Electronic Health Record Standards - A Brief Overview

Article · December 2006

DOI: 10.1109/ITICT.2006.358222

CITATIONS
20

READS
947

5 authors, including:

Marco Eichelberg
OFFIS
142 PUBLICATIONS **1,138** CITATIONS
SEE PROFILE

Thomas Aden
OFFIS
11 PUBLICATIONS **381** CITATIONS
SEE PROFILE

Jörg Riesmeier
58 PUBLICATIONS **517** CITATIONS
SEE PROFILE

Asuman Dogac
Middle East Technical University
125 PUBLICATIONS **2,200** CITATIONS
SEE PROFILE

ELECTRONIC HEALTH RECORD STANDARDS – A BRIEF OVERVIEW

M. Eichelberg[1], T. Aden[1], J. Riesmeier[1], A. Dogac[2], G. B. Laleci[2]

1 OFFIS – Institute for Information Technology
Escherweg 2, 26121 Oldenburg, Germany
Phone: +49 (441) 9722-0 Fax: +49 (441) 9722-102
E-mail: {eichelberg|aden|riesmeier}@offis.de,

2 Dept. of Computer Eng., Software R&D Center
Middle East Technical University (METU)
06531, Ankara, Turkey
Phone: +90 (312) 210 5598 Fax: +90 (312) 210 5572
E-mail: {asuman|gokce}@srdc.metu.edu.tr

Abstract: Most medical information systems store clinical information about patients in proprietary format. To address the resulting interoperability problems, several electronic health record (EHR) standards that enable structured clinical content for the purpose of exchange are currently under development. In this article, we present a brief overview of the most relevant EHR standards, examine the level of interoperability they provide and assess their functionality in terms of content structure, access services, multimedia support and security.

Keywords: E-health, EHR standards, Interoperability.

1. INTRODUCTION

The electronic health record (EHR), which has been a key research field in medical informatics for many years, is defined by Iakovidis [1] as “digitally stored health care information about an individual's lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times”. The EHR includes information such as observations, laboratory tests, diagnostic imaging reports, treatments, therapies, drugs administered, patient identifying

information, legal permissions, and allergies. Currently, this information is stored in various proprietary formats through a multitude of medical information systems available on the market. Typical formats include relational database tables; structured document-based storage in various formats and unstructured document storage such as digitized hardcopies maintained in a classical document management system. This results in a severe interoperability problem in the healthcare informatics domain.

Making EHRs interoperable will contribute to more effective and efficient patient care by facilitating the retrieval and processing of clinical information about a patient from different sites. Transferring patient information automatically between care sites will speed delivery and reduce duplicate testing and prescribing. Automatic reminders will reduce errors, improve productivity, and benefit patient care. Furthermore, one of the prominent research directions in the medical field is about using genomics data for improving health knowledge and processes for prevention, diagnosis, treatment of diseases and personalization of health care. This also necessitates the interoperability of biomedical information and the EHRs.

2. EHR STANDARDS

To address the EHR interoperability problem, several standards and technical specifications are currently under development. These specifications aim to structure and markup the clinical content for the purpose of exchange.

2.1 CEN prEN 13606 EHRcom

The CEN standard EN 13606 “Electronic Healthcare Record Communication” (EHRcom) is a comprehensive EHR standard that is currently under development at the technical committee on Health Informatics of the European Committee for Standardization (CEN/TC 251). EHRcom is based on the older pre-standard (ENV 13606) and many concepts that have been adopted from openEHR. EHRcom, will be a five-part standard consisting of:

- The Reference Model,
- Archetype Interchange Specification,
- Reference Archetypes and Term Lists,
- Security Features, and
- Exchange Models.

Currently, however, only the reference model (EN 13606-1) is stable, whereas parts 2 through 5 are still working drafts. Therefore, the following discussion mostly focuses on the reference model as defined in [2]. The EHRcom reference model consists of four packages, which together describe the aspects of an EHR that are relevant for communication of EHR extracts between information systems. The *Extract* package defines the root class of the reference model (called "EHR_EXTRACT") and the data structures for EHR content. The *Demographics* package provides a minimal data set to define the various persons, software agents, devices and organizations that are referenced within the EHR extract. The *Access Control* package, which is under development as EN 13606-4, will define a representation for EHR access policies (such as consents for disclosure). The *Message* package, which is under development as EN 13606-5, will define the attributes required to communicate the EHR extract to a requesting process via a message or other serialized form.

