

Clinical Terminologies: A Solution for Semantic Interoperability

Hyeoun-Ae Park¹, Nick Hardiker²

College of Nursing, Seoul National Univ.¹,
Univ. of Salford²

Abstract

To realize the benefits of electronic health records, electronic health record information needs to be shared seamlessly and meaningfully. Clinical terminology systems, one of the current semantic interoperability solutions, were reviewed in this article. Definition, types, brief history, and examples of clinical terminologies were introduced along with phases of clinical terminology use and issues on clinical terminology use in electronic health records. Other attempts to standardize the capture, representation and communication of clinical data were also discussed briefly with future needs. (*Journal of Korean Society of Medical Informatics 15-1, 1-11, 2009*)

Key words: Semantic Interoperability, Electronic Health Record, Clinical Terminology, Reference Terminology, Interface Terminology

Corresponding Author: Hyeoun-Ae Park, RN, Ph.D, Associate Dean, Professor, Biostatistics and Health Informatics, College of Nursing
Seoul National University 28 Yongon-dong Chongno-gu Seoul, 110-799, Korea
IMIA VP WG & SIGs
IMIA-NI VC Adm & Finances
Tel: +82-2-740-8827, **Fax:** +82-2-765-4103, **E-mail:** hapark@snu.ac.kr

Introduction

Electronic health records (EHRs) are being implemented worldwide with the promises that they will increase patient safety, reduce medical errors, improve efficiency and reduce medical costs. To realize these benefits, however, EHR systems must be adopted by a critical mass, and different systems must be interoperable with one another¹⁾. The problem is that solutions that foster adoption may decrease interoperability and vice-versa.

A key element in this conflict is the way terminology is used in healthcare information systems. An obstacle to the widespread adoption of EHR systems is the difficulty associated with capturing structured clinical information from health care providers who prefer to document clinical findings, procedures, and outcomes using “free text” natural language. However, standardized terminology and capturing structured clinical data are a prerequisite for interoperability and sharing and exchanging healthcare information, decision support and health service research. Unfortunately, standardized terminologies are unfamiliar and unnatural to most health care providers, who struggle with application programs using terminologies and become frustrated. As a result, user adoption falls.

Current attempts to standardize the capture, representation and communication of clinical data rely upon three layers of artifacts to represent the meaning of clinical data. Current semantic interoperability solutions include generic reference models for representing clinical data such as HL7 Clinical Document Architecture (CDA)²⁾ and the EHR Reference Model³⁾; agreed clinical data structure definitions such as *openEHR* archetypes⁴⁾, HL7 templates⁵⁾; and clinical terminology systems such as Logical Observation Identifiers Names and Codes (LOINC)⁶⁾ and Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT)⁷⁾.

This paper limits its scope to clinical terminologies and is prepared with the basic assumption that electronic health data standards - including a standard clinical vocabulary - are a part of the information infrastructure needed for health care, clinical and health services research, and public health. There are many clinical terminologies available; however, each terminology differs in scope, granularity

levels and inherent quality. Thus, true interoperability is almost impossible if all these different terminologies are used in one system, since the multiple combinations of sending and receiving EHR data would require the creation and maintenance of dozens of mapping solutions.

In this article, we would like to review terminologies about clinical terminologies, brief history of clinical terminologies, examples of clinical terminologies, and the use of clinical terminologies. We also would like to briefly cover future directions for ensuring full semantic interoperability.

Clinical terminologies

Clinical terminology is defined as “standardized terms and their synonyms which record patient findings, circumstances, events, and interventions with sufficient detail to support clinical care, decision support, outcomes research, and quality improvement; and can be efficiently mapped to broader classifications for administrative, regulatory, oversight and fiscal requirements”⁸⁾. These have provided nomenclature, controlled vocabularies and hierarchical classifications of diseases, etiologies, and treatments to facilitate the entry and analysis of healthcare data. Examples of clinical terminologies include International Classification of Diseases (ICD)⁹⁾, Current Procedural Terminology (CPT)¹⁰⁾, and SNOMED CT⁷⁾.

