Standards for Computerized Clinical Data: Current Efforts and Future Promise

By Jonathan Y Lukoff, MD; Robert H Dolin, MD

Abstract
Use of an electronic health record (EHR) will help us to realize the full potential of modern medical care. To optimize the functionality of a “virtual” record, universal informatics standards are needed. Standards for coded medical terminologies and for a common representation of clinical data will allow patient information to be transmitted clearly and unambiguously between different computers and different software applications in a secure form which is easily searched, interpreted, and manipulated—and thus most useful. Many of these standards are key components of Kaiser Permanente’s national Clinical Information System (KP CIS).

Introduction
As practicing clinicians in the 21st century, we have become used to change. Just a few years ago, many of us discovered the value of applying to our practice the concepts and terminology taught in business school: “seamless,” “Total Quality Management,” and “transparent.” Now we find ourselves confronted with a set of unfamiliar terms from a new branch of medicine, medical informatics—a field created to study and advance the science of efficiently recording and retrieving medical information.

An increasingly familiar creation of medical informatics is the electronic health record (EHR) containing medical data, ie, information from patient charts, laboratory reports, and radiology reports. To ensure optimal functionality of this electronic record, it must be unambiguous, universally available, transmissible, exchangeable with other EHRs, searchable and researchable, manipulable, secure, and must conform to governmental requirements set forth in regulations. (By “manipulable” we mean that the EHR should allow for automated reminders based on the data being processed and stored. In addition, the EHR will facilitate outcomes research,1 enable more complete documentation of quality of care delivered, and enable automatic documentation of our level of service to help assure appropriate compensation for services delivered.) These features can be achieved by development and implementation of universal standards for medical informatics.

A new vocabulary of acronyms has been developed to represent medical informatics standards...
ized. Records will be directly input via keyboard devices; structured data entry will be automated by use of templates; and manual input will be bypassed through use of optical character recognition scanning, automatic voice recognition, direct transmission from laboratory machines, and other means.

Health Level 7 (HL7)
Accredited by the American National Standards Institute (ANSI), Health Level Seven (HL7)\(^4\) is an organization whose mission is to develop standards (not software) for unambiguous transmission of clinical and administrative health care information between computers. According to the organization’s mission statement, HL7 works ‘‘to provide standards for the exchange, management, and integration of data that support clinical patient care and the management, delivery, and evaluation of health care services. Specifically, to create flexible, cost-effective approaches, standards, guidelines, methodologies, and related services for interoperability between health care information systems.’’\(^4\)

**Tools for Standardizing Transmission of Electronic Medical Data**

The Reference Information Model (RIM)
The most widely used standard being developed by HL7 is a messaging standard that enables disparate software applications to exchange clinical and administrative health care data. While interpreting medical communications as multiple discrete messages, HL7 will assign varied types of data (eg, laboratory test results) to predefined locations to show clearly the type of information intended by the user. HL7 will also define relations between data; thus, a given laboratory value can remain correctly linked with a specific patient. HL7 has recognized that designing a complete and usable standard requires regulated criteria for establishing vocabulary and for transmitting data.

As part of its development process, HL7 has created an object model—the HL7 Reference Information Model (RIM)—to represent clinical data pictorially and to identify the life cycle of events carried by a message or by groups of related messages. The RIM thus is used to create a messaging standard. Stated simply, the RIM defines fields (blank areas) that are designed to contain standardized vocabularies meeting certain requirements.\(^4\)

The RIM encompasses the entire domain of health care services, including laboratory and pharmacy services as well as patient admission, discharge, and transfer to and from health care facilities. The RIM has been applied most widely to laboratory data allowing information to be clearly and precisely located so that each laboratory result is clearly associated with a specific laboratory test and with a specific patient: For example, a practitioner must be certain that the potentially ambiguous phrase “patient X’s potassium” designates a laboratory result and not a prescription—and that it refers to the laboratory value of patient X and not someone else’s. HL7 has expanded the RIM to allow unambiguous transmission of more types of information within messages and clinical documents.