Table 1. Logical building blocks of EHRcom

EHR	The electronic healthcare record for one person
Folders	High-level organization of the EHR, e. g. per episode, per clinical specialty
Compositions	A clinical care session, encounter or document e. g. test result, letter
Sections	Clinical headings reflecting the workflow and consultation process
Entries	Clinical "statement" about Observations, Evaluations, and Instructions
Clusters	Nested multi-part data structures (tables and interval time series) e. g. audiogram
Elements	Leaf nodes with single data values, e. g. reason for encounter, body weight
Data Values	Data types for instance values, e. g. coded terms, measurements with units

Table 1 shows the logical building blocks of EHR content according to [3]. The top level is a directory of possibly nested folders for a patient, allowing for a high-level organization of the EHR, for example, per episode or per clinical specialty. Folders contain zero or more ``compositions'' by reference. A composition (which roughly corresponds to one clinical document) may contain sections with section headers and entries which consist of elements or clusters of elements. Each element has a single value of a single data type. Content in the EHR extract is always added or replaced as a complete composition –

versioning, ownership and audit trail in EHRcom are based on the composition.

The second important building block for EHRcom is the archetype concept, which uses a two-level methodology to model the EHR structure. In the first level, a generic reference model that is specific to the healthcare domain but still very general is developed. This model contains only relatively few classes and must be stable over time. In the second level, healthcare and application specific concepts such as blood pressure, lab results etc. are modeled as archetypes, that is, constraint rules that specialize the generic data structures that can be implemented using the reference model. Archetypes allow to describe specific clinical concepts such as blood pressure or ECG measurements, as constraint rules that restrict the possible types, relationships and values of the record components in an EHRcom composition, or a part thereof. EN 13606-2 will define an archetype description language (ADL), that is, a formal language that is related to the EHRcom reference model. EN 13606-3 will contain a library of archetypes for various purposes, similar in principle to the library of Structured Reporting templates in part 16 of the DICOM standard. At this time it is not yet possible to make any statement on implementations or market acceptance of EHRcom since only the reference model is stable and EHRcom parts 2–5 are still under development. However, improvement of the architecture based on the lessons learned with ENV 13606 will certainly improve usability and acceptance over the 1999 pre-standard.

2.2 HL7 Clinical Document Architecture (CDA)

The Clinical Document Architecture (CDA) is a document markup standard developed by the Health Level Seven (HL7) organization. CDA defines structure and semantics of medical documents for the purpose of exchange. The documents are encoded in Extensible Markup Language (XML). They derive their meaning from the HL7 Reference Information Model (RIM) and use the HL7 Version 3 Data Types, which are part of the HL7 RIM.

Table 2. Levels of document granularity in CDA Release One and Release Two

CDA Release One	*CDA Release Two*
CDA Level One	Unconstrained CDA specification
CDA Level Two	CDA specification with section-level templates applied
CDA Level Three	CDA specification with entry-level templates applied

CDA distinguishes three different levels of granularity as shown in Table 2, where each level iteratively adds more markup to clinical

documents, although the clinical content remains constant at all levels. “Level One” focuses on the content of narrative documents with high-level context such as parties, roles, dates and time, places and structural organization of headings. It consists of two parts, the CDA Header and the CDA Body, which are based on the HL7 data types. The document header is derived from RIM and unambiguously defines the semantics of each entry in the document. The body contains the clinical document content, and can be either an unstructured text, or can be comprised of nested containers such as sections, paragraphs, lists, and tables through structured markup. Hence there is no semantics in “Level One” body; it offers interoperability only for human-readable content. In fact, CDA “Level One” describes a kind of HTML document with a standardized header that contains additional information on the document. “Level Two” CDA models the fine-grained observations and instructions within each heading through a set of RIM Act classes. With “Level Two”, it is possible to constrain both structure and content of a document by means of a template and thereby increase interoperability since the receiver ``knows what to expect''. However, a completely structured document where the semantics of each information entity is specified by a unique code will only be possible with “Level Three” providing for machine processing. Although CDA Release Two [4] does not use the term “Level” anymore, the basic architecture with structured documents of different granularity remains. This approach is intended to facilitate the migration from current free text documents to more structured CDA documents.

Unlike other standards HL7 CDA does not specify services or protocols that are used to exchange a document. From the perspective of HL7 messages, a CDA document is just a multimedia object than can be exchanged as a MIME (Multipurpose Internet Mail Extensions) package. Many national and international pilot projects use HL7 CDA Release One as a format for clinical documents. Also commercial products implementing CDA are starting to become available. Strictly speaking, the HL7 Clinical Document Architecture (CDA) is not an EHR standard since it only defines parts of an EHR architecture. However, CDA may form an important component of an EHR.

