Canonical document for medical data exchange

Sanae Mazouz[1], Ouçamah Mohammed Cherkaoui Malki[1], El Habib Nfaoui[1]
[1]Department of Computer Science, Faculty of Science Dhar El Mahraz, Fez, Morocco

Abstract. In the healthcare domain, the number of heterogeneous standards and medical applications is large. In consequence, the decision about which system and standard are more appropriate is difficult. The most widely deployed healthcare standard is Health Level Seven. Unfortunately, this standard has problems. The version 3 has seen slow adoption and the version 2.x had interoperability problems because of the variety of implementations by healthcare providers. To reduce the health interoperability problem between health information systems, we propose a canonical document for exchanging medical information. This document will be used as mediator between healthcare systems and shall allow health providers to share data among healthcare institutions without any adjustment or requirements. As result, the proposed mediator can make the exchange of health data simpler and efficient.

Keywords: Health information system; Interoperability; Health Level Seven; Canonical document; Mediator.

1 Introduction

Exchanging information across healthcare stakeholders is a hard task. The existence of various Health Information Systems (HIS) [5] makes the exchange of Electronic Health Record (EHR) [3] very difficult. Many standards are created to enable the exchange of patient medical information such as HL7 [1], openEHR [4], DICOM [2], etc. The most deployed healthcare standard is Health Level Seven (HL7), adopted by more than 55 countries around the world. There are two major HL7 versions, HL7 v2.x and HL7 v3. The version 3 introduces new technologies like Extensible Markup Language1 (XML) and Unified Modeling Language2 (UML) languages, which makes it more extensible and open. However, most medical applications use HL7 v2.x. The reason is the incompatibility between HL7 v2.x versions and the large number of

1 http://www.w3.org/XML/

2 http://www.uml.org/

implementations. Also, the use of standards requires changing the structure of existing patient's records.
Several countries have created and developed their own EHR for example: France "DMP", Canada "Health Infoway", Taiwan "TMT", UK "NHS", and Australia "NHETA"; and they are engaged to standardize a national Electronic Health Record. These efforts did not attempt the objective.
Therefore, there is an important need to communicate all systems through a simple and easy solution without any requirements or adjustment.
In this context, we propose a canonical document used as a mediator between healthcare systems. Our objective is to provide a health document mediator for exchanging health information across healthcare systems using XML language to encode the documents. The structure of the canonical document is simple and can be read by a non-expert. The mediator does not impose any training or adjustment on the system architecture. The use of the canonical document mediator will reduce the number of interfaces between HIS from [n×(n-1)/2] interfaces to (n) interfaces.
This paper summarizes some standards to deal with the problem of interoperability in healthcare domain and explains the mediator structure.
The paper is structured as follows. In the next section, we summarize some standards used to build interoperability into healthcare domain. Section 3 contains a review of related works. In section 4, we describe the structure of the mediator. Section 5 gives some experimental. Finally, section 6 concludes our paper.

2 Standards for Healthcare Interoperability

Aguilar [10] defines interoperability as the ability of two or more systems or components to exchange information and to use the information that has been exchanged. Interoperability depends upon two important concepts: syntax and semantics.

- Syntactic interoperability: is the ability of systems to communicate the structure of information basing on rules for spelling and grammar.
- Semantic interoperability: is the ability for information shared by systems to be correctly interpreted on the receiving end.

2.1 Standards

Standards for healthcare were created by a variety of organizations for various types or categories of interoperability in HIS. These standards allow HIS to communicate in the same way across system. Below is a summary of key standards at the syntactic and semantic level.

• **Health Level 7 (HL7)** [1] - is a standard for the exchange of data between healthcare applications. There are two major HL7 Versions, HL7 V2.x and HL7 V3. The Version 3 introduces a new approach to clinical information exchange: the Clinical Document Architecture (CDA), the Reference Information Model (RIM), and Clinical Context Object Workgroup (CCOW). The CDA is a document markup standard that specifies the structure and semantics of clinical documents. However, even with the innovations of the version 3 it has a seen slow adoption. In addition, the version 2.x had interoperability problems because of the variety of implementations by healthcare providers.

• **Digital Imaging and Communications in Medicine (DICOM)** [2] - is an international standard for the communication of medical images in radiology, cardiology, dentistry, and pathology. Developed by the DICOM Standards Committee and under the umbrella of National Electrical Manufacturers Association (NEMA).

