



Public Health Informatics Fellowship Program
October 14, 2005

The Unified Medical Language System
Integrating Biomedical Terminology

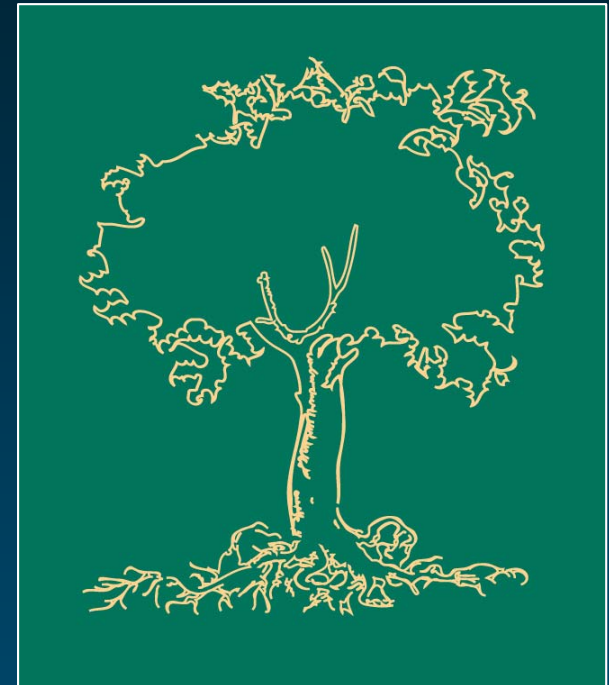


Olivier Bodenreider

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

What does UMLS stand for?

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



UMLS[®]
Unified Medical Language System[®]
UMLS Metathesaurus[®]



Motivation

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

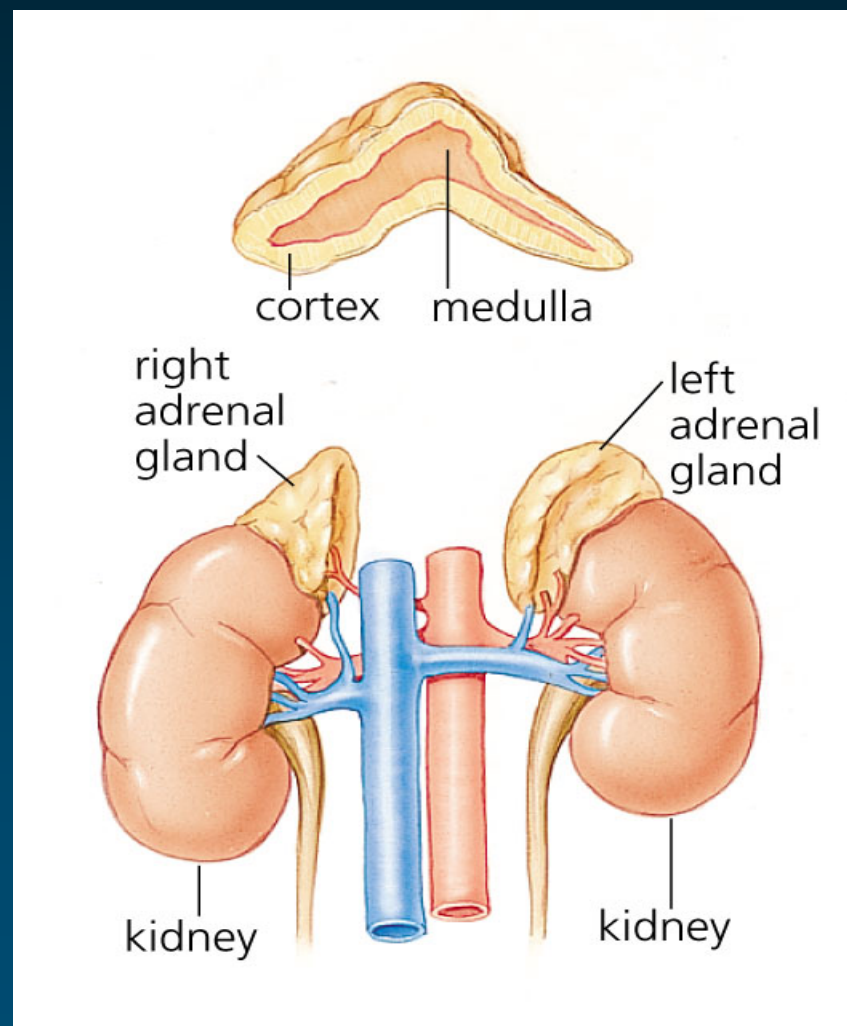
«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- 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.»

