Customizing the UMLS Metathesaurus for Your Applications

William T. Hole, MD
Betsy L. Humphreys, MLS
Suresh Srinivasan, MS
Laura Roth, MLS

National Library of Medicine
Saturday, November 4, 2000 - 1:00-4:30 pm
Outline of Tutorial

- Why customize - Betsy Humphreys
- How to customize - Bill Hole
- A tool to help you customize - Laura Roth and Suresh Srinivasan
- Adding “local” terminology - Bill Hole
A few Metathesaurus Ingredients...
UMLS Source “Vocabularies”

- Widely varying purposes, structures, properties that do not add up to single ontology or view of the world:
  - Thesauri, e.g., MeSH
  - Statistical Classifications, e.g., ICD
  - Billing Codes, e.g., CPT
  - Clinical coding systems, e.g., SNOMED
  - Lists of controlled terms, e.g., HL7 valid values
How to combine them?

SNOMED, MeSH, ICD9, READ …

Meta Processor,
Alpha 0.001
Not really ….

“The Metathesaurus preserves the meanings, hierarchical connections, and other relationships between terms present in its source vocabularies, while adding certain basic information about each of its concepts and establishing new relationships between concepts and terms from different source vocabularies.”
It keeps getting bigger...
2000 UMLS Metathesaurus

- 730,000 concepts
- 1,338,650 “terms” (Eye, Eyes, eye = 1)
- 1,593,730 “strings”/concept names - (Eye, Eyes, eye = 3)
- >50 source vocabularies
Why Customize?

3 basic reasons

- Because *nobody* needs or wants all of it for any specific set of purposes
  - extraneous vs. pernicious content

- Because you don’t have the licenses required for operational use of all source vocabularies

- Because the default “preferred name” is not best for your applications
Possibly Extraneous, e.g.,

- Terms in languages other than English
- Redundant minor variations
- Procedure codes, when your application is focused on problems
Possibly Pernicious, e.g.,

- Terms that lack face validity
- Abbreviations and short forms
- Other less than beautiful “suppressible synonyms” already identified by NLM
- Unhelpful “views of the world”
Your query term is "prostate"

This query term has multiple concepts associated with it in the Metathesaurus. Select a concept and click on submit button to obtain information about that concept.

- **Prostate**
  - Semantic Type: Body Part, Organ, or Organ Component

- **Prostatic Diseases**
  - Semantic Type: Disease or Syndrome

- **Benign neoplasm of prostate**
  - Semantic Type: Disease or Syndrome
  - Semantic Type: Neoplastic Process

- **Carcinoma in situ of prostate**
  - Semantic Type: Neoplastic Process

- **Neoplasm of uncertain or unknown behavior of prostate**
  - Semantic Type: Neoplastic Process
Your query term is "ER"

This query term has multiple concepts associated with it in the Metathesaurus. Select a concept and click on submit button to obtain information about that concept.

- **Endoplasmic Reticulum**
  - Semantic Type:
    - Cell Component

- **Estrogen Receptors**
  - Semantic Type:
    - Amino Acid, Peptide, or Protein
  - Semantic Type:
    - Receptor

**Definition:**

Cytoplasmic proteins that bind estrogens and migrate to the nucleus where they regulate DNA transcription. Evaluation of the state of estrogen receptors in breast cancer patients has become clinically important.

**Definition:**

ER. Protein found on some breast cancer cells to which estrogen will attach. Breast cancer cells that are estrogen receptor positive (ER+) need the hormone estrogen to grow and usually respond to hormone treatment.

SUBMIT
Examples of customization

- Natural Language Processing version of the Metathesaurus, *excluding* “suppressible synonyms”, highly structured terms such as those in LOINC, etc.
Home

Metathesaurus

Semantic Network

SPECIALIST Lexicon

Expert Search

Download Results

Comments

Help

---

**UMLS Knowledge Source Server**

- AMECillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMEDINOCillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMIKacin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMOxicillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMOxicillin and CLAVULANATE:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMPHOTERICIN B:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMPICillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AMPICillin and SULBACTAM:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AZITHROMYCin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AZLOCillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- AZTREONAM:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- BACAMPICillin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
- BUTirosin:Susceptibility:Point in Time:Isolate and Serum:Ordinal:Serum Bactericidal Titer
Examples of customization

- PubMed’s UMLS/MeSH mapping table
Incidence of acute myocardial infarction and cause-specific mortality after transurethral treatments of prostatic hypertrophy.

Weisman KM, Larijani GE, Goldberg ME

Publication Types:
- Letter

MeSH Terms:
- Cause of Death
- Human
- Male
- Myocardial Infarction/mortality
- Myocardial Infarction/complications*
- Prostatic Hyperplasia/mortality
- Prostatic Hyperplasia/complications*
Examples of customization

- National Cancer Institute’s Metaphrase™ implementation, which changes default “preferred name” precedence
**BASIC CONCEPT INFORMATION**

**Concept Name:** Persian Gulf Syndrome

**UI:** C0282550

**Semantic Type:** Disease or Syndrome

**Definition:** Unexplained symptoms reported by veterans of the Persian Gulf War with Iraq in 1991. The symptoms reported include fatigue, skin rash, muscle and joint pain, headaches, loss of memory, shortness of breath, gastrointestinal and respiratory symptoms, and extreme sensitivity to commonly occurring chemicals. (Nature 1994 May 5;369(6475):8)

**Synonyms:**
- Gulf War Syndrome

**Sources:** MSH2000

**Other Languages:**
- Persianlahden oireyhtymae - Finnish
- GOLFE PERSIQUE, SYNDROME - French
- GUERRE GOLFE, SYNDROME - French
- SYNDROME GOLFE PERSIQUE - French
Disease or Syndrome

Definition(s)

MSH2000 Unexplained symptoms reported by veterans of the Persian Gulf War with Iraq in 1991. The symptoms reported include fatigue, skin rash, muscle and joint pain, headaches, loss of memory, shortness of breath, gastrointestinal and respiratory symptoms, and extreme sensitivity to commonly occurring chemicals. (Nature 1994 May 5;369(6475):8)

Synonym(s)

- Gulf war syndrome, NOS
- Persian Gulf Syndrome

Sources

MSH2000 SNMI98

Broader Concepts

Occupational Diseases

TOXIC EFFECTS OF NONMEDICINAL SUBSTANCES
Examples of customization

- CliniWeb International, which uses MeSH and several of its translations
Welcome to the new CliniWeb International!

