

Ontology and Taxonomy Coordinating Working Group

MITRE - McLean, VA

October 5, 2005

## The Unified Medical Language System

### *Overview*

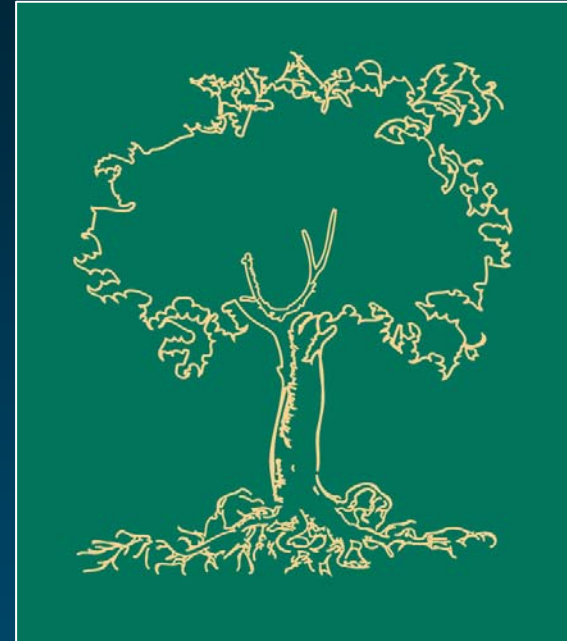


*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA

# What does UMLS stand for?

- ◆ Unified
- ◆ Medical
- ◆ Language
- ◆ System



UMLS<sup>®</sup>  
Unified Medical Language System<sup>®</sup>  
UMLS Metathesaurus<sup>®</sup>



# Motivation

- ◆ Started in 1986
- ◆ National Library of Medicine
- ◆ “Long-term R&D project”

«[...] 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.»



# Source Vocabularies

(2005AB)

- ◆ 133 source vocabularies contributing concept names
- ◆ ~80 families of vocabularies
  - multiple translations (e.g., MeSH, ICPC, ICD-10)
  - variants (American-English equivalents, Australian extension/adaptation)
  - subsequent editions usually considered distinct families (ICD: 9-10; DSM: IIR-IV)
- ◆ Broad coverage of biomedicine
- ◆ Common presentation



# Biomedical terminologies

## ◆ General vocabularies

- anatomy (UWDA, Neuronames)
- drugs (RxNorm, First DataBank, Micromedex)
- medical devices (UMD, SPN)

## ◆ Several perspectives

- clinical terms (SNOMED CT)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- data exchange terminologies (HL7, LOINC)

