas of different systems, an entity-relationship concept modeling process is carried out and results in a group of 26 concepts. Semantic overlapping between concepts is avoided to our best to comply with the single archetype for single concept principle. The derived clinical concepts are shown in Figure 1.

5. Map concepts to archetypes

To achieve maximum reusability, the public archetype repository Clinical Knowledge Manager (CKM) is searched for matching archetypes with each derived concept with the name or other key words. The candidate archetypes filtered by searching are analyzed in details. If all the data items of the clinical concept are covered by the existing archetype, this archetype can be directly adopted without any modification. If only part of the data items of clinical concept is covered by the existing archetype, this archetype needs to be modified and extended. If more than one matching archetypes exist, the most semantically suited one was selected and then checked for whether it could completely cover or partially cover the concept. If there is no matching archetype for the clinical concept, a new one is defined. For archetype modification and extension, several operations are illustrated in the openEHR specification [15] such as revision, specialization, and new version.

6. Model relationships between concepts

After mapping concepts to archetypes for the CDR, the relationships between the concepts also need to be modeled. There are two general methods to express relationships between archetypes, archetype slot and link. Archetypes can be composed to express valid possibilities for larger structures of data from different levels of ontological hierarchy of the reference model. Such compositional connections are termed as 'slots'. For example, "Section" and "Entry" archetypes can be composed through slots to represent the structure similar to clinical document with head, body, and content. Currently, almost all the archetypes in CKM use slots to express relationships between archetypes. "Link" is an attribute root in the deep architecture of the openEHR reference model and can refer to many other archetype structures.

Results

26 archetypes are identified to cover all the data requirements of the CDR, shown in Figure 2. 9 archetypes, listed in Table 5, are from CKM and can be directly reused without any modification, which account for 34.6% of all 26 archetypes. 10 archetypes, listed in Table 6, are modified and extended, among which 2 are revisions by replacing compatible data types with some data items, 8 are modified by adding data items, no specialization and no new version archetypes. 7 archetypes, listed in Table 7, are newly defined since not all the clinical concepts are covered by the existing archetypes, which include 2 for laboratory test, 2 for examination, 2 for medication orders, and 1 for transfer management.

So far, the 26 archetypes have been used to implement the CDR based on the relational database that provides data access service in the hospital. Clinical data has been integrated from the heterogeneous systems into the CDR and a clinical data viewer for clinicians has been developed and used in the clinical practice.

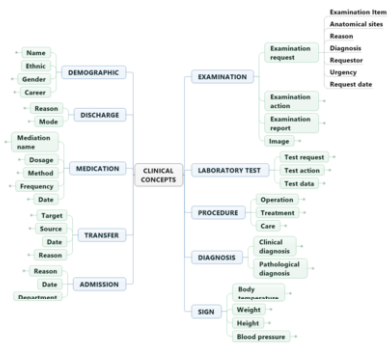

Figure 1 – The clinical concepts of the CDR

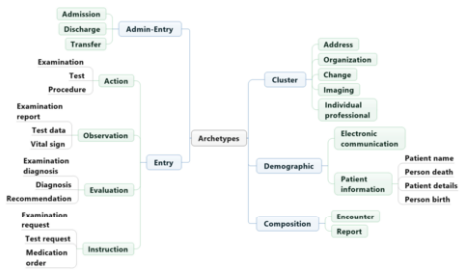

Figure 2 – The archetype structure of the CDR

Table 5 – Not changed archetypes

Archetype
ADMIN_ENTRY.discharge_admin_info.v3
CLUSTER.address.v1
CLUSTER.lab_result_annotation.v1
CLUSTER.medication_amount.v1
CLUSTER.organisation.v1
CLUSTER.specimen.v1
DEMOGRAPHIC-ADDRESS.electronic_communication.v1
DEMOGRAPHIC-PARTY_IDENTITY.person_name.v1
DEMOGRAPHIC-PERSON.person-patient.v1

Table 6 – Modified and extended archetypes

Archetype
DEMOGRAPHIC-CLUSTER.person_identifier.v1.1
DEMOGRAPHIC-ITEM_TREE.person_details.v1
ACTION.imaging_exam.v1.1
ACTION.medication.v1.1
CLUSTER.medication_admin.v1
ADMIN_ENTRY.admission.v1.1
COMPOSITION.encounter.v1.1
INSTRUCTION.medication.v1.1
INSTRUCTION.request-imaging_exam.v1.1
INSTRUCTION.request-lab_test.v1.1

Table 7 – Newly developed archetypes

Archetype
ACTION.lab_test.v1
ACTION.medication_order_schedule.v1
ADMIN_ENTRY.transfer.v1
INSTRUCTION.medication_order_schedule.v1
OBSERVATION.imaging_exam_image_series.v1
OBSERVATION.imaging_exam_report.v1
OBSERVATION.lab_test_single.v1

Discussion

Although all the clinical concepts retrieved from the CDR of Chinese hospital can be modelled using the openEHR archetype approach, several issues were encountered and discussed below.

