

Electronic Health Record (EHR) Standards Survey

Compiled by: Simona Cohen, Amnon Shabo

Date: August 1st, 2001

This report is a short survey about the main emerging standards that relate to EHR - Electronic Health Record. EHR, a major component of the health informatics domain, is defined as digitally stored healthcare information about an individual's lifetime with the purpose of supporting continuity of care, education and research. It includes such things as observations, laboratory tests, diagnostic imaging reports, treatments, therapies, drugs administered, patient identifying information, legal permissions, and so on. EHR systems will contribute to more effective and efficient patient care by facilitating the unification of clinical information from large populations of patients across care 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.

EHR is sometimes also called Computer-based Patient Record (legacy CPR), or Electronic Patient Record (EPR), or Electronic HealthCare Record (EHCR), or electronic medical record, or longitudinal health record.

Achieving implementation of EHR systems throughout healthcare is complex. Information is currently scattered across many computers (laboratory, billing, ECG carts, word processors) within institutions and across institutions (nursing facilities, pharmacies, hospitals, physician offices). To make this information useful, standards are needed for the three EHR parameters:

- healthcare messages exchange
- EHR object model (i.e. content and structure)
- healthcare terminology/ vocabulary.

In addition, the EHR system should incorporate a proper security mechanism.

There are three main organizations that create standards related to EHR- HL7, CEN TC 215 and ASTM E31. HL7, operating in the United States, develops the most widely used health care-related electronic data exchange standards in North America, while CEN TC 215, operating in 19 European member states, is the preeminent healthcare information technology standards developing organization in Europe. There is a memorandum to intensify collaboration between the two groups and move toward the development of technically identical and interchangeable U.S. and European standards. Both HL7 and CEN collaborate with the ASTM that operates in the United States and is mainly used by commercial laboratory vendors.

Our survey shows that while there exist standards for healthcare message exchange and terminology, there is no standard for EHR object model. HL7 is in the initial phases of defining an EHR content. CEN TC 215 suggest an EHR structure but does not define its content. ASTM E31 is the closest to an EHR object model, but it still lacks a lot of essential information and it's not clear how to extend the given model.

The human genome has been sequenced and the research community is working on correlating patient diseases with the patient genes to discover new treatments and develop personalized

medications. The EHR object model should include the patient genomic information to facilitate such treatments. None of the standards has addressed yet the challenge of encoding DNA, proteins, and polymorphism features in the EHR.

The following table summarizes how the three standards organizations support the three parameters of EHR:

	HL7	CEN TC 215	ASTM E31
Messaging	HL7 V3	ENV 13606 - 4	ASTM E1238, ASTM1394, ASTM E1467
EHR Object Model	no standard	partially: ENV 13606 - 1, ENV 13606 - 2	partially: ASTM 1384
Terminology	LOINC, SNOMED, UNLS, etc.	no standard	ICD9, SNOMED, etc.

Sections 1-3 details how HL7, CEN TC 215, and ASTM E31 respectively support each EHR parameter. Section 4 details some external coding schemes that are used by HL7 and ASTM E31 and may be used by CEN TC215 too.

1. HL7

Founded in 1987, Health Level Seven, Inc. (<http://www.HL7.org>) is a not-for-profit, ANSI-Accredited Standards Developing Organization that provides standards for the exchange, management and integration of data that supports clinical patient care and the management, delivery and evaluation of healthcare services. Its 2,200 members represent over 400 corporate members, including 90 percent of the largest information systems vendors serving healthcare.

1.1 HL7 Messaging

HL7 final standard is version 2.4 while version 3 is under development and slated for publication in December 2001. The V2.x series of messages were widely implemented and very successful. These messages evolved over several years using a "bottom-up" approach that has addressed individual needs through an evolving ad-hoc methodology. Version 3 uses an object-oriented development methodology and a Reference Information Model (RIM) to create messages. The RIM is an essential part of the HL7 Version 3 development methodology, as it provides an explicit representation of the semantic and lexical connections that exist between the information carried in the fields of HL7 messages. It is an all-encompassing, open-umbrella look at the entire scope of healthcare IT, containing more than 100 classes and more than 800 attributes used to create HL7 messages.

The new features in HL7 version 3 are:

- Using a Reference Information Model (RIM)

- Reducing optionally
- Testable criteria for application conformance
- XML-based messages
- Message Development Framework
- Object Oriented (main entities are classes)
- Interaction models
- Expanding the HL7-recognized coding schemes (e.g., SNOMED)

1.2 HL7 EHR Object Model

HL7 does not have a standard yet for EHR object model. HL7 V2.x states that it is silent on the actual logical and physical construction of the patient longitudinal health record. In version 3 this intention is going to change with the establishment of a new EHR SIG that was chartered on May 2001 to create a high level architecture that supports EHR requirements. There is also a Medical Records/Information Management technical committee that its main purpose is to produce an accurate, legal, and legible document that serves as a comprehensive account of healthcare services provided to a patient. However, the committee is currently focused on document management, i.e., messages that contain information about a clinical document and sometimes contain the document itself.

