INTEGRATING KEY BIOMEDICAL TERMINOLOGIES

NCI Metathesaurus

NATIONAL® CANCER INSTITUTE

Enterprise Vocabulary Services

July 3, 2006

TABLE OF CONTENTS

Executive Summary

Intended Audience

This paper is for those who need to understand the structure and use of NCI Metathesaurus data, for those who will download the NCI Metathesaurus Rich Release Format (RRF) distribution, and for informaticians who use the caCORE object model-based Application Program Interface (API) and who want to understand how the data accessed by the API is structured and the logic behind those structures.

Strategic Goals for NCI Metathesaurus
• To provide a central uniform access point for relevant biomedical terminologies • To clarify which terms are equivalent • To facilitate consistent annotation and retrieval of data • To facilitate the facilitate the consistent annotation and retrieval of data.

Historical Overview

In the past, cancer researchers and others have found it difficult to exchange information because of the disparate terminologies being used. Terminologies often covered cancer poorly and were not machine-readable. Translations among terminologies were rare.

To facilitate cancer research, it was necessary to provide users with a vocabulary tool that enabled them to communicate clearly with one another as well as to conduct their research more easily and efficiently. In addition, due to the complexity and nature of the terms themselves as well as the confusing patchwork of clinical and research terms in current use by various institutions, it was imperative to have a tool that provided mappings of terms across vocabularies.

Although there was a tool already in existence that mapped terms across medical vocabularies—the National Library of Medicine (NLM) Unified Medical Language System (UMLS) Metathesaurus—it was not cancer-centric.

Rather than duplicate a method already in existence, the National Cancer Institute (NCI) obtained permission from the NLM to use the UMLS Metathesaurus as the foundation for a new tool called *NCI Metathesaurus.* While it is based on the UMLS version, the NCI Metathesaurus drops the use of less relevant sources, adds various cancer-oriented special purpose terminologies and makes some changes in cancer-

related concepts. The NCI Metathesaurus assists users in finding appropriate terms and translations corresponding to relevant biomedical terminologies.

The NCI Metathesaurus is published monthly and provides an up-to-date consolidated resource, making it easier for users to remain current with the terminologies in cancer prevention, treatment and research.

Focus of Paper

This paper explores some of the challenges users face when searching terminologies and explains how the NCI Metathesaurus addresses these issues. It also describes the NCI Metathesaurus content sources and its organization. It outlines the differences between the NCI Metathesaurus, the UMLS Metathesaurus and the NCI Thesaurus, and provides an overview of how to access and work with the NCI Metathesaurus.

NCI Metathesaurus Overview

The NCI Metathesaurus is a biomedical terminology database. It contains more than 50 distinct terminology products. While the NCI Metathesaurus is based on the UMLS Metathesaurus, the NCI version is more cancer-centric and includes a growing number of cancer-centric biomedical vocabularies. The NCI Metathesaurus is intended to provide users with basic, translational, clinical and population-based terminologies, plus a great deal of general biological, medical and epidemiological terminology, as well.

The NCI Metathesaurus contains public domain vocabularies from the UMLS Metathesaurus, NCI-specific vocabularies and some proprietary vocabularies. Certain vocabularies such as ICD-10 and ICD-O3 are made available by permission for noncommercial use only. Others, specifically MedDRA, are made available subject to license restrictions (see http://www.nlm.nih.gov/research/umls/license.html for details).

The Enterprise Vocabulary Services Project

The NCI Metathesaurus arose from an internal research project called the Science Information System (SIS). SIS was intended to provide knowledge management tools to NCI executive management. Work on the knowledge management revealed that a source of controlled terminology was a *sine qua non*. NCI Metathesarus was designed to meet the terminology needs of SIS, but survived long after the SIS project ended.

The NCI Center for Bioinformatics (NCICB) and the NCI Office of Communications (NCIOC) collaborated to create a set of services and resources that would address NCI's terminology needs and so formed the Enterprise Vocabulary Services (EVS) Project around the NCI Metathesaurus.

NCICB provides bioinformatics infrastructure, including the Common Ontologic Resource Environment (caCORE), which includes a metadata repository (caDSR), data models (caBIO), common security infrastructure (CSM) and a toolkit for the easy creation of semantically integrated caCORE-like systems (caCORE SDK). In addition to caCORE, NCICB supports the national data and application sharing resource (caGRID) that links together NCI and the cancer centers and other organizations that make up the cancer Biomedical Informatics Grid (caBIG) project.

NCIOC provides evidence-based cancer information services, including Physician Data Query (PDQ), the Clinical Trials Portal and the Cancer Information Service. The NCIOC also provides the NCI Web Portal, which includes access to the cancer information services.

EVS is a key component of caCORE and caBIG and supports the NCI Web Portal and PDQ cancer information services. EVS also facilitates the standardization of vocabulary across NCI as well as the larger biomedical domain by providing access to controlled terminologies. The main EVS products include:

- **NCI Thesaurus,** http://nciterms.nci.nih.gov/ **–** EVS provides a logic-based reference terminology and biomedical ontology used in a growing number of NCI and other systems. The NCI Thesaurus provides rich textual and ontologic descriptions of more than 50,000 key biomedical concepts and is provided under an open content license.

- **NCI Metathesaurus,** http://ncimeta.nci.nih.gov **–** EVS provides a comprehensive biomedical terminology database which cross-links meanings from more than 60 terminologies. It currently contains most public domain vocabularies from the National Library of Medicine's UMLS Metathesaurus as well as a growing number of other biomedical vocabularies. Some proprietary vocabularies are included that have restrictions on their use.

- **Informatics Infrastructure –** EVS develops software to create, manage and update terminology data, and operates a number of servers that provide access to terminology data. Software professionals and terminologists can visit NCI's open development site to join with the NCI in these efforts, http://gforge.nci.nih.gov/softwaremap/trove_list.php?form_cat=434.

EVS licenses terminology products such as MedDRA for NCI use and arranges permission for use of certain terminologies by the cancer community.

EVS collaborates with a number of NCI and affiliated organizations, such as the U.S. Food and Drug Administration (FDA), U.S. Department of Veterans Affairs (VA) and the Centers for Disease Control and Prevention (CDC); major standards organizations, including Health Level Seven (HL7), the Clinical Data Interchange Standards Consortium (CDISC) and the World Wide Web Consortium (W3C); cancer centers and cooperative groups; and research collaborators such as the Microarray Gene Expression Data Society (MGED).

Areas of collaboration in terminology development and publishing include drugs, devices and clinical trials with the FDA; drugs, clinical trial semantics and terminology operations with the VA; cancer incidence and prevention as well as terminology operations with the CDC; clinical trials, experimental organisms, micronutrients and open terminology servers with cancer centers and caBIG; clinical research data standards with CDISC/HL7& RCRIM; and more.

Uses of the NCI Metathesaurus

The NCI Metathesaurus is useful when:

- Users know the code for something such as a clinical finding in one terminology, but need to find the corresponding code in another terminology.
- Users come across a biomedical term, but do not know what it means.
- Users know a term, for example "breast-cancer," and want to find the more specific terms subsumed by it.
- Users need a comprehensive source of cancer-oriented terminology for text processing of clinical, scientific or administrative documents.

The NCI Metathesaurus includes mappings among SNOMED, MedDRA, PDQ and the NCI Thesaurus terminologies, as well as sources that are not included in UMLS, such as Clinical Bioinformatics Ontology (CBO).

The NCI Metathesaurus includes NLM's Semantic Network, which provides semantic types (or broad classification categories) that group concepts together. Many biomedical terms are ambiguous. For example, the term "mole" can refer to a mammal, a unit of measure, a skin nevus, or a clinical finding. Each of these senses of "mole" are represented in the NCI Metathesaurus as semantic types. If you encounter the term mole in a chemistry paper, you can use the semantic type "quantitative concept" to identify the correct meaning of mole.

The NCI Metathesaurus is a reference vocabulary source for:

- Creation of metadata describing data models and research and clinical processes and entities, all components of semantic integration in caCORE.
- Processing biomedical text, as in caTIES (Cancer Text Information Extraction System), which evolved out of the Shared Pathology Informatics Network (SPIN). caTIES processes pathology reports that try to identify types of cancers and procedures.
- Public information (online dictionary and references), using the rich synonymy to facilitate searching in other applications.

Accessing the NCI Metathesaurus

The NCI Metathesaurus can be accessed at http://ncimeta.nci.nih.gov. The data files can be downloaded or a browser is provided to access the Metathesaurus content. More details on accessing the data files can be found in *For assistance, telephone NCICB Application Support at 301-451-4384 or 888-478-4423 (toll-free) or email ncicb@pop.nci.nih.gov.* on page 16.

User Support

User support is available via email at ncicb@pop.nci.nih.gov and also via phone at 301-451-4384 or toll free at 888-478-4423.

Structure Overview

The NCI Metathesaurus is based on the UMLS Metathesaurus, which is a large, multipurpose, multilingual vocabulary database containing information about

biomedical and health-related concepts, their various names and the relationships among them. The UMLS Metathesaurus is built from source vocabularies comprising electronic versions of many different thesauri, classifications, code sets and lists of controlled terms used in patient care, health services billing, public health statistics, indexing and cataloging biomedical literature, and/or basic, clinical and health services research. Most of the contents of the UMLS Metathesaurus is included in the NCI Metathesaurus except for certain proprietary vocabularies and other vocabularies whose content has not been updated over the years. In addition, the NCI Metathesaurus includes terminologies developed by cancer researchers to meet the specific needs of the cancer community.

The NCI Metathesaurus is organized by ***concept***, described in the following section.

Overview of a Concept

Often it is the meaning of a word and not the word itself that is significant; therefore, the NCI Metathesaurus is organized by ***concepts,*** or meanings. Each term in the NCI Metathesaurus is assigned to a concept, which forms the basis or mechanism for mapping a term in one source with one or more corresponding terms in other sources. Each concept has its own distinct meaning. The meaning of an NCI Metathesaurus concept should be thought of as the sum of the meanings of each term associated with it. For example, if the term "red" means "energy at wavelength 440 nm," and in another terminology it is defined as "visible light perceived to be opposite green on the color wheel," then the concept for "red" would be defined to include both of these meanings.