Overview through an example

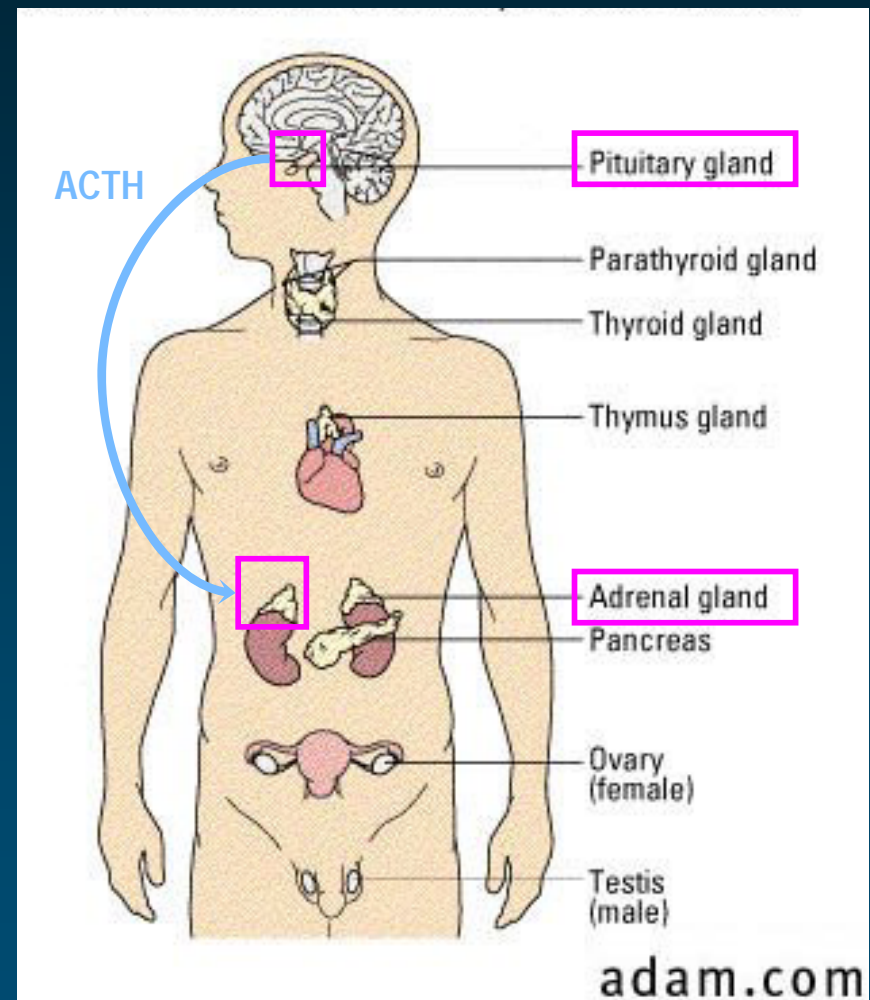
Addison's disease

- ◆ Addison's disease is a rare endocrine disorder
- ◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- ◆ For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism



Adrenal insufficiency Clinical variants

- ◆ Primary / Secondary
 - Primary: lesion of the adrenal glands themselves
 - Secondary: inadequate secretion of ACTH by the pituitary gland
- ◆ Acute / Chronic
- ◆ Isolated / Polyendocrine deficiency syndrome



Addison's disease: Symptoms

- ◆ Fatigue
- ◆ Weakness
- ◆ Low blood pressure
- ◆ Pigmentation of the skin (exposed and non-exposed parts of the body)
- ◆ ...

