

Unified Medical Language System®

SNOMED CT® and the UMLS®

7th Annual SNOMED® International Users Group September 9, 2005

Tammy Powell Jan Willis National Library of Medicine National Institutes of Health

www.nlm.nih.gov







Outline

- UMLS overview
- UMLS license agreement
- Accessing the Metathesaurus
- SNOMED CT® in the Metathesaurus
- Mappings
- RxNorm



What is the UMLS?

- Unified Medical Language System
- Objective: Help interpret and understand medical meaning across systems
- Overcome:
 - disparities in language
 - disparities in granularity and perspective
 - problems in mapping and aggregating information across systems



The UMLS consists of ...

Metathesaurus



Biomedical Concepts

Semantic Network



Broad Categories + Relationships

SPECIALIST Lexicon & Lexical Tools



Lexical information, language processing programs

... 3 Knowledge Sources used together or separately



Knowledge Sources are ...

- Machine-readable data files (relational tables)
 - Records (rows)
 - Fields (columns)
- Applications
 - MetamorphoSys (installation and customization)
 - RRF Browser (view)
 - Ivg (lexical programs)

... sets of data files and programs



the UMLS is not an end-user application

UMLS Uses

- Information retrieval
- Thesaurus construction
- Natural language processing
- Automated indexing
- Electronic health records (EHR)
- Distribution mechanism for U.S. vocabulary standards



License Agreement

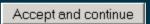


First step license agreement

Online Web-based license:

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

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







Submit

- Verification and turnaround:
 - Receive e-mail from NLM and respond within 72 hours
 - NLM official countersigns, license added to database
 - Receive 2nd e-mail from NLM with new license number



- 11. Some of the Material in the UMLS Metathesaurus is from copyrighted sources. If LICENSEE uses any material from copyrighted sources from the UMLS Metathesaurus:
 - a) the LICENSEE is required to display in full, prior to providing user access to the Metathesaurus or any of the vocabulary sources within the UMLS, the following wording in order that its users be made aware of these copyright constraints:

"Some material in the UMLS Metathesaurus is from copyrighted sources of the respective copyright holders. Users of the UMLS Metathesaurus are solely responsible for compliance with any copyright, patent or trademark restrictions and are referred to the copyright, patent or trademark notices appearing in the original sources, all of which are hereby incorporated by reference.";

to display a list of all of the vocabularies contained within the UMLS Metathesaurus that are used in the LICENSEE's application; and to indicate for each vocabulary any appropriate copyright notice and whether the entire contents is present or only a portion of it.

- b) the LICENSEE is prohibited from altering UMLS and other vocabulary source content contained within the UMLS Metathesaurus, but may include content from other sources in applications that also contain content from the UMLS Metathesaurus. The LICENSEE may not imply in any way that data from other sources is part of the UMLS Metathesaurus or of any of its vocabulary sources.
- c) the LICENSEE is required to include in its applications identifiers from the UMLS Metathesaurus such that the original source vocabularies for any data obtained from the UMLS Metathesaurus can be determined by reference to a complete version of the UMLS Metathesaurus.

 Netathesaurus

 **Neta
- 12. For material in the UMLS Metathesaurus obtained from some sources additional restrictions on LICENSEE's use may apply. The categories of additional restrictions are described below. The list of UMLS Metathesaurus Vocabulary Sources, which is part of this Agreement and is updated when each version of the Metathesaurus is released, indicates the category of additional restrictions, if any, that apply to each vocabulary source.

 additional licenses may be necessary

LICENSEE should contact the copyright holder directly to discuss uses of a source vocabulary beyond those allowed under this license agreement. If LICENSEE or LICENSEE's end user has a separate agreement with the copyright holder for use of a UMLS Metathesaurus source vocabulary, LICENSEE or LICENSEE's end user may use vocabulary source content obtained from the UMLS Metathesaurus in accordance with the terms of the separate agreement.

12. 1. Category 1:

LICENSEE is prohibited from translating the vocabulary source into another language or from producing other derivative works based on this single vocabulary source.

