HIR
Healthcare Informatics Research

Convergence of Health Level Seven Version 2 Messages to Semantic Web Technologies for Software-Intensive Systems in Telemedicine Trauma Care

Pedro Monteiro Menezes, MSc[1], Timothy Wayne Cook, MSc[2], Luciana Tricai Cavalini, MD, PhD[3]

[1]Post-Graduation Program in Civil Defense and Security, Rio de Janeiro, Brazil; [2]Emergent Group for Research and Innovation on Healthcare Information Technology, Rio de Janeiro State University, Rio de Janeiro, Brazil; [3]Department of Epidemiology and Biostatistics, Fluminense Federal University, Niterói, Brazil

Objectives: To present the technical background and the development of a procedure that enriches the semantics of Health Level Seven version 2 (HL7v2) messages for software-intensive systems in telemedicine trauma care. **Methods:** This study followed a multilevel model-driven approach for the development of semantically interoperable health information systems. The Pre-Hospital Trauma Life Support (PHTLS) ABCDE protocol was adopted as the use case. A prototype application embedded the semantics into an HL7v2 message as an eXtensible Markup Language (XML) file, which was validated against an XML schema that defines constraints on a common reference model. This message was exchanged with a second prototype application, developed on the Mirth middleware, which was also used to parse and validate both the original and the hybrid messages. **Results:** Both versions of the data instance (one pure XML, one embedded in the HL7v2 message) were equally validated and the RDF-based semantics recovered by the receiving side of the prototype from the shared XML schema. **Conclusions:** This study demonstrated the semantic enrichment of HL7v2 messages for intensive-software telemedicine systems for trauma care, by validating components of extracts generated in various computing environments. The adoption of the method proposed in this study ensures the compliance of the HL7v2 standard in Semantic Web technologies.

Keywords: Telemedicine, Health Information Systems, Information Technologies and Communication Projects, Information Exchange, Health Level Seven

I. Introduction

Healthcare information systems (HISs) can be regarded as an important asset to support telemedicine trauma care, since they can be related to the special features of emergency care, such as short response latency; the smaller the latency, the higher the likelihood of a favorable outcome. Taking this scenario into account, the information exchange implemented for a telemedicine trauma care HIS cannot be subjected to misinterpretation or the loss of the semantic meaning of its data [1]. By definition, trauma care involves at least two actors, emergency services and the receiving hospital. In this

sense, interoperable, software-intensive HISs [2] are necessary for achieving full functionality of telemedicine trauma care.

In a broader sense, the implementation of information technology (IT) has been proposed to increase the effectiveness of healthcare systems, but these expectations have not been met yet. Over the last 50 years, many software companies have tried to realize semantic interoperability in distributed systems to provide a information platform for consistent healthcare [3] and computable, clinical decision support.

The achievement of such levels of interoperability between HISs remains a challenge when the current standards are used. Currently, there is a multitude of companies and government institutions whose task is to develop HISs, each of them implementing its own data model, which is specific to that system's functionality [4]. These data models differ from system to system and are always changing according to the changing requirements of the systems based on continuous scientific developments in medicine, government policies, and the standards of healthcare insurance companies [5].

This constant change is an expensive component of health information management systems. In addition to the expense, the semantic context of healthcare information is incorporated into both the database structure and the application source code when the information is recorded. Thus, when trying to perform data sharing among different HISs, even in the simplest situation (when the data types are the same), the full temporal, spatial, and ontological contexts in which the data was recorded remain unknown to the receiving system [6].

Many solutions have been proposed to the problem of HIS semantic interoperability, which includes a vast and variable set of knowledge representation models [7]. The most widely adopted solution for data exchange to date is the Health Level 7 version 2 (HL7v2) standard [8]. This standard has already gone through all the steps of the life cycle necessary for its consolidation, and it has become almost mandatory in HISs adopted by healthcare facilities in the United States [9]. Being a very elastic standard, in its definition, it allows the inclusion of virtually any data in a message. However, this elasticity always requires the receiving software to establish a one-to-one mapping to the sending system, and there is no semantic integrity assurance in this mapping [8].