Concepts are linked to one or more ***semantic types,*** which help classify all the concepts under a semantic network. For example, "cold" with the semantic type "Physiologic Function" specifies that this sense of "cold" denotes the sensation of being chilly, while "cold" with the semantic type "Disease or Syndrome" denotes a respiratory infection.

Useful ***relationships*** between different concepts are represented in the NCI Metathesaurus. For example, the concept for "rat" is related to the concept for "rodent" by a narrower-than relationship because "rodent" is a more general concept than "rat."

New concepts are created when a new vocabulary source is inserted into the NCI Metathesaurus and during the process of editing when a concept is split into two.

Anatomy of a Concept

Concepts contain **atoms, contexts, attributes, semantic types,** and **relationships.**

- **Atoms**—Atoms can be one word, two words or whole phrases that come from a terminology to represent a meaning. Terms make up the synonymy that is present in a concept. In the Metathesaurus data, each term is tagged to indicate from which terminology it came.
    - **Lexical variants** are names that are identical after a series of well-defined manipulations are done computationally, such as making all characters lower case, putting all words in a defined order and changing all plural forms to a singular.
    - **Terms** are names with lexical variants or minor variations that break down to the same lexical term. A term is the group of all strings that are lexical variants of each other. (For example: Eye, Eyes, eye = 1 term.)

- **Strings** are unique concept names that take into account variations in upper-lower case, plurality and punctuation. (For example: Eye, Eyes, eye = 3 strings.)

- **Contexts**—Contexts come from the sources that provide them. There is no overall Metathesaurus context. For example, MeSH (Medical Subject Headings) is in a tree structure that is represented as context in the Metathesaurus.

  Note: Contexts are the same as hierarchies. See *Hierarchies* on page 11.

- **Attributes**—Attributes provide more information regarding terms from a source, such as additional codes, mappings to other sources and references.

- **Semantic Types**—Semantic types come from the UMLS Semantic Network (one of two knowledge bases connected by the UMLS Metathesaurus) that provides consistent categorization for all concepts contained in the UMLS Metathesaurus and the NCI Metathesaurus, and defines a set relationship among those concepts. There are both chemical and non-chemical semantic types. The major groupings of semantic types include organisms, anatomical structures, biologic function, chemicals, events, physical objects and concepts or ideas. Each NCI Metathesaurus concept is assigned at least one semantic type and, in some cases, several. In all cases, the most specific semantic type available in the network hierarchy is assigned to the concept.

- **Relationships**—Relationships link concepts to each other. Some relationship types are defined in the UMLS Semantic Network. Other relationships are derived from sources themselves or are created by editors during editing. *Table 2.1* shows an example of relationship types among geographical concepts. There are many relationship types in the Metathesaurus. Several of the most common are *isa*, *part_of* or *result_of*.

Maryland Preferred Concept Name CUI C0024858 CUI STY Geographic Area Semantic Type Atoms: (name and source it came from) Maryland - NCI2006_01C Maryland - MSH2005_2004_10_12 Maryland - AOD2000 Contexts: Alcohol and Other Drug Thesaurus geographic area Americas North America United States Southeastern US Maryland Relationships: Southeastern United States is Broader than Maryland Regions of United States of America is Broader than Maryland

Table 2.1 Example of a simplified concept and its relationship types

Identifying Concepts, Strings, Atoms & Terms

Since the NCI Metathesaurus is organized by concept, there is a need to ensure the unique identification of all concepts, strings, terms and atoms. Each of these concept components is assigned identifiers in addition to retaining any identifiers already present in the source vocabularies.

Concept Identifiers

Because the NCI Metathesaurus incorporates the UMLS Metathesaurus, concept identification in both metathesauri is substantially the same; both identify concepts by the ***concept unique identifier*** (CUI). The CUI has no intrinsic meaning. UMLS Metathesaurus CUIs are represented by the letter **C** followed by seven digits (for example, C0010028). This identifier remains the same across versions of the UMLS Metathesaurus, irrespective of the term designated as the preferred name of the concept. Since the January 2001 release, the UMLS Metathesaurus has included the MRCUI file that contains mappings for CUIs that disappear. See *Appendix A* to view an excerpt of the MRCUI file.

CUIs are never re-used; once a meaning is associated with a CUI, the CUI will always signify that meaning. However, CUIs can be 'retired'. Retired CUIs are removed from the main body of UMLS Metathesaurus. For this reason, CUIs should not be used to code instance data in data repositories when search and retrieval operations depend on the stored data codes being available in the namespace.

Because the NCI Metathesaurus contains NCI local sources in addition to those that appear in the UMLS Metathesaurus, users of the NCI Metathesaurus will see concept numbers of the form CL###### as well as C#######. The CL number lets users know that the concept occurs only in a NCI local source[1].

CUIs help support data exchange and linking as well as assist migration among individual source vocabularies. CUIs also help track meaning across versions of the NCI Metathesaurus. CUIs can be used to track concepts, but not atom names since atom names may change if a source changes its terminology.

String Identifiers

Each unique string has a ***string unique identifier*** (SUI). SUIs are formatted as an S followed by seven digits. Variations in character sets, case or punctuation are considered to be separate strings, with separate SUIs. There are also separate SUIs if strings are published in a different language. If a string has more than one meaning, the SUI will be linked to more than one CUI.

Atom Identifiers

Each and every occurrence of a name in each source vocabulary is assigned a ***unique atom identifier*** (AUI). AUIs are formatted as A followed by seven digits. If a string appears more than once in the same vocabulary source (for example, both the long string and the short string), a unique AUI is assigned for each occurrence. If a string appears in one or more source vocabularies, each string will have a unique AUI, however, each AUI will be linked to a single string identifier (SUI) since they represent

1. The terms "local source" or "local terminology" indicate that the NCI has inserted the terms into the metathesaurus, as opposed to getting the terms from the UMLS. NCI does not create the terms itself, but it might obtain them from medical information systems or a foreign scientist, for example.

occurrences of the same string. Unlike string identifiers, a single AUI is always linked to a single concept identifier because each occurrence of a string in a source can only have one meaning.

Terms and Lexical Identifiers

For English-language entries only, each string is linked to all of its lexical variants or minor variations by a ***lexical unique identifier*** (LUI). LUIs are formatted as an L followed by seven digits. The LUI for an English string may be linked to one or more concepts. This occurs when strings that are lexical variants of each other have different meanings. LUIs may be valuable in detecting and dealing with minor variations in language.

Concept, String, Atom and Term Identifier Uses

Identifiers in the NCI Metathesaurus have a logical relationship in that:

- Every CUI (concept) is linked to at least one AUI (atom), SUI (string) and LUI (term).
- Every AUI (atom) is linked to a single SUI (string), a single LUI (term) and a single CUI (concept).
- Each SUI (string) can be linked to many AUIs (atoms), to a single LUI (term) and to more than one CUI (concept).
- Each LUI (term) can be linked to many AUIs (atoms), many SUIs (strings) and more than one CUI (concept).

For example, as shown in *Table 2.2*, Atrial Fibrillation appears as an atom in more than one source vocabulary and has a distinct AUI for each occurrence. Since each of these atoms has an identical string or concept name, they are linked to a single SUI. Atrial Fibrillations, the plural of Atrial Fibrillation, has a different string identifier. Since the singular and plural are lexical variants of each other, both are linked to the same LUI. There is a different LUI and different SUIs and AUIs for Auricular Fibrillation and its plural Auricular Fibrillations. Since Atrial Fibrillation and Auricular Fibrillation have been judged to have the same meaning, they are linked to the same CUI.

Identifiers	***Examples***
Concepts (CUIs)	**C0004238** Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations

Table 2.2 Identifier Uses; also shows relationship types among geographical concepts

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

Table 2.2 Identifier Uses; also shows relationship types among geographical concepts

CUIs link information that correlates with particular concepts; therefore, CUIs can be used to retrieve all the concept names, relationships, and attributes for a particular concept. CUIs also serve as permanent, publicly available identifiers for biomedical concepts or meanings to which many individual source vocabularies are linked.

Overview of Synonomy

Synonomy is a list or collection of synonyms (words or expressions that have the same or nearly the same meaning). ***Concepts*** link various names of the same meaning. If there are synonymous strings within a source and among sources, they will appear in the same concept.

How Synonymy is Determined

Determining synonymy can be difficult and there is no foolproof method. Examples of how synonymy is determined are shown in *Table 2.3*.

Determinants	Examples
Asserted by a vocabulary source	UWDA asserts that canine tooth is synonymous with cuspid.

Strings with only lexical variation (varying only in singular/plural form, direct/indirect form, punctuation, and so on)
 	
Note: Not all lexical variants are synonymous.
	• **Singular/plural:** Tooth and teeth, tumor and tumors, neoplasm and neoplasms
	• **Direct/indirect**:
	  ◦ Malignant Neoplasm & Neoplasm, malignant
	  ◦ Bipolar Disorder & Disorder, Bipolar
	• **Punctuation:**
	  ◦ Insulin-like receptor & insulin like receptor
	  ◦ Breast Feeding and Breast-Feeding

Synonymy between unlike strings	• Breast Feeding and Nursing • ICP and Intracranial Pressure • PEEP and Positive End Expiratory Pressure • CHOP and Cyclophosphamide/Doxorubicin/Prednisone/Vincristine
Clues from the source such as contexts, definitions and scope notes	• MedDRA does not contain chemicals so no names from that source should be in a chemical concept. • Context gives clues to meaning: SNOMED example referring to Marital Separation: SNOMEDCT_2005_07_31/ SNOMEDCT_2006_01_31/PT1/160788005 SNOMED CT Concept Clinical finding Finding by method History finding Social and personal history finding Life event finding <Separation>

Table 2.3 Examples of How Synonymy is Determined

Determinants	Examples
Relying on biomedical knowledge	• Burkitt's Lymphoma and Small Non-Cleaved Cell Lymphoma, Burkitt's Type are the same neoplastic disease. • Diazepam is the generic name for the drug Valium. • Developmental Articulation Disorder is an old term for Phonologic Disorder. • Vulpes vulpes is the scientific name for the Red fox.

