Introduction to the Unified Medical Language System

Save to myBoK

by Susan H. Fenton, MBA, RHIA

Coded data is the basis for knowledge transfer in 21st-century healthcare. The Unified Medical Language System (UMLS) is often called the Rosetta stone of international terminologies used in healthcare worldwide.

The purpose of the UMLS is to "facilitate the development of computer systems that behave as if they 'understand' the meaning of the language of biomedicine and health."[1] This goal is achieved through three "knowledge sources":

- The UMLS Metathesaurus
- The SPECIALIST Lexicon
- The UMLS Semantic Network

Coders need to understand these new tools because they introduce other types of code sets designed for roles in e-HIM™. This article examines each of these sources, as well as the MetamorphoSys installation and customization tool.

What Is the UMLS?

The UMLS (www.nlm.nih.gov/research/umls) is a US government–funded project from the National Library of Medicine (NLM). In development since 1986, the UMLS has played a large role in emerging national standards for the electronic health record. It is specifically designed to facilitate development of systems that can retrieve and integrate electronic biomedical information from a variety of sources. The familiar ICD-9-CM, HCPCS, and CPT coding systems used for claims processing are found within UMLS, as are more than 100 other terminology and vocabulary systems.

UMLS is the source for the US license for SNOMED CT and contains the SNOMED to ICD-9-CM maps. The system is an internationally recognized clinical vocabulary resource unlike any other. Some health information managers will use its data as they participate in organization-wide standard efforts or the implementation of an electronic health record; however, all HIM professionals need to be aware of the UMLS and its evolution.

UMLS Metathesaurus

The Metathesaurus is a vocabulary database that contains biomedical concepts and terms from more than 100 controlled vocabularies and classifications used in patient records, administrative health data, bibliographic and full-text databases, and expert systems. It "preserves the names, meanings, hierarchical contexts, attributes, and inter-term relationships present in its source vocabularies; adds certain basic information to each concept; and establishes new relationships between terms from different source vocabularies."[2]

The Metathesaurus source vocabularies include terminologies designed for use in patient record systems; large disease and procedure classifications used for statistical reporting and billing; more narrowly focused vocabularies used to record data related to psychiatry, nursing, medical devices, and adverse drug reactions; disease and finding clinical terminologies from expert diagnostic systems; and some thesauri used in information retrieval.

The Metathesaurus can assist information professionals in creating computer programs to interpret user inquiries, interact with users to refine their questions, identify which databases contain information relevant to particular inquiries, and convert the users' terms into the vocabulary used by relevant information sources. The scope of the Metathesaurus for any given application "is determined by the combined scope of the selected vocabularies."[3]

Unlike ICD-9-CM, which is organized by disease, "the Metathesaurus is organized by concept or meaning. Alternate names for the same concept (synonyms, lexical variants, and translations) are linked together. Each Metathesaurus concept has attributes that help to define its meaning, e.g., the semantic type(s) or categories to which it belongs, its position in the

hierarchical contexts from various source vocabularies, and, for many concepts, a definition. A number of relationships between different concepts are represented. Some of these relationships are derived from the source vocabularies; others are created during the construction of the Metathesaurus."[4]

Each concept, idea, or meaning in the Metathesaurus has a unique concept identifier (CUI), which has no intrinsic meaning. Although it may be useful to think of a CUI as a code number, CUIs were developed for computer systems. They are not organized for human readability. For example, "cold" with a meaning relating to temperature would have a different CUI than "cold" with a meaning of disease.

Each concept name, or string, in each source vocabulary is identified using a unique atom identifier (AUI). AUIs can be thought of as the building blocks of the Metathesaurus. Each concept or string from each source vocabulary will have an AUI, even if it is synonymous with other concepts or strings. Thus, if the same concept exists in exactly the same format with exactly the same meaning in two different vocabularies, it will receive two separate AUIs—one AUI for the concept in each source vocabulary.

Each unique concept name, or string, in each language in the Metathesaurus has a unique string identifier (SUI). Any variation in upper-lower case is a separate string with a separate SUI. For example, "Cold" would have a different string identifier than "cold." The same string in different languages (e.g., English and Spanish) will have a different string identifier for each language. If a given string—"Cold," for example—appears in several source vocabularies, it is possible to have multiple AUIs linked to one SUI.

For English-language entries in the Metathesaurus, each string is linked to all of its lexical variants or minor variations by means of a common term identifier or a unique lexical identifier (LUI). All string and term identifiers are linked to at least one concept identifier. Different terms with the same meaning are linked to the same concept identifier. "Unique Identifier Relations," below, illustrates how a CUI might be related to AUIs, SUIs, and LUIs.