The HL7 version 3 (HL7v3) standard was proposed to overcome the limitations of HL7v2. However, in practical terms, since HL7v3 is not compatible with HL7v2, the maintainance of two types of interfaces for both standards has

lead to increased costs and the possibility of systemic errors in the software. This, in turn, has resulted in slower uptake of HL7v3 by developers, to date, and the overall complexity of HL7v3 has also been cited as a point of failure [10].

eXtensible Markup Language (XML) technologies are very promising in the scenario of semantic interoperability, but they should be combined with solutions already proven in software to ensure semantic interoperability, such as the constraint-based multilevel model-driven (MMD) approach, originally proposed by the openEHR Foundation [11]. The MMD approach defines a common reference model (RM) and allows only constraints on the RM components in a domain model (DM). This insures that all DM concept expressions can be interpreted in the context of the RM. Thus, a combination of XML and MMD technologies could provide full semantic interoperability for an HIS and its compliance to the emerging Semantic Web and the Internet of Things [12]. Given the current scenario of the HIS industry worldwide, it is desirable to combine the ubiquitous HL7v2 infrastructure with the emerging Semantic Web technologies, specifically XML syntactic processing with concept semantics expressed in Resource Description Framework (RDF) structures.

This is especially critical for clinical documentation, where the multiplicity of concepts and the low level of consensus on what would be the 'maximal data model' for each of them. This means that a comprehensive HIS ecosystem is a complex network of data silos, paper persistence, and other workarounds [4,13]. In life threatening situations, such as pre-hospital trauma care, the current situation is likely to be harmful to patients [14]. The search for implementable solutions to the semantic interoperability problem for HISs is a necessity for 21st century healthcare and beyond.

Given the scenario described above, this study had the objective of developing a method to embed the point-of-care contextual semantics of captured information in HL7v2 messages. The proposed method uses an XML implementation of the MMD principles with embeded RDF to express semantic relationships. This exploits the existing HL7v2 infrastructure, thus reducing the barrier to adoption of improved semantic interoperability. The Pre-Hospital Trauma Life Support (PHTLS) protocol is the use case for the proof of the concept.

II. Methods

The use case scenario devised for the telemedicine application proposed here comprises a mobile application for decision support on advanced trauma care, in which the PHTLS

model is a fraction of the full implementation. The purpose of the message hybridization presented in this study is to allow the full semantics and structure of the data collected at the pre-hospital point-of-care to be sent to the reference hospital or any other healthcare facility to which the patient is going to be transported.

The methodological approach adopted to develop the prototype application included the following: 1) the implementation of the PHTLS initial examination as an HL7v2 message and as a Concept Constraint Definition (CCD) [15]; 2) the generation of an instance of the HL7v2 message for the same data model; 3) the generation of an XML data instance that is valid against the CCD; 4) the creation of a hybrid HL7v2-XML message; 5) the transmission of the message; and 6) the receipt of the message, persisting in three formats: pure HL7v2, pure XML, and hybrid, while the semantics from the pure XML instance and hybrid are recovered.

The PHTLS initial examination is known as the ABCDE protocol, covering the parameters of airway, breathing, circulation, disability, and exposure [16]. This protocol was modeled as a CCD, an XML schema file with the semantics embeded as RDF statements, which defines constraints on the reference model in accordance with the MMD principles established by the Multilevel Healthcare Information Modeling (MLHIM) specifications, which are described in detail elsewhere [12,15]. The Concept Constraint Definition Generator (CCD-Gen; www.ccdgen.com) was used to create the CCD. The CCD-Gen also produced an example XML data instance, which was edited to contain clinically significant data and revalidated to its correspondent CCD by the use of oXygen XML editor version 16.1.

The XML instance was incorporated into an HL7v2 message containing the same data model. The HL7v2 was implemented manually in a text editor. To enable the incorporation of the XML instance into the HL7v2 message, a new type of message segment, called ZML, was created. The ZML segment was created specifically to allow MLHIM-based data instances to be embedded in HL7v2 messages. The HL7v2 standard [17] provides for 'Z' segments that are user defined, as custom segments of a message outside of the scope of the strict HL7v2 standard.

A prototype application was created to receive the hybrid message, validate it, and write it to disk using the Mirth Connect middleware [18], which is a widely used, cross-platform, open-source application that allows two-way messaging targeting different connection points as well as the filtering and processing of HL7 messages.

To perform the message exchange, the HL7 browser was used. The HL7 browser is an open-source application writ-ten in Java to display, send, and receive HL7v2 messages via TCP/IP socket. The HL7 browser was used in this study to simulate the exchange of hybrid message between distributed systems.

The validation prototype was composed of the Mirth application and the third party backwards validation from the MLHIM XML data instance, the CCD XML schema, and the MLHIM reference model schema using two independent XML schema 1.1 validators, Saxon-EE and Xerces.

