AMIA 2006 Tutorial T12
Unified Medical Language System®
UMLS® Overview

Rachel Kleinsorge
Jan Willis
Allen Browne
Alan Aronson
National Library of Medicine

National Institutes of Health
U.S. Dept. of Health & Human Services
Schedule

1:00 - 1:30  UMLS Overview
1:30 - 2:00  Metathesaurus
2:00 - 2:15  Electronic Health Data Standards
2:15 - 2:30  UMLS Knowledge Source Server
2:30 - 3:00  MetamorphoSys:
  Customizing the Metathesaurus

3:00 - 3:30  - break -

3:30 - 4:00  SPECIALIST Lexicon and Lexical Tools
4:00 - 4:30  MetaMap Technology Transfer (MMTx)
The UMLS consists of

3 Knowledge Sources + associated tools

- 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 processing language
UMLS Objectives

- Intellectual middleware

- A set of multi-purpose tools for system developers

- Knowledge Sources used to overcome:
  - disparities in language and 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)
The UMLS is not an end-user application
UMLS Access

- **Remote access**
  - UMLS Knowledge Server (UMLSKS)
      - Browsers, Navigators
      - APIs
      - Download files and programs
      - Documentation

- **Local access**
  - MetamorphoSys: install files locally, create customized Metathesaurus subsets
  - Subset Browser: search, browse, view customized subsets
Metathesaurus License Agreement

Online Web-based license:

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

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
License Agreement Restrictions

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.

See list of current sources in Appendix A.1.
The Metathesaurus is

- very large
- multi-purpose
- multi-lingual

It contains information about

- biomedical and health related concepts
- their various names and associated codes
- the relationships among them
Metathesaurus: clusters terms by meaning

- Concepts contain synonymous terms
- Preferred term is chosen (default can be changed)
- Unique identifier (CUI) is assigned

<table>
<thead>
<tr>
<th>term</th>
<th>source</th>
<th>term type</th>
<th>source ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison’s disease</td>
<td>Metathesaurus</td>
<td>PN</td>
<td></td>
</tr>
<tr>
<td><strong>Addison’s disease</strong></td>
<td>SNOMED CT</td>
<td>PT</td>
<td>363732003</td>
</tr>
<tr>
<td>Addison’s Disease</td>
<td>MedlinePlus</td>
<td>PT</td>
<td>T1233</td>
</tr>
<tr>
<td><strong>Addison Disease</strong></td>
<td>MeSH</td>
<td>PT</td>
<td>D000224</td>
</tr>
<tr>
<td>Bronzed disease</td>
<td>SNOMED Intl</td>
<td>SY</td>
<td>DB-70620</td>
</tr>
<tr>
<td><strong>Primary Adrenal Insufficiency</strong></td>
<td>MeSH</td>
<td>EN</td>
<td>D000224</td>
</tr>
<tr>
<td>Primary hypoadrenalism syndrome, Addison</td>
<td>MedDRA</td>
<td>LT</td>
<td>10036696</td>
</tr>
</tbody>
</table>

C0001403 Addison’s disease
Organization of Concepts

**Concept**
- C0001621

**Term**
- L0001621
  - S0011231 Adrenal Gland Disease
    - A0020266 MeSH
    - A7568579 NCI Thesaurus
  - S0000441 Disease of adrenal gland
    - A0001264 SNOMED 1982
    - A6917004 SNOMED Clinical Terms
  - S0481705 Diseases of Adrenal Gland
    - A0014499 SNOMED 1982
  - S0220090 Diseases, adrenal gland
    - A0049924 MeSH

**Term**
- L0181041
  - S0632950 Disorder of adrenal gland
    - A0688820 Read Codes
    - A4778687 SNOMED Clinical Terms
  - S0354509 Adrenal Gland Disorders
    - A6996540 MedlinePlus
    - A7576253 NCI Thesaurus
    - A7561794 Psychological Index Terms

**Term**
- L1279026
  - S1520972 Nebennierenkrankheiten
    - A7500884
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
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
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
SPECIALIST Lexicon

- Syntactic English lexicon of common words, biomedical terms (250K+ words, 400K+ 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

Adrenal gland diseases
Diseases of the adrenal glands
Adrenal disorder
Disorder of adrenal gland
C0001621
SPECIALIST Lexicon | lexical records

{base=Kaposi's sarcoma
 spelling_variant=Kaposi
sarcoma
 entry=E0003576
   cat=noun
   variants=uncount
   variants=reg
   variants=glreg
 }

