

# Unified Medical Language System<sup>®</sup> (UMLS<sup>®</sup>) Basics

December 5, 2008

Instructors:  
Rachel Kleinsorge  
Jan Willis

National Library of Medicine  
National Institutes of Health  
U.S. Department of Health & Human Services



# Schedule

---

<b>8:30 - 9:30</b>	<b>Welcome and Introduction</b>
<b>9:30 - 10:15</b>	<b>Metathesaurus</b>
<b>10:15 - 10:30</b>	<b>BREAK</b>
<b>10:30 - 11:45</b>	<b>Metathesaurus, continued</b>
<b>11:45 - 12:45</b>	<b>LUNCH</b>
<b>12:45 - 2:15</b>	<b>MetamorphoSys &amp; UMLS Tools</b>
<b>2:15 - 2:30</b>	<b>BREAK</b>
<b>2:30 - 3:00</b>	<b>SPECIALIST Lexicon &amp; Lexical Tools</b>
	<b>Semantic Network</b>
<b>3:00 - 3:30</b>	<b>NLM, the UMLS, and Health Data Standards</b>
<b>3:30 - 4:00</b>	<b>Questions, Help, and Evaluation</b>
<b>4:00 - 5:00</b>	<b>Open Lab Time</b>

# Acknowledgements

---

Thanks to NLM staff and contractors who contributed to this presentation:

- Alan Aronson, PhD
- Vivian Auld
- Anantha Bangalore
- Olivier Bodenreider, MD
- Allen Browne
- Brian Carlsen
- Guy Divita
- William Hole, MD
- Kin Wah Fung, MD
- Betsy Humphreys
- John Kilbourne, MD
- Stephanie Lipow
- Chris Lu
- Stuart Nelson, MD
- Tammy Powell
- Tom Rindfleisch, PhD
- Laura Roth
- David Sheretz
- Suresh Srinivasan
- Karen Thorn
- Mark Tuttle
- Steve Emrick

# UMLS Video

---

<http://www.nlm.nih.gov/research/umls/UMLS300kbps.swf>

# The UMLS consists of

---

## Metathesaurus

1 million+  
biomedical  
**concepts**  
from over 100  
sources

## Semantic Network

135 broad  
**categories** and  
54 **relationships**  
between  
categories

## SPECIALIST Lexicon & Tools

lexical  
information and  
programs for  
**language  
processing**

**3 Knowledge Sources**  
used separately or together

# History of the UMLS

[Lindberg & al., *Methods*, 1993]  
[Humphreys & al., *JAMIA*, 1998]

- Started at National Library of Medicine, 1986
- “Long-term R&D project”
- Complementary to IAIMS

(Integrated Academic Information Management Systems)



«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»

# UMLS Objectives

---

- Intellectual “middleware”
- A set of multi-purpose tools for system developers
- Knowledge Sources used to overcome:
  - disparities in language format  
Ex: atrial fibrillation, auricular fibrillation, af
  - disparities in granularity and perspective  
Ex: Contusions, hematoma, bruise  
Ex: Instruct patient to promptly report nosebleeds and excessive bruising (NIC), Epistaxis (MeSH)

# UMLS Made available as:

---

- Data files
  - 3 separate sets of relational files
- Tools
  - MetamorphoSys (installation and customization)
  - RRF Subset Browser
  - lvg (lexical programs)
- Distributed on DVD; downloaded from UMLSKS

The UMLS is *not* an end-user application



# UMLS Access

---

- UMLS Knowledge Source Server
  - <http://umlsks.nlm.nih.gov/>
  - Browser, Navigators
  - APIs
  - Download data files and programs
  - Documentation and other resources
- Local storage and customization (Metathesaurus)
  - MetamorphoSys
  - RRF Browser

# UMLS Uses

---

- Information retrieval
- Thesaurus construction
- Natural language processing
- Automated indexing
- Electronic health records (EHR)
- Distribution mechanism for HIPAA, CHI, PHIN regulatory standards

# 2008AB UMLS: November 2008

---

- Metathesaurus:
  - 148 sources
  - 1,885,896 concepts
  - 17 languages
- Semantic Network:
  - 135 Semantic Types
  - 54 Semantic Relationships
- SPECIALIST Lexicon:
  - Over 330K records (over 557K inflected forms)

# A guided tour of UMLS Resources

---

- <http://www.nlm.nih.gov/research/umls/>
- <http://www.nlm.nih.gov/research/umls/documentation.html>
- [http://www.nlm.nih.gov/research/umls/listserv\\_info.html](http://www.nlm.nih.gov/research/umls/listserv_info.html)
- <https://list.nih.gov/archives/umlsusers-l.html>
- [http://www.nlm.nih.gov/research/umls/new\\_users.html](http://www.nlm.nih.gov/research/umls/new_users.html)

What is the UMLS?

# UMLS 3 Knowledge Sources

---

- Metathesaurus
  - Source vocabularies
  - Concepts
  - Relationships, Attributes
- Semantic Network
  - Semantic Types (categories)
  - Semantic Relationships
- Lexical resources
  - SPECIALIST Lexicon
  - Lexical tools

# Metathesaurus: clusters terms by meaning

- Synonymous terms clustered into a concept
- Preferred term is chosen
- Unique identifier (CUI) is assigned

Addison's disease	Metathesaurus	PN	
Addison's disease	SNOMED CT	PT	363732003
Addison's Disease	MedlinePlus	PT	T1233
Addison Disease	MeSH	PT	D000224
Bronzed disease	SNOMED Intl 1998	SY	DB-70620
Deficiency; corticorenal, primary	ICPC2-ICD10 Thesaurus	PT	MTHU021575
Primary Adrenal Insufficiency	MeSH	EN	D000224
Primary hypoadreanlism syndrome, Addison	MedDRA	LT	10036696

C0001403

Addison's disease

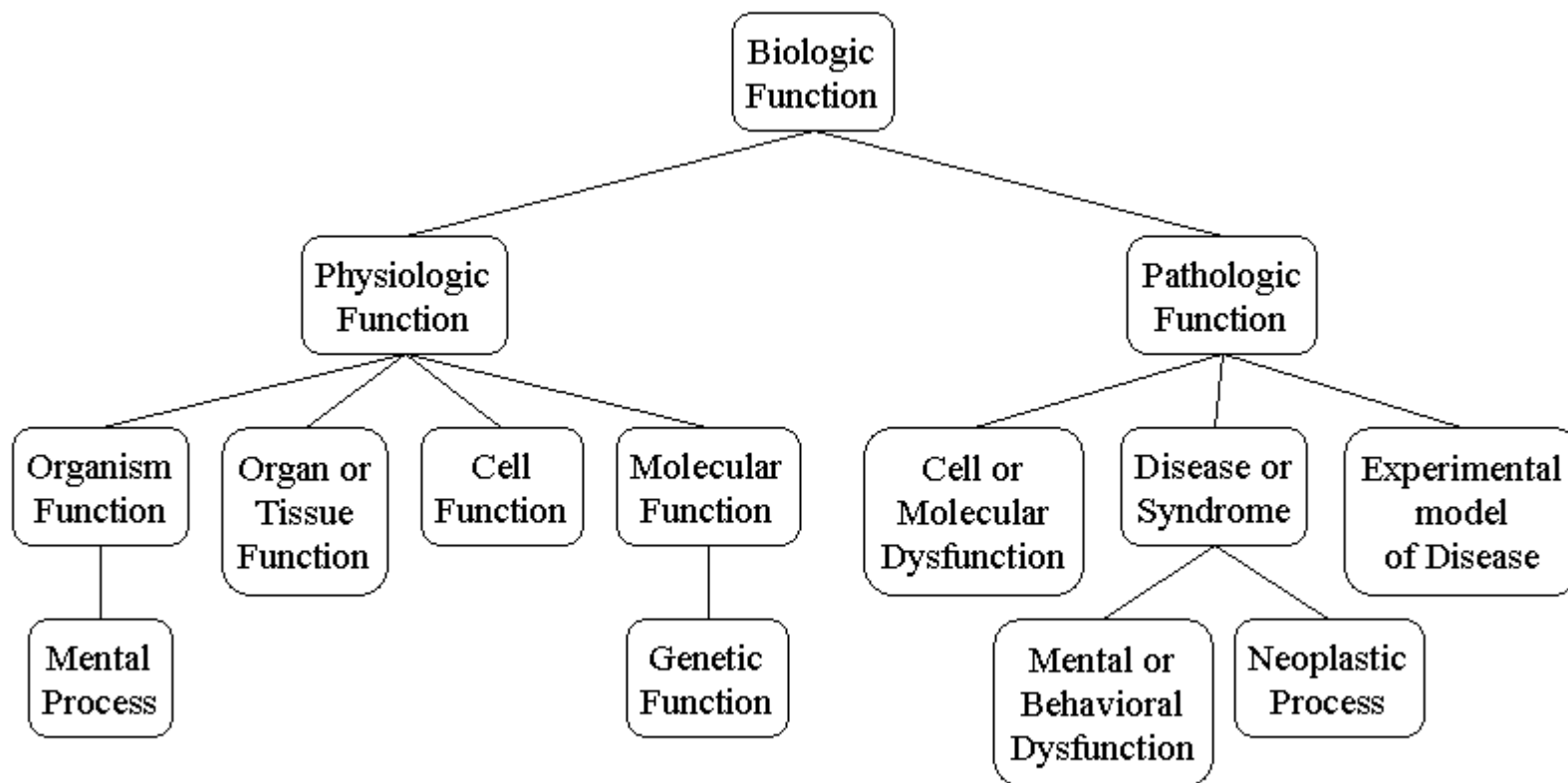
# Semantic Network

---

- 135 Semantic Types
  - Broad subject categories (Clinical Drug, Virus)
  - Ex:
    - **Addison's Disease**
    - **Semantic Type: Disease or Syndrome**
- 54 Semantic Relationships
  - Links between categories (isa, causes, treats)
  - Ex:
    - Virus **causes** Disease or Syndrome
- Types + Relationships
  - Form the structure of the semantic network
  - Broadly categorize the biomedical domain



# Semantic Network: Biologic Function



# SPECIALIST Lexicon & Lexical Tools

---

- General English lexicon of common words, biomedical terms
- Used by SPECIALIST Natural Language Processing System
- Lexical records contain encoded information:
  - syntactic (how words are put together),
  - morphological (inflection, derivation, and compounding), and
  - orthographic (spelling) information
- Lexical tools process terms
- Used in Metathesaurus production

Adrenal gland diseases  
Adrenal disorder  
Disorder of adrenal gland  
Diseases of the adrenal  
glands  
C0001621

# Lexical Tools

---

- Manage lexical variation in biomedical terminologies and text
- Used separately or with SPECIALIST Lexicon
- Perform transformations selected and ordered by users
- 3 primary programs: normalizer, word index generator, lexical variant generator

# Summary

---

## Metathesaurus:

- clusters terms into concepts – assigns unique identifier

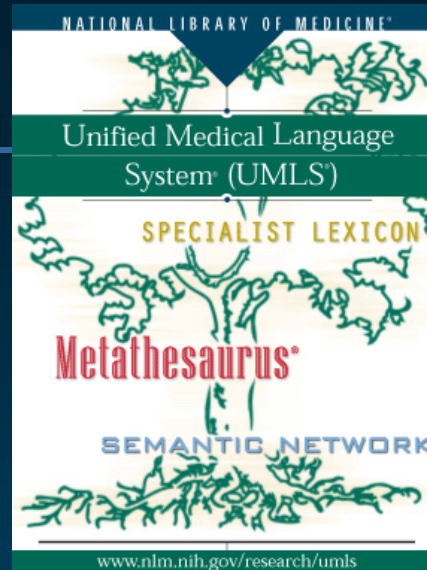
## Semantic Network

- defines relationships between concepts, organizes concepts into categories

## Lexicon and Lexical Tools:

- processes terms for entry into the Metathesaurus

# UMLS Bookmark



The UMLS is a collection of tools and resources that enable **systems** to link biomedical information from different sources, such as patient record systems, bibliographic databases, factual databases, decision support systems and other biomedical and health applications.

The UMLS also improves the ability of computer programs to understand biomedical meaning in user inquiries, and to use that understanding to retrieve and integrate relevant information for users.

The UMLS consists of data files and programs organized into three Knowledge Sources. Those sources are:

**METATHESAURUS** — Contains information about biomedical concepts and terms from over 100 controlled vocabularies and classification, including U.S. clinical and administrative standards. Primarily in English, with some foreign language translations. Accompanied by MetamorphoSys, a program that can be used to produce customized versions for particular applications.

**SEMANTIC NETWORK** — Defines broad subject categories to which Metathesaurus concepts are assigned, as well as the useful relationships that exist between categories.

**SPECIALIST LEXICON** — Includes commonly occurring English and biomedical words, as well as information necessary for natural language processing, such as word form, structure, and spelling variants. Accompanied by a suite of lexical programs.

Together these Knowledge Sources and programs provide useful tools to overcome differences in biomedical terminologies that often create barriers to effective retrieval and integration of biomedical information.

The UMLS Knowledge Source Server ([umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov)) provides Web access to the component parts of the UMLS as well as the ability to download UMLS files and programs for local use.

The UMLS does not include end-user applications beyond the Knowledge Source Server. Using the UMLS requires appropriate computing resources and technical expertise.

License Agreement — The UMLS is free of charge. Users complete a license agreement that protects the rights of the providers of component vocabularies. A separate license agreement is necessary with some sources.

#### ADDITIONAL INFORMATION

— [umlsinfo.nlm.nih.gov](http://umlsinfo.nlm.nih.gov)

- Fact Sheets
- Online Documentation
- License Agreement
- Knowledge Source Server
- Other UMLS Information, e.g., FAQs, Learning Resources, UMLS Users Listserv

— [custserv@nlm.nih.gov](mailto:custserv@nlm.nih.gov) — NLM's e-mail address for questions on the UMLS and all other NLM resources



U.S. DEPARTMENT OF HEALTH  
AND HUMAN SERVICES

4/2004

CB158

WWW.NLM.NIH.GOV

1-BBB-346-3656 • 8600 ROCKVILLE PIKE, BETHESDA, MD 20894

# Metathesaurus

# Metathesaurus

---

- The Metathesaurus is a
  - very large
  - multi-purpose
  - multi-lingual biomedical vocabulary
- It contains information about
  - biomedical and health related concepts
  - their various names
  - the relationships among them

# Metathesaurus Sources

---

- Over 100 controlled vocabularies, code sets, and mappings
- Determine scope of Metathesaurus
- Source meanings, hierarchies preserved
- Relationships, Semantic Types added



# Metathesaurus Sources 2

---

- 126 vocabularies in 17 languages
- Broad coverage of general, specialized biomedicine
- Varying purposes, structures, properties
- Sets of valid values:
  - Thesauri, e.g., MeSH, CRISP, NCI
  - Statistical classifications, e.g., ICD-9-CM
  - Billing codes, e.g., CPT, ABC Codes
  - Clinical coding systems, e.g., SNOMED CT
- Complete source lists in Appendixes A.1, B.4  
<http://www.nlm.nih.gov/research/umls/documentation.html>

# Metathesaurus Sources 3

---

- Sources reflect regular and irregular update schedules
- NLM resources determine currency of sources in the Metathesaurus
- Priority given to maintaining critical standards vocabularies

# Why Not Combine Them Into One?

---

- One size does not fit all
- NLM supports coordination when possible
- Growing awareness of benefits of standardization

“The UMLS approach assumes continuing diversity in the formats and vocabularies of different information sources and in the language employed by different elements of the biomedical community. It is not an attempt to build a single standard biomedical vocabulary.”



**Humphreys, BL and PL Schuyler**, The Unified Medical Language System: Moving beyond the vocabulary of bibliographic retrieval. In: Broering NC, ed. High- Performance Medical Libraries: advanced information management for the virtual era. Westport (CT): Meckler; 1993, p. 33.

# Metathesaurus Source Evaluation 1

---

1. Who does the vocabulary serve?
2. Is the vocabulary authoritative?
3. Is the thesaurus available in a well-structured, computable electronic form?
4. Have copyright ownership and license restriction levels been considered?

# Metathesaurus Source Evaluation 2

---

5. What thesaurus characteristics and principles are used?
6. How fully descriptive are the terms when seen alone, rather than in the associated context?
7. Will there be many cases where the 'face' meaning of your terms are different from the meaning of the same term in other areas of biomedicine?
8. Who would be available to assist in editing and review?

# Cimino's Desiderata

---



- Dr. James J. Cimino, M.D.
- <http://www.dbmi.columbia.edu/cimino/>
- Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. *Methods Inform Med* 1998; 37: 394-403.
- Content
- Concept Orientation
- Concept Permanence
- Meaningless Concept Identifiers
- [.....]