CliniWeb is an index and table of contents to clinical information on the World Wide Web. It now allows search terms to be entered in five different languages (English, German, French, Spanish, Portuguese) and has direct links to MEDLINE searches via the PubMed system at the National Library of Medicine. The CliniWeb database can be accessed by:

- **Searching** - using the SAPHIRE look-up system
- **Browsing** - through the MeSH hierarchy

Additional information about CliniWeb is also available.

We welcome your feedback by e-mail. Please notify us if you know of clinical sites we have overlooked or disagree with any indexing term selections.
Customization is critical, but it **requires** a clear understanding of:

- Your functional requirements
- Characteristics of relevant UMLS source vocabularies
- Your license arrangements
- **-- and** Technical expertise
- Therefore, it is usually a team sport.
Outline of Tutorial

- Why customize - *Betsy Humphreys*
- *How to customize* - *Bill Hole*
- A tool to help you customize - *Laura Roth and Suresh Srinivasan*
- Adding “local” terminology - *Bill Hole*
How to Customize your Metathesaurus

- Limit vocabularies
- Limit languages
- Limit Relationships
- Aggregate by Relationships
- Add and Use Suppressibility
- Change Naming Precedence
- Select by attribute e.g. STYs
- Add local terminology
To work with Meta files:

- You will have to select the rows and columns you need, and combine them in ways which best meet your needs.

- You need your own programming and data management tools to work with the files, e.g.:
  - Unix: grep, sed, cut, awk, sort, join ...
  - RDBMS: Mysql, Access, Oracle, Ingres…
  - Other languages and extensions, e.g. Java, C++, Visual Basic, Perl; interfaces to B-trees, RDBMS …

- We supply MetamorphoSys, a tool to exclude vocabularies and alter naming, described later.

- Some helpful files, such as RDBMS load scripts, are being developed - check http://umlsinfo.nlm.nih.gov
Using Unique Identifiers:

Simplified example: Concept Unique Identifier **C0002871** links tables:

**MRCON: Concept names**
- C0002871|ENG|P|L0002871|PF|S0013742|Anemia|0|
- C0002871|ENG|S|L0376533|PF|S0500659|Oligocythemia of red blood cells|3|
- C0002871|FRE|P|L0162748|PF|S0227229|ANEMIE|2|

**MRSO: Sources (vocabulary information)**
- C0002871|L0002871|S0013742|MSH2000|MH|D000740|0|
- C0002871|L0376533|S0500659|SNMI98|SY|DC-10010|3|
- C0002871|L0162748|S0227229|INS2000|MH|D000740|3|
- C0002871|L0162748|S0227229|WHOFRE|PT|0544|2|

**MRSTY: Semantic Types**
- C0002871|T047|Disease or Syndrome|

**MRDEF: Definitions**
- C0002871|CSP98|subnormal levels or function of erythrocytes, resulting in symptoms of tissue hypoxia.|
- C0002871|MSH2000|A reduction in the number of circulating erythrocytes or in the quantity of hemoglobin.|
- C0002871|PDQ99|A condition in which the number of red blood cells is below normal.|
First, Select the data you need:

The ‘Big Five’ Relational files:

- **MRCON**: Concept Names
- **MRSO**: Source vocabulary information
- **MRSTY**: Semantic Types
- **MRREL**: Most Relationships
- **MRSAT**: Simple Attributes
### MRCON: Concept Names

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>TS</th>
<th>LUI</th>
<th>STT</th>
<th>SUI</th>
<th>STR</th>
<th>LRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0154009</td>
<td>ENG</td>
<td>P</td>
<td>L0180842</td>
<td>PF</td>
<td>S0245368</td>
<td>Benign neoplasm of prostate</td>
<td>0</td>
</tr>
<tr>
<td>C0154009</td>
<td>ENG</td>
<td>S</td>
<td>L0524759</td>
<td>PF</td>
<td>S0598915</td>
<td>Benign prostatic tumour</td>
<td>3</td>
</tr>
<tr>
<td>C0154009</td>
<td>ENG</td>
<td>S</td>
<td>L0524759</td>
<td>PF</td>
<td>S0598915</td>
<td>Prostate prostatic tumour</td>
<td>3</td>
</tr>
<tr>
<td>C0154009</td>
<td>ENG</td>
<td>S</td>
<td>L0033572</td>
<td>PF</td>
<td>S0999020</td>
<td>Prostate &lt;3&gt;</td>
<td>0</td>
</tr>
<tr>
<td>C0154009</td>
<td>ENG</td>
<td>S</td>
<td>L0033572</td>
<td>VO</td>
<td>S0077252</td>
<td>Prostate</td>
<td>3</td>
</tr>
<tr>
<td>C0154009</td>
<td>GER</td>
<td>P</td>
<td>L1258213</td>
<td>PF</td>
<td>S1500159</td>
<td>Gutartige Neubildung: Prostata</td>
<td>1</td>
</tr>
</tbody>
</table>

**CUI**  Unique identifier for concept (“The name that never changes”);
- Links ALL Concept information

**LAT** Language of Term

**TS** Term status

**LUI** Unique identifier for term (lexically similar strings, defined by lvgl)

**STT** String type

**SUI** Unique identifier for string

**STR** String

**LRL** Least Restriction Level (lowest license restriction* for a source of this string)
- lowest license restriction level for this string; see MRSO Source Restriction level

---

*License restriction levels are defined in the UMLS License Agreement. A copy of the License Agreement is provided in the Appendix*
Customizing with MRCON - Examples:

• Select or exclude sets of concepts by *CUI* identified elsewhere
• Select or exclude a language by *LAT*
• Select or exclude by Term Status TS, e.g.:
  
  TS="P" Preferred;
  TS="S" Synonym
  TS="s" Supressible Synonym

• Use Preferred Terms (*TS="P"*) and preferred form (*STT="PF"*)

• Select or exclude by Least Restriction Level (*LRL*), e.g.