{base=chronic
 entry=E0016869
   cat=adj
   variants=inv
   position=attrib(1)
   position=pred
   stative
 }

{base=aspirate
 entry=E0010803
   cat=verb
   variants=reg
   tran=np
   nominalization=aspiration|noun|E0010804
 }

{base=in
 entry=E0033870
   cat=prep
 }

18
Lexical Tools

- **JAVA programs to manage lexical variation, indexing, normalization in biomedical text**

- **Wordind**
  - breaks strings into words
  - produces the Metathesaurus word indexes (MRXW)

- **lvg**
  - performs various lexical transformations
  - 58 flow components and 38 options in 2006

- **NORM**
  - a selection of LVG transformations
  - produces Metathesaurus normalized word and string indexes (MRXNW & MRXNS)
  - used to access those indexes
Normalization

- Remove genitive
- Remove stop words
- Lowercase
- Strip punctuation
- Uninflect
- Sort words

```
Hodgkin’s diseases, NOS
Hodgkin diseases, NOS
Hodgkin diseases,
hodgkin diseases,
hodgkin diseases
hodgkin disease
disease hodgkin
```
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
2006AD UMLS

Metathesaurus:
- 120 sources
- 1,352,403 concepts
- 17 languages

Semantic Network:
- 135 Semantic Types
- 54 Semantic Relationships

SPECI ALIST Lexicon:
- Over 297K records (over 482K inflectional forms)
Metathesaurus
Metathesaurus Sources

- Wide range of general and specialized biomedical terminologies
- Used in variety of settings and purposes:
  - Clinical
  - Administrative
  - Public Health Reporting
  - Research
Metathesaurus Sources

- 120 vocabularies in 17 languages
- Sets of valid values
  - Thesauri, e.g., MeSH, CRISP, NCI
  - Statistical classifications, e.g., ICD-9-CM
  - Billing codes, e.g., CPT
  - Clinical coding systems, e.g., SNOMED CT

- See License Appendix, documentation
Metathesaurus: not a single vocabulary

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."

Metathesaurus highlights

- Concept based
- Represents the meaning in each source
- Represents and delivers data in common format
- Adheres to principle of “source transparency”

- Source information tagged
- Context-free unique identifiers added

- Normalized word and string indexes included
Source Data to Metathesaurus Files

- Names, Synonyms
- Terms, Codes → MRCONSO
- Relationships → MRREL
- Hierarchies → MRHIER
- Attributes → MRSAT
- Definitions → MRDEF
- Mappings → MRMAP, MRSMAP, (MRREL, MRCONSO)

CUIs links concept data across files
## MRCONSO (sample rows 1..5)

<table>
<thead>
<tr>
<th></th>
<th>CUI</th>
<th>LAT</th>
<th>TS</th>
<th>LUI</th>
<th>SUI</th>
<th>ISPREF</th>
<th>AUI</th>
<th>SAUI</th>
<th>SCUI</th>
<th>SDUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C0001403</td>
<td>ENG</td>
<td>P</td>
<td>L0001403</td>
<td>PF S0354372</td>
<td>Y</td>
<td>A4367951</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C0001403</td>
<td>ENG</td>
<td>P</td>
<td>L0001403</td>
<td>PF S0354372</td>
<td>N</td>
<td>A2922421</td>
<td>485624014363732003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C0001403</td>
<td>ENG</td>
<td>P</td>
<td>L0001403</td>
<td>VC S0010794</td>
<td>Y</td>
<td>A0019740</td>
<td></td>
<td>M0000346</td>
<td>D000224</td>
</tr>
<tr>
<td>4</td>
<td>C0001403</td>
<td>ENG</td>
<td>S</td>
<td>L0494851</td>
<td>PF S2164152</td>
<td>N</td>
<td>A2018589</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C0001403</td>
<td>FRE</td>
<td>P</td>
<td>L3246333</td>
<td>PF S3773545</td>
<td>Y</td>
<td>A3996251</td>
<td></td>
<td>D000224</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SAB</th>
<th>TTY</th>
<th>CODE</th>
<th>STR</th>
<th>SRL</th>
<th>SUPPRESS</th>
<th>CVF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MTH</td>
<td>PN</td>
<td>NOCODE</td>
<td>Addison's disease</td>
<td>0</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SNOMEDCT</td>
<td>PT</td>
<td>363732003</td>
<td>Addison's disease</td>
<td>4</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MSH</td>
<td>MH</td>
<td>D000224</td>
<td>Addison’s Disease</td>
<td>0</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MDR</td>
<td>LT</td>
<td>10052381</td>
<td>Primary adrenal insufficiency</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>MSHFRE</td>
<td>MH</td>
<td>D000224</td>
<td>Addison, maladie</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