License Restriction Levels 0-4

Level 0

(28.2%)

unrestricted

Level 1

(1.6%)

negotiate to translate

Level 2 (0.4%)

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

67%

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



Access



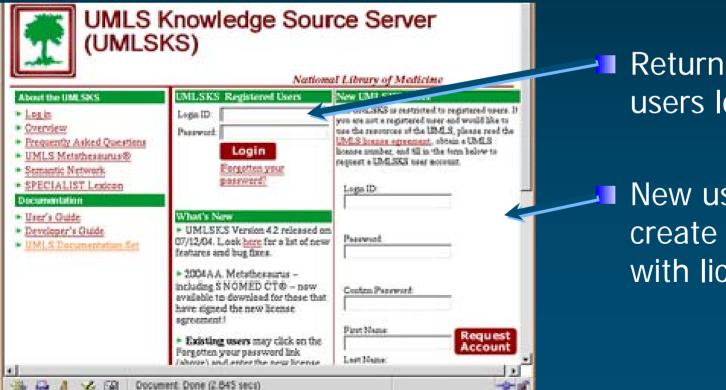
Accessing the Metathesaurus

- Remote access
 - UMLS Knowledge Source Server (UMLSKS)
 - download files
 - search and browse
- Local access
 - MetamorphoSys: install files locally, create customized Metathesaurus subsets
 - RRF Browser: search, browse, view customized subsets



Remote Access UMLSKS

- UMLS Knowledge Source Server (UMLSKS): http://umlsks.nlm.nih.gov
- Web search interface



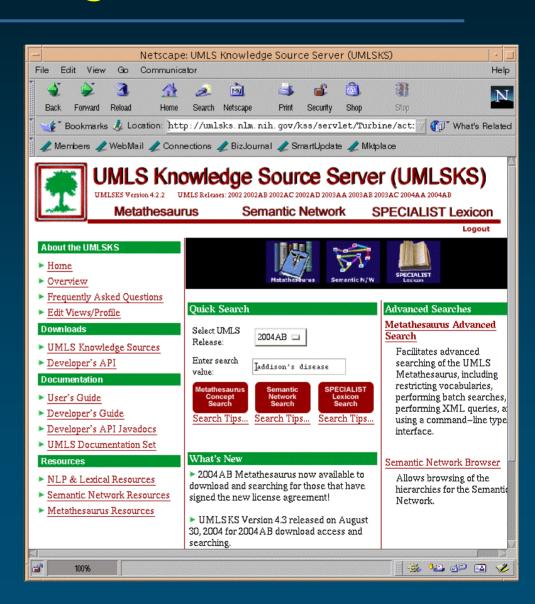
- Returning users log in
- New users create account with license #

UMLSKS Home Page

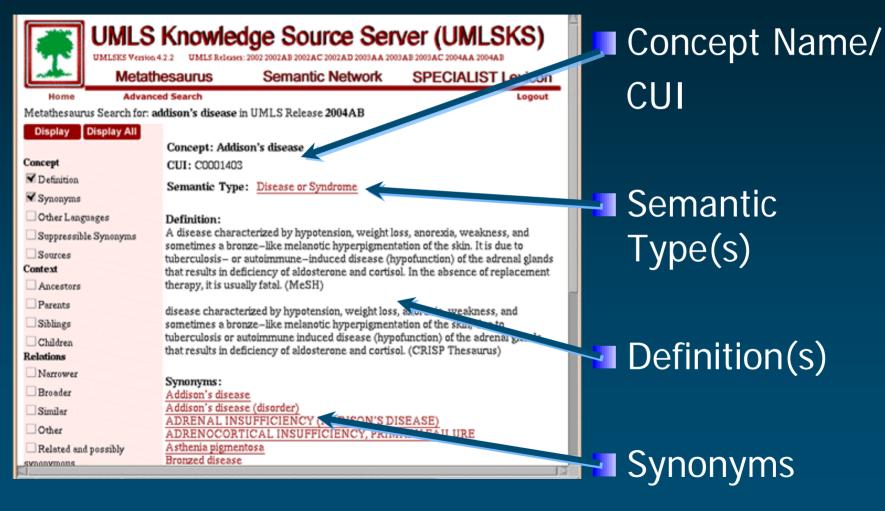
Tabs across top access basic searching of 3 Knowledge Sources