Although the HL7 lacks a standard for EHR object model, there are some evolving standards and proposals that give some ideas regarding an EHR object model, namely the Clinical Document Architecture (CDA) standard and the Document Ontology Task Force (DOTF) proposal. The CDA standard, which was until recently known as the Patient Record Architecture (PRA), provides an exchange model for clinical documents (such as discharge summaries and progress notes), and brings the healthcare industry closer to the realization of an electronic medical record.

CDA documents can be structured in three levels: CDA level 1 represents a general clinical document, CDA levels 2&3 represent levels of specialization. At the time this report is written, only level 1 is approved while level 2&3 are still in the initial development phase. Another effort in HL7 is to develop Domain Information Model (DIM), which represents a detailed model of a single 'area of interest' such as pharmacy, and Refined Message Information Model (RMIM) for each domain. CDA level 3 will probably use the DIM and RMIM of the domain, and constitutes a good basis for defining the EHR content.

The DOTF proposal suggests a classification of the CDA documents. It proposes a polyhierarchical taxonomy of document names, where each name is derived from a set of terminology axes. The axes that are currently proposed are service such as hospital administration, condition such as endocrine system disease, clinical category such as cardiology, and practice setting such as hospital unit. This classification constitutes a good basis of an EHR structure.

1.3 HL7 Terminology

The RIM which is the data model used across all HL7 standards support the use of terminology from any external coding scheme such as LOINC for observations, SNOMED for procedures, UMLS for medication treatments, etc. HL7 has also formed a vocabulary technical committee that provides an organization and repository for maintaining coded vocabulary for various domains.

2. CEN TC 215

CEN TC 215 on health informatics (<http://www.centc251.org>) is an European organization with the scope to achieve compatibility and interoperability between independent health systems and to enable modularity. This includes requirements on health information structure to support clinical and administrative procedures, technical methods to support interoperable systems as well as requirements regarding safety, security and quality. TC 215 includes four working groups: WG1 “Information Models”, WG2 “Terminology”, WG3 “Security, Safety, and Quality”, WG4 “Technology for interoperability”. The pre-standard ENV 13606 “Electronic healthcare record communication” is part of WG1 and has four parts that relates to EHR. The standards are in accordance with the ISO standard organization.

2.1 CEN TC 215 Messaging

The pre-standard ENV 13606 part 4 “Messages for the exchange of information” specifies messages that enable exchange of electronic healthcare record information between healthcare parties responsible for the provision of clinical care to an individual patient. There are three types of messages: request EHR message, provide EHR message, and EHR notification message. All messages contain the attributes: identification of message by originator, issue date and time of message, urgency of message, message receipt acknowledgment request, and optionally the attributes: comment on message, and language . All messages contain the classes: EHR source, EHR destination, patient matching information, and optionally the classes: EHR message related agent, healthcare agents directory, and message reference.

The Domain Information Model (DIM) is a set of classes that consist of the conceptual model describing common concepts and their relationship for all EHR messages. The DIM is divided to eight subsystems such as communicating parties subsystem, message subsystem, structured record subsystem, etc. Specialization of the DIM is provided by a General Message Description (GMD) which contains a subset of the DIM prescribing the information content and semantic structure of a message used to meet one or more identified information interchange requirements. There is a separate GMD for each type of message that shows the hierarchical decomposition of each message into its main classes.

2.1 CEN TC 215 EHR Object Model

CEN TC 215 defines the structure of an EHR but not its content. Pre-standard ENV 13606 part 1 “Extended architecture” suggests to structure a health record in mainly Original Component Complex (OCC) components. Each OCC represents an aggregation of other record components that is determined by the time and situation in which they were originally added to the EHR. The OCC is an abstract class with four types of specialisation:

- Folder OCC - high-level subdivisions of the entire EHR for a patient, usually grouping entries over long time-spans within one organisation or department, or for a particular health problem. For example, second pregnancy, nursing record, GP record, inpatient stay, diabetes care record.
- Composition OCC - a set of record entries relating to one time and place of care delivery; grouped contributions to an aspect of health care activity; composed reports and overviews of clinical progress. For example, consultation record, orthopedic record, operation notes, discharge summary, vital signs chart.
- Headed Section OCC - subdivisions used to group entries with a common theme or derived through a common healthcare process. For example, current medication, past medical history, presenting symptoms, examination findings, treatment plan.
- Cluster OCC - low-level aggregations of elementary entries (record items) to represent a compound clinical concept. For example, blood pressure measurement (systolic & diastolic), heart sounds, differential white cell count, insulin schedule.