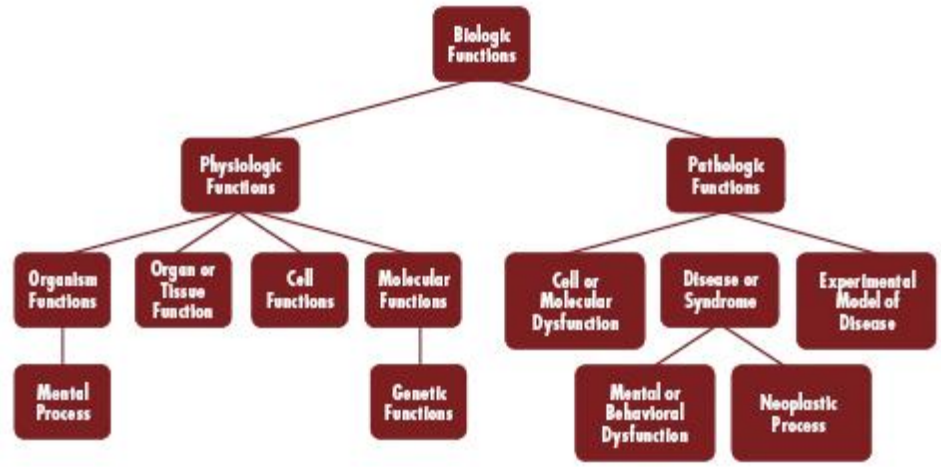
Content of the Metathesaurus

The 2004AC edition of the Metathesaurus includes more than one million concepts and five million concept names from more than 100 different controlled vocabularies, some in multiple languages. It includes US standards such as SNOMED CT and LOINC and is updated quarterly.

Unique Identifier Relations

Concept (CUI)	Terms (LUIs)	Strings (SUIs)	Atoms (AUIs) *RRF Only
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)	A0027665 Atrial Fibrillation (from MSH) A0027667 Atrial Fibrillation (from PSY)
		S0016669 Atrial Fibrillations	A0027668 Atrial Fibrillations (from MSH)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016899 Auricular Fibrillation (preferred)	A0027930 Auricular Fibrillation (from PSY)
		S0016900 (plural variant) Auricular Fibrillations	A0027932 Auricular Fibrillations (from MSH)

Biologic Function Hierarchy



UMLS Semantic Network

The Semantic Network, through its 135 semantic types or meanings, provides "a consistent categorization of all concepts represented in the UMLS Metathesaurus."[5] The 54 links, or relationships, between the semantic types "provide the structure for the network and represent important relationships in the biomedical domain."[6]

"All information about specific concepts is found in the Metathesaurus; the network provides information about the basic semantic types"[7] that are assigned to these concepts, and it defines the relationships that may be present between the semantic types. The Semantic Network is the road map of the UMLS. Examples of semantic types include organisms, anatomical structures, chemicals, and physical objects. "Biologic Function Hierarchy," above, shows a portion of the network.

"The primary link or relationship is the 'isa' link. This establishes the hierarchy of types within the network and is used for deciding the most specific semantic type available for assignment to a Metathesaurus concept."[8]

More information on the Semantic Network, as well as a complete list of semantic types and relations between semantic types, can be found at www.nlm.nih.gov/research/umls/meta3.html.

SPECIALIST Lexicon

The SPECIALIST lexicon is an English-language lexicon with many biomedical terms. It was developed in the context of the SPECIALIST natural language processing project at NLM.

The lexicon entry for each word or term records the following information: syntactic (formal properties of languages); morphological (study and description of word formation in a language, including inflection, derivation, and compounding); and orthographic (correct spelling or the representation of the sounds of a language by written or printed symbols). Lexical entries may be single or multiword terms. Entries sharing their base form and spelling variants, if any, are collected into a single lexical record. The base form is the uninflected form of the lexical item; the singular form in the case of a noun, the infinitive form in the case of a verb, and the positive form in the case of an adjective or adverb.

For more information on the SPECIALIST Lexicon, go to http://specialist.nlm.nih.gov/Specialist LexiconFactSheet.pdf.

MetamorphoSys

MetamorphoSys is an installation and customization software that enables UMLS users to exclude any vocabulary they do not need or for which they have not negotiated the additional license arrangements when required by the license agreement for

their intended uses of the Metathesaurus. The software also makes it easier for UMLS users to select from a variety of data output options and filters. Because of the size of the UMLS, it is strongly recommended that users request a copy of the DVD to run MetamorphoSys.

Obtaining the UMLS

The UMLS tools and knowledge sources are available at no charge to users. Users are asked to sign a licensing agreement consenting to appropriately acknowledge any use of the UMLS in their research or other efforts. The license agreement can be found at www.nlm.nih.gov/research/ umls/access.html.

The UMLS has become a recognized source of clinical and biomedical vocabularies and classifications. Further, the NLM has developed software tools to assist developers, as well as an international distribution mechanism via the Internet. Because of this, it is important for health information managers to familiarize themselves with its content, structure, and potential uses.

Notes

1. National Library of Medicine, National Institutes of Health. “About the UMLS Resources.” Available online at www.nlm.nih.gov/research/umls/about_umls.html.
2. NLM, NIH. “Fact Sheet: UMLS Metathesaurus.” Available online at www.nlm.gov/pubs/factsheets/umlsmeta.html.
3. Ibid.
4. Ibid.
5. NLM, NIH. “About the UMLS Resources.”
6. NLM, NIH. “Fact Sheet: UMLS Semantic Network.” Available online at www.nlm.nih.gov/pubs/factsheets/umlssemn.html.
7. NLM, NIH. “About the UMLS Resources.”
8. NLM, NIH. “Fact Sheet: UMLS Semantic Network.”

Susan Fenton *(susan.fenton@ahima.org) is a professional practice manager at AHIMA.*

Article citation:
Fenton, Susan. "An Introduction to the Unified Medical Language System." *Journal of AHIMA* 76, no.5 (May 2005): 60-62.

Driving the Power of Knowledge

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.