Table 2.3 Examples of How Synonymy is Determined (Continued)

Examples of Synonymy

The following examples of synonymy were excerpted from the NCI Metathesaurus:

- Neoplasm is synonymous with tumor, tumors, neoplastic growth and neoplasia.
- AB-DIC is synonymous with Bleomycin/Dacarbazine/Doxorubicin/Lomustine/Prednisone.
- Breast cancer is synonymous with Carcinoma of Breast.
- Emesis is synonymous with Vomiting.
- IV Fluid is synonymous with Intravenous Fluid.
- Meperidine is synonymous with Demerol is synonymous with Pethidine.
- CML is synonymous with Chronic Myelogenous Leukemia.
- Liver Cancer is synonymous with Malignant neoplasm of liver.

Hierarchies

Hierarchies that are present in the Metathesaurus are source-specific and are maintained by the source and are relevant only to that source. For example, a term in the MeSH hierarchy may display a different parent than a corresponding term in SNOMED.

Hierarchies are source-specific.

Not all sources possess hierarchies; for example, LOINC is a flat terminology. However, various hierarchies and role relationships are preserved and are available in the RRF files. See *For assistance, telephone NCICB Application Support at 301-451-4384 or 888-478-4423 (toll-free) or email ncicb@pop.nci.nih.gov.* on page 16 for more information.

Broader, Narrower and Related Concepts

Broader (or More General) and Narrower (or More Specific) relationships are Metathesaurus relationships among concepts that are derived from those sources that contain hierarchy structures.

It is important to note that the Broader Than (BT)/Narrower Than (NT) relationships do not represent hierarchy. There is no inherent correspondence between the BT and NTs and any *specific* hierarchy.

> Broader Than or Narrower Than relationships do not represent hierarchies.

Taken across all sources with hierarchies in which the concept occurs:

- Each concept may have one or more broader concepts whose semantic content is a generalization of the selected concept.
- A concept may have 0 to many descendants, where each descendant concept is a specialization of the selected concept.

Thus, the list of ***broader*** concepts is the compendium of antecedent concepts from all of the sources that have hierarchies and the list of ***narrower*** concepts is the set of all descendant concepts over all such vocabularies.

The list of ***related*** concepts encompasses a broader and less well-defined set of relations, as it depends on the semantic relations defined in the contributing vocabularies. Some vocabularies, such as the NCI Thesaurus, define very sophisticated and specific relations, such as the fact that a particular bacterium is the etiologic agent of a specific disease. Other sources provide only primitive relations indicating that two concepts depend on one another in unspecified ways.

Note: The NCI Metathesaurus does not differentiate among relationships but instead amasses sophisticated precise relationships from sources such as the NCI Thesaurus with vague relationships from other sources; for example, inverse (has_location is_location_of), narrow relationship (is_a, part_of) or related relationship (is_mapped_from). This information is displayed when it is available in the source to help users determine if they want to follow the link to view the information on the concept pointed to by the link.

Sources of Metathesaurus Content

Sources of Metathesaurus content come from many different entities. Examples of source vocabularies include terminologies used 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, adverse drug reactions, and so on; disease and finding terminologies from expert diagnostic systems; and some thesauri used in information retrieval, such as MeSH.

Whenever data is included in the NCI Metathesaurus, the source is identified by a unique abbreviation; for example, MSH2001 means MeSH from 2001.

UMLS vs. NCI Sources

UMLS vocabularies broadly cover four communities: medicine, nursing, drugs and certain other clinical disciplines and biological science.

The NCI Metathesaurus contains most public domain vocabularies from NLM's UMLS Metathesaurus, as well as a growing number of other biomedical terminologies needed by NCI and its partners, such as the Mitelman Database of Chromosome Aberrations in Cancer. The NCI Metathesaurus also contains a limited set of proprietary vocabularies. Certain vocabularies, such as ICD-10 and ICD-O3, are made available, by permission, for noncommercial use only. Others, specifically MedDRA, are made available subject to license restrictions. See http://www.nlm.nih.gov/research/umls/license.html for details about license restrictions.

Accessing UMLS Sources

To access the current list of sources for the UMLS Metathesaurus, go to http://www.nlm.nih.gov/research/umls/ and click the **List of Sources** link. You can also get a list of UMLS sources included in NCI Metathesaurus by going to the same link as displayed in the following section, *Accessing NCI Sources*. Note that these links direct you only to the *list* of sources, not to the sources themselves.

Accessing NCI Sources

NCI local sources are included to address specific needs of the National Cancer Institute[2]. To access these sources, go to the main NCI Metathesaurus Browser page, http://ncimeta.nci.nih.gov/indexMetaphrase.html and click the **Sources** link.

Note: Most NCI local terminologies are incorporated into the NCI Metathesaurus in their entirety, but other terminologies are only partially included. For example, only a fragment of the full NCI Developmental Therapeutics Program (DTP) terminology is present in the NCI Metathesaurus. In contrast, the Common Terminology Criteria for Adverse Events (CTCAE) is included in the NCI Metathesaurus in its entirety. NCI local terminologies are color-coded in the NCI Metathesaurus sources list. Users can view this list from the NCI Metathesaurus Browser page located at http://ncimeta.nci.nih.gov/indexMetaphrase.html by clicking the **Sources** link.

Note: Appendix B provides a representative comparison of UMLS and NCI sources.

NCI Metathesaurus Updates

The EVS Project updates the NCI Metathesaurus on a monthly basis. The update is a minor update in some months, while in other months, the update is considered to be a major update.

- Major updates indicate that the UMLS Metathesaurus release, on which the NCI Metathesaurus is based, has been changed.

2. The terms "local sources" or "local terminologies" indicate that the NCI has inserted the terms into the metathesaurus, as opposed to getting the terms from the UMLS. NCI does not create the terms itself, but it might obtain them, for example, from a medical information system or a foreign scientist.

- Minor updates indicate that the version of the NCI Thesaurus or versions of other NCI-maintained sources have changed, or a new source has been added, but the underlying UMLS release remains the same.

NCI Metathesaurus releases are named using a convention that reflects the release of the NCI Thesaurus that the Metathesaurus release contains. For example, the Metathesaurus release “Metathesaurus 2006O1C.RRF” contains the NCI Thesaurus release 06.O1C. Each monthly update may also include updates to other sources in the NCI Metathesaurus or new insertions.

The EVS Project takes two or three “builds” (which correspond roughly to a major release of a software product) of the UMLS Metathesaurus data and incorporates this data into the NCI Metathesaurus each year. UMLS builds may contain new sources and existing sources may be dropped or supplanted.

How Builds are Numbered
The NCI Metathesaurus is built on a specific base release of UMLS. The NCI Thesaurus browser identifies both the UMLS build, and the NCI Thesaurus build that it contains. For instance, "UMLS build M05AA, covers up to NCI Thesaurus version 06.01C" means that the release is based on the UMLS 2005AA build, and includes the NCI Thesaurus release that was the first build of January 2006. The second January build would be 06.01B and so forth. There is no way to determine if an NCI Metathesaurus release is a major or minor release by inspecting the release name. The release documentation specifies that information.

Relationship with the UMLS Metathesaurus

The NCI Metathesaurus is not in competition with UMLS Metathesaurus; rather, the National Cancer Institute is a principal customer of the National Library of Medicine and uses NLM’s UMLS Metathesaurus in mission-critical operational infrastructure systems.

NCI modifies the UMLS Metathesaurus data to enable NCI to incorporate cancer-specific sources and to exclude general-purpose sources that cause problems with cancer-centric semantics. While the NCI Metathesaurus is based on the NLM’s UMLS Metathesaurus, it drops less useful sources and adds special purpose terminologies.

Relationship with NCI Thesaurus

The NCI Thesaurus is designed for database coding, search and data-mining. It serves as a reference terminology for NCI and its partners. The NCI Thesaurus provides the base semantics for caCORE and caBIG metadata and models, and is a federal standard terminology.

While not all of the sophisticated semantic relationships that exist in the NCI Thesaurus are available in the NCI Metathesaurus, all of the NCI-specific terms in NCI Thesaurus are mapped to corresponding terms used in standard biomedical vocabularies.

The NCI Thesaurus Has Richly Structured Content

The NCI Thesaurus has richly structured content for people and computers with 50,000 concepts comprising more than 130,000 synonyms and codes, 30,000 text definitions and 150,000 role relationships that semantically connect concepts. The NCI Thesaurus is built using description logic formalisms, which enables use of automated classification techniques. Automated classification assures that the Thesaurus contains no duplicate concepts or orphan concepts and reduces the possibility that the NCI Thesaurus contains tautologies.

Broad coverage of cancer-related clinical, research and other subject areas is provided in the NCI Thesaurus. For example, the following are included:

- Neoplastic and other diseases, findings and abnormalities
- Anatomy, tissues and subcellular structures
- Agents, drugs and chemicals
- Genes, gene products and biological processes
- Clinical and research procedures, tools, trials, management and more
- Animals models

In addition, the NCI Thesaurus contains terminology contributed by FDA, CDISC and other collaborators.

The NCI Thesaurus is published monthly, with pre-release and release candidate versions available for continuous test cycles. It is available via Web browsers, the caCORE API at http://ncicb.nci.nih.gov/NCICB/infrastructure, downloadable files in OWL, Ontylog XML and flat-file formats. The NCI Thesaurus is also freely available under open content licenses.

- **NCI Thesaurus Downloads:** The content of the NCI Thesaurus is available for download at http://ncicb.nci.nih.gov/download/ as compressed (Zipped) files in ASCII flat-file format, in Ontology Web Language format and in Apelon's proprietary Ontylog XML format. The current content of the NCI Thesaurus is also always available as uncompressed OWL at http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl.

  Note: This OWL file is about 65MB; and takes a while to load even over a fast connection.

The NCI Thesaurus is provided under an open content license. The license and a description of the semantics of the NCI Thesaurus are available at the download site.

- **NCI Thesaurus Website:** Web browsing of the NCI Thesaurus and several other public terminologies is provided at the NCI Terminology Browser at http://nciterms.nci.nih.gov/ NCIBrowser/Dictionary.do. To access the NCI Thesaurus, select the **NCI Thesaurus** radio button and then click the **Connect** button located at the bottom of the page.

