MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-161

Building a Logical EHR architecture based on ISO 13606 standard and Semantic Web Technologies

Marcelo R. Santos[a], Marcello P. Bax[a], Dipak Kalra[b]

[a] *Information Sciences School, Minas Gerais Federal University, Belo Horizonte, Brazil*
[b] *CHIME, University College London, United Kingdom*

Abstract

Among the existing patterns of EHR interoperability, the ISO 13606 standard is an important consideration. It is believed that the use of this norm, in conjunction with semantic technologies, may aid in the construction of a robust architecture, keeping in mind the challenges of semantic interoperability. The objective of this paper is to present a proposal for an EHR architecture, based on ISO 13606 and on the utilization of semantic technologies, for a real EHR scenario. In order to accomplish that, a real EHR scenario is described, as well as its main interoperability requirements and a candidate architecture is proposed to solve the presented challenges of interoperability. The ability of the ISO 13606 EHR reference model to accommodate the scenario was highlighted, together with the support provided by the use of the ontology specification languages – RDF and OWL – in respect to the maintenance of a controlled vocabulary.

Keywords:

Electronic Health Records, Semantic Interoperability, ISO 13606 standard, Two-level Modelling, Archetypes, EHR Service Architecture

Introduction

The issue of interoperability between EHR systems has become, in recent years, a theme of great relevance for the international clinical and informatics communities. The realization that the construction of an integrated EHR system will always involve, at some level, interoperation between different systems, underlines the relevance of interoperability patterns and of semantic technologies in the current context.

In order to understand this assertion, let us consider the creation of an integrated EHR service for a city, state or country. Although it is possible to develop a single EHR system, which fulfills the needs of every health institution in a city, this system will need to be integrated with those of other cities, which will then form the EHR service at the state or regional level. Likewise, it is possible to have a single EHR system for a given state, that will in turn be integrated to those of other states, in order to obtain as EHR service for the country. The same reasoning applies in the case of the obtainment of an EHR service in a global scale.

In this paper the name Logical EHR is given to this arrangement of different EHR systems capable of interoperating so as to offer an integrated service, preserving the existing semantics in the knowledge domain, updating clinical data for every patient in a consistent way, rather than ambiguously, all in accordance with the legal ethical regulations and obligations, following the safety and confidentiality patterns of information. The basic premises for the creation of a Logical EHR are the existence of interoperability patterns and the utilization of semantic technologies.

An example interoperability pattern, which may enable the creation of a logical EHR, is the ISO 13606 standard. It is an international standard published by ISO that specifies the information models and vocabularies needed for the interoperability of EHR systems. Initially developed by the European Committee for Standardization (CEN), with the denomination EN 13606 (*Electronic Health Record Communication*), and is a subset of the architecture proposed by the *openEHR*[1] Foundation [8][9][5]. In summary, it proposes a model of EHR statements that represents the complete or partial information of a patient's health record, extracted from an EHR system, for the purpose of communication to another EHR system or other requesting system.

Concerning semantic technologies, one might observe that, although the Internet offers sufficient flexibility and space in terms of connectivity for reaching interoperability, the techniques traditionally used for integration and interoperability at the application level involve only the use of data exchange formats whose success in terms of semantic has not been very significant. However, the use of semantic technologies in order to facilitate the integration and interoperability of EHR systems may bring significant benefits [4].

[1] 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

This paper aims at presenting a proposal for an EHR architecture, based on the ISO 13606 standard and on the utilization of semantic technologies, for a real EHR scenario.

ISO 13606 standard

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. It is a subset of the reference model proposed by *openEHR*. Based on the two-level modeling approach, it defines an information architecture to communicate part or the entire electronic health record of a given patient: preserving the original clinical meaning intended by the author; and reflecting the confidentiality of each data as intended by both author and patient [8].

The *two-level modelling* approach is based on the separation of knowledge and information levels in information systems [1][2]. The Information level represented by the reference model that are statements which apply to all entities of a class [1].The Knowledge level is represented by clinical archetypes that are statements about specific entities.