The knowledge of clinical concepts and the way in which these interrelate has traditionally been implemented in healthcare through clinical coding schemes. The coding schemes associated with these terminologies were primarily required to assist with the subsequent analysis and aggregation of the data across clinical systems and health care facilities, furthermore across different countries. A clinical terminology primarily serves to provide a systematized and controlled vocabulary of clinically relevant phrases that can be used during data entry to provide a more precise and sharable expression than might be obtained by using free text. Being controlled expressions, the translation of a terminology to another natural language is moderately scalable, permitting EHR data to be interpreted across languages. Different fine-grained terms can be cross-mapped to a coarser one in order to permit

them to be processed homogeneously, if this is sufficiently precise for the particular purpose. Unfortunately, some terminologies have historically failed to distinguish the requirements of a classification system from that of a clinical vocabulary¹¹⁾.

History of clinical terminologies

Even though modern terminology began with the modest classification developed for the London Bills of Mortality in the latter part of the 16th century, most terminologists credit the birth of the Standardized Nomenclature of Disease (SND) as the beginning of a modern era for clinical description⁸⁾. SND introduced the notion of multi-axial coding with the two axes, topology (anatomy) and etiology (pathophysiology). With this terminology, users could model complex concepts by mixing the two codes together. For example, “Chest pain” can be composed from an anatomy axis concept for “chest” and a pathophysiology axis concept for “pain”.

The number and size of available clinical terminologies have expanded over time as the result of three phenomena: the exponential growth of medical knowledge in the second half of the 20th century; the availability of computer systems that are able to handle the complexities of knowledge representation, storage, retrieval, and maintenance; and the expansion of intended uses for clinical terminologies beyond diagnostic classification¹²⁾. Representative terminologies such as ICD for statistical reporting and billing, CPT for billing, and Medical Subject Headings (MeSH)¹³⁾ for indexing, cataloging, and searching biomedical literature have been operational as *de facto* standards in various revisions for decades.

In the 1990s, excellent clinical vocabularies emerged, but none became *de facto* standards. By contrast, individual terminologies have been identified by standards organizations as candidates for specific uses. From early 2000, there have been efforts to bring together multiple terminologies. For example, two leading candidates - SNOMED and NHS Clinical Terms - merged and became SNOMED CT. In addition, the US national Library of Medicine has assembled the Unified Medical Language System (UMLS)

as a metathesaurus to encompass multiple terminologies and vocabularies. Currently, the UMLS contains over one million biomedical concepts and five million synonyms from more than 100 terminologies.

Types of Clinical Terminologies

A terminology consists of a collection of words or phrases, called terms, aggregated in a systematic fashion to represent the conceptual information that makes up a given knowledge domain such as clinical pathology or nursing¹²⁾. Terms in a terminology generally correspond to actual events or entities and to their cognitive representations in people's minds called concepts¹⁴⁾. For example, “myocardial infarction” or “heart attack” in a given terminology might represent the event of an “ischemic injury and necrosis of heart muscle cells resulting from absent or diminished blood flow in a coronary artery.” A health care provider evaluating a patient with prolonged chest pain may consider the diagnosis of an ischemic heart muscle injury and could use the terms “myocardial infarction” or “heart attack” to consider and to communicate such a possibility. Terminologies also typically contain hierarchical organizations and other representations of linkages among concepts, such as the “is-a-type-of” relationship between “myocardial infarction” and “heart disease.”

Terminologies can be further broken down based on their characteristics and purpose⁸⁾ (Table 1). Administrative terminology is used primarily for the classification of information and the administration of healthcare delivery or reimbursement. Two examples of administrative terminologies are International Classification of Disease (ICD) and the Current Procedural Terminology (CPT). Neither is well-suited for capturing clinical data for EHRs due to their lack of granularity, but both have been widely adopted in non-clinical settings.

A reference terminology is a concept-based, controlled medical terminology which allows for the complex organization and aggregation of clinical information. Examples include SNOMED CT and International Classification for Nursing Practice (ICNP) Version 1.1.¹⁵⁾

Interface terminology, sometimes called colloquial or

entry terminology, is used to describe lists of terms and phrases which are a systematic collection of health care-related phrases (terms) that support clinicians' entry of patient-related information into computer, such as narrative nursing notes or decision support tools. An example of interface terminology is the NANDA Taxonomy II.

The phases of terminology use are widely regarded as entry terms, then reference terminologies, and finally aggregate or administrative classifications⁸⁾. Entry terms are colloquial expressions or terms that are familiar to health care providers or users and convey sufficient specificity to say what is meant. These are translated into an underlying reference terminology, which is capable of semantic closure and unambiguous representation. Finally, the formal reference terms can be aggregated using explicit inclusion, exclusion, and cross-referencing rules into a high-level classification such as ICD for reporting and billing purposes (Fig. 1).