  No additional restrictions *LRL* = 0

<table>
<thead>
<tr>
<th>CUI</th>
<th>LAT</th>
<th>TS</th>
<th>LUI</th>
<th>STT</th>
<th>SUI</th>
<th>STR</th>
<th>LRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>ENG</td>
<td>P</td>
<td>L0002871</td>
<td>PF</td>
<td>S0013742</td>
<td>Anemia</td>
<td>0</td>
</tr>
</tbody>
</table>
**MRSO: Source vocabulary information**

<table>
<thead>
<tr>
<th>CUI</th>
<th>LUI</th>
<th>SUI</th>
<th>SAB</th>
<th>TTY</th>
<th>SCD</th>
<th>SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0013742</td>
<td>LCH90</td>
<td>PT</td>
<td>U000235</td>
<td>0</td>
</tr>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>MH</td>
<td>D000740</td>
<td>0</td>
</tr>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0013742</td>
<td>MTH</td>
<td>PT</td>
<td>U000161</td>
<td>0</td>
</tr>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0013742</td>
<td>PSY94</td>
<td>PT</td>
<td>024503</td>
<td>0</td>
</tr>
</tbody>
</table>

...  

CUI       Unique identifier for concept  
LUI       Unique identifier for term  
SUI       Unique identifier for string  
SAB       Source abbreviation  
TTY       Term type in source  
SCD       Unique Identifier or code for string in source  
SRL       Source Restriction Level  

* on this slide and many which follow, trailing spaces have been added to align columns for readability. They do not exist in the Metathesaurus files.*
Source Restriction Level vs. Least Restriction Level

• **Source (Vocabulary) Restriction Level (SRL), in MRSO:**

  0 Means there are no additional restrictions

  1 - 3 Means there are additional restrictions as defined in the License Agreement.

• **Least Restriction Level (LRL), in MRCON:**

The lowest restriction level (SRL) for all vocabularies providing this string

*All Restrictions are defined and listed for each vocabulary in the UMLS License Agreement*
Customizing with MRSO - Examples:

• Better to use MetamorphoSys to select or exclude sources
  * it removes ALL source information - demo later

• Select or exclude a TTY, e.g.
  
  TTY=“PT” : Preferred term in source vocabulary;
  
  TTY=“AB” : Abbreviation in source vocabulary;

• Select or exclude or order by SCD, e.g.
  
  ICD-9-CM codes with 2 digits before decimal are Procedures

• Select or exclude by SRL (Source Restriction Level)

* SABs are listed in Manual Section B.2 and the SABDOC file
  
  TTYs are described Manual Section B.4 and the TTYDOC file
**Don’t assume you know what Semantic Types mean!**
Use Knowledge Source Server (http://umlsks.nlm.nih.gov)

- Browser display for STY “Finding”:
  ```
  Semantic Types
  | Entity
  | Conceptual Entity
  | Finding ...
  | Laboratory or Test Result
  Sign or Symptom
  ```

- Definitions: ...
- See a list of Metathesaurus concepts with Semantic type: “Finding”.
- Relations and Related Types for "Finding".
- View records for Finding in relational format

* **hint:** Tree numbers in SRDEF can be useful for walking hierarchies*
MRSTY: Semantic Types

<table>
<thead>
<tr>
<th>CUI</th>
<th>TUI</th>
<th>STY</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>T047</td>
<td>Disease or Syndrome</td>
</tr>
<tr>
<td>C0004057</td>
<td>T109</td>
<td>Organic Chemical</td>
</tr>
<tr>
<td>C0004057</td>
<td>T121</td>
<td>Pharmacologic Substance</td>
</tr>
</tbody>
</table>

CUI       Unique identifier of concept
TUI       Unique identifier of Semantic type
STY       Semantic type, defined in the Semantic Network.
Customizing with MRSTY - Examples:

• Select or exclude by Semantic Type or sets of Types

• Use Knowledge Source Server to see definitions, hierarchy, examples: http://umlsks.nlm.nih.gov

  e.g. select or exclude all Laboratory or Test Result

<table>
<thead>
<tr>
<th>CUI</th>
<th>TUI</th>
<th>STY</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0005798</td>
<td>T034</td>
<td>Laboratory or Test Result</td>
</tr>
</tbody>
</table>
### MRREL: Most Relationships

<table>
<thead>
<tr>
<th>CUI1</th>
<th>REL</th>
<th>CUI2</th>
<th>RELA</th>
<th>SAB</th>
<th>SL</th>
<th>MG</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>CHD</td>
<td>C0002891</td>
<td>isa</td>
<td>MSH2000</td>
<td>MTH</td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>RB</td>
<td>C0221016</td>
<td></td>
<td>MTH</td>
<td>MTH</td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>RL</td>
<td>C0002886</td>
<td>mapped_to</td>
<td>SNMI98</td>
<td>SNMI98</td>
<td></td>
</tr>
<tr>
<td>C0000665</td>
<td>RN</td>
<td>C0612003</td>
<td>mapped_to</td>
<td>MSH2000</td>
<td>MSH2000</td>
<td>G</td>
</tr>
</tbody>
</table>

- **CUI1**: Unique identifier of first concept
- **REL**: Relationship of second to first concept
- **CUI2**: Unique identifier of second concept
- **RELA**: Relationship attribute
- **SAB**: Abbreviation of the source of relationship
- **SL**: Source of relationship labels
- **MG**: Machine-generated and unverified indicator (optional)
MRREL: Most Relationships

C0002871 | CHD | C0002891 | isa | MSH2000 | MTH |

Anemia   Anemia, Neonatal

“Neonatal Anemia” is child of “Anemia” (REL = CHD)
“Neonatal Anemia” is a “Anemia” (RELA= isa)

SAB - Source of REL is MeSH;
SL - Source of labels (type of relationship) is MTH
MG - Not Machine generated.
MRREL: Most Relationships

Picking the Relationships you want:

*RELDOC file, A bonus file available on*  

<table>
<thead>
<tr>
<th>SAB</th>
<th>SL</th>
<th>REL</th>
<th>RELA</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD2000</td>
<td>ICD2000</td>
<td>CHD</td>
<td></td>
<td>19579</td>
</tr>
<tr>
<td>ICD2000</td>
<td>ICD2000</td>
<td>PAR</td>
<td></td>
<td>19579</td>
</tr>
<tr>
<td>MSH2000</td>
<td>MSH2000</td>
<td>CHD</td>
<td>connected_to</td>
<td>43</td>
</tr>
<tr>
<td>MTH</td>
<td>MTH</td>
<td>RB</td>
<td>ingredient_of</td>
<td>112</td>
</tr>
<tr>
<td>MTH</td>
<td>MTH</td>
<td>RB</td>
<td></td>
<td>157580</td>
</tr>
<tr>
<td>RAM99</td>
<td>RAM99</td>
<td>RO</td>
<td>clinically_similar</td>
<td>8200</td>
</tr>
<tr>
<td>UWDA99</td>
<td>UWDA99</td>
<td>CHD</td>
<td>isa</td>
<td>16591</td>
</tr>
<tr>
<td>UWDA99</td>
<td>UWDA99</td>
<td>CHD</td>
<td>part_of</td>
<td>7697</td>
</tr>
</tbody>
</table>
Customizing with MRREL - Examples:

• Select or exclude RELS, RELAs, and SAB or SL use RELDOC or determine suitability empirically

• Use selected RELS and RELAs to aggregate e.g., all ‘clinically_similar’

• Combine with other criteria
  e.g., selected RELs to concepts with selected Semantic Types
# MRSAT: Simple Attributes

<table>
<thead>
<tr>
<th>CUI</th>
<th>LUI</th>
<th>SUI</th>
<th>SCD</th>
<th>ATN</th>
<th>SAB</th>
<th>ATV</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0013742</td>
<td>D000740</td>
<td>MN</td>
<td>MSH2000</td>
<td>C15.378.71</td>
</tr>
<tr>
<td>C0002871</td>
<td>L0002871</td>
<td>S0414880</td>
<td>208/04453</td>
<td>SOS</td>
<td>PDQ99</td>
<td>secondary related condition</td>
</tr>
</tbody>
</table>

**CUI**  Unique identifier for concept or, for LT element only, LUI for term  
**LUI**  Unique identifier for term (optional)  
**SUI**  Unique identifier for string (optional)  
**SCD**  Unique identifier or code for entry in the source of the attribute  
**ATN**  Attribute name *(see manual, ATNDOC file)*  
**SAB**  Abbreviation of the source of the attribute  
**ATV**  Attribute value *(see manual, ATNDOC file)*

*note: few attribute values exceed 2,200 characters.*
Customizing with MRSAT - Examples:

- Wide variety of Attributes, many very detailed, described in manual or in *ATNDOC file* available on umlsinfo; e.g.:

<table>
<thead>
<tr>
<th>SAB</th>
<th>ATN</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH2000</td>
<td>SOS</td>
<td>88355</td>
</tr>
</tbody>
</table>

  -- Details of scope of concept in MeSH

<table>
<thead>
<tr>
<th>MTH</th>
<th>AM</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTH</td>
<td>ST</td>
<td>730155</td>
</tr>
</tbody>
</table>

  -- Ambiguous string indicator

  -- Status - ‘R’viewed or ‘U’nreviewed

- Attributes may be attached to CUIs, LUIs, or SUIs
- Choose the attributes you need!
Customizing with MRSAT, continued ...

You may also identify *sets* of concepts by attribute

- For example, to Select or exclude by Attribute for *Reviewed Status*
  - ATN “Status” = ST,
  - Attribute Value ATV = ‘R’ (Values are ‘R’viewed or ‘U’nreviewed)

- All Unreviewed Concepts (which are MeSH Supplementary concepts) have ATN ‘ST’ and ATV ‘U’

- To use or exclude these concepts, select their CUIs from the MRSAT rows with ATN ‘ST’ and ATV ‘U’
Select Other data you need:

- MRDEF: Definitions
- MRCXT: Contexts
- MRLO: Locator
- MRATX: Mapping one vocabulary to Another
- MRCOC: Co-occurrences

Indexes:
- MRXW.ENG, MRXW.FRE, Other languages ...
- MRXNW.ENG
- MRXNS.ENG
### MRDEF: Definitions

<table>
<thead>
<tr>
<th>CUI</th>
<th>SAB</th>
<th>DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>MSH2000</td>
<td>A reduction in the number of circulating erythrocytes or in the quantity of hemoglobin.</td>
</tr>
</tbody>
</table>

**CUI** Unique identifier for concept  
**SAB** Abbreviation of the source of the definition  
**DEF** Definition

**Note - there are nearly 32,000 definitions.**  
Some definitions are very long, up to 6 kb
MRCXT: Contexts - Examples:

**MSH2000**

Diseases (MeSH Category) [C]
- Virus Diseases [C2]
  - RNA Virus Infections [C2.782]
    - Arenaviridae Infections [C2.782.82]
      - Lassa Fever [C2.782.82.545]

**MSH2000**

Diseases (MeSH Category) [C]
- Virus Diseases [C2]
  - RNA Virus Infections [C2.782]
    - Hemorrhagic Fevers, Viral [C2.782.417]
      - Lassa Fever [C2.782.417.505]

**ICD10**

Certain infectious and parasitic diseases
- Arthropod-borne viral fevers and viral haemorrhagic fevers
  - Arenaviral haemorrhagic fever
- Lassa fever [A96.2]
### Using MRCXT: Contexts

-- See the documentation!

<table>
<thead>
<tr>
<th>CUI</th>
<th>SUI</th>
<th>SAB</th>
<th>SCD</th>
<th>CXN</th>
<th>CXL</th>
<th>RNK</th>
<th>CXS</th>
<th>CUI2</th>
<th>HCD</th>
<th>REL</th>
<th>XC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>ANC</td>
<td>1</td>
<td>MeSH</td>
<td>C0220876</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>ANC</td>
<td>2</td>
<td>Diseases (MeSH Category)</td>
<td>C0012674</td>
<td>C15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>ANC</td>
<td>3</td>
<td>Hemic and Lymphatic Diseases</td>
<td>C0018981</td>
<td>C15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>ANC</td>
<td>4</td>
<td>Hematologic Diseases</td>
<td>C0018939</td>
<td>C15.378</td>
<td>isa</td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>CCP</td>
<td></td>
<td>Anemia</td>
<td>C0002871</td>
<td>C15.378.71</td>
<td>isa</td>
<td>+</td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>CHD</td>
<td></td>
<td>Anemia, Aplastic</td>
<td>C0002874</td>
<td>C15.378.71.85</td>
<td>isa</td>
<td>+</td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>SIB</td>
<td></td>
<td>Blood Protein Disorders</td>
<td>C0005830</td>
<td>C15.378.147</td>
<td>isa</td>
<td>+</td>
</tr>
<tr>
<td>C0002871</td>
<td>S0013742</td>
<td>MSH2000</td>
<td>D000740</td>
<td>1</td>
<td>SIB</td>
<td></td>
<td>Bone Marrow Diseases</td>
<td>C0005956</td>
<td>C15.378.190</td>
<td>isa</td>
<td>+</td>
</tr>
</tbody>
</table>

**CUI**  Unique identifier of concept  
**SUI**  Unique identifier for string used in this context  
**SAB**  Source abbreviation  
**SCD**  Unique Identifier or code for string in that source  
**CXN**  The context number (to distinguish multiple contexts, same source, same SUI).  
**CXL**  Context member label, e.g., ANC for ancestor of this concept\  
**RNK**  For rows with a CXL value of ANC, the rank of the ancestors  
**CXS**  String for context member.  
**CUI2**  Unique concept identifier of context member (may be empty)  
**HCD**  Hierarchical number or code of context member in this source (optional).  
**REL**  Relationship of concept to parent, if applicable and known.  
**XC**  Plus(+) sign indicating that the concept has children (see manual)
MRCXT: Contexts - Caveats:

**MRCXT hierarchies are viewable and incomplete.**

- MRCXT was created to provide *reasonably viewable* hierarchies.
- Many vocabularies do not have full, principled hierarchies.
- Many hierarchies have too many children or siblings to be viewable.