<table>
<thead>
<tr>
<th>Table 1. Glossary of computer science terms relevant to medical informatics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acronym</strong></td>
</tr>
<tr>
<td>ANSI</td>
</tr>
<tr>
<td>CDA</td>
</tr>
<tr>
<td>EHR/EMR</td>
</tr>
<tr>
<td>HIPAA</td>
</tr>
<tr>
<td>HL7</td>
</tr>
<tr>
<td>LOINC</td>
</tr>
<tr>
<td>Metastructure</td>
</tr>
<tr>
<td>SNOMED</td>
</tr>
<tr>
<td>Syntax</td>
</tr>
<tr>
<td>W3C</td>
</tr>
<tr>
<td>XML</td>
</tr>
</tbody>
</table>
The Clinical Document Architecture (CDA)

The expanded capability of the RIM includes use of the Clinical Document Architecture (CDA), a model for exchanging clinical documents (ie, medical records). Derived from the RIM, the CDA converts documents into a format which can be read by machines (ie, for electronic processing) as well as by humans.\textsuperscript{4,5} The CDA standards being developed by HL7 can be used to represent clinical documents such as progress notes, discharge summaries, and results of physical examinations.

It is hoped that computerized medical records (ie, the EHR) will be designed to use the CDA standard. The CDA organizing framework can be used to ensure clear, unambiguous representation of all patient information which is input into a computer and displayed via any software (ie, an EHR developed by the same or a different vendor) adhering to the same standard (ie, HL7’s CDA). Thus, by following the HL7 CDA, any programmer will be able to design an EHR which can be transmitted over computer networks such as the Internet and which can be automatically integrated into any other EHR written to the HL7 CDA standard.

XML: A tool for Enhancing Data Transmission over the Internet

To be widely available, information must use a syntax, or rules governing construction of a machine language, which allows transmission over the Internet. The World Wide Web Consortium (W3C)\textsuperscript{6} created XML (Extensible Markup Language),\textsuperscript{7,8} a data representation standard (or open-standard metastructural computer language) which allows information transmitted over the Internet to be clearly interpreted by the receiver of that information.

XML is also a proper, easier-to-use subset of the Standard Generalized Markup Language (SGML), which is used to create HyperText Markup Language (HTML)—the programming code used to encode material for visual presentation as Web pages. (‘Surfing the Web’ thus involves transparent interaction with SGML.) A standardized syntax like XML enables transmission of HL7 information over the Internet. Computer metastructures such as XML extend the capabilities of computer languages, enhance representation of structured messages, and improve syntactic interoperability. Metastructures embed data “tags” (field names) into the data so that they are hidden from the clinician. These tags automatically instruct the computer where and in what format to place the data to be received by the person using the information (eg, laboratory test results or radiology reports).\textsuperscript{9,10} These metastructure tags enable Web browser software to display information clearly and unambiguously (eg, as text headings) (Figure 1). The content to be displayed (eg, each field value) is contained within the opening and closing tags.

KP CIS currently uses HL7 Version 2 messages. HL7 Version 3 standards, which are derived from the Reference Information Model and are transmitted in XML (including both messaging standards and CDA), are fairly new and are not currently part of KP CIS.

<table>
<thead>
<tr>
<th>Example of XML format</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;PHYSICIAN&gt;</code></td>
</tr>
<tr>
<td><code>&lt;FIRST-NAME&gt;BOB&lt;/FIRST-NAME&gt;</code></td>
</tr>
<tr>
<td><code>&lt;LAST-NAME&gt;DOLIN&lt;/LAST-NAME&gt;</code></td>
</tr>
<tr>
<td><code>&lt;/PHYSICIAN&gt;</code></td>
</tr>
<tr>
<td><code>&lt;PATIENT&gt;</code></td>
</tr>
<tr>
<td><code>&lt;FIRST-NAME&gt;BOB&lt;/FIRST-NAME&gt;</code></td>
</tr>
<tr>
<td><code>&lt;LAST-NAME&gt;DOLIN&lt;/LAST-NAME&gt;</code></td>
</tr>
<tr>
<td><code>&lt;/PATIENT&gt;</code></td>
</tr>
</tbody>
</table>

Figure 1. Example of XML (Extensible Markup Language) format: Patient Bob Dolin and physician Jon Lukoff could be represented in XML in this way. The XML conventions enable programmers to specify options that determine a document’s display format, semantic content, and context. This format, authored by another, is no more difficult to use than many other computer programs. Work completed in the standardized format is transmissible over the Internet for display on World Wide Web sites.