**Appendix - Metathesaurus relational files (RRF)**
## MRCONSO (sample rows 6..10)

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>TS</th>
<th>LUI</th>
<th>LTS</th>
<th>SUI</th>
<th>ISPREF</th>
<th>AUI</th>
<th>SAUI</th>
<th>SCUI</th>
<th>SDUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0001403</td>
<td>FRE</td>
<td>S</td>
<td>L1272481</td>
<td>PF</td>
<td>S1514427</td>
<td>Y</td>
<td>A1464383</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td>GER</td>
<td>P</td>
<td>L1229627</td>
<td>PF</td>
<td>S1471573</td>
<td>Y</td>
<td>A4030156</td>
<td>D000224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td>GER</td>
<td>S</td>
<td>L1239271</td>
<td>PF</td>
<td>S1481217</td>
<td>Y</td>
<td>A4034094</td>
<td>D000224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td>JPN</td>
<td>P</td>
<td>L3437833</td>
<td>PF</td>
<td>S3965327</td>
<td>Y</td>
<td>A4264008</td>
<td>D000224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td>JPN</td>
<td>S</td>
<td>L3465347</td>
<td>PF</td>
<td>S3992841</td>
<td>Y</td>
<td>A4291522</td>
<td>D000224</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Appendix - Metathesaurus relational files (RRF)

<table>
<thead>
<tr>
<th>SAB</th>
<th>TTY</th>
<th>CODE</th>
<th>STR</th>
<th>SRL</th>
<th>SUPPRESS</th>
<th>CVF</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHOFRE</td>
<td>IT</td>
<td>0410</td>
<td>MALADIE D'ADDISON</td>
<td>2</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>MSHGER</td>
<td>MH</td>
<td>D000224</td>
<td>Addison-Krankheit</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>MSHGER</td>
<td>SY</td>
<td>D000224</td>
<td>Bronzehautkrankheit</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>MSHJ PN</td>
<td>MH</td>
<td>D000224</td>
<td>Addison病</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>MSHPJ N</td>
<td>SY</td>
<td>D000224</td>
<td>副腎性黒皮症</td>
<td>3</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
### MRREL (sample rows)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>REL</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C0001403</td>
<td>CUI</td>
<td>RB</td>
<td>C0001621</td>
<td>CUI</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C0001403</td>
<td>A0019738</td>
<td>SY</td>
<td>C0001403</td>
<td>A0049628</td>
<td>AUI</td>
</tr>
<tr>
<td>3</td>
<td>C0001403</td>
<td>A2922421</td>
<td>SCUI</td>
<td>CHD</td>
<td>C0085859</td>
<td>A2977940</td>
</tr>
<tr>
<td>4</td>
<td>C0001403</td>
<td>A6326321</td>
<td>SCUI</td>
<td>RO</td>
<td>C0688490</td>
<td>A6339383</td>
</tr>
<tr>
<td>5</td>
<td>C0001403</td>
<td>A0019743</td>
<td>AUI</td>
<td>PAR</td>
<td>C0935495</td>
<td>A1988358</td>
</tr>
</tbody>
</table>

### Appendix - Metathesaurus relational files (RRF)

<table>
<thead>
<tr>
<th>RELA</th>
<th>RUI</th>
<th>SRUI</th>
<th>SAB</th>
<th>SL</th>
<th>RG</th>
<th>DR</th>
<th>SUPPRESS</th>
<th>CVF</th>
</tr>
</thead>
<tbody>
<tr>
<td>isa</td>
<td>R02837989</td>
<td>MTH</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>may_treat</td>
<td>R18849683</td>
<td>MSH</td>
<td>MSH</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>has_member</td>
<td>R19859511</td>
<td>1658795027</td>
<td>SNOMEDCT</td>
<td>SNOMEDCT</td>
<td>0</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>may_treat</td>
<td>R27600039</td>
<td>NDFRT</td>
<td>NDFRT</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>has_member</td>
<td>R08110401</td>
<td>PSY</td>
<td>PSY</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### MRHIER (sample rows)