**MRREL relationships are authoritative and complete.**
### MRLO: Locator

<table>
<thead>
<tr>
<th>CUI</th>
<th>ISN</th>
<th>FR</th>
<th>UN</th>
<th>SUI</th>
<th>SNA</th>
<th>SOUI</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0002871</td>
<td>DXPLAIN</td>
<td></td>
<td></td>
<td>S0352787</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>MEDLINE (1990–1995)</td>
<td>1666</td>
<td>*CITATIONS</td>
<td>S0013742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>MEDLINE (1996–Fall 1998)</td>
<td>707</td>
<td>*CITATIONS</td>
<td>S0013742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>PDQ</td>
<td></td>
<td></td>
<td>S0414880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0002871</td>
<td>QMR</td>
<td></td>
<td></td>
<td>S0013787</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CUI**  Unique identifier of concept  
**ISN**  Name of information source or database in which concept appears  
**FR**  Frequency count of number of occurrences of concept in the information source (optional)  
**UN**  Meaning of frequency (optional)  
**SUI**  Unique identifier of string if name used in information source appears in MRCON (optional)  
**SNA**  Actual name that occurs in the information source if not otherwise present in the Metathesaurus (optional)  
**SOUI**  Unique identifier of record in which the concept appears in source (optional)
## MRATX: Mapping one vocabulary to another

<table>
<thead>
<tr>
<th>CUI</th>
<th>SAB</th>
<th>REL</th>
<th>ATX</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0494958</td>
<td>MSH2000</td>
<td></td>
<td><strong>&lt;Kidney Failure, Acute&gt; AND &lt;Kidney Papillary Necrosis&gt;</strong></td>
</tr>
</tbody>
</table>

“Aacute renal failure with medullary necrosis”
in ICD-9-CM and ICD-10

- **CUI**: Unique identifier of concept to which the expression is related
- **SAB**: Abbreviation of source of terms *in expression*
- **REL**: Relationship of meaning of expression to main concept
- **ATX**: Associated expression (with complex internal syntax)

Most Metathesaurus concept *do not* have ATXs.
Most are *FROM ICD-9-CM TO* MeSH

Need feedback on usefulness
MRCOC: Co-occurrences

Co-occurrence information:

Most are MEDLINE co-occurrences

Also includes:

Patient record co-occurrences from CCPSS, the Canonical Clinical Problem Statement System from Vanderbilt University

Knowledge base co-occurrences from AI Rheum, a Rheumatology expert system
MRCOC: Co-occurrences

*(see manual)*

<table>
<thead>
<tr>
<th>CUI1</th>
<th>CUI2</th>
<th>SOC</th>
<th>COT</th>
<th>COF</th>
<th>COA</th>
</tr>
</thead>
<tbody>
<tr>
<td>C00002871</td>
<td>C0000727</td>
<td>MBD</td>
<td>L</td>
<td>1</td>
<td>BL=1, ET=1</td>
</tr>
<tr>
<td>C00002871</td>
<td>C0000737</td>
<td>MBD</td>
<td>L</td>
<td>1</td>
<td>ET=1</td>
</tr>
<tr>
<td>C00002871</td>
<td>C0000772</td>
<td>MBD</td>
<td>L</td>
<td>2</td>
<td>CN=2</td>
</tr>
<tr>
<td>C0000731</td>
<td>C0000737</td>
<td>CCPSS99</td>
<td>PP</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

- **CUI1**: Unique identifier of first concept
- **CUI2**: Unique identifier of second concept
- **SOC**: Abbreviation of the Source of co-occurrence information if applicable
- **COT**: Type of co-occurrence
- **COF**: Frequency of co-occurrence, if applicable
- **COA**: Attributes of co-occurrence, if applicable
Indexes:
MRXW.ENG, MRXW.FRE, Other languages … lowercased words
MRXNW.ENG … lowercased normalized words
MRXNS.ENG … lowercased normalized strings

Example:
MRCON:
C0026927|ENG|P|L0026927|PF|S1396311|Mycobacteria, Atypical|0|

MRXW.ENG:
ENG|atypical|C0026927|L0026927|S1396311|
ENG|mycobacteria|C0026927|L0026927|S1396311|

MRXNW.ENG:
ENG|atypical|C0026927|L0026927|S1396311|
ENG|mycobacterium|C0026927|L0026927|S1396311|

MRXNS.ENG:
ENG|atypical mycobacterium|C0026927|L0026927|S1396311|
Using Indexes: The Specialist Lexical Programs

• Very powerful, standardized tools are available as part of the UMLS Specialist Lexicon and Lexical Programs

• Source code and binary programs for Unix and NT are included

• “norm” will process your terms and queries to match the Meta indexes

• “lvg” does much more …

• All are included in the UMLS distribution
Recap: How to Customize the Metathesaurus

- Limit vocabularies, languages
  - MetamorphoSys does this best (demo is next!)
- Use Semantic Types
- Limit and aggregate by Relationships
  - MRREL
- Add and Use Suppressibility
  - MRCON “TS”
- Change Naming Precedence
  - MetamorphoSys - GUI or custom configuration
- Select by attribute e.g. ST
  - MRSAT
- Use other MR Files as needed …
- Adding “local” terminology (after demo)
Outline of Tutorial

- Why customize - Betsy Humphreys
- How to customize - Bill Hole
- A tool to help you customize - Laura Roth and Suresh Srinivasan
- Adding “local” terminology - Bill Hole
MetamorphoSys

- A tool distributed for use with the UMLS Knowledge Sources
  - Already present in UMLS distribution

- Multi-platform Java software

- Creates a customized version of the Metathesaurus
Why Use MetamorphoSys?