# Source data in Metathesaurus files

---

Names, Synonyms

Terms, Codes → MRCONSO

Relationships → MRREL

**Hierarchies** → MRHIER

**Mappings** → MRREL, MRMAP,  
MRSMAP

Attributes → MRSAT

**Definitions** → MRDEF

***CUIs links concept data across files***

# Metathesaurus files

---

- Concepts (MRCONSO)
  - Synonymous and near-synonymous terms
  - Codes
  - Concept Unique Identifier (CUI), Semantic Type added
- Relationships and mappings (MRREL, MRHIER, MRMAP, MRSMAP)
  - Concepts are related to other concepts
  - Source hierarchies are represented
  - Relationship attributes specify type, source of relationship
  - Simple and complex maps
- Attributes (MRSAT), (MRDEF)



# Concept

cluster of synonymous terms

Concept  
C0001621

Term  
adrenal disease gland  
L0001621

**S0011232** *Adrenal Gland Diseases*  
**S0011231** Adrenal Gland Disease  
**S0000441** Disease of adrenal gland  
**S0481705** Disease of adrenal gland, NOS  
**S0220090** Disease, adrenal gland  
**S0044801** Gland Disease, Adrenal

Term  
adrenal disorder gland  
unspecified  
L0041793

**S0860744** *Disorder of adrenal gland, unspecified*  
**S0217833** Unspecified disorder of adrenal glands

Term  
adrenal disorder  
L0161347

**S0225481** *ADRENAL DISORDER*  
**S0627685** DISORDER ADRENAL (NOS)

Term  
adrenal disorder gland  
L0181041

**S0632950** *Disorder of adrenal gland*  
**S0354509** Adrenal Gland Disorders

Term  
L0162317

**S0226798** *SURRENALE, MALADIES* FRE

# Concept structure

Concept (1.3 M) CUI  
set of synonymous names

Term (> 4.6 M) LUI  
set of normalized names

String (> 5.1 M) SUI  
distinct concept name

Atom (> 6.2 M) AUI  
concept name in source

**A1412439** headaches (BI)  
**S1459113** **headaches**

**A2882187** Headache (SNOMED)  
**A0066000** Headache (MeSH)  
**S0046854** **Headache**

**L0018681** **headache**

**A1641293** Cranial Pain (MeSH)  
**S1680378** **Cranial Pain**

**L1406212** **cranial pain**

**A0418053**  
HEAD PAIN CEPHALGIA (DxP)  
**S0375902**  
**HEAD PAIN CEPHALGIA**

**L0290366** **cephalgia head pain**

**C0018681** **Headache**

# Metathesaurus

## metadata, history and indexes

---

- Metadata files
  - MRSAB sources
  - MRFILES files
  - MRCOLS data elements
  - MRDOC finite values of selected data
  - MRRANK source/term type precedence
- Change files deleted, merged CUI, LUI
- MRCUI.RRF retired CUIs
- Concept name indexes

# Representing ambiguity multiple meanings

---

- Same string, different meanings
- Cold (L0009264)
  - cold temperature (C0009264)
  - common cold (C0009443)
  - chronic obstructive lung disease (C0024117)
  - Cold sensation (C0234192)
- AMBIGLUI.RRF (AMBIG.LUI)
  - L0009264|C0009443|
  - L0009264|C0024117|
  - L0009264|C0234192|
- AMBIGSUI.RRF (AMBIG.SUI)
  - S0026353|C0009443|
  - S0026353|C0234192|

# Metathesaurus Distribution Formats

---

- Pipe-delimited text files
- UTF-8 character encoding
  
- Original Release Format (ORF) **or** Rich Release Format (RRF)
  
- Original Release Format (ORF)
  - **Metathesaurus-Concept-Centric view**
  - Explicit concept-based connection between terms in different sources
  - Most information represented at concept (CUI) level

# Rich Release Format (RRF)

---

- **Source-Centric View** supports source transparency
- Represents original source information at atom (AUI) level
- Atom = unit of meaning (name) in a source
- New data fields and files to capture source specific identifiers, mappings

# Metathesaurus highlights

---

- Concept based
- Represents the meaning in each source
- Represents data in common format
- Adheres to principle of “source transparency”
  
- Tags source information
- Adds context-free unique identifiers
  
- Includes normalized word and string indexes produced using UMLS lexical tools

# Accessing the Metathesaurus

---

- UMLSKS <http://umlsks.nlm.nih.gov/>
  - Simple and advanced searches
  - View concept reports and raw relational records
  
- RRF Browser in MetamorphoSys
  - Generate and view customized subset
  - Search and browse
  - View concept reports and raw relational records



# Metathesaurus License Agreement 1

---

- Online Web-based license:

<http://www.nlm.nih.gov/research/umls/license.html>

- Read license
- Read appendix 1
- Read appendix 2
- Print a copy for your records
- Complete the Web form

Accept and continue

Accept and continue

Accept

 **Printer-friendly Version**

Submit

- Verification and turnaround:

- Receive e-mail from NLM and respond within 72 hours
- NLM official countersigns, license added to database
- Receive 2<sup>nd</sup> e-mail from NLM with new license number

# Metathesaurus License Agreement 2

---

- 2. No charges, usage fees or royalties will be paid to NLM.
- 5. Within 30 days of the end of any calendar year ... provide NLM with a brief report
- 11.c. required to include ... identifiers from ... the original source vocabularies
- 12. For material ... from some sources additional restrictions ... may apply.
- Special restrictions for SNOMED CT in light of the International Agreement

*See list of current sources in Appendix A.1.*

# License Restriction Levels 0-4

- Level 0 (28.2%)
  - *unrestricted*
- Level 1 (1.6%)
  - *negotiate to translate*
- Level 2 (0.4%)
  - *negotiate to use in health data creation*
- Level 3 (30.6%)
  - *negotiate to use in production*
  - *explicitly prohibited to provide Internet access*
- Level 4 (39.2%)
  - *unrestricted for U.S. use and distribution*

There may be additional restrictions, or separate license fees, associated with usage of specific vocabularies. Read the UMLS License Agreement, including the Appendix!

67%

# Metathesaurus Data Files

# MRCONSO (sample rows 6..10) <sup>10</sup>(2008AA)

	1	2	3	4	5	6	7	8	9	10	11
	CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI
6	C0001403	FRE	S	L1272481	PF	S1514427	Y	A1464383			
7	C0001403	GER	P	L1229627	PF	S1471573	Y	A7481335			D000224
8	C0001403	GER	S	L1239271	PF	S1481217	Y	A7485281			D000224
9	C0001403	JPN	P	L3437833	PF	S3965327	Y	A7888016			D000224
10	C0001403	JPN	S	L3442091	PF	S3969585	Y	A7890846			D000224

	12	3	14	15	6	17	8
	SAB	TRY	CODE	STR	SRL	SUPPRESS	CVF
6	WHOFRE	IT	0410	MALADIE D'ADDISON	2	N	
7	MSHGER	MH	D000224	Addison-Krankheit	3	N	
8	MSHGER	EP	D000224	Bronzehautkrankheit	3	N	
9	MSHJPN	MH	D000224	Addison病	3	N	
10	MSHPJN	SY	D000224	副腎性黒皮症	3	N	

# MRHIER (sample rows)

10 (2008AA)

	1	2	3	4	5	6
	CUI	AUI	CXN	PAUI	SAB	RELA
1	C0001403	A6954527	1	A6993207	MSH	
2	C0001403	A6954527	2	A0028022	MSH	
3	C0001403	A7559242	1	A7559213	PSY	
4	C0001403	A2922421	10	A3307650	SNOMEDCT	isa
5	C0001403	A2922421	11	A3307650	SNOMEDCT	isa

	7	8	9
	PTR	HCD	CVF
1	A0434168.A2367943.A2366890.A0135391.A6967433.A0020267.A6993207	C19.053.500.263	
2	A0434168.A2367943.A2366890.A0135391.A6970450.A0028022	C20.111.163	
3	A0449751.A7559213		
4	A3684559.A3886745.A3456474.A3456963.A3459284.A3473498.A6938225.A6919956.A6938229.A3307650		
5	A3684559.A3886745.A3456474.A3456963.A3459284.A6938487.A6938225.A6919956.A2933400.A2989549.A3307650		

# MRSAT (sample rows)

10 (2008AA)

	1	2	3	4	5	
	CUI	LUI	SUI	METAUI	STYPE	CODE
1	C0001403	L0001403	S0010792	A6954527	SDUI	D000224
2	C0001403	L0001403	S0010794	A6326321	SCUI	C712
3	C0001403	L0001403	S0354372	A2922421	SAUI	363732003
4	C0001403	L0001403	S0354372	A2922421	SAUI	363732003
5	C0001403				CUI	

	7	8	9	10	11	14	15
	ATUI	SATUI	ATN	SAB	ATV	SUPPRESS	CVE
1	AT38111026		DC	MSH	1	N	
2	AT33411754		MESH_UI	NDFRT	D000224	N	
3	AT24166602		DESCRIPTIONS TATUS	SNOMEDCT	0	N	
4	AT24529925		LANGUAGECODE	SNOMEDCT	en	N	
5	AT44908841		ST	MTH	R	N	

# MRSTY

⑩(2008AA)

---

CUI	TUI	STN	STY	ATUI	CVF
C0001403	T047	B2.2.1.2.1	Disease or Syndrome	AT32681465	1792



# MRDEF

10 (2008AA)

CUI	AUI	ATUI	SATUI	SAB	DEF	SUPPRESS	CVF
C0001403	A0388277	AT51219985		CSP	Disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin; due to tuberculosis or autoimmune induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol.	N	
C0001403	A6954527	AT43116350		MSH	An adrenal disease characterized by the progressive destruction of the ADRENAL CORTEX, resulting in insufficient production of ALDOSTERONE and HYDROCORTISONE. Clinical symptoms include ANOREXIA; NAUSEA; WEIGHT LOSS; MUSCLE WEAKNESS; and HYPERPIGMENTATION of the SKIN due to increase in circulating levels of ACTH precursor hormone which stimulates MELANOCYTES.	N	

MetamorphoSys

# MetamorphoSys

---

- Multi-platform Java software
- Included in each UMLS release
- Unzips native Metathesaurus compressed files
- Installs Knowledge Sources to local storage
- Customizes a local Metathesaurus

# Download from UMLSKS ...

---

- High speed Internet connection required
- Files must be stored in the same folder
- 2006AD UMLS Files
  - **mmsys.zip (zipped MetamorphoSys application)**
  - 2006AD-1-meta.nlm (compressed Metathesaurus data)
  - 2006AD-2-meta.nlm (compressed Metathesaurus data)
  - 2006AD-otherks.nlm (compressed Semantic Network and SPECIALIST Lexicon)
  - 2006AD.CHK
  - 2006AD.MD5
  - Copyright\_Notice.txt
  - **README.txt** ````

- ... or request DVD
- [umls\\_support@nlm.nih.gov](mailto:umls_support@nlm.nih.gov)
- Include your license number
- Run MetamorphoSys from DVD

# Machine Requirements

---

- A fast CPU – 1 GHz or higher
- 2 GB RAM recommended (512 MB min.)
- 6x (or better) DVD drive
- **20** GB minimum free disk space
  
- Runs on Sun Solaris 8 & 9, Windows XP, NT, 2000, Linux, and Mac
- 1-10 hours run time on platforms tested

# Customize the Metathesaurus

---

- Use MetamorphoSys
  - To comply with terms of license agreement
  - To remove unhelpful or harmful content
  - To change default settings (precedence, output)
- Customization is critical and requires understanding of:
  - Selected vocabularies
  - Functional requirements, purpose and perspective
- Technical expertise requires multidisciplinary team

# How MetamorphoSys Works

---

- Removes all information from all relational files from excluded vocabularies
  - atoms, strings, relations, hips, attributes, mappings
- Applies additional options selected by user
  - Changes to suppressibility or precedence
- Produces custom set of Metathesaurus relational files reflecting selected criteria
- Log file records subset details
- Output directory: set of Metathesaurus files

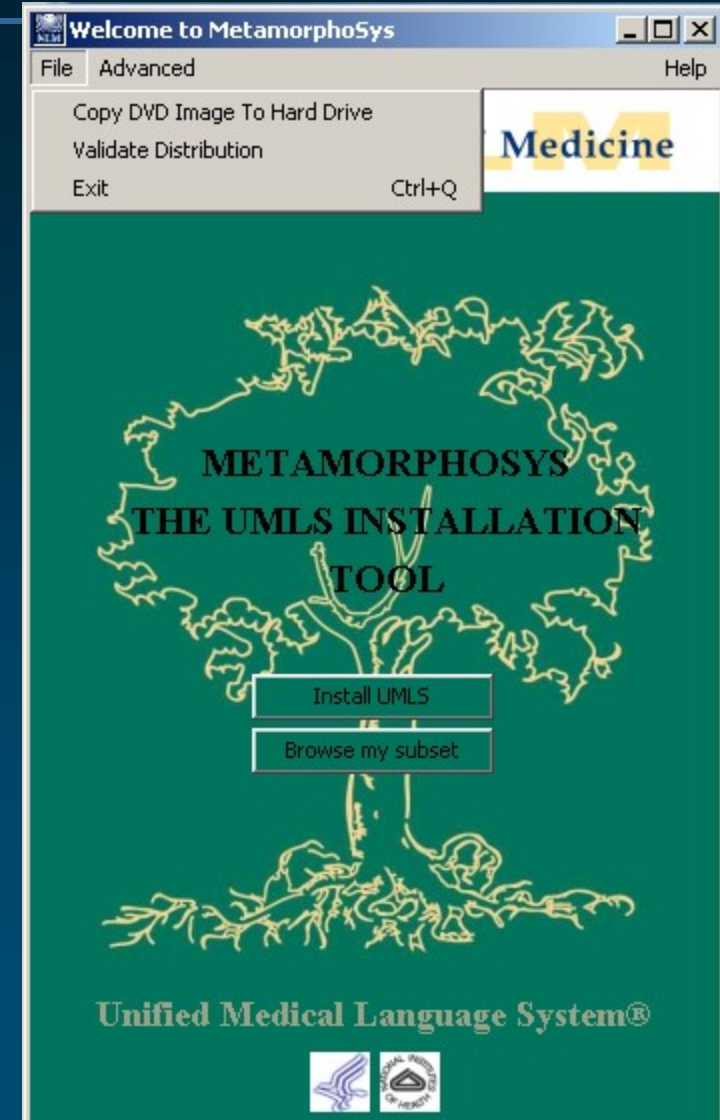
# MetamorphoSys Welcome Screen

## ■ Validate Distribution

- Ensures that all files were downloaded
- Process takes @ 30 minutes
- Writes validation.log file

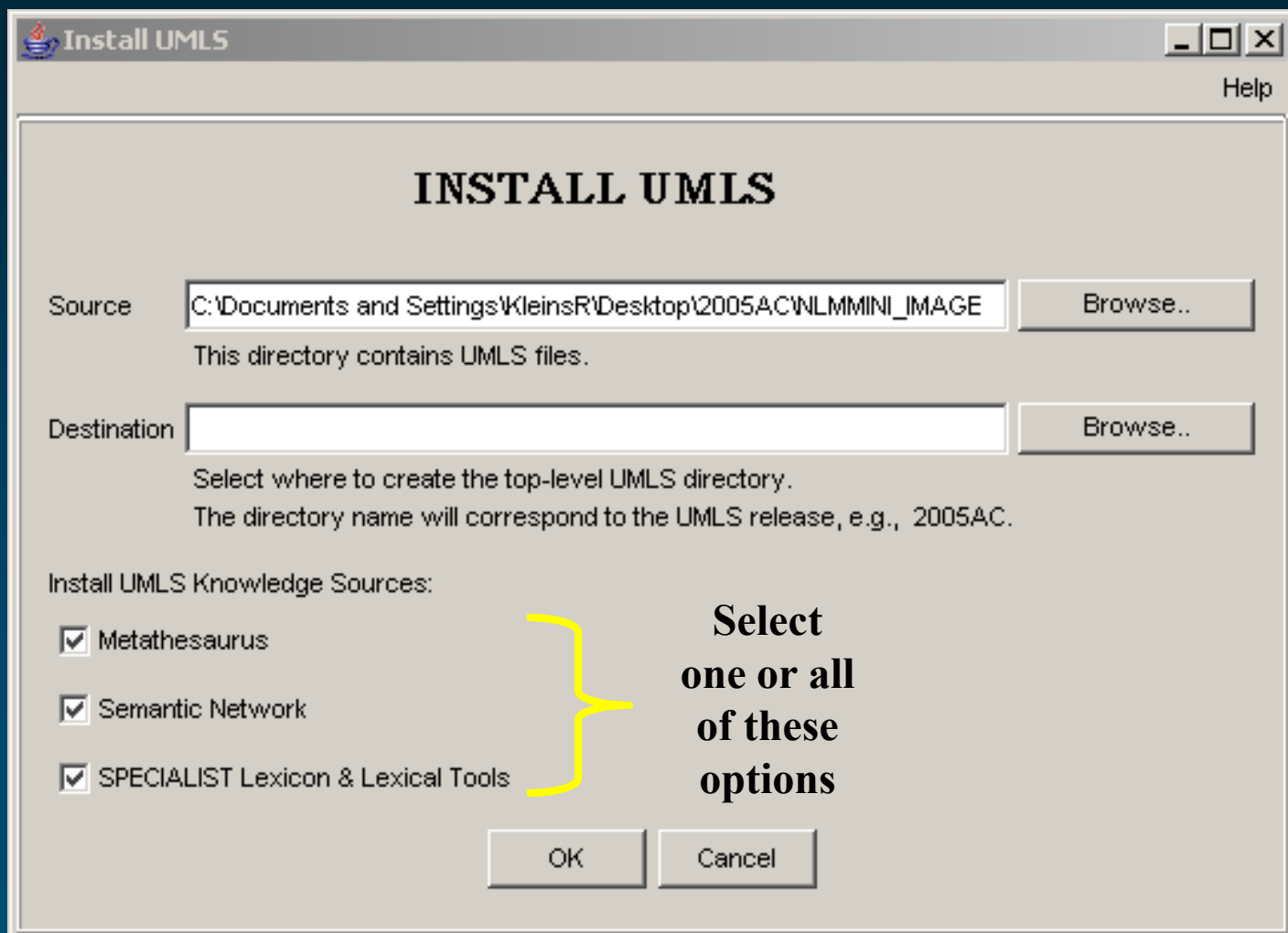
## ■ Copy DVD to hard drive

- Copies all files to local storage
- Allows multiple people to use one DVD
- May improve run time






# Install UMLS



# Validation Summary

Validation Log

 **Validation Summary**

1862	files passed validation
0	files are missing
992	files have unexpected byte counts
0	files have unexpected content (the files may be corrupt or of the wrong version)

Validation started at: Wed Sep 12 16:24:37 EDT 2007  
Validation ended at: Wed Sep 12 16:24:37 EDT 2007

The following file(s) have unexpected byte counts:

- Autorun.inf
- Copyright\_Notice.txt
- MMSYS/config/att\_types.dat
- MMSYS/config/mmsys.a.prop
- MMSYS/config/mmsys.b.prop
- MMSYS/config/mrpluscolsfiles.dat
- MMSYS/config/rel\_types.dat
- MMSYS/config/snomed\_rela\_map.dat
- MMSYS/config/timestamp.dat