<table>
<thead>
<tr>
<th></th>
<th>CUI</th>
<th>AUI</th>
<th>CXN</th>
<th>PAUI</th>
<th>SAB</th>
<th>RELA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C0001403</td>
<td>A0019740</td>
<td>1</td>
<td>A0020270</td>
<td>MSH</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C0001403</td>
<td>A0019740</td>
<td>2</td>
<td>A0028022</td>
<td>MSH</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C0001403</td>
<td>A0019743</td>
<td>3</td>
<td>A1988358</td>
<td>PSY</td>
<td>member_of_cluster</td>
</tr>
<tr>
<td>4</td>
<td>C0001403</td>
<td>A2922421</td>
<td>1</td>
<td>A3307650</td>
<td>SNOMEDCT</td>
<td>isa</td>
</tr>
<tr>
<td>5</td>
<td>C0001403</td>
<td>A2922421</td>
<td>2</td>
<td>A3307650</td>
<td>SNOMEDCT</td>
<td>isa</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>A0449751.A1988279.A1988358</td>
</tr>
<tr>
<td></td>
<td>A2872360.A3307650</td>
</tr>
<tr>
<td></td>
<td>A2933400.A2989549.A3307650</td>
</tr>
</tbody>
</table>

**Appendix - Metathesaurus relational files (RRF)**
### MRSAT (sample rows)

<table>
<thead>
<tr>
<th>CUI</th>
<th>LUI</th>
<th>SUI</th>
<th>METAUI</th>
<th>STYPE</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0001403</td>
<td>L0001403</td>
<td>S0010792</td>
<td>A0019738</td>
<td>AUI</td>
<td>D000224</td>
</tr>
<tr>
<td>C0001403</td>
<td>L0001403</td>
<td>S0010794</td>
<td>A6326321</td>
<td>SCUI</td>
<td>C712</td>
</tr>
<tr>
<td>C0001403</td>
<td>L0001403</td>
<td>S0354372</td>
<td>A2922421</td>
<td>SAUI</td>
<td>363732003</td>
</tr>
<tr>
<td>C0001403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0001403</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Appendix - Metathesaurus relational files (RRF)

<table>
<thead>
<tr>
<th>ATUI</th>
<th>SATUI</th>
<th>ATN</th>
<th>SAB</th>
<th>ATV</th>
<th>SUPPRESS</th>
<th>CVF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT15321482</td>
<td></td>
<td>DI</td>
<td>MSH</td>
<td>D000224</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>AT33411754</td>
<td>MESH_UI</td>
<td></td>
<td>NDFRT</td>
<td>D000224</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>AT24166602</td>
<td>DESCRIPTIONS</td>
<td></td>
<td>SNOMEDCT</td>
<td>0</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>AT27438950</td>
<td>REFINABILITY</td>
<td></td>
<td>SNOMEDCT</td>
<td>0</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>AT02925340</td>
<td>ST</td>
<td></td>
<td>MTH</td>
<td>R</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.
Metathesaurus Metadata Files

- **MRFILES**: files authority
- **MRCOLS**: data elements
- **MRDOC**: data element valid values
- **MRSAB**: source vocabularies
- **MRRANK**: source/term types ranks

- Generate UMLS release documentation
- Full set of metadata files in subset
Metathesaurus Indexes

- **Word**: each word in each unique string in each language
  - **Heart disease** yields
    - ENG| disease |C0024117|L0024117|S0058458|
    - ENG| heart |C0018787|L0018787|S0047194|

- **Normalized word**: each normalized word in each unique English string
  - **Disease, diseases, diseased** yields
    - ENG| disease |C0024117|L0024117|S0058458|

- **Normalized string** **obstructive lung disease**
  - ENG| disease lung obstructive | C0024117| L0024117| ... S0058458|
Different Formats for Different Purposes

- Original Release Format (ORF)
- **Lexical** View
- Natural language processing

- Rich Release Format (RRF)
- Atomic View
- Greater specificity
  - Facilitates maintenance
  - Enables other types of changes and applications

- RRF Browser in MetamorphoSys
How do I use the Metathesaurus?

- Identify useful sources
- Identify useful content from specific sources
- Create customized Metathesaurus using MetamorphoSys

- Use UMLS Release Documentation to understand file content and structure
Metathesaurus Use Cases

• Information retrieval
• Thesaurus construction
• Natural language processing
• Automated indexing
• Electronic patient records

• Distribution of health data standard vocabularies
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
Electronic Health Data Standards
Health IT

Health IT initiatives harness current and emerging information technologies to improve patient safety and convenience while reducing the cost of providing care. Some of the numerous benefits of health IT initiatives will include: a reduction in medical errors, avoidance of costly duplicate testing, and elimination of unnecessary hospitalizations.