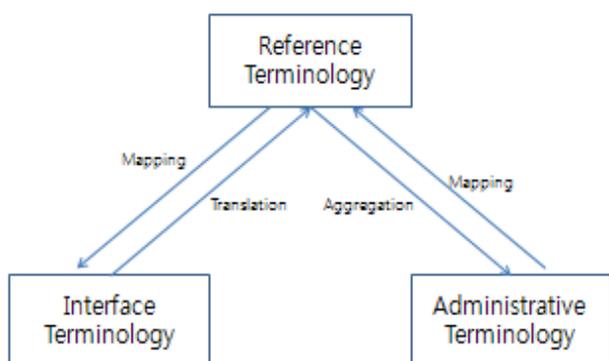


Figure 1. Phases of Terminology Use

Examples of Clinical Terminologies

International Classification of Disease

The International Classification of Disease⁹⁾ is a core classification of the WHO Family of International Classification (WHO-FIC) and the most widely used classification system worldwide. It is being used for morbidity and mortality statistics, and reimbursement and decision support in health care. Use of the ICD promotes international comparability in collecting, processing, classifying and presenting these statistics. The core classification of ICD-10, the latest edition adopted by the WHO in 1990 and distributed worldwide in 1994, is a three-character code, which is the mandatory level of coding for international reporting. The four- and five-digit sublevels are recommended for other purposes, such as a further subdivision or a classification of other aspects. For example, “Acute myocardial infarction” has the code I21 and “Acute transmural myocardial infarction of anterior wall” has the code I21.0. Each health condition can be assigned to a unique category and given a code. Such categories can include a set of similar diseases.

ICD-11 is planned for release in 2015, and will be built using an ontological approach to enable standardized information processing by computers in an e-health application. The revision process for ICD-11 will make use of distributed web-based tools to collate suggestions, discussions and evidence¹⁶⁾.

Current Procedural Terminology

The Current Procedural Terminology (CPT)¹⁰⁾ is a

Table 1. Types of Clinical Terminologies

Types	Definition	Examples
Reference Terminology	Set of canonical concepts, their structure, relationships, and - if present - their systematic and formal definitions. It is intended for both human and computer.	ICNP Version 1.1 SNOMED-CT
Interface terminology, (Entry Terminology, Colloquial Terminology, Presentation Terminology)	Set of designations that are mainly intended for human use, and map to concepts in a terminological system	NANDA, NIC
Administrative Classification (Statistical Classification)	A set of mutually exclusive categories to aggregate data at a pre-prescribed level of granularity for a specific purpose	ICD 10, CPT

coding system used mainly in the United States for billing and reimbursement. It provides a coding scheme for medical, surgical and diagnostic services and procedures to communicate uniform information about them among physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytical purposes. The CPT code set is maintained by the American Medical Association through the CPT Editorial Panel. The current version of CPT is the CPT 2009. There are three types of CPT codes: Category I CPT Code(s) - Medical services and Procedures; Category II CPT Code(s) - Performance Measurement; and Category III CPT Code(s) - Emerging Technology.

Systematized Nomenclature of Medicine -Clinical Terminology

SNOMED CT⁷⁾ is a systematically organized computer-processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, and etc.. Components of the SNOMED CT are shown in Table 2.

SNOMED CT is a multidisciplinary terminology system that has been recognized as a standard terminology by many countries. This terminology system is based on clinical concepts, and each concept is based on a specific definition. The defining relationships among the concepts create the SNOMED CT hierarchical structure. The July 2008 release of SNOMED CT contains more than 311,000 active clinical concepts and more than 800,000 descriptions to express clinical concepts, including both the preferred name and related synonyms. There also are more than 1,360,000

defining relationships supporting the meaning of the concepts within the SNOMED CT hierarchies. These defining relationships enhance data retrieval and can be used to support research.

SNOMED CT concepts are organized in 19 hierarchies with multiple levels of granularity. This parent-child relationship is known as an “Is-A” relationship, where one class is a subclass of another class. In SNOMED CT, an Is-A relationship exists when a parent concept has a broader meaning than its child concept. For example, the procedure “Verification of allergy status (procedure)” is a child of and therefore has an Is-A relationship with the procedure “Verification routine (procedure).” Similar to a tree, SNOMED CT continues to branch into finer levels of granularity where the lowest level concepts in the hierarchy contain the most granular meanings.