Continue Cancel

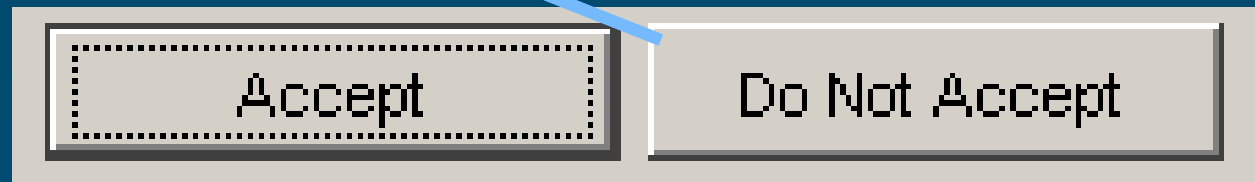
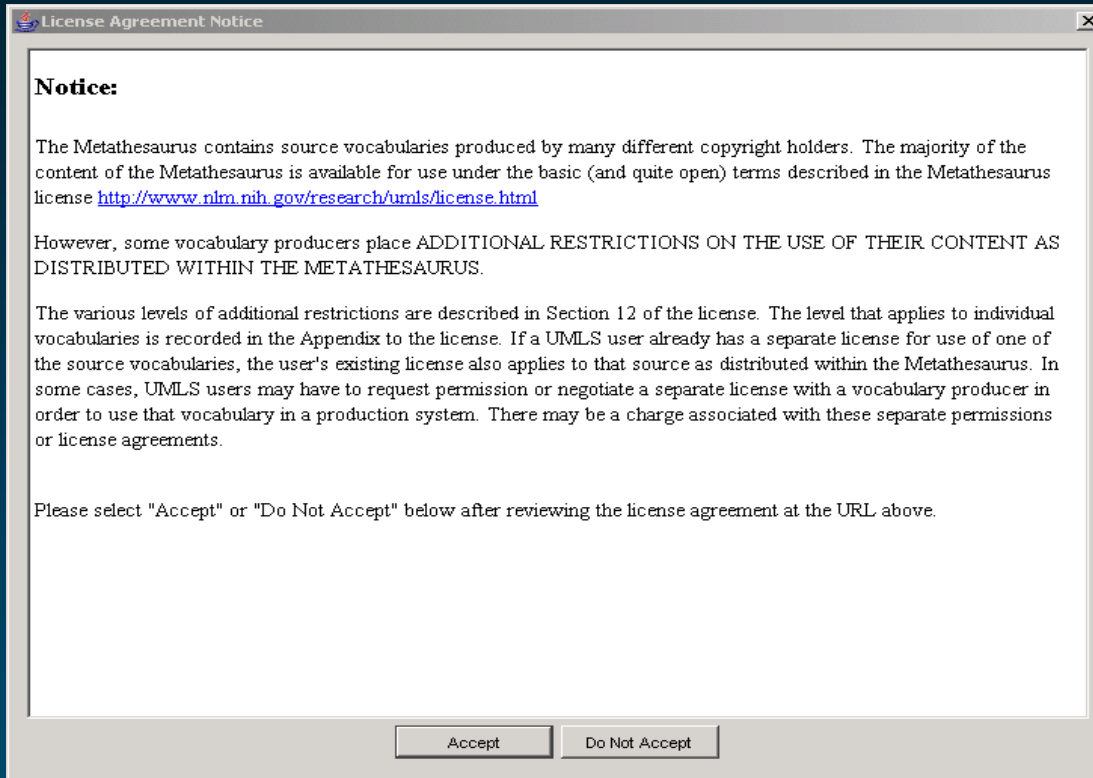
# Configuration Files

---

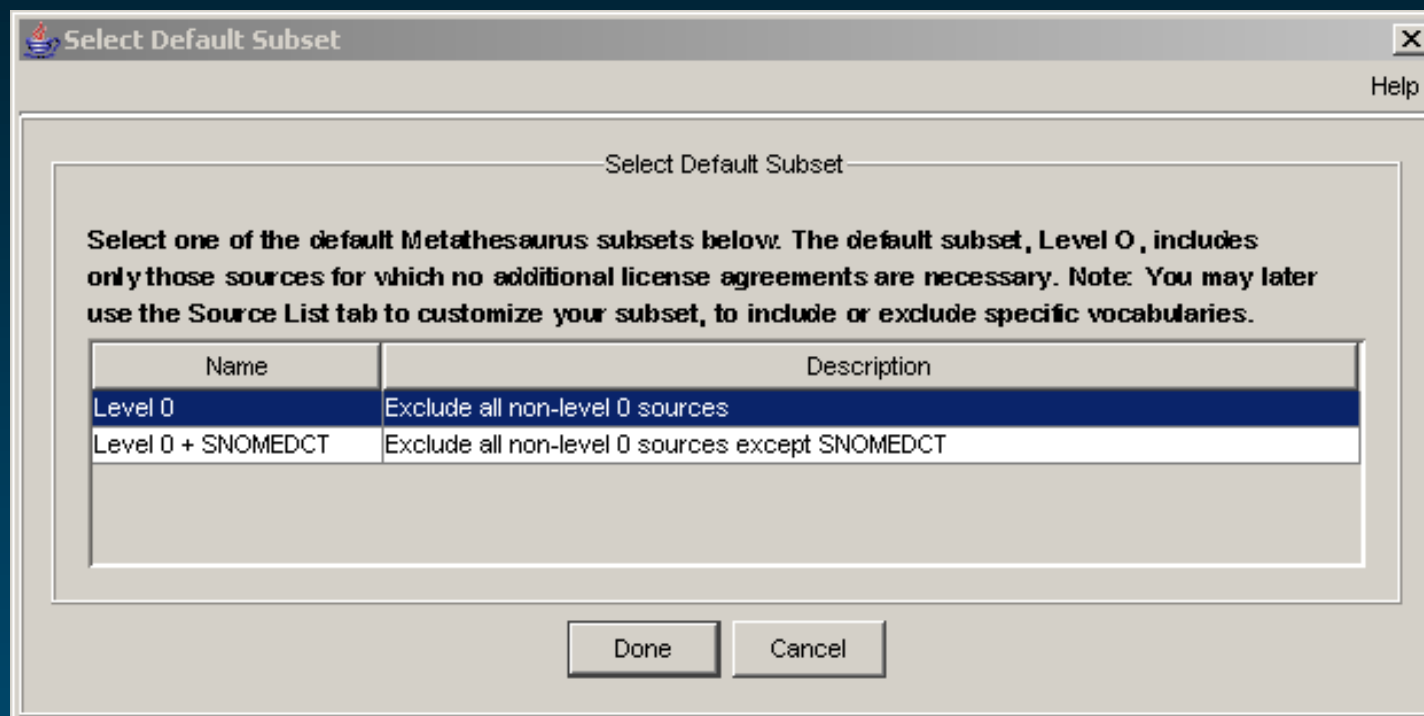


- New users must select "New Configuration"
- Returning users may select "Open Configuration" to open a previously saved MetamorphoSys Configuration
- This screen only appears if you install the Metathesaurus

# UMLS License Notice



# Select a default subset



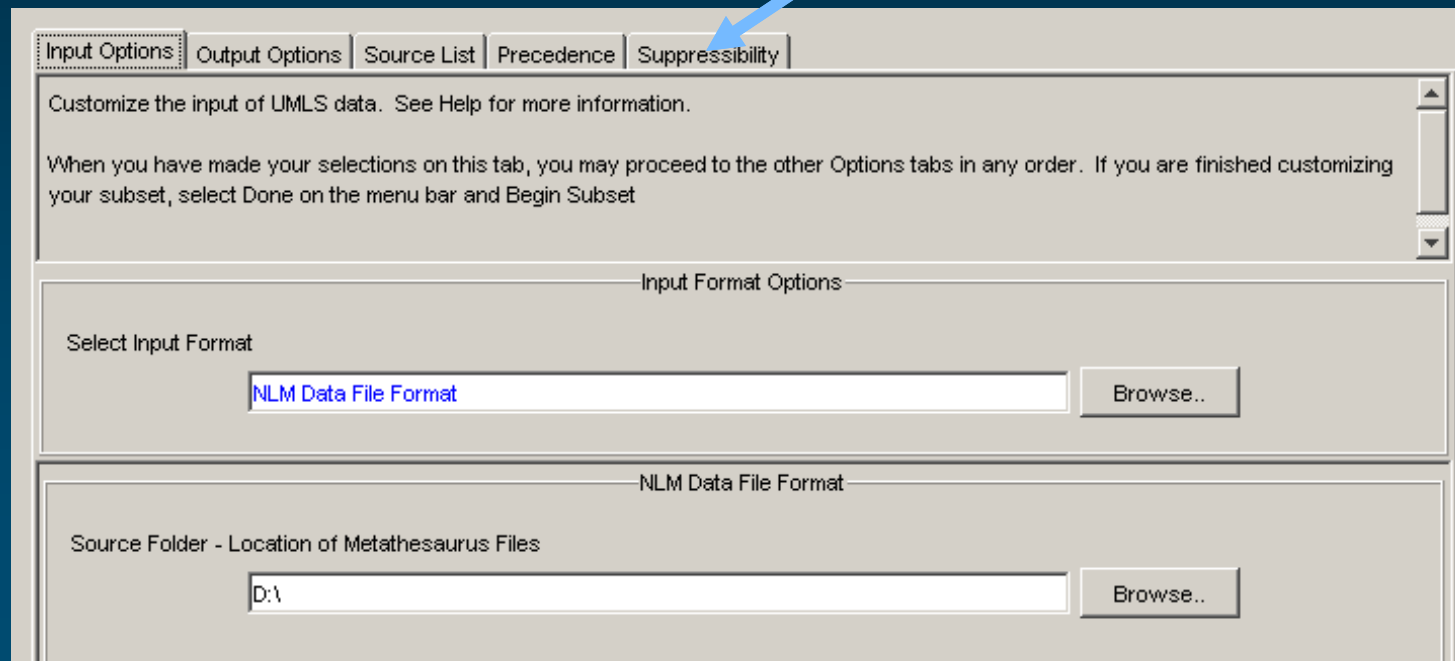
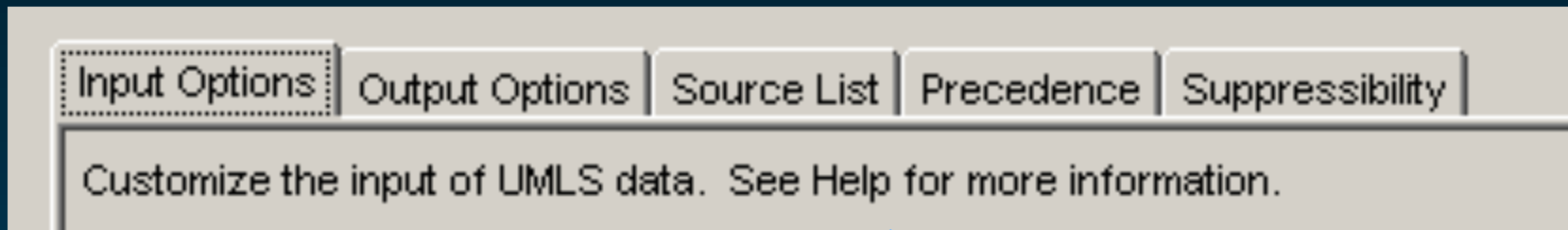
Level 0

→ no separate additional license agreements

Level 0 + SNOMEDCT

→ Non-U.S. users must have separate license agreements

# Input Options Tab



# Output Options Tab

Input Options

Output Options

Source List

Precedence

Suppressibility

Select data output options for your local application. See Help for more information.

Calculate MD5s for output files - writes mmsys.md5 file.

Select Output Format

Remove concepts containing only MTH atoms.

Add UTF-8 BOM characters to output files.

Subset Folder - Location of Subset Files

Write Oracle load script.

Remove records containing extended UTF-8 characters.

Write Mysql load script.

Truncate long fields to  characters.

Output versioned source abbreviations rather than versionless source abbreviations.

# Source List Tab

Input Options | Output Options | **Source List** | Precedence | Suppressibility

Indicate below to INCLUDE or EXCLUDE selected sources.

SELECTED sources appear on a dark background, e.g., AIR/RHEUM, 1993.

To undo selections and return to default source list, select "Reset Source List" from Reset menu at top.

Select sources to EXCLUDE from subset

Select sources to INCLUDE in subset

Exclude or  
Include

Sources to Exclude

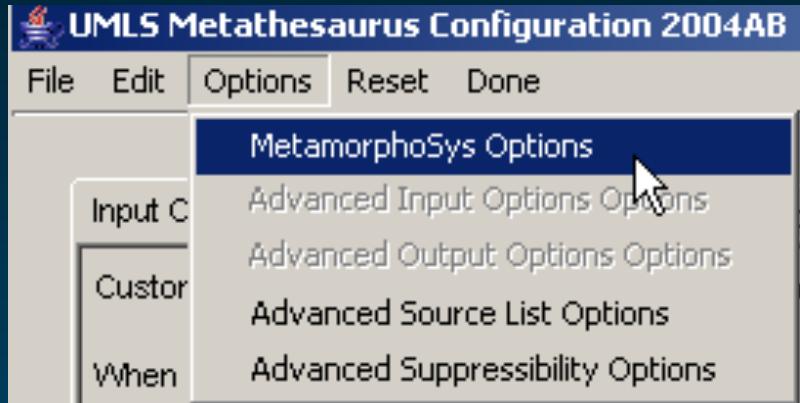
Full Source Name	Source Abbreviation	Source Family	Language	Level
AIR/RHEUM, 1993	AIR93	AIR	ENG	0
Alternative Billing Concepts, 2004	ALT2004	ALT	ENG	3
Alcohol and Other Drug Thesaurus, 2000	AOD2000	AOD	ENG	0
Beth Israel Vocabulary, 1.0	BI98	BI	ENG	2
Canonical Clinical Problem Statement System, 1999	CCPSS99	CCPSS	ENG	3
Clinical Classifications Software, 2003	CCS2003	CCS	ENG	0
Current Dental Terminology 2005 (CDT-5), 5	CDT5	CDT	ENG	3
COSTAR, 1989-1995	COSTAR_89-95	COSTAR	ENG	0
Medical Entities Dictionary, 2003	CPM2003	CPM	ENG	2
Physicians' Current Procedural Terminology, Spanish Translation, 2001	CPT01SP	CPT	SPA	3
Physicians' Current Procedural Terminology, 2005	CPT2005	CPT	ENG	3
CRISP Thesaurus, 2004	CSP2004	CSP	ENG	0
COSTART, 1995	CST95	CST	ENG	0
Diseases Database, 2000	DDB00	DDB	ENG	3
German translation of ICD10, 1995	DMDICD10_1995	ICD10	GER	1
German translation of UMDNS, 1996	DMDUMD_1996	UMD	GER	1
DSM-III-R, 1987	DSM3R_1987	DSM3R	ENG	3
SM-IV, 1994	DSM4_1994	DSM4	ENG	3
Xplain, 1994	DXP94	DXP	ENG	0
Gene Ontology, 2004_12_20	GO2004_12_20	GO	ENG	0
CPCS Version of Current Dental Terminology 2005 (CDT-5), 5	HCDT5	CDT	ENG	3
...	...	...	...	...

Highlighted  
rows are  
excluded  
from the  
subset.



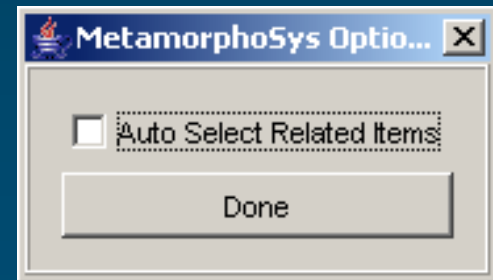


# MetamorphoSys Option Tab

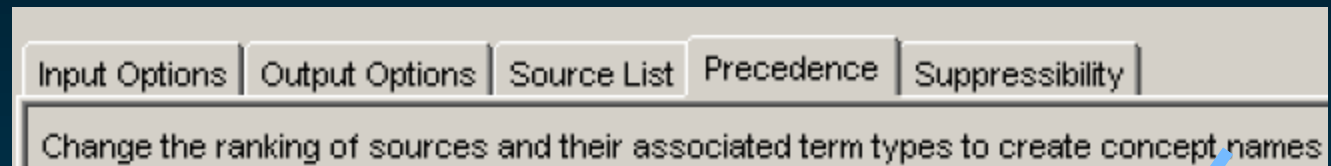


Source list behavior can be changed using the MetamorphoSys Option Tab

If you wish to Auto Select Related Items check this box



# Precedence Tab



- Ranks names by types of terms within sources
- Highest ranking name determines the Preferred Name

Input Options | Output Options | Source List | **Precedence** | Suppressibility

Change the ranking of sources and their associated term types to create concept names that are more useful in your local application. Concept names are determined by the term with highest ranking source/term type.

To move rows, either cut and paste rows, or drag and drop.