- Exclude vocabularies as required by the UMLS License Agreement
  - Default is to select only vocabularies that have no special restrictions (category zero)
  - Example: Remove SNOMED if user does not have an agreement with CAP for use

- Exclude vocabularies not needed for users’ specific purposes
  - Example: Nursing vocabularies
Use (cont’d)

- Alter “preferred name” precedence
  - Precedence is determined by MRRANK table supplied by NLM
  - Highest precedence source provides the “preferred name” of the concept name
  - User may want the preferred names of concepts to be other than the default provided
  - Example: User wants Read vocabulary to be highest precedence source
Use (cont’d)

- Add suppressibility to any Source-TTY combination
  - Metathesaurus already marks some combinations as suppressible by setting TS=s in MRCON
  - Suppressibility allows users to program their applications to remove certain groups of terms that may not be suitable
  - Example: Remove LOINC for Natural Language Processing purposes
How does MetamorphoSys work?

- What it does: removes all information from MR* files that is supplied by the excluded vocabularies
  - Allows added source-termgroup suppressibility
  - Allows altered “preferred name” precedence

- What results: A full Metathesaurus of the remaining vocabularies
How to Use MetamorphoSys

- Machine requirements
- Configuration interface
- Demo on subset of Metathesaurus
Machine Requirements

- A minimum of 256 MB of physical memory, as well as 8 GB recommended free disk space
  - Full UMLS distribution needs to be present

- Can be run on: Sun Solaris, Windows NT, Linux and Windows 98
Configuration Interface

- User sees a Java graphical user interface
- Interactively prompts for information
- Started by the MetamorphoSys program once UMLS distribution has been unpacked
  - Found in UMLS2000/META/METAMSYS/directory
  - MetamorphoSys.sh in the UNIX environment
  - MetamorphoSys.bat in Windows
Configuration

- Simple to use

- No configuration needed to exclude all non-zero category vocabularies
  - Default is a Metathesaurus of just category zero vocabularies, i.e., to exclude those with category 1, 2 or 3 restrictions (see License Agreement in Appendix for details)
Configuration

- Four tabs present in the MetamorphoSys interface
  - Files/folders
  - Sources
  - Precedence
  - Term Status
Files/Folders

- Indicate where UMLS distribution is present on user’s machine
- Indicate where the customized Metathesaurus should go once MetamorphoSys is finished
- Default directories are provided
Please choose the folders to use for the following purposes:
(Note that folder (2) below will require about 2699918K bytes of free space.)
(1) The UMLS installation folder/directory and
Choose UMLS installation folder...
/UMLS2000/META

Choose target folder...
/UMLS2000/METASUBSET
Sources

- All sources are alphabetically listed
- Ones highlighted are the ones to be excluded
- Can change to include or exclude any vocabulary
Sources Configuration
Precedence

- MTH source is the default highest precedence source
- Sources are arranged in list alphabetically and not by precedence
- Highlighting a source will select it as the highest precedence source
  - Only one source can be chosen
Select a single source whose terms you want to have the highest precedence, overriding the default. This will cause terms from this source to be used to represent the name of concepts in which they occur. For more complex ordering, please refer to the documentation.
Term Status

- Used to add suppressibility
- All Source-TTY combinations that are suppressible are highlighted
- Cannot change these to be non-suppressible in interface
- New combinations can be highlighted to make them suppressible
Select one or more source and term type combinations that you wish made suppressible. Default selections are highlighted and are shown with a leading asterisk.
(Note: the default selections must remain suppressible.)

<table>
<thead>
<tr>
<th>Select one or more suppressible term types</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD10AE</td>
</tr>
<tr>
<td>ICD10</td>
</tr>
<tr>
<td>ICD10AE</td>
</tr>
<tr>
<td>ICD10</td>
</tr>
<tr>
<td>*ICD10AE</td>
</tr>
<tr>
<td>*ICD10</td>
</tr>
<tr>
<td>ICDMAE</td>
</tr>
<tr>
<td>ICD10AM</td>
</tr>
<tr>
<td>ICDMAE</td>
</tr>
<tr>
<td>ICD10AM</td>
</tr>
<tr>
<td>*ICDMAE</td>
</tr>
<tr>
<td>*ICD10AM</td>
</tr>
<tr>
<td>ICD91</td>
</tr>
<tr>
<td>ICD89</td>
</tr>
<tr>
<td>PDQ99</td>
</tr>
<tr>
<td>PDQ99</td>
</tr>
<tr>
<td>*ICD9000</td>
</tr>
</tbody>
</table>
Running MetamorphoSys

- Once configuration is defined, a simple file selection starts subsetting of Metathesaurus
Output

- Once started, a message tells user that the program has started and the interface disappears.
- Program exits automatically upon completion of data subsetting.
- Your customized Metathesaurus will be present in defined directory.
  - UMLS2000/METASUBSET by default.
Outline of Tutorial

- Why customize - Betsy Humphreys
- How to customize - Bill Hole
- A tool to help you customize - Laura Roth and Suresh Srinivasan
- Adding “local” terminology - Bill Hole
MetamorphoSys Output
Distributed 2000 Metathesaurus
Data for C0221233
### Distributed Data for C0221233...

<table>
<thead>
<tr>
<th>C0221233</th>
<th>ENG</th>
<th>P</th>
<th>L0294378</th>
<th>PF</th>
<th>S0625803</th>
<th>Cyclic edema</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
<td>VW</td>
<td>S0368795</td>
<td>EDEMA, CYCLIC</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
<td>VW</td>
<td>S0369867</td>
<td>Edema, cyclic</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>PF</td>
<td>S0368798</td>
<td>EDEMA, PERIODIC</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VC</td>
<td>S0369871</td>
<td>Edema, periodic</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VW</td>
<td>S0395649</td>
<td>Periodic, edema</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0803934</td>
<td>PF</td>
<td>S0852190</td>
<td>Cyclic oedema</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C0221233</th>
<th>L0294378</th>
<th>S0368795</th>
<th>DXP94</th>
<th>SY</th>
<th>NOCODE</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0369867</td>
<td>SNM2</td>
<td>PT</td>
<td>M-36680</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0625803</td>
<td>SNMI98</td>
<td>PT</td>
<td>M-36480</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0368798</td>
<td>DXP94</td>
<td>SY</td>
<td>NOCODE</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0369871</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0395649</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0803934</td>
<td>S0852190</td>
<td>SNMI98</td>
<td>SY</td>
<td>M-36480</td>
<td>3</td>
</tr>
</tbody>
</table>
Distributed Data for C0221233...