Advanced searching options on right-hand side





UMLSKS Basic Concept Report

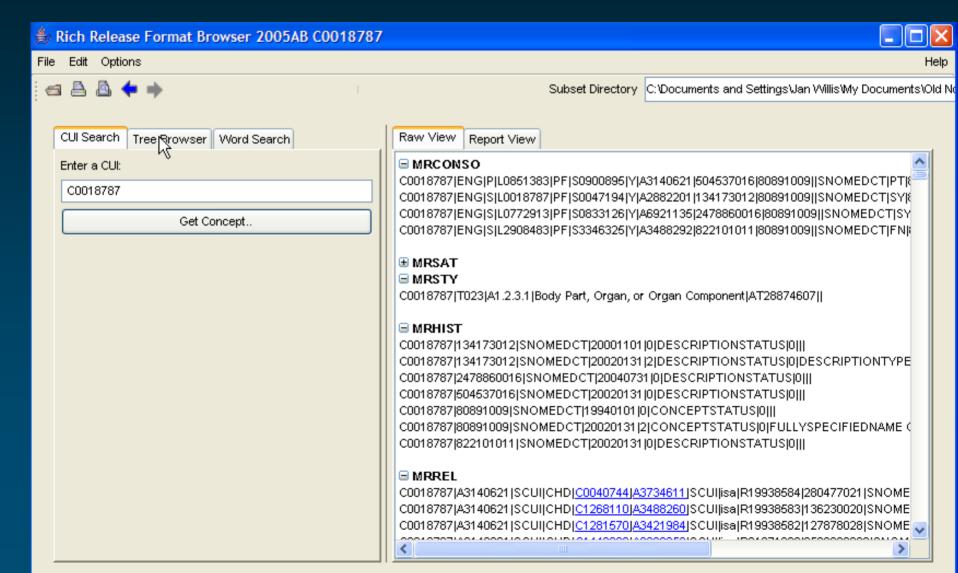


MetamorphoSys

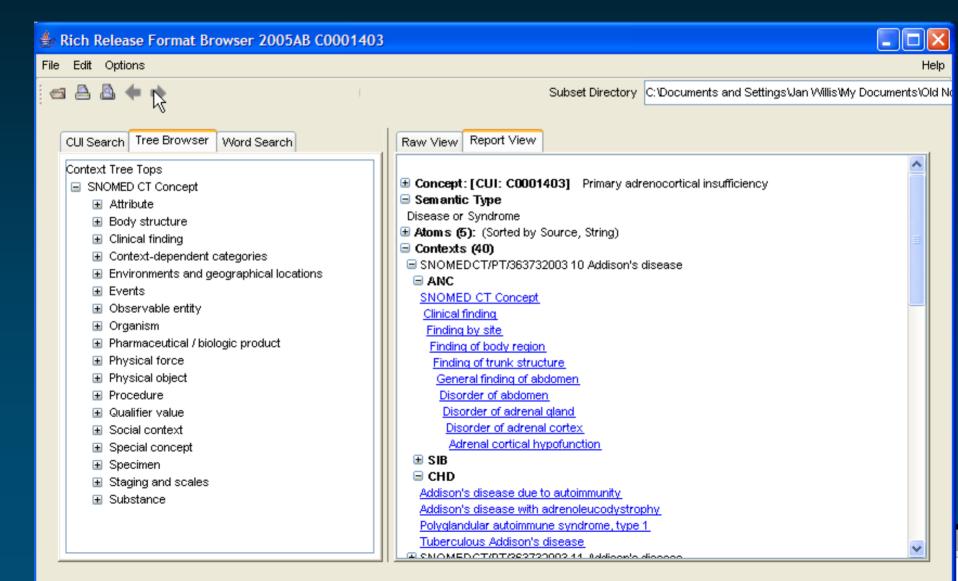
- Multi-platform Java software with each UMLS release
- Installs Knowledge Sources to local storage
- Customizes a local Metathesaurus
- Customization is critical and requires:
 - understanding of source vocabularies
 - license arrangements
 - functional requirements, purpose and perspective
 - technical expertise



RRF Browser search and view subset



RRF Browser tree browser



UMLS 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, and
 - the relationships among them.