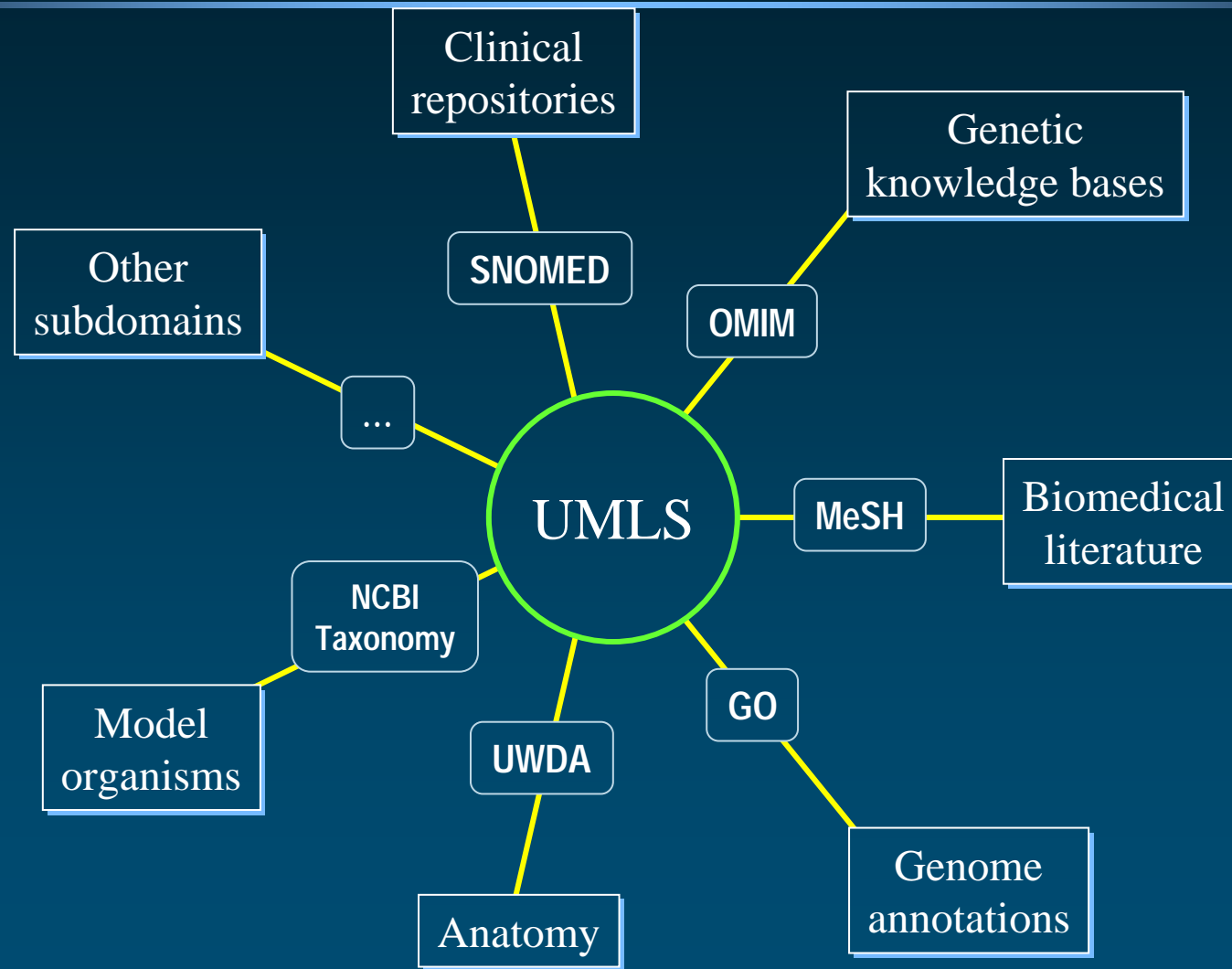


# Biomedical terminologies (cont'd)

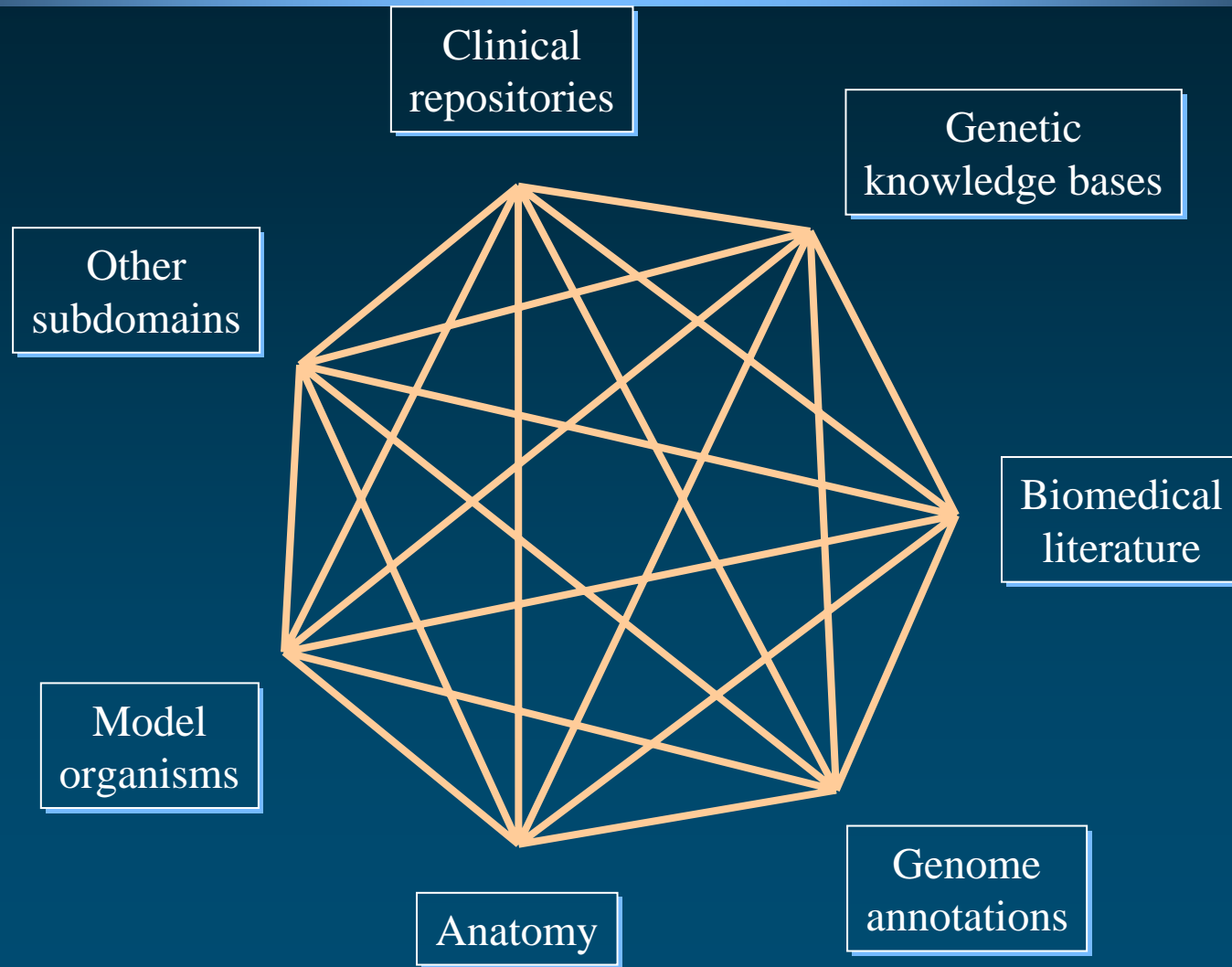
- ◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
  - genomics (GO, OMIM, HUGO)
- ◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)