2.3 DICOM Structured Reporting

DICOM Structured Reporting (SR) is an extension to the DICOM (Digital Imaging and Communications in Medicine) standard that covers medical reports and other clinical data. Structured Reporting [5,6] is a general model for encoding medical reports in a structured manner in

the tag-based format used by the DICOM standard. SR can utilize the existing DICOM network infrastructure in order to archive and to communicate, to encrypt and to digitally sign structured reports with only relatively small changes to existing systems. In addition to the header information that is also used for DICOM images, the actual content of a structured report is represented by a document tree. Each content item (node) of the tree contains some piece of information, for example, a text paragraph or a reference to an image. A set of well-defined relationships describe how “parent” and “child” content items in the hierarchical document structure are related to each other. The semantics of most content items in the SR document tree is described by a machine-readable code and, therefore, enables computer-supported automatic evaluation and processing. The various terms and measurements can be encoded in a machine-readable way through annotation with codes taken from controlled vocabularies such as SNOMED or LOINC. DICOM SR provides a very flexible model to store almost any kind of data ranging from simple free text reports to completely structured documents with numeric measurement values and codes. In order to enhance interoperability in practice, the DICOM standard specifies document classes and other types of constraints for different medical applications. For example, the standard defines templates to harmonize the document structure and groups of codes to limit the choice for a particular context. This collection of standard templates, context groups and codes is called the DICOM Content Mapping Resource.

The DICOM standard does not specify how an SR document is actually rendered by an application. The visualization of reports for a human reader is regarded to be out of scope. Nevertheless, the application has to make sure that the full meaning of the report is conveyed in an unambiguous manner.

The most important fields of application for DICOM SR currently are the encoding of measurements performed at a modality and the documentation of CAD (Computer Assisted Detection) results. Correspondingly, current products focus on medical fields that are well-covered by the DICOM, that is, radiology and cardiology. Outside of the “imaging world” DICOM is not that common and, therefore, it is rather unlikely that Structured Reporting will become accepted as an EHR standard.

2.4 IHE Retrieve Information for Display (RID)

Retrieve Information for Display (RID) is a technical specification published by the Integrating the Healthcare Enterprise (IHE) initiative

[7]. It provides a simple and rapid read-only access to patient-centric clinical information that is located outside the user's current application. It supports access to existing persistent documents in well-known presentation formats. It also provides access to specific key patient-centric information such as allergies, current medications, summary of reports for presentation to a clinician.

In technical terms RID is a simple with a binding to HTTP GET and a description using the Web Service Description Language (WSDL). An “information source” is a system that provides an RID web service through which clients can access documents, and a “display” is a system that accesses the information source, retrieves patient-centric information or persistent documents, and displays them to a human observer. The focus of the integration is visual presentation, not a complete integration of the structured databases on which the actors might be based. Documents are exchanged in well-known presentation formats such as HL7 CDA Level One, PDF or JPEG. It is the responsibility of the information source to convert the healthcare specific semantics into a suitable presentation format. The display, on the other hand, may process and render this presentation format with only generic healthcare semantics knowledge, but will in general not be able to provide any processing of the healthcare information beyond document display. When combined with other IHE specifications such as Audit Trail and Node Authentication (ATNA) for network security, Cross-enterprise User Authentication (XUA) for the exchange of user credentials and Patient Identifier Cross-referencing (PIX) for patient ID look-ups, RID can be used to provide access to clinical documents both within and between enterprise boundaries.

The RID profile was initially published in August 2003 and has seen a rather quick market uptake. Prototype implementations from 22 different vendors have been successfully tested for their interoperability at the IHE cross-vendor testing events between 2003 and 2005, and several commercial products are already available on the market.

2.5 IHE Cross-Enterprise Document Sharing (XDS)

Cross-enterprise Document Sharing (XDS) is another IHE specification aimed at providing a document archive for the “longitudinal”, that is, life-long, cross-institutional healthcare record. XDS is document centric and “content agnostic” in the sense that any kind of document can be stored in an XDS archive, provided that the metadata for the document (for which XDS has a detailed specification) is available.

XDS uses an ebXML registry with one or more attached repository systems to implement the EHR archive.

