Health Informatics: The Relevance of Open Source and Multilevel Modeling

Luciana T. Cavalini[1] and Timothy W. Cook[2]

[1] Institute of Community Health, Fluminense Federal University (UFF)
Rua Marquês de Paraná, 303 - 3rd Floor Annex HUAP
Niterói, RJ – Brazil 24220-331
`lutricav@vm.uff.br`
[2] National Institute of Science and Tecnology – Medicine Assisted by Scientific Computing (INCT-MACC)
`timothywayne.cook@gmail.com`
`http://www.mlhim.org`

Abstract. Health information features significant spatial-temporal and domain complexities, which brings challenges to the implementation of patient-centered, interoperable and semantically coherent healthcare information systems. This position paper supports the idea that the multilevel modeling approach is essential to ensure interoperability at the semantic level, but true interoperability is only achieved by the adoption of open standards, and open source implementations are needed for promote competition based on software quality. The Multilevel Healthcare Information Modelling (MLHIM) specifications are presented as the fully open source multilevel modeling reference implementation, and best practices for the development of multilevel-based open source healthcare applications are suggested.

Keywords: Health informatics, open source software, multilevel modeling.

1 Introduction

The information related to human health is inherently complex at the intersections of space, time and knowledge [17]. This complexity precludes the feasibility of one, single, all encompassing electronic health record for individuals in a population, since various pieces of information from a number of different applications may be needed at any point in time [21]. Semantic interoperability is crucial in recording information in purpose specific applications that need to synchronize to larger databases. This is the way to provide useful information from the point of care to the healthcare services in a timely manner [15].

Regarding public health, the typical turnaround time now for acquiring the healthcare status of a population can be weeks or even months while experts pour over the data and try to merge pieces from various applications and paper forms into something meaningful [14]. By basing applications on a common information model and using a constraint based approach to define the knowledge components, we can achieve semantic interoperability and near real-time information regarding the

healthcare status of a population and its individuals so that faster action can be taken to meet the needs in that area. This approach also helps empowering application developers at the local level, so they can develop healthcare applications fitted to the very specific local needs without losing semantic coherence and interoperability with other local services and with the highest levels of the healthcare systems [11].

Healthcare inequalities will not be easily solved by adopting 21st Century healthcare based on 20th Century information systems [16]. There are no remaining obstacles related to hardware, including mobile computing and pervasive medicine, but software based on traditional data models are not fitted to deal with the significant spatial and temporal complexities of healthcare information [4].

That is the case because health information systems based on traditional data models are not interoperable and have high maintenance costs. These problems have a significant negative impact on the use of these systems to the emerging situations and dynamics currently found in healthcare [5,25]. In fact, the development of healthcare applications is a complex challenge, especially given the large number of concepts in constant evolution, which makes it difficult to reach a consensus on any concept [9].

Some solutions to these problems have been proposed over the past two decades, such as the work of Yoder et al. [35]. However, the solution most fitted to the specific features of healthcare information involves the separation between domain model and persistence of data. This multilevel modeling approach proposes the definition of at least two levels: the Reference Model, which defines generic types of data and data structures and a Domain Model, defined by restrictions on the Reference Model [19].

Health information systems based on multilevel modeling are more easily interoperable and can be implemented on any hardware. The adoption of a common Reference Model and a Domain Model for different systems allows a transparent and shareable interface with geographic information systems and statistical analysis tools that can analyze information collected from various remote systems [22].

Nowadays, multilevel modeling specifications for healthcare information systems are openly available and proven in software. Based on these specifications, it is possible to develop healthcare applications centered on the citizen, with the capability of recording longitudinal data [12].

Furthermore, decision support systems and standardized reports can be implemented in the systems and still ensuring semantic interoperability at any level, since the development of algorithms for decision support based on a common domain model allows the reuse of decision rules in different implementations. Thus, in the point and time of care, control measures can be implemented immediately, allowing for greater effectiveness of healthcare and, at the governance level, larger areas can be monitored and priority areas can be identified for intervention [2].

