XML in Health Information Systems

Justin Brewton, Xiaohong Yuan, Francis Akowuah
Department of Computer Science, North Carolina A&T State University, Greensboro, North Carolina, USA

Abstract
Advancing technologies in the healthcare industry has led to the idea of an electronic health record. This form of document will allow healthcare institutions to store patient information more efficiently. The technology that allows hospitals to create such a document is XML. This paper discusses the emergence of XML in the healthcare field and also the HL7 standard, which provides guidelines for the creation and sharing of these documents. Also discussed will be current issues regarding securing the XML language.

Keywords
Health information systems, security and privacy, XML, HL7

1. Introduction

The development of the Hypertext Markup Language (HTML) brought about a significant change in the way electronic documents were exchanged. The flexibility and simplicity of the language was key part in the growth of the World Wide Web. HTML focuses on separating text information from presentation information through the use of a tagging system. As websites became more widespread, the shortcomings of HTML began to be exposed. The major problem was that HTML had no means of representing structured data. Data elements that had a hierarchical relationship could not be efficiently represented in the language. In an effort to mitigate these problems, the Extensible Markup Language (XML) was created [11].

Initially XML was to take the place of HTML as the norm for the exchange of data and documents over the internet. However, HTML remained the standard for internet exchanges and XML found it's calling in facilitating exchanges in transaction-based systems and various other disparate systems. XML is considered a meta-language, meaning that it can be used to define a language [11]. A user constructs a new language by creating custom tags that are tailored for the type of data being manipulated.

In recent years, there has been a rapid increase in the development of health information systems motivated by legislation intended to protect patients' information and privacy, and the government's interests in reducing the cost and improving the quality of healthcare. Electronic health record allows healthcare institutions to store patient information more efficiently. XML has become a basic technology for implementing electronic health record and health information systems.

This paper introduces the basics of XML, and discusses Health Level 7 (HL7), an organization that sets standards. The Clinical Document Architecture (CDA) defined by HL7 is introduced, which provides guidelines for the creation and sharing of electronic health records. Current issues regarding securing the XML language is also discussed.

This paper is organized as follows. Sections 2 and 3 introduce the basics of XML and the advantages and disadvantages of XML. Section 4 discusses the history of patient records. HL7 is introduced in Section 5. Section 6 discusses security issues in XML and Section 7 concludes the paper.

2. XML Basics

The creation of an XML document may consist of three parts. The first of which is the data type definition (DTD). This layer describes the version of the data format, element descriptions, data structures, and some of the restrictions placed on the data. Essentially the overall format of the document is specified by the DTD. Here is a very simple example of a DTD that could hold a list of basketball players on a team:

```
1. < !ELEMENT player_list (player) *>
2. < !ELEMENT player (name, age, school? , country)>
3. < !ELEMENT name (#PCDATA) >
4. < !ELEMENT age (#PCDATA) >
5. < !ELEMENT school (#PCDATA) >
6. < !ELEMENT country (#PCDATA) >
```