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

# Integrating subdomains



# Integrating subdomains

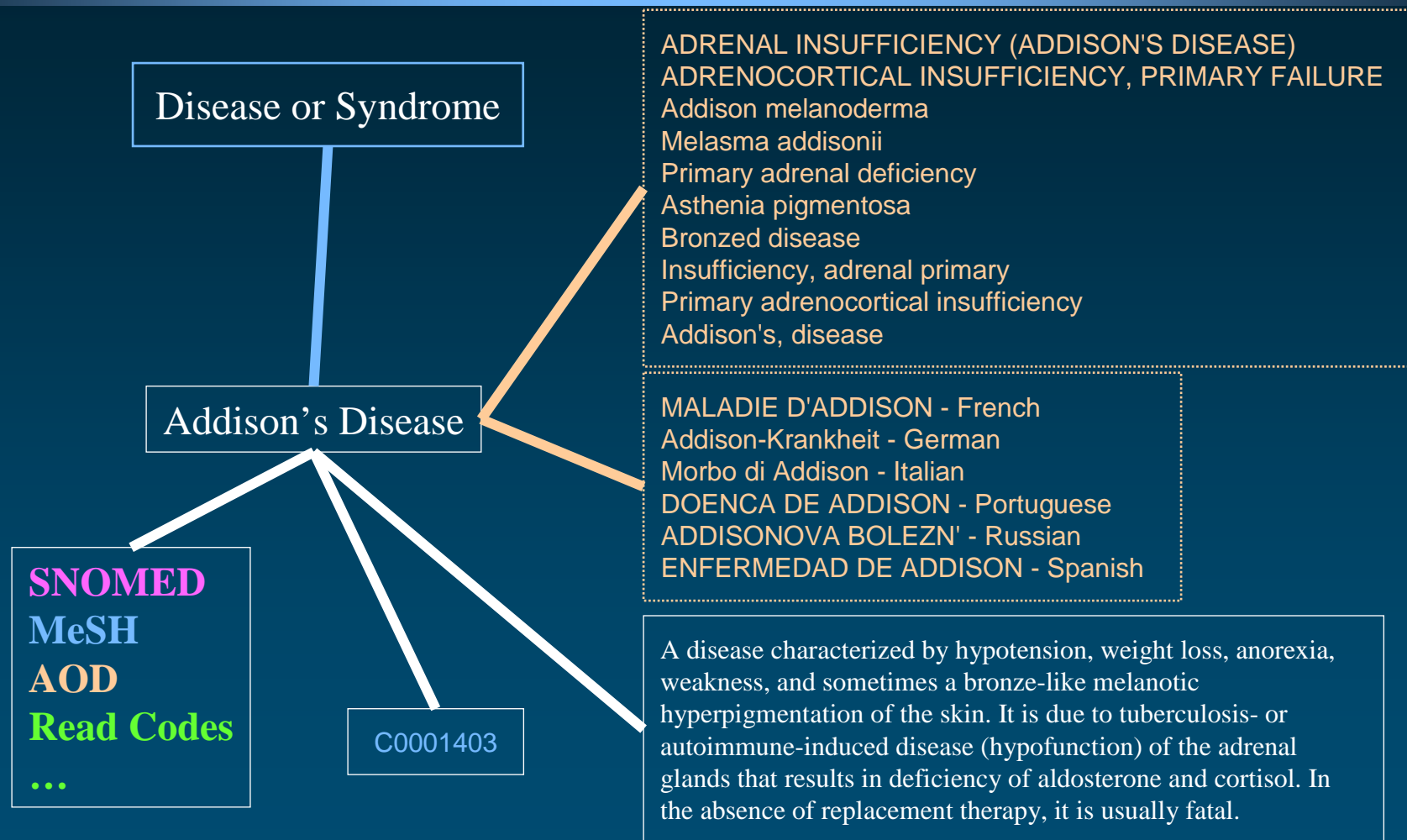




# UMLS: 3 components

- ◆ Metathesaurus
  - Concepts
  - Inter-concept relations
- ◆ Semantic Network
  - Semantic types
  - Semantic network relations
- ◆ Lexical resources
  - SPECIALIST Lexicon
  - Lexical tools

# Addison's Disease: Concept



# Metathesaurus Concepts (2005AB)

- ◆ Concept (~ 1.2 M) CUI
  - Set of synonymous concept names
- ◆ Term (~ 4.2 M) LUI
  - Set of normalized names
- ◆ String (~ 4.8 M) SUI
  - Distinct concept name
- ◆ Atom (~ 5.6 M) AUI
  - Concept name in a given source

A0000001 headache (source 1)  
A0000002 headache (source 2)  
**S0000001**

A0000003 Headache (source 1)  
A0000004 Headache (source 2)  
**S0000002**

**L0000001**

A0000005 Cephalgia (source 1)  
**S0000003**

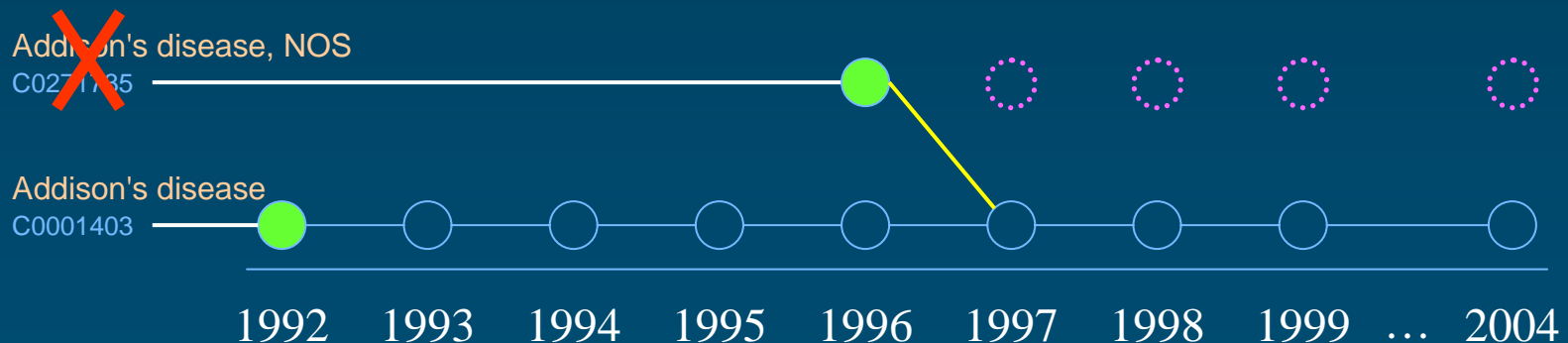
**L0000002**

**C0000001**



# Metathesaurus Evolution over time

- ◆ Concepts never die (in principle)
  - CUIs are permanent identifiers
- ◆ What happens when they do die (in reality)?
  - Concepts can merge or split
  - Resulting in new concepts and deletions



# Metathesaurus Relations

- ◆ Symbolic relations: ~9 M pairs of concepts
- ◆ Statistical relations : ~7 M pairs of concepts  
(co-occurring concepts)
- ◆ Mapping relations: 100,000 pairs of concepts

- 
- ◆ Categorization: Relationships between concepts and semantic types from the Semantic Network

# Symbolic relations

## ◆ Relation

- Pair of “atom” identifiers
- Type
- Attribute (if any)
- List of sources (for type and attribute)