Metathesaurus Sources

- Varying purposes, structures, properties
- "Sets of valid values" for data elements:
 - Thesauri, e.g., MeSH
 - Statistical Classifications, e.g., ICD
 - Billing Codes, e.g., CPT
 - Clinical coding systems, e.g., SNOMED
 - Lists of controlled terms
- Full list in Appendix to License Agreement



The UMLS serves as a vehicle for regulatory standards (HIPAA, CHI, PHIN)

Metathesaurus highlights

- Represents the meaning in each source
- Tags source information
- Adds context-free unique identifiers
 - "the name that never changes"
- Adheres to principle of "source transparency"
- Includes normalized word and string indexes produced using UMLS lexical tools



Metathesaurus organization

- Concepts (MRCONSO)
 - Synonymous terms are clustered into a concept
 - Additional information
 - From source (definitions, attributes)
 - Supplied by NLM (CUI, Semantic Type)
- Relationships (MRREL)
 - Concepts are related to other concepts
 - Relationship attributes (type, source of relationship)
- Attributes (MRSAT)



Metathesaurus

- Concept (1.2 M) CUI
 - Set of synonymous concept names
- Term (> 4.1 M) LUI
 - Set of normalized names
- String (> 4.7 M) SUI
 - Distinct concept name
- Atom (> 5.4 M) AUI
 - Concept name in a given source

A1412439 headaches (BI) S1459113 headaches

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

L0018681 headache

A1641293 Cranial Pain (MeSH) S1680378 Cranial Pain

A0418053
HEAD PAIN CEPHALGIA (DxP)
S0375902
HEAD PAIN CEPHALGIA

L0290366 cephalgia head pain

C0018681 Headache



Relationships types

- Hierarchical
 - Parent / Child
 - Broader / Narrower than
- Associative: other

Source asserted or editor-assigned



Relationships Examples: Hierarchical

SNOMED Clinical Terms

Clinical finding

Clinical history and observation findings

General finding of observation of patient

Symptom

Pain

Pain finding at anatomical site Pain of head and neck region

Headache



Relationships Examples: Children

Ocular headache
Aural headache
Nasal headache
Posttraumatic headache
Analgesic overuse headache
Benign exertional headache
Cough headache syndrome
Generalized headache
Headache due to cold exposure



Relationships: Associative

Other source relationships to Headache:

Facial Pain (SNOMED Clinical Terms)

Seizures (Clinical Problem Statements)

Sinusitis (Clinical Problem Statements)

Sinusitis (MedlinePlus)



Semantic Network categorizes

135 high-level categories (semantic types)

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 Network relations

- Relations define useful relations between types
- Types + Relations = Network

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



SNOMED CT in the Metathesaurus 1

- Core content U.S. English, Spanish
 - Concepts, Descriptions, Relationships tables
 - Generic drugs
 - History table
 - ICD-9-CM mapping
- Not included:
 - US, UK SNOMED branded drug extensions
 - non-English, non-Spanish content



SNOMED CT in the Metathesaurus 2

- Different senses of concepts, synonyms, hierarchies
- UMLS concepts vs. SNOMED CT concepts
 - Useful level of distinction in biomedical practice
- UMLS hierarchies vs. SNOMED CT hierarchies
 - Single vs multiple hierarchies
- Occasional disagreement in synonymy and concepts

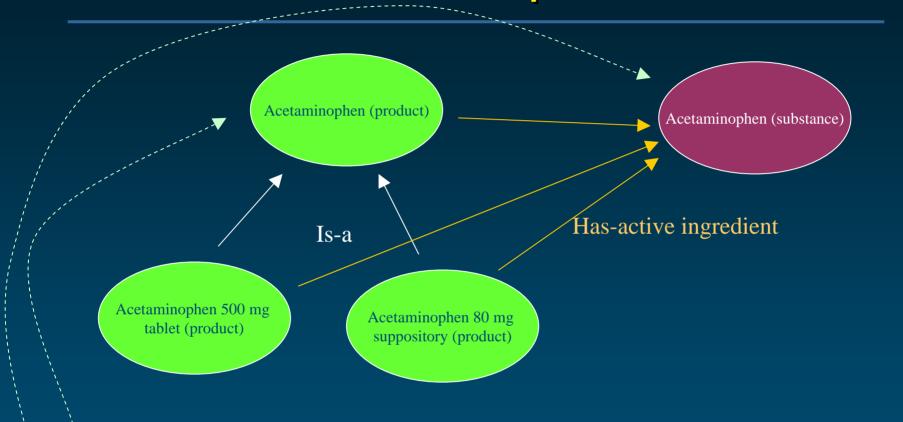


Different Concept Views 1

- Jan. 2005 data
- SNOMED CT differs from UMLS view (14%)
 - Most differences are intentional
 - Some undiscovered synonymy fixed via collegial working relationship
- SNOMED CT 2 concepts / UMLS 1 concept
- For example: substance / product
 - Antacid (substance)
 - Antacid (product)



SNOMED CT Concept View



'This patient is on acetaminophen.'