The benefits of health IT initiatives range from consumer convenience, as patients will not have to fill out repetitive paper work, to life-saving early detection of an infectious disease outbreak, as anonymous data from emergency rooms is sent to public health systems instantly. These initiatives aid in fulfilling the President’s goal for most Americans to have electronic health records by the year 2014.

American Health Information Community (AHIC) Topics

The American Health Information Community (AHIC) is a federal advisory body, chartered to make recommendations to the Secretary of HHS on how to accelerate the development and adoption of Health IT. The Community will make recommendations to the Secretary of HHS to enable advancement in four areas of focus by the end of 2006.

- Consumer Empowerment - Make available a consumer-directed and secure electronic record of health care registration information and a medication history for patients.
- Chronic Care - Allow the widespread use of secure messaging, as appropriate, as a means of communication between doctors and patients about care delivery.
- Regulational - Enable the transfer of standardized and anonymized health data from the point of health care delivery to authorized public health agencies within 24 hours of its collection.
- Electronic Health Records - Create an electronic health record that includes laboratory results and interpretations, that is standardized, widely available and secure.

Highlighted Federal Efforts

Health & Human Services

Health & Human Services (HHS) - HHS is facilitating the development of standards for Health IT systems that will improve patient care and increase efficiency across the health care system. HHS, through several of its agencies, also provides funding to organizations engaged in building and testing Health IT systems, standards and projects.

Department of Defense

Department of Defense (DoD) - Currently, thousands of military medical providers use the DoD's electronic health record system, AHLTA, and nearly 300,000 outpatient visits are captured digitally every week. DoD's vision is to provide each patient with a continuously updated digital medical record from the point of first contact. The Federal health information initiative is supported by the United States Department of Veterans Affairs (VA).

Veterans Health Administration (VHA) - The VA is a division of the U.S. Department of Veteran's Affairs, and provides care for over five million veterans of the United States Armed Services. The VA's goal is to provide state-of-the-art medical care to those who served our country.

http://www.hhs.gov/healthinformationtechnology/
Health Information Technology at NLM

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 HHS Secretary Mike Leavitt and President Bush’s goal for the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care.

Health Data Standards

- Clinical Vocabularies supported, licensed, or developed by NLM:
  - SNOMED CT | ICD-9-CM | ICD-10 | ICHOM
- Uniform distribution mechanism for HIPAA and clinical vocabulary standards through the UMLS Metathesaurus

NLM coordinates efforts to develop mappings between HIPAA code sets and standard clinical vocabularies.
- SNOMED CT to ICD-9-CM | SNOMED CT to CPT | LOINC to CPT

NLM promotes harmonization between standards.
- Contract with HL7 to align HL7 message standard with CCHT standard vocabularies and create implementation guides for exchange of entire EHRs.

System Development Tools

- UMLS Resources
- Language and Knowledge Processing
- Image Processing

Grants and Funding

Education and Training

- Funded Support for Academic Training in Biomedical Informatics & Bioinformatics - Formal programs and individual fellowships are offered to assist medical informaticians in pursuit of a degree.
- Medical Informatics Training Program - Support for visiting scientists and students to participate in research at the U.S. National Center for Biomedical Communications. Medical Informatics Course at Woods Hole sponsored by NLM and the Woods Hole Marine Biology Laboratory.

Key Reports Supported or Produced by NLM

- Ending the Document Game (2005)
- Networking Health: Prescriptions for the Internet (2000)
- Evaluating the Coverage of Controlled Health Data Terminologies: Report on the Results of the NLM/HCPR Large Scale Vocabulary Test (1997)
- For the Record: Protecting Electronic Health Information (1997)
- Making a Powerful Connection (1995)

Electronic Health Data Standards

- Include standard vocabularies
- Key element of health information technology infrastructure for:
  - Effective decision support
  - Safe, evidence-based, and coordinated health care
  - Cost-effective care
  - Increased/informed choice
  - More efficient clinical, public health, and health services research
  - Timely public health and bioterrorism surveillance
NLM-led Support for Development and Maintenance