SNOMED CT concepts in the two different hierarchies are linked with attributes. In the example below, “Lumbar discitis (disorder)”, a concept in the Clinical finding hierarchy, is related to a concept in the Body structure hierarchy through two attributes: Finding Site and Associated Morphology.

<i>Lumbar discitis (disorder)</i>
FINDING SITE <i>Structure of lumbar intervertebral disc (body structure)</i>
ASSOCIATED MORPHOLOGY <i>Inflammation (morphologic abnormality)</i>

The two attributes FINDING SITE and ASSOCIATED MORPHOLOGY and their assigned values provide definition for the concept “Lumbar discitis (disorder)”.

Table 2. Components of the SNOMED CT

Components	Definition	Examples	Number
Concepts	Basic unit of meaning designated by a unique numeric code, unique name (fully specified name), and descriptions, including preferred term and one or more synonyms	Gastric ulcer (finding)	311,000 active concepts
Descriptions	Terms or names (synonyms) assigned to a concept	Gastric ulcer (finding) Stomach ulcer (finding) Gastric ulceration (finding)	800,000 descriptions
Relationships	Link concepts either within a hierarchy or across hierarchies. These relationships provide formal definitions and other characteristics of the concept.	Gastric ulcer (finding) IS_A Disease of Stomach (finding) Gastric ulcer (finding) IS_A Gastrointestinal ulcer (finding) Gastric ulcer (finding) has_findingsite Stomach (body structure)	1,360,000 links or semantic relationships

SNOMED CT used description logics for its representation. In practice SNOMED CT concepts can be used as building blocks for describing other SNOMED CT concepts. For some SNOMED CT concepts, the set of relations to other concepts provided is necessary and sufficient to fully define the concept. Other concepts, called primitives, are incomplete definitions, sometimes limited to one subclass (IS-A) relation. SNOMED CT has the extra property of term co-ordination. This means that basic terms can be combined to compose more complex expressions: for example, a “headache” can be stated to be located in the “frontal region” of the head and to be “left sided” and “severe”, all in a single terminological expression. It theoretically also permits terms to be members of more than one hierarchy, which could allow terms to be grouped by use case as well as by pathological process.

Logical Observation Identifiers Names and Codes

LOINC is a set of universal names and codes for identifying laboratory tests and clinical observations⁶⁾. The laboratory portion of the LOINC database contains the usual categories of chemistry, hematology, serology, microbiology (including parasitology and virology), and toxicology, as well as categories for drugs and cell counts, antibiotic susceptibilities, and more. The clinical portion of the LOINC database includes entries for vital signs, hemodynamics, intake/output, EKG, obstetric ultrasound, cardiac echo, urologic imaging, gastroendoscopic procedures, pulmonary ventilator management, selected survey instruments

(e.g. Glasgow Coma Score, PHQ-9 depression scale, CMS-required patient assessment instruments), and other clinical observations.

Each LOINC record has two main types of entities; laboratory tests and clinical observations on the one hand, and the entities necessary for their description (sometimes referred to as “parts”) on the other. In fact, LOINC “parts” concepts serve as building blocks for the description of tests and observations, in association with a set of semantic relations. For example, “Sodium: SCnc:Pt:Ser/Plas:Qn”, the laboratory test in which the molar concentration of sodium is measured in the plasma (or serum) is identified by 2951-2. Major parts of a test or observation name are shown in Table 3. For example, the “parts” concept “Sodium” is linked to this test by the relationship “Component”. Each part of the name can be further divided into subparts.

