Relevant Tools for Tackling Interoperability Problems on Heterogeneous Electronic Health Record Systems: An Exploratory Research

Herramientas Relevantes para Abordar Problemas de Interoperabilidad en Sistemas de Expedientes Clínicos Electrónicos Heterogéneos: una Investigación Exploratoria

D. Ojeda-Carreño, M. A. Cosío-León, J. I. Nieto-Hipólito

1Universidad Autónoma de Baja California.

ABSTRACT

Patients’ medical records have been originally based on paper and since 1865 the medical field recognized the need to improve the means for storing and retrieving disperse information. Electronic Health Records (EHR) systems face similar problems, with interoperability being one of them, it defined by IEEE standard glossary of software engineering terminology as: the ability among information systems to exchange data. Mainly, two perspectives of interoperability emerge: a) syntactic, which refers to the ability of an information system to import and validate grammar, as well as a set of construction rules utterances created by another system; and b) semantic, the ability to exchange data among systems and understand it in the same way, regardless of the source system, also called Semantic Interoperability on Information and Communication Technologies or SIICT. The use of standards, vocabularies and terminologies is a common practice for sharing data among heterogeneous EHR systems in order to face interoperability problems. This heterogeneity starts with the manner to collect and store data. This paper presents to e-health practitioners and researches relevant tools used to achieve interoperability in heterogeneous EHR systems, as well as challenges and future trends to manage EHRs with the intention to achieve SIICT.

RESUMEN
Los registros clínicos de pacientes se basaron en papel originalmente, desde el año de 1865 se reconoció en el entorno médico la necesidad de mejorar los medios para el almacenamiento y recuperación de información dispersa. Los Sistemas de Expediente Clínicos Electrónicos (ECE) enfrentan problemas similares, siendo uno de estos problemas, la interoperabilidad. La cual es definida por “IEEE standard glossary of software engineering terminology” como: la habilidad de intercambiar datos entre sistemas de información, donde dos perspectivas de interoperabilidad emergen: a) sintáctica, la cual se refiere a la habilidad de un sistema de información de importar y validar la gramática de datos, así como de un conjunto de reglas de expresiones creadas por otro sistema; y b) semántica, como la habilidad de intercambiar datos entre sistemas y de entenderlos de la misma forma, independientemente del sistema fuente, llamando a esto como Interoperabilidad Semántica sobre Tecnologías De La Información y Comunicación (ISTIC). El uso de estándares, vocabularios y terminologías es una práctica común para poder compartir datos entre sistemas ECE con el fin de enfrentar problemas de interoperabilidad. Esta heterogeneidad se da desde la forma en cómo se colecta y almacenan los datos. Este artículo presenta a los profesionales de salud e investigadores de e-salud acerca de herramientas relevantes utilizadas para lograr la interoperabilidad de sistemas ECE heterogéneos, así como retos y futuras tendencias para administrar sistemas ECE con la intención de lograr ISTIC.


Correspondencia
DESTINATARIO: Dolores Ojeda Carreño  
DIRECCIÓN: Carretera Transpeninsular Ensenada-Tijuana #3917, Col. Playitas, C.P. 22860, Ensenada, Baja California, México  
CORREO ELECTRÓNICO: ojeda.dolores@uabc.edu.mx

Fecha de recepción: 10 de junio de 2016
Fecha de aceptación: 27 de septiembre de 2016
INTRODUCTION

The IEEE standard glossary of software engineering terminology\[1\] defines the term interoperability as the ability between two or more Information Systems (IS) to exchange information and use it. There are two major focus of interoperability: a) semantic, which is understood as the ability to exchange information among IS and understand it in the same way, regardless of the source system\[2\]. In the same context, unambiguous data exchange is described as Semantic Interoperability on Information and Communication Technologies or SIICT; b) syntactic, which refers to the ability of an information system to import and validate grammar and/or a set of construction rules utterances created by another IS\[3\]. The lack of both produces problems related with storing and retrieving disperse medical information. This paper’s contributions are two-fold, mainly focusing on describing standards, terminologies and controlled vocabularies to achieve both interoperability types found in the literature. Secondly, we analyzed Electronic Health Record Systems and the Hospital Information System (HIS) modules\[4\], their relation to data context and the impact of standards, terminologies and controlled vocabularies regulating data transactions to achieve interoperability.