Mapping with the NCI Metathesaurus

The NCI Metathesaurus can be used to map individual terms and perform bulk mapping exercises.

Mapping can be carried out using the RRF data itself, the RRF data converted to Oracle load files, the caCORE APIs or even the NCI Metathesaurus interactive Web browser.

For example, assume you have the SNOMED code 9861002 and wish to find the mapping to a corresponding code from another terminology. Using the NCI Metathesaurus RRF files (see *Appendix C*), you could proceed as follows.

1. Search for 9861002 in the CODE column of the MRCONSO.RRF table to identify the concept (CUI) that subsumes the SNOMED code 9861002.
2. Then search the table for the row that contains that concept (CUI) and the abbreviated source name (SAB) of the terminology to which you wish to map 9861002, say for example, MeSH (MSH).
3. Finally, read the CODE column value from that row, which would be the MeSH code D013296.

In addition to using the RRF flat files, if you wish to run SQL queries, when you run Metamorphosys, you have the option to create an Oracle Load Script which populates tables. You edit the script for your machine setup, and so forth, then you can run SQL queries.

The RRF files themselves are complex, and require database resources and programming to be useful. For those who wish to use NCI Metathesaurus to perform mapping but who do not want to invest in these resources, NCI provides alternatives.

Individual or small sets of codes can be conveniently mapped using the interactive Web browser found at http://ncimeta.nci.nih.gov. Use the **Advanced Search** tab and the **Code** search option following the instructions in the Users Guide: http://ncimeta.nci.nih.gov/Metaphrase/NCIMetaphraseUserGuideV1.02Final.pdf.

Larger mapping jobs can be conveniently performed using the NCI caCORE APIs. See the caCORE Technical Guide ftp:/ftp1.nci.nih.gov/pub/cacore/caCORE3.1 Tech Guide.pdf for details.

Finally, NCI EVS will perform mapping jobs for NCI Metathesaurus customers upon request, subject to availability of resources. In addition, for large mapping jobs Apelon produces a web services product called TermWorks that can be used to map codes in two different UMLS or NCI Metathesaurus vocabularies. Based on its DTS terminology server, it can provide an Excel-like interface that can be pre-populated with specific vocabularies and configured to use the UMLS synonymy to facilitate matching. See: http://apelon.com/products/termworks.htm.

Note: For assistance, telephone NCICB Application Support at 301-451-4384 or 888-478-4423 (toll-free) or email ncicb@pop.nci.nih.gov.

Working with RRF Files

Rich Release Format (RRF) files are the release mechanism for the NCI Metathesaurus and the UMLS Metathesaurus. RRF files enable the extraction of source vocabularies and facilitate mapping across different vocabularies. The content of UMLS and NCI RRF files are different from each other but the actual field structures and tables are the same.

- RRF files are distributed from the Downloads area located here: http://ncicb.nci.nih.gov/download/

NCI uses the MetamorphoSys/RRF package for downloading Metathesaurus files. MetamorphoSys is also used to install and customize applications distributed with the UMLS Metathesaurus. NCI updates the MetamorphoSys configuration files for the NCI Metathesaurus monthly on the FTP Downloads site. These files contain:

- A full distribution of MetamorphoSys that will be updated periodically (approximately annually)
- http://www.nlm.nih.gov/research/umls/documentation.html
- A second partial distribution of MetamorphoSys that is applicable to each monthly RRF release

MetamorphoSys can be used to generate a subset of the NCI RRF distribution for a user's local system, for example, in the case where the user is interested only in SNOMED, NCI Thesaurus, and MedDRA rather than all 50+ vocabularies.

Note: See *Appendix C* for a list of tables that are included in RRF downloads. General documentation on Metamorphosys may be found at http://www.nlm.nih.gov/research/umls/documentation.html.

Once users have initially downloaded the full MetamorphoSys distribution, they can simply overwrite the appropriate files in MetamorphoSys with the files that come in the "partial" MetamorphoSys distribution matching that RRF release to get a working copy of MetamorphoSys for a specific RRF release.

Note: For assistance with RRF files, telephone NCICB Application Support at 301-451-4384 or 888-478-4423 (toll-free) or email ncicb@pop.nci.nih.gov.

Summary

The NCI Metathesaurus provides users with a central uniform access point for relevant biomedical terminologies from basic, translational, clinical and population-based terminology domains thus facilitating research with the most appropriate cancer terminology. The NCI Metathesaurus also facilitates users' collaboration and sharing of data with colleagues in other disciplines by resolving for them which terms are equivalent among the various sources. The NCI Metathesaurus promises to be an increasingly valuable tool as the content evolves to meet the changing needs of the cancer research community.

APPENDIX A
MRCUI FILE

The chart below lists the columns and their descriptions that are included in the MRCUI file sample that follows.

Column	Description
CUI1	Unique identifier for first concept - Retired CUI - was present in some prior release, but is currently missing
VER	The last release version in which CUI1 was a valid CUI
REL	Relationship
RELA	Relationship attribute
MAPREASON	Reason for mapping
CUI2	Unique identifier for second concept - the current CUI that CUI1 most closely maps to
MAPIN	Is this map in current subset? Values of Y, N, or null. MetamorphoSys generates the Y or N to indicate whether the CUI2 concept is or is not present in the subset. The null value is for rows where the CUI1 was not present to begin with (i.e., REL=DEL)

Column headings: CUI1|VER|REL|RELA|MAPREASON|CUI2|MAPIN|
Data:

C0001899	200509	SY			C0376147	Y
C0002346	200509	SY			C1148475	Y
C0002590	200509	SY			C0002591	Y
C0002752	200509	SY			C0009010	Y
C0003234	200509	SY			C0282564	Y
C0003731	200509	SY			C0042172	Y
C0003830	200509	SY			C0701016	Y
C0004906	200509	SY			C0004905	Y
C0005763	200509	SY			C0040233	Y
C0006710	200509	SY			C0030342	Y
C0008429	200505	SY			C0728810	Y
C0008846	200509	SY			C0591261	Y
C0008877	200509	SY			C0031955	Y
C0008994	200509	SY			C0162357	Y
C0009951	200505	SY			C0036572	Y
C0010285	200509	SY			C0700579	Y
C0011155	200505	SY			C0442798	Y
C0012792	200509	SY			C0700978	Y
C0013360	200509	SY			C0058180	Y
C0013881	200509	SY			C0700528	Y
C0015845	200509	SY			C0546864	Y
C0017449	200509	SY			C0699935	Y
C0022613	200509	SY			C0700541	Y
C0024080	200509	SY			C0700455	Y
C0024453	200509	SY			C0809894	Y
C0025573	200509	SY			C0591074	Y
C0027673	200505	SY			C0008267	Y
C0028221	200509	SY			C0939465	Y
C0029275	200509	SY			C0002260	Y
C0030311	200509	SY			C0030715	Y
C0031008	200509	SY			C0701028	Y
C0031270	200509	SY			C0546862	Y
C0031465	200509	SY			C0281398	Y
C0032918	200509	SY			C0043440	Y
C0033390	200509	SY			C0700567	Y
C0035750	200509	SY			C0007561	Y
C0037523	200509	SY			C0733804	Y
C0038155	200509	SY			C0887175	Y
C0039184	200509	SY			C0733367	Y
C0039543	200509	SY			C0591192	Y
C0039654	200509	SY			C0879362	Y
C0042049	200509	SY			C0546865	Y
C0043251	200509	SY			C0043250	Y
C0043443	200509	SY			C0878100	Y
C0047995	200509	SY			C0700482	Y
C0050156	200509	SY			C1257880	Y
C0050163	200509	SY			C0017066	Y
C0050910	200509	SY			C0093437	Y
C0052511	200509	SY			C0700606	Y
C0053604	200509	SY			C0209211	Y

APPENDIX

B

REPRESENTATIVE COMPARISON OF SOURCES

This appendix provides a representative comparison of UMLS and NCI sources.

Since UMLS and NCI sources are dynamic, the sources listed in Appendix B will not necessarily be current. To see an up-to-date listing of sources, refer to the following links:

- To access the current list of sources for the NCI Metathesaurus, go to http://ncimeta.nci.nih.gov and click the **Sources** link.
- To access the current list of sources for the UMLS Metathesaurus, go to http://www.nlm.nih.gov/research/umls/ and click the **List of Sources** link.

Source	*UMLS*	*NCI*	*NCI Local*
AIR - Published by the National Library of Medicine, AI/RHEUM is an experimental expert system for support in the diagnosis of rheumatic diseases.	X		
ALT - Published by ABC Coding Solutions (formerly Alternative Link LLC), the Alternative Billing Concepts (Altlink) covers ABC codes and related solutions that allow more than 3 million healthcare practitioners to file electronic claims for healthcare services that are not adequately described in other national code sets.	X		
AOD - Published by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and Center for Substance Abuse Prevention (CSAP), the Alcohol and Other Drug Thesaurus is a guide to concepts and terminology in substance abuse and addiction.	X	X	

Source	UMLS	NCI	NCI Local
BI - Published by the Center for Clinical Computing, Beth Israel Deaconess Medical Center, Harvard University, the Beth Israel OMR Clinical Problem List Vocabulary.	X		
BioC - Published by BioCarta, the BioC is based on BioCarta online maps of molecular relationships adapted for NCI use.			X
CBO - Published by Cerner Corporation, the Clinical Bioinformatics Ontology is a curated semantic network utilizing RefSeq information (part of the NCBI collection of bioinformatics resources) that describes clinically significant genomics concepts, including concepts appropriate for both molecular diagnostics and cytogenetics.			X
CCPSS - Published by the Eskind Biomedical Library, Vanderbilt University Medical Center, the Canonical Clinical Problem Statement System.	X	X	
CCS - Published by the Agency for Healthcare Research and Quality (AHRQ), Clinical Classifications Software is a diagnosis and procedure categorization scheme that can be employed in many types of projects analyzing data on diagnoses and procedures. CCS is based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), a uniform and standardized coding system. The ICD-9-CM's multitude of codes-- over 12,000 diagnosis codes and 3,500 procedure codes--are collapsed into a smaller number of clinically meaningful categories that are sometimes more useful for presenting descriptive statistics than are individual ICD-9-CM codes. For example, CCS can be used to identify populations for disease- or procedure-specific studies or to develop statistical reports providing information (such as charges and length of stay) about relatively specific conditions. CCS was formerly called the Clinical Classifications for Health Policy Research (CCHPR).	X	X	
CDC - The Centers for Disease Control and Prevention is one of the 13 major operating components of the Department of Health and Human Services (HHS), which is the principal agency in the United States government for protecting the health and safety of all Americans and for providing essential human services. The CDC provides the CDC Ontology of Reportable Conditions and CDC WONDER (a system for disseminating public health data and information).			X