Line one says that player_list is a valid element name and any instance of such element contains any number of player elements. The * signifies that there can be 0 or more player elements within the player_list element. The next line states that player is a valid element and any instance of this element should be followed by elements of type name, then age, then school (optional), and finally country. The ? character following an element signifies that the element is optional. Lines three, four, five, and six merely declare the elements name, age, school, and country as valid element types. The tag (#PCDATA) stands for parsed character data, meaning that the data is taken from what is entered by the author of the document. The following is an example of a document that conforms to this DTD:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE people_list SYSTEM "example.dtd">
<player_list>
  <player>
    <name>John Hooper</name>
    <age>23</age>
    <country>USA</country>
  </player>
</player_list>
```

The second part of the document is a detailed explanation of what the user created tags mean. The last layer of an XML document, which is optional, defines how the information will be presented [1]. Documents can be linked to use CSS or XSLT style sheet.

3. Advantages and Disadvantages of XML

The advantages of XML make it a viable solution to many of the data exchange problems that plague modern systems. There are many advantages to using a language like XML, but the major ones are:

- The ability to support user created tags allows the language to be fully extensible and void of any type of tag limitations. Since the language does not actually "do" anything, compatibility between systems is not an issue. As long as both systems can support the XML application that actually uses the document then the exchange of data is possible.

- Another key advantage of XML is its versatility. Any type of data can be modeled and tags can be created for very specific contexts.

There are also limitations to the XML that must be considered, such as:

- The lack of powerful applications that can process XML data and actually make the data useful is a primary disadvantage of the language. Only in recent history have browsers began to have the ability to read XML. Even now, these browsers still make use of HTML to render the XML document. This means that as of now, XML cannot be used as a language that is independent of HTML.

- Another disadvantage of XML results from the unlimited flexibility of the language. The tags implemented in a document are solely chosen by the creator. There is not a standard or generally accepted set of tags to be used in an XML document. As a result of this, designers can not just create general applications because each company will invariably have their own set of special tags and unique meaning for those tags.

4. History of Patient Records

The majority of medical institutions initially used paper to record various transactions that occurred. Doctors used and still use the traditional pen and pad to record any medical notes about a patient. The notes included general observations, possible diagnosis, and information about any follow up visits that need to be scheduled. In addition to medical notes, medical centers also needed to keep financial information about each patient for billing purposes. When considering the potentially high number of patients a doctor's office or hospital could encounter, the cost of materials to store their records could easily reach a very high value.