Medical information is collected on a document known as medical record; this record was originally based on paper and physically stored. Actually, every patient has one since his/her birth, named on this paper as original medical record. This is consulted by physicians and healthcare professionals (nurses, and physician assistants, among other) who are allowed to access medical records. The aforementioned document has brought numerous problems for retrieving and storing medical information. These problems were identified since 1865, when the medical field recognized the need for “better means of storing and retrieving medical information”\[5\].

The above mentioned problems have been partially resolved through the use of Information Communication Technologies (ICT). These allow a decrease of medical records based on paper\[6\], avoided paper limitations\[7\] and empower the electronic health known as e-health\[8\] to make way to digital medical records\[9\], called Electronic Health Record (EHR), which allows healthcare professionals to retrieve and store digital medical information\[2\]. Although, problems for sharing data persist, and the consequences of having disperse electronic medical records impact inaccurate clinical diagnosis, which could generate serious consequences for the patient’s health; another consequence is that, it would be necessary to do diagnostic studies again, due to the disperse information. These consequences are reflected in the economic costs for patients due to duplication of medical studies\[7\].

This paper addresses the question: Currently, which are the relevant Medical Standards, vocabularies and terminologies for tackling interoperability problems? Through the following sections: on Section Electronic Health Record, it provides relevant information about EHR and the rest of HIS sub-domain interactions.
In the Research Methodology section, it describes the methodology followed to retrieve information about tools to achieve interoperability in the literature. In the Section named Standards, Terminologies and Controlled Vocabularies related to EHR we describe our relevant findings; there is a discussion about interoperability challenges in the Section named Discussion and, finally, we present our final analysis in the Section Conclusions and Future Work.

Electronic health record

A HIS manages administrative, clinical and financial issues in a hospital[10]. It has different sub-domains, as shown in Fig. 1, one of them is EHR also called EMR (Electronic Medical Records) in the literature[17, 5], but as mentioned by Gaynor Mark et al. in[11] the definitions adopted by the U.S. Department of Health & Human Services, EHR is: the patient’s complete healthcare information and for EMR is: the medical information from a single medical intervention. In this document, we consider using the acronym EHR instead of mixing both terms.

The mechanism designed to manage the EHR and its data sources is called EHR system[12, 5], which includes: a) the EHR creation, b) gathering data from physicians/healthcare professionals, c) storing, consultation, retrieving and updating medical information, and d) the sharing of EHR among EHR Systems.

A brief summary of data sources for EHR addressed in this work are the following systems:

- **Laboratory Information System** (LIS): is a system that manages/chemistry, hematology, immunology, micro-biology, genetics, and other histopathological markers[7].

- **Radiology Information System/Picture Archiving and Communication System** (RIS/PACS): the RIS handles functions as reporting, scheduling of patient studies, imaging study management, among others[7]. On the other hand, PACS has the following functions: manages capturing, storing and displaying medical images, as well as distributing and retrieving medical images, and reports[13].

- **Electronic Prescribing System** (EPS): paper prescribing has been substituted in the electronic environment by electronic prescribing and the latter is handled by EPS also known as e-prescribing, with the intention to reduce medication errors; where the communication between prescribers, pharmacies, pharmacists and patients is predominant[14].

- **Personal Health Record System** (PHRS) manages[15] the complete record of a patient’s medical history owned and maintained by the patient[11], which includes a summary of patients’ health history[16], medications; measuring physiological parameters such as body temperature, blood pressure and medical information[16], including promoting healthy habits[7].

From patient’s perspective, the benefits of EHR systems to improve healthcare quality are many, for example: a) they promote proactive attitude on patients[8], b) they reduce error medication[14], and c) they promote healthy habits[7].

On other hand, the benefits for the physicians are: a) EHR is accessible at any time and place, b) they increase quality of medical services, and c) they allow expedited attention without neglecting efficiency in healthcare[17].