Source	*UMLS*	*NCI*	*NCI Local*
CDISC - Clinical Data Interchange Standards Consortium (CDISC) is an open, multidisciplinary, non-profit organization committed to the development of industry standards to support the electronic acquisition, exchange, submission and archiving of clinical trials data and metadata for medical and biopharmaceutical product development. The Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM) is an XML specification that is a vendor neutral, platform independent format for interchange and archive of data collected in clinical trials. The model represents study metadata, data and administrative data associated with a clinical trial. Only the information that needs to be shared among different software systems during a trial, or archived after a trial is included in the model.			X
CDT - Published by the American Dental Association, the Current Dental Terminology is a reference manual published by the American Dental Association that contains a number of useful components, including the Code on Dental Procedures and Nomenclature (Code), instructions for use of the Code, Questions and Answers, the ADA Dental Claim Form Completion Instructions, and Tooth Numbering Systems.	X		
COSTAR - Published by Massachusetts General Hospital, Computer-Stored Ambulatory Records. **Note:** The UMLS Metathesaurus includes terms that were used frequently at three COSTAR sites and supplied to NLM by Massachusetts General Hospital.	X	X	
CPM - Published by the Columbia Presbyterian Medical Center, the Medical Entities Dictionary is a large repository of medical concepts that are drawn from a variety of sources either developed or used at the New York Presbyterian Hospital, including the UMLS, ICD9-CM and LOINC. Currently numbering over 69,000, these concepts correspond to coded terms used in systems and applications throughout both medical centers (Columbia-Presbyterian and New York-Cornell). It continues to grow at about 6,000 terms per year, although accelerated growth is anticipated as additional network hospitals are integrated into the NYPH system. **Note:** The UMLS Metathesaurus includes a relatively small number of terms created at Columbia Presbyterian Medical Center for the MED, which also includes terms obtained from the UMLS Metathesaurus and other sources.	X		
CPT - Published by the American Medical Association, the Current Procedural Terminology (CPT) provides guidelines and coding rules necessary to correctly interpret and report medical procedures and services.	X		
CPT01SP - Published by the American Medical Association, the Spanish translation of the Current Procedural Terminology (CPT).	X		

Source	UMLS	NCI	NCI Local
CSP - Published by the National Institutes of Health, Division of Research Grants, Research Documentation Section, the Computer Retrieval of Information on Scientific Projects (CRISP) Thesaurus is a controlled vocabulary used to assign indexing terms or keywords to research projects.	X	X	
CST - Published by U.S. Food and Drug Administration, Center for Drug Evaluation and Research, CST is the Coding Symbols for Thesaurus of Adverse Reaction Terms (COSTART). **Note:** COSTART has been superseded by the Medical Dictionary for Regulatory Activities (MedDRA) Terminology.	X	X	
CTCAE - Published by Cancer Therapy Evaluation Program (CTEP), National Cancer Institute, the Common Terminology Criteria for Adverse Events is a descriptive terminology which can be utilized for Adverse (AE) reporting. A grading (severity) scale is provided for each AE term.	X		X
CTEP - Published by the National Cancer Institute, the Cancer Therapy Evaluation Program provides support for clinical trials.			X
DCP - The Division of Cancer Prevention is the primary unit of the National Cancer Institute devoted to cancer prevention research. DCP provides the Dictionary of Cancer Terms (cancer-related terms) and the NCI Drug Dictionary (definitions, names, and links).			X
DDB - Published by the Medical Object Oriented Software Enterprises Ltd., the Diseases Database is a cross-referenced index of human disease, medications, symptoms, abnormal investigation findings, and so on, including general internal medical disorders, symptoms and signs, congenital and inherited disorders, infectious diseases and organisms, drugs and medications and common haematology and biochemistry investigation abnormalities.	X		
DMDICD10 - Published by the Deutsches Institut fuer Medizinische Dokumentation und Information, the German translation of ICD10.	X		
DMDUMD - Published by the Deutsches Institut fuer Medizinische Dokumentation und Information, the German translation of UMDNS.	X		
DSM - Published by the American Psychiatric Association, the Diagnostic and Statistical Manual of Mental Disorders covers diagnostic criteria for the most common mental disorders including: description, diagnosis, treatment, and research findings.	X	X	

Source	UMLS	NCI	NCI Local
DTP - The National Cancer Institute, NCI Developmental Therapeutics Program (the drug discovery and development arm of NCI which plans, conducts, and facilitates development of therapeutic agents for cancer and AIDS) provides Web-accessible data sets and data mining tools from in-house screening programs and extramural collaborations. Bulk data available for download includes human tumor cell line assay, AIDS antiviral assay data, yeast assay data, chemical structures, characterizations of molecular targets in cell lines, data from the Botstein/Brown Stanford U. Microarray Project. Tools for searching and analyzing data include searches by chemical name, NSC number, substructure, and COMPARE searches.			X
DXP - Published by Massachusetts General Hospital, the DXplain program is a decision support system that has the characteristics of both an electronic medical textbook and a medical reference system.	X		
ELC - Published by the National Center for Biotechnology Information, the Expression Library Classification covers tissue classifications in genomics and proteomics.			X
FDA - The U.S. Food and Drug Administration is the government agency responsible for regulating food (human and animal), dietary supplements, drugs (human and animal), cosmetics, medical devices (human and animal), biologics, and blood products in the United States. The FDA is divided into five major Centers: the Center for Biologics Evaluation and Research (CBER), the Center for Devices and Radiological Health (CDRH), the Center for Drug Evaluation and Research (CDER), the Center for Food Safety and Applied Nutrition (CFSAN) and the Center for Veterinary Medicine (CVM). Each center has its own origins and history, the oldest being CBER, which oversees blood products, vaccines, and newer therapeutics related to stem cells and gene therapy. CDER, which regulates human pharmaceuticals, receives considerable public scrutiny.			X
GO - Published by the Gene Ontology Consortium, the Gene Ontology project is a collaborative effort to address the need for consistent descriptions of gene products in different databases. The project began as a collaboration between three model organism databases: FlyBase (Drosophila),the Saccharomyces Genome Database (SGD) and the Mouse Genome Database (MGD) in 1998. Since then, the GO Consortium has grown to include many databases, including several of the world's major repositories for plant, animal and microbial genomes.	X	X	
HCDT - Published by the U.S. Centers for Medicare & Medicaid, the HCPCS Version of the Current Dental Terminology (CDT).	X		
HCPCS - Published by the U.S. Centers for Medicare & Medicaid Services, the Healthcare Common Procedure Coding System.	X	X	

Source	UMLS	NCI	NCI Local
HCPT - Published by the U.S. Centers for Medicare & Medicaid Services, the version of Physicians' Current Procedural Terminology (CPT) included in the Healthcare Common Procedure Coding System (HCPCS).	X		
HHC - Published by Georgetown University, the Home Health Care Classification covers nursing diagnoses and interventions.	X	X	
HL7 - Health Level Seven is one of several American National Standards Institute (ANSI)-accredited Standards Developing Organizations (SDOs) operating in the healthcare arena. Health Level Seven's domain is clinical and administrative data. The Reference Information Model (RIM) is the cornerstone of the HL7 Version 3 development process. An object model created as part of the Version 3 methodology, the RIM is a large pictorial representation of the clinical data (domains) and identifies the life cycle of events that a message or groups of related messages will carry.	X	X	
HLREL - Published by the University of Amsterdam, HLREL covers the ICPC2E-ICD10 relationships from Dr. Henk Lamberts (HLREL).	X	X	
HUGO - Published by the HUGO Gene Nomenclature Committee, Department of Biology, University College London, the HGNC database, previously known as Genew, contains over 22,000 public records with approved human gene nomenclature and associated information.	X	X	
ICD10 - Published by the World Health Organization, the International Statistical Classification of Diseases and Related Health Problems.	X	X	
ICD10AE - Published by the World Health Organization, the ICD-10, Americanized Version.	X		
ICD10AM - Published by the National Centre for Classification in Health, University of Sydney, Faculty of Health Sciences, the ICD-10,Australian Modification version.	X		
ICD10AMAE - Published by the National Library of Medicine, UMLS project, the Australian Modification (ICD-10-AM), Americanized English Equivalents.	X		
ICD10DUT - Published by the Department of General Practice, Academic Medical Center/University of Amsterdam, Dutch College of General Practitioners (NHG), the ICD-10, Dutch Translation.	X		

Source	*UMLS*	*NCI*	*NCI Local*
ICD9CM - Published by the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services, the ICD-9-CM [computer file] provides an international classification of diseases, clinical modification. **Note:** the National Library of Medicine has generated fully specified titles for ICD-9-CM codes in cases in which the official ICD- 9-CM titles consist of extensions to higher levels in the ICD-9-CM hierarchy. The fully specified names were produced with reasonable care, but have not yet been reviewed and approved by the producers of ICD-9-CM.	X	X	
ICDO3 - Published by the International Association of Cancer Registries, the International Classification of Diseases for Oncology (ICD) is used for cancer genome activities and incidence reporting.			X
ICPC - Published by the World Organisation of Family Doctors, the International Classification of Primary Care (ICPC) includes the following languages: Basque (ICPCBAQ), Danish (ICPCDAN), Dutch (ICPCDUT), Finnish (ICPCFIN), French (ICPCFRE), German (ICPCGER), Hebrew (ICPCHEB), Hungarian (ICPCHUN), Italian (ICPCITA), Norwegian (ICPCNOR), Portuguese (ICPCPOR), Spanish (ICPCSPA), and Swedish (ICPCSWE).	X	X	
ICPC2ICD10ENG - Published by the International Classification of Primary Care/prepared by the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors. The ICPC2 - ICD10 Thesaurus.		X	
JABL - Published by the National Library of Medicine, JABL provides Online Congenital Multiple Anomaly/Mental Retardation Syndromes.	X	X	
JAX - JAX is an NCI-modified picklist used by the Jackson Laboratories Tumor Biology Database. It is also used by NCI's Cancer Models Database.			X
KEGG - Published by Kanehisa Laboratories, the KEGG Pathway Database is a collection of graphical pathway maps, ortholog group tables and molecular catalogs.			X
LCH - Published by the Library of Congress, the Library of Congress Subject Headings is a controlled vocabulary for subject indexing of the Library of Congress collection and can be used in subject cataloging and indexing by other libraries or indexing agencies. **Note:** There are later editions of this source that are not reflected in the UMLS Metathesaurus. This source has considerable non-biomedical content and will never be included in the Metathesaurus in its entirety.	X		
LNC - Published by The Regenstrief Institute, the Logical Observation Identifier Names and Codes (LOINC) covers clinical laboratory terminology.	X	X	

