SNOMED: A Controlled Vocabulary for Computer-based Patient Records

by Karen M. Kudla, MBA, MT (ASCP), and Marjorie C. Rallins, DPM, MT

SNOMED is a detailed and specific coded vocabulary of names and descriptions used in healthcare -- explicitly designed for use in the computerized patient record. Here's an overview of this system and a look at its next generation.

Methods to collect data about human morbidity and mortality, and healthcare practices in general, can be traced as far back as the early 18th century. During the past 300 years, they have expanded to a profusion of diverse and highly specialized systems, each built to fill a particular need. The vast majority of these systems are classifications -- systematic groupings into predetermined categories. The ICD-9-CM classification of diseases and the CPT-4 classification of medical procedures are the most common examples of classifications in widespread use throughout medicine in the United States.

SNOMED ® is the Systematized Nomenclature of Human and Veterinary Medicine.

The purpose of SNOMED is to index, store, and retrieve information about a patient in a computerized medical record. It was built on the premise that a detailed and specific nomenclature is essential in order to accurately reflect the complexity and diversity of information found in a medical record. SNOMED is not a classification. At its most basic level, it is a coded vocabulary of names and descriptions used in healthcare.

The granular, detailed concepts codified in SNOMED can be easily combined into the broader classifications used for statistical and administrative purposes, but it can do much more. The compositional nature of SNOMED and its hierarchical systematized structure as a multiaxial controlled terminology transforms the notion of a coded vocabulary into a powerful tool for outcomes analysis, cost-effectiveness studies, knowledge-based practice guidelines and support systems, comparison of healthcare plans, and many other applications.

SNOP: The Foundation for SNOMED

SNOMED is owned and copyrighted by the College of American Pathologists (CAP), a medical specialty organization of board-certified pathologists. In 1965, the CAP's Committee on Nomenclature and Classification of Disease published SNOP -- the Systematized Nomenclature of Pathology. Its purpose was to codify the information in a surgical pathology report so that the medical case could be easily retrieved at a later date.

While SNOP could have been developed as a classification, it was determined from the outset that the pathologist should be given the flexibility to codify the pathology findings as accurately and in as much or little detail as he or she desired, without being limited to a predetermined classification. It was built on the premise that a detailed and specific nomenclature is essential in order to accurately reflect the complexity and diversity of information found in a pathology report. SNOP was a tool to help pathologists organize their records. While there was a suggested methodology on how to code a report in SNOP, the actual implementation was left to the individual pathologist or pathology group. There was never any requirement to use SNOP; however, its utility was widely apparent, and it was immediately adopted by the vast majority of pathologists. Pathology reports are typically free-text descriptions of the gross and microscopic findings in, for example, a tissue biopsy, a surgically removed organ, an autopsy, etc. SNOP encoded reports enabled pathologists to organize their pathology reports based on key findings so that specific cases could be efficiently retrieved for use in conferences, research projects, quality assurance, etc. For example, to find all cases of fine needle aspirates from the thyroid, the pathologist would search a database for all cases containing the SNOP codes for "thyroid" and "fine needle aspirate."

The Creation of SNOMED

Ten years of successful and widespread use in pathology led to the expansion of SNOP into other specialties. Furthermore, the growing prevalence of computers in the mid-1970s provided a notable stimulus to build upon the foundation of SNOP and develop a nomenclature for the computerized patient record. In 1977, the CAP published SNOMED, a systematized multiaxial and hierarchically organized nomenclature of medically useful terms. The original SNOMED and its successor, SNOMED II (published in 1979), consisted of six nomenclature axes or categories representing various aspects of the human being from a pathophysiologic point of view:

- *Topography (anatomy)*: the parts and regions of the body -- all tissues, organs, anatomic sites, and structures
- *Morphology:* descriptions of any change in the normal anatomy of the body at birth or during life, including changes at the cellular level, such as carcinomas
- *Etiology:* the causes or causal agents of disease, illness, or injury, including chemicals and drugs which can either attack or cure the body
- *Function:* descriptions of the normal and abnormal functions of the human mind and body, including functional states and functional units
- *Disease/diagnosis:* general and complex disease terms and syndromes with cross-mapping to ICD-9-CM
- *Procedures:* the administrative, preventive, diagnostic, and therapeutic actions taken to prevent or cure a disease, illness, or injury
- *Occupations:* this seventh axis, introduced in SNOMED II, consisted of the classification of occupations from the WHO International Labour Office and was included in SNOMED to enable studies on the effects occupation and environment have on health

At the time SNOMED was being developed for publication in 1976, the World Health Organization (WHO) embarked on a project to adapt the American Cancer Society's Manual of Tumor Nomenclature and Coding system into an international coding system for oncology. A joint committee between the CAP Committee on Nomenclature and Classification of Diseases and the WHO adopted a dual axis code based on ICD-9 topography and SNOMED morphology. Since 1979, the morphology section of ICD-O and Sections 8 and 9 of the SNOMED morphology axis have been identical.