• **International Statistical Classification of Diseases (ICDx)** [7] - is an international standard for epidemiology, health management, and clinical purposes. Used to identify diseases, signs, symptoms, abnormal findings, complaints, and social circumstances for billing purposes list by the World Health Organization (WHO)[3].

• **Clinical Context Object Workgroup (CCOW)** [11] – is a standard for providing comprehensive view and single sign-on capability across systems without integrating databases. CCOW specify technology-neutral architectures, component interfaces, and data definitions as well as an array of interoperable technology-specific mappings of these architectures, interfaces, and definitions. It is an independent vendor developed by the HL7 organization

• **OpenEHR** [4] - is an open international standard specification in health informatics describing the health data in EHRs.

• **Logical Observation Identifiers Names and Codes (LOINC)** [6] - is a universal standard for identifying individual laboratory results and clinical observations. It facilitates the exchange of test results for clinical care, healthcare management, and research.

• **National Council for Prescription Drug Programs (NCPDP)** [9] - is a standard for transmitting prescription requests and fulfillment from pharmacies to payers.

• **Clinical of Care Document (CCD)** [11] - is a standard for specifying the encoding, structure, and semantics of a patient summary clinical document for exchange. CCD allows physicians to send electronic medical information to other providers without loss of meaning and enabling improvement of patient care. CCD is a US version of CDA.

[3] www.who.int

2.2 Classification of healthcare standards

In healthcare domain, a number of standards were created to address the requirements of interoperability problems at both semantic and syntactic layer. These standards are organized into six categories [12]. Table 1 provides a complete classification of these standards.

Table 1. Classification of healthcare standards

Functional and Syntactic level			Semantic level		
Messaging standards	Terminology standards	Document standards	Conceptual standards	Application standards	Architecture standards
HL7 DICOM NCPDP	LOINC ICDx	CDA CCD	RIM	CCOW	OpenEHR

- **Messaging standards** – outline the structure, content and data requirements of electronic messages to enable the effective and accurate sharing of information.
- **Terminology standards** – provide specific codes for terminologies and classifications for clinical concepts such as diseases, allergies and medications. Terminology systems assign a unique code or value to a specific disease or entity.
- **Document Standards** – indicate the type of information included in a document and also the location of the information.
- Conceptual standards – allow the transmission of information between systems without losing meaning and context.
- **Application standards** – determine the implementation rules for software systems to interact with each other. For example, application standards using single sign-on allow users to logs into multiple information systems within the same environment.
- **Architecture standards** – define a generic model for health information systems. They allow the integration of health information systems by providing guidance to aid the planning and design of new systems and also the integration of existing systems.

3 Related works

Many pervious works have proposed solution to promote interoperability in healthcare domain. Bicer et al [13] proposed semantic mediation of exchanged messages. He demonstrates how to mediate between HL7 V2.x and HL7 V3

messages. First, messages exchanged in the healthcare domain are in EDI (Electronic Data Interchange) or XML format. They are transformed into OWL (Web Ontology Language) ontology instances. Then they are mediated through an ontology-mapping tool named OWLmt, which is used to reason over the source ontology instances while generating the target ontology instances according to the graphically defined mapping patterns. Lopez et al [15] proposed a framework as a set of principles and guidelines as well as methodologies and techniques for realizing semantic interoperability in Health Information Systems using Rational Unified Process (RUP) and formal software processes engineering methods. To achieve this objective, he analyzed approaches for information systems architecture and he harmonized them towards the framework. Jian et al [14] reported a national level standard called Taiwan Electronic Medical Record Template (TMT) that aims to achieve semantic interoperability in EHR exchanges. The TMT provides a basis for building a portable, interoperable information infrastructure for EHR exchange in Taiwan.
These projects are limited to a specific use. Therefore, the need to build a new solution for general use and without any requirement or adjustment is an urgent priority.