Source	UMLS	NCI	NCI Local
MBD - Published by the National Library of Medicine, MBD provides MEDLINE backfiles from 1996 to 2000.	X	X	
MCM - Published by the McMaster University, MCM provides the Glossary of Methodologic Terms for Clinical Epidemiologic Studies of Human Disorders.	X	X	
MDBCAC - Published by the Cancer Genome Anatomy Project, the Mitelman Database of Chromosome Aberrations in Cancer relates chromosomal aberrations to tumor characteristics, based either on individual cases or associations. The Cancer Genome Anatomy Project has developed 5 tools to help with analysis of information in this database.			X
MDDB - Published by Medi-Span, the Master Drug Data Base covers comprehensive National Drug Codes (NDCs), drug names, ingredients, and pricing information.	X		
MDR - Published by Northrop Grumman, MedDRA MSSO. The Medical Dictionary for Regulatory Activities (MedDRA) terminology covers all phases of drug development, excluding animal toxicology. It also applies to the health effects and malfunction of devices.	X		X
MDRAE - Published by Northrop Grumman, Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English Equivalents.			X
MDRDUT - Published by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the Medical Dictionary for Regulatory Activities Terminology (MedDRA) is the Dutch edition of MedDRA.	X		
MDREA - Published by Northrop Grumman, the Medical Dictionary for Regulatory Activities Terminology (MedDRA), American English Equivalents with expanded abbreviations.			X
MDREX - Published by Northrop Grumman, the Medical Dictionary for Regulatory Activities Terminology (MedDRA), with expanded abbreviations.			X
MDRFRE - Published by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the Medical Dictionary for Regulatory Activities Terminology (MedDRA), French edition.	X		
MDRGER - Published by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the Medical Dictionary for Regulatory Activities Terminology (MedDRA), German edition.	X		
MDRPOR - Published by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), the Medical Dictionary for Regulatory Activities Terminology (MedDRA), Portuguese edition.	X		

Source	*UMLS*	*NCI*	*NCI Local*
MED - Published by the National Library of Medicine, MED covers the current MEDLINE files, an online database of more than 11 million citations and abstracts from health and medical journals and other news sources.	X	X	
MEDLINEPLUS - Published by the National Library of Medicine, MEDLINEPLUS covers MedlinePlus Health Topics.	X	X	
MIM - Published by Johns Hopkins University, Center for Biotechnology Information, MIM provides the Online Mendelian Inheritance in Man (OMIM). **Note:** To date the UMLS Metathesaurus contains a relatively small amount of data from this source.	X	X	
MMSL - Published by Multum Information Services, Inc., the Medisource Lexicon provides drug names, drug product information, disease names and coding systems.	X	X	
MMX - Published by Micromedex, the Micromedex DRUGDEX covers FDA-approved and investigational prescription and nonprescription drugs, as well as non-U.S. preparations. Areas discussed include dosage, pharmacokinetics, cautions, interactions, clinical applications, adverse effects, comparative efficacy, drug of choice information, and orphan drug status.	X		
MSH - Published by the National Library of Medicine, the Medical Subject Headings (MeSH) is the NLM's controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. **Note:** This source has been translated into many languages. To date, eight of the translations have been incorporated into the UMLS Metathesaurus.	X	X	
MSHCZE - Published by the National Library of Medicine, the Czech translation of the Medical Subject Headings (MeSH).	X	X	
MSHDUT - Published by the Nederlands Tijdschrift voor Geneeskunde [Dutch Journal of Medicine], the Dutch translation of the Medical Subject Headings (MeSH).	X	X	
MSHFIN - Published by the Finnish Medical Society Duodecim, the Finnish translations of Medical Subject Headings (MeSH).	X	X	
MSHFRE - Published by the Institut National de la Sante et Recherche Medicale, the Thesaurus Biomedical Francais/Anglais is the French translation of MeSH.	X	X	
MSHGER - Published by the Deutsches Institut fur Medizinische Dokumentation und Information, the German translation of Medical Subject Headings (MeSH).	X	X	
MSHITA - Published by the Istituto Superiore di Sanita, Settore Documentazione, the Italian translation of Medical Subject Headings (MeSH).	X	X	
MSHJPN - Published by the Japan Medical Abstracts Society, Igaku-Chuo-Zasshi, the JAMAS Japanese Medical Thesaurus (JJMT).	X	X	

Source	UMLS	NCI	NCI Local
MSHPOR - Published by the Latin American and Caribbean Center on Health Sciences Information, the Descritores em Ciencias da Saude is the Portuguese translation of Medical Subject Headings (MeSH).	X	X	
MSHRUS - Published by the State Central Scientific Medical Library, the Russian Translation of Medical Subject Headings (MeSH).	X	X	
MSHSPA - Published by the Latin American and Caribbean Center on Health Sciences Information, the Descriptores en Ciencias de la Salud is the Spanish translation of Medical Subject Headings (MeSH).	X	X	
MSHSWE - Published by the Karolinska Institutet, the Swedish translation of Medical Subject Headings (MeSH).	X	X	
MTH - Published by the National Library of Medicine, concept names with this source abbreviation were created by NLM to facilitate creation of the UMLS Metathesaurus. There are relatively few of them.	X	X	
MTHCH - Published by the National Library of Medicine, Metathesaurus Hierarchical CPT Terms were created by the NLM to provide contextual information for CPT.	X		
MTHFDA - Published by the National Library of Medicine, the Metathesaurus Forms of FDA National Drug Code Directory comprises concept names created by NLM to provide contextual information for FDA NDC terms.	X	X	
MTHHH - Published by the National Library of Medicine, the Metathesaurus Hierarchical HCPCS Terms comprises concept names created by NLM to provide contextual information for HCPCS.	X		
MTHHL7 - Published by the National Library of Medicine, the HL7 Vocabulary comprises 7-bit equivalents created by the National Library of Medicine.	X		
MTHICD9 - Published by the National Library of Medicine, MTHICD provides Metathesaurus additional entry terms for ICD-9-CM.	X	X	
MTHICPC2EAE - Published by the International Classification of Primary Care/prepared by the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors, the International Classification of Primary Care provides American English Equivalents. Concept names with this source abbreviation were created by NLM to provide contextual information for ICPC2E terms.	X		

Source	UMLS	NCI	NCI Local
MTHICPC2ICD107B - Published by the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors, the International Classification of Primary Care provides 7-bit Equivalents.	X	X	
MTHICPC2ICD10AE - Published by the Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA), known more briefly as the World Organization of Family Doctors, the International Classification of Primary Care provides American English Equivalents.	X	X	
MTHMDRSPA71 - Published by the National Library of Medicine, the Methathesaurus Forms of Medical Dictionary for Regulatory Activities Terminology (MedDRA), is the Spanish Edition.	X		
MTHMST - Published by the National Library of Medicine, the Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy.	X	X	
MTHMSTFRE - Published by the National Library of Medicine, the Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy, French Translation.	X		
MTHMSTITA - Published by the National Library of Medicine, the Metathesaurus Version of Minimal Standard Terminology Digestive Endoscopy, Italian Translation.	X		
MTHPDQ - Published by the National Cancer Institute, the Methathesaurus Forms of Physician Data Query (PDQ) provides terms created by the National Library of Medicine to provide contextual information for PDQ terms.	X		
MTHSCT - Published by the National Library of Medicine, the Metathesaurus forms of SNOMED Clinical Terms.	X	X	
NAN - Published by NANDA International, NAN provides NANDA nursing diagnoses, definitions and classifications.	X		
NCBI - Published by the U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, the NCBI Taxonomy database contains the names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence.	X	X	
NCI - Published by the National Cancer Institute, National Institutes of Health, the NCI Thesaurus knowledgebase contains the working vocabulary used in NCI data systems. It covers clinical, translational and basic research as well as administrative terminology. (Subset only.)	X		X
NCI-CTCAE - Published by the National Cancer Institute, the NCI modified Common Terminology Criteria for Adverse Events (CTCAE).	X	X	

Source	UMLS	NCI	NCI Local
NCI-GLOSS - Published by the National Cancer Institute, NCI-GLOSS is the Cancer.gov Dictionary.			X
NCISEER - Published by the National Cancer Institute, the NCI Surveillance, Epidemiology, and End Results (SEER) provides conversions between ICD-9-CM and ICD-10 neoplasm codes.	X	X	
NDDF - Published by First DataBank, the National Drug Data File Plus source vocabulary combines granular descriptive and pricing information with a selection of advanced clinical decision-support modules. These modules are used with a structured vocabulary for drugs and medical conditions that can be linked to external terminologies, thus facilitating application development.	X		
NDFRT - Published by the U.S. Department of Veterans Affairs, Veterans Health Administration, the National Drug File - Reference Terminology is a description logic-based resource created to support clinical operations at one of the largest healthcare providers in the U.S.	X	X	
NEU - Published by the University of Washington, Primate Information Center, the Neuronames Brain Hierarchy is a structured system of neuroanatomical terminology that provides a comprehensive representation of virtually all human and non-human primate brain structures that are identifiable either grossly or in Niss1-stained histological sections. This system was devised for computer applications to address deficiencies in the brain terminology presented in Nomina Anatomica. English terms are listed for 783 structures in nine levels of hierarchical ranking. Abbreviations are provided for all superficial and primary volumetric structures.	X		
NIC - Published by Mosby-Year Book, the Nursing Interventions Classification:Iowa Intervention Project is a comprehensive, research-based, standardized classification of interventions that nurses perform. It is useful for clinical documentation, communication of care across settings, integration of data across systems and settings, effectiveness research, productivity measurement, competency evaluation, reimbursement, and curricular design.	X		
NLM-MED - Published by the National Library of Medicine, NLM-MED provides NLM Medline data.	X	X	
NOC - Published by Mosby-Year Book, the Nursing Outcomes Classification is a comprehensive, standardized classification of patient/client outcomes developed to evaluate the effects of nursing interventions. Standardized outcomes are necessary for documentation in electronic records, for use in clinical information systems, for the development of nursing knowledge and the education of professional nurses.	X		