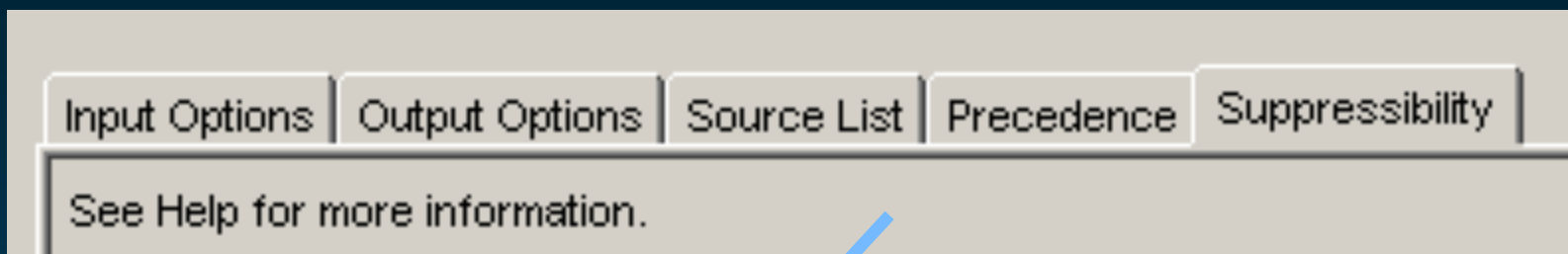
When you have made your selections on this tab, you may proceed to the other Options tabs in any order. If you are finished customizing

Precedence

Full Source Name	Source Abbreviation	Term Type
UMLS Metathesaurus	MTH	PN
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	MH
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	TQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EP
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EN
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	XQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	NM
RXNORM Project, META2004AB	RXNORM_04AB	SCD

Cut and paste rows to alter the preferred name

# Suppressibility Tab



Input Options | Output Options | Source List | Precedence | Suppressibility

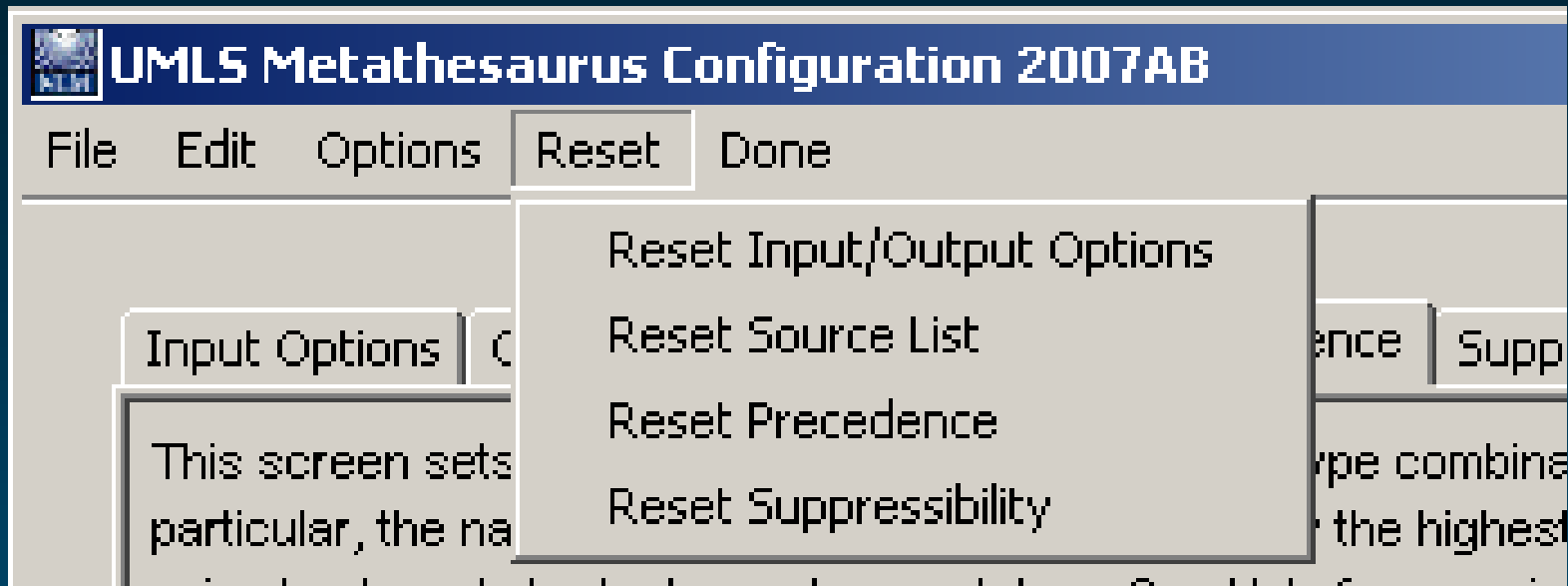
See Help for more information.

When you have made your selections on this tab, you may proceed to the other Options tabs in any order. If you are finished customizing your subset, select Done on the menu bar and Begin Subset.

International Classification of Primary Care, 1993	ICPC93	CP
International Classification of Primary Care, 1993	ICPC93	CS
International Classification of Primary Care, 1993	ICPC93	CX
International Classification of Primary Care, 1993	ICPC93	HT
International Classification of Primary Care, 1993	ICPC93	PC
International Classification of Primary Care, 1993	ICPC93	PS
International Classification of Primary Care, 1993	ICPC93	PT
International Classification of Primary Care, 1993	ICPC93	PX
ICPC, Basque Translation, 1993	ICPCBAQ_1993	CP

Highlighted source term types will be marked suppressible

# Reset menu



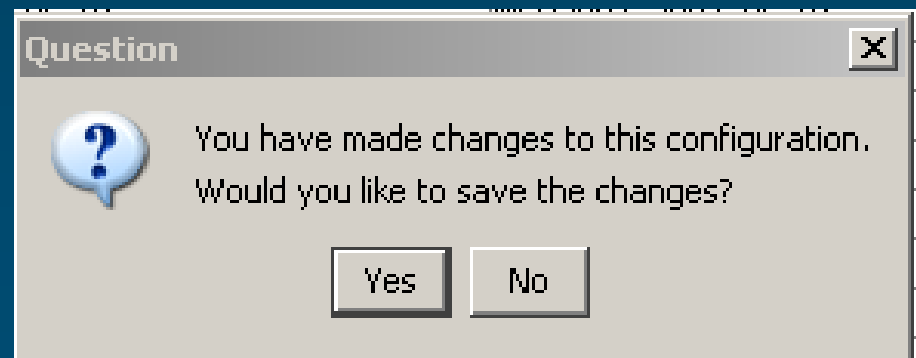
- Returns all filters to default selections
- Default selections in "mmsys.prop.default file" in config folder
- mmsys.prop.default contains properties in last run

# Done – Begin Subset

---



- Complete configuration options
- Done menu
- Begin Subset



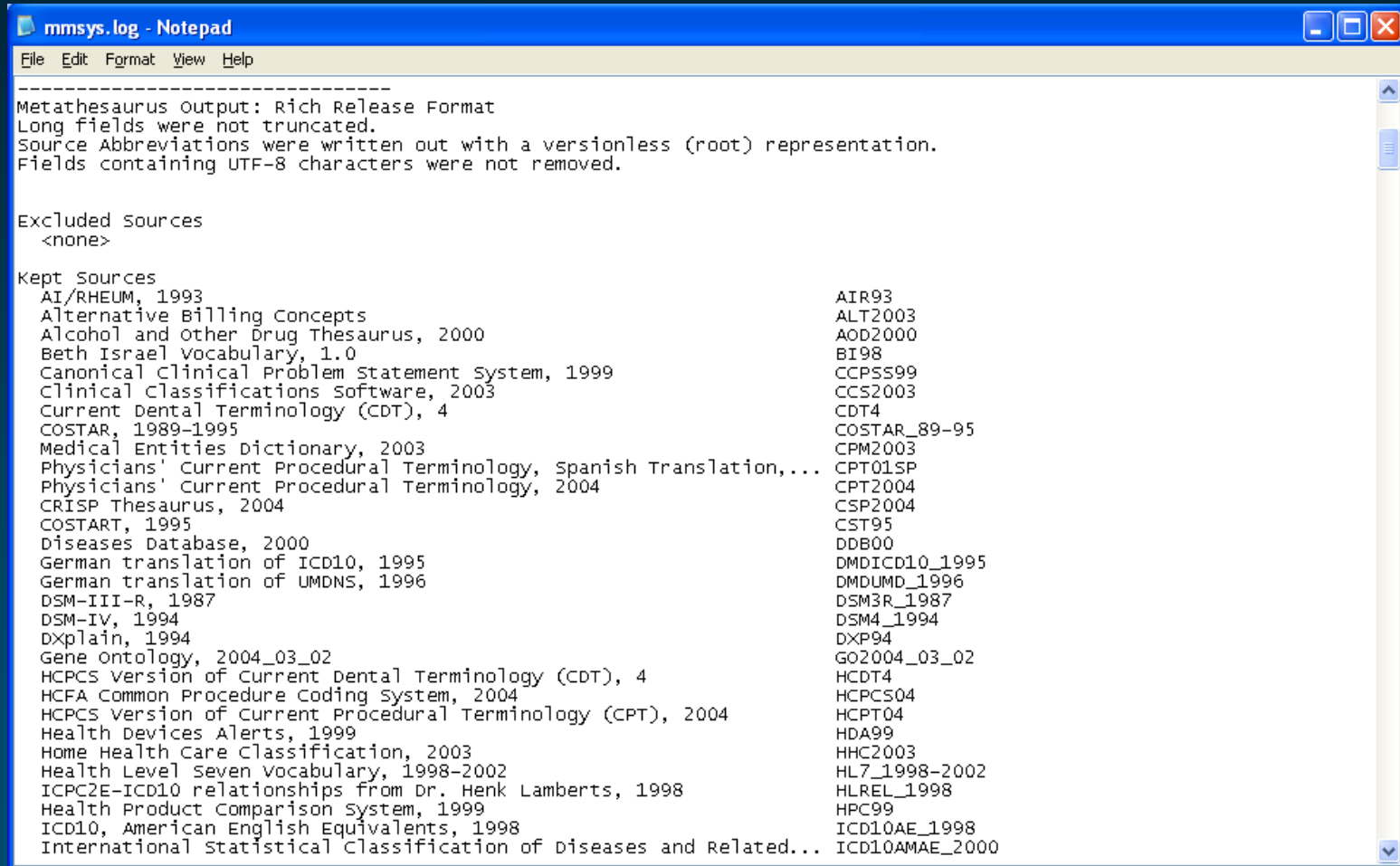
# MetamorphoSys log

mmsys.log - Notepad

File Edit Format View Help

```
MetamorphoSys version:..... 5.21
MetamorphoSys Build Date:..... 2004_08_30_14_47_11
UMLS Build Date:..... 2004_07_12_09_57_26
Release Version:..... 2004AB
Release Date:..... 20040720
Release Description:..... July 2004 Release
Metathesaurus source paths:..... C:\UMLS\DVDIMAGE
Subsetted Metathesaurus folder:..... C:\UMLS\DVDIMAGE\2004AB\META
Start at:..... wed Sep 01 13:04:18 EDT 2004
Initialize CUI List completed:..... wed Sep 01 13:06:20 EDT 2004
Subset Metathesaurus completed:..... wed Sep 01 14:13:02 EDT 2004
Subset Index Files completed:..... wed Sep 01 14:21:11 EDT 2004
Subset Release Metadata completed:..... wed Sep 01 14:21:17 EDT 2004
Finished at:..... wed Sep 01 14:22:48 EDT 2004
Concepts in source:..... 1078246
Concepts in subset:..... 1078246
Time elapsed:..... 01:18:29
```

# MetamorphoSys log



```
mmsys.log - Notepad
File Edit Format View Help
-----
Metathesaurus output: Rich Release Format
Long fields were not truncated.
Source Abbreviations were written out with a versionless (root) representation.
Fields containing UTF-8 characters were not removed.

Excluded Sources
<none>

Kept Sources
AI/RHEUM, 1993
Alternative Billing Concepts
Alcohol and Other Drug Thesaurus, 2000
Beth Israel vocabulary, 1.0
Canonical Clinical Problem Statement System, 1999
Clinical Classifications Software, 2003
Current Dental Terminology (CDT), 4
COSTAR, 1989-1995
Medical Entities Dictionary, 2003
Physicians' Current Procedural Terminology, Spanish Translation,...
Physicians' Current Procedural Terminology, 2004
CRISP Thesaurus, 2004
COSTART, 1995
Diseases Database, 2000
German translation of ICD10, 1995
German translation of UMDNS, 1996
DSM-III-R, 1987
DSM-IV, 1994
Dxplain, 1994
Gene ontology, 2004_03_02
HCPCS Version of Current Dental Terminology (CDT), 4
HCFA Common Procedure Coding System, 2004
HCPCS Version of Current Procedural Terminology (CPT), 2004
Health Devices Alerts, 1999
Home Health Care Classification, 2003
Health Level Seven vocabulary, 1998-2002
ICPC2E-ICD10 relationships from Dr. Henk Lamberts, 1998
Health Product Comparison System, 1999
ICD10, American English Equivalents, 1998
International Statistical Classification of Diseases and Related...

AIR93
ALT2003
AOD2000
BI98
CCPSS99
CCS2003
CDT4
COSTAR_89-95
CFM2003
CPT01SP
CPT2004
CSP2004
CST95
DDB00
DMDICD10_1995
DMDUMD_1996
DSM3R_1987
DSM4_1994
DXP94
GO2004_03_02
HCDT4
HCPCS04
HCPT04
HDA99
HHC2003
HL7_1998-2002
HLREL_1998
HPC99
ICD10AE_1998
ICD10AMAE_2000
```

**Kept Sources**

# Output directory contents

The screenshot shows a Windows Explorer window titled "C:\UMLS\DVDIMAGE\2004AB\META". The address bar shows the path "C:\UMLS\DVDIMAGE\2004AB\META". The main pane displays a list of files and folders with columns for Name, Size, Type, and Date Created. The left pane shows "File and Folder Tasks" and "Other Places". The details pane shows information for the "META" folder.

Name	Size	Type	Date Created
CHANGE		File Folder	9/1/2004 2:21 PM
indexes		File Folder	9/1/2004 1:06 PM
release.dat	1 KB	DAT File	9/1/2004 2:21 PM
config.prop	8 KB	PROP File	9/1/2004 2:22 PM
AMBIGLUI.RRF	1,225 KB	RRF File	9/1/2004 2:21 PM
AMBIGSUI.RRF	955 KB	RRF File	9/1/2004 2:21 PM
MR.COC.RRF	809,207 KB	RRF File	9/1/2004 1:06 PM
MRCOLS.RRF	21 KB	RRF File	9/1/2004 1:06 PM
MRCONSO.RRF	596,528 KB	RRF File	9/1/2004 1:06 PM
MRCUI.RRF	9,221 KB	RRF File	9/1/2004 1:06 PM
MRCXT.RRF	9,391,778 KB	RRF File	9/1/2004 1:06 PM
MRDEF.RRF	17,172 KB	RRF File	9/1/2004 1:06 PM
MRDOC.RRF	88 KB	RRF File	9/1/2004 1:06 PM
MRFILES.RRF	4 KB	RRF File	9/1/2004 1:06 PM
MRHIER.RRF	899,786 KB	RRF File	9/1/2004 1:06 PM
MRHIST.RRF	70,843 KB	RRF File	9/1/2004 1:06 PM
MRMAP.RRF	9,362 KB	RRF File	9/1/2004 1:06 PM
MRRANK.RRF	9 KB	RRF File	9/1/2004 1:06 PM
MRREL.RRF	1,499,350 KB	RRF File	9/1/2004 1:06 PM
MRSAB.RRF	54 KB	RRF File	9/1/2004 1:06 PM
MRSMAP.RRF	6,141 KB	RRF File	9/1/2004 1:06 PM
MRSTY.RRF	69,771 KB	RRF File	9/1/2004 1:06 PM
MRXNS_ENG.RRF	229,014 KB	RRF File	9/1/2004 1:06 PM
MRXNW_ENG.RRF	485,810 KB	RRF File	9/1/2004 1:06 PM
MRXW_BAO.RRF	108 KB	RRF File	9/1/2004 1:06 PM

**File and Folder Tasks**

- Make a new folder
- Publish this folder to the Web
- Share this folder

**Other Places**

- 2004AB
- My Documents
- Shared Documents
- My Computer
- My Network Places

**Details**

**META**  
File Folder  
Date Modified: Today, September 02, 2004, 12:47 PM



# RRF Subset Browser

---

- Must have an RRF Subset
- View customized subsets
  - Tree, report, or raw data for each term
  - Expandable Concept Report View
  - Search for CUI or more information
- Quickly find terms within a subset
- Hyperlinked concepts

*[http://www.nlm.nih.gov/research/umls/rrf\\_help.html](http://www.nlm.nih.gov/research/umls/rrf_help.html)*

# RRF Browser word search & report view

The screenshot displays the 'Rich Release Format Browser 2005AB C0001403' application. The interface includes a menu bar (File, Edit, Options, Help) and a toolbar with navigation icons. The 'Word Search' tab is active, showing a search for 'addison's disease' with 15 results. The 'Report View' tab is also active, displaying a detailed report for the selected concept: 'C0001403 Primary adrenocortical insufficiency'. The report includes semantic types, atoms, contexts, and a list of related concepts with links to their respective pages.

Rich Release Format Browser 2005AB C0001403

File Edit Options Help

Subset Directory C:\Documents and Settings\Jan Willis\My Documents\Old No

CUI Search Tree Browser Word Search

Enter search terms: (English)

addison's disease

Search..