The fully specified name of a test or clinical observation is described in reference to the analyte measured (component), the property under investigation, the time aspect, the origin of the sample (system) and the type of scale used. Additionally, the method used is reported when appropriate. The LOINC terminology does not use any particular formalism, such as description logics. However, the formal definitions provided by LOINC all conform to the 6-axis template presented in the example above and make use of named semantic relations, which makes them amenable to automatic processing. In addition to simple tests, LOINC also defines complex concepts, including panels (i.e., collections of tests) and concepts involving a challenge (e.g., glucose measurement, 90 minutes after oral administration of 75g of glucose). In LOINC Version 2.2,

Table 3. Major LOINC Parts of a Test/Observation Name

Six Parts	Definition	Examples
Component (Analyte)	Name of analyte or measurement	Potassium, Hemoglobin, Hepatitis C Antigen
Kind of property (Kind of quantity)	Distinguishes between different kinds of quantities relating to the same substance	mass concentration or substance concentration of sodium in the urine sample
Time Aspect	Whether the measurement is an observation at a moment of time or an extended duration of time	A moment of time, An extended duration of time
System (Sample) type	System or sample used for measurement	Urine, blood
Type of Scale	Whether the measurement is quantitative, ordinal, nominal or narrative	A true measurement, A ranked set of options, E.Coli, Dictation result from x-rays
Type of Method	Method used to produce the result, or other observations	

the total number of tests and observations is 50,809, of which 37,767 correspond to laboratory tests and 34,767 to simple laboratory tests. The total number of “parts” concepts is 44,314, of which 13,794 are used as values for the 6 main axes. All LOINC concepts are in the UMLS. While the “part” concepts are generally well integrated with equivalent concepts from other terminologies, concepts for laboratory tests and clinical observations are not, due to the peculiarity of their names.

International Classification of Nursing Practice Version 1.1

International Classification of Nursing Practice (ICNP)¹⁵ is a logic-based compositional terminology developed by the International Council of Nurses to serve as a unified nursing language system, to articulate nursing's contribution to health and health care globally, and to promote harmonization among existing nursing standards.

ICNP Version 1.1 represents nursing phenomena and actions and is designed to be comprehensible to both computers and humans. ICNP Version 1.1 is maintained in the Web Ontology Language (OWL). One public “view” of the OWL ontology is the so-called 7-axis model. In this representation, thousands of terms and definitions are arranged into one of seven axes: focus, judgment, means, action, time, location, and client (ICNP, 2009).

A primary aim of the ICNP is “to provide nurses with a unified nursing language system to represent and document what nurses do in a variety of nursing settings.” The ICNP supports documentation across all phases of the nursing process. Thus, the ICNP provides a means to capture atomic level data that can be employed to support the application of evidence to practice and the building (or extraction) of evidence and nursing knowledge from practice.

The new content of ICNP Version 1.1 includes ICNP® catalogue pre-coordinated statements for nursing diagnoses, interventions and outcomes. ICNP® catalogues are subsets of the terminology that facilitate software developers and nurses working in a specialty area (e.g., community nursing) or a focus area of nursing (e.g. promoting adherence to treatment), to more readily integrate ICNP® into practice.

Nursing statements such as nursing diagnoses, interventions and outcomes can be composed by combining terms from various axes of ICNP Version 1.1, based on the International Organization for Standardization (ISO) standard 18104: 2003, *Integration of a Reference Terminology for Nursing*¹⁷. For example, a nursing diagnosis “Medication Non-adherence Risk” can be composed with “Medication Non-adherence” from the Focus axis and “Risk” from the Judgment axis, a nursing intervention “Monitoring medication adherence with pill box” can be composed with “Monitoring” from the Action axis, “Medication adherence” from the Focus axis and “Pill box” from the Means axis, and a nursing outcome “Medication adherence” can be composed with “Medication adherence” from the Focus axis and “Actual” from the Judgment¹⁸.

NANDA Taxonomy II

NANDA is the most widely used classification of patient phenomena in the nursing domain, developed and maintained by NANDA International. NANDA taxonomy II has a hierarchical structure of 12 domains, 47 classes and 188 nursing diagnoses¹⁹. Taxonomy II has a code structure with 5-digit codes. A new code is assigned to the nursing diagnosis when it is approved by the NANDA Board of Directors. Taxonomy II is multiaxial in its form, thereby substantially improving the flexibility of the nomenclature and allowing for easy additions and modifications. There are seven axes: the diagnostic concept; time (acute, chronic, intermittent, continuous); subject of diagnosis (individual, family, group, community); age (fetus to old-old adult); health status (wellness, risk, actual); descriptor (limits or specifies the meaning of the diagnosis concept); and topology (parts/regions of the body and related functions).

The axes are represented in the named/coded nursing diagnosis through their values. In some cases they are named explicitly, e.g., “Ineffective community coping” in which the subject of the diagnosis, “community”, is named. “Ineffective” is from the descriptor axis. In other cases the axis is implicit, e.g., “Activity intolerance”, in which the individual is the subject of the diagnosis.