- **1999** – LOINC (lab tests/instrument observations) - contract support
- **2002** – RxNorm (clinical drugs) - direct development
- **2003** – SNOMED CT contract & license for U.S-wide use *(as distributed by NLM in UMLS)*
- **2004** – NLM designated central coordinating body for clinical terminology standards within U.S. Dept. of Health and Human Services (HHS)
- **2006** – Draft LOINC to CPT mapping distributed
SNOMED CT

- CHI recommended standard
- Comprehensive clinical vocabulary
- Substantially increased Metathesaurus content
  - Concepts: ~300K (+37K to Metathesaurus)
  - Descriptions: ~737K (+350K to Metathesaurus)
  - Relationships: 1.3M
- Generic drugs, History table

- Updates
  - January SNOMED CT update → Spring UMLS
  - July SNOMED CT update → Fall UMLS
**RxNorm**

- CHI recommended standard
- Developed to address **missed synonymy** in UMLS clinical drugs
- Provides standardized (normalized) forms of U.S. clinical drugs
- Supports
  - **effective sharing of drug data across systems**
  - electronic health record (EHR)
  - computerized physician order entry (CPOE)
Mappings

- Link specific terms or codes between two sources
- NLM given the responsibility for funding, coordinating, and/or performing official mappings between standard clinical terminologies and HIPAA code sets
- Represented in Metathesaurus files MRMAP, MRSMAP
Mapping projects planned/underway

Clinical standards → HI PAA code sets:
- LOINC → CPT (Draft now available for testing!)
- SNOMED CT → ICD-9-CM, ICD-10-CM
- SNOMED CT → CPT

SNOMED CT → “other” vocabularies:
- International Classification of Functioning, Disability and Health (ICF)
- International Classification of Primary Care (ICPC)
- Medcin
- Medical Dictionary for Regulatory Affairs (MedDRA)
- Medical Subject Headings (MeSH®)
- Nursing Vocabularies (NIC, NOC, NANDA)
Key NLM Assumptions about Mapping

- Mapping is an R & D problem
- Iteration required to build highly functional maps
- Testing, validation, and use in real world settings
- Participants include producers of vocabularies on both ends, and prospective users, recipients of maps
- Mapping will prompt changes/corrections to vocabulary content and update schedules
- Mappings must be updated when either end is updated
- Mappings will be distributed in the UMLS (not exclusively)
Draft LNC215 to CPT2005 Mappings

Use case assumes
- LOINC codes used to order lab tests, observations
- CPT codes used for billing (one-way)

2000+ most common mappings contributed

Created by Intermountain Healthcare
Reviewed by Regenstrief Institute and AMA

Future developments to include:
- Radiology tests
- Document names (consults, progress notes)
- More Laboratory LOINC mappings
- Clinical findings (vital signs, height, weight)
## Mapping examples – LOINC → CPT

<table>
<thead>
<tr>
<th>LOINC Code</th>
<th>CPT Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1795-4 AMYLASE:CCNC:PT:FLU:QN</td>
<td>82150 Amylase (blood or urine)</td>
</tr>
<tr>
<td>1798-8 AMYLASE:CCNC:PT:SER:QN</td>
<td></td>
</tr>
<tr>
<td>1799-6 AMYLASE:CCNC:PT:UR:QN</td>
<td></td>
</tr>
<tr>
<td>LNC215_TO_CPT2005</td>
<td>MRSMAP</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>C1704201</strong></td>
<td>LNC</td>
</tr>
<tr>
<td><strong>1795-4</strong></td>
<td>CODE</td>
</tr>
</tbody>
</table>

| C1704201 | LNC | ATX104051418 |
| 1798-8 | CODE | RN | 82150 | CODE |

| C1704201 | LNC | ATX104051419 |
| 1799-6 | CODE | RN | 82150 | CODE |
LNC215_TO_CPT2005  MRCONSO

C1704201|ENG|P|L6107611|PF|S7015223|Y
|A10924448| |LNC| XM| NOCODE| LNC21
5 to CPT2005 Mappings |0|N|
| ATX104051002 | MAPSETNAME | LNC | LNC215 to CPT2005 Mappings | N |
| ATX104051003 | MAPSETTYPE | LNC | LOINC has associated CPT code | N |
| ATX104051007 | FROMVSAB | LNC | LNC215 | N |
| ATX104051008 | TOVSAB | LNC | CPT2005 | N |
| ... |  |
| ATX104051018 | MAPSETVERSION | LNC | Sept 2006 | N |
Unified Medical Language System

Home > Biomedical Research & Informatics > UMLS

9/30/06: Draft LOINC to CPT Mappings now available for download from the UMLSKS. *** New to the UMLS? Register now.