4 System description

Traditional information systems reach a functional interoperability but not semantic. This means that the information arrives at its destination, but it is not understood. To reach semantic interoperability, terminological references are indispensable [16].
Our mediator includes these terminological standards such as SNOMED CT [18], LOINC and ICDx. Therefore, data coded with distinct terminologies can be related to each other.
Figure 1 illustrates the class diagram of our mediator. It is based on the canonical model [19, 20]. We have extracted the more prominent classes for mediation, and we propose a new canonical class diagram [8].
The canonical model contains three elements: (i) Post Production of Medical Information (PPMI) (ii) Medical Activity, and (iii) Pathological Case. The PPMI element contains the classes (MatrialPost, Team, and Actor) responsible for the production of medical information. The Medical Activity element include the classes (MedicalAction, ReferenceActivity, and MedicalActivity) used to represent the care plan realized by the PPMI. Finally, the Pathological Case consists of the classes (Patient and Pathology) that show the association between patient and their historical records.
The advantage of this representation is that the internal structure of each of the component can be left to the free conceptual choice of each healthcare system. In addition, this canonical model will be the basis for the creation of the canonical health document.

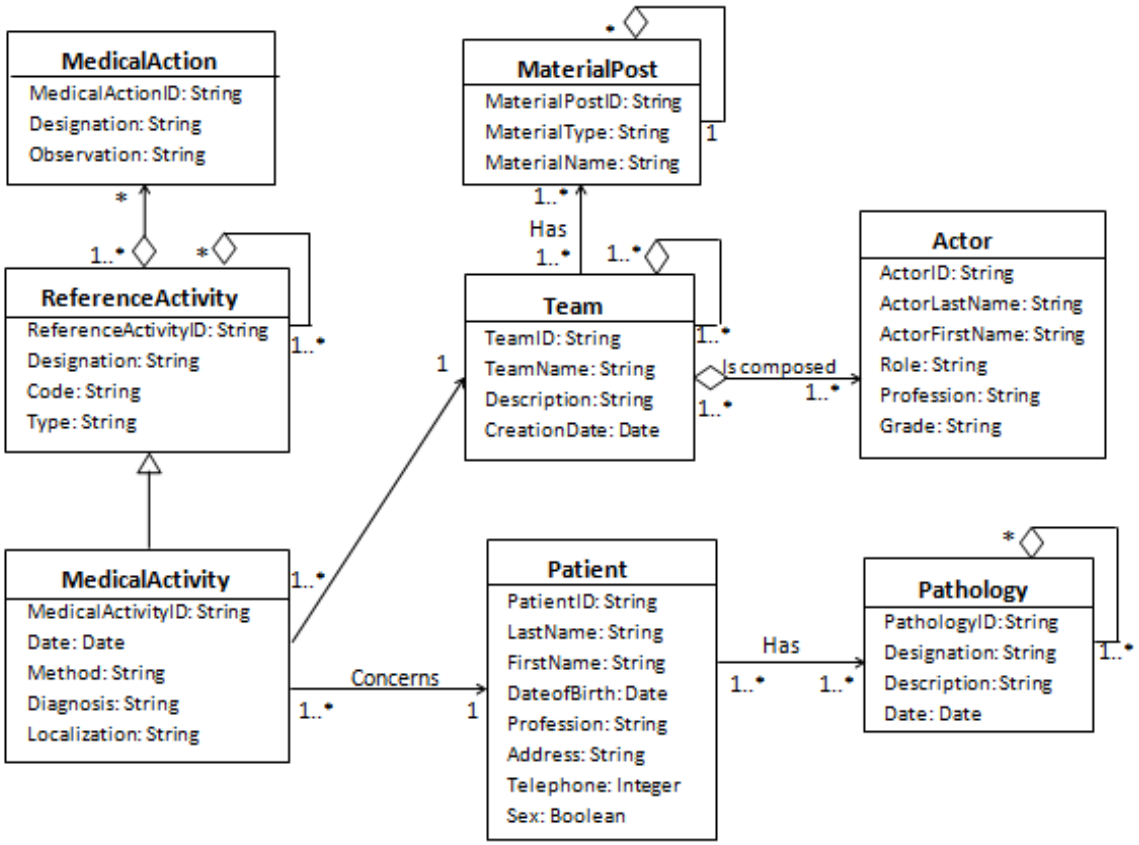

Fig.1 HIS Mediator Class Diagram

The classes of our mediator are:

- Medical Activity represents information about diagnosis and treatment;
- Reference Activity contains description of methods of execution and the site on which it takes place;
- Medical Action is the simplest action of Medical Activity such as a question during a medical check;
- Patient is an individual awaiting or under medical care and treatment;
- Pathology contains historical diseases related to the patient.
- Actor describes the medical personnel;
- Team is a group of Actors that intervene in a Medical Activity.
- Material Post is the equipment used by Team to practice a Medical Activity.