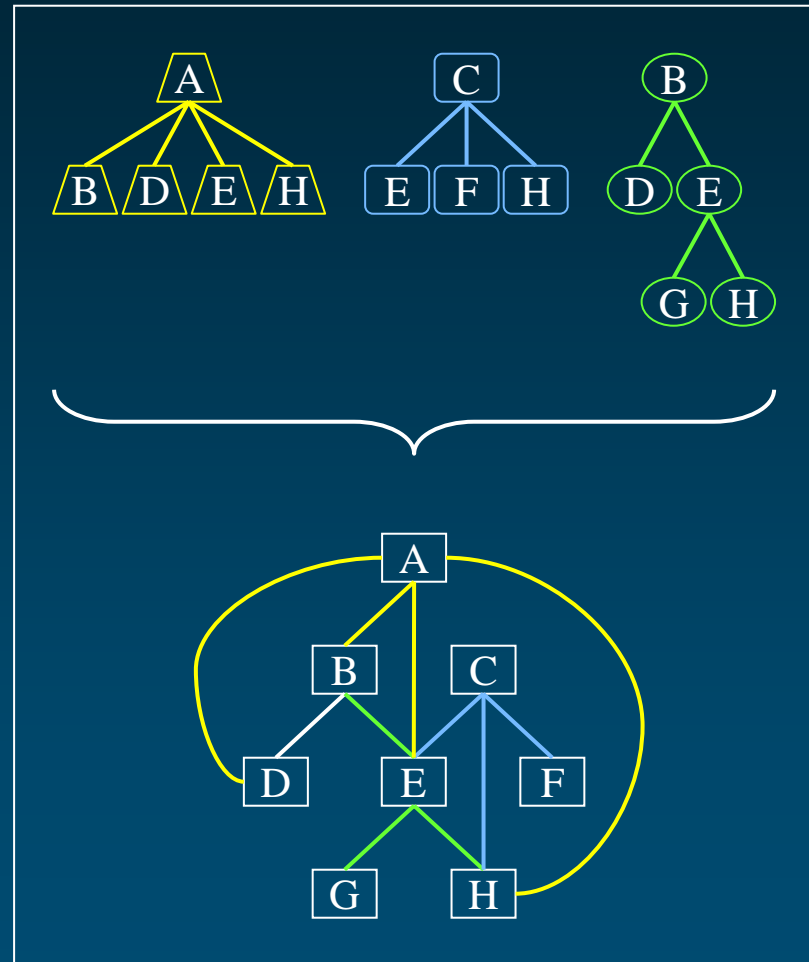
## ◆ Semantics of the relationship: defined by its *type* [and *attribute*]

Source transparency: the information  
is recorded at the “atom” level



# Organize concepts

- ◆ Inter-concept relationships: hierarchies from the source vocabularies
- ◆ Redundancy: multiple paths
- ◆ One graph instead of multiple trees (multiple inheritance)