The reference model represents the global characteristics of components of an EHR, how they are assembled, and the information context required to meet the ethical, legal and originality requirements. It defines the following 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.

This reference model is complemented by clinical data structure definitions, known as archetypes, which represent the clinical concepts agreed upon within a community, with the objective of guaranteeing interoperability, consistency and data quality. An archetype is a formal definition of prescribed combinations of basic classes in the reference model for a specific clinical domain or for organizations, expressed in the form of data restriction. In the ISO 13606 standard they are defined in Part 2, supported by Part 1 but not mandatory [7].

The following may be quoted as benefits of the use of archetypes [6].

- Enable the formal definition of clinical content by domain experts without the need for technical understanding;
- Preserve the meaning of data by explicitly specified and well structured clinical content;
- Can safely evolve and thus deal with ever-changing health knowledge using a two-level approach;
- Can simplify the use of clinical terminologies.

Definitively the ISO 13606 contributes to the Logical EHR approach, as we can see at the Figure 1.

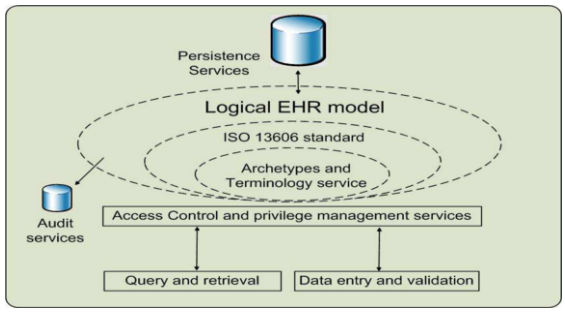

Figure 1 – Logical EHR architectural components

Semantic Technologies

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"[12].

The initiatives in the field of Semantic Web establish a set of protocols and technologies which promise to improve data categorization and association through an increase in the ability to create relationships and to generate inferences between systems and data. Adding intelligence to the traditional services offered by the web – based on Service Oriented Architecture (SOA) – and make stored data accessible by a set of semantic technologies [13].

In a typical interoperability scenario based on the SOA architecture, systems interact via Web Services using XML and SOAP messages on top of an HTTP network protocol [11]. The challenge of this architecture is to enable heterogeneous web applications, written in different platforms, to be integrated, exchanging data in a user-transparent way. This approach handles messages in an efficient way, but it does not consider the compatibility of the vocabulary being used and does not guarantee that the message recipient is able to comprehend its content [13]. This difficulty has led to W3C's[2] efforts to incorporate semantic technologies as part of its Semantic Web Services initiative.

Two W3C specifications are highlighted on the Semantic Technologies: Resource Description Framework (RDF) and Ontology Web Language (OWL). The RDF provides a framework to establish relationships among data elements, and the OWL improves the RDF with an ability to make constraints on different data elements and their relations with others. RDF and OWL can operate together depending on the scenario. Another important W3C specification is the Web Ontology Language Service specification (OWL-S) that provides a flexible framework for describing and initiating web services [13].

[2] World Wide Web Consortium – http://www.w3c.org

Materials and Methods

This study was carried out at the Department of Health for the State of Minas Gerais (SES-MG[3]), specifically within the scope of the electronic records of the Family Health Program. The Project was begun in the last quarter of 2008 and the main development phase to continue to 2011. Based on the results of grade tests which took place previously [10], the ISO 13606 standard was used as a starting point for the construction of the EHR statements and modeling of the archetypes for the central repository. The proposed architecture was developed from the direct participation of the author in the technical team of SES-MG and PRODEMGE and it was officially accepted in August 2009.

Firstly, the analysis of requirements for interoperability for the EHR project was performed. These were identified by means of brainstorm meetings and specific interviews involving the clinical team at the SES-MG – composed of ten members – at least fifteen technical visits to the health units in the state and research of previous projects and inside documents. These activities were undertaken in the initial 6 months of the project. At the same time, three members of the clinical team ran workshops with the clinical area participants of SES-MG in order to identify the data elements that should be part of the Patient Clinical Summary. After the architecture's approval, the archetype modeling process commenced. The archetypes were developed with the help of editor *linkEHR*[4] and the terminologies to be part of the terminology service were identified while modeling each of the archetypes.