III. Results

The knowledge modeling process uses the MLHIM reference model to model the concepts of the ABCDE protocol of the PHTLS model. As a result, a CCD was modeled, composed of the following pluggable complex types [12]: 'A' airway obstruction; 'B' breathing; 'C' circulation; 'D' disability, measured by the Glasgow Coma Scale; and 'E' exposure, and one specific PcT to provide the mapping between the OBX segment of the HL7v2 message and the Type 4 universal unique identifier (UUID) of the corresponding PcTs in the CCD. The resulting CCD modeling of the PHTLS concepts is presented in Table 1.

The 'A' PcT consists of DvBooleanType; the 'B, C, D' and mapping PcTs are clusters; and the 'E' PcT is composed of DvStringType. The 'C' cluster consists of four other clusters: skin, blood pressure, pulse, and bleeding. The mapping cluster is composed of two PcTs that contain the combination of the HL7v2 segment with the corresponding PcT UUID, thereby creating a semantic identity for the HL7v2 segments.

The HL7v2 message was modeled according to the version 2.5.1 of the standard, composed of the standardized segments MSH, PID, PV1, and OBX as well as the user-defined segment ZML that contains the XML instance, excluding the line break character. Thus, the hybrid message is composed of the standard HL7v2 segments with the ZML segment. The resulting HL7v2 message, the MLHIM XML data instance, the corresponding CCD in XML schema and the hybrid HL7v2/XML message can be downloaded at https://github.com/lutricav/Academic/tree/master/menezespm.

The message sent by the HL7 Browser was received via a data channel created in Mirth Connect. The channel was configured with a data source TCP listener on port 6661. Three destinations were created, and the results were saved to disk. The first destination wrote the message as received; thus, if necessary, it would be possible to check whether the filter of the other two targets operated appropriately. The second target was set to retain only the MLHIM XML instance (Figure 1), and the third part removed the ZML segment

Table 1. CCD modeling of the PHTLS concepts related to 'ABCED'

Concept	Data type	Valid values
A: Airway obstruction	Boolean	Yes = TRUE
		No = FALSE
B: Breathing	Ordinal	Apneic
		Slow
		Normal
		Fast
		Abnormally fast
	Count	0 to 500 ipm
C: Circulation		
Skin		
Skin color	String	Normal
		Pale
		Cyanotic
		Others
Capillary refill	String	<2 seconds
		≥2 seconds
Moist	String	Dry
		Moist
Temperature	Ordinal	High
		Normal
		Low
	Quantity	0°C to 200°C
Blood pressure		
Diastolic blood pressure	Quantity	0 to 2,000 mmHg
Systolic blood pressure	Quantity	0 to 2,000 mmHg
Pulse		
Arrhythmia	Boolean	Yes = TRUE
		No = FALSE
Pulse perception	Ordinal	Imperceptible
		Slow
		Normal
		Accelerated
Pulse rate	Count	0 to 1,000 bpm
Bleeding		
Type of bleeding	String	Capillary
		Venous
		Arterial

and retained only the HL7v2 content (Figure 2). The architectural prototype is described in Figure 3.

The original hybrid message was sent from a MacBook Air computer to an Ubuntu 14.10 computer and Windows 8 Home Basic computer. All three segments (HL7v2 message, MLHIM XML instance, and hybrid message) were success- fully validated on three platforms.

IV. Discussion

This study demonstrated the semantic enrichment of HL7v2 messages by combining them with an XML implementation of

Table 1. Continued

Concept	Data type	Valid values
D: Disability (Glasgow Coma Scale)		
Eye	Ordinal	Does not open eyes
		Opens eyes in response to painful stimuli
		Opens eyes in response to voice
		Opens eyes spontaneously
Motor	Ordinal	Makes no movements
		Extension to painful stimuli (decerebrate response)
		Abnormal flexion to painful stimuli (decorticate response)
		Flexion/withdrawal to painful stimuli
		Localizes painful stimuli
		Obeys commands
Verbal	Ordinal	Makes no sounds
		Incomprehensible sounds
		Utters inappropriate words
		Confused, disoriented
		Oriented, converses normally
Glasgow Coma Scale score	Ordinal	3 to 15
		(reference range: 3–8 severe; 9–12 moderate, 13–15 mild)
E: Exposure	String	Free text
Mapping cluster		
OBX segment	String	OBX description
PcT UUID	String	PcT UUID

