health systems

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 in an abbreviated form. But what do abbreviations such as HL7, XML, LOINC, and SNOMED stand for? Why should we care what they mean? What could these acronyms do for—or, worse, to—us? This article briefly explains some of the most important medical informatics terms and concepts in the context of clinical practice (Table 1).

The EHR and Use of Medical Informatics Standards

We are rapidly progressing beyond handwritten medical information—and even beyond medical reports typed from dictation. Medical information such as medical records, laboratory results, and radiology reports is increasingly being generated and stored on computers—and this trend can be expected to continue. The Health Insurance Portability and Accountability Act of 1996 (HIPAA)[2] is federal legislation which requires formation and acceptance of standards for clinical terminology used in each EHR to impose order and uniformity in health information as well as to assure adequate security and confidentiality of this information. Since 1968, Lawrence Weed[3]—developer of the problem-oriented medical record—has taught us how to organize medical information logically. Standards for electronic records can be expected to incorporate logical systems such as these. Soon, the EHR will be simultaneously created and computer-

A new vocabulary of acronyms has been developed to represent medical informatics standards …

Jonathan Y Lukoff, MD, is a Partner Pediatrician with the Southern California Permanente Medical Group and member of the Medical Informatics-Convergent Medical Terminology Project. E-mail: Jonathan.Y.Lukoff@kp.org.
Robert H Dolin, MD, is a Partner Internist with the Southern California Permanente Medical Group, Physician Lead for the KP CIS Convergent Medical Terminology team, and the SCPMG Physician Consultant to Permanente Knowledge Connection.

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 1. Glossary of computer science terms relevant to medical informatics

Acronym	Complete term	Definition
ANSI	American National Standards Institute	
CDA	Clinical Document Architecture	A standardized representation of clinical documents (eg, reports of medical history and physical examination, Progress Notes)
EHR/EMR	Electronic Health Record/Electronic Medical Record	Computerized medical record; medical record stored in electronic form
HIPAA	Health Insurance Portability and Accountability Act	Federal regulations enacted in 1991
HL7	Health Level 7	An ANSI-accredited standards organization that develops methods for electronically transmitting medical data and information unambiguously
LOINC	Logical Observation Identifier Names and Codes	A standardized set of codes for representing laboratory result terms
Metastructure		A universally understood abstraction underlying an information management solution for creating and exchanging views of content
RIM	Reference Information Model	A Clinical Data Object Model produced by HL7
SNOMED	Systematized Nomenclature of Medicine	A standardized, logically interrelated medical vocabulary
Syntax		Rules governing construction of a machine language
W3C	World Wide Web Consortium	The standards organization that developed XML
XML	Extensible Markup Language	A standardized syntax used to tag information for transmission over the Internet

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.[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)[6] created XML (Extensible Markup Language),[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).[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.

Example of XML format

```
<PHYSICIAN>
   <FIRST-NAME>JON</FIRST-NAME>
   <LAST-NAME>LUKOFF</LAST-NAME>
</PHYSICIAN>
<PATIENT>
   <FIRST-NAME>BOB</FIRST-NAME>
   <LAST-NAME>DOLIN</LAST-NAME>
</PATIENT>
```

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)[11] and SNOMED (Systematized Nomenclature of Medicine)[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.[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.[11]

SNOMED is a reference medical terminology set developed more than 20 years ago and enhanced continuously ever since.[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.[13] LOINC has more complete defining characteristics for laboratory result data than SNOMED, but the two terminology sets are complementary.[12]

SNOMED defines codes for a wide spectrum of clinical concepts: Diseases and Findings; Procedures; Biological Functions; Body Structures; Living Organisms; Physical Agents, Activities, and Forces; Substances; Specimens; Occupations; Social Contexts; and Modifier Concepts.

Current Efforts to Further Standardize Clinical Vocabularies

The Convergent Medical Terminology (CMT)[a] Project began as a venture conducted jointly by the College of American Pathologists, the Kaiser Permanente Medical Care Program, the Mayo Clinic, and the National Library of Medicine.[14,15 b] This working group has revised SNOMED into the RT (Reference Terminology) version (released in November 2000) by using description logic, which allows us to interrelate terms parsed (divided) into their component parts (eg, "Pneumococcal Pneumonia" is both a "Pulmonary Disease" and an "Infectious Disease" and is caused by the organism *"Streptococcus pneumoniae"*).[16] Definitions, synonyms, and hierarchical relations are fully defined in SNOMED RT. Definitions from the International Classification of Diseases, 9th revision[17] are mapped to SNOMED, and LOINC concepts are incorporated into the laboratory procedure axis of SNOMED. This incorporation has permitted creation of a reference terminology useful for clinical medicine and will allow KP CIS to capture the richness of SNOMED, whereas mapping to ICD-9-CM enables semi-automated extraction of administrative billing codes. This semi-automated process should allow us to relieve our clinicians from the burden of coding their patient encounters.

… KP CIS will provide point-of-care information to clinicians along with a wealth of clinical data that promises to have great impact on our ability to enhance patient care.