Through an analysis of interoperability requirements and research on the utilization of semantic technologies for the creation of a terminology service, it was possible to elaborate the candidate architecture presented here.

Electronic Health Record Scenario

The Department of Health for the State of Minas Gerais, interested in the benefits brought about by the use of integrated clinical information, intends to create a state-wide EHR repository, with the objective of consolidating each citizen's demographic data and clinical summary. Initially the scope of the project will include only primary care centers in the Family Health Program. However, the objective of the government is to create the foundation for the inclusion of secondary and tertiary care units in the EHR repository.

The State of Minas Gerais is the largest in the country in number of cities, totaling, according to the Brazilian Institute of Geography and Statistics (IBGE[5]), 853 cities. For delivering primary care services, the State has over 5000 health units distributed among its cities, according to information from the SES-MG. The government's proposal is not to develop a single EHR system for the State to be imposed on the cities, conversely, it is to create macroregions and deploy EHR system contracted from the market to service each one. The solution is based on the creation of a central repository and on the creation of a message infra-structure aiming at the interoperability of the EHR systems to be contracted. The repository shall contain the demographic data and a clinical summary of the patients, necessary to support to the Family Health Program.

Each of the afore mentioned macroregions, composed of several health units, will use its own choice of EHR system and will have its own datacenter. The architecture must clearly establish the rules of data exchange so that the solutions access and provide data updates to the central repository, as well as prioritize the utilization of a controlled vocabulary.

Interoperability requirements

As a result from the data survey done with members of the SES-MG, a series of interoperability requirements were identified. A summary of the key requirements is outlined below.

A set of general requirements:

- The central repository should centralize the demographic data and each patient's clinical summary ;
- The EHR systems should feed the central repository and not interoperate directly with each other – EHR system integration should always occur via the central repository;
- Every time clinical data is recorded in the EHR system, the central repository should be updated;
- Before a patient is entered in the EHR system, the central repository should be checked. If the citizen already exists in the central repository, the centrally-held data should be copied to the requesting EHR system.

A set of specific requirements for the central repository:

- It should allow for the structuring of its data elements through a standardised reference model and archetypes;
- It should provide a mechanism which enables the recording and retrieval of each patient's clinical information;
- It should provide a mechanism which enables the recording and retrieval of each patient's demographic information;
- It should provide the EHR system with an authentication mechanism for the identification of the system and its operators;
- It should provide a repository of terminology services, which would serve as basis for the creation of the common vocabulary for the exchange of information with the EHR systems;
- It should provide auditing and reporting services.

[3] Department of Health for the State of Minas Gerais http://www.saude.mg.gov.br

[4] LinkEHR was developed by the Biomedical Informatics Group at the Technical University of Valencia, Spain. http://www.linkEHR.com

[5] Brazilian Institute of Geography and Statistics http://www.ibge.gov.br

Set of requirements for each EHR system:

- It should implement mechanisms for clinical and demographic data exchange with the central repository;
- It should implement mechanisms for the authentication and data security to ensure the data communication with the central repository;
- It should be able to exchange data with the central repository according to the terminology predicted by the terminology service;
- It should be able to exchange data with the central repository, respecting the restrictions defined by the archetypes, which represent the data elements from the central repository;
- It should be able to supply information for audits of any nature.

These requirements served as the basis for the design of the central repository and established key elements for the construction of the service architecture presented below.

EHR Services Architecture

The proposed architecture uses archetypes for the representation of its data elements and was designed with four basic services: clinical data services, demographic data services, archetype services and terminology services. See Figure 2.