In 1993, a new edition of SNOMED was published. SNOMED International expanded to its present level of 11 axes by splitting the etiology axis into four distinct axes and adding a new axis for linking concepts:

- *Living organisms:* all types of living organisms, from viruses to mammals and from fungi to higher-level plants of all phylogenetic groups
- *Chemicals:* common elements and their isotopes, chemical compounds, industrial chemicals, pesticides, plant and animal products, and toxins that affect the health of both man and animals
- *Physical agents, forces, and activities:* agents, activities, or forces that may cause injury. Also included are term codes for prostheses, artificial organs, hospital equipment, and devices
- *Social context:* social conditions and circumstances that may have a bearing on the patient's health and medical condition. These include ethnic and religious heritage, family status, economic conditions, etc.
- *General linkage modifiers:* terms that link the detailed elements found in each of the other modules in a meaningful way to accurately reflect the medical events they are intended to represent. Also includes terms that relate to the need to modify or qualify the diagnostic entries found in each of the other modules

Exhibit 1—Sample Hierarchy

T-28000	Lung, NOS
T-28001	Costal surface of lung, NOS
T-28002	Medial surface of lung, NOS
T-28004	Mediastinal part of medial surface of lung
T-28006	Diaphragmatic surface of lung
T-28007	Interlobar surfaces of lung
T-28008	Anterior margin of lung
T-28009	Inferior margin of lung

The American Veterinary Medical Association copublished SNOMED International, and a number of veterinary terms were added. SNOMED International has been updated annually between 1993 and 1997. Version 3.4, released in 1997, contains more than 150,000 terms and is provided on a CD-ROM in an ASCII tab-delimited format for incorporation into any database application. Numerous computer vendors have developed customized applications that use SNOMED and have licenses to incorporate SNOMED in their systems. The scientific content of SNOMED is managed by an 18-person SNOMED editorial board, which includes physicians from such specialties as pathology, radiology, and internal medicine, and representatives from the American Dental

T-28010	Pulmonary alveoli
T-28012	Alveolar sac of lung
T-28014	Alveolar duct of lung
T-28016	Atrium of alveolus
T-28020	Alveolar epithelium
T-28030	Alveolar wall

medical terms and codes.

Systematized Structure
Wherever possible, the terms in SNOMED are all arranged in a hierarchy, represented by an alphanumeric term code where each digit represents a specific location in the hierarchy. For example, T-20000 represents the respiratory system; T-28000 the lungs; and T-28010 the alveoli. An example of a typical SNOMED hierarchy is shown in Exhibit 1. In cases where a strict hierarchy is not possible, SNOMED incorporates a systematic coding scheme based on certain common characteristics among the terms.

Exhibit 2

The compositional nature of SNOMED International enables one to code the condition "contact dermatitis due to self-tanning cream" in a variety of ways:

Pre-coordinated:
This is the closest single SNOMED concept that would cover this disorder:

> D0-10322 contact dermatitis due to cosmetic

Compositional:
Depending on the level of specificity desired, there are numerous ways to code this condition by combining single SNOMED concepts. These are a few examples:

> [D0-10200] + [GC001]+ [C-52550] => [contact dermatitis, NOS] [due to] [cosmetic]
>
> [T-01000] [M-4000] + [GC001] + [C-52550] => [skin] [inflammation] [due to] [cosmetic]
>
> [D0-10200] + [GC001] + [C-92800] => [contact dermatitis, NOS] [due to] [pigmenting agent]
>
> [T-02120][M-40006] + [G-C001] + [C-92800] => [skin of face][focal inflammation] [due to] [pigmenting agent]

SNOMED-RT: The Next Generation

Association, American Veterinary Medical Association, American Nurses Association, American Academy of Ophthalmology, and the Centers for Disease Control and Prevention.

Structure of SNOMED

Systematized, multiaxial, and controlled vocabulary are three characteristics that form the essence of SNOMED and differentiate it from a simple list of

Multiaxial
Multiaxial refers to the ability of the nomenclature to express the meaning of a concept across several axes. This feature is most apparent in the disease/diagnosis axis, where each term is associated with a cross-reference field -- a series of codes from the other axes of SNOMED that comprise the essential characteristics of the disease.

Any given disease in SNOMED can be defined by one or more of the following essential characteristics: topography (the part of the anatomy that is affected by the disease); morphology (the morphologic change characterized by the disease); function (the functional characteristics of the disease); and etiology (the cause of the disease). For example, the disease "acute rheumatic fever" is identified with term code D3-17112. In the cross-reference field we find the following codes: (T-15000)(M-41000). T-15000 indicates that this disease is associated with the heart, and M-41000 indicates that its morphology is acute inflammation.