CCD: Concept Constraint Definition, PHTLS: Pre-Hospital Trauma Life Support, OBX segment: observation segment, UUID: universal unique identifier.
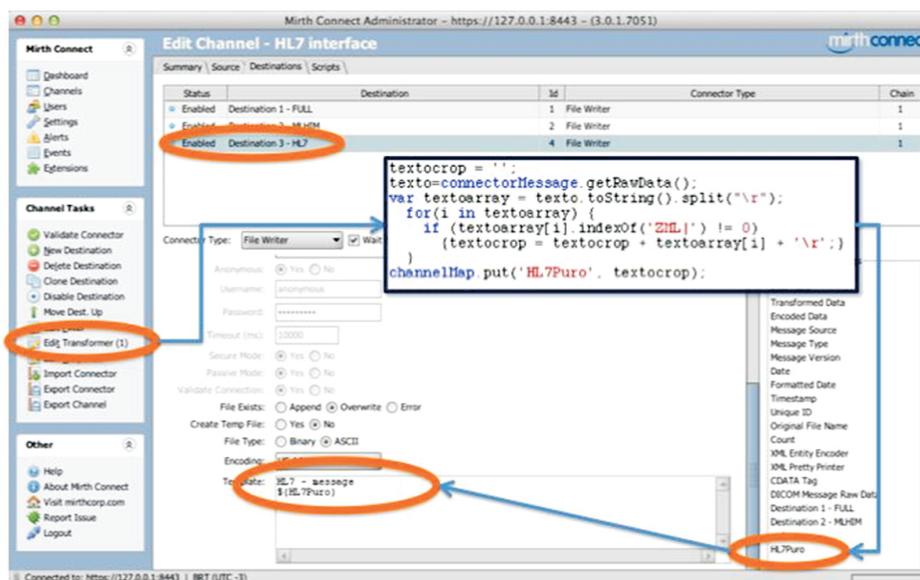

Figure 1. Specific destination of the XML data instance.

the MMD principles. The results demonstrate that a message containing MLHIM and HL7v2 segments is valid for both specifications; in some cases directly, in other cases with the development of a small piece of software, which can be developed in Mirth or directly on the legacy application, for separating the hybrid message into pure HL7v2 and pure XML.
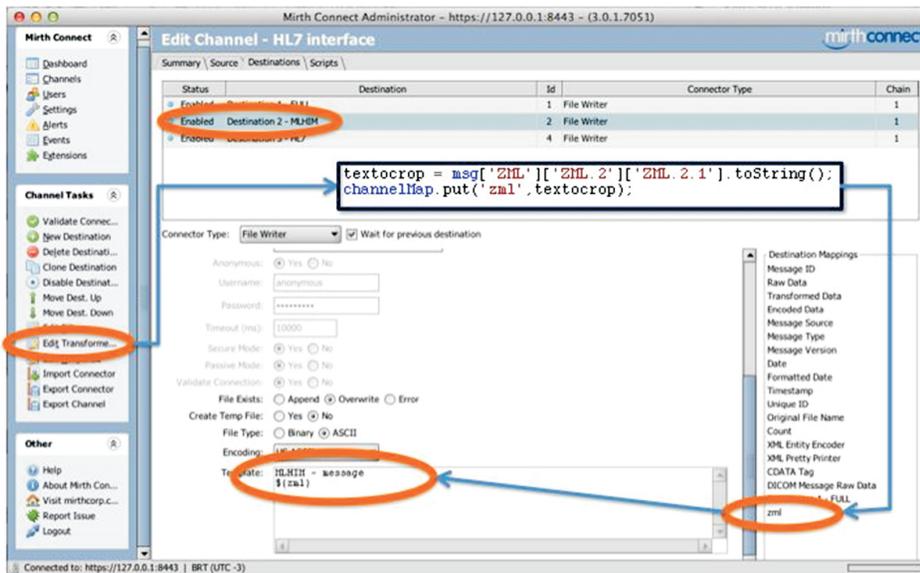
Figure 2. Specific destination of the Health Level 7 version 2 (HL7v2) message.
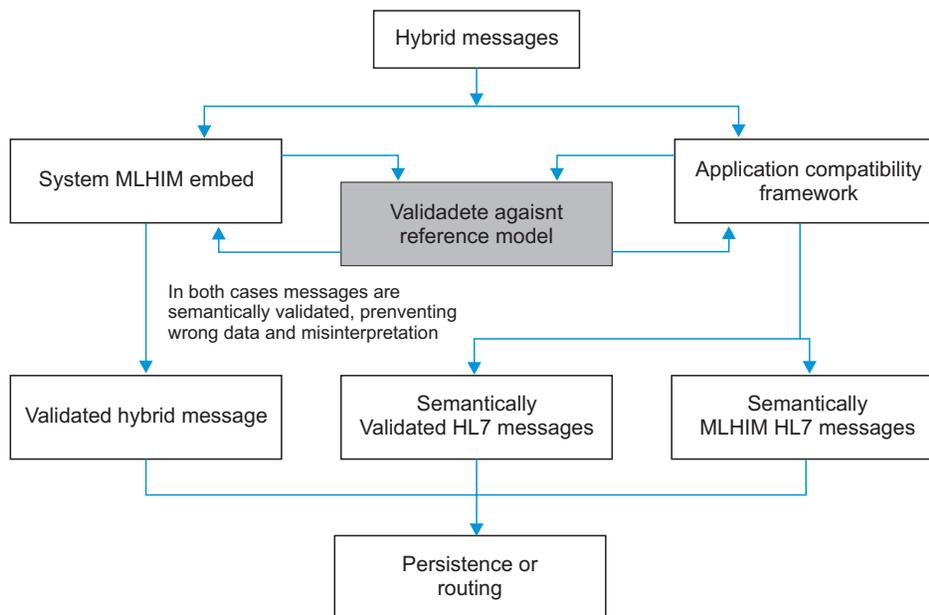