# Semantic Network

## ◆ Semantic types (135)

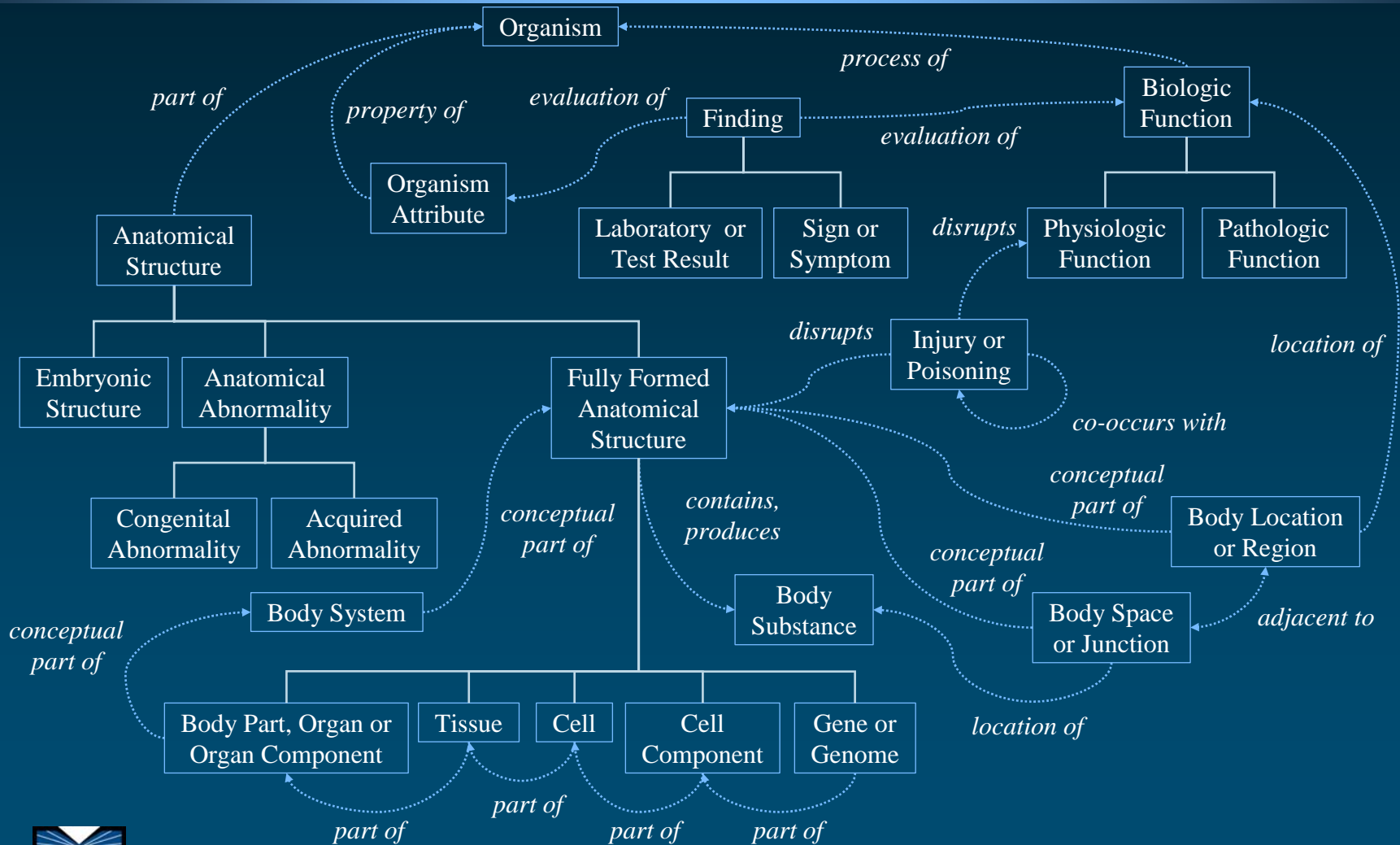
- tree structure
- 2 major hierarchies
  - Entity
    - Physical Object
    - Conceptual Entity
  - Event
    - Activity
    - Phenomenon or Process