Pre-standard ENV 13606 part 2 “Domain term list” provides a set of categories for some record component names that can be used to identify or to group similar kinds of record entries. This is primarily to allow EHR entries or extracts originating from a range of different clinical teams to be processed and handled in a homogeneous fashion for record navigation, for searching, and for high-level aggregation analyses. The pre-standard defines also archetypes that suggest how to construct categorical structures in a controlled way, and some fine grained descriptors that may be organized according to the predefined categorical structures. While ENV 13606-1 describes the structure of EHR, these archetypes and fine grained descriptors described in ENV 13606-2 partially suggests some content of the EHR.

2.1 CEN TC 215 Terminology

The provisions within parts 1 and 4 of ENV 13606 anticipate that record components will be named using controlled vocabularies. These will be determined by local enterprises or national and international bodies, and ENV 13606 does not provide any such list.

3. ASTM E31

Established in 1970, American Standards for Testing and Materials (ASTM) Committee E31 on Healthcare Informatics develops standards related to the architecture, content, storage, security, confidentiality, functionality, and communication of information used within healthcare and healthcare decision making, including patient-specific information and knowledge (see <http://www.astm.org/cgi-bin/SoftCart.exe/COMMIT/COMMITTEE/E31.htm?L+mystore+txxf8000+970175387>)

3.1 ASTM E31 Messaging

ASTM has several standards for messaging in the healthcare domain as follows:

- ASTM E1238 “Standard Specification for Transferring Clinical Observations Between Independent Computer Systems” - used by most of the largest commercial laboratory vendors in the US to transmit laboratory results.
- ASTM 1394 “Clinical Laboratory Instruments to Computers” - has been developed by a consortium consisting of most US manufacturers of clinical laboratory instruments and is being implemented in the current laboratory instruments generation.
- ASTM E1467 “Standard Specification for Transferring Digital Neurophysiological Data Between Independent Computer Systems” - defines codes and structures needed to transmit electrophysiologic signals and results produced by electroencephalograms and electromyograms . The standard is similar in structure to ASTM 1238 and HL7, and is being adopted by all of the EEG systems manufacturers.

3.2 ASTM E31 EHR Object Model

ASTM E1384 “Standard Guide for Content and Structure of the Electronic Health Record” partially defines an EHR object model. The clinical heart of the EHR is the core of the entities (objects): patient, provider, problem, encounters, orders, services and observations. Fig 4 in the standard define the relationship among those entities. Then, there are about 700 attributes, divided into categorized segments that are mapped to the different objects. Table 4 in the standard defines which segment of attributes belongs to which entity e.g. demographics segment belongs to the patient entity, health history segment belongs to the observation entity.

3.3 ASTM E31 Terminology

ASTM facilitate the use of external coding schemes such as ICD9, SNOMED, READ. Also each attribute has a code that its prefix is the segment number to which this attribute belongs. For example, segment 5, the problem list segment, includes attributes such as 05001. Problem ID, 05001.01 Problem Statement, 05001.05 Problem Cause, etc.

4. Coding Schemes

This section describes some popular coding schemes that may be used within other standards, in particular within HL7, CEN 215, and ASTM E31.

4.1 ICD

ICD stands for "International Classification of Disease" and its 9th revision is the most common one (the 10th edition was not well adopted). It was originally published by the World Health Organization (WHO). The ICD-9-CM (Clinical Modification) was developed by the National Center for Health Statistics for use in the United States. It is based on the WHO international

ICD-9. A version based on ICD-10 (ICD-10-CM) is in preparation. For an online version see <http://www.eicd.com/EICDMain.htm>.

4.2 SNOMED / ReadCodes

SNOMED stands for "Systematized Nomenclature of Medicine" and is developed by SNOMED International - a division of the College of American Pathologists (CAP). SNOMED is aimed at being a comprehensive, multi-axial, controlled terminology created for the indexing of the entire medical record. The new version is called SNOMED RT Reference Terminology).

SNOMED CT (Clinical Terms) is aimed at specifying the core file structure of SNOMED Clinical Terms. This new collaborative terminology is being developed jointly by the National Health Service (NHS) and the College of American Pathologists, from a basis of Clinical Terms (the Read Codes) and SNOMED RT. It will be available by the end of 2001. For an on line version see <http://www.snomed.org>.

4.3 LOINC

LOINC stands for Logical Observation Identifiers Names and Codes and includes identifying individual laboratory results (e.g. hemoglobin), clinical observations (e.g. discharge diagnosis), diagnostic study observations (e.g. chest x-ray impression).

4.4 UMLS

UMLS stands for "Unified Medical Language System" and is developed by the USA National Library of Medicine (NLM). The UMLS project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs. The UMLS **Metathesaurus** is one of three knowledge sources developed and distributed by the NLM as part of the UMLS project. The Metathesaurus contains information about biomedical concepts and terms from many controlled vocabularies and classifications used in patient records, administrative health data, bibliographic and full-text databases and expert systems. It preserves the names, meanings, hierarchical contexts, attributes, and inter-term relationships present in its source vocabularies; adds certain basic information to each concept; and establishes new relationships between terms from different source vocabularies. For an on line version see <http://www.nlm.nih.gov/research/umls/umlsmain.html>