NANDA diagnoses have been modeled into SNOMED CT and integrated into the National Library of Medicine's Unified Medical Language System.

Discussion

Scrutiny of the abovementioned classifications and nomenclatures of diseases and even procedures will show that one is a refinement of the other. If a logical classification of a disease carries codes that are so arranged that each entity included in a particular class has its own code, it becomes a nomenclature. On the other hand, if a very detailedly coded nomenclature has all its entries arranged in affinity groups or classes, then at this higher level it may be called a classification. It would appear that nomenclature and classification are at opposite ends of the same spectrum.

For quite a few people, a nomenclature such as SNOMED was viewed as the competitor to a classification such as ICD, but a nomenclature does not compete with a classification, it enhances it by furnishing it with greater detail. Although this fact was not comprehended by many statisticians until recently, it was clear to a group of medical specialists on medical language. It was believed that the generality of a classification and the specificity of a nomenclature could be combined within a single system. Thus, it is very important to recognize that nomenclatures might complement and not compete with classifications²⁰⁾. This recognition will resolve what has been a very longstanding controversy. Many researchers have derided one modality or another, failing to recognize that each serves its purpose. It is self-evident that well-defined nomenclatures can be “rolled up” into aggregating classifications, although the rules and logic about how exactly to undertake this are not always obvious or explicit.

The adoption of nomenclature such as SNOMED CT is being recommended for the EHR because it captures clinical information at the level of detail needed by clinicians for the provision of care (the reference terminology level) in most healthcare disciplines and most healthcare settings. Because SNOMED CT was developed to support the recording and retrieval of electronic clinical

information, its use does not replace the need for classification systems. Terminologies at different levels of granularity complement one another and can be aggregated from the detailed level to the more generic level. An example of an aggregation is from SNOMED CT to International Classification of Diseases initiated by a task force in the AHIMA²¹⁾.

In clinical information systems, terminologies used in patient records need to be interoperable with terminologies used in subsystems, such as laboratory systems. The most tangible example is the adoption by reference within SNOMED of codes from LOINC for laboratory tests and clinical findings²²⁾. Similarly, the National Library of Medicine has organized and facilitated the formal reference of SNOMED CT in CPT construction as well as the mapping of SNOMED content into ICD classifications.

Terminology integration systems, such as the Unified Medical Language System (UMLS) play an important role in creating *post hoc* mapping between these terminologies and contribute to the interoperability of systems relying on these terminologies. However, a key element to identifying equivalent concepts across terminologies in the UMLS is the lexical resemblance among concept names. As a consequence, concepts whose names are not amenable to national language processing, such as the names of laboratory tests in LOINC, generally cannot be mapped to equivalent concepts in other terminologies. However, SNOMED CT and LOINC provide formal definitions for their concepts in the form of a rich set of relations to other concepts. Comparing such sets of relations also provides the basis for comparing these concepts, provided there are enough shared relations between the two terminologies. Other terminologies with the same specificity and similar concepts may be able to be mapped to each other. Examples of this type of mapping include between ICNP and other nursing classifications²³⁾.

Convergence and openness have also influenced terminology development. For example, NANDA's decision to create NANDA Taxonomy II with a multiaxial system¹⁹⁾, and the WHO's decision to develop ICD-11 based on an ontology system to link ICD with terminologies¹⁶⁾. This linkage in the ICD-11 will require an ontological

definition of the entity including its taxonomic status (i.e. in what chapter, section in the classification tree, whether it is a disease, disorder, injury, syndrome, sign, symptom, other; its possible level of use such as in Primary care, Clinical care, Research; and other characteristics such as episodicity, severity, chronicity and etc.). An Expert Drafting Group will be given the mandate of identifying core constructs and concepts of ICD-11 using terminology/ontology tools to formalize the concepts and constructs using SNOMED and/or any other terminology.

Conclusion

Even though clinical terminology systems are used to promote the semantic interoperability of the EHR, there are a number of problems with clinical terminologies.

First, modern clinical terminologies attempt to provide a comprehensive coverage of health care to classify all terms, even though many terms cannot be defined, or agreed upon. If many of the terms in a terminology are not used consistently, it is useless to have them there. Terminologies also often include multiple representations for the same clinical concept. This also adds challenge to semantic interoperability.