Controlled Vocabulary
In the English language there are many ways to say the same thing. This certainly applies to medicine. Take, for example, the concept "increased body temperature." In SNOMED, the following terms are classified as synonyms of increased body temperature and carry the same term code: fever, pyrexia, hypothermia, febrile. The terms feverish and hyperthermic are also given the same term code, but are classified as adjectival forms. If patient data were stored in a computer without a controlled vocabulary, a search for all patients with the condition "increased body temperature" would miss the numerous cases where this condition was recorded as fever, febrile, hyperthermic, etc. A controlled vocabulary allows individuals to record data in the patient's record using any one of a variety of synonyms, but references it back to a single primary concept. A versatile computer interface would even allow the individual clinician to use a local or personalized word, such as "temp," by providing an option that allows the clinician to map this local term as a synonym to the concept "increased body temperature" in the controlled vocabulary.

Two studies conducted by the Computer-based Patient Record Institute (CPRI) evaluated the ability of current coded nomenclatures and classification systems to accurately and completely reflect the information found in patients' medical records. The first study evaluated the ability of major coding systems to capture all concepts in the medical record.[1] Because classifications are not meant to code all the contents of a medical record, the second study was limited to controlled vocabularies.[2] In addition to completeness, the second study evaluated features such as clarity, definitions, administrative mapping, and clinical taxonomy. In both studies, SNOMED came out as the most comprehensive controlled vocabulary for coding the contents of a patient record.

Although it was widely recognized for its completeness, the studies pointed out certain weaknesses which are recognized by the developers of SNOMED and are being addressed in SNOMED-RT. RT stands for reference terminology. A reference terminology is "a set of concepts and relationships that provides a common reference point for comparison and aggregation of data about the entire healthcare process, recorded by multiple different individuals, systems, or institutions."[3]

Two of the chief strengths of SNOMED International -- its hierarchical structure and its compositional nature -- are also weaknesses. SNOMED-RT will correct these deficiencies through the use of description logic and explicit hierarchy tables.

The Compositional Structure of SNOMED

One of the major advantages of SNOMED over classification systems is its ability to describe a condition in fine detail. For example, consider the case of a patient who presents with a rash following application of a self-tanning cream. There are multiple ways to code this finding in SNOMED. It can be coded as a single concept D0-10322 "contact dermatitis due to cosmetic" or in numerous other combinations using the compositional approach shown in Exhibit 2. A search for patients with contact dermatitis due to cosmetics would easily miss any cases that were coded using the compositional approach. Humans may easily recognize that the precoordinated term and the various compositional arrangements are similar, but the computer would not. In SNOMED-RT, this deficiency will be corrected through the use of description logic.

The use of description logic will enable SNOMED-RT to explicitly define the essential characteristics of a given concept so that the equivalence of a composite concept and its constituent parts can be readily identified. The essential characteristics of a disease are typically its anatomy, morphology, and etiology. In SNOMED-RT, the essential characteristics of "contact dermatitis due to cosmetic," for example, will be explicitly defined in a computer-readable format so that whether coded as a single concept or multiple individual concepts combined together, the SNOMED-RT description logic will identify them as the same condition. An example is shown in Exhibit 3a.

The Hierarchical Structure of SNOMED

In SNOMED, the term code limits the concept to a particular hierarchical position in the nomenclature. For example, is systemic lupus erythematosus a connective tissue disease or an autoimmune disorder? It is both, but the old SNOMED coding scheme requires that it be classified as either one or the other. In this case, it was classified as a connective tissue disease. A search for patients with autoimmune disorders, if limited only to the term code, would easily miss those patients with systemic lupus erythematosus. In SNOMED-RT, the term code no longer carries the hierarchical meaning that it did in SNOMED International. In its place will be a series of relational tables with explicit relationships between the term and its parent term. These relationships can be "is a" relationships. For example, systemic lupus erythematosus "is a" connective tissue disease; and systemic lupus erythematosus "is a" autoimmune disease. Two relationships, "is a" and "part of," will form the core of expressing hierarchies in the first versions of SNOMED-RT, but numerous other relationships are possible and can be added to suit individual needs. An example of a hierarchy is shown in Exhibit 3b.

Exhibit 3

SNOMED-RT will define the essential characteristics of a disease and the hierarchies of each term so that whether coded as a single concept or multiple individual concepts combined together, the computer will recognize them as the same condition.