Select a concept. (1 to 12 of 15 results)

- C0001403 Primary adrenocortical insufficiency
- C0271737 Addison's disease due to autoimmunity
- C0546992 Tuberculous Addison's disease
- C0342477 Addison's disease with adrenoleucodystrophy
- C0270989 Myopathy in Addison's disease
- C0268785 Salt-losing nephropathy
- C0271739 Familial adrenocortical hypoplasia
- C0002892 Pernicious anemia
- C0036420 Unspecified circumscribed scleroderma
- C0271740 Congenital primary adrenocortical hypofunction
- C0085859 Polyglandular autoimmune syndrome, type 1
- C0085860 Polyglandular autoimmune syndrome, type 2

Raw View Report View

- Concept: [CUI: C0001403]** Primary adrenocortical insufficiency
- Semantic Type**
  - Disease or Syndrome
- Atoms (5):** (Sorted by Source, String)
- Contexts (40)**
  - SNOMEDCT/PT/363732003 10 Addison's disease
    - ANC**
      - [SNOMED CT Concept](#)
      - [Clinical finding](#)
      - [Finding by site](#)
      - [Finding of body region](#)
      - [Finding of trunk structure](#)
      - [General finding of abdomen](#)
      - [Disorder of abdomen](#)
      - [Disorder of adrenal gland](#)
      - [Disorder of adrenal cortex](#)
      - [Adrenal cortical hypofunction](#)
  - SIB**
  - CHD**
    - [Addison's disease due to autoimmunity](#)
    - [Addison's disease with adrenoleucodystrophy](#)
    - [Polyglandular autoimmune syndrome, type 1](#)
    - [Tuberculous Addison's disease](#)
  - SNOMEDCT/PT/363732003 11 Addison's disease

# RRF Browser tree browser & report view

The screenshot displays the 'Rich Release Format Browser 2005AB C0001403' application window. The interface includes a menu bar with 'File', 'Edit', and 'Options', and a 'Help' button. The main area is divided into two panes. The left pane, titled 'Tree Browser', shows a 'Context Tree Tops' section with a tree structure of categories: SNOMED CT Concept, Attribute, Body structure, Clinical finding, Context-dependent categories, Environments and geographical locations, Events, Observable entity, Organism, Pharmaceutical / biologic product, Physical force, Physical object, Procedure, Qualifier value, Social context, Special concept, Specimen, Staging and scales, and Substance. The right pane, titled 'Report View', displays detailed information for the selected concept: 'Concept: [CUI: C0001403] Primary adrenocortical insufficiency'. It also shows 'Semantic Type' (Disease or Syndrome), 'Atoms (5)' (Sorted by Source, String), and 'Contexts (40)'. The contexts listed include 'SNOMEDCT/PT/363732003 10 Addison's disease', which is expanded to show 'ANC' with various related SNOMED CT concepts such as 'Addison's disease due to autoimmunity', 'Addison's disease with adrenoleucodystrophy', 'Polyglandular autoimmune syndrome, type 1', and 'Tuberculous Addison's disease'. The application title bar shows the file path: 'C:\Documents and Settings\Jan Willis\My Documents\Old No...'. The window also features standard navigation icons and a 'Subset Directory' field.

Rich Release Format Browser 2005AB C0001403

File Edit Options Help

Subset Directory C:\Documents and Settings\Jan Willis\My Documents\Old No...

CUI Search Tree Browser Word Search

Raw View Report View

Context Tree Tops

- [-] SNOMED CT Concept
  - [+] Attribute
  - [+] Body structure
  - [+] Clinical finding
  - [+] Context-dependent categories
  - [+] Environments and geographical locations
  - [+] Events
  - [+] Observable entity
  - [+] Organism
  - [+] Pharmaceutical / biologic product
  - [+] Physical force
  - [+] Physical object
  - [+] Procedure
  - [+] Qualifier value
  - [+] Social context
  - [+] Special concept
  - [+] Specimen
  - [+] Staging and scales
  - [+] Substance

[+] **Concept: [CUI: C0001403]** Primary adrenocortical insufficiency

[-] **Semantic Type**  
Disease or Syndrome

[+] **Atoms (5):** (Sorted by Source, String)

[-] **Contexts (40)**

- [-] SNOMEDCT/PT/363732003 10 Addison's disease
  - [-] **ANC**
    - [SNOMED CT Concept](#)
    - [Clinical finding](#)
    - [Finding by site](#)
    - [Finding of body region](#)
    - [Finding of trunk structure](#)
    - [General finding of abdomen](#)
    - [Disorder of abdomen](#)
    - [Disorder of adrenal gland](#)
    - [Disorder of adrenal cortex](#)
    - [Adrenal cortical hypofunction](#)
  - [+] **SIB**
  - [-] **CHD**
    - [Addison's disease due to autoimmunity](#)
    - [Addison's disease with adrenoleucodystrophy](#)
    - [Polyglandular autoimmune syndrome, type 1](#)
    - [Tuberculous Addison's disease](#)
- [+] SNOMEDCT/PT/363732003 11 Addison's disease

# RRF Browser -- search by code

Rich Release Format Browser 2007AB-AC-20070904 C0001002

File Edit Options Help

Monosp... ..

C:\Documents and Settings\kleinsr\Desktop\test rxnorm\2007AB-AC-20070904\M...

CUI Search Code Search Tree Browser Word Search

Enter a CODE:

D000096

Get Concepts..

Select a concept. (1 result)

C0001002 Acetone

Raw View Report View

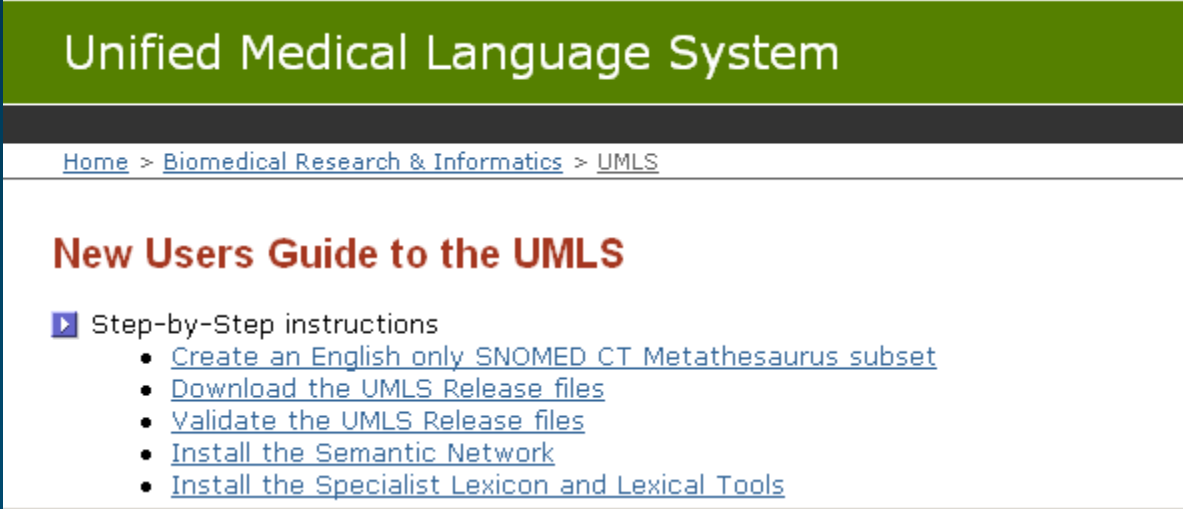
- ⊕ **Concept:** [CUI C0001002] Acetone
- ⊖ **Semantic Types**
  - Organic Chemical
  - Pharmacologic Substance
- ⊖ **Definitions**

MSH/MH|A colorless liquid used as a solvent and an antiseptic. It is one of the ketone bodies produced during ketoacidosis.

MSHSWE/MH|En brandfarlig, färglös, flyktig vätska med behaglig eterisk doft. Den är ett allmänt använt lösningsmedel och används som sådant inom farmacin; vid koncentrationer över 80% används den som antiseptis kt medel. Aceton är ett av de ketonämnen som bildas vid ketoacidosis, som i diabetes eller vid svält.
- ⊖ **Atoms (16):** (Sorted by Source, String)
  - 2-Propanone [A0527815/MSH/N1/D000096]
  - ⊕ Acetone [A0018351/MSH/MH/D000096]
  - ACETON [A3910037/MSHCZE/MH/D000096]
  - Asetoni [A9072482/MSHFIN/MH/D000096]
  - Acétone [A7439779/MSHFRE/MH/D000096]
  - Aceton [A7481179/MSHGER/MH/D000096]
  - Azeton [A7483923/MSHGER/EP/D000096]

# MetamorphoSys Step-by-Step help

- MS Word files
  - Screen grabs with explanation
  - Cover common MMSYS tasks such as installing the UMLS Knowledge Sources and creating a custom subset
  - Suggestions for new Step-by-Steps requested!



The screenshot shows a web page for the Unified Medical Language System (UMLS). The page has a green header with the text "Unified Medical Language System". Below the header is a navigation breadcrumb: "Home > Biomedical Research & Informatics > UMLS". The main content area features a red heading "New Users Guide to the UMLS". Underneath this heading is a section titled "Step-by-Step instructions" with a right-pointing triangle icon. This section contains a bulleted list of five links: "Create an English only SNOMED CT Metathesaurus subset", "Download the UMLS Release files", "Validate the UMLS Release files", "Install the Semantic Network", and "Install the Specialist Lexicon and Lexical Tools".

⑩ [http://www.nlm.nih.gov/research/umls/new\\_users.html](http://www.nlm.nih.gov/research/umls/new_users.html)

How do I?

# Specify sources for a customized subset?

Input Options | Output Options | **Source List** | Precedence | Suppressibility

Indicate below to INCLUDE or EXCLUDE selected sources.  
SELECTED sources appear on a dark background, e.g., AIRRHEUM, 1993.

To undo selections and return to default source list, select "Reset Source List" from Reset menu at top.

Select sources to EXCLUDE from subset  
 Select sources to INCLUDE in subset

Sources to Exclude

10 Exclude or Include

Sources to Exclude

Full Source Name	Source Abbreviation	Source Family	Language	Level
AIRRHEUM, 1993	AIR93	AIR	ENG	0
Alternative Billing Concepts, 2004	ALT2004	ALT	ENG	3
Alcohol and Other Drug Thesaurus, 2000	AOD2000	AOD	ENG	0
Beth Israel Vocabulary, 1.0	BI98	BI	ENG	2
Canonical Clinical Problem Statement System, 1999	CCPSS99	CCPSS	ENG	3
Clinical Classifications Software, 2003	CCS2003	CCS	ENG	0
Current Dental Terminology 2005 (CDT-5), 5	CDT5	CDT	ENG	3
COSTAR, 1989-1995	COSTAR_89-95	COSTAR	ENG	0
Medical Entities Dictionary, 2003	CPM2003	CPM	ENG	2
Physicians' Current Procedural Terminology, Spanish Translation, 2001	CPT01SP	CPT	SPA	3
Physicians' Current Procedural Terminology, 2005	CPT2005	CPT	ENG	3
CRISP Thesaurus, 2004	CSP2004	CSP	ENG	0
COSTART, 1995	CST95	CST	ENG	0
Diseases Database, 2000	DDB00	DDB	ENG	3
German translation of ICD10, 1995	DMDICD10_1995	ICD10	GER	1
German translation of UMDNS, 1996	DMDUMD_1996	UMD	GER	1
DSM-III-R, 1987	DSM3R_1987	DSM3R	ENG	3
DSM-IV, 1994	DSM4_1994	DSM4	ENG	3
DXplain, 1994	DXP94	DXP	ENG	0
Gene Ontology, 2004_12_20	GO2004_12_20	GO	ENG	0
HCPCS Version of Current Dental Terminology 2005 (CDT-5), 5	HCDT5	CDT	ENG	3
Healthcare Common Procedure Coding System, 2005	HCPCS05	HCPCS	ENG	0

10 Highlighted rows are excluded from the subset.

# Create a custom database load script?

- Select the Output Options tab
- Check the box next to the type of load script you require

Input Options | **Output Options** | Source List | Precedence | Suppressibility

Select data output options to customize the subset and create additional files if desired. See Help for more information.

Users may proceed to the other Options tabs in any order by clicking on the tabs across the top of the screen or by clicking "Done" on the File bar and selecting "Begin Subset".

Output Format Options

Select Output Format

Rich Release Format

Write Database Load Scripts

Write Oracle load script.  Write MySQL load script.

- Other Options on this tab



# Change how preferred term is set?

- Select the Precedence tab
- Cut and paste or drag and drop source and term types to reflect your preferred ranking order

Input Options | Output Options | Source List | **Precedence** | Suppressibility

Change the ranking of sources and their associated term types to create concept names that are more useful in your local application. Concept names are determined by the term with highest ranking source/term type.

To move rows, either cut and paste rows, or drag and drop.

When you have made your selections on this tab, you may proceed to the other Options tabs in any order. If you are finished customizing

Precedence

Full Source Name	Source Abbreviation	Term Type
UMLS Metathesaurus	MTH	PT
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	MH
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	TQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EP
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EN
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	XQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	NM
RXNORM Project, META2004AB	RXNORM_04AB	SCD

# Remove specific term types from subset?

Input Options | Output Options | Source List | Precedence | **Suppressibility**

This screen contains the Suppressibility Filter, which specifies source/term type combinations to be suppressed. Users can customize the subset by selecting and deselecting source/term type combinations. See Help for more information.

Users may proceed to the other Options tabs in any order by clicking on the tabs across the top of the screen or by clicking "Done" on the File bar and selecting "Begin Subset".

Select One or More Suppressible Term Types

Source	Source Abbreviation	Term Type
AIRRHEUM, 1993	AIR93	DI
AIRRHEUM, 1993	AIR93	FI
AIRRHEUM, 1993	AIR93	HT
AIRRHEUM, 1993	AIR93	SY

- Select Suppressibility tab
- Highlight rows to exclude

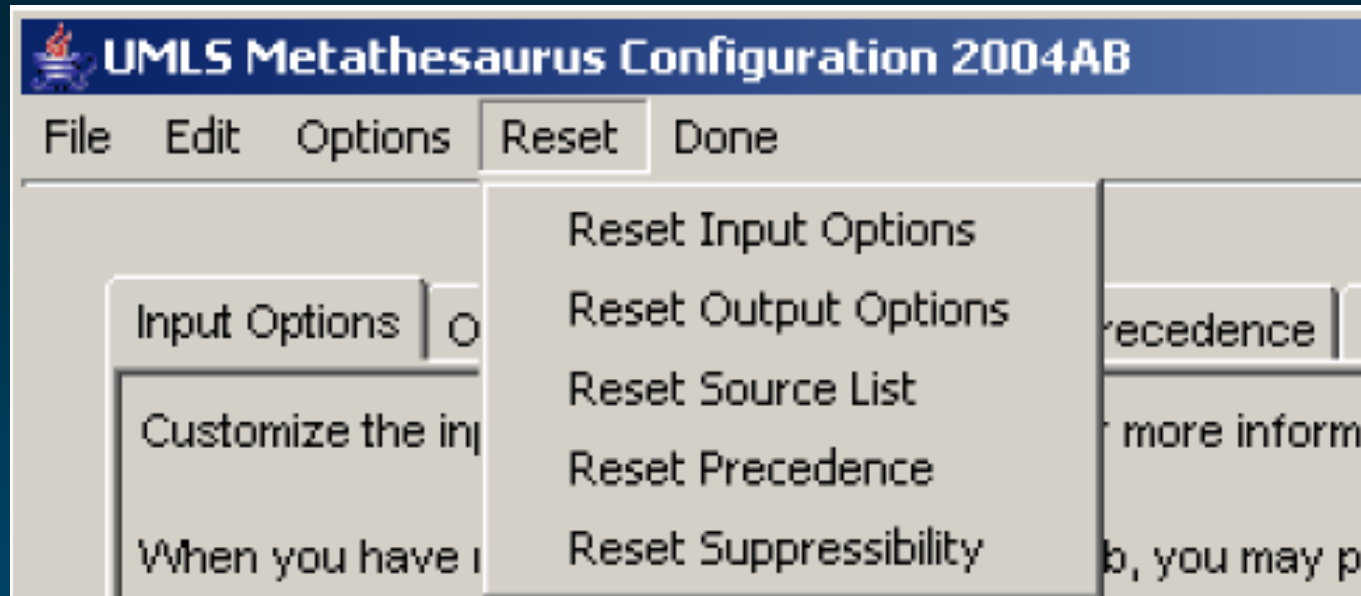
- Select Options → Advanced Suppressibility Options
- Select the term types to remove

Advanced Suppressibility Options

- Remove Source Term Type Suppressible Data.
- Remove Editor Assigned Suppressible Data.
- Remove Obsolete Data.

Done

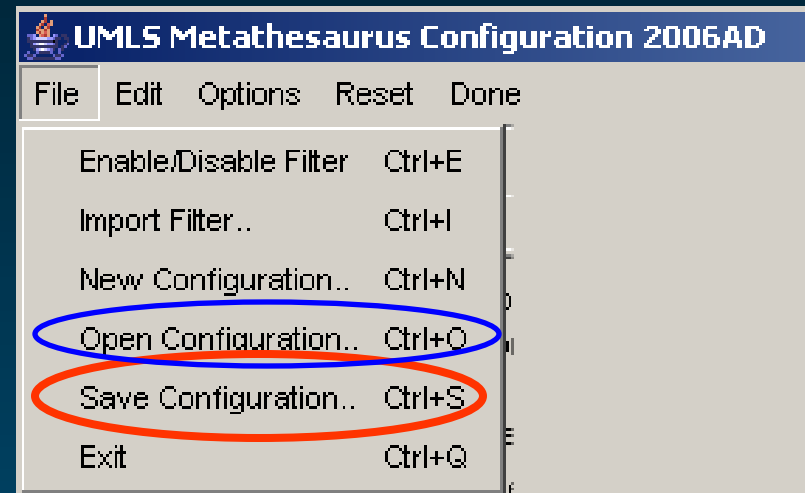
# Reset default MMSYS Options?



- Returns all filters to default selections
- Default selections in "mmsys.prop.default file" in config folder
- mmsys.prop.default contains properties in last run

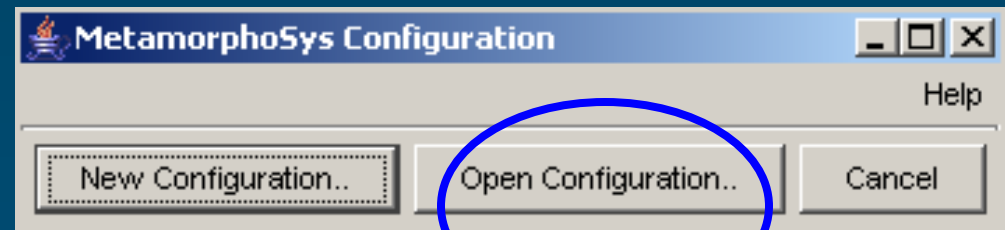
# Ensure all team members have the same subset?