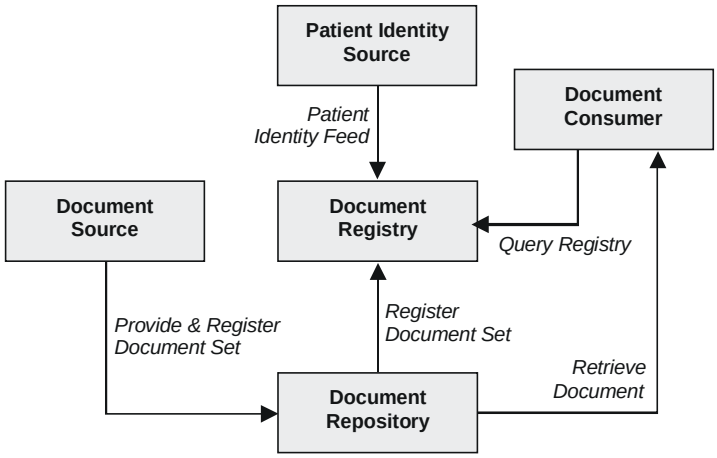

Figure 1. IHE XDS actors and transactions

The actors (systems) and transactions (interfaces) defined by the XDS profile are depicted in Figure 1. A *Document Source* represents a healthcare point of service system where care is provided and associated clinical information is collected. A document source provides clinical documents and metadata to one of the *Document Repositories* ("provide and register document set") which stored the document and forwards the metadata to the central *Document Registry* ("register document set"). A *Document Consumer* is a service application system where care is provided and access to clinical documents is needed. The document consumer queries the central registry for certain documents and receives, in response, a list of matching documents and their locations. Access to the documents is then possible through direct access to the document repository or repositories where the documents are stored ("retrieve document"). The *Patient Identity Source*, finally, is the central system that assigns and manages patient identifiers for one XDS installation (a so-called "affinity domain").

An affinity domain is defined as a group of healthcare enterprises that agree to work together for clinical document sharing. Such institutes agree on a common set of policies such as how the patients are identified, how the access is controlled, and the common set of coding terms to represent the metadata of the documents.

Since XDS handles healthcare documents in a content neutral way, a document may include any type of information in any standard format such as simple text, formatted text, images, or structured and vocabulary coded clinical information. Given this, to ensure the interoperability between the document sources and the document consumers, each clinical affinity domains also defines the set of supported document formats, their structure and the content.

Even though the XDS profile is rather new (the ``trial implementation draft'' was published in August 2004 and the final text has been released in August 2005), XDS prototype implementations from 28 distinct vendors have been successfully tested for their interoperability at the IHE cross-vendor testing events, including 5 different registries and 12 different repositories. The first commercial products are also already available. This indicates a significant interest from industry and indicates a quick market uptake to be very likely.

2.6 Medical Markup Language (MML)

The Medical Markup Language (MML) [8] has been developed since the mid 1990s by the EHR Research Group of the Japanese Ministry of Health and Welfare. Its purpose is to provide a standardized way to exchange medical documents and other clinical data. The current version 3.0 uses the XML-based HL7 Clinical Document Architecture Release One (CDA) format with a local header extension to store MML specific header fields and local markup to store the MML specific content.

MML documents can be exchanged via HL7 messages or by any other means of electronic communication. The local header contains many fields that are also stored in the CDA header (e. g. patient demographics, document creator and diagnostic information) but in a different format. Even the content of the document is duplicated to some extent in the local markup section of the document body. So currently, HL7 CDA is merely used as a standardized container that carries MML information. However, compared to CDA Release One, MML specifies more restrictions on the structure and the content of a document. The Medical Markup Language was never really used outside Japan and with the appearance of HL7 CDA there seems to be no reason that this will change in the future. However, within Japan, MML seems to be actively used and there are commercial products on the market that support MML.

3. DISCUSSION

The surveyed EHR related standards vary widely regarding their scope and content. While CDA and MML only specify a content format but no communication protocol, RID and XDS only specify communication protocols and are "content agnostic", that is, they do not define any content format. Only DICOM SR and EHRcom define both.

3.1 EHR Content Structure

Table 3 below summarizes the functionality of the EHR standards regarding content structure. RID and XDS are not shown in the table since they do not define a content structure of their own.

Table 3. Comparison of EHR standard content structures

	EHRcom	*CDA*	*SR*	*MML*
EHR contains persistent documents	⊕	⊕	⊕	⊕
EHR can contain multimedia documents	⊕	⊕	⊕	⊕
References to multimedia data in documents	⊕	⊕	⊕	⊕
Structured content suitable for processing	⊕	⊕	⊕	⊕
EHR supports archetypes / templates	⊕	⊕	⊕	⊕
Library of archetypes / templates	⊕	⊕	⊕	⊕
EHR specifies distribution rules	⊕	–	–	⊕
EHR standard covers visualization	–	⊕	–	–
Digital signatures on persistent documents	–	–	⊕	–

The properties of the four content standards are remarkably similar. All of them can be used to store persistent structured documents as well as multimedia content (images, signals and movies), which can also be references from other documents. All standards support the concept of a two-level modeling of EHR content using a simple reference model and an additional set of constraint rules (called archetypes, templates or content modules) that describe how certain clinical observations can be expressed in an unambiguous manner using structures of the basic reference model. All of the standards aim at specifying a library of standard archetypes or templates.