'Tylenol contains acetaminophen.'



Representing and Resolving Differences

Merging

- Two or more SNOMED CT concepts merged into one UMLS concept
- 38,485 SNOMED CT concepts involved (12.9%)

Splitting

- One or more SNOMED CT atoms split into different UMLS concepts
- 3037 SNOMED CT synonyms in 1,970 concepts (0.7%)



SNOMED CT in the Metathesaurus

- SNOMED CT January 2004 (English)
 - Concepts: 298,090
 - Descriptions: 736,946
 - Relationships: 1,010,873
- Represented in 2004AA UMLS Release as:
 - Concepts: 37,089
 - Descriptions: 350,464 (mostly fully-specified names e.g. Pneumonia (disorder))



SNOMED CT Concepts in the UMLS

SNOMED CT Concepts Table:

ConceptId|ConceptStatus|FullySpecifiedName|CTV3ID|SNOMEDID |IsPrimitive|

271737000 O Anemia (disorder) XM05A DC-10009 1

MRCONSO.RRF:

```
CUI|...|AUI|...|SCUI|...|SAB|TTY|...|STR|...|
C0002871|A3597593|271737000|SNOMEDCT|FN|Anemia
  (disorder)
C0002871|A2878480|271737000|SNOMEDCT|PT|Anemia|
```

C0002871 | A2952250 | 27173 7000 | SNOMEDCT | SY | Absolute anemia |

C0002871|A3095181|271737000|SNOMEDCT|PTGB|Anaemia|

C0002871|A3089808|271737000|SNOMEDCT|SYGB|Absolute anaemia



Additional information on NLM Web site

- SNOMED CT® in the UMLS® Metathesaurus®: Inversion Source Transparency Achieved Jan. 2005
- SNOMED CT in the 2005AB Release
- <u>SNOMED CT® in the UMLS®</u> <u>Metathesaurus®:Release Source Transparency</u> <u>Achieved Jan. 2005</u>
- SNOMED CT in the 2005AB Release
- SNOMED CT Information from the College of American Pathologists



NLM and Health Data Standards

- Health Data Standards
 - Administrative (HIPAA)
 - Clinical (CHI)
 - Public Health (PHIN)
- NLM's current contributions
 - Maintains UMLS as distribution mechanism for standards
 - Negotiated and maintains US license for SNOMED CT
- NLM's evolving role
 - Reduce overlap
 - Establish and support mappings between standard clinical vocabularies (SNOMED, LOINC, RxNorm), and between other administrative and clinical standards



Mappings



Mappings

Link concepts between two or more specific vocabularies

Inter-terminology mappings facilitate reuse of information

- Source-asserted
 - SNOMED CT-NIC/NOC/NANDA (coming)
 - ICD-10-ICPC



Code to Code Mappings

- NLM mapping projects underway:
 - SNOMED CT to CPT/HCPCS
 - SNOMED CT to ICD-9-CM (specifically to support reimbursement)
 - LOINC to CPT
- Mapping projects being discussed:
 - SNOMED CT to MedDRA
 - SNOMED CT to MeSH



Mappings

examples

- LOINC CPT
 - LOINC
 - 1795-4 AMYLASE:CCNC:PT:FLU:QN
 - 1798-8 AMYLASE:CCNC:PT:SER:QN
 - 1799-6 AMYLASE:CCNC:PT:UR:QN
 - All map to CPT
 - 82150 Amylase (blood or urine)
- SNOMED CT CPT
 - SNOMED CT
 - 313500004 Urine protein/creatinine ratio measurement
 - Maps to two CPT codes
 - 82570 Creatinine; other source
 - 84155 Protein, total, except by refractometry; serum



Key NLM Assumptions about Mapping

- Participants must include producers on both ends, and users of output
- Mappings must keep current at both ends
- Mappings will be distributed in the UMLS
- Mapping is an R & D problem; iteration is required to build highly functional maps