The first step to solving the cost problem was to incorporate information technology into the health care industry. Offices began to electronically deal with back-office operations such as billing. Dramatic reductions in cost resulted from this shift towards the use of electronic business systems. The success of the electronic business model sparked an even stronger focus on finding ways to integrate the latest technologies in information systems. The next major advancement was the creation of a system to digitize the process of Admitting, Discharging, and Transferring of a patient (ADT). These ADT systems provided health care facilities the ability to not only locate patients but also keep an accurate count of them [10].

The next logical step to create a fully digitalized health care system is to develop an electronic system that is capable of storing a patient's entire health history. It is at this point that the idea of the Electronic Health Record (EHR) becomes the focus of research. Imagine an electronic record that displays a patient's lab results, billing information, allergies etc. This type of record would serve to minimize costs and medical errors while increasing data accuracy and integrity. The ultimate goal of all this work on EHR is to create a system where information can be shared between patients and medical institutions and also back and forth between independent medical practices. This system does not require some huge data center because each practice will store its information remotely. However, there is a need for a standard that details how EHRs should be formatted.

5. Health Level 7 (HL7)

Health Level 7 (HL7) is an organization that sets standards and is accredited by the American National Standards Institute. This group is responsible for many communication standards used across America. Some of the standards created by this organization consist of:

- Arden Syntax – a grammar for representing medical conditions and recommendations
- Structured Product Labeling – the published information that accompanies a medicine
- Clinical Context Object Workgroup – an interoperability specification for the visual integration of user applications
- Claims Attachments – a standard health care attachment to augment another healthcare transaction

Their goal for the healthcare field is to provide standards for the exchange, management and integration of data that support clinical patient care and the management, delivery, and evaluation of healthcare services. In addition to creating messaging standards HL7 is also working on developing standards for the representation of clinical documents such as discharge summaries and progress notes. As a whole, these standards collectively make up the HL7 Clinical Document Architecture (CDA) [2].

The CDA aims at solving the previously discussed problem of finding a reliable and standardized means of storing and exchanging clinical documents. By specifying a mark-up and semantic structure through XML, the architecture works toward creating a universal way of allowing medical institutions to share clinical documents.

5.1 Clinical Document

A clinical document is defined as having these qualities:

- *Persistence* – A clinical document remains in an unaltered state for a user specified amount of time
- *Stewardship* – An entrusted person or party must have the responsibility of maintaining the document
- *Potential for authentication* – The document is intended to be legally authenticated
- *Wholeness* – Authentication applies to the whole document and not to just portions of the information
- *Human readability* – A clinical document should be human readable

5.2 Reference Information Model (RIM)