Figure 2 illustrates the architecture of our mediator. The canonical transformer component will enable data to be converted into a canonical document. It transforms the message to a common canonical format. This canonical transformation requires the use of auxiliary information: thesaurus. Our mediator supports two types of thesaurus; terminological thesauri: WordNet [21] and biomedical thesauri: UMLS [17].

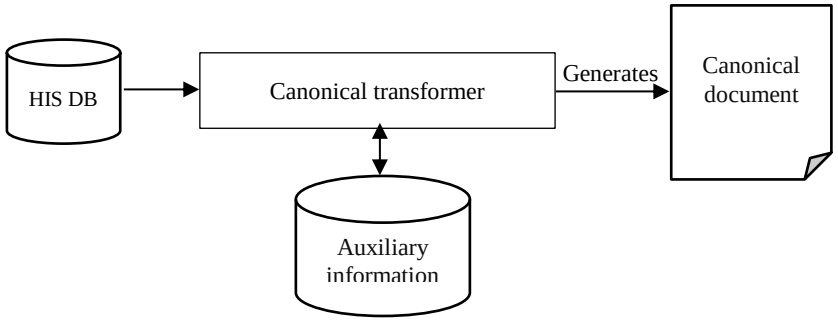

Fig.2 The Mediator Architecture

5 Experiment

To demonstrate the simplicity of our mediator, we chose to work with HL7v 2 standards. The reason for this choice is that HL7 is the most popular standard in healthcare domain. Figure 3 shows an example of HL7 message in both versions.
It is clear that we should be an expert to understand HL7 version 2 message; even if it is the most widely used standard in the world today. Figure 4 shows message after canonical transformation.
The HL7 v2 message is converted to the canonical document using the canonical transformation. As result, the message can now easily be read by human eye and the content is understandable. Therefore, the use of the canonical mediator facilitates the exchange of medical data across healthcare systems.

```
<?xml version="1.0" encoding="UTF-8"?>
<ORU_R01>
…
  <ORU_R01.PATIENT>
   <PID>
    <PID.3>
     <CX.1>010-11-1111</CX.1>
    </PID.3>
    <PID.5>
     <XPN.1>
      <FN.1>Estherhaus</FN.1>
     </XPN.1>
     <XPN.2>Eva</XPN.2>
    </PID.5>
    <PID.7>
     <TS.1>19720520</TS.1>
    </PID.7>
    <PID.8>F</PID.8>
<XAD.1>
      <SAD.1>256 Sherwood Forest.</SAD.1>
     </XAD.1>
    ...
   </PID>
…
<CE.2>GLUCOSE</CE.2>
     <OBX.5>175</OBX.5>
     <OBX.6>
      <CE.1>mg/dl</CE.1>
     </OBX.6>
  …
</ORU_R01>
```

Fig.3 – HL7 v2 message

```
…
<MedicalActivity>
    <MedicalActivityID>948642</MedicalActivityID>
    <Diagnosis>GLUCOSE</Diagnosis>
    <Date>200502150730</Date>
    <Patient>
        <PatientID>010-11-1111</PatientID>
        <LastName>Estherhaus</LastName>
        <FirstName>Eva</FirstName>
        <DateofBirth>19720520</DateofBirth>
        <Sex>F</Sex>
        <Adress>256 Sherwood Forest 70809</Adress>
        <Telephone>(225)334-5232</Telephone>
  </Patient>
  <Team>
     <Actor>
        <ActorID>020-22-2222</ActorID>
        <ActorLastName>Levin-Epstein</ActorLastName>
        <ActorFirstName>Anna</ActorFirstName>
        <Role>Ordering Provider</Role>
        <Grade>MD</Grade>
    </Actor>
..
    <MedicalAction>
        <MedicalActionID>1554-5</MedicalActionID>
        <Designation>GLUCOSE </Designation>
        <Observation>175 mg/dl </ Observation >
    </MedicalAction>
    </ReferenceActivity>
</MedicalActivity>
…
```

Fig.4 –HL7 version 2 message after transformation

6 Conclusion

In this paper, we briefly summarized standards used to support interoperability in healthcare domain. It is clear that the number of standards in health domain is large. However, it is extremely difficult to except the use of a single worldwide standard without further integration, and development. In this context, we propose a canonical document to mediate between health information systems.