Hence, the importance of using tools that allow interoperability among heterogeneous systems, with the aim to have the EHR available at any time and in any place[17].
Research Methodology

The exploratory research on the literature was performed following directives of PRISMA methodology\(^{[18]}\). The first step was searching by definitions in medical records, such as: interoperability standards, semantic, syntactic, health information, EHR, EMR, among others; using MeSH terminology\(^{[19]}\) and related literature, with the intention to define search queries. The rest of the methodology’s steps were:

- Eligibility criteria, which includes a publishing period from 1990 to 2015, to ensure that relevant interoperability standards, terminologies and controlled vocabularies were not excluded.

- Process of selecting Databases: ACM, EBSCO, ELSEVIER, IEEE and SPRINGER were the selected database on this step.

- Building search queries, (((Interoperability Standards) AND (semantic OR syntactic )) OR (Interoperability Standards AND semantic AND syntactic)) AND (Health Information Systems) OR (telemonitoring).

- The relevant paper selection process has three phases:
  - Filtering by topic and title around EHR, resulting in 53 papers.
  - Filtering by relevant information on abstract, 31 papers, different articles were obtained; books and book chapters from America, Europe, Asia, and Oceania. We organized papers by country, with resulting works from Canada, U.S.A., Spain, France, China, Japan and Australia, and lastly we generated a conceptual map in order to find relevant concepts.
  - Meeting with an expert to analyze initially rejected papers (M.E. Cortés-M from Pacific Health Solutions S.A.), we recovered 4 papers, resulting in a total of 35 papers.

- Selected papers were fully read, obtaining the main interoperability standards, terminologies and controlled vocabularies related to EHR; described in the next section.

Standards, Terminologies and Controlled Vocabularies related to EHR

The technical report Health informatics electronic health record definition, scope and context\(^{[20]}\) defines standard as: a document approved by a recognized body, which contains common and repeated use, rules, guidelines or characteristics for activities or their results in a given context. On the other hand, a controlled vocabulary consists in a specific list of terms whose meaning is unalterable\(^{[19]}\). A terminology consists in a specific list of technical terms or expressions used in a specific field\(^{[19]}\), for this case in the medical field. In the previous sections, we described EHR issues related with storing and retrieving data, as well as some solutions; it is also necessary that the information has the same meaning for both, the source system and the target system. The following subsections describe tools in the literature to achieve interoperability.

Standards

Health Seven Level (HL7), openEHR, ISO 13606 and DICOM are considered the major standards\(^{[21, 2, 12]}\) to achieve internal or external interoperability. We will describe them in the following subsections.

HL7

HL7 acronym means Health Level Seven, which is related to the application layer of the Open Systems Interconnection (OSI) Model, implemented into the
field of health systems. HL7 is formed by an international community of experts in the field of health and information technologies, collaborating together to create standards for the exchange, management and integration of electronic health information. Initial steps of HL7 were to specify the format and structure of information to be exchanged; that is, oriented to the syntax, although HL7 gives recommendations for data exchange between systems, it does not delve into technical details. Following, we described five members of the HL7 standards’ family:

› HL7 Version 2 standard (HL7 V2): published in 1988. It was the first HL7 messaging protocol, it was developed to provide a common format for exchanging textual healthcare data among information systems. HL7 V2 is defined by the use of pipe symbol as delimiter of text within each message. Until now, is the most used in healthcare information systems.

› HL7 Version 3 standard (HL7 V3) was published in 2004. HL7 V3 is a set of specifications to work with the full set of messages, data types and terminologies. HL7 V3 produces messages and electronic documents. An object-oriented development methodology, and Reference Information Model (RIM) are the key stone in HL7 V3. HL7 RIM includes detailed data types, classes, state diagrams, use case models, and terminology to derive domain specific information models with the purpose of solving semantic interoperability problems.

› HL7 PHR-System Functional Model (HL7 PHR-S FM): HL7 PHR-S FM Release 1 was published in 2008. It defines the features, functions and security necessary to create and effectively manage PHRs.