Currently HL7 version 3 is being developed. This family of standards includes The Clinical Data Architecture as well as rules for messaging. The newly developed version 3 allows clinical documents to contain not only text but also images, sounds, and other types of multimedia [3]. Both standards are implemented with XML and are derived from the Reference Information Model. The Reference Information Model or RIM is an object-oriented graphical depiction of clinical data and aids to understanding the life-cycle of events that messages and documents go through [6]. It focuses on five major themes:

- Ensure coverage of HL7 version 2.x. It ensured that it included all the information content of HL7 version 2.x.
- Remove unsubstantiated content from the model. It removed content from the draft that the technical committee did not originate and could find no rationale for retaining.
- Unified service action model (USAM). It introduced a concise, well-defined set of structures and vocabularies that address the information needs of a wide variety of clinical scenarios.
- Ensure quality. It addressed inconsistencies in the draft model and conflicts between the model and the modeling style guide.
- Address the "left hand side" of the model. It introduced powerful structures and vocabularies for the non-clinical portions of the model (patient administration, finance, scheduling).

Figure 1 [4] shows an example of RIM represented in graphical form.

5.3 The Hierarchical Structure of CDA

The actual architecture of the CDA can be thought of as a set of hierarchically related XML Document Type Definitions. As of now, only the top node, known as Level One, has been defined. Level one is designed to include enough detail to mark up narrative clinical notes. The objective of this level is to ease users into RIM. It is intentionally not very complex to allow deeper levels the ability to mark up the document even more. As seen in Figure 2 [6], level one material consists of the raw data gathered from an encounter. There are no high level medical codes or terminologies used [8].

Level Two, which has not been developed, will be a set of templates that can be layered on top of Level One. Level Two is envisioned to provide constraints to documents by

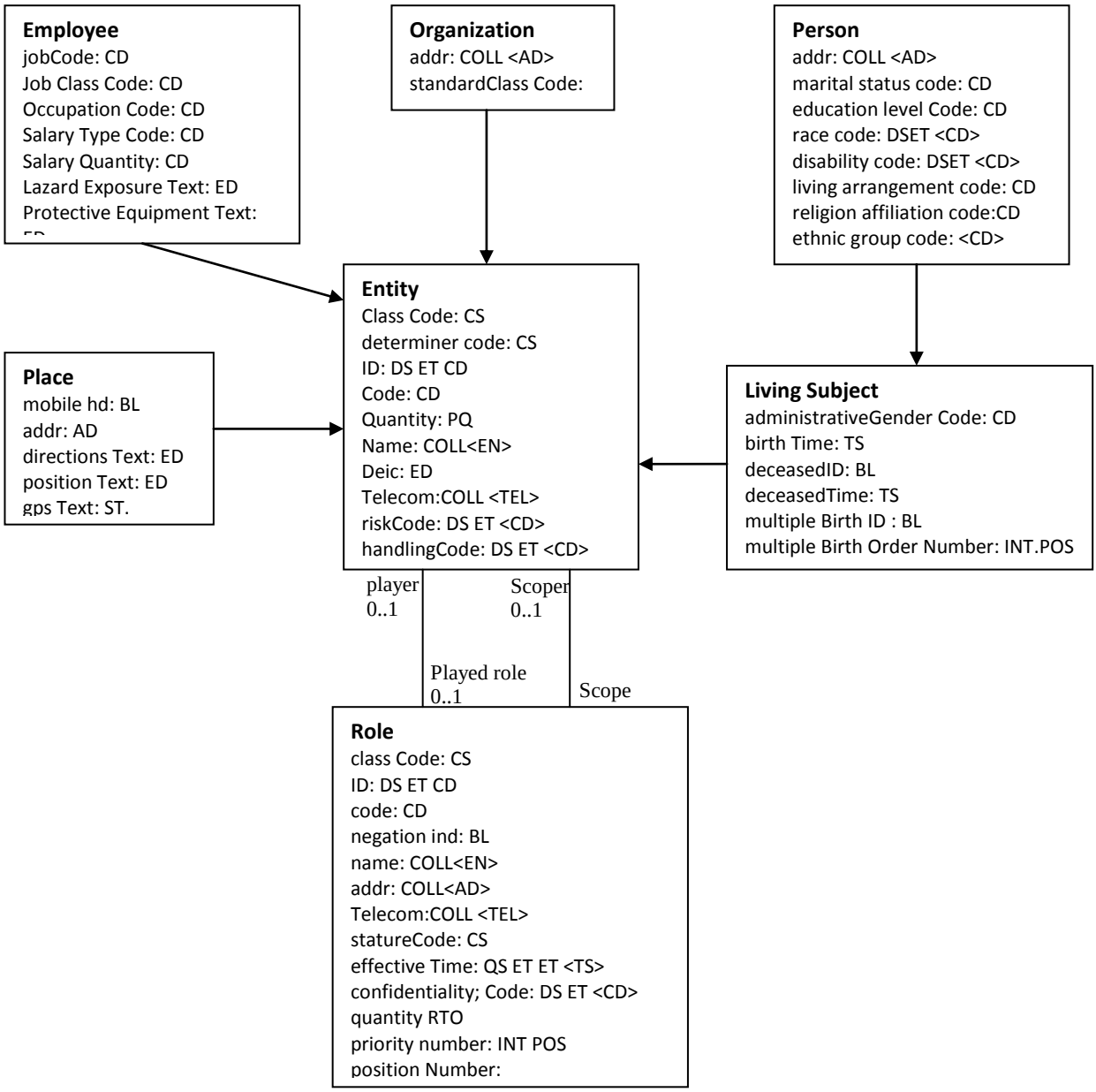

Figure 1. An example of Reference Information Model

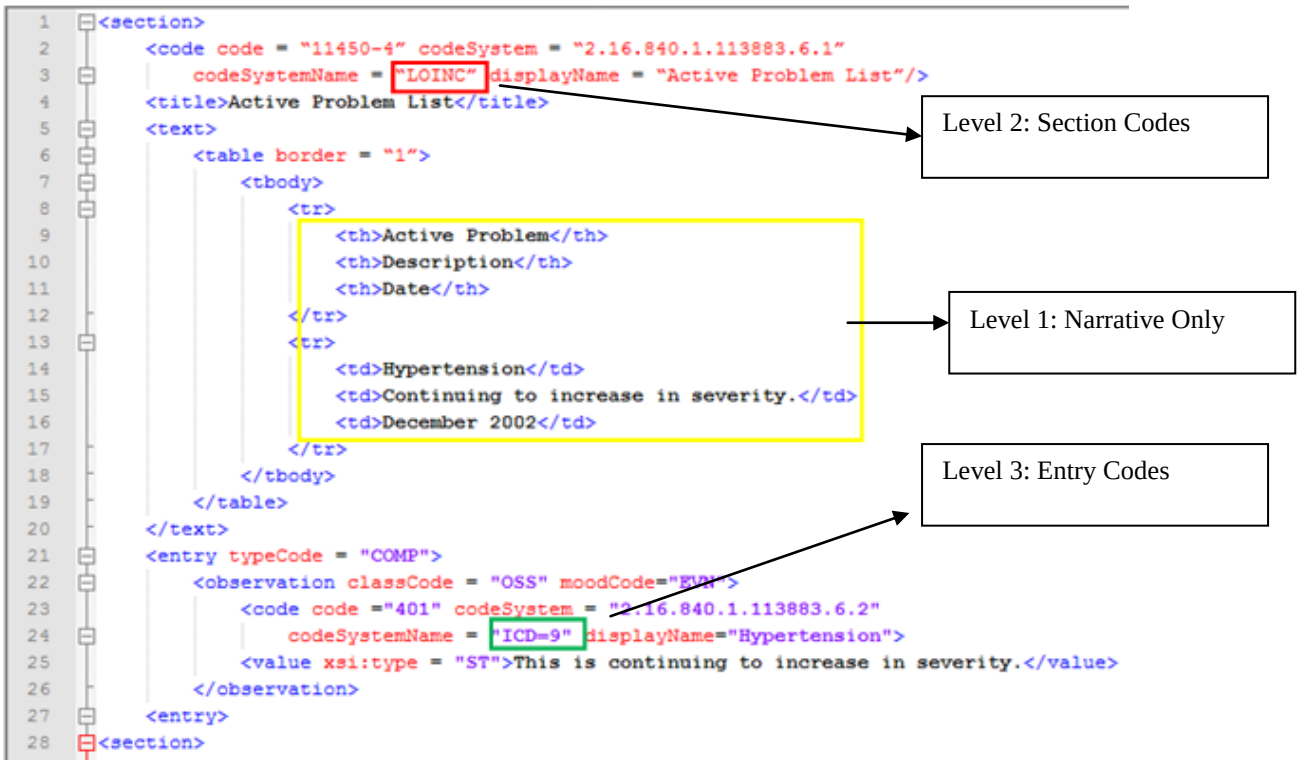

Figure 2. The hierarchical structure of CDA

requiring that specific types of medical documents contain a certain piece of information. For example one of these templates might require that a document of type "blood work" requires an "insulin level" section. As imagined, this type of structure would necessitate input from various professional groups to come up with an agreed upon template. Level Three information consists of specific medical codes used by healthcare institutions. A proper structure for information of such a deep level will require extensive collaboration between healthcare offices worldwide and the HL7 group.

5.4 CDA Document Structure

CDA documents are composed of a header and a body. The header is used to describe the context in which the document was created. CDA document headers serve three purposes:

1. Make document exchange possible within the same institutions and between separate institutions
2. Facilitate document management
3. Facilitate the compilation of an individual's complete medical history

The body of the document is made up of paragraphs, lists, and tables. Each of these sections can contain data, medical codes, and multimedia that describe the patient health care based transactions [4].

6. XML Security