About the UMLS Resources
- Metathesaurus; Semantic Network; SPECIALIST Lexicon and lexical programs; MetamorphoSys

Accessing UMLS Knowledge Sources
- Metathesaurus license; Semantic Network; SPECIALIST Lexicon; DVD

Knowledge Source Server
- Download files; searching; additional tools and resources

Documentation

Help
- Training; contact us; FAQs; listserv

About
- NLM's Unified Medical Language System (UMLS) project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs for system developers. Researchers will find the UMLS products useful in investigating knowledge representation and retrieval questions.

Metathesaurus Source Vocabularies
- SNOMED CT
- LOINC
- RxNorm
- MeSH
- List of Sources
- Source FAQs
- Mappings

More Resources
- Metathesaurus License
- Tools
- Learning Resources
- MetaMap Transfer (MMTx)
- Archives
UMLS Knowledge Source Server (UMLSKS)
UMLS Knowledge Source Server

- Licensed users access online:

- Web search interface
  for the three Knowledge Sources

- Application Programming Interface (API)
Using the UMLSKS

- Create login ID and password using the form on the right.

- Each UMLS license can have multiple login IDs.

- NLM does not maintain a copy of the password or Login ID.

- Passwords may be reset, not login IDs.
UMLS Knowledge Source Server (UMLSKS)

Includes:
- Definitions
- Synonyms
- Other Languages
- Sources
- Context Information
- Relations
- Co-occurring Concepts

Can be used to find:
- Source IDs
- Other Language
- Synonyms
- Term Type Information
- Term Variants
Metathesaurus Advanced Search Options

- **Focused Search**
  
  Search by release, source vocabulary or language
  
  Partial string matching with right or left truncation

- **XML Query**
  
  Send Standard API commands or specialized requests
  
  Run against the UMLSKS, results returned in tab delimited text

- **Request relational records**
  
  View records by CUI for the files MRCONSO, MRDEF, MRSTY, MRCXT, MRREL, MRSAT, & MRCOC as tab delimited text
Semantic Network search

- Enter search term or select from drop down box
- View Type, TUI, and Definition
- Select checkboxes to see Meta Concepts, Relations And Raw Records
SPECIALIST Lexicon Search

- Search by entering the complete term
- View record in relational or unit record format
### Application Programming Interfaces

- Remote server at NLM
- Local application connected through

<table>
<thead>
<tr>
<th><strong>Java RMI</strong></th>
<th><strong>TCP/IP socket</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Java-based applications</td>
<td>XML-based queries</td>
</tr>
<tr>
<td>Developer’s Guide: Chapter 3</td>
<td>Developer’s Guide: Chapter 5</td>
</tr>
<tr>
<td>Set of Java classes (part of the UMLSKS API download)</td>
<td>XML schema</td>
</tr>
<tr>
<td>Detailed <em>Javadoc</em> documentation online and with API download</td>
<td>Socket server</td>
</tr>
</tbody>
</table>

- **Host:** umlsks.nlm.nih.gov
- **Port:** 8042

You MUST register your IP Address!
New UMLSKS expected Spring 2007

- Web services based
- Create your own tabs
- Forms auto-complete, learning as you go
- Looking for beta testers
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
- Include your license number
- Run MetamorphoSys from DVD
Machine Requirements

- A fast CPU – 1 GHz or higher
- 1 GB RAM recommended (512 MB min.)
- 6x (or better) DVD drive
- 13 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
How do I?
Specify sources for a customized subset?

Exclude or Include

Highlighted rows are excluded from the subset.
Extract an RxNorm subset?

- On the File menu select “Enable/Disable Filter”
  Selecting any filter opens a new tab for that filter.
- Make your selections on the tab for that filter.

Other filters allow users to remove terms:
- Attribute Type List
- Relationship Type List
- Semantic Types List
- Source Term Types
Create a custom database load script?

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

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
Remove specific term types from subset?