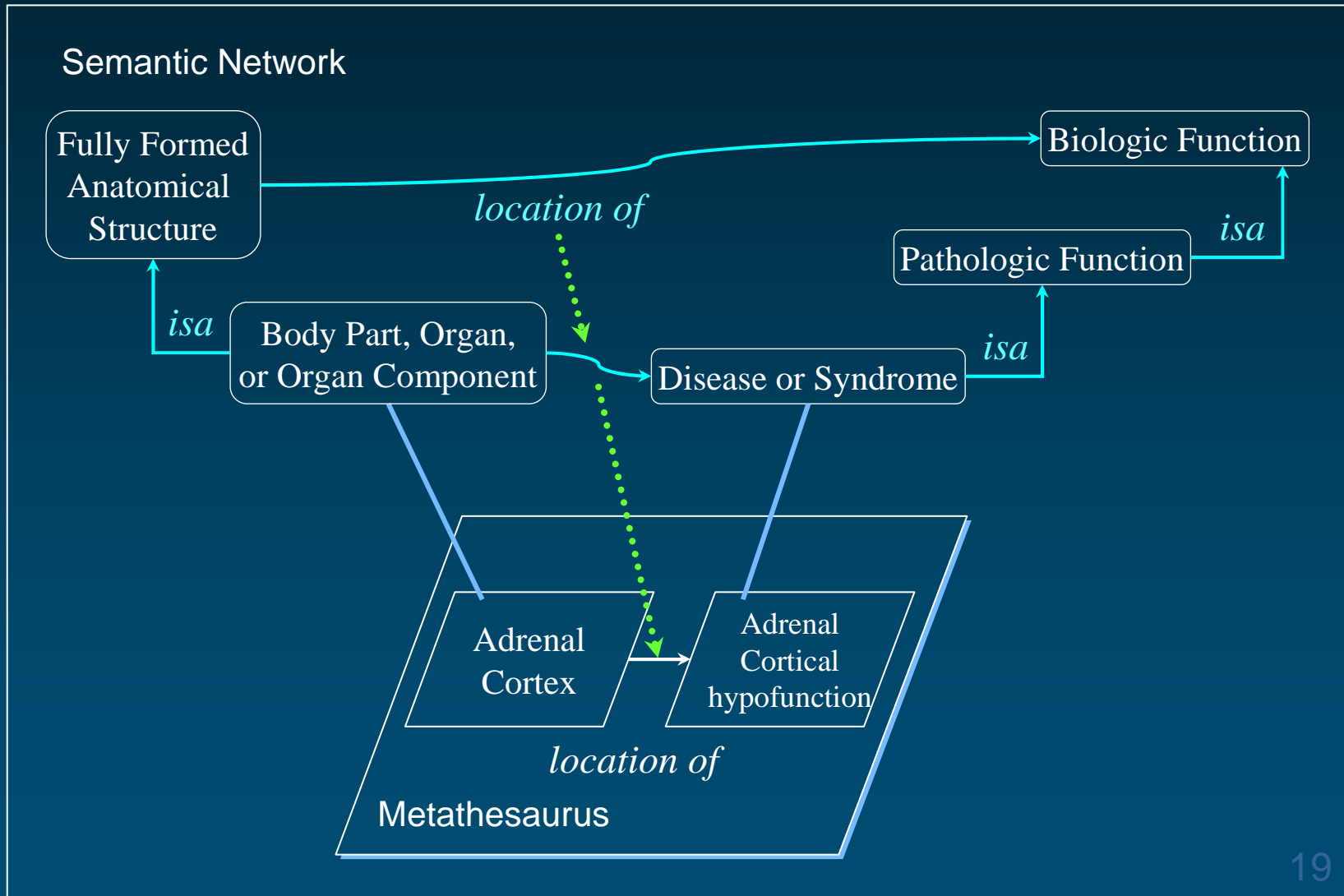
# Semantic Network

- ◆ Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - *Animal isa Organism*
      - *Enzyme isa Biologically Active Substance*
    - among relations
      - *treats isa affects*
  - non-hierarchical
    - *Sign or Symptom diagnoses Pathologic Function*
    - *Pharmacologic Substance treats Pathologic Function*