Source	UMLS	NCI	NCI Local
OMS - Published by W.B. Saunders, the Omaha System: Applications for Community Health Nursing is a standardized classification (taxonomy) recognized by the American Nurses Association. It is a comprehensive practice and documentation tool that can be used by multidisciplinary health care practitioners in any setting from the time of client admission to discharge. The three components of the Omaha System include an assessment component (Problem Classification Scheme), an intervention component (Intervention Scheme), and an outcomes component (Problem Rating Scale for Outcomes).	X	X	
PCDS - Published by Vanderbilt University School of Nursing, the Patient Data Care Set.	X		
PDQ - Published by the National Cancer Institute, the Physician Data Query terminology is a source of clinical cancer disease and drug terminology.	X		X
PMA - Published by the National Cancer Institute (NCI) Division of Cancer Control and Population Studies (DCCPS), the Portfolio Management Application is a DCCPS Grants Portfolio Management application.			X
PNDS - Published by AORN, the Perioperative Nursing Data Set covers perioperative nursing vocabulary.	X		
PPAC - Published by the American Pharmaceutical Association, the Pharmacy Practice Activity Classification taxonomy organizes pharmacy activity classifications.	X		
PSY - Published by the American Psychological Association, PSY is a thesaurus of psychological index terms.	X		
QMR - Published by First Databank, the Quick Medical Reference is a diagnostic decision-support system for internists.	X	X	
RAM - Published by Vanderbilt University, RAM covers QMR clinically related terms from Dr. Randolf A. Miller, Dept. of Biomedical Informatics, Vanderbilt University.	X	X	
RCD - Published by the National Health Service National Coding and Classification Centre, the Clinical Terms Version 3 (CTV3) (Read Codes) comprises a set of files containing Read coded clinical concepts (and their representative clinical terms) in a hierarchical relationship, together with associated cross-references to the clinical classifications. Additional information relating to the concepts and to the terms is also included. When implemented in a software system, the files can be used for the storage, retrieval, cross mapping and analysis of patient information.	X		
RCDAE - Published by the National Library of Medicine, UMLS project, the American English equivalent of the Clinical Terms Version 3 (CTV3).	X		
RCDSA - Published by the National Library of Medicine, UMLS project, the American English equivalent of synthesized terms from the Clinical Terms Version 3 (CTV3).	X		

Source	UMLS	NCI	NCI Local
RCDSY - Published by the National Library of Medicine, UMLS project, RCDSY covers Synthesized Read terms (without initial bracketed letters) of the Clinical Terms Version 3 (CTV3).	X		
RXNORM - Published by the National Library of Medicine, RXNORM contains concepts created by the National Library of Medicine which express the meaning of a drug name in a normalized form. These concepts relate the names of orderable medications to a dose form and the components of those medications.	X	X	
SCTSPA - Published by the College of American Pathologists, SNOMED Clinical Terms, Spanish Language Edition. SNOMED CT has been created by combining SNOMED RT and a computer based nomenclature and classification known as Clinical Terms Version 3 (CTV3), formerly known as Read Codes, Version 3, which was created on behalf of the U.K. Department of Health.	X		
SNM2 - Published by the College of American Pathologists, the Systematized Nomenclature of Medicine (SNOMED) is a multiaxial, hierarchical classification system which allows for the coding of several aspects of a disease. A diagnosis in SNOMED may consist of a topographic code, a morphology code, a living organism code, and a function code. When a well-defined diagnosis for a combination of these four codes exists, a dedicated diagnostic code is defined.	X		
SNMI - Published by the American Veterinary Medical Association, SNOMED International, the Systematized Nomenclature of Human and Veterinary Medicine, provides information about SNOMED's modules, updates, conversion tables, microglossaries, translations, costs and licensing, and who is using SNOMED.	X		
SNOMEDCT - Published by the College of American Pathologists, Systematized Nomenclature of Medicine Clinical Terms consists of SNOMED CT content loaded by Apelon from files received directly from the College of American Pathologists (CAP). SNOMED CT was created by combining SNOMED RT and a computer based nomenclature and classification known as Clinical Terms Version 3 (CTV3), formerly known as Read Codes, Version 3, which was created on behalf of the U.K. Department of Health.	X		X
SPN - Published by the U.S. Food and Drug Administration, the Standard Product Nomenclature was developed by the FDA to carry out its regulatory responsibilities for medical devices. This standard product nomenclature is used by the industry in identifying devices for which premarket submissions are being made and for which postmarket reporting is required. Various nongovernmental organizations and the Canadian government have adopted the FDA standard product nomenclature for use in their device activities.	X	X	

Source	UMLS	NCI	NCI Local
SRC - Published by the National Library of Medicine, the UMLS Metathesaurus Source Terminologies.	X	X	
ULT - Published by the Brigham & Womens Hospital, Ultrasound Structured Attribute Reporting (UltraSTAR) allows a sonographer to record reports using a hierarchic standardized vocabulary. Reports are constructed interactively by using checkboxes and radio buttons in a graphical user interface to select concepts from the vocabulary. Concepts are presented in small windows that open when needed to record further detail. Preliminary reports are printed for the referring physician. Reports are later finalized by an attending radiologist and uploaded to the hospital's information system.	X		
UMD - Published by ECRI, the Universal Medical Device Nomenclature System (UMDNS) is a standard international nomenclature and computer coding system for medical devices. UMDNS facilitates identifying, processing, filing, storing, retrieving, transferring, and communicating data about medical devices. The nomenclature is used in applications ranging from hospital inventory and work-order controls to national agency medical device regulatory systems and from e-commerce and procurement to medical device databases.	X	X	
USPMG - Published by the United States Pharmacopeia (USP), the Medicare Prescription Drug Benefit Model Guidelines covers drug categories and classes in Part D.	X		
UWDA173 - Published by the University of Washington, the University of Washington Digital Anatomist (also known as the Foundational Model of Anatomy) consists of the Anatomy taxonomy (At) and selected structural relationships (part-of, branch-of, tributary of).	X	X	
VANDF - Published by the U.S. Department of Veterans Affairs, the Veterans Health Administration National Drug File terminology represents dictated medication list phrases from the Mayo Clinic. NDFRT is a description logic-based resource created to support clinical operations.	X		
WHO - Published by the World Health Organization Collaborating Centre for International Drug Monitoring , the WHO Adverse Drug Reaction Terminology (WHOART). WHO is published in several languages: French (WHOFRE), German (WHOGER), Portuguese (WHOPOR) and Spanish (WHOSPA).	X		

APPENDIX

C

TABLES INCLUDED IN RRF DOWNLOADS

Table	*Table Description*	*Columns/Column Descriptions*
AMBIGLUI.RRF	Ambiguous Term Identifiers	LUI - Lexical Unique Identifier
		CUI - Concept Unique Identifier
AMBIGSUI.RRF	Ambiguous String Identifiers	SUI - String Unique Identifier
		CUI - Concept Unique Identifier
MRAUI.RRF	Unique Identifier for String	CUI1 - Concept unique identifier
		VER - Version in which this change to the AUI occurred
		REL - Relationship
		RELA - Relationship attribute
		MAPREASON - Reason for mapping
		AUI2 - Unique identifier for second atom
		CUI2 - Unique identifier for second concept
		MAPIN - Mapping in current subset

Table	*Table Description*	*Columns/Column Descriptions*
MRCOC.RRF	Co-occurring Concepts	CUI1 - Unique identifier of first concept
		AUI1 - Unique identifier of first atom
		CUI2 - Unique identifier of second concept or not present
		AUI2 - Unique identifier of second atom
		SAB - Abbreviation of the source of co-occurrence information
		COT - Type of co-occurrence
		COF - Frequency of co-occurrence, if applicable
		COA - Attributes of co-occurrence, if applicable
		CVF - Content View Flag
MRCOLS.RRF	Data Elements	COL - Column or data element name
		DES - Descriptive Name
		REF - Documentation Section Number
		MIN - Minimum Length, Characters
		AV - Average Length
		MAX - Maximum Length, Characters
		FIL - Physical FILENAME in which this field occurs
		DTY - SQL-92 data type for this column
MRCONSO.RRF	Concept Names and Sources	CUI - Unique identifier for concept
		LAT - Language of term
		TS - Term status
		LUI - Unique identifier for term
		STT - String type
		SUI - Unique identifier for string

Table	Table Description	Columns/Column Descriptions
		ISPREF - Atom status
		AUI - Unique identifier for atom
		SAUI - Source asserted atom identifier
		SCUI - Source asserted concept identifier
		SDUI - Source asserted descriptor identifier
		SAB - Abbreviated source name
		TTY - Abbreviation for term type in source vocabulary
		CODE - Most useful source asserted identifier
		STR - String
		SRL - Source restriction level
		SUPPRESS - Suppressible flag. Values = O (obsolete), E (non-obsolete marked suppressible by an editor), Y (non-obsolete marked suppressible during inversion), or N (none of the above).
		CVF - Content View Flag
MRCUI.RRF	Retired CUI Mappings	CUI1 - Unique identifier for first concept - Retired CUI - was present in some prior release, but is currently missing
		VER - The last release version in which CUI1 was a valid CUI
		REL - Relationship
		RELA - Relationship attribute
		MAPREASON - Reason for mapping
		CUI2 - Unique identifier for second concept - the current CUI that CUI1 most closely maps to