However, despite its technical advantages, multilevel modeling-based solutions have not been widely implemented in real healthcare settings, except for some few academic projects [6,20,24].

There is one aspect that is essential but seldom addressed regarding interoperability of healthcare information systems, which is related to the modality of software licensing. In fact, the general business model of proprietary software companies may be considered unfriendly to interoperability, since the competition between companies has the goal to establish hegemony or monopoly, in order to concentrate capital, and that is based on the secrecy of the software source code. Actually, one can state there

is no proven interoperability without the development of, at least, open specifications, since it is necessary for one system to be compliant to the other system's features that are related to interoperability, and that can only be attained if the systems were developed based on a common set of specifications. Thus, expanding this argument, it is possible to deduce that a complete condition of interoperability between all systems is only possible if they share a common set of specifications, at least at the level of data extract exchange. Therefore, full interoperability requires open standards and open source software [1,8].

This business model has not shown any differences when applied to healthcare. Actually, it is stated that open standards facilitate competition between open source software and proprietary software, since it allows the competition between different implementations of the same specification [30]. This is a key issue related to software quality, which is crucial in healthcare, since the quality of the software is directly related to the quality of care [2].

Taking into account the centrality of open source and multilevel modeling to ensure the development of high quality, citizen-centered, interoperable, semantically coherent health information systems, our objective is to describe the essential features of a open specification for multilevel healthcare information modeling, and to propose a set of best practices for the development of healthcare applications based on those specifications.

2   Method

2.1   Summary of the Specifications

The "Multilevel Healthcare Information Modelling" (MLHIM) specifications are a fully open set of specifications for the development of health information systems based on multilevel modeling. The MLHIM documentation is published under the Open Document Format (ODF) at http://www.mlhim.org.

The technological choices for the development of the MLHIM specifications were made because of the distributed and diverse nature of healthcare information systems; thus, its goal is interoperability and standardization is the path. The basis of MLHIM are the dual-level openEHR specifications [3] and the healthcare-specific data types as defined by the ISO 21090 standard. These specifications and standards are articulated in a single specification, with the specific purpose of creating a path for semantic interoperability among different health applications, including legacy systems.

In the MLHIM specifications, the classes of the Reference Model are persistent and should be kept as stable as possible over time. In the Domain Model, the Constraint Definitions on the Reference Model provide the semantic interpretation of the objects stored by the Reference Model.

The idea behind the multilevel modeling is that changes in structure and rules of inference are reflected on the Constraint Definitions and not on the Reference Model. Thus, change requests on the persistence mechanisms of the information systems are reduced. Furthermore, the Constraint Definitions are created and edited by domain experts and not by computer scientists, which avoids the need for interpretation of the

knowledge extracted from an *ad hoc* interaction. Once the domain expert is responsible for modeling the knowledge, concepts are thoroughly and accurately expressed as Concept Constraint Definitions (CCDs).

This approach is compliant to the following standards developed by the International Organization for Standardization (ISO):

- ISO/TS 18308:2004 - Health informatics - Requirements for an electronic health record architecture;
- ISO/TR 20514:2005 - Electronic health record - Definition, scope and context;
- ISO 13606-1:2008 - Health informatics - Electronic health record communication - Part 1: Reference model;
- ISO 13606-2:2008 - Health informatics - Electronic health record communication - Part 2: Archetype interchange specification;
- ISO 13606-3:2009 - Health informatics - Electronic health record communication - Part 3: Reference archetypes and term lists;
- ISO/TS 13606-4:2009 - Health informatics - Electronic health record communication - Part 4: Security;
- ISO 13606-5:2010 - Health informatics - Electronic health record communication - Part 5: Interface specification;
- ISO/FDIS 21090:2011 - Health informatics - Harmonized data types for information interchange;

2.2 Knowledge Modeling

The MLHIM specifications adopt XML Schema Documents (XSDs) for the elaboration of the Concept Constraint Definitions (CCDs). A CCD is a XSD file that expresses a defined healthcare concept. This concept is expressed on the CCD as constraint definitions on the MLHIM Reference Model.