When considering the security of XML documents, all the traditional qualities are desired: integrity, confidentiality, authorization etc. To achieve these goals XML data is treated much like any other types of data, in terms of security. XML data that is used to make up an individual's health record must be secure [9]. A patient's electronic health record could be potentially sent to many different institutions to be viewed by various doctors. Patients need to be sure that their personal information is only seen by an authorized party.

6.1 Digtial Signatures

XML Signatures operate identically to regular digital signatures [5]. A signature contains three sections:

- **SignedInfo:** Contains information about what part of the document is actually signed.
- **SignatureValue:** This is the output of the encryption of the data. It is the actual digital signature.
- **KeyInfo:** Provides the key or information on finding the key that validates the signature.

An XML signature allows the signing of a whole or specific section of a document. This standard provides integrity, message authentications as well as authentication for the signer of the document. Consider a patient that has been instructed by her physician to see a cardiologist. If the cardiologist makes any changes to the patient's EHR, only the section changed should be digitally signed by the cardiologist.

6.2 XML Encryption

The recommended encryption techniques to provide confidentiality for XML documents are not a replacement for security protocols such as SSL/TLS. Instead, XML encryption mandates requirements for areas not covered by SSL. More specifically XML allows for certain parts of the data to be encrypted and also provides security for sessions between more than two parties. Along with those two new areas covered, XML encryption still provides the traditional encryption methods. The need for an efficient encryption method, when dealing with healthcare documents is evident. The ultimate goal of this EHR revolution is to facilitate the exchange and storage of medical information. As new technologies make these tasks easier, the measures for securing this type of sensitive information must be strengthened.

6.3 Attribute Based Encryption (ABE)

Attribute Based Encryption (ABE) is an encryption method that works well with XML. ABE allows only users who have a specific set of attributes, which also match with the attribute set associated with a message, to decrypt the contents of that message. Just like with the traditional Identity Based Encryption method, user will be assigned a secret key by a central authority. However, the ABE secret key is based on the specific attributes of each user. When messages are created, the author creates a policy that corresponds with the ciphertext. The policy is just a Boolean statement that specifies the attributes a user must have to decrypt the information [7]. ABE secures message passing between separate entities. The following example explains the ABE method.

- Mary and John work for a company
- Mary is a *sales manager* and John work in the *IT department*
- When each employee's private key is assigned by the authority, the key contains attributes about their position(Mary- *Sales AND Manager* ; John – *IT*)
- A message is sent with a policy that maintains that only worker in the IT department are allowed to view it
- John's private key fulfils the policy, as a result, his key can decrypt the message. Mary is unable to view the message because her private key attributes do not satisfy the policy of the message.