<table>
<thead>
<tr>
<th>MRREL</th>
<th>C0013604</th>
<th>CHD</th>
<th>C0221233</th>
<th>SNM2</th>
<th>SNM2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C0013604</td>
<td>CHD</td>
<td>C0221233</td>
<td>SNMI98</td>
<td>SNMI98</td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>PAR</td>
<td>C0013604</td>
<td>SNM2</td>
<td>SNM2</td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>PAR</td>
<td>C0013604</td>
<td>SNMI98</td>
<td>SNMI98</td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>RB</td>
<td>C0238094</td>
<td>MTH</td>
<td>MTH</td>
</tr>
<tr>
<td></td>
<td>C0238094</td>
<td>RN</td>
<td>C0221233</td>
<td>MTH</td>
<td>MTH</td>
</tr>
</tbody>
</table>

| MRSAT   | C0221233 | || DA | MTH | 19950417 |
|---------|----------|------|------|------|-------|
|         | C0221233 | || MR | MTH | 20000101 |
|         | C0221233 | || ST | MTH | R     |
1. Default Subset for C0221233

Removing all restricted sources
## Default Subset for C0221233 (Before)

<table>
<thead>
<tr>
<th>Code</th>
<th>Eng</th>
<th>P</th>
<th>L0294378</th>
<th>PF</th>
<th>S0625803</th>
<th>Cyclic edema</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
<td>VW</td>
<td>S0368795</td>
<td>EDEMA, CYCLIC</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
<td>VW</td>
<td>S0369867</td>
<td>Edema, cyclic</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>PF</td>
<td>S0368798</td>
<td>EDEMA, PERIODIC</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VC</td>
<td>S0369871</td>
<td>Edema, periodic</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VW</td>
<td>S0395649</td>
<td>Periodic, edema</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0803934</td>
<td>PF</td>
<td>S0852190</td>
<td>Cyclic oedema</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>L0294378</th>
<th>S0368795</th>
<th>DXP94</th>
<th>SY</th>
<th>NOCODE</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0369867</td>
<td>SNM2</td>
<td>PT</td>
<td>M−36680</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0625803</td>
<td>SNMI98</td>
<td>PT</td>
<td>M−36480</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0368798</td>
<td>DXP94</td>
<td>SY</td>
<td>NOCODE</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0369871</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0395649</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0803934</td>
<td>S0852190</td>
<td>SNMI98</td>
<td>SY</td>
<td>M−36480</td>
<td>3</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0368795</td>
<td>DXP94</td>
<td>SY</td>
<td>NOCODE</td>
<td>0</td>
</tr>
</tbody>
</table>
Default Subset for C0221233 (After)

<table>
<thead>
<tr>
<th>MRCON</th>
<th>C0221233</th>
<th>ENG</th>
<th>P</th>
<th>L0294378</th>
<th>PF</th>
<th>S0368795</th>
<th>EDEMA, CYCLIC</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>PF</td>
<td>S0368798</td>
<td>EDEMA, PERIODIC</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VC</td>
<td>S0369871</td>
<td>Edema, periodic</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
<td>VO</td>
<td>S0395649</td>
<td>Periodic, edema</td>
<td>0</td>
</tr>
<tr>
<td>MRSO</td>
<td>C0221233</td>
<td>L0294378</td>
<td>S0368795</td>
<td>DXP94</td>
<td>SY</td>
<td>NOCODE</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>L0298620</td>
<td>S0368798</td>
<td>DXP94</td>
<td>SY</td>
<td>NOCODE</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>L0298620</td>
<td>S0369871</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C0221233</td>
<td>L0298620</td>
<td>S0395649</td>
<td>ICD91</td>
<td>IT</td>
<td>995.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>MRREL</td>
<td>C0221233</td>
<td>RB</td>
<td>C0238094</td>
<td></td>
<td>MTH</td>
<td>MTH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C0238094</td>
<td>RN</td>
<td>C0221233</td>
<td></td>
<td>MTH</td>
<td>MTH</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Altering Precedence for C0221233

Making ICD91 the highest precedence source
### Changing Precedence for C0221233 (Before)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Precedence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>P</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
<td>S</td>
<td>L0803934</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Precedence</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0368795</td>
<td>DXP94</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0369867</td>
<td>SNM2</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
<td>S0625803</td>
<td>SNM198</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0368798</td>
<td>DXP94</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0369871</td>
<td>ICD91</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
<td>S0395649</td>
<td>ICD91</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0803934</td>
<td>S0852190</td>
<td>SNM198</td>
</tr>
</tbody>
</table>
## Changing Precedence for C0221233 (After)

| MRCON | |  |
|--------|--------|  |
| C0221233 | ENG | P | L0298620 | PF | S0369871 | Edema, periodic | 0 |  |
| C0221233 | ENG | P | L0298620 | VC | S0368798 | EDEMA, PERIODIC | 0 |  |
| C0221233 | ENG | P | L0298620 | VO | S0395649 | Periodic, edema | 0 |  |

| MRSO | |  |
|--------|--------|  |
| C0221233 | L0294378 | S0368795 | DXP94 | SY | NOCODE | 0 |  |
| C0221233 | L0294378 | S0369867 | SNM2 | PT | M-36680 | 3 |  |
| C0221233 | L0294378 | S0625803 | SNM198 | PT | M-36480 | 3 |  |
| C0221233 | L0298620 | S0368798 | DXP94 | SY | NOCODE | 0 |  |
| C0221233 | L0298620 | S0369871 | ICD91 | IT | 995.1 | 0 |  |
| C0221233 | L0298620 | S0395649 | ICD91 | IT | 995.1 | 0 |  |
| C0221233 | L0803934 | S0852190 | SNM198 | SY | M-36480 | 3 |  |
3. Adding Suppressibility for C0221233