- On File menu "Save Configuration"
- Share configuration file with team members
- Have team members select "Open Configuration" from File Menu



/or/

Select "Open Configuration" from Configuration Screen



# Search for a term in my RRF subset?

- Select "Browse my Subset" from welcome screen
- Browse to your subset location
- Search by term, string or CUI
- Reports include:
  - Hyperlinked concepts
  - Raw data view
  - Attributes and Relations

The screenshot displays a search interface with four tabs: CUI Search, Code Search, Word Search, and Tree Browser. The 'Word Search' tab is selected. Below the tabs is a search input field containing 'Addisons' and a 'Search..' button. Below the search field, it says 'Select a concept. (2 results)' with left and right arrow buttons. A list of results is shown, with 'C0001403 Addison's disease' selected. To the right, the 'Raw View' tab is active, showing details for the selected concept: 'Concept: [CUI C0001403] Addison's disease', 'Semantic Type: Disease or Syndrome', and 'Definitions: CSP/PT|disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin; due to tuberculosis or autoimmune induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol.'

# Search for a code in my RRF subset?

- Select "Browse my Subset" from welcome screen
- Browse to your subset location
- Select "Code search"

The screenshot shows the 'Rich Release Format Browser 2008AA C0001403' application window. The 'Code Search' tab is active, and the search input field contains 'D000224'. A red circle highlights the search input field and the 'Get Concepts..' button. Below the search input, the text 'Select a concept. (1 result)' is displayed, followed by a list of search results: 'C0001403 Addison's disease'. The right pane shows the details for the selected concept, including the 'Concept' name, 'Semantic Type' (Disease or Syndrome), and 'Definitions' (CSP/PT|disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin; due to tuberculosis or autoimmune induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol).

Rich Release Format Browser 2008AA C0001403

File Edit Options Help

Monospaced 11 C:\Documents and Settings\kleinsr\Desktop\2008AA\2008AA\META

CLI Search Code Search Word Search Tree Browser

Enter a CODE:

D000224

Get Concepts..

Select a concept. (1 result)

C0001403 Addison's disease

Raw View Report View

**Concept:** [CUI C0001403] Addison's disease

**Semantic Type**  
Disease or Syndrome

**Definitions**  
CSP/PT|disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin; due to tuberculosis or autoimmune induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol.

# More information and help

---

- MetamorphoSys Documentation

<http://www.nlm.nih.gov/research/umls/meta6.html>

- Readme file on the DVD or downloaded from the UMLSKS
- Help Menu from any page in MetamorphoSys

# UMLS Tools



# UMLS Knowledge Source Server

---

- Licensed users access online:  
<http://umlsks.nlm.nih.gov>  
Web search interface for the three Knowledge Sources (Metathesaurus, Semantic Network, Specialist Lexicon)
- Application Programming Interface (API)
- File downloads
- Documentation

# Create a UMLSKS Account

## Registered Users

Login ID:

Password:

**Login**

Need a UMLS License? Click [here](#)

[Forgot your password?](#)  
[Request an account \(New Users\)](#)  
[Request a developer certificate \(Existing Users\)](#)

## New UMLSKS Users

The UMLSKS is restricted to registered users. If you have not yet accepted the [UMLS license agreement](#), obtain a UMLS license. Once you have successfully created the account, you will be mailed a certificate. Please click on the 'Request a developer's certificate' link.

\* Login ID:

\* Password:

\* Confirm Password:

\* First Name:

\* Last Name:

\* Email:

\* Organization:  Please do not use any special characters including " , \*

City:

State:

Country:

\* UMLS License #:

# UMLSKS Search Options

Home Help Log Off

## UMLS Knowledge Source Server (UMLSKS)

Home Metathesaurus SPECIALIST Lexicon Semantic Network UMLS and Source View Tree Browser myUmlsks

You have been logged out

### Downloads

- UMLS Knowledge Sources
- RxNorm Files
- SNOMED CT ®
- Mappings
- VA/KP Problem List

### Home

#### Overview

Welcome to the new UMLS Knowledge Source Server (UMLSKS) version 6.0 (beta).

NLM is working to resolve a number of problems associated with this new implementation. We

#### What's New

- **Web Services Implementation of the API.**
- **Web Interface based on Portal technology.**

# UMLSKS v5 Search Options

## UMLS Knowledge Source Server (UMLSKS)

6.0 UMLS Releases: 2006AA 2006AB 2006AC 2006AD 2007AA 2007AB 2007AC 2008AA 2008AB

[Metathesaurus](#)

[Semantic Network](#)

[SPECIALIST Lexicon](#)

[Home](#)

[Advanced Search](#)

[Logout](#)

### Quick Search

Select UMLS

Release:

2008AB

Enter search

value:

[Metathesaurus  
Concept  
Search](#)

[Semantic  
Network  
Search](#)

[SPECIALIST  
Lexicon  
Search](#)

[Search](#)

[Tips...](#)

[Search](#)

[Tips...](#)

[Search](#)

[Tips...](#)

# Search Options

Home Metathesaurus SPECIALIST Lexicon Semantic Network UMLS and Source View Tree

Choose a Section: Metathesaurus

- Metathesaurus:
- Contexts
- Relations
- Raw View
- Co-occurrence Info
- Lexical View

Downloads

### Metathesaurus Search

Enter term

Input type:  Term  CUI  Code

Term:  Release: 2008AA Index: Normalized String

OK

Sources: All sources

SPECIALIST Lexicon Semantic Network UMLS and

### SPECIALIST Lexicon

#### Lexicon Search

Enter term:

OK

SPECIALIST Lexicon Semantic Network UMLS and Source View Tree

### Semantic Network

#### Semantic Network Search

Please select a Semantic type from the list below to find more information

Acquired Abnormality

Submit Query

# UMLSKS: Downloading files

## UMLS Knowledge Sources: File Downloads

<b>2008AA UMLS Files</b> <b>April 01, 2008 10:00:00 ET</b>
<a href="#">2008AA.CHK</a>
<a href="#">2008AA.MD5</a>
<a href="#">2008aa-1-meta.nlm</a>
<a href="#">2008aa-2-meta.nlm</a>
<a href="#">2008aa-otherks.nlm</a>
<a href="#">mmsys.zip</a>
<a href="#">Copyright Notice.txt</a>
<a href="#">README.txt</a>

Download Notes:

### Downloads

UMLS Knowledge Sources  
RxNorm Files  
SNOMED CT ®  
Mappings  
VA/KP Problem List

## SNOMED CT® Release Files

Download SNOMED CT Releases (English and Spanish) from this page. Click on the file name to start the download.

### SNOMED CT: International Release July 2008

File Name	Contents	Notes
<a href="#">SNOMED_CT_Essential_20080731.zip (136M)</a>	SNOMED CT terminology, cross maps, and subsets. User Guide (content, principles and uses); Technical Implementation Guide (design of applications using SNOMED CT) and Technical Reference Guide (file layouts, field sizes, required values, data diagrams)	SNOMED CT terminology also available in Metathesaurus format in UMLS 2008AB Release, November 2008
<a href="#">SNOMED_CT_Optional_20080731.zip (2.8M)</a>	ICD-9-CM cross map and US drug extensions	ICD-9-CM cross map also available in Metathesaurus format in UMLS 2008AB Release, November 2008

# UMLSKS v5: Downloading files

**About the UMLSKS**

- ▶ [Home](#)
- ▶ [Overview](#)
- ▶ [Frequently Asked Questions](#)
- ▶ [Edit Views/Profile](#)

**Downloads**

- ▶ [UMLS Knowledge Sources](#)
- ▶ [RxNorm Files](#)
- ▶ [SNOMED CT®](#)
- ▶ [Mappings](#)
- ▶ [VA/KP Problem List](#)
- ▶ [Developer's API](#)

**Documentation**

## UMLS Knowledge Sources: File Downloads

### 2008AB UMLS Files

November 13, 2008 11:00:00 ET

[2008AB.CHK](#)

[2008AB.MD5](#)

[2008AB-1-meta.nlm](#)

[2008AB-2-meta.nlm](#)

[2008AB-otherks.nlm](#)

[mmsys.zip](#)

[Copyright Notice.txt](#)

[README.txt](#)

# The SPECIALIST Lexicon and Lexical Tools

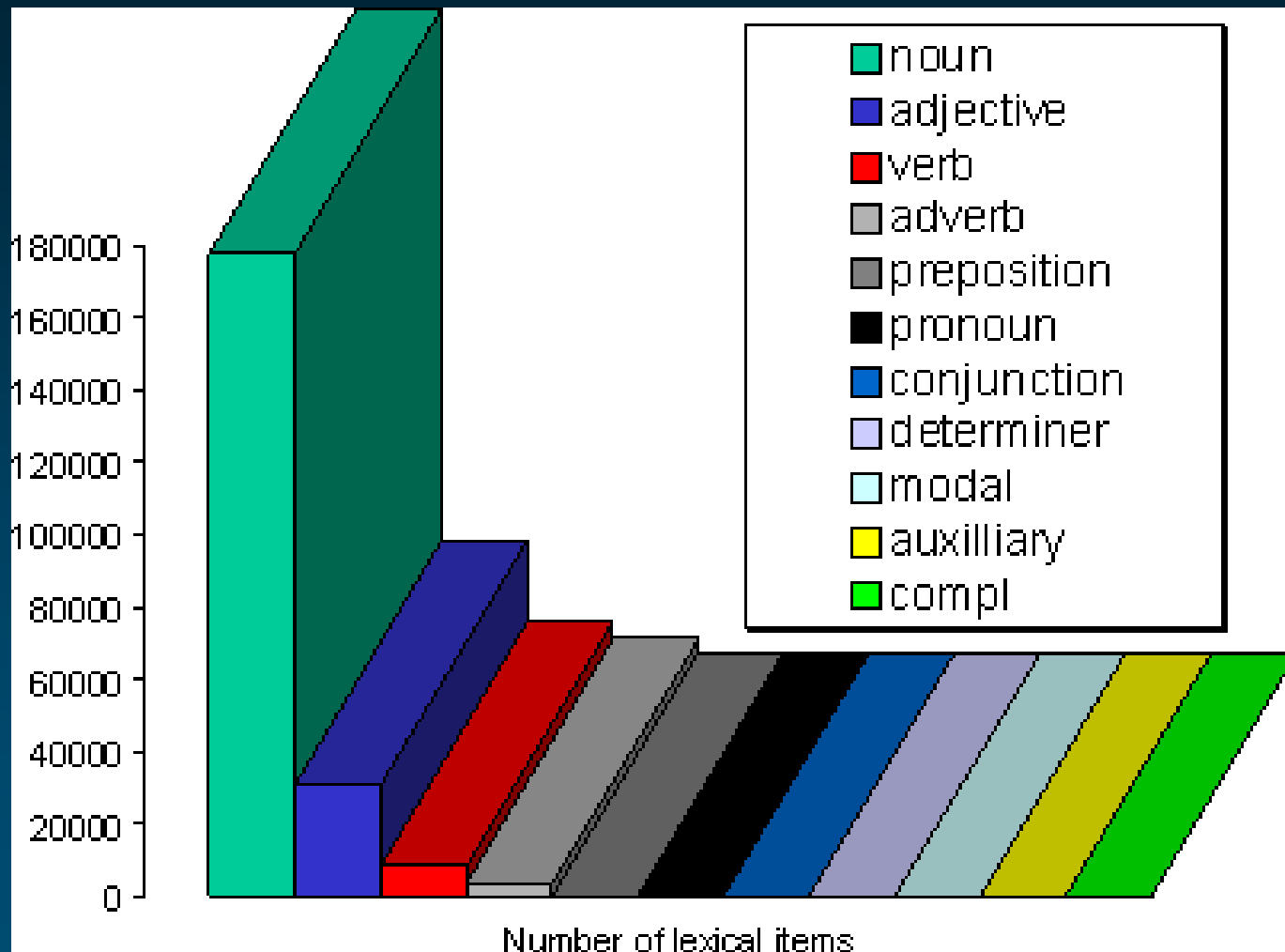


# SPECIALIST Lexicon

---

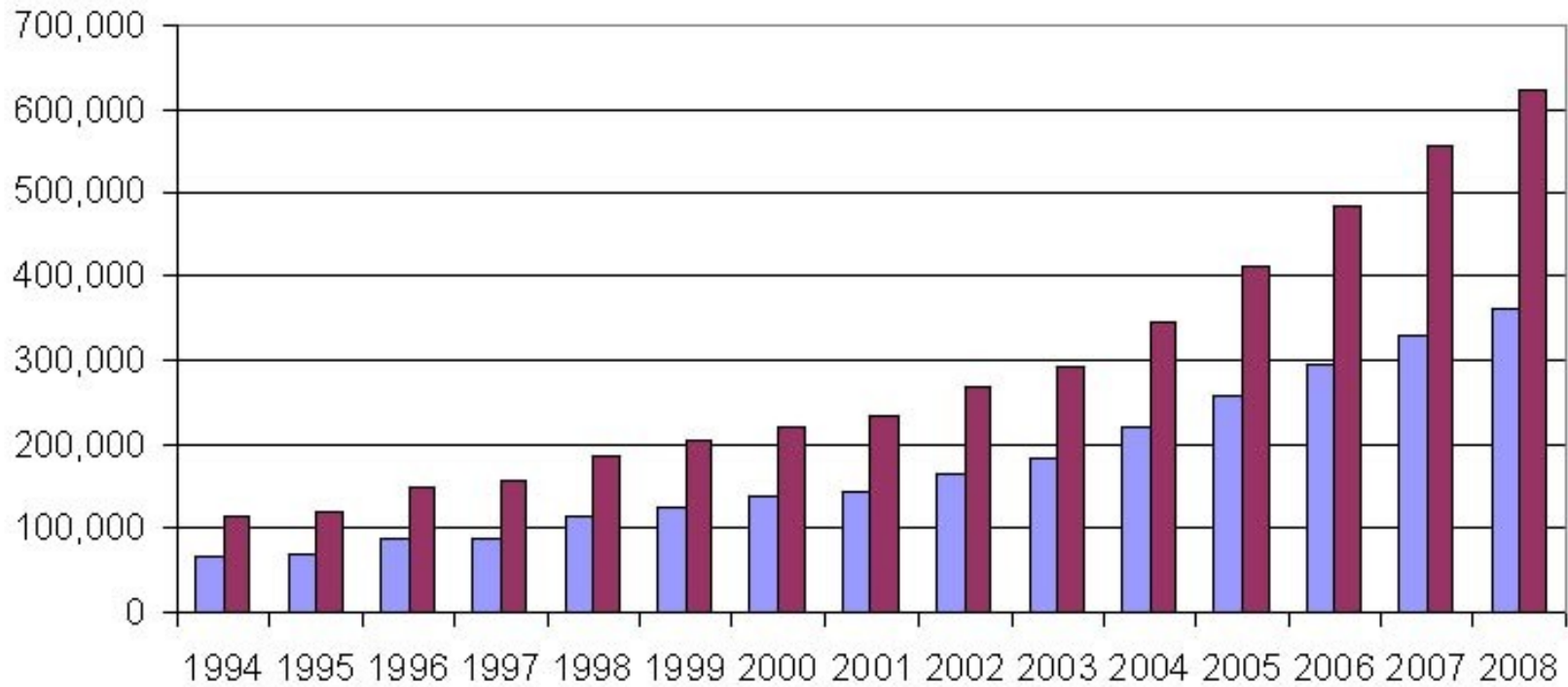
- Syntactic English lexicon of common words, biomedical terms (330K+ words, 550K+ variants)
  - Word properties
    - Syntax (how words are put together)
    - Morphology (inflection, derivation, and compounding)
    - Orthography (spelling)
  - Used by SPECIALIST Natural Language Processing System to process text and terms
    - Customizable
    - Used to maintain Metathesaurus, indexes
- ⑩ Adrenal gland diseases
  - ⑩ Diseases of the adrenal glands
  - ⑩ Adrenal disorder
  
  - ⑩ Disorder of adrenal gland
  - ⑩ C0001621

# SPECIALIST Lexicon by part of speech



# Specialist Lexicon Growth

## SPECIALIST LEXICON Growth



Year

Lexical Items

Inflected Forms

# Lexicon Growth by Year

---

<b>Release</b>	<b>Items</b>	<b>Forms</b>	<b>New Items</b>	<b>New Forms</b>
<a href="#"><u>2008</u></a>	360,688	625,041	30,233	67,644
<a href="#"><u>2007</u></a>	330,455	557,397	33,141	75,169
<a href="#"><u>2006</u></a>	297,314	482,268	40,515	69,755
<a href="#"><u>2005</u></a>	256,799	412,473	35,916	68,490
<a href="#"><u>2004</u></a>	220,883	343,983	37,582	51,004
<a href="#"><u>2003</u></a>	183,301	292,979	19,413	24,187

# SPECIALIST Lexicon Lexical Entry

---

```
{base=disease
  entry=E0023270
  cat=noun
  variants=reg
  variants=uncount
  compl=pphr(of,np|bone|)
  compl=pphr(of,np|breast|)
  compl=pphr(of,np|liver|)
  compl=pphr(of,np|ovary|)}
```

⑩ Base form

⑩ Unique identifier

⑩ Part of speech

⑩ Lexical variants

⑩ Prepositional phrase  
complements

# Orthography

---

- Spelling variants

- oe/e

- ae/e

- ise/ize

- genitive mark

- British-American variants

- ⑩ oesophagus - esophagus

- ⑩ anaemia - anemia

- ⑩ cauterise - cauterize

- ⑩ Addison's disease

- ⑩ Addison disease

- ⑩ Addisons disease

- ⑩ criticise -- criticize

- ⑩ centre -- center

- ⑩ foetus -- fetus



# SPECIALIST Lexicon

# lexical records

```
{base=Kaposi's sarcoma  
spelling_variant=Kaposi  
sarcoma  
entry=E0003576
```

```
    cat=noun  
    variants=uncount  
    variants=reg  
    variants=glreg
```

```
}  
{base=aspirate  
entry=E0010803
```

```
    cat=verb  
    variants=reg  
    tran=np
```

```
nominalization=aspiration|noun|E0010804
```

```
}
```

```
{base=chronic  
entry=E0016869
```

```
    cat=adj  
    variants=inv  
    position=attrib(1)  
    position=pred  
    stative
```

```
}
```

```
{base=deja vu  
spelling_variant=deja-vu  
spelling_variant=déjà vu  
entry=E0021340
```

```
    cat=noun  
    variants=uncount
```

```
}
```