XML is regarded as the most widely adopted solution to system interoperability and semantic coherence; therefore, in order to fit its purposes, XML must be open source [26]. There are some recent publications (since 2005) about the use of XML Schema languages as an attempt to perform *a posteriori* standardization of data types and metadata [28,32], development of templates for structured documents [27,35], or any combination of the techniques cited above [10,13,29], what can be understood as solutions to promote data interchange between one-level based information systems.

Some of those studies adopt the concept of the domain expert as the author of the knowledge modeling [18]. Although that approach solves some of the semantic loss derived from *ad hoc* interactions between the domain expert and the system developer, it still does not ensure interoperability and semantic coherence for the attempts of data interchange between one-level model applications.

On the other hand, the use of XML W3C Schemas for knowledge modeling based on the dual-model ISO 13606 standard was tested and validated by Rinner *et al.* [31], allowing semantic validation of knowledge components, conditional to the definition of a "fully generic validation" provided by the Reference Model. On the other hand, the authors describe the technical difficulties regarding the specific transformations (or constraints) on the ISO 13606 Reference Model classes that are needed to express a given healthcare concept. That suggests the need for a knowledge modeling editor, which is a common concern for multilevel modeling projects [23,33].

In the MLHIM specifications, a Constraint Definition Designer has been developed, using Mind Maps, which are proven efficient as concept definition tools [7]. XMind (XMind Ltd.), an open source Mind Map editor, was used to build the CCD template for the MLHIM Constraint Definition Designer. In this template, instead of using the Mind Map nodes to directly constraint the domain concepts, they were defined as classes of the MLHIM Reference Model, being the constraint definitions applied by the specific arrangement of some classes required for a specific concept, and by defining constraints on the attributes of those classes, expressed as sub-items of a given Mind Map node. The resulting XMind file should be transformed into a XSD file, which is the CCD for a particular healthcare concept and it can be validated against the MLHIM Reference Model. Furthermore, it can be combined to other CCDs to inform the development of Graphic User Interfaces for MLHIM-based applications. The MLHIM Constraint Definition Designer is available at https://launchpad.net/cdd.

3   Application Development

This section proposes a set of best practices for the multilevel modeling-based application development. The underlying reasoning expressed here is that interoperability and semantic coherence are ensured by open specifications based on multilevel modeling, such as the MLHIM specifications, which include a generic, stable, standard-compliant Reference Model and the rules for the Concept Constraint Definitions on the Reference Model. In order to increase the probability of building CCDs that are valid against the MLHIM Reference Model, the Constraint Definition Designer was devised as a CCD editing tool.

Any other particular feature such as the combination of CCDs in templates, the definition of GUI, the choice of the object-oriented programming language and the correspondent application framework, the choice of the No-SQL or object-oriented database to persist the data and the query algorithms are considered as implementation choices and do not interfere on the technical aspects regarding interoperability and semantic coherence, already addressed in a comprehensive manner by the MLHIM specifications.

However, in order to guarantee that the interoperability and the semantic coherence ensured by the specifications will be attained by multilevel modeling-based systems, it is necessary to develop high quality applications. Given this reasoning, a non exhaustive set of best practices for healthcare application development based on multilevel modeling is presented below.

3.1   Application Framework and Data Persistence

The MLHIM specifications provide the Open Source Health Information Platform (OSHIP) (https://launchpad.net/oship), a open source implementation of the MLHIM specifications. OSHIP is implemented in Eclipse Modeling Framework (EMF) (The Eclipse Foundation), which allows code export to the main object-oriented languages, such as Java, Python and Ruby. This raw code can be wrapped into many application development frameworks based on those programming languages, allowing a wider adoption of the specifications, since the application framework is not an additional learning curve for the developers.

Usually, the chosen application framework will guide the choice of the type of database to be adopted for the persistence of data. It is important to notice that, since the MLHIM specifications are object-oriented, the more obvious choice would be an object-oriented database for data persistence. However, object-oriented databases may present performance or query issues. Taking that technical difficulty into account, No-SQL or hierarchical databases can be chosen instead, in order to circumvent those technical complexities in real life applications, although the persistence of data origined from MLHIM-based applications is trivial if an object-oriented database is chosen.