EHRcom and MML allow for a specification "distribution rules", that is, statements specifying under which circumstances the EHR content may be communicated. Distribution rules are a different concept from access control in that they are part of the EHR content, not part of an EHR access protocol.

Most EHR standards do not specify how EHR content should be visualized. The CDA specification gives at least recommendations on how documents are to be structured and encoded in order to facilitate the visualization process.

A final property in which the standards differ is the ability to attach digital signatures to EHR documents. DICOM SR is the only standard yet in which this is explicitly specified, although the other standards could make use of XML signatures.

3.2 EHR Access Services

Table 4 below summarizes the functionality of the EHR standards regarding access services. CDA and MML which do not currently define access services are not shown in the table.

Table 4. Comparison of EHR standard access services

	EHRcom	*SR*	*RID*	*XDS*
Service for querying EHR content	⊕	⊕	⊕	⊕
Service for retrieving EHR content	⊕	⊕	⊕	⊕
Service for submitting EHR content	⊕	⊕	–	⊕
Document-centric storage / retrieval	–	⊕	⊕	⊕
Content format agnostic	–	–	⊕	⊕

All four standards support the retrieval of EHR content, but RID is a “read only” service, that is, it does not support the submission of new documents to a repository. In all standards, except EHRcom, the persistent document is the basic unit of information that is queried, retrieved or submitted to the EHR. EHRcom instead uses the concept of an ``EHR extract'', which may contain several documents, for query and retrieval (but not necessarily for submission).

RID and XDS are content format agnostic, i. e. they treat documents as opaque byte streams and only process the metadata accompanying the document. While this makes the services easier to define and implement, it may also prevent support for advanced services beyond document visualization such as document processing, mediation or automated translation services. XDS addresses this problem with the concept of so-called XDS document content profiles, which restrict document formats and encoding options for specific clinical applications and, therefore, improve interoperability.

4. CONCLUSION

The evaluation of the EHR standards reveals no clear "winner". The content formats used are surprisingly similar in concept and capabilities, based on a two-level modeling approach with a simple reference model and constraint rules (archetypes, templates) for mapping clinical data onto the model. The standards differ in the progress achieved in the standardization process, but in principle, each of the content formats

seems to be suitable for implementing electronic healthcare records. However, the most likely candidates for a comprehensive EHR solution are either IHE XDS (with structured content in CDA or EHRcom format) or the complete EHRcom architecture including the EHRcom security features and exchanges models that are still under development.

5. ACKNOWLEDGEMENT

This work is supported by the European Commission through the RIDE project and in part by the Scientific and Technical Research Council of Turkey (TÜBITAK), Project No: EEEAG 104E013.

REFERENCES

[1] I. Iakovidis, “Towards Personal Health Record: Current situation, obstacles and trends in implementation of Electronic Healthcare Records in Europe”, International Journal of Medical Informatics vol. 52 no. 128 (1998), pp. 105 –117.

[2] CEN prEN 13606-1:2004. “Health informatics – Electronic health record communication – Part 1: Reference model”, Draft European Standard for CEN Enquiry, European Committee for Standardization, Brussels, Belgium.

[3] D. Kalra, “CEN prEN 13606, draft standard for Electronic Health Record Communication, and its introduction to ISO TC/215.” Document CEN/TC 251/WG I/N04-52, 2004.

[4] The HL7 Version 3 Standard: Clinical Document Architecture, Release 2.0, ANSI Standard.

[5] R. Hussein, U. Engelmann, A. Schröter, H.-P. Meinzer, "DICOM structured reporting: Part 1. Overview and characteristics" Radiographics vol. 24 no. 3 (2004), pp. 891–96.

[6] R. Hussein, U. Engelmann, A. Schröter, H.-P. Meinzer, "DICOM structured reporting: Part 2. Problems and challenges in implementation for PACS workstations", Radiographics vol. 24 no. 3 (2004), pp. 897–909.

[7] Integrating the Healthcare Enterprise, “IT Infrastructure Technical Framework Revision 2.0”. URL:
“http://www.ihe.net/Technical_Framework/index.cfm”.

[8] J. Guo, A. Takada, K. Tanaka, et al., “The Development of MML (Medical Markup Language) Version 3.0 as a Medical Document Exchange Format for HL7 Messages”, Journal of Medical Systems vol. 28 no. 6 (2004), pp. 523–533.