When considering ABE's application to electronic health records, think about a patient that has all of his medical history contained in one document. After a routine visit of his regular physician an appointment to see a dermatologist is made because of a rash found on the patient's arm. As a result some of the patient's medical history needs to be sent to the dermatologist's office. By making use of ABE, the patient's entire medical record can be sent and the patient can be assured that only the dermatologist is able to view his personal information. To accomplish this, the record needs to be sent with a policy that allows only the individual with the dermatologist's credentials to view the document. In addition, ABE can also be used to guarantee that only information pertinent to the skin problem can be seen by the dermatologist. To achieve this all the information that the dermatologist is not allowed to see should be encrypted with a key that is different from the one used with the ABE.

7. Conclusion

It seems that in the future, the way medical information is stored and shared between institutions will be revolutionized. The extensibility and versatility of XML will be used as a catalyst for this advancement. The overall goal of creating an environment in which medical institutions can freely share information is far from becoming a reality but it is not impossible. The most important key for achieving this goal will be creating a standard for the storage of documents and for the method of sharing these documents between independent medical entities.

Organizations, like the HL7 group are essential to this process. As with many new technologies, as more people began to make use of this system, the necessity for creating a secure environment will increase. Equally important as creating a structure for storing and sharing medical information is the issue of securing this information. As developers continue to create more powerful process and actually make use of XML data, the usefulness of electronic medical applications will grow.

Acknowledgements

This work is partially supported by NSF under grant HRD-1137516, and by Department of Education under grant P120A090049. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation and Department of Education

References

[1] Wasim A Al-Hamdani, "XML Security in Healthcare Web Systems," *Information Security Curriculum Development Conference*, Kennesaw GA, 2010.

[2] Dolin RH, Alschuler L, Behlen F, et al. HL7 document patient record architecture: an XML document architecture based on a shared information model. Proc AMIA Symp. 1999:52–56.

[3] Dolin RH, Alschuler L. Approaching Semantic interoperability in Health Level Seven. Journal of the American Medical Informatics Association;Jan2011, Vol. 18 Issue 1, p99.

[4] Corepoint Health. CDA Architecture. [Online]. http://www.corepointhealth.com/resource-center/hl7-resources/hl7-cda.

[5] ABB Corporate Research. Standards for XML and Web Services Security. [Online]. http://www.tik.ee.ethz.ch.

[6] HL7 Standards. CDA Levels of Interoperability.[Online]. http://www.hl7standards.com/blog/2011/06/02/cda-levels-of-interoperability/

[7] Jin Li, Man Ho Au, Willy Susilo, Dongqing Xie, and Kui Ren. 2010. Attribute-based signature and its applications. In *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security* (ASIACCS '10). ACM, New York, NY, USA, 60-69. DOI=10.1145/1755688.1755697 http://doi.acm.org/10.1145/1755688.1755697

[8] Chien-Tsai Liu, Ann-Ging Long, Yu-Chuan Li, Kuo-Ching Tsai, Hsu-Sung Kuo, "*Sharing patient care records over the World Wide Web*," International journal of medical informatics 1 May 2001 (volume 61 issue 2 Pages 189-205)

[9] Hsiao, Tsung-Chih. *A Secure Integrated Medical Information System*. Journal of Medical Systems. *2011*.

[10] Gupta, V. and Murtaza, M. B. (2009), "Approaches To Electronic Health Record Implementation", The Review of Business Information Systems, 13(4), pg. 21

[11] Seals M. The use of XML in healthcare information management. J Healthc Inf Manag. 2000 Summer;14(2):85–95.