# Lexical Tools

---

- Manage lexical variation in biomedical terminologies and text
- Used separately or with SPECIALIST Lexicon
- Perform transformations selected and ordered by users
- 3 primary programs: normalizer, word index generator, lexical variant generator



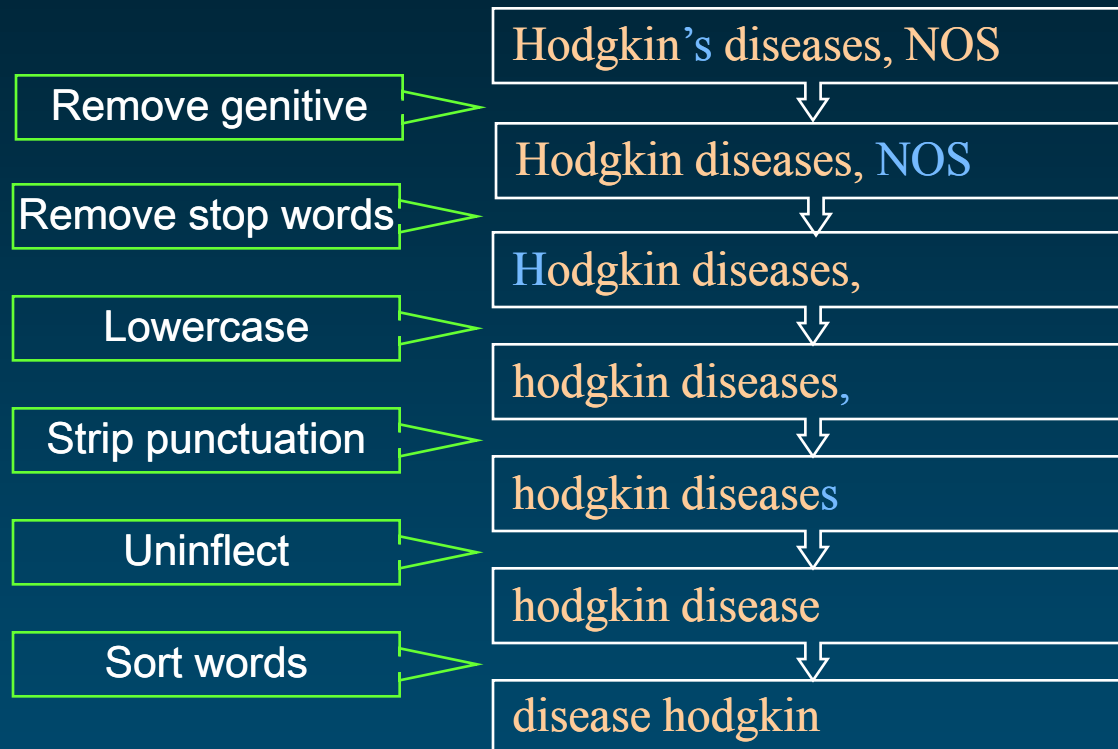
# Lexical Tools

---

- Wordind
  - breaks strings into words
  - produces the Metathesaurus word indexes (MRXW)
- lvg
  - performs various lexical transformations
  - 58 flow components and 32 options in 2005
- NORM
  - a selection of LVG transformations
  - used for Metathesaurus indexing
  - produces Metathesaurus normalized word and string indexes (MRXNW & MRXNS)
  - used to access those indexes

# Normalization

---



# Normalization: Example

Hodgkin Disease  
HODGKINS DISEASE  
Hodgkin's Disease  
Disease, Hodgkin's  
Hodgkin's, disease  
HODGKIN'S DISEASE  
Hodgkin's disease  
Hodgkins Disease  
Hodgkin's disease NOS  
Hodgkin's disease, NOS  
Disease, Hodgkins  
Diseases, Hodgkins  
Hodgkins Diseases  
Hodgkins disease  
hodgkin's disease  
Disease, Hodgkin



Normalized term is not necessarily readable

# Lexical Systems Group

---

Official Websites	<a href="http://SPECIALIST.nlm.nih.gov">http://SPECIALIST.nlm.nih.gov</a> <a href="http://mmtx.nlm.nih.gov">http://mmtx.nlm.nih.gov</a>
Mail Addresses	Allen Browne - <a href="mailto:browne@nlm.nih.gov">browne@nlm.nih.gov</a> Guy Divita - <a href="mailto:divita@nlm.nih.gov">divita@nlm.nih.gov</a> Chris Lu - <a href="mailto:lu@nlm.nih.gov">lu@nlm.nih.gov</a> Lexical Systems group - <a href="mailto:umlslex@nlm.nih.gov">umlslex@nlm.nih.gov</a>

# MetaMap/MMTx

---

- To map biomedical text to concepts in the UMLS Metathesaurus
- To find Metathesaurus concepts in text
- MMTx was created to provide a distributable version of MetaMap

# The MetaMap/MMTx Algorithm

---

- Parsing
  - Using SPECIALIST minimal commitment parser, SPECIALIST lexicon, a part of speech tagger
- Variant generation
  - Using SPECIALIST lexicon, Lexical Variant Generation (LVG)
- Candidate retrieval
  - From the Metathesaurus
- Candidate evaluation
- Mapping construction

# MetaMap/MMTx Example

● Text: Termination of clinical trials: the beta-blocker heart attack trial ...

● Concepts

- Termination
- Clinical Trials
- Adrenergic beta-Antagonists
- Myocardial Infarction

⑩ Heart attack (Myocardial Infarction)

⑩ Clinical Trials  
⑩ Trial (Clinical Trials)

⑩ Beta-blocker (Adrenergic beta-Antagonists)

# Semantic Network



# Semantic Network

---

- **Semantic Types**

- 135 broad subject categories (Clinical Drug, Virus)
- Organized into 2 hierarchies: entity, event
  - Addison's Disease (concept)
  - Semantic Type: Disease or Syndrome

- **Semantic Relationships**

- 54 possible, useful links between categories
- Hierarchical: isa
- Associative, e.g. causes
  - Virus **causes** Disease or Syndrome

- **Types + Relationships = Semantic Network**

- <http://www.nlm.nih.gov/research/umls/meta3.html>

# Why have a Semantic Network?

---

- Semantic Types
  - **High level categories** assigned to Metathesaurus concepts
  - Independent of position in source hierarchies
- Semantic Relations
  - **Useful links between** Semantic Types
  - Relationships may hold at the concept level
  - Other relationships may apply at the concept level

# Semantic Types categorize

---

- Concept: Discipline of Nursing
- Semantic Type: Biomedical Occupation or Discipline
  
- Concept: Nursing Homes
- Semantic Type: Health Care Related Organization  
Manufactured Object
  
- Concept: Home Nursing
- Semantic Type: Health Care Activity

# Semantic Type: Disease or Syndrome

---

- Natural Phenomenon or Process

- Biologic Function

- Physiologic Function

- Organism Function

- » Mental Process

- Organ or Tissue Function

- Cell Function

- Molecular Function

- » Genetic Function

- Pathologic Function

- ***Disease or Syndrome***



- Mental or Behavioral Dysfunction

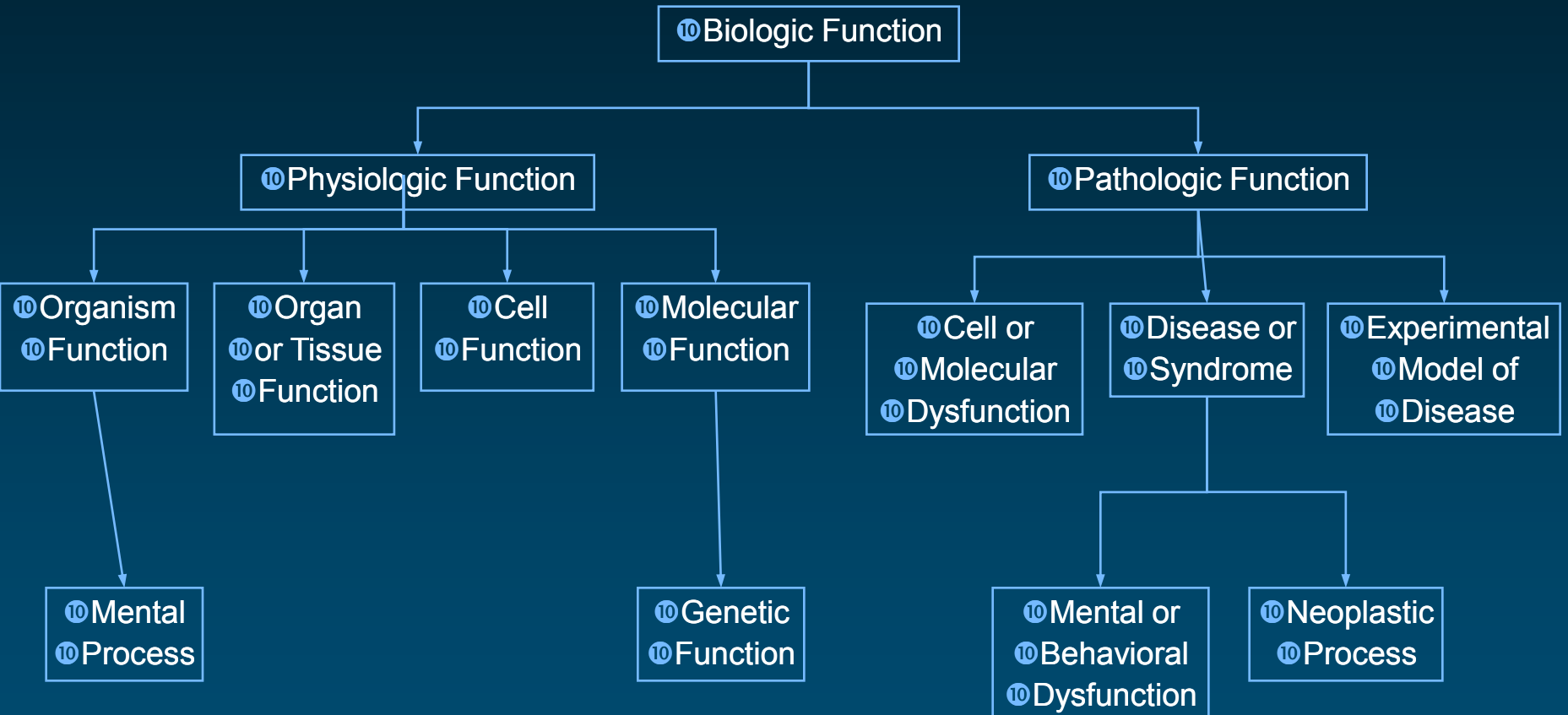
- Neoplastic Process

# 54 Semantic Relationships

---

- **Hierarchical** (isa = is a kind of)
  - among types
    - Animal *isa* Organism
    - Enzyme *isa* Biologically Active Substance
  - among relationships
    - prevents *isa* affects
- **Non-hierarchical**
  - Sign or Symptom *diagnoses* Pathologic Function
  - Pharmacologic Substance *treats* Pathologic Function

# "Biologic Function" hierarchy (isa)



# Semantic Relationships

---

Define useful relations between types

Biomedical Occupation or Discipline

**ASSOCIATED\_WITH** Social Behavior

Language

**ISSUE\_IN** Biomedical Occupation or  
Discipline

Manufactured Object

**CAUSES** Injury or Poisoning

Diagnostic Procedure

**USES** Manufactured Object

Health Care Activity

**ISA** Occupational Activity

Health Care Activity

**AFFECTS** Disease or Syndrome

# Semantic Relationships

---

Disease or Syndrome **associated\_with** Finding

Disease or Syndrome **result\_of** Pathologic Function

Body Part, Organ, or Organ Component **location\_of**  
Disease or Syndrome

Hormone **affects** Disease or Syndrome

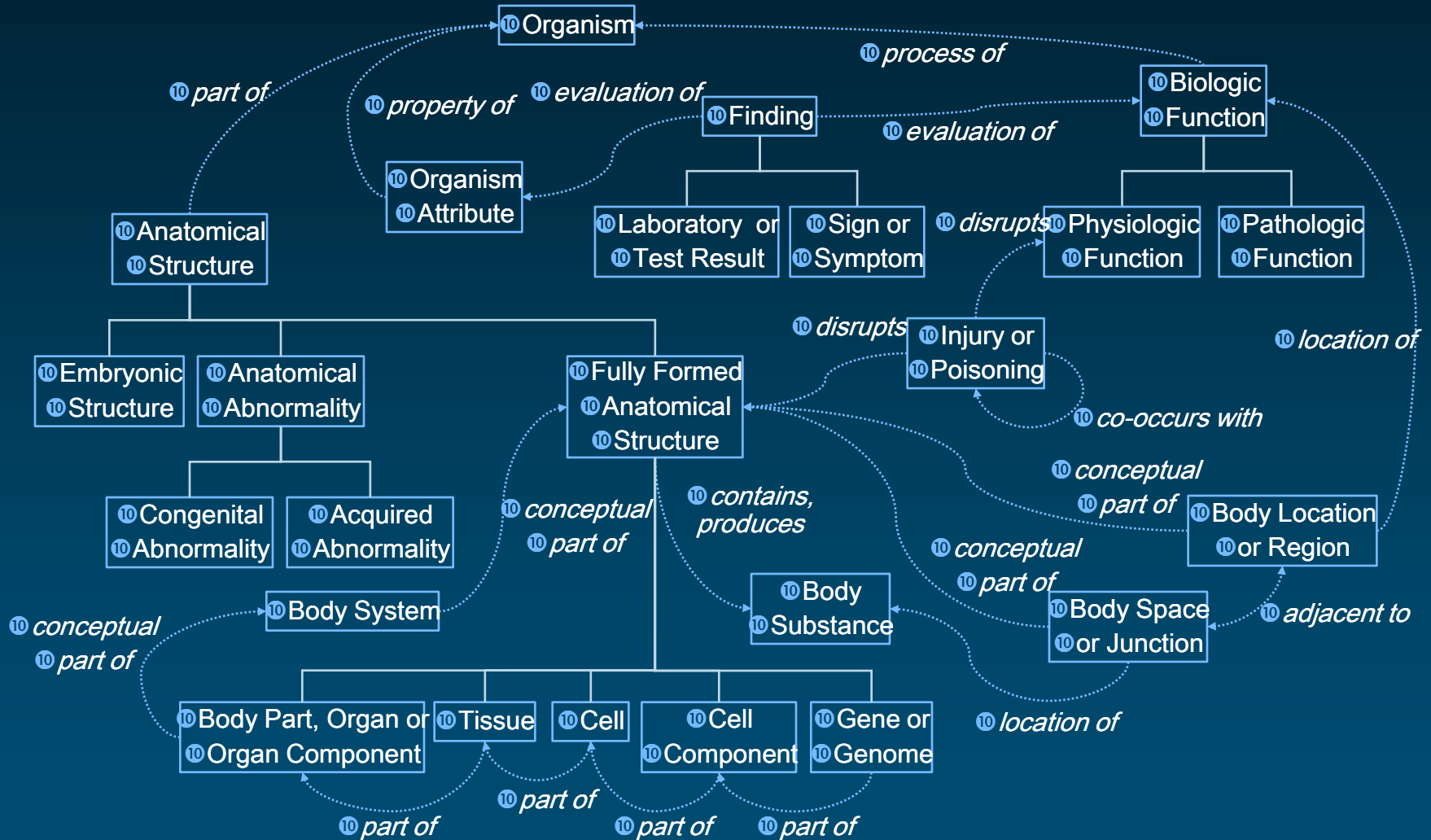
Hormone **complicates** Disease or Syndrome

Laboratory Procedure **diagnoses** Disease or Syndrome

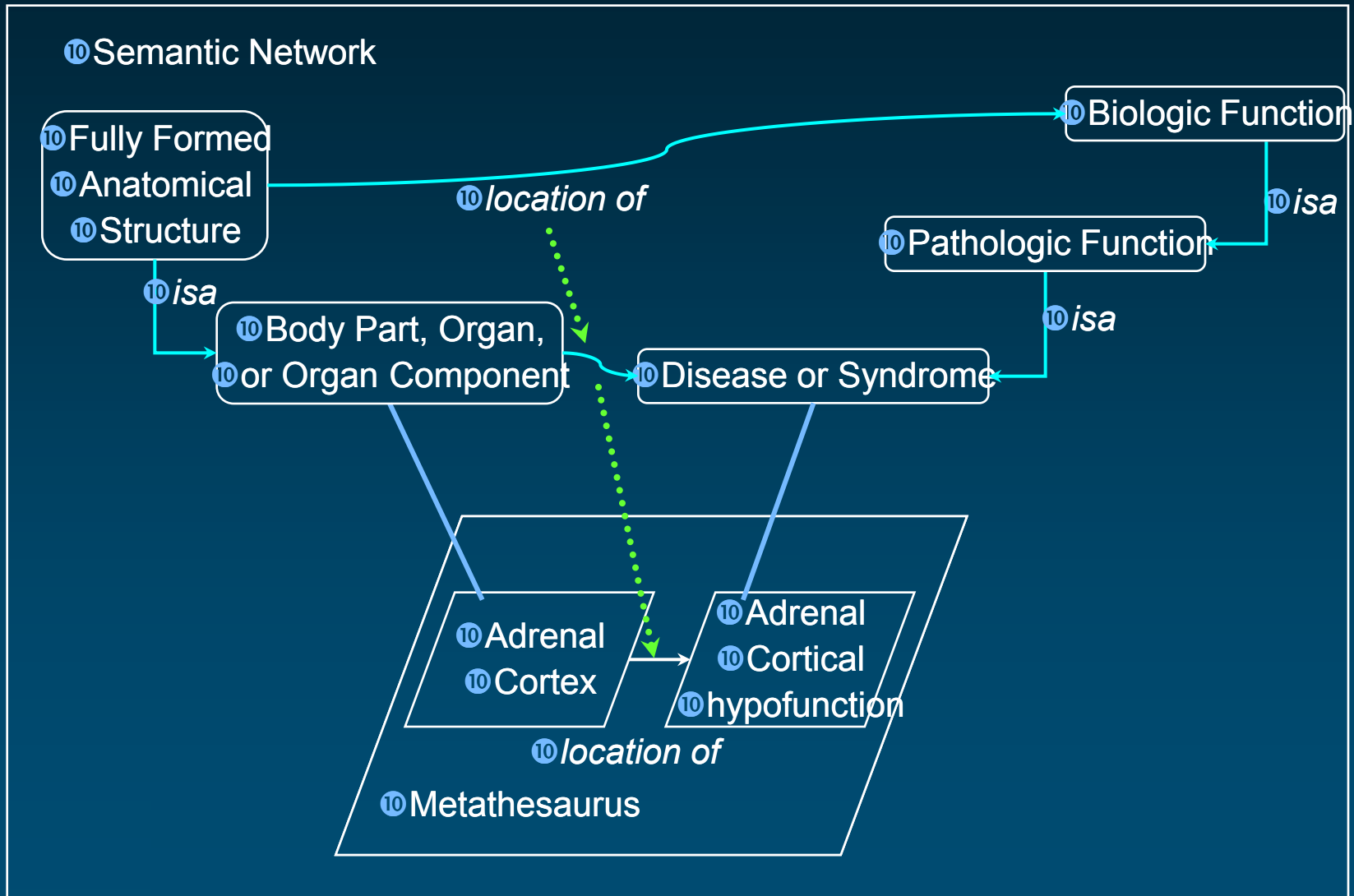
Pharmacologic Substance **treats** Disease or Syndrome