AD in medical vocabularies

◆ Synonyms: different terms

- Addisonian syndrome
 - Bronzed disease
 - Addison melanoderma
 - Asthenia pigmentosa
 - Primary adrenal deficiency
 - Primary adrenal insufficiency
 - Primary adrenocortical insufficiency
 - Chronic adrenocortical insufficiency
- } eponym
- } symptoms
- } clinical variants

◆ Contexts: different hierarchies



Organize terms

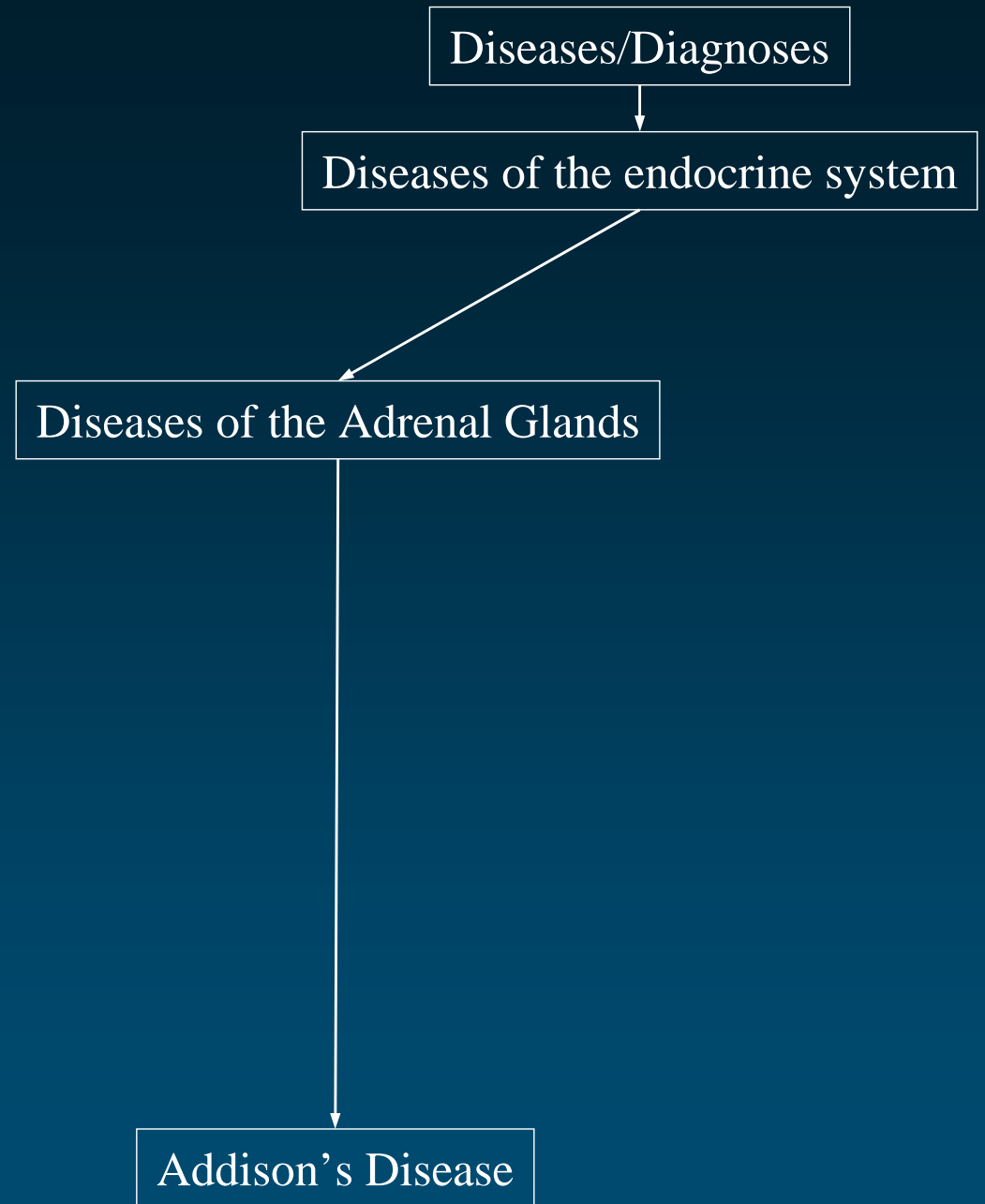
- ◆ Synonymous terms clustered into a concept
- ◆ Preferred term
- ◆ Unique identifier (CUI)