RxNorm



NLM's RxNorm Project

- Addresses problems:
 - missed synonymy in UMLS clinical drugs
 - patient safety
- CHI recommended standard Designated standard for use in U.S. Federal Government systems for the electronic exchange of clinical health information
- Supports e-prescribing



RxNorm Source and Subset

- RxNorm Source
 - Normalized forms of clinical drugs
 - Created by NLM
 - Forms only from existing commercial and government clinical drug vocabularies
- RxNorm Subset
 - RxNorm forms from NLM
 - Clinical drug content including SNOMED CT, First DataBank NDDF, FDA NDC
- Users must respect license restrictions



RxNorm Normalized Forms

- Active standardized (generic) ingredients
- Strengths
- Units of measure
- Dosage form

As administered to a patient



Creating RxNorm Concepts

- PALIFERMIN 6.25 MG INTRAVENOUS INJECTION, POWDER, LYOPHILIZED, FOR SOLUTION [KEPIVANCE]
- Standardize measurement and strength
 - Convert 6.25 MG to mg/ml
 - 5 mg/ml solution (per prescribing information)
- Identify brand name, base ingredient, dose form
 - Brand name (BN) = Kepivance
 - Base Ingredient (IN) = Palifermin
 - Dose Form = Injectable Solution
- SBD = Palifermin 5 mg/ml Injectable Solution [Kepivance]

RxNorm Term Types

SBD = Palifermin 5 mg/ml Injectable Solution [Kepivance] → 7 additional concepts

■ IN= Palifermin

■ BN= Kepivance

SBDC = Palifermin 5 mg/ml [Kepivance]

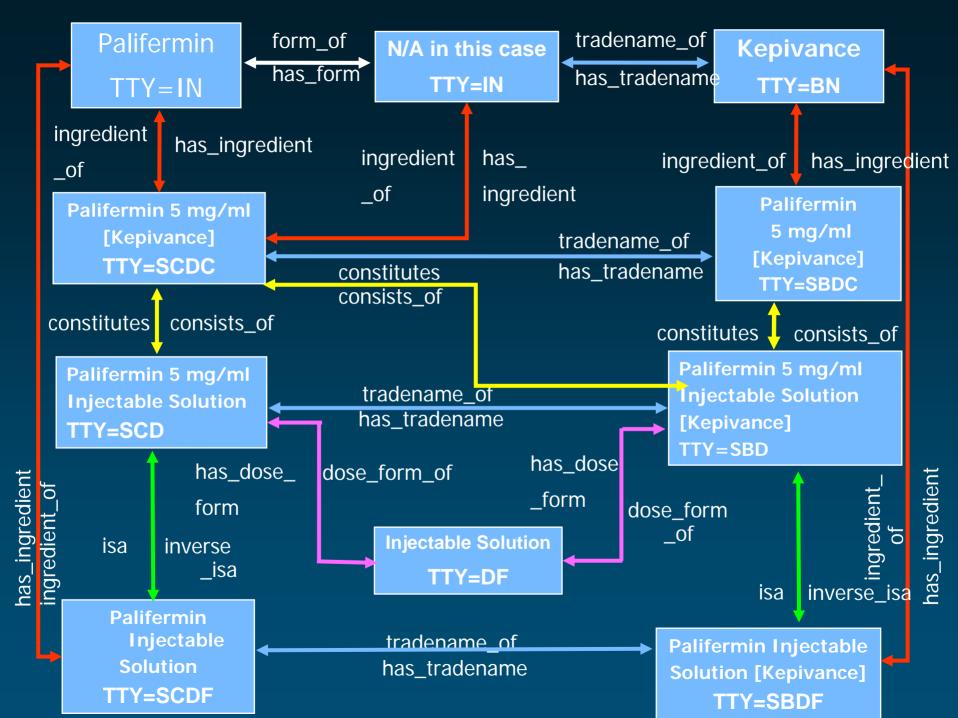
SBDF = Palifermin Injectable Solution [Kepivance]