3.2 Communication Layer

The communication layer is proposed to be built on any software component based on distributed technology, available in object-oriented programming languages, running on distributed or wireless communication networks. The association between the layers of modeling and data communication can be done through a model-driven approach. In this approach, the overall architecture of the system will be specified at a high level, using an Architecture Description Language (ADL), implicitly or explicitly annotated with the CCDs used in the data layer, allowing the development of the Graphical User Interfaces, which are likely to be persisted in a database, and which are communicated through the network from each system component. This specification will therefore be the basis for the generation of code for different parts of the application.

The proposed ADL library to be adopted is the ADL Acme (http://www.cs.cmu.edu/~acme/). The library AcmeLib can be used as a basis for building code generators for different parts of the application.

3.3 System Integration

The adoption of multilevel modeling for the development of healthcare applications brings a great deal of flexibility for application developers. There is no need for the development of a monolithic Electronic Medical Record (EMR) for the entire healthcare setting, irregardless of its size and complexity (primary care, outpatient clinic, hospital). Multilevel modeling allows the development of purpose-specific applications, no matter how restricted the data is (which includes applications for specialized scientific research); the data can be exchanged from any application to any other, and the data extracts of both are still valid. However, some institutions might require a higher level of system integration due to the specificities of its particular workflow.

In order to allow system integration, the use of a Service Oriented Architecture (SOA) is proposed. SOA allows better integration between different languages and platforms, which is necessary since MLHIM-based applications can be developed on any object-oriented language. Additionally, SOA makes easier the management of scalability, reusability, distribution and storage of applications. The development of this application integration is proposed by the adoption of the Representational State Transfer (REST) architectural style. In order to achieve system integration via REST, it is proposed the use open source libraries for the development of clients and servers

based on REST with portability for the main object-oriented languages (e.g., Restfulie, RIP). Those libraries allow code breaking so that it creates a very flexible service that is able to evolve with minimal change on the client side. Thus, it is possible to integrate different applications developed in different languages and platforms in a cost-effective way, reducing risks and costs associated with traditional system integration tasks.

3.4   Decision Support

The C Language Inference Processing System (CLIPS) is proposed as the inference mechanism for the development of decision support algorithms in MLHIM-based applications. CLIPS provides important benefits for the development of decision support engines in healthcare systems for the following reasons: (a) it supports a forward-chaining (or data-driven) mode of processing inference rules, which means that whenever new data become available, all available states of the system are checked again, and (b) it provides a powerful applications programming interface in C/C++, which allows, for the processing of a rule, the addition of routines that are interactive with the user and the management of dependencies between the actions of dynamic control. Both features are essential for modeling rules for dynamic scenarios such as healthcare. In addition, CLIPS supports multiple persistence layers, allowing a fully object-oriented system design, essential for information systems based on multilevel modeling.

3.5   Data Aggregation and Business Intelligence

It is proposed that the preparation of reports based on aggregated data from the local level to regional and national levels be based on the Statistical Data and Metadata Exchange - Health Domain (SDMX-HD) specifications, which are being developed under the auspices of the World Health Organization (WHO) for the standardization of aggregate data formats, in order to facilitate the exchange of statistical measurements and health indicator definitions. These specifications are inspiring some of the requirements of ISO 14639 - eHealth Architecture (now at the stage of Draft Technical Report).

For the execution of Business Inteligence (BI) tasks, open source solutions such as the Pentaho BI tool (http://www.pentaho.com) (Pentaho Corporation) are suggested for the preparation of pre-defined management reports, monitoring reports, custom data analysis, data aggregation and formatting data for export to legate management systems required by national and regional healthcare authorities that are not compliant to the SDMX-HD specifications.