› HL7 Fast Healthcare Interoperability Resources (HL7 FHIR): it was published in 2015. It is the newest member of the HL7 standards’ family. It combines the best features of HL7 V2, HL7 V3 and HL7 CDA; it mainly uses XML format, also it can work with JavaScript Object Notation (JSON) format. The HL7 FHIR calls resources to documents and messages; HL7 FHIR also works on mobile phone apps, cloud communications, EHR-based data sharing, among other. Both HL7 CDA and HL7 FHIR are based in a model stead of archetype based standards.

openEHR

openEHR Foundation has published abstract specifications using Unified Modeling Language (UML) notation and formal textual class specifications for the Reference Model (RM), the Service Model (SM) and Archetype Model (AM) to ensure that components (as documents in HL7) and systems achieve semantic reasoning and validation in health information. Following is a brief description of openEHR specifications:

› openEHR RM guarantees that key attributes for information in health records (who, when and where) are already taken care of.

› openEHR AM defines models to describe the semantics of archetypes and templates including Archetype Definition Language (ADL).

› openEHR SM includes definitions of basic services in the health information environment. These are centered on the EHR.
ISO 13606

openEHR/ISO 13606, EN ISO 13606, EN13606 and EN/ISO 13606 are some of the names to make reference to ISO 13606 standard\[27, 24, 28, 12\]. The standard was developed by the European Committee for Standardization (CEN), based on openEHR specification to achieve semantic interoperability among heterogeneous systems of EHRs\[12, 27\].

Below we present a brief description of the ISO 13606 standard (Health informatics Electronic health record communication) parts:

- **ISO 13606 Reference model** (year of release 2008): the openEHR RM is considered a super-set of ISO 13606 Reference model\[28\]. ISO 13606 Reference model defines the general characteristics of an EHR\[12\], named by ISO standards, as ISO 13606-1\[29\].

- **ISO 13606 Archetype interchange specification** (year of release 2008): this archetype specification was derived of openEHR AM\[26, 28\] also known as ISO 13606-2\[30\], where it includes the information’s architecture\[12\] and the ADL adoption\[24\].

- **ISO 13606 Reference archetypes and term lists** (year of release 2009): provides specification of archetype repositories and vocabularies for other parts of the standard\[12\], denominated as ISO 13606-3\[31\].

- **ISO 13606 Security** (year release 2009): deals with security requirements into EHR communications \[12\], named as ISO 13606-4\[32\].

- **ISO 13606 Interface specification** (year of release 2010): specifies the information architecture to enable the communication of EHR data, deals with message and service interfaces\[12\], designated as ISO 13606-5\[33\].

Both openEHR and EN ISO 13606 are archetype-native\[24\], since archetypes are considered a tool to gather professional consensus on how the clinical data must be represented in a consistent manner, and they are indispensable to achieve interoperable EHR\[27\].

openEHR and EN ISO 13606 propose a two-level modeling paradigm to represent EHR contents composed by RM and AM, denominated dual-model\[27\]. Where RM makes reference to the ISO 13606-1 and openEHR RM. The AM makes reference to openEHR AM and ISO 13606-2, where there are no differences related to each other because they are identical\[28\]. Dual-model aims to achieve semantic interoperability.

In 2012, the *Semantic Interoperability in Standardized Electronic Health Record Databases*\[34\] announced that HL7 was in the process of adopting the dual-model approach, and also disclosed that openEHR was trying to improve its ability to accommodate data in CDA form, making reference with CDA to the HL7 V3 CDA Release 2.

**Digital Imaging and Communications in Medicine (DICOM)**

DICOM, is one of the most widely deployed healthcare messaging standards for medical images (ISO 12052). DICOM is implemented in radiology, cardiology imaging, and radiotherapy devices, and its implementation is increasing in devices from other medical domains\[35\]. HL7 standards’ family and DICOM work closely, and both are promoted by the Integrating the Healthcare Enterprise (IHE) association\[36\].

Constructs seen in the DICOM data model are based in two ideas; they are object classes and service classes. In the first one, data as a patient’s demographics, the image data itself, among other data are represented within an informational object class. The second refers to a process where the data is generated, operated, or communicated; as storage, querying, retrieval, and printing of image services\[7\].
Terminologies and controlled vocabularies

To understand the information around an EHR in the same manner, it is not only necessary to have interoperability standards, it is also necessary to use medical terminologies and controlled vocabulary. In fact HL7 standards’ family use SNOMED-CT, LOINC and ICDx (where x can be version 9 or 10) as tools to achieve semantic interoperability\[23, 22\]. Also openEHR architecture is designed to use the same health terminologies\[26\].