Figure 3. Description of the architectural prototype. MLHIM: Multilevel Healthcare Information Modeling.

It is important to emphasize that there are two types of data constraint concepts required to effect clinical decision support. The structural semantics and the ontological semantics (there is also the temporal element) are both required, and this is the key advantage of MLHIM; it provides them in one modular, sharable package. There has been a much wider socio-political acceptance of HL7, including HL7v3 implementations, such as the Clinical Document Architecture (CDA), but these are beyond the scope of this paper. Rather, we chose to focus on HL7v2 because of its still wide implementation in devices and systems, and its retro-compatibility in new devices. However, despite all the industry and academic efforts around those standards, it did not achieve the ability

to provide full computable semantic interoperability in any real distributed and heterogeneous healthcare information ecosystem. Therefore, it is not able to guarantee computerized, operational decision support. The top-down nature of its design requires such a level of flexibility in the message modeling that there are no effective constraints represented on the data. The real advantage of adopting XML and RDF technologies for the development of this method was the potential for the development of semantically interoperable applications for real healthcare settings, regardless of the application size or how it will be used. By incorporating an ZML segment, HL7v2 messages can be verified and validated from the CCDs referenced in embedded XML instances. Through

the use of CCDs, models can be created freely by domain experts, specifically addressing the needs of any particular application, while still enabling semantic interoperability. The simple adoption of XML does not guarantee full semantic representation of the data. A combination of XML schema and RDF, modeled according to the MLHIM principles, can represent structural and semantic data details that can be shared across sender and receiver applications. The only limitations regarding this approach are related to resistance to innovation and the difficulties faced by any paradigm-breaking technology in the scope of a scientific revolution. These issues were extensively addressed elsewhere [12].

Since XML is a ubiquitous standard and all major programming languages have tools for handling XML schema, this allows application developers to work in their preferred language, using persistence models of their choice, and yet not build 'data silos' [4]. With this harmonization between MMD and HL7, HL7v2 loses its hermetic features because hybrid messages will be accompanied by their semantic context, enabling safer mapping of the input data for persistence and machine-based decision support.

Systems that would consume hybrid messages are not required to implement the MLHIM reference model or the CCD in their application source code to ensure that messages carry semantics. The only requirement is that applications are able to import and export messages validated according to the MLHIM reference model adopted to model the CCD using standard XML validation. This feature allows messages to be easily validated at a communication point or diluted in steps according to the pieces of software used at each endpoint. The message validation is done in three steps: the XML instance against the CCD, the CCD against the MLHIM reference model schema, and this last one against the W3C XML schema specifications. The first two steps are required to receive messages from unknown CCDs, but once the CCD is incorporated into the system, that is sufficient to match it with new data instances to attest the validity of new messages. This validation method is not available in any other current healthcare IT standard.

