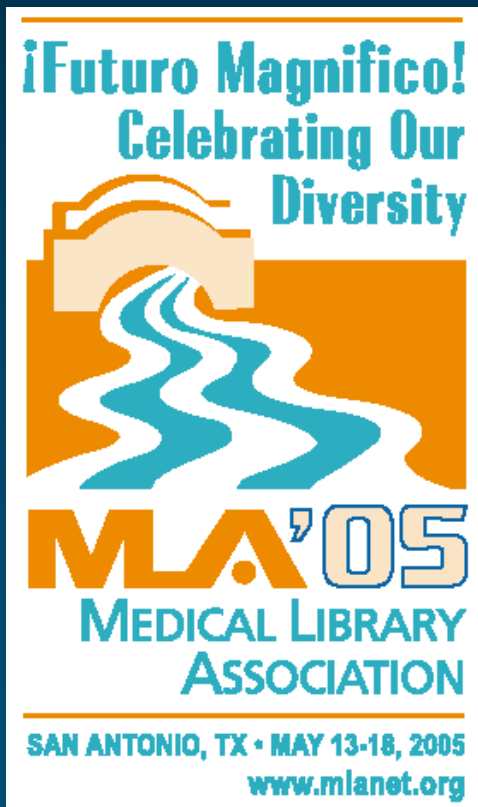# Unified Medical Language System® (UMLS®)

NLM Presentation Theater
MLA 2005
May 16 & 17, 2005

Rachel Kleinsorge

What does UMLS stand for?

- **U**nified
- **M**edical
- **L**anguage
- **S**ystem

UMLS®
Unified Medical Language System®

NLM

The UMLS is:

Metathesaurus   Semantic Network   SPECIALIST Lexicon & Lexical Tools

Concepts   Categories and relationships   Lexicon, lexical databases and programs

The Knowledge Sources
(delivered as machine readable files)

NLM

UMLS 3 Knowledge Sources

◆ Metathesaurus
- Over 100 source vocabularies
- Over 1 million concepts
- Inter-concept relationships

◆ Semantic Network
- 135 Semantic types (broad categories)
- 54 Semantic relations (between categories)

◆ Lexical resources
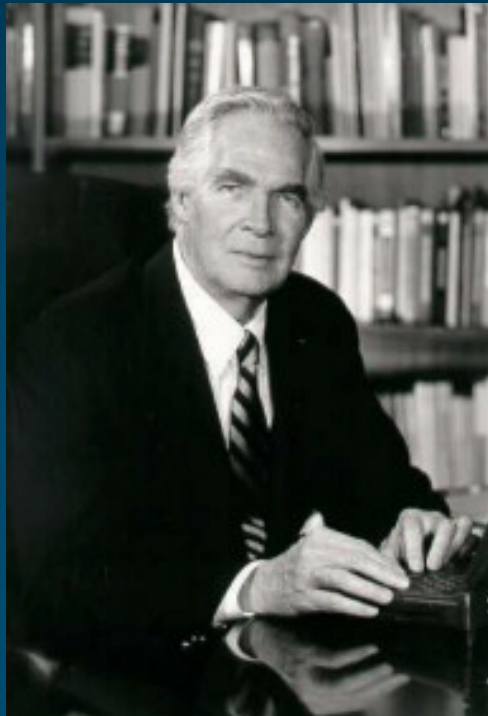- SPECIALIST Lexicon
- Lexical tools (programs and databases)

History of the UMLS

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

◆ Started at National Library of Medicine, 1986

◆ "Long-term R&D project"

◆ Complementary to IAIMS

(Integrated Academic Information Management Systems)

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

· The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.

· The second is the distribution of useful information among many disparate databases and systems.»

NLM

UMLS Objectives

◆ Knowledge Sources used to overcome:

- disparities in language format
- Ex: atrial fibrillation, auricular fibrillation, af

- disparities in granularity and perspective
- Ex: Contusions, hematoma, bruise
- Ex: Instruct patient to promptly report nosebleeds and excessive bruising (NIC), Epistaxis (MeSH)

- problems in mapping and aggregating within and across databases and systems

UMLS in Practice

◆ Intellectual "middleware" = a set of multi-purpose tools for system developers

◆ Databases: 3 separate sets of relational files

◆ Tools:

  ● **MetamorphoSys (installation and customization)**

  ● **Web interface Knowledge Source Server (UMLSKS)**

  ● **Application programming interfaces**

  ● **lvg (lexical programs)**

  ● **RRF Subset Browser**

The UMLS is *not* an end-user application

UMLS Uses

◆ Information retrieval

◆ Thesaurus construction

◆ Natural language processing

◆ Automated indexing

◆ Electronic patient records

License Agreements

◆ Semantic Network, SPECIALIST Lexicon, and Lexical Programs
  - terms and conditions of use online

◆ Metathesaurus
  - liscense agreement process
  - some 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.

NLM

What is the UMLS?

Overview through an example

Addison's Disease in medical vocabularies

◆ Synonyms: different terms

- Addisonian syndrome   ] eponym

- Bronzed disease
- Addison melanoderma   symptoms
- Asthenia pigmentosa

- Primary adrenal deficiency
- Primary adrenal insufficiency
- Primary adrenocortical insufficiency   clinical variants
- Chronic adrenocortical insufficiency

◆ Contexts: different hierarchies

NLM

Metathesaurus gathers and organizes terms

◆ Synonymous terms clustered into a concept

◆ Preferred term is chosen

◆ Unique identifier (CUI) is assigned

Adrenal gland diseases	MeSH	D000307
Adrenal disorder	AOD	0000005418
Disorder of adrenal gland	Read	C15z.
Diseases of the adrenal glands	SNOMED	DB-70000

C0001621

Adrenal Gland Diseases

NLM

Cluster of synonymous terms

Concept
C0001621

Term
L0001621

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

[…]

Term
L0041793

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

Term
L0161347

S0225481 *ADRENAL DISORDER*
S0627685 DISORDER ADRENAL (NOS)

[…]

Term
L0181041

S0632950 *Disorder of adrenal gland*
S0354509 Adrenal Gland Disorders

[…]

Term
L0368399

S0586222 *Adrenal disease*
S0466921 ADRENAL DISEASE, NOS

[…]

Term
L1279026

S1520972 *Nebennierenkrankheiten*

GER

Term
L0162317

S0226798 *SURRENALE, MALADIES*

FRE

[…]

Semantic Network

◆ Organizes terms into 135 broad subject categories
  - Semantic Types (Clinical Drug, Virus)
  - Addison's Disease **Semantic Type: Disease or Syndrome**

◆ Defines 54 Semantic Relationships
  - Links between categories (isa, causes, treats)
  - Ex: Virus **causes** Disease or Syndrome

◆ Together, types and relations:
  - Form the structure of the semantic network
  - Broadly categorize the biomedical domain

NLM

Lexical Tools
process terms and text

Organization is accomplished using:

◆ Normalization

◆ Semantic pre-processing

◆ UMLS editors

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

NLM

Summary

Metathesaurus:

◆ clusters terms into concepts – assigns unique identifier

Semantic Network:

◆ defines relationships between concepts, organizes concepts into categories

Lexicon and Lexical Tools:

◆ process terms for entry into the Metathesaurus

NLM