A summary of terminologies and controlled vocabularies in medical environment will be described in the next subsections:

The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)

SNOMED-CT is the most comprehensive and precise clinical health terminology in the world\[37\]. It is developed, owned and distributed by the International Health Terminology Standards Development Organization (IHTSDO)\[12\]. The SNOMED-CT terminology handles a wide range of biomedical domains allowing a proper integration with the EHR information system\[27\]. The SNOMED-CT Components are\[37\]:

- **Concepts:** they represent clinical ideas.
- **Descriptions:** they link human readable terms to concepts. A concept can have several associated descriptions, where every description represents a synonym that describes the same clinical idea.
- **Relationships:** link each concept to other concepts that have a related meaning.

In 2008, SNOMED-CT release had 311,000 active concepts. The number of concepts continues growing\[37\], where every concept has a unique identifier that has been assigned\[23\], which is designated as *ConceptID*. SNOMED-CT was developed around a model of illnesses\[23\]; it provides explicit links to health related classifications and coding schemes as ICD-9 and ICD-10\[37\]. The International Classification of Diseases (ICD) is defined by the World Health Organization (WHO) as the standard diagnostic tool for epidemiology, health management, and clinical purposes, including the analysis of the general health situation of population groups.

Logical Observation Identifiers Names and Codes (LOINC)

LOINC was launched in 1994 by the researcher Clem McDonald from the Regenstrief Institute. Regenstrief organized the LOINC committee to develop LOINC. This is a common language for clinical and laboratory observations that enables the exchange of clinical results, outcome management and research\[38\].

LOINC clinical observation names are derived using six major axes\[39\], as follows\[38\]:

- The component/analyte, which consists of three sub-componentes: a) principal name, b) challenge of provocation, and c) any standardization or adjustment.
- Property measured: What is measure and its type.
- Timing: interval time of measured items.
- System, consists of two sub-parts: a) names the system, b) names a sub-part of the sample (not the patients).
- Scale: Specifies the scale of the measure, divided in four types: a) quantitative (Qn), b) ordinal (Ord), c) nominal (Nom), and d) narrative (Nar).
- Method: Specifies the type of method used to perform the test.
**RxNorm**

National Library of Medicine (NLM) created the RxNorm (the abbreviation of prescription is Rx). This is a standardized nomenclature for clinical drugs, that has been designated for use in the U.S. federal government systems, for the exchange of electronic clinical health information, across messages exchanged between systems that use different vocabularies[40]. The RxNorm is based on a model developed by NLM together with HL7, and the Veterans Administration (VA). This model establishes what the physician is going to prescribe, as well as the type of prescription that will be received by the pharmacy. Eventually RxNorm will cover all prescription medications approved for use in the United States[40].

**The relationship of HL7 with HIS’ sub-domains, terminologies and controlled vocabularies**

Our findings show that: HL7 standards’ family is considered the major EHR standards, because this family is more widely used in healthcare to resolve interoperability challenges[12, 21]; it is also one of the most adaptable standards family in the field of healthcare[6], in Figure 2, we propose the relationship of HL7 standards’ family with HIS’ sub-domains.

A HIS based on HL7 standard needs to generate messages for its sub-domains, they should include LOINC codes for an easy integration in clinical and research repositories[39]. One example of HL7 message is in the area of radiology, when it is necessary to obtain a type of imaging study from RIS to PACS[7] using DICOM to retrieve radiology reports[41].

Regarding PHRS, it can use HL7 PHR-S FM to create and manage PHRs, given that the PHR information is expected to be sent, received, or exchanged from multiple systems[16].

The records generated by LIS, RIS, PACS and PHRS are the data source for EHR Systems[7, 42] as shown in Fig. 2. Through EHR Systems, the above mentioned records are stored in EHRs, and then they are managed to achieve internal or external interoperability among HIS.