4   Conclusions

The issues regarding interoperability and semantic coherence are more relevant for health information systems than for any other economic sector of the society. That is so due to the need for the maintenance of the citizen's longitudinal health record all through his life. However, the extreme conceptual, spatial and temporal dynamics of

the healthcare activities require a high level of diversity between information systems for different healthcare settings and purposes.

Traditional one-level data model applications, which fit the needs of almost any other economic activity of human society are being used for 45 years in the healthcare sector and have not been able to provide a citizen-centered, interoperable and semantically coherent health record.

Over the last 20 years, multilevel modeling specifications have been developed in order to address those important issues. Over that development process, it became evident that true interoperability will only be achieved if the multilevel modeling specifications were openly available. The implementation of the multilevel specifications in open source software has the potential to increase the competition for the development of good quality software, which is critical in the healthcare sector.

This paper presented the state of the art of the open source multilevel modeling that are currently available, which demonstrates the practical possibility of development of open source healthcare applications based on multilevel modeling. Thus, by contributing to those projects, the open source community can help improving the quality of healthcare on a global basis.

References

1. Almeida, F., Oliveira, J., Cruz, J.: Open standards and open source: enabling interoperability. Int. J. Soft. Eng. App. 2(1) (2011), doi:10.5121/ijsea.2011.2101
2. Ammenwerth, E., Shaw, N.: Bad informatics can kill: is evaluation the answer? Methods Inf. Med. 44(1), 1–3 (2005)
3. Barretto, S.A., Warren, J., Goodchild, A., Bird, L., Heard, S., Stumptner, M.: Linking guidelines to Electronic Health Record design for improved chronic disease management. In: AMIA Annu. Symp. Proc., pp. 66–70 (2003)
4. Beale, T., Heard, S.: openEHR Architecture overview. openEHR Foundation, London (2008)
5. Blobel, B.: Comparing concepts for electronic health record architectures. Stud. Health Technol. Inform. 90, 209–214 (2002)
6. Cantiello, J., Cortelyou-Ward, K.H.: The American Recovery and Reinvestment Act: lessons learned from physicians who have gone electronic. Health Care Manag (Frederick) 29(4), 332–338 (2010)
7. Chen, R., Klein, G.: The openEHR Java reference implementation project. Stud. Health Technol. Inform. 129(Pt 1), 58–62 (2007)
8. D'Antoni, A.V., Zipp, G.P., Olson, V.G., Cahill, T.F.: Does the mind map learning strategy facilitate information retrieval and critical thinking in medical students? BMC Med. Educ. 10, 61 (2010)
9. Dutton, W.: Key enablers for eTransformation? eID, interoperability and open source. Eur. J. ePractice 6, 2 (2009)
10. Eccles, M., Mason, J.: How to develop cost-conscious guidelines. Health Technol. Assess. 5(16), 1–69 (2001)
11. Gao, S., Mioc, D., Yi, X., Anton, F., Oldfield, E., Coleman, D.J.: Towards web-based representation and processing of health information. Int. J. Health Geographics 8, 3 (2009)
12. Garde, S., Chen, R., Leslie, H., Beale, T., McNicoll, I., Heard, S.: Archetype-based knowledge management for semantic interoperability of electronic health records. Stud. Health Technol. Inform. 150, 1007–1011 (2009)