Second, with concept-based terminologies such as SNOMED CT and ICNP Version 1.1, perhaps the area with the highest risk is post-coordination. This is a mechanism to permit users to combine terms in an *ad hoc* way during data entry to document a particular situation in a given patient. This seems at first to be appealing because it provides complexity and diversity of health care with a modest number of core terms plus a coordination mechanism. The drawback is that users are relatively free to combine terms in any way they choose, and we are not sure how much more systematic this will prove to be than writing free text. Post-coordination means EHR systems receiving data from another system have to cater for the unexpected. If more than hundreds of thousands of terms can be combined in many different ways, we cannot ensure the semantic equivalence of data from other systems.

Finally, a new generation terminology system trying to replace free text may have lost track of why we needed a

clinical terminology in the first place. We need to remind ourselves that our primary need is to systematize the vocabulary for the clinical data items that computers are able to read and process.

Future needs

SemanticHELAH's project on semantic interoperability suggests that clinical terminology systems such as SNOMED CT and LOINC are not sufficient to ensure full semantic interoperability across heterogeneous EHRs in order to gain the benefits of decision support, workflow management and evidence-based health care²⁴. Additional solutions recommended include EHR reference models for representing clinical data and agreed clinical data structure.

EHR reference models

The reference models define the high-level logical model for any kind of EHR and the information properties that will be common to all of the entries contained in it. They are dates and times of when observations occurred, health events took place and when information was recorded; the person who provided, composed, entered or authorized particular entries, or played particular roles in a health care process; version management information, including who changed any of the entries, when and why; who the information is about, if not the patient; the degree of sensitivity of the information and who should be allowed to access it; the ability to label each point in the record hierarchy i.e., to include a name for each folder, document, heading and the parts of each detailed entry; a standard way of representing coded clinical terms, measured quantities, dates, times and various kinds of multimedia data. Provided that the reference model to be is known by both sending and receiving information systems, any health record extract exchanged between them will contain all of the structure, names and medico-legal information required for it to be presented faithfully on receipt even if the nature of clinical content has not been agreed in advance. The two global candidates in selecting a reference model for EHR interoperability are ISO/EN 13606 part 1³, and HL7 CDA².

In addition, terminology content needs to fit within a framework of other standards and manage their implications for mutual update and harmonization, including overarching or reference information models. International Standardization efforts by CEN and ISO related to Electronic Health Records and semantic interoperability have resulted in a number of categorial structures which are a step towards supporting healthcare terminologies with an ontology that in turn will support multipurpose uses and safe communication¹⁷⁾²⁵⁻²⁶⁾.

A categorial structure supports interoperability by providing common frameworks with which to develop terminologies that are able to be related to each other and to analyze the properties of different terminologies to establish the relationship between them,

Clinical Data Structure

The sharing of health records and their meaningful analysis across distributed sites also requires that consistent clinical data structures are used through the reference model, so that equivalent clinical information is represented consistently, or at least can be mapped to consistent representation for interpretation. Clinical archetypes provide a systematic approach to representing the definition of any EHR data structure. The archetype approach is itself generic, and archetypes can represent data structures for any profession, specialty or service. The formalism takes into account that the clinical data structures and value sets required by a different care domain will be diverse, complex and will change frequently as clinical practices and medical knowledge advance. An archetype instance is a knowledge artifact that defines how the EHR reference model hierarchy should be organized to represent the data for one clinical entry or care scenario. Because these archetype definitions are represented in a standardized form, they can be shared and used across record-sharing communities to define how locally organized clinical data should be mapped consistently. The kinds of meaning that are represented using archetypes are a clinical headings framework, fine grained clinical data structures, and relevant data value sets or terminology constrains and a

specification optionality and multiplicity. ISO/EN 13606 part 2²⁷⁾ and HL7 Templates⁵⁾ are examples of clinical data structures. Currently the Korean Center for Interoperable EHR is developing archetypes called Clinical Content Model based on years of research in Korea²⁸⁾.

Over the past few years it has become clear that the greatest areas of challenge in full semantic interoperability for the EHR lie first in the defining of the data structure (e.g. clinical archetypes) and secondly in the binding of data structure nodes (archetype nodes) to terminology. By binding the nodes of an archetype to a part of a terminology system (for example, to specify that the value for a node called "location of fracture" must be a term from a hierarchy of bones in the skeletal system in terminology), it should promote consistency and reliability in how EHR data are represented, exchanged and interpreted. However, it is not easy to bind nodes of an archetype to a constrained set of terms, since data structure and terminology systems have been developed in relative isolation, with very little or no cooperation on their mutual requirements or scope, resulting in overlapping coverage and a clumsy fit. Thus, it is important for data structure and terminology researchers to work closely to achieve full semantic interoperability.