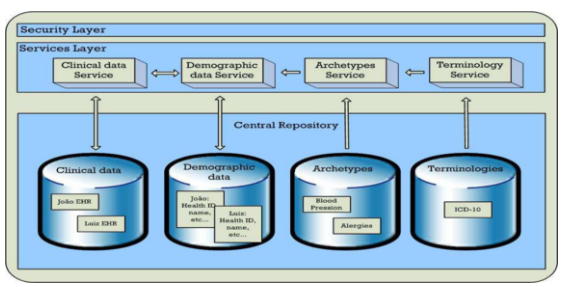

Figure 2 – EHR Services Architecture

Demographic data services are divided into: Demographic data inclusion service, alteration of demographic data service, user deactivation service, service of user consultation by code and service of user consultation with parameters. The Brazilian government has a published standard for sending and receiving data from the national health card. It was decided to utilize this standard, which is based on the utilization of Web Services, for exchanging messages in XML language.

The clinical data services were designed considering the utilization of EHR statements constructed from the reference model of the ISO 13606 standard and *datatypes* specified by the CEN/TS 14796 [3]. Two services were proposed: one for sending data to the central repository and another for enabling the retrieval of data from the central repository.

For sending data to the central repository, the EHR systems will have daily to generate the clinical statements (within *Compositions)* which took place for each patient at the health units, and update the central repository overnight. For that, they will have to consult the archetype structure through the archetype service available, and research the *Sections, Entries* and *Cluster/Elements* that will make up each *Composition.* For the update of the central repository, the statements shall be compacted according to a compacting algorithm yet to be established, and sent via FTP protocol. Since the architecture specifies the utilization of a repository with EHR version control, every time there is a correction or deletion in the clinical data, the EHR system shall send an update statement with the altered data, according to standard established by the ISO 13606 standard.

With regard to the retrieval of clinical data, this will only occur whenever a patient moves from one macroregion to another, or when the EHR system needs, for whatever reason, to update its local database. For that, the EHR system shall send a request to the central repository, and receive an XML file online with the current version of the Patient Clinical Summary (within a *Composition*). For this reason the architecture includes the design of two clinical databases: Patient Clinical Summary database and Clinical History database. The archetypes will include the rules to keep the each data element up to date at the Patient Clinical Summary database. Keeping these databases it will be possible also retrieve all *Compositions* of a Patient from the Clinical History database if it is required. In both cases, because of the version control process, only the most recent version of each object will be considered (if there have been corrections made to an original version).

The terminology service was designed to make it possible that all of the vocabulary content to be used by the EHR systems for integration with the central repository is made available through a unified source. Terminology such as ICD-10, procedure tables, medications, laboratory exams and terminology internal to the SES/MG will be available. The idea is to facilitate access to the terminology by the EHR system, promoting the development of the Logical EHR environment expected by the State. Furthermore, whenever the central repository receives EHR statements from the EHR software, there will be the need to perform syntax and semantic validations on the statement. Within this context, the use of semantic technologies, through the encoding of terminology in ontology representation languages such as RDF(s) and OWL will be a point of great relevance for the project. All terminology will be made available through the same service.

The archetype service was designed to make it possible for contracted EHR systems to request the versions of published archetypes. This is a critical point, for it will enable the gradual publication of new versions of archetypes, facilitating the deployment of the central repository in its initial phase. It will also facilitate the implementation of concepts which will be used later for other health care settings such as public hospitals.

Discussion

The utilization of a unified terminology server, based on the use of semantic technology, in conjunction with the single reference model published by the ISO 13606 standard will enable the establishment of a Logical EHR for the State. But that the work undertaken to date has shown that the architecture alone would not be sufficient to guarantee that the semantics present in the knowledge domain be, in fact, homogeneous. There will be the need for introduction of descriptive terminologies (e.g. SNOMED[6]) and monitoring on the part of the State's clinical team, in terms of audits focusing on the quality of information recorded in the central repository.