In addition, the National Health Service [United Kingdom] READ Codes[18] 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."[19:389]

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."[20:9] 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, researchable, 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.[20] 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 information model and is transferred using standardized syntax (eg, XML) can be used as the basis of a universal EHR. As R Dolin stated, this standardization "go[es] a long way toward ensuring that what the sender thought was being sent (and said) equals what the receiver thought was being received …"[19:416] The HIPAA requires endorsement of some standards, and US Government requirements for recording and reporting encourage widespread acceptance of these standards. Along with other factors, these standards will give clinicians greater access to important patient information easily and seamlessly—even between different platforms—from any networked computer terminal. Other technical and nontechnical factors facilitating this increased clinical access include widespread deployment of secure data networks and interoperable clinical information systems (adopted through cooperation of vendors). Health information will be exchanged among health care delivery systems only after fully informed consent is given by patients and by their health care providers. In part through adherence to these standards, KP CIS will provide point-of-care information to clinicians along with a wealth of clinical data that promises to have great impact on our ability to enhance patient care. ❖

[a] The KP CIS team is now also referred to as the CMT team.

[b] The Mayo Clinic and the National Library of Medicine no longer participate in this project.

Acknowledgments

Yvonne G Lin, MS, contributed to developing the draft manuscript and made many useful suggestions. Louise Gordon, MD; Amy L Thach, RN, BSN, NP student; and My-Lan Le-Nguyen, MD, reviewed the manuscript.

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

References

1. Dolin RH. Outcome analysis: Considerations for an electronic health record. MD Computing 1997 Jan-Feb:14(1):50-4.
2. HIPAA Insurance Reform. Health Insurance Portability and Accountability Act. [Internet site by the United States Health Care Financing Administration]. Available on the World Wide Web (accessed Oct 18, 2000): www.hcfa.gov/medicaid/hipaa/default.asp.
3. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl.
4. Health Level Seven. Available on the World Wide Web (accessed May 5, 2000): www.hl7.org/.
5. Dolin RH, Mattison JE, Alschuler L, et al. The HL7 Clinical Document Architecture. J Am Med Inform Assoc 2001 Nov-Dec;8(6):552-69.
6. W3C, World Wide Web Consortium. About the World Wide Web Consortium. Available on the World Wide Web (accessed Oct 11, 2000): www.w3.org/Consortium/.
7. Bosak J, Bray T. XML and the second-generation Web: the combination of hypertext and a global Internet started a revolution: a new ingredient, XML, is poised to finish the job. Sci Am 1999 May;280(5):89-93.
8. Walsh N. A technical introduction to XML. Available on the World Wide Web (accessed Oct 18,2000): www.isgmlug.org/n4-2-49.htm.
9. Dolin R, Lukoff J, McKinley CS, Fuller LM, Minkler T, Biron P. Clinical practice guidelines made easy on the intranet: smoking cessation intervention for pediatrics and adults [abstract]. Pediatrics 1998 Sept;102(3 Suppl):691.
10. Lukoff JY, Dolin RH, McKinley CS, Fuller LM, Biron PV. Validation of an XML-based process to automatically Web-enable clinical practice guidelines: experience with the smoking cessation guideline. Proc AMIA Symp 1999:311-4.
11. Logical Observation Identifier Names and Codes. (LOINC) Available on the World Wide Web (accessed June 14, 2000): www.mcis.duke.edu/standards/termcode/loinclab.
12. Spackman K, Campbell K, Côté R. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
13. Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. J Am Med Inform Assoc 1998 Mar-Apr;5(2):203-13.
14. Dolin RH, Alschuler L, Biron PV, et al. Clinical practice guidelines on the Internet: a structured, scalable approach. MD Comput 1999 Mar/Apr;16(2):60-4.
15. CMT Project. Systematized Nomenclature of Medicine (SNOMED). Available on the World Wide Web (accessed June 14, 2000): www.snomed.org/Campbell/tsld013.
16. Dolin RH, Spackman K, Abilla A, et al. The SNOMED RT Procedure Model. Proc AMIA Symp 2001:139-43.
17. 1998 ICD-9-CM: International classification of diseases, 9th revision, clinical modification. Chicago: American Medical Association; 1997.
18. Harding A, Stuart-Buttle C. The development and role of the Read Codes. J AHIMA 1998 May;69(5):34-8.
19. Dolin RH. Advances in data exchange for the clinical laboratory. Clin Lab Med 1999 Jun:19(2);385-419, vii.
20. Chute CG. Moving toward international standards in primary care informatics. In: Moving toward international standards in primary care informatics: clinical vocabulary. Conference summary report of the US Department of Health and Human Services; 1995 Nov 1-2; New Orleans, LA. (AHCPR Pub. No. 96-0069, Oct 1996) p 9.
21. Lum AM, Zuiderveer MD. Clinician-patient communication: the electronic medical record: barrier or bridge to effective clinician-patient communication? Perm J 2002 Winter;5(1):52-4.