Immaturity of archetype modification operations

openEHR specifies three operations for archetype modification: revision, specialization, and new version [15]. Other than these three operations, there is another operation with subtle differences. So far, the only way to add data items to existing archetypes is through specialization. Since each archetype represents a concept, this requires creating a new specialized archetype, which represents a new concept. There is no way to add these data items to existing archetypes. Since the archetype development is still an ongoing process, there are great demands for this requirements. We define a new operation for this situation named extension shown in Table 8.

Table 8 – Operations for archetype modification

Operation	Modification	Compatibility
Revision	Modify description part	Ensure backward compatibility
	Expand attributes, range of value sets, terminology	Data created by pre-revised archetype is compatible with the revised version
Specialization	Strengthen the constraints	Ensure the new specialized archetype must create data that conforms to the parent
	Redefine and add nodes	
	The range of value sets and semantics of nodes conform to the previous archetype	
New version	Change mandatory item to optional	Modifications are incompatible with the previous archetype
	Adjust value range or coded term set	
Extension	Add missing data items to existing archetypes	Compatible with the original archetype
	Use semantic version as the naming rule	

To take an example, if two data items of memo and report identifier need to be added in the archetype INSTRUCTION.request-imaging_exam.v1, it can be easily versioned as INSTRUCTION.request-imaging_exam.v1.1 through the "Extension" operation.

With the extension operation, archetypes can be clearly managed using the semantic versioning mechanism. Applying archetypes to local context will be much easier since there is no need to figure out a proper name each time to add data items through specialization.

The granularity of archetype

Differences in the granularity of archetypes between the CDR data requirements and CKM archetypes can cause problems of information representation in clinical practices. Take imaging examination sub-domain as an example. The concepts extracted from the CDR include Request, Request Item, Result, Report, DICOM Study, and Image shown in Figure 3a. Two coarse-grained archetypes (a) INSTRUCTION.request-imaging_exam.v1 which contains Request, Request Item and (b) OBSERVATION.imaging_exam.v1 which contains Result, Report, DICOM Study, Image are found in CKM and analyzed to extract corresponding concepts in Figure 3b. By comparing Figure 3a with Figure 3b, several important relationships between concepts are missing or improper, such as ③ one-to-many relationship between Report and Request Item, ⑤ one-to-one relationship between Request Item and DICOM Study, improper relationship cardinality between ④ Report and DICOM Study. The problem is that OBSERVATION.imaging_exam.v1 contains multiple concepts and the relationships between these concepts are greatly limited and not suitable for the CDR requirements. These differences result in structural modifications and reorganization of the archetypes which are demonstrated in Figure 3c. To represent ④, OBSERVATION.imaging_exam.v1 is split into two archetypes, (c) OBSERVATION.imaging_exam_report.v1 for Result and Report and (d) OBSERVATION.imaging_exam_image_series.v1 for DICOM Study and Image.

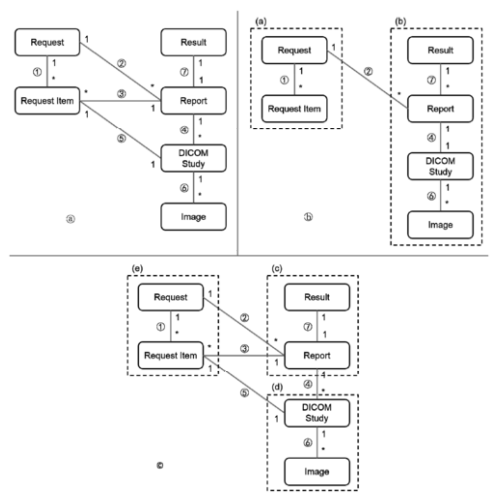

Figure 3 – Imaging examination concept relationships.

Bearing the principle of one archetype for one clinical concept in mind, with more and more clinical requirements appended, fine-grained archetypes is more flexible and easier to represent concepts and its relationships, and will improve the modelling capability of archetypes eventually.

Metadata-level versus data-level modelling

Problems are also encountered due to mismatches between metadata-level modelling and data-level modelling which happen in candidate archetypes and the CDR data requirements. For instance, all the results of the laboratory test have the similar structure, which consists of test item, test value and unit. In comparison to the clinical practice, for each laboratory test subject, such as full blood count and liver function, there is an archetype that contains certain data items of that subject. There are over 200 laboratory test subjects in the CDR, while only 19 of which have been defined with archetypes. Given the low scale of archetype coverage to clinical concepts, it is necessary to use an archetype that has the common structure for the general concept of laboratory test results.There is a OBSERVATION.lab_test.v1 archetype in CKM as the basis for all the laboratory test archetypes. We define a specialized archetype OBSERVATION.lab_test_single.v1 with data items test item, result, and result unit to represent the general strucutre of the laboratory test results.

Representing archetype relationships

Most of the archetypes in CKM use archetype slots to represent relationships between archetypes, but it has significant limitations that it can only express relation between certain types of archetypes. For example, archetype slot is not allowed between Entry based archetypes. The semantics of archetype slot is the major obstacle in this paper, and we choose links to express relationships between archetypes. For example, among imaging examination archetypes in Figure 3c, to relationship ④, a link node Image series is added to OBSERVATION.imaging_exam_report.v1 to refer to OBSERVATION.imaging_exam_image_series.v1. For ③, a node Report identifier is added under node activities to (e) INSTRUCTION.request-imaging_exam.v1.1. For ⑤, a link node Examination requested is added to OBSERVATION.imaging_exam_image_series.v1. Although link is a general method to represent relationships between archetyps, the usage of the link is not well explained in openEHR specifications and there are few examples. As relationship is an important aspect in information modelling, flexible relationship representation can greatly facilitate the application of the archetype approach to clinical practice.

Conclusion

This paper is the first research that builds a CDR and develops a clinical application with the openEHR methodology in clinical practice in China. A 6-step archetype modeling method was proposed, which refers to the MOH standards and exisiting database schemas for collecting required data items in CDR. It provides an important lesson for including experiments with the openEHR approach adoption in China, which also facilitates the openEHR adoption.

Although the feasibility of the openEHR methodology has been verified by the case study, some limitations of openEHR when implemented in China have also been identified, including immaturity of archetype modification operations, the granularity of archetype, metadata-level versus data-level modelling and representation of archetype relationships. The findings of the case study will facilitate the openEHR adoption in China.

Acknowledgments

This work was supported by the National High-tech R&D Program of China (2012AA02A601) and National Science and Technology Major Project of China (2013ZX03005012).

References

[1] Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. Journal of the American Medical Informatics Association 2010; 17(2): 131-5.
[2] Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. AMIA Annual Symposium Proceedings 2012; 2012: 189-198.
[3] Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE–An integrated standards-based translational research informatics platform. AMIA Annual Symposium Proceedings 2009; 2009: 391-5.
[4] Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, Tao C, Parker CG, Haug PJ, Huff SM, Chute CG. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. Journal of biomedical informatics 2012; 45(4): 763-71.
[5] Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. AMIA summits on translational science proceedings 2010; 2010: 1-5.
[6] Beale T, Heard S. openEHR Architecture, Architecture Overview. Revision: 1.1. 2007.
[7] Beale T, Heard S, Kalra D, Lolyd D. The openEHR Reference Model, EHR Information Model. Revision: 5.1.1. 2008.
[8] Späth MB, Grimson J. Applying the archetype approach to the database of a biobank information management system. International journal of medical informatics 2011; 80(3): 205-26.
[9] Garde S, Hovenga E, Buck J, Knaup P. Expressing clinical data sets with openEHR archetypes: A solid basis for ubiquitous computing. International journal of medical informatics 2007; 76(3 Suppl): 334-41.
[10]Buck J, Garde S, Kohl CD., Knaup-Gregori P. Towards a comprehensive electronic patient record to support an innovative individual care concept for premature infants using the openEHR approach. International journal of medical informatics 2009; 78(8): 521-31.
[11]Zhang X, Yao Z. Modelling personal health record based on openEHR. Computer Applications and Software 2013; 30(5): 71-2.
[12]Li B. Discussion and research on the integration between SNOMED CT and opneER. Journal of medical informatics 2012; 33(1): 46-9.
[13]Yang Z, Liu D, Xu Y, Liang Y, Zhang H. The study of Clinical Information Model and Structured Clinical Data. Chinese journal of health informatics and management 2014; 11(1): 18-20.
[14]Xu W, Guan Z, Sun J, Wang Z, Geng Y. Development of an open metadata schema for prospective clinical research (openPCR) in China. Methods of information in medicine 2014; 53(1): 39-46.
[15]Beale T. Archetype Semantics. Revision: 0.5. 2007.

*Address for correspondence

Xudong Lu: E-mail address: lvxd@zju.edu.cn.