# Semantic Network relations



# Relationships can inherit semantics



Semantic Types

Anatomical Structure

Fully Formed Anatomical Structure

Embryonic Structure

Body Part, Organ or Organ Component

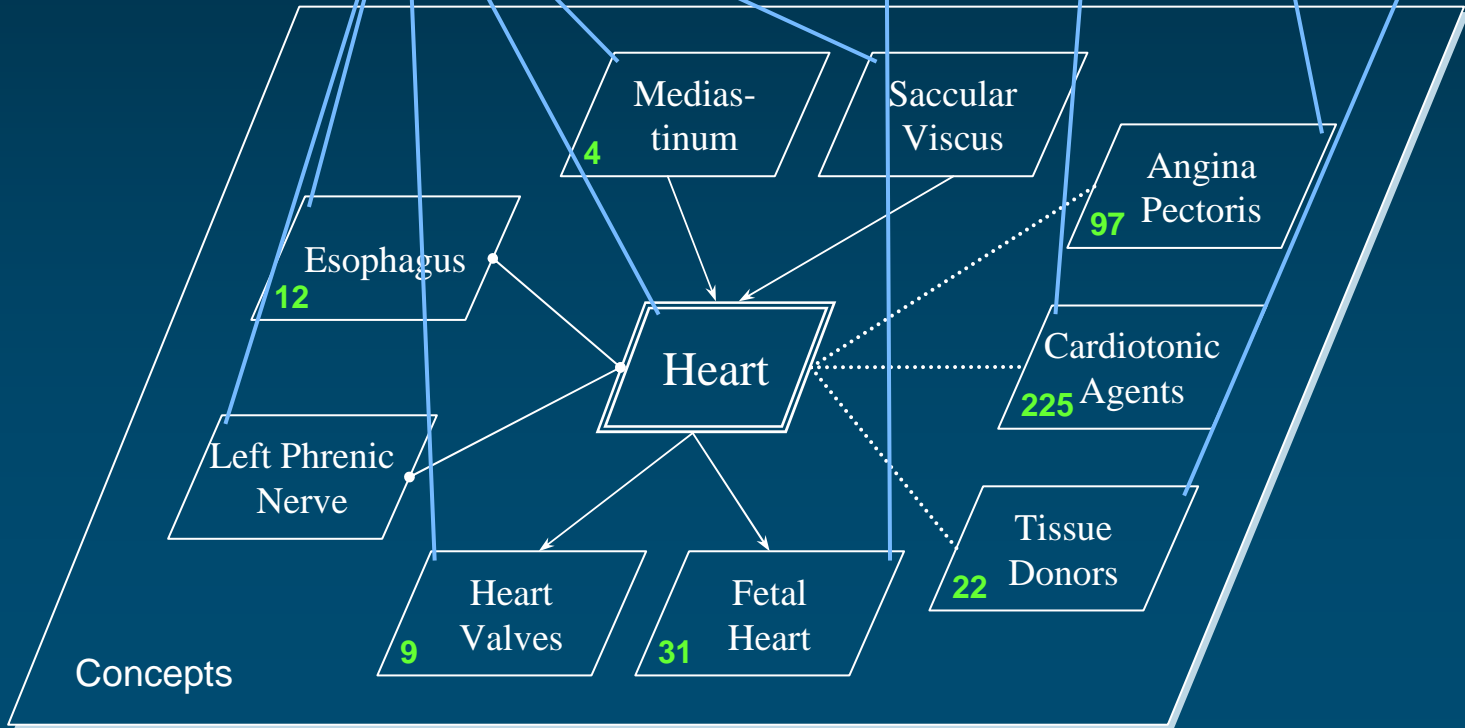
Disease or Syndrome

Pharmacologic Substance

Population Group

Semantic Network

Metathesaurus



Concepts

# Lexical tools

- ◆ To manage lexical variation in biomedical terminologies
- ◆ Major tools
  - Normalization
  - Indexes
  - Lexical Variant Generation program (lvg)
- ◆ Based on the SPECIALIST Lexicon
- ◆ Used by noun phrase extractors, search engines

# UMLS distribution

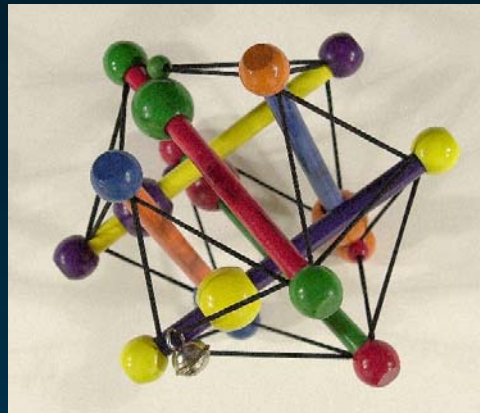
- ◆ License agreement
  - Part of the content is copyrighted
- ◆ 3-4 releases per year
  - Some components released more frequently
- ◆ Availability of UMLS data
  - Distributed on DVD
  - Downloaded from NLM website
  - APIs available for Java and XML-TCP/IP



# UMLS tools developed at NLM

- ◆ Several browsers
- ◆ MetamorphoSys
  - Install and customize the UMLS
  - Part of the UMLS distribution
- ◆ Lexical Variant Generation tools
  - Manage term variation
  - Part of the UMLS distribution and available separately
- ◆ MetaMap (MMTx)
  - Identify Metathesaurus concepts in text
  - Available separately (requires UMLS license)





# Medical Ontology Research

Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)

Web: [mor.nlm.nih.gov](http://mor.nlm.nih.gov)



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA



# References

## ◆ UMLS

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

## ◆ UMLS browsers

(free, but UMLS license required)

- Knowledge Source Server: [umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov)
- Semantic Navigator:  
<http://mor.nlm.nih.gov/perl/semnav.pl>
- RRF browser  
(standalone application distributed with the UMLS)



# References

## ◆ Recent overviews

- Bodenreider O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*; D267-D270.
- Nelson, S. J., Powell, T. & Humphreys, B. L. (2002 ). The Unified Medical Language System (UMLS) Project. In: Kent, Allen; Hall, Carolyn M., editors. *Encyclopedia of Library and Information Science*. New York: Marcel Dekker. p.369-378.

# References

## ◆ UMLS as a research project

- Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods Inf Med, 32*(4), 281-91.
- Humphreys, B. L., Lindberg, D. A., Schoolman, H. M., & Barnett, G. O. (1998). The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc, 5*(1), 1-11.

# References

## ◆ Technical papers

- McCray, A. T., & Nelson, S. J. (1995). The representation of meaning in the UMLS. *Methods Inf Med*, 34(1-2), 193-201.
- Bodenreider O. & McCray A. T. (2003). Exploring semantic groups through visual approaches. *Journal of Biomedical Informatics*, 36(6), 414-432.