Select Options → Advanced Suppressibility Options
Select the term types to remove
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
More information and help

- MetamorphoSys Documentation at:

- Readme file on the DVD or downloaded from the UMLSKS

- Help Menu from any page in MetamorphoSys
SPECIALIST Lexicon and Lexical Tools
MetaMap
and
MetaMap Transfer
(MMTx)
Outline

- Purpose of MetaMap/MMTx
- The MetaMap/MMTx Algorithm
- Availability
- Demo
Purpose of MetaMap/MMTx

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

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

Concepts:
- Termination
- Clinical Trials
- Adrenergic beta-Antagonists
- Myocardial Infarction
- Heart attack (Myocardial Infarction)

Clinical Trials
- Trial (Clinical Trials)

Beta-blocker (Adrenergic beta-Antagonists)
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
Technical Details and Availability

http://skr.nlm.nih.gov/
- Click ‘Research Information’ for technical details

http://mmtx.nlm.nih.gov/
- Click ‘Documentation’ and ‘Prerequisites’

Use restrictions
- Sign UMLS license agreement; then access MetaMap and download MMTx using UMLS ID/password
- Respect UMLS constituent vocabulary copyrights
MetaMap Demo

Example: normal processing

Phrase: “lung cancer.”

Meta Candidates (8):
1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]
1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]
861 Cancer (Malignant Neoplasms) [Neoplastic Process]
861 Lung [Body Part, Organ, or Organ Component]
861 Cancer (Cancer Genus) [Invertebrate]
861 Lung (Entire lung) [Body Part, Organ, or Organ Component]
861 Cancer (Specialty Type - cancer) [Biomedical Occupation or Discipline]
768 Pneumonia [Disease or Syndrome]

Meta Mapping (1000):
1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

Meta Mapping (1000):
1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]
Example: Variants (-v)

Phrases: “lung cancer.”

lung cancer [noun] variants (n=1):
lung cancer{[noun], 0=[{}]} 

lung [noun] variants (n=9):
lung{[noun], 0=[{}]} lungs{[noun], 1="i"} pneumonia{[noun], 5="ds"} 
Pneumoniae{[noun], 5="ds"} pneumonias{[noun], 5="ds"} 
Pneumonic{[adj], 2="s"} Pulmonal{[adj], 4="ss"} Pulmonary{[adj], 2="s"} Pulmonic{[adj], 2="s"} 

Cancer [noun] variants (n=4):
cancer{[noun], 0=[{}]} cancerous{[adj], 3="d"} cancers{[noun], 1="i"} 
Carcinomatous{[adj], 2="s"} 
...

Example: Compound mappings

Phrase: “obstructive sleep apnea.”
Meta Candidates (8):
...
Meta Mapping (1000):
1000 Obstructive sleep apnoea (Sleep Apnea, Obstructive) [Disease or Syndrome]
Meta Mapping (901):
827 Obstructive (Obstructed) [Functional Concept]
901 Apnea, Sleep (Sleep Apnea Syndromes) [Disease or Syndrome]
Meta Mapping (851):
827 Obstructive (Obstructed) [Functional Concept]
827 Sleep [Organism Function]
827 APNOEA (Apnea) [Pathologic Function]
...

without
--best_mappings_only
Example: show sources (-G)

Phrase: “scorpion sting."

Meta Candidates (4):
- 1000 Scorpion sting \{MDR, DXP\} [Injury or Poisoning]
- 861 Sting (Sting Injury
  \{MTH, MSH, MDR, RCD, SNM, SNOMEDCT, SNMI, WHO\}) [Injury or Poisoning]
- 694 Scorpion (Scorpions
  \{LCH, MSH, MTH, SNM, SNOMEDCT, SNMI, CSP, RCD, NCBI\}) [Invertebrate]
- 694 SCORPION (Scorpion antigen \{MTH, LNC\}) [Immunologic Factor]

Meta Mapping (1000):
- 1000 Scorpion sting \{MDR, DXP\} [Injury or Poisoning]
Example: restrict to sources (-GR LCH)

Phrase: “scorpion sting.”

Meta Candidates (1):
   694 Scorpion (Scorpions {LCH}) [Invertebrate]

Meta Mapping (694):
   694 Scorpion (Scorpions {LCH}) [Invertebrate]
Example: restrict to STs (-J neop)

Phrase: “lung cancer.”

Meta Candidates (3):
1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]
1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]
861 Cancer (Malignant Neoplasms) [Neoplastic Process]

Meta Mapping (1000):
1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

Meta Mapping (1000):
1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]
Questions?
UMLS Documentation and Support

- UMLS homepage
- UMLSKS homepage
- UMLSUSERS-L
  - subscribe to discussion list
- NLM Customer Service email:
  - 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
Thank you