Suppressing SNOMED synonyms (SNMI98-SY) strings
| C0221233 | ENG | P | L0294378 | PF | S0625803 | Cyclic edema | 3 |
| C0221233 | ENG | P | L0294378 | VW | S0368795 | EDEMA, CYCLIC | 0 |
| C0221233 | ENG | P | L0294378 | VW | S0369867 | Edema, cyclic | 3 |
| C0221233 | ENG | S | L0298620 | PF | S0368798 | EDEMA, PERIODIC | 0 |
| C0221233 | ENG | S | L0298620 | VC | S0369871 | Edema, periodic | 0 |
| C0221233 | ENG | S | L0298620 | VW | S0395649 | Periodic, edema | 0 |
| C0221233 | ENG | S | L0803934 | PF | S0852190 | Cyclic oedema | 3 |

| C0221233 | L0294378 | S0368795 | DXP94 | SY | NOCODE | 0 |
| C0221233 | L0294378 | S0369867 | SNM2 | PT | M-36680 | 3 |
| C0221233 | L0294378 | S0625803 | SNMI98 | PT | M-36480 | 3 |
| C0221233 | L0298620 | S0368798 | DXP94 | SY | NOCODE | 0 |
| C0221233 | L0298620 | S0369871 | ICD91 | IT | 995.1 | 0 |
| C0221233 | L0298620 | S0395649 | ICD91 | IT | 995.1 | 0 |
| C0221233 | L0803934 | S0852190 | SNMI98 | SY | M-36480 | 3 |
### Adding Suppressibility for C0221233 (After)

<table>
<thead>
<tr>
<th>MRCON</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
<tr>
<td>C0221233</td>
<td>ENG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MRSO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C0221233</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0294378</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0298620</td>
</tr>
<tr>
<td>C0221233</td>
<td>L0803934</td>
</tr>
</tbody>
</table>


MetamorphoSys Configuration

- GUI and “Batch” Components
- Data in the “config” Subdirectory,
  e.g., sources.to.remove, prec.order
- GUI creates “Batch” config files
- Or Hand-Edit and Re-run
Why Hand-Edit?

- Better precedence control
- Saving configuration between runs
- Not for novices!
- Details are on http://umlsinfo.nlm.nih.gov
### MetamorphoSys Configuration Files

<table>
<thead>
<tr>
<th>File</th>
<th>Use</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources.to.remove</td>
<td>Batch</td>
<td>Abbreviations of all sources that will be removed – one per line;</td>
</tr>
<tr>
<td>Prec.order</td>
<td>Batch</td>
<td>Source abbreviation and term type in order of precedence; one per line. Leading asterisk implies suppressible combination</td>
</tr>
<tr>
<td>Mmsys.properties</td>
<td>both</td>
<td>Properties file; attribute=value format, e.g., begin=true</td>
</tr>
</tbody>
</table>
## MetamorphoSys Help Files

<table>
<thead>
<tr>
<th>File</th>
<th>Use</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source.txt</td>
<td>GUI</td>
<td>Abbreviation and description of all sources</td>
</tr>
<tr>
<td>Sources.category&lt;n&gt;</td>
<td>n/a</td>
<td>Abbreviations of sources for each restriction level: 0, 1, 2, 3</td>
</tr>
<tr>
<td>Prec.txt</td>
<td>GUI</td>
<td>Source and term types in NLM’s default order of precedence</td>
</tr>
</tbody>
</table>
sources.to.remove file

- Default is to remove all sources with restriction level > 0
- Edit with any text editor - add additional source abbreviations (or remove existing ones)
- Ensure that added ones are in source.txt
- Save file prior to re-run
prec.order file

- Order of Source-TTY lines decides precedence (lines in decreasing order of precedence)
- Edit with any text editor to change order
- Make additional ones suppressible - by adding a leading ‘*’
- Save file prior to re-run
Expert(!) Mode Summary

- Edit sources.to.remove, prec.order and mmsys.properties as needed
- Change to installation directory
- Run the “batch” component directly
- Full details on umlsinfo
Post Subset Examples

- Removing Suppressible Synonyms
  - awk -F'|' '$3!=“s”{print $1}' MRCON
  - SELECT CUI FROM MRCON WHERE TS!=‘s’

- Select concepts by STY
  - join -F'|' MRSTY MRCON | awk -F'|' '$2==“Disease or Syndrome”{print $1}'
  - SELECT MRCON.CUI FROM MRCON, MRSTY WHERE MRCON.CUI=MRSTY.CUI AND MRSTY.STY=‘Disease or Syndrome’
Notes

- Resulting STT not always accurate without LVG (see umlsinfo.nlm.nih.gov)
- Sort order - By CUI and ASCII within CUI
- Save your configuration files if needed
- Next version will be better!
Outline of Tutorial

- Why customize - Betsy Humphreys
- How to customize - Bill Hole
- A tool to help you customize - Laura Roth and Suresh Srinivasan
- *Adding “local” terminology* - Bill Hole
Adding Terminology

Create Distinct Unique Identifiers for *your* Terminology...

- e.g., for your concepts, use: ‘CA000001 …’ as CUIs instead of Meta’s ‘C0000001 ….’ for CUIs
- Similarly, use ‘LA000001 …’ for LUIs and ‘SA000001 …’ for SUIs, as needed
- Create a table recording your UIs, and for mapping them to UMLS UIs
Adding Terminology, Continued

Which of your terms are Meta Synonyms?

- Normalize them and look for matches in the Normalized String Index (MRXNS).
- Use other sensible approaches to searching:
  - normalized word searches;
  - explore alternate naming styles and conventions
- Use Meta CUIs for Synonyms

* See “Discovering Missed Synonymy in a Large Concept-Oriented Metathesaurus” - AMIA 2000 paper (Session S7)
Adding Terminology, Continued

**Bonus - add relationships ...**

- As you look for Meta Synonyms, add *relationships to Meta*

- Assign a REL and RELA to label the particular kinds of relationships you need and will use, e.g. to map or aggregate
Updating to a New Meta Release

- Repeat MetamorphoSys processing scripts from previous release
- Re-use previously found UIs for your terms to map synonyms, etc.
- Check for new synonyms of your terms which were not Meta synonyms before
- Check for any deleted CUIs and map them
Incremental updates are coming!

Same procedures will apply to the changes

- Update consists of deletes and adds only;
- Any changed concept is deleted, then added with the changes
- Any deleted concept is deleted
- Any new Concept is added
Online Resources:

WWW:
http://www.nlm.nih.gov/research/umls/

E-mail:
umlsmeta@nlm.nih.gov
umlslex@nlm.nih.gov
umlsnet@nlm.nih.gov
umlsks@nlm.nih.gov

umls-users listserv:

To subscribe to the listserv, send a message to listserv@nlm.nih.gov which includes the following line:

subscribe umls-users

To post a message to the umls-users listserv AFTER subscribing: send email to: umls-users@nlm.nih.gov