Call for a Standard Clinical Vocabulary

The medical informatics community—vendors and users—have been seeking a common, comprehensive clinical vocabulary for the past decade. New models for health care and new reporting requirements increase the necessity and desirability of such a vocabulary set. Much of the negotiation time between systems interested in exchanging data is related to matching vocabulary and code sets. Unfortunately, that is only the beginning and on-going work is required to keep the vocabulary sets synchronized. At the same time, increased emphases on decision support, clinical guidelines, integrated delivery systems, and data reporting for performance evaluation require more precise and understandable vocabulary terms.

Much work has been done on creating controlled vocabularies. There are more than 100 vocabulary sets defined internationally for some purpose. In this country, institutions are forced to deal with at least 5–10 of these sets to meet reporting or reimbursement requirements. The two vocabulary-related papers published in this issue of the journal[1,2] highlight problems in trying to use existing controlled-vocabulary sets or even combinations of sets in representing clinical data. Both papers make it clear that existing controlled-vocabularies are unable to meet the requirements for clinical data representation.

There are many reasons that existing controlled vocabularies do not meet the requirements for clinical data storage and data interchange. First, no existing set has been created for the purpose of representing clinical data. The purpose of most sets, in fact, is to address a higher level requirement which requires groupings or classifications of terms to meet design objectives. Many of the existing controlled vocabularies are proprietary, exist in a closed or controlled environment, and provide economic benefit to the developers. Motivations to expand sets to meet broader needs are not forth-coming. The Unified Medical Language System solved some problems in providing cross-linkage among the major classification systems, but it was never intended to provide a clinical vocabulary.

Fundamental to the problem of creating a clinical vocabulary is the understanding of what level of granularity is data to be stored in computer-based patient records and at what level is data to be interchanged between clinical units, third party payers, reporting

agencies, and others. Henry and Mead distinguish "raw" data and coded or classified data. If clinical content is to be preserved, it is likely that data must be stored at the atomic or "raw" data level—perhaps at the level of "clinical utterances." At this level of very fine granularity, we need to recognize that the vocabulary set will be precisely the words that are used in describing and recording clinical events— what we observe, what we think, and what we do. Importantly, however, is that the vocabulary is defined to be non-redundant and non-ambiguous. We need to identify preferred terms and include a precise definition. We need to recognize the value of synonyms and provide for the translation of those synonyms into preferred terms.

How do we create this clinically-rich formal terminology? The existence of so many controlled-vocabularies makes it difficult to start at the beginning. If no coded sets existed today, the solution to this problem would be more obvious. Circumstances force us to pursue an almost backwards solution to the problem. It is clear that we must take advantage of all the work that has been done to produce the many sets of controlled-vocabularies. Licensing and royalty agreements must be solved in order to incorporate existing sets into a broader vocabulary. In an integrated delivery system, a consultant may view data on a workstation which is coded in the sending system and translated dynamically for display. Do not use charges apply to both? I believe that the atomic formal terminology must be freely available for use, most likely available on a Web site.

Health Level Seven (HL7), an ANSI-accredited data interchange standards body, has recently established a Special Interest Group (SIG) for Clinical Vocabulary. Although the group is still understanding its charter and what it proposes to do, the purpose of the SIG is to define a vocabulary set for HL7 messages with which clinical data may be exchanged with full understanding. My interpretation of what the SIG proposes to do is that HL7 will define explicitly the vocabulary set which will be used in each data field for each specific data element. Not only will each data element vocabulary set be defined, but vocabulary sets for value or result, when appropriate, will be defined. Already a number of interesting observations have been made. The first is the obvious value of tying the clinical vocabulary to the data model being developed by HL7. Bi-directional liaison between the groups should produce better and more complete work by both groups. Tying the vocabulary set to use (use case) is important. For example, for an administrative use of the data element gender, the vocabulary set of male and female may be sufficient. For clinical

representation in an obstetric or pediatric setting, the same data element might require a vocabulary set of male, female, hermaphrodite, male dominant, hermaphrodite, female dominant; unknown; etc.

The HL7 master vocabulary set will include an internal HL7 reference number, which may be used in the data exchange, the HL7 preferred term, a definition, and a linkage to UMLS. Hierarchical linkages or classification links may be added as necessary to support decision support modules. HL7 will be able to avoid existing problems relating to the deletion of codes, in changes in the meaning of codes, and in the changes of codes to mean the same thing. Cross-mapping will be done in UMLS. The HL7 vocabulary set will provide a complete and stable set of codes which may be used in the storage and exchange of clinical data, and the translation of those codes into other coding systems through the cross linkages provided by UMLS.

The time has come to create a standard clinical vocabulary which provides atomic-level granularity ad-equate for storage and meaningful and unambiguous exchange of clinical data. The HL7 SIG-Vocabulary provides an opportunity for the medical informatics community to come together to solve one of the major barriers to the computer-based patient record.

W. EDWARD HAMMOND, PHD

References ■

1. Campbell JR, Carpenter P, Sneideman C et al. Phase II Evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. J Am Med Inform Assoc. 1997;4:238–251.
2. Henry SB, Mead CN. Nursing Classification Systems: necessary but not sufficient, for representing "what nurses do" for inclusion in computer-based patient record systems. J Am Med Inform Assoc. 1997;4:222–232.

■