13. Garde, S., Knaup, P., Hovenga, E., Heard, S.: Towards semantic interoperability for electronic health records. Methods Inf. Med. 46(3), 332–343 (2007)
14. Hägglund, M., Scandurra, I., Moström, D., Koch, S.: Bridging the gap: a virtual health record for integrated home care. Int. J. Integr. Care 7, 26 (2007)
15. Hammond, W.E., Bailey, C., Boucher, P., Spohr, M., Whitaker, P.: Connecting information to improve health. Health Aff. (Millwood) 29(2), 284–288 (2010)
16. Haughton, J.: Look up: the right EHR may be in the cloud. Major advantages include interoperability and flexibility. Health Manag. Technol. 32(2), 52 (2011)
17. Haux, R.: Medical informatics: past, present, future. Int. J. Med. Inform. 79(9), 599–610 (2010)
18. Hudson, D.L.: Cohen ME. Uncertainty and complexity in personal health records. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 6773–6776 (2010)
19. Hulse, N.C., Rocha, R.A., Del Fiol, G., Bradshaw, R.D., Hanna, T.P., Roemer, L.K.: KAT: A flexible XML-based knowledge authoring environment. J. Am. Med. Inform. Assoc. 12, 418–430 (2005)
20. Kalra, D., Beale, T., Heard, S.: The openEHR Foundation. Stud. Health Technol. Inform. 115, 153–173 (2005)
21. Kashfi, H.: An openEHR-based clinical decision support system: a case study. Stud. Health Technol. Inform. 150, 348 (2009)
22. Kelley, J.: The interoperability hang-up. When it comes to information exchange, how should precede what. Health Manag. Technol. 32(2), 32–34 (2011)
23. Kohl, C.D., Garde, S., Knaup, P.: Facilitating secondary use of medical data by using openEHR archetypes. Stud. Health Technol. Inform. 160(Pt 2), 1117–1121 (2010)
24. Maldonado, J.A., Moner, D., Bosca, D., Fernandez-Breis, J.T., Angulo, C., Robles, M.: LinkEHR-Ed: a multi-reference model archetype editor based on formal semantics. Int. J. Med. Inform. 78(8), 559–570 (2008)
25. Martinez-Costa, C., Menarguez-Tortosa, M., Fernandez-Breis, J.T.: An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. J. Biomed. Inform. 43(5), 736–746 (2010)
26. Michelsen, L., Pedersen, S.S., Tilma, H.B., Andersen, S.K.: Comparing different approaches to two-level modelling of electronic health records. Stud. Health Technol. Inform. 116, 113–118 (2005)
27. Neeser, A.E.: XML: The open source solution to interoperability. Open Lib. Class J. 1(2) (2009), `http://www.infosherpas.com/ojs/index.php/openandlibraries/article/view/31/40` (accessed on March 6, 2011)
28. Norlin, C., Kerr, L.M., Rocha, R.A.: Using clinical questions to structure the content of a web-based information resource for primary care physicians. In: AMIA Symp. Proc., pp. 482–486 (2009)
29. Paterson, T., Law, A.: An XML transfer schema for exchange of genomic and genetic mapping data: implementation as a web service in a Taverna workflow. BMC Bioinformatics 10, 252 (2009)
30. Qian, Y., Tchuvatkina, O., Spidlen, J., Wilkinson, P., Gasparetto, M., Jones, A.R., Manion, F.J., Scheuermann, R.H., Sekaly, R.P., Brinkman, R.R.: FuGEFlow: data model and markup language for flow cytometry. BMC Bioinformatics 10, 184 (2009)
31. Reynolds, C.J., Wyatt, J.C.: Open Source, open standards, and health care information systems. J. Med. Internet Res. 13(1), e24 (2011)

32. Rinner, C., Janzek-Hawlat, S., Sibinovic, S., Duftschmid, G.: Semantic validation of standard-based electronic health record documents with W3C XML schema. Methods Inf. Med. 49(3), 271–280 (2010)
33. Seibel, P.N., Krüger, J., Hartmeier, S., Schwarzer, K., Löwenthal, K., Mersch, H., Dandekar, T., Giegerich, R.: XML Schemas for common bioinformatic data types and their application in workflow systems. BMC Bioinformatics 7, 490 (2006)
34. Sundvall, E., Qamar, R., Nystrom, M., Forss, M., Petersson, H., Karlsson, D., Ahlfeldt, H., Rector, A.: Integration of tools for binding archetypes to SNOMED CT. BMC Med. Inform. Decis. Mak. 8(suppl. 1), S7 (2008)
35. Yoder, J.W., Balaguer, F., Johnson, R.: Architecture and design of adaptive object-models, `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.9077&rep=rep1&type=pdf`
36. Zhao, L., Lee, K.P., Hu, J.: Generating XML Schemas for DICOM structured reporting templates. J. Am. Med. Inform. Assoc. 12, 72–83 (2005)