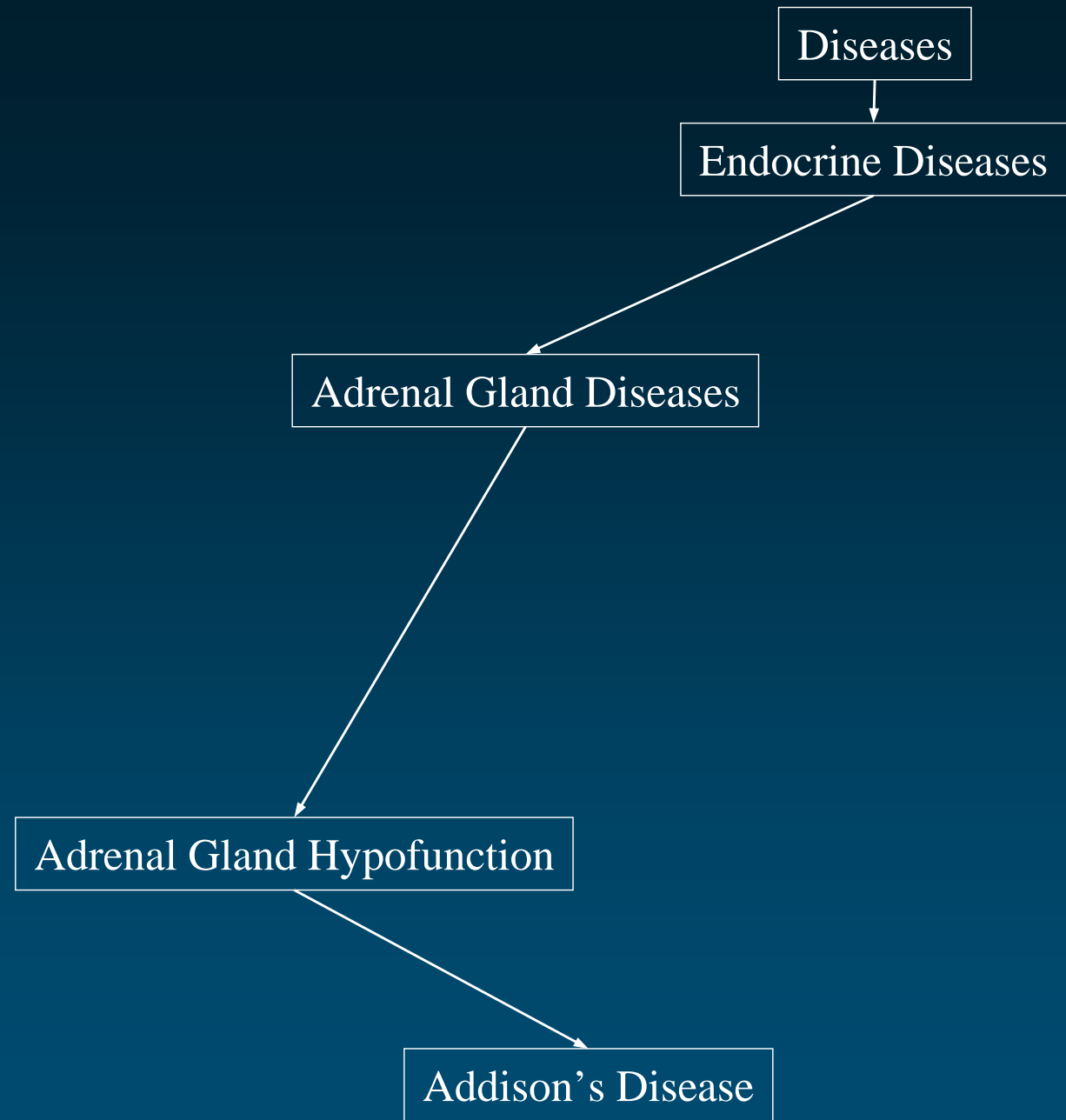
| | | |
|--------------------------------------|-----------|-----------|
| Addison Disease | MeSH | D000224 |
| Primary hypoadrenalism | MedDRA | 10036696 |
| Primary adrenocortical insufficiency | ICD-10 | E27.1 |
| Addison's disease (disorder) | SNOMED CT | 363732003 |

C0001403