REFERENCE

1. Kanter AS, Wang AY, Masarie FE, Naeymi-Rad F, Safran C, Interface Terminologies: Bridging the Gap between Theory and Reality for Africa, Proceedings of MIE2008, Goteburg, Sweden.
2. Dolin R et al, HL7 Clinical Document Architecture Release 2.0. Health Level 7; May 2005.
3. Kalra, D, Lloyd, D. EN 13606 Health informatics - Electronic health record communication - Part 1: Reference model. Technical report. European Committee for Standardisation; 2007.
4. Chen R, Beale T. Towards archetype-based semantic interoperability of the EHR, Proceedings of MIE2008, Goteburg, Sweden.
5. Grieve G, Hamm R, Shafarman M, Mulrooney G. HL7 Template Specification. Health Level 7; 2007.
6. Logical Observation Identifiers Names and Codes (LOINC). Available at: <http://loinc.org/>. Accessee March 20, 2009.

7. SNOEMD Clinical Terms. Available at: <http://www.ihtsdo.org/snomed-ct/> Accessed March 20, 2009.
8. Chute CG. Clinical classification and terminology: Some history and current observations. *J of Am Med Inform Assoc.* 2000;7:298-303.
9. International Classification of Diseases, Available at: <http://www.who.int/classifications/icd/en/>. Accessed on March 20, 2009.
10. Current Procedural Terminology. Available at: <http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.shtml>. Accessed on March 20, 2009.
11. Ingram D, Lloyd D, Baille O, Grubb P. The GEHR. Requirements for Communication Capacity. The Good European Health Record Project: Deliverable 9. Brussels: European Commission; 1992.
12. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc.* 2006 May-Jun; 13(3):277-88.
13. MeSH, Available at: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>. Accessed March 20, 2009.
14. Campbell KE, Oliver DE, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. *J Med Inform Assoc.* 1988;5:421-31.
15. International classification for nursing practice. Available at: <http://www.icn.ch/icnp.htm>, Accessed March 20, 2009.
16. Production of ICD-11: The overall revision process (2007). Available at: <http://www.who.int/classifications/icd/ICDRevision.pdf>. Accessed March 20, 2009.
17. ISO 18104: 2003 Health Informatics - Integration of a reference terminology model for nursing. Geneva: ISO; 2003.
18. Guidelines for ICNP catalogue development (2008). Available at: http://www.icn.ch/icnp_Catalogue_Development.pdf, Accessed March 20, 2009.
19. NANDA International. Nursing Diagnoses: Definitions and classification 2007-2008. Philadelphia: Author; 2007.
20. Cote RA. Editorial: Ending the classification Versus Nomenclature Controversy. *Informatics for Health and Social Care*, 1983;8:1, 1-4.
21. Brouch K. AHIMA Project Offers Insights into SNOMED, ICD-9-CM Mapping Process. *Journal of AHIMA.* 2003;74(7):52-55.
22. Bodenreider O. Issues in mapping LOINC laboratory tests to SNOMED CT, *AMIA Annu Symp Proc* 2008; 51-55.
23. Hyun S, Park H.A. Cross-mapping the ICNP with NANDA, HHCC, Omaha System and NIC for unified nursing language system development, *International Nursing Review*, 2002;49(2):99-110(12).
24. Kalra D. Barriers, approaches and research priorities for semantic interoperability in support of clinical care delivery – EHR, *SemanticHEALTH. D4.1, EC-supported FP6 SA*, December 2007.
25. EN 1828: 2002 Health Informatics Categorical Structure for Terminologies of Surgical Procedures. CEN, 2002.
26. ISO/TC 215 WG 3 - Health informatics - Conceptual framework for patient findings and problems in terminologies. ISO. 2009.
27. ISO 13606-2:2008. Health informatics - Electronic health record communication - Part 2: Archetype, Geneva: ISO; 2008.
28. Center for Interoperable EHR, Available at: <http://ehrkorea.org/eng/>. Accessed March 20, 2009.