SCD= Palifermin 5 mg/ml Injectable Solution

SCDC = Palifermin 5 mg/ml

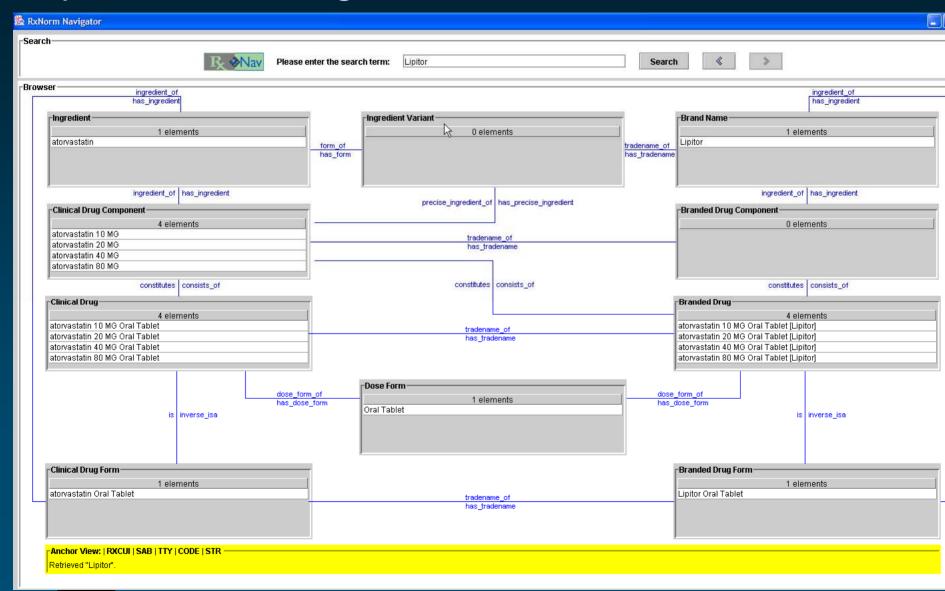
SCDF= Palifermin Injectable Solution





RxNorm Nav

http://mor.nlm.nih.gov/download/rxnav/



RxNorm Subset

- More than 8 full and partial sources
- Full sources
 - FDA, Medispan, Micromedex, NDDF, MDDB
 - More to come
- Partial sources
 - SNOMED CT, MeSH
- NOT in RxNorm
 - USP Medicare Model Guidelines

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



RxNorm Subset 2 ways

- Download RxNorm Source zip files
 - More frequent updates than UMLS
 - Monthly schedule to begin soon
 - Regular updates of individual sources coming
- Extract RxNorm Subset from UMLS
 - 3-4 releases per year
 - MetamorphoSys extracts preconfigured subset
 - Default setting includes Level 0 sources and SNOMED CT
 - Option to select specific sources



What's Not in RxNorm

- Inactive ingredients (e.g., lactose, preservatives, alcohol)
- Pack size
- Flavors



RxNorm Development

- Clinical Drugs from sources are released but RxNorm forms are not created:
 - Radiopharmaceuticals
 - Multi-vitamins
 - More than 3 ingredients
 - Drug Delivery Devices



Conclusion



Current UMLS Release

- UMLS 2005AB Release (June)
 - Metathesaurus
 - 133 sources including SNOMED CT, RxNorm
 - 1,196,265 concepts; 4,752,383 names
 - Semantic Network
 - SPECIALIST Lexicon
- Next release: 2005AC (November)
 - Updated sources include July 2005 SNOMED CT, MeSH 2006, RxNorm, ICD-9-CM 2006
 - New sources include HL7, PNDS
 - MetamophoSys, Browser enhancements



Current RxNorm Release

- August 2005 (7/29/05)
 - Consistent with 2005AB UMLS Metathesaurus
 - Updates: SNOMED CT (Jan. 2005), NDDF, VA NDF, NDF-RT
- September 2005 (on or about 9/12)
 - FDA NDC



Documentation and Support

- UMLS homepage:
 - http://umlsinfo.nlm.nih.gov/
 - links to all other UMLS information
- UMLSKS homepage:
 - http://umlsks.nlm.nih.gov
 - links to the User's and Developer's guides
- UMLSUSERS-L
 - subscribe to discussion list
- NLM Customer Service:
 - custserv@nlm.nih.gov

Additional training

- UMLS Basics
 - NLM
 - American Medical Informatics Association (AMIA) tutorials
 - Medical Library Association (MLA) CE class
 - American Health Information Management Association (AHIMA)
- More SNOMED CT content coming





Center Drive

Library of Medicine

Thank you!