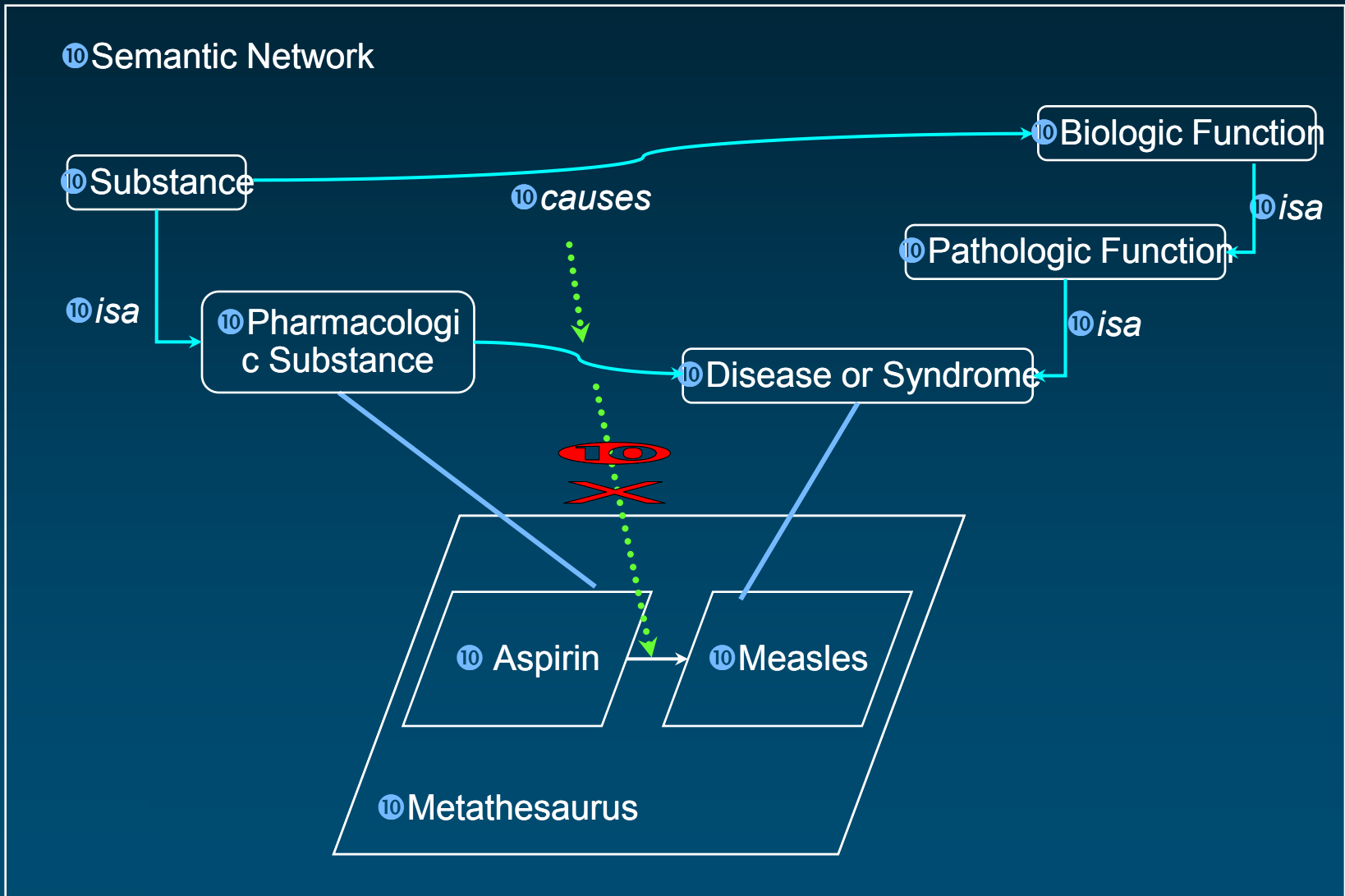
# Associative (non-isa) relationships



# Relationships can inherit semantics



# Relationships don't always inherit semantics



# Why have a Semantic Network?

---

- Semantic Type serves as a **high level category** assigned to each Metathesaurus concept, *independent of its position in its original source hierarchy*
- A Semantic Relation is a **possible link between 2 concepts** assigned those Semantic Types
  - The relationship may hold at the concept level
  - Other relationships may apply at the concept level

# UMLS and Health Data Standards

# Toward a nationwide system of health information

- Adoption
  - Incentives, Regulatory Reform, Workforce Needs and Impacts, Public Awareness
- Interoperability
  - Product Certification
  - Data Standards
  - Standard Product Identifiers and Vocabulary
  - Drug Records
- Connectivity
  - Privacy Standard and Criminal Sanctions, Nationwide Health Information Network (NHIN), Consumer Protections

# Electronic Health Records (EHR)

---

- National collaboration toward secure, Internet-accessible EHR in 10 years
- Federal and private-sector to develop standards with agreement from all stakeholders
- Electronic health records and other information technology to transform health care system
  - reduce medical errors, improve quality of care
  - minimize paperwork hassle, lower costs

Secretary Leavitt Takes News Steps to Advance Health IT, June 15, 2005 ([www.hhs.gov/healthit](http://www.hhs.gov/healthit))

# The call for standards

---

- 1991 – IOM: *The Computer-Based Patient Record*
- 1995 – PHS: *Making a Powerful Connection*
- 1996 – Congress: *HIPAA*
- 2000 – IOM: *To Err is Human*
- 2000 – NCVHS: *Report on Standards for PMRI*
- 2001 – IOM: *Crossing the Quality Chasm*
- 2002 – NCVHS: *PMRI recommendations (1<sup>st</sup> Set)*
- 2002 – AMIA Congress: *Drug Nomenclature Information*
- 2002 – AAMC conference: *IT Enabling Clinical Research*
- 2003 – CHI eGov: *U.S. gov't target standards (1<sup>st</sup> set)*
- 2003 – Connecting for Health: *Standards Wk.Gr.Report*



# Strategy for U.S. health data standards

---

- Establish mechanisms for designating U.S. standards
  - ✓ Pick best available as starting point
  - ✓ Support development, maintenance, distribution
- Coordinate development of selected standards to achieve non-overlapping, interlocking set
- Broaden participation in standards development
- Promote use and improvement

# Federal standards selection mechanisms

---

- **HIPAA** (1996) - Health Insurance Portability and Accountability Act of 1996 requires administrative standards
- **NCVHS** - National Committee on Vital and Health Statistics, a long-standing (50+ years) advisory committee to HHS expanded by HIPAA, recommends standards
- **CHI** (2001) Consolidated Health Informatics project, a cross-agency eGov initiative designates U.S. gov't-wide clinical data standards
- **Medicare Modernization Act** (2003) - requires e-Prescribing standards; establishes Commission on Systemic Interoperability

# Standards implementation

---

- **CHI (clinical)** *LOINC*
  - E.g., lab test results, problems, diagnoses, history, physical
  - Electronic exchange of clinical health information in U.S. Government systems
- **HIPAA (administrative)** *CPT*
  - e.g., health insurance claims, billing, ordering
  - HIPAA Administrative Simplification provisions
  - Designated DHHS national standards for electronic healthcare transactions
- **PHIN (public health)** *ICD-9-CM*
  - e.g., disease surveillance, immunization rates, environmental monitoring
  - CDC designated standards for public health reporting

# NLM role

---

- HHS central coordinating body for Patient Medical Record Information (PMRI) terminologies
- Work with agencies, organizations to support standards
- Support mappings between standard clinical vocabularies (SNOMED, LOINC, RxNorm) and others
- Reduce overlap between standards vocabularies

# NLM role 2

---

- Maintain UMLS as uniform distribution mechanism for
  - HIPAA code sets, recommended PMRI terminologies
  - **mappings** between standard codes and PMRI terminologies
  - information on valid vocabularies for transaction standards

# Health Information Technology Homepage

Health Information Technology and Health Data Standards at NLM - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.nlm.nih.gov/healthit.html>



United States  
**National Library of Medicine**  
National Institutes of Health

Search NLM Web Site

Go

[NLM Home](#) | [Contact NLM](#) | [Site Map](#) | [FAQs](#)

[Home](#) > [Health Information Technology](#)

## Health Information Technology and Health Data Standards at NLM

[Printer-friendly Version](#)

**NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS). NLM works closely with the Office of the National Coordinator for Health Information Technology (ONC) to ensure NLM's efforts are aligned with the goal of the President and HHS Secretary for the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care.**

### Health Data Standards

NLM supports the development, enhancement, and distribution of clinically specific vocabularies to facilitate the exchange of clinical data and improve retrieval of health information.

- Clinical Vocabularies supported, licensed, or developed by NLM:  
[SNOMED CT](#) | [LOINC](#) | [RxNorm](#)
- Uniform distribution mechanism for HIPAA and clinical vocabulary standards through the [UMLS Metathesaurus](#)
- [UMLS Enhanced VA/KP Problem List Subset of SNOMED](#)

NLM distributes FDA approved Structured Product Labels linked to RxNorm and related medication information:

### Related Sites

[HealthIT](#) (U.S. Government Health Information Technology Web Site)

[Office of the National Coordinator for Health Information Technology \(ONC\)](#)

[AHRQ National Resource Center for Health IT](#)

[USHIK](#) (United States Health Information Knowledgebase)

### NIH Initiatives

# International Health Terminology Standards Development Organisation (IHTSDO)

---

- National Library of Medicine (NLM) is U.S. member



- Promotes SNOMED CT as one of key US standards
- Distributes SNOMED CT

# SNOMED CT in the UMLS

---

- English descriptions and relationships (2004)
  - Concepts: 298,090
  - Descriptions: 736,946 (synonyms)
  - Relationships: 1,315,910
- Added to Metathesaurus:
  - Concepts: 37,089
  - Descriptions: 350,464
- Spanish terms added in 2004AB



# SNOMED CT in the Metathesaurus

---

- Core content U.S. English, Spanish
  - Concepts, Descriptions, Relationships tables
  - Generic drugs
  - History table
  - ICD-9-CM mapping
- SNOMED updates in UMLS
  - January SNOMED CT update → Spring UMLS
  - July SNOMED CT update → Fall UMLS

# Additional information on NLM Web site

---

- [SNOMED CT® in the UMLS® Metathesaurus®: Inversion Source Transparency Achieved Jan. 2005](#)
- [SNOMED CT in the 2005AB Release](#)
- [SNOMED CT® in the UMLS® Metathesaurus®:Release Source Transparency Achieved Jan. 2005](#)
- [SNOMED CT in the 2005AB Release](#)
- [SNOMED CT Information from the College of American Pathologists](#)

RxNorm

# RxNorm Project

---

- Developed by NLM to address
  - missed synonymy in UMLS clinical drugs
  - medication errors
- CHI recommended standard
- Intended to support
  - effective sharing of drug data across systems
  - electronic health record (EHR)
  - computerized physician order entry (CPOE)

# RxNorm Contents

---

- Full sources
  - FDA, NDDF, Medispan, Micromedex, Multum
  - More to come
- Partial sources
  - SNOMED CT, MeSH
- NOT in RxNorm
  - USP Medicare Model Guidelines
  - MedDRA (adverse reactions)

RxNorm (NLM)  
SNOMED CT (CAP)  
MTHFDA (FDA NDC)  
VANDF (VA)  
Gold Standard  
NDDF (FirstDataBank)  
MDDDB (Medispan)  
MMX (Micromedex)  
Multum

# RxNorm Release Files

---

- RxNorm (standalone vocabulary)
  - Normalized forms of clinical drugs
  - Created by NLM
  - From commercial and government clinical drug vocabularies
  - Full monthly, and weekly additional updates
- RxNorm subset (extracted from UMLS)
  - RxNorm forms (NLM)
  - Clinical drug content including SNOMED CT, VANDF, FDA NDC

Respect license restrictions

# RxNorm Normalized Forms

---

- Semantic Normal Form (SNF)
  - Active standardized (generic) ingredients
  - Strengths
  - Units of measurement
  - Dosage form
- As administered to a patient
  - Acetaminophen 500 MG Oral Tablet
  - Acetaminophen 500 MG Oral Tablet [Tylenol]

# RxNorm Term Types

---

- SBD = Palifermin 5 mg/ml Injectable Solution [Kepivance] → 7 additional concepts
- IN= Palifermin
- BN= Kepivance
- SBDC= Palifermin 5 mg/ml [Kepivance]
- SBDF = Palifermin Injectable Solution [Kepivance]
- SCD= Palifermin 5 mg/ml Injectable Solution
- SCDC= Palifermin 5 mg/ml
- SCDF= Palifermin Injectable Solution

all are separate RxNorm concepts



# Mappings

# Mappings

---

- Synonymy within concept (implied)
- Explicit links between concepts, terms, codes or other content in specified vocabularies
- Source-asserted
  - SNOMED CT → ICD-9-CM
- Created/supported by NLM
- Represented in
  - MRREL (simple only)
  - MRMAP, MRSMAP (simple and complex)

# UMLS – MeSH mapping file

---

Used in MEDLINE/PubMed searching

Based on synonymy

...

myocardial infarction|attack coronary

myocardial infarction|attack heart (nos)

myocardial infarction|cardiac infarction

myocardial infarction|cardiac infarction, nos

myocardial infarction|cardiac; infarction

myocardial infarction|heart attack

myocardial infarction|heart attack, nos

myocardial infarction|heart attacks

myocardial infarction|heart infarction

# Other Types of Mappings

---

## LOINC - CPT

- 1795-4 AMYLASE:CCNC:PT:FLU:QN →
- 1798-8 AMYLASE:CCNC:PT:SER:QN →
- 1799-6 AMYLASE:CCNC:PT:UR:QN →
- 82150 Amylase (blood or urine)

## SNOMED CT – CPT

- 313500004 Urine protein/creatinine ratio measurement →
- 82570 Creatinine; other source
- 84155 Protein, total, except by refractometry; serum

# Mapping projects planned/underway

---

- CHI standards → HIPAA code sets
  - SNOMED CT → ICD-9-CM (support reimbursement)
  - SNOMED CT → ICD-10-CM
  - SNOMED CT → CPT/HCPCS
  - LOINC → CPT
- SNOMED CT →
  - MedDRA
  - MeSH
  - Nursing (NIC, NOC, NANDA)

# Key NLM Assumptions about Mappings

---

- Participants must include **producers** on both ends and **users**
- Mappings may prompt changes to content and adjustment to update schedules
- Mappings must keep current at both ends
- Mappings will be distributed in the UMLS with license terms applicable to both ends (MRMAP, MRSMAP)
- Mapping is an R & D problem; iteration is required to build highly functional maps
- Webcast:  
<http://www.nlm.nih.gov/research/umls/umlswebcasts/index.html>

Questions?

# Summary

---

- UMLS = 3 knowledge sources
  - Metathesaurus
  - Semantic Network
  - SPECIALIST Lexicon and Lexical Tools
- UMLSKS
  - Remote access and browser
  - Download files
  - Documentation
- MetamorphoSys
  - Installs UMLS, customizes Metathesaurus
  - RRF Browser



# Bibliography of UMLS Articles

---

- Search MEDLINE®/PubMed® for recent research journal articles on the UMLS (1997 to present)
- Comprehensive Bibliography 1986-1996
  - Selden, Catherine  
Unified Medical Language System (UMLS): January 1986 through December 1996 : 280 citations / prepared by Catherine R. Selden, Betsy L. Humphreys. -- Bethesda, Md. (8600 Rockville Pike, Bethesda 20894) : U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine, Reference Section ; Pittsburgh, PA : Sold by the Supt. of Docs., U.S. G.P.O., 1997.  
**-- (Current bibliographies in medicine ; 96-8)**

# References

---

- UMLS as a research project
  - Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods Inf Med*, 32(4), 281-91.
  - Humphreys, B. L., Lindberg, D. A., Schoolman, H. M., & Barnett, G. O. (1998). The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc*, 5(1), 1-11.
- Short presentation
  - Bodenreider, O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*, 32(Database issue), D267-70.

# UMLS Documentation and Support

---

- UMLS homepage
  - <http://umlsinfo.nlm.nih.gov/>
- UMLSKS homepage
  - <http://umlsks.nlm.nih.gov>
- UMLSUSERS-L
  - subscribe to discussion list
- NLM Customer Service email:
  - [custserv@nlm.nih.gov](mailto:custserv@nlm.nih.gov)

# Explore

---

- Register: sign the license agreement
- Create UMLSKS account
- Explore Knowledge Sources
- Download files or request DVD
- Create subsets using MetamorphoSys





Center  
Drive

Library  
of  
Medicine

Thank you