Table	Table Description	Columns/Column Descriptions
		MAPIN - Is this map in current subset? Values of Y, N, or null.
MRCXT.RRF	Concept contexts	CUI - Unique identifier of concept
		SUI - Unique identifier of string used in this context
		AUI - Unique identifier of atom that has this context
		SAB - Abbreviated source name
		CODE - Unique identifier or code for string in that source
		CXN - The context number (if the atom has multiple contexts)
		CXL - Context member label
		RNK - For rows with a CXL value of ANC, the rank of the ancestors
		CXS - String or concept name for context member
		CUI2 - Concept identifier of context member
		AUI2 - Atom identifier of context member
		HCD - Source hierarchical number or code of context member (if present)
		RELA - Additional relationship label providing further categorization of the CXL, if applicable and known.
		XC - A plus (+) sign indicates that the CUI2 for this row has children in this context. If this field is empty, the CUI2 does not have children in this context.
		CVF - Content View Flag
MRDEF.RRF	Definitions	CUI - Unique identifier for concept

Table	Table Description	Columns/Column Descriptions
		AUI - Unique identifier for atom
		ATUI - Unique identifier for attribute
		SATUI - Source asserted attribute identifier
		SAB - Abbreviated source name of the source of the definition
		DEF - Definition
		SUPPRESS - Suppressible flag. Values = O (obsolete), E (non-obsolete marked suppressible by an editor), Y (non-obsolete marked suppressible during inversion), or N (none of the above).
		CVF - Content View Flag
MRDOC.RRF	Content Views	KEY - Data element or attribute
		VALUE - Abbreviation that is one of its values
		TYPE - Type of information in EXPL column
		EXPL - Explanation of VALUE
MRFILES.RRF	Metathesaurus Files	FIL - Physical FILENAME
		DES - Descriptive Name
		FMT - Comma separated list of column names
		CLS - # of COLUMNS
		RWS - # of ROWS
		BTS - Size in bytes in this format
MRHIER.RRF	Representation of Hierarchies	CUI - Unique identifier of concept
		AUI - Unique identifier of atom
		CXN - Context number
		PAUI - Unique identifier of atom's immediate parent within this context

Table	Table Description	Columns/Column Descriptions
		SAB - Abbreviated source name of the source of atom
		RELA - Relationship of atom to its immediate parent
		PTR - Path to the top or root of the hierarchical context from this atom, represented as a list of AUIs, separated by periods (.)
		HCD - Source asserted hierarchical number or code for this atom in this context; this field is only populated when it is different from the code (unique identifier or code for the string in that source).
		CVF - Content View Flag
MRHIST.RRF	History	CUI - Unique identifier for concept
		SOURCEUI - Source asserted unique identifier
		SAB - Abbreviated source name
		SVER - Release date or version number of a source
		CHANGETYPE - Source asserted code for type of change
		CHANGEKEY - CONCEPTSTATUS (if history relates to a SNOMED CT concept) or DESCRIPTIONSTATUS (if history relates to a SNOMED CT atom)
		CHANGEVAL - CONCEPTSTATUS value or DESCRIPTIONSTATUS value after the change took place.
		REASON - Explanation of change if present
		CVF - Content View Flag
MRMAP.RRF	Mappings	MAPSETCUI - Unique identifier for the map set concept to which this mapping belongs

Table	Table Description	Columns/Column Descriptions
		MAPSETSAB - Source abbreviation (SAB) for the provider of the map set
		MAPSUBSETID - Map sub-set identifier used to group together related mappings within a map set
		MAPRANK - Order in which mappings in a subset should be applied
		MAPID - Unique identifier for this mapping
		MAPSID - Source asserted identifier for this mapping
		FROMID - Identifier for the entity being mapped from
		FROMSID - Source asserted identifier for the entity being mapped from
		FROMEXPR - Entity being mapped from
		FROMTYPE - Type of entity being mapped from FROM-RULE Machine processible rule applicable to the entity being mapped from
		FROMRES - Restriction applicable to the entity being mapped from
		REL - Relationship of the entity being mapped from to the entity being mapped to
		RELA - Additional relation-ship label
		TOID - Identifier for the entity being mapped to
		TOSID - Source asserted identifier for the entity being mapped to
		TOEXPR - Entity being mapped to
		TOTYPE - Type of entity being mapped to

Table	Table Description	Columns/Column Descriptions
		TORULE - Machine processible rule applicable to the entity being mapped to
		TORES - Restriction applicable to the entity being mapped to
		MAPRULE - Machine processible rule applicable to this mapping
		MAPRES - Restriction applicable to this mapping
		MAPTYPE - Type of mapping
		MAPATN - The name of the attribute associated with this mapping
		MAPATV - The value of the attribute associated with this mapping
		CVF - Content View Flag
MRRANK.RRF	Order of precedence of vocabulary source and terms types	RANK- Numeric order of precedence, higher value wins
		SAB - Abbreviated source name for source vocabulary.
		TTY - Abbreviation for term type in source vocabulary
		SUPPRESS - NLM-recommended Source and Term Type (SAB/TTY) Suppressibility. Values = Y or N.
MRREL.RRF	Related Concepts	CUI1 - Unique identifier of first concept
		AUI1 - Unique identifier of first atom
		STYPE1 - The name of the column in MRCONSO.RRF that contains the identifier used for the first concept or first atom in source of the relationship.
		REL - Relationship of second concept or atom to first concept or atom

Table	Table Description	Columns/Column Descriptions
		CUI2 - Unique identifier of second concept
		AUI2 - Unique identifier of second atom
		STYPE2 - The name of the column in MRCONSO.RRF that contains the identifier used for the second concept or second atom in the source of the relationship
		RELA - Additional (more specific) relationship label (optional)
		RUI - Unique identifier of relationship
		SRUI - Source asserted relationship identifier, if present
		SAB - Abbreviated source name of the source of relationship
		SL - Source of relationship labels
		RG - Relationship group
		DIR - Source asserted directionality flag. Y indicates that this is the direction of the relationship in its source; N indicates that it is not; a blank indicates that it is not important or has not yet been determined.
		SUPPRESS - Suppressible flag. Values = O (obsolete), E (non-obsolete marked suppressible by an editor), Y (non-obsolete marked suppressible during inversion), or N (none of the above)
		CVF - Content View Flag
MRSAB.RRF	Pre-computed data provided as a convenience to users	CUI - CUI of the versioned SRC concept for a source
		RCUI - Root CUI
		VSAB - Versioned Source Abbreviation

Table	Table Description	Columns/Column Descriptions
		RSAB - Root Source Abbreviation
		SON - The official name for a source
		SF - Source Family
		SVER - Source Version
		VSTART - Meta Start Date
		VEND - Meta End Date
		IMETA - Meta Insert Version
		RMETA - Meta Remove Version
		SLC - Source License Contact
		SCC - Source Content Contact
		SRL - Source Restriction Level
		TFR - Term Frequency
		CFR - CUI Frequency
		CXTY - Context Type
		TTYL - Term Type List
		ATNL - Attribute Name List
		LAT - Language
		CENC - Character Encoding
		CURVER - Current Version
		SABIN - Source in Subset
		SSN - Source Short Name
		SCIT - Source Citation
MRSAT.RRF	Simple Concepts and Atom Attributes	CUI - Unique identifier for concept
		LUI - Unique identifier for term
		SUI - Unique identifier for string
		METAUI - Metathesaurus atom identifier

Table	Table Description	Columns/Column Descriptions
		STYPE - The name of the column in MRCONSO.RRF or MRREL.RRF that contains the identifier to which the attribute is attached, e.g., SAUI, SCUI, SRUI, CODE, CUI, AUI
		CODE - Most useful source asserted identifier (if the source vocabulary contains more than one) or a Metathesaurus-generated source entry identifier (if the source vocabulary has none)
		ATUI - Unique identifier for attribute
		SATUI - Source asserted attribute identifier
		ATN - Attribute name.
		SAB - Abbreviated source name
		ATV - Attribute value
		SUPPRESS - Suppressible flag. Values = O (obsolete), E (non-obsolete marked suppressible by an editor), Y (non-obsolete marked suppressible during inversion), or N (none of the above)
		CVF - Content View Flag
MRSMAP.RRF	Simple Mappings	MAPSETCUI - Unique identifier for the map set
		MAPSETSAB - Source abbreviation for the map set
		MAPID - Mapping ID (ATUI value of XMAP)
		MAPSID - Source asserted Mapping ID
		FROMEXPR - Mapped_from expression
		FROMTYPE - Type of mapped_from expression
		REL - Relationship

Table	Table Description	Columns/Column Descriptions
		RELA - Additional relationship label
		TOEXPR - Mapped_to expression
		TOTYPE - Type of mapped_to expression
		CVF - Content View Flag
MRSTY.RRF	Semantic Types	CUI - Unique identifier of concept
		TUI - Unique identifier of Semantic Type
		STN - Semantic Type tree number
		STY - Semantic Type.
		ATUI - Unique identifier for attribute
		CVF - Content View Flag
MRXNS_ENG.RRF	Word Index	LAT - Abbreviation of language of the string
		NSTR - Normalized string in lowercase
		CUI - Concept identifier
		LUI - Term identifier
		SUI - String identifier
MRXNW_ENG.RRF	Word Index	LAT - Abbreviation of language of the string in which the word appears
		NWD - Normalized word in lowercase
		CUI - Concept identifier
		LUI - Term identifier
		SUI - String identifier
MRXW_ENG.RRF	Word Index	LAT - Abbreviation of language of the string in which the word appears
		WD - Word in lowercase
		CUI - Concept identifier
		LUI - Term identifier
		SUI - String identifier

Table	Table Description	Columns/Column Descriptions
NCIMEME_200510_history.txt	NCIMEME history	conceptcode - Unique, permanent, alphanumeric identifier of the concept conceptname - Current text name of the concept editaction - Activity being recorded (create, modify, split, merge, retire) editdate - Date the activity occurred referencecode - Identifier for a concept affected by or influencing the action referencename - Current text name of the reference concept in all cases except concept name changes. For these, the referencename is the old name of the current concept.