In summary, the method proposed in this paper could be useful for the connection of HL7v2 messages with the Semantic Web processing technologies. Since the MLHIM specifications can be seen as an appropriation of the best features of HL7v3, especially the adoption of XML technologies and some of the features of the ISO 21090 standard, and the MMD principles first proposed, in implementation, by the openEHR specifications, there are potential advantages in converging these technologies. This can be realized without the limitations inherent in these two approaches when con-

sidered in isolation and without creating a new communications infrastructure. This use of MLHIM solves the primary issues of the current approaches to semantic interoperability. In the HL7v2/v3 case, the absence of an implemented reference model and constraint-based domain models does not allow the validation of its schemas. In openEHR, a domain-specific language is adopted for knowledge representation. In the case of the emerging HL7 FHIR proposal, the communications/message infrastructure must be completely rebuilt. In addition, both standards have in common the top-down approach for the modeling of clinical concepts, which is not compatible with the complexity and dynamics of knowledge in healthcare. MLHIM provides the ability of any domain expert to express the information they wish to exchange.

Considering HL7v2, a 'pure' system, the method proposed in this study can be used as a middleware that will validate incoming messages before delivery to the system, and the output XML messages can be created according to each specific data model. This approach reduces deployment costs while keeping legacy databases and adding CCDs to enhance the exchange of data semantics for decision support and other uses, such as big data processing and analysis. The convergence of the HL7 standard with the syntactic and semantic soundness of the MLHIM specifications, as proposed in this study, provides a new horizon in the development of HIS in telemedicine.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

References

1. Lenz R, Reichert M. IT support for healthcare processes: premises, challenges, perspectives. Data Knowl Eng 2007;61(1):39-58.
2. Broy M. The 'grand challenge' in informatics: engineering software-intensive systems. Computer 2006;39(10): 72-80.

3. De Leon S, Connelly-Flores A, Mostashari F, Shih SC. The business end of health information technology: can a fully integrated electronic health record increase provider productivity in a large community practice? J Med Pract Manag 2010;25(6):342-9.

4. Raths D. Shifting away from silos: the interoperability challenges that hospitals face pale in comparison to the headaches plaguing state departments. Healthc Inform 2010;27(1):32-3.

5. Hufnagel SP. Interoperability. Mil Med 2009;174(5S):43-50.

6. Blobel BG, Pharow P. Analysis and evaluation of EHR approaches. Stud Health Technol Inform 2008;136:359-64.

7. Blobel B. Ontologies, knowledge representation, artificial intelligence: hype or prerequisites for international pHealth Interoperability? Stud Health Technol Inform 2011;165:11-20.

8. Lopez DM, Blobel B. Architectural approaches for HL7-based health information systems implementation. Methods Inf Med 2010;49(2):196-204.

9. Quinn J. An HL7 (Health Level Seven) overview. J AHIMA 1999;70(7):32-4.

10. Ingenerf J, Kock AK, Poelker M, Seidl K, Zeplin G, Mersmann S, et al. Standardizing intensive care device data to enable secondary usages. Stud Health Technol Inform 2012;180:619-23.

11. Kalra D, Beale T, Heard S. The openEHR Foundation. Stud Health Technol Inform 2005;115:153-73.

12. Cavalini LT, Cook TW. Use of XML Schema Definition for the development of semantically interoperable healthcare applications. In: Gibbons J, MacCaull W, editors. Foundations of health information engineering and systems. Heidelberg, Germany: Springer-Verlag; 2014. p. 125-45.

13. Saleem JJ, Russ AL, Neddo A, Blades PT, Doebbeling BN, Foresman BH. Paper persistence, workarounds, and communication breakdowns in computerized consultation management. Int J Med Inform 2011;80(7):466-79.

14. Moulton SL, Haley-Andrews S, Mulligan J. Emerging technologies for pediatric and adult trauma care. Curr Opin Pediatr 2010;22(3):332-8.

15. Cavalini LT, Cook TW. Knowledge engineering of healthcare applications based on minimalist multilevel models. Proceedings of IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom); 2012 Oct 10-13; Beijing, China. p. 431-4.

16. National Association of Emergency Medical Technicians. PHTLS: prehospital trauma life support. 8th ed. Burlington (MA): Jones & Bartlett Learning; 2014.

17. Health Level Seven International. Section 2.5.2 Segments and segment groups [Internet]. Ann Arbor (MI): HL7 International; 2012 [cited at 2015 Dec 20]. Available from: http://wiki.hl7.org/index.php?title=Section_2.5.2_Segments_and_ segment_groups.

18. Ceron JD, Gomez GA, Lopez DM, Gonzalez C, Blobel B. A personal health record system for diabetes care conforming to the ISO 16527 interoperability requirements. Stud Health Technol Inform 2014;200:49-55.