SNOMED term code and English nomenclature:

In SNOMED International, not only is hierarchy of a term implied by its term code, but its meaning is as well. Is the iris a flower or part of the eye? Term code T-AA553 implies that it is a topography term and therefore is part of the human anatomy. Iris" and "part of" relationships in SNOMED-RT will eliminate the guesswork and explicitly define that term code T-AA553 is "part of" the eye, and L-DC900 "is a" flower.

SNOMED in the Computer-based Patient Record

Imagine the power of a controlled terminology when used in a computer-based patient record system. Take the hypothetical case of a healthcare

D0-10322 contact dermatitis due to cosmetic

a. Defining the essential characteristics of the concept enables the computer to recognize that this disease affects the skin, presents as an inflammation, and is caused by a cosmetic.

   concept: contact dermatitis due to cosmetic
   has topography skin
   has morphology inflammation
   has etiology cosmetic cream

2. Hierarchy tables in SNOMED-RT create linkages between the more specific child term and the less specific parent term. Shown here are a few sample hierarchies for inflammation, skin, and cosmetic:

child term	relationship	parent term
focal inflammation	is a	inflammation
skin of face	part of	skin of head
skin of head	part of	skin
pigmenting agent	is a	cosmetic cream
cosmetic cream	is a	cosmetic
cosmetic	is a	skin agent
skin agent	is a	chemical

enterprise that routinely tracks the diagnoses of its patients. Let's say this enterprise notices a steady but gradual increase in the number of patients treated for atopic dermatitis, an allergic reaction affecting the skin. Why the sudden increase? Records encoded in SNOMED-RT will enable the institution to easily isolate all patients with this disorder and further probe their records to determine what they have in common. Were they taking common medications? Did they have any common underlying disorders? Was only a particular part of the body affected? Did they have a history of being exposed to the same chemical? By traversing the essential characteristics and hierarchies in SNOMED-RT, the search could easily be expanded to other related conditions -- to all patients who, for example, complain of itching or present with a rash.

Outcomes research, knowledge-based decision support systems, practice guideline protocols, and cost-effectiveness studies are just a few of the areas of research that would benefit from SNOMED-RT.

The power of a controlled vocabulary has been recognized by the clinical information systems leaders in the Kaiser Permanente Medical Group as early as 1993. In order to ensure comparability of data across all its facilities, Kaiser initiated the Convergent Medical Terminology (CMT) project. Its purpose was to build a comprehensive and functional terminology that could be integrated into its clinical information systems using SNOMED as the base. Beginning with the diseases/diagnoses and procedures axis, physicians and nurses in the Kaiser CMT project are building the explicit hierarchies and defining the essential characteristics of each SNOMED term, thus converting SNOMED International into SNOMED-RT. All the work of the Kaiser CMT project will be incorporated into SNOMED-RT.

SNOMED and the Health Information Manager

What is the relevance of SNOMED-RT to health information professionals? The advent of computerized databases containing information about a patient's health record will bring an unprecedented demand for information by healthcare providers, payers, and consumers that is impossible to gather with today's nonelectronic medical record. This research will require highly skilled individuals trained in coding and searching medical records and retrieving relevant data.

It is inevitable that the role of medical records professionals will change as the management of health information continues to automate. The profession will demand a more highly skilled individual. Instead of spending valuable time chasing down doctors for signatures or dictation, the HIM professional will spend managing collected information. Through the use of crosswalks from SNOMED to ICD and CPT, it will be possible to code clinical details in SNOMED and automatically derive the appropriate billing codes. When outcomes data and billing data derive from the same source -- the encoded patient record -- it will enable physicians, hospitals, and managed care organizations to make more consistent healthcare management decisions. Health information professionals are at the core of this model of the future of healthcare.

Notes

1 Chute, C., S. Cohn, K. Campbell, D. Oliver, and J.R. Campbell. "The Content Coverage of Clinical Classifications." *Journal of the American Medical Informatics Association* 3, no. 3 (1996): 224-33.

2 Campbell, J.R., P. Carpenter, C. Sneiderman, S. Cohn, C.G. Chute, and J. Warren. "Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity." *Journal of the American Medical Informatics Association* 4, no. 3 (1997): 238-51.

3 Spackman, K.A., K.E. Campbell, and R.A. Cote. "SNOMED-RT: a reference terminology for healthcare." *Proceedings of the 1997 AMIA Fall Symposium*, ed. D. Masys. Philadelphia, PA: Hanley & Belfus, 1997: 640-644.

Karen Kudla *is SNOMED program manager and* ***Marjorie Rallins*** *is SNOMED program analyst at the College of American Pathologists, Northfield, IL.*

Article citation:
Kudla, Karen M. "SNOMED: a Controlled Vocabulary for Computer-based Patient Records." *Journal of AHIMA* 69, no.5 (1998): 40-44.