![FIGURE 2: Sequence diagram HL7.](image-url)
DISCUSSION

Varied tools to achieve interoperability are used worldwide. It was found that different standards may be used in the same continent; an example of that are two European projects, where one of them describes using the HL7 standards’ family in Portugal[21], while Sweden proposed a solution under the openEHR standard[28]. This is due because each country may impose or suggest using specific interoperability standards to achieve exchange of clinical data among institutions. This scenario generates interoperability problems in systems using different types of standards.

In Mexico’s case, the HL7 standards’ family is considered as the most common and flexible in the field of healthcare[6, 21]. Its use was established since project PROY-NOM-024-SSA3-2007[43] over the “Norma Oficial Mexicana” (NOM), where HL7 V3 and HL7 V3 CDA standards were included in guides and formats for the exchange of health information.

During a second phase and with the purpose of consolidating the use of standards, NOM-024-SSA3-2010 was issued[43], which is one of the most commonly used nowadays. The last version of this NOM was NOM-024-SSA3-298-2012[43]. Nevertheless, the normalized use of standards has been modified with time; in item a), section 3.2.1 of NOM-024-SSA3-2010, its use is defined by the word “must”, within the following paragraph: “Must use HL7 (Versión México) V.3.0 for interoperability purposes”[43]. However, when NOM-024-SSA3-2012 was issued, in subsection 6.1.3.1, the word “must” was substituted by: “the standards may be taken in consideration”, within the paragraph “Based on the extent of information exchange, the following standards may be taken in consideration: HL7 CDA, HL7 V3, XML and/or the proven standard determined by the Department...”[43]; leaving an option for developers and software users that manage clinical files to invest and participate in resolving interoperability problems, as is done in developed countries, while taking advantage of government actions to achieve this, and thus avoiding information systems to become isolated. Derived from the revision to NOM-024-SSA3-2010 and to NOM-024-SSA3-2012[43], certain incongruences were observed, such as the proposal to use “HL7 (Versión México) V.3.0.” or “HL7 (Capítulo México) V.3.0.” since no standards identified by that name were found, nor with an adaptation for Mexico. Engineer Adrián Pacheco López, Director of Telesalud of Centro Nacional de Excelencia Tecnológica en Salud (CENETEC-Salud), explained that that was a transcription error and that, for Mexico’s situation, its chapter has been closed since 2013.

Taking in consideration the current international status of using standards and the need for integration of developed systems to be used in Mexico, it is important to adopt intelligent algorithms supporting managing of large data volumes and diversity of standards, terminologies and controlled vocabularies to convert EHR systems in intelligent entities capable of deciding on documents received.

Up to this point, two problems affecting interoperability have been described: a) varied standards, and b) legislation lacking clear guidelines to define the use of standards; a third one is related to the technological basis used to store data and communications contributing to a free flow of large data volumes. These issues require collective responsibility to adopt normativity in the use of standards that goes beyond a simple suggestion, such as is presented in NOM-024-SSA3-2012.

CONCLUSIONS AND FUTURE WORK

The most relevant interoperability standards, controlled vocabularies and terminologies are described in this paper, in order to provide syntactic and semantic interoperability to EHR systems; with the understanding that it is important to first solve syntactic interoperability and then semantic interoperability. The type of semantic interoperability is very sensitive in all
areas of applicability, but even more so in a health environment; so it is vitally important to properly use interoperability tools. Further research is needed on this topic in order to learn how these tools work, either together or combined. In conclusion, and in order to achieve an interoperability environment, it is first necessary to solve the following: a) ethical, political and legal issues surrounding an HIS; b) issues that limit the adoption of different interoperability standards, terminologies and controlled vocabularies, and c) matters limiting the adoption of intelligent algorithms to manage data volumes, as well as standards for interoperability, terminologies and controlled vocabularies.

ACKNOWLEDGEMENTS

We would like to thank CONACyT (Consejo Nacional de Ciencia y Tecnología) and MyDCI UABC (Maestría y Doctorado en Ciencias e Ingeniería Universidad Autónoma de Baja California) for the assistance provided to make this paper, also a special thank you to Jeniffer Espinoza Hernández for her help with graphic design and figures throughout this paper.
REFERENCES


[38] Logical Observation Identifiers Names and Codes [Internet]. cited Feb 10, 2016. Available from: http://loinc.org