Structural Components of Standardized Clinical Vocabularies

LOINC (Logical Observation Identifier Names and Codes)\textsuperscript{11} and SNOMED (Systematized Nomenclature of Medicine)\textsuperscript{12} are standardized medical vocabularies that have been accepted internationally and are foundational components of KP CIS. LOINC is a standardized set of names and codes for laboratory tests and clinical observations which was developed in mid-1995 and which has gained wide acceptance.\textsuperscript{11} The LOINC database encompasses more than 14,000 codes. To completely characterize the components of laboratory terminology, they are classified into six axes (subject headings): component or analyte (ie, what is measured), property of the component or analyte (eg, its concentration), time aspect of test, system (sample) type, type of measurement scale (ie, quantitative or qualitative), and type of test method.\textsuperscript{11}

SNOMED is a reference medical terminology set developed more than 20 years ago and enhanced continuously ever since.\textsuperscript{12} Intended to completely and logically interrelate groupings of defined medical terms, SNOMED is a formalized, information-packed set of more than 300,000 coded medical terms.\textsuperscript{13} LOINC has more complete defining characteristics for laboratory result data than SNOMED, but the two ter-
Semi-automated process should allow us to relieve our clinicians from the burden of coding their patient encounters. In addition, the National Health Service [United Kingdom] READ Codes have been combined with SNOMED RT to form SNOMED CT (Clinical Terminology). Other specialized vocabularies will be integrated or mapped to SNOMED CT as necessary to allow for full interoperability of information systems across the broadest possible range of medical needs. Participants in the CMT Project plan to develop a “comprehensive strategy for representing detailed laboratory terms as well as appropriately classifying terms.”

Discussion

Further refinement and widespread application of standards for medical informatics will give authorized personnel access to this medical information anytime through the Internet. Why should we—and how will we—further this goal?

Medical informatics standards are critical for design of terminologies, which are increasingly used to populate clinical databases. These databases affect data retrieval for many clinical purposes, such as patient care, audit, research, decision support, epidemiology, and management. In addition, terminologies designed from informatics standards are important for populating databases such as those used for determining eligibility for insurance or employment.

Chris Chute, MD, DrPH writes, “The emphasis on characterizing patient information—including presenting conditions, findings, symptoms, working diagnoses, interventions, and outcomes—is manifest in a broad spectrum of health analyses. Clinical epidemiology, outcomes analysis, health services research, guideline development, continuous quality improvement [CQI], and health economics are among the traditions that rely fundamentally on a consistent representation of underlying patient data.” The body of work Dr Chute describes will lead to better and more rational delivery of medical care. When executed correctly, electronic delivery of medical data will add built-in decision support to our medical records and will enable them to be searchable, re-searchable, interpretable, transmissible, available, clear, and thus more useful. All these processes require standards for clinical data representation and transmission.

Conclusion

Our goal is for each patient to have an EHR which can be used across computer platforms. The combination of clear definitions and interrelations of medical terms (as in LOINC and SNOMED) used to populate an HL7 standardized “message” or document using standardized syntax (eg, XML) will allow medical information to be transmitted to and retrieved from any telecommunication system connected to the World Wide Web. In turn, this achievement could enable a clinician to retrieve any patient’s medical chart, laboratory and radiology reports, and other necessary information anywhere, anytime, given proper security—if, that is, we can all agree on and use these same standards. Information represented in this format will allow manipulation of data to facilitate advanced functions, including record searches, patient-specific guidelines, outcomes research, or other functions. Standardized, precise, logically interrelated and searchable terminology (ie, SNOMED and LOINC) which populates a standardized in-
Standards for Computerized Clinical Data: Current Efforts and Future Promise

The Medical Editing Department, Kaiser Foundation Research Institute, provided editorial assistance.

References