The proposed mediator is simple and for general use. It does not impose any requirement or adjustment on the system architecture. Also, it is open, flexible, and platform independent.

References

1. HL7. Available from: (http://www.hl7.org/).
2. DICOM. Available from: (http://dicom.nema.org/).
3. HOERBST, A. et AMMENWERTH, E. Electronic health records. Methods Inf Med, 2010, vol. 49, no 4, p. 320-336.
4. KALRA, Dipak, BEALE, Thomas, et HEARD, Sam. The openEHR foundation.Studies in health technology and informatics, 2005, vol. 115, p. 153-173.
5. HAUX, Reinhold. Health information systems–past, present, future.International journal of medical informatics, 2006, vol. 75, no 3, p. 268-281.
6. MCDONALD, Clement J., HUFF, Stanley M., SUICO, Jeffrey G., et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update.Clinical chemistry, 2003, vol. 49, no 4, p. 624-633.
7. STEINDEL, Steven J. International classification of diseases, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. Journal of the American Medical Informatics Association, 2010, vol. 17, no 3, p. 274-282.
8. Mazouz, S., Malki, M.O.C. and Nfaoui, E.H. 'An XML-based mediator for health information systems', Int. J. Medical Engineering and Informatics, Accepted.
9. LIU, Hangsheng, BURKHART, Q., et BELL, Douglas S. Evaluation of the NCPDP Structured and Codified Sig Format for e-prescriptions. Journal of the American Medical Informatics Association, 2011, p. amiajnl-2010-000034.
10. AGUILAR, Antonio. Semantic Interoperability in the context of eHealth. In :Research Seminar, DERI Galway, December 15th. 2005. "The Continuity of Care Document. Changing the Landscape of Healthcare Information Exchange". 2009 Corepoint Health.
11. SELIGER, Robert. Overview of HL7's CCOW Standard. Health Level Seven, Inc. URL:http://www.hl7.Org/library/committees /sigvi/ ccow_overview_2001. doc, 2001.
12. KIM, Katherine. Clinical data standards in health care: Five case studies.California HealthCare Foundation Julio, 2005.
13. Bicer, Veli, et al. "Artemis message exchange framework: semantic interoperability of exchanged messages in the healthcare domain." ACM Sigmod Record 34.3 (2005): 71-76.
14. Jian, W. S., Hsu, C. Y., Hao, T. H., Wen, H. C., Hsu, M. H., Lee, Y. L., ... & Chang, P. (2007). Building a portable data and information interoperability infrastructure—framework for a standard Taiwan Electronic Medical Record Template. Computer methods and programs in biomedicine, 88(2), 102-111.
15. Lopez, D. M., & Blobel, B. G. (2009). A development framework for semantically interoperable health information systems. International journal of medical informatics, 78(2), 83-103.
16. Bruno Frandji, et all. Interopérabilité des données de santé : comment passer à l'acte ?. 13èmes Journées Francophones d'Informatique Médicale Qualité, Risques, Information et Informatique de Santé 28, 29 et 30 avril 2009 - Nice, France.
17. BODENREIDER, Olivier. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, 2004, vol. 32, no suppl 1, p. D267-D270.
18. ELKIN, Peter L., BROWN, Steven H., HUSSER, Casey S., et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. In : Mayo Clinic Proceedings. Elsevier, 2006. p. 741-748.
19. Cherkaoui, M.M.O., D. Doukkali, I. El Azami and C. Tahon, 2008. Communication de l'activite medicale entre SIH: Une representation canonique et un modele de patron. La Conférence Internationale GISEH'08 Lausanne en Suisse, Sept. 4-6, EPFL, pp: 1-7.
20. Azami, Ikram El. Ingéniérie des Systèmes d'Information Coopératifs, Application aux Systèmes d'Information Hospitaliers Ressource électronique. [S.l.] : [s.n.], 2012

21. Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.