Despite the fact that the central repository is based on a reference model and on archetypes - which will allow for a robust and yet adaptable setting for the establishment of a Logical EHR environment and greater flexibility for the introduction of new concepts in the repository - the architecture needs to anticipate the participation of EHR systems that do not use the archetype approach. The EHR system need to, at least, be able to:

- identify the restrictions to which each data element must be submitted, through formal representation in ADL;
- create and interpret messages in XML according to the pattern presented by ISO 13606;
- validate the data elements to be sent to the central repository using the terminology service, by means of RDF and OWL schemes.

It is believed that the frequency of publication of new archetype versions by the SES-MG might be able to sparkle the interest of software suppliers in automatically interpreting the formal specifications predicted in the archetypes. This way, they might be able to obtain a cost reduction in the process of updating new clinical concepts and the terminology adoption in their EHR system.

Conclusion

This paper has made evident the importance of the Logical EHR concept, since the scenario of EHR systems utilization normally involves the integration between different software applications. It was possible to observe the adjustment of the two-level modeling approach, which was the basis for the construction of the proposed architecture, regarding the level of information represented by the reference model, as well as the level of knowledge represented by the archetypes. These allowed for the mapping of all the data elements that constitute the central repository and provided a greater flexibility to the proposed architecture. From the terminologies point of view, the semantics technologies propitiated a favorable setting for the use of a controlled vocabulary, increasing the chances of preserving the semantics of the knowledge domain. It was verified that the reference model for the ISO 13606 standard was appropriate for fulfilling the requirements of interoperability of the case researched, and its statement (*Entry*) model was enough to fulfill the EHR specificities of a Brazilian State. Furthermore, it allowed the participation of EHR systems that do not use the archetype approach.

Acknowledgments

We would like to thank FAPEMIG that funded the travel costs for participation in the MEDINFO 2010 congress.

References

[1] Beale T. Archetypes: Constraint-based Domain Models for Future-proof Information Systems. OOPSLA 2002 workshop on behavioural semantics, 2002.

[2] Bax MP; Augsten, Eduardo. A Two-level Modeling and Implementation Approach for CDA Using Plone/Archetypes. In: 2nd Int Conf on the CDA, Acapulco, 2004.

[3] CEN TS 14796. Health informatics - Data types. Technical Specification TS 14796, European Committee for Standardization. Brussels, Belgium. 2003.

[4] Fernandes-Breis JT, Vivancos-Vicente PJ, Menárgues-Totosa M, Moner D, Maldonado JA, Valencia-Garcia R, Miranda-Mena TG. Using semantic technologies to promote interoperability between electronic healthcare records' information models. Proceedings of the 28th IEEE, 2006.

[5] Garde S; Knaup P; Hovena EJS; Heard S. Towards Semantic Interoperability for Electronic Health Records Domain Knowledge Governance for openEHR Archetypes. Schattauer GmbH, 2007.

[6] Garde S, Hovengal E, Buck J, Knaup P. Expressing clinical data sets with openEHR archetypes: a solid basis for ubiquitous computing. Int J of Med Informatics, 2007.

[7] ISO/TC251 13606 Health informatics - Electronic record communication - Part 1: Reference Model and Part 2: Archetype interchange. ISO, 2008.

[8] Kalra, D. Electronic Health Record Standards. IMIA, 2006.

[9] Kalra D. Barriers, approaches and research priorities for semantic interoperability in support of clinical care delivery. Semantic Health Project, 2008.

[10] Munoz A, Somolinos R, Pascual M, Fragua JA, González MA, Moteagudo JL, Salvador CH. Proof of concept Design and Development of an EN13606-based EHR Service. J of the Am Medical Informatics, 2007.

[11] W3C working Group Note. Webservices Architecture. W3C Working Group note 11, Fev. 2004.

[12] Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific American, May 2001.

[13] SICOP. Introducing Semantic Technologies and the Vision of the Semantic Web. - Semantic Interoperability Community of Practice. White Paper Series Module 1, 2005.

Address for correspondence

Marcelo Rodrigues dos Santos, Information Sciences School - UFMG, Brazil. marrsantos@ufmg.eci.br, marrsantos@terra.com.br

[6] SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms) http://www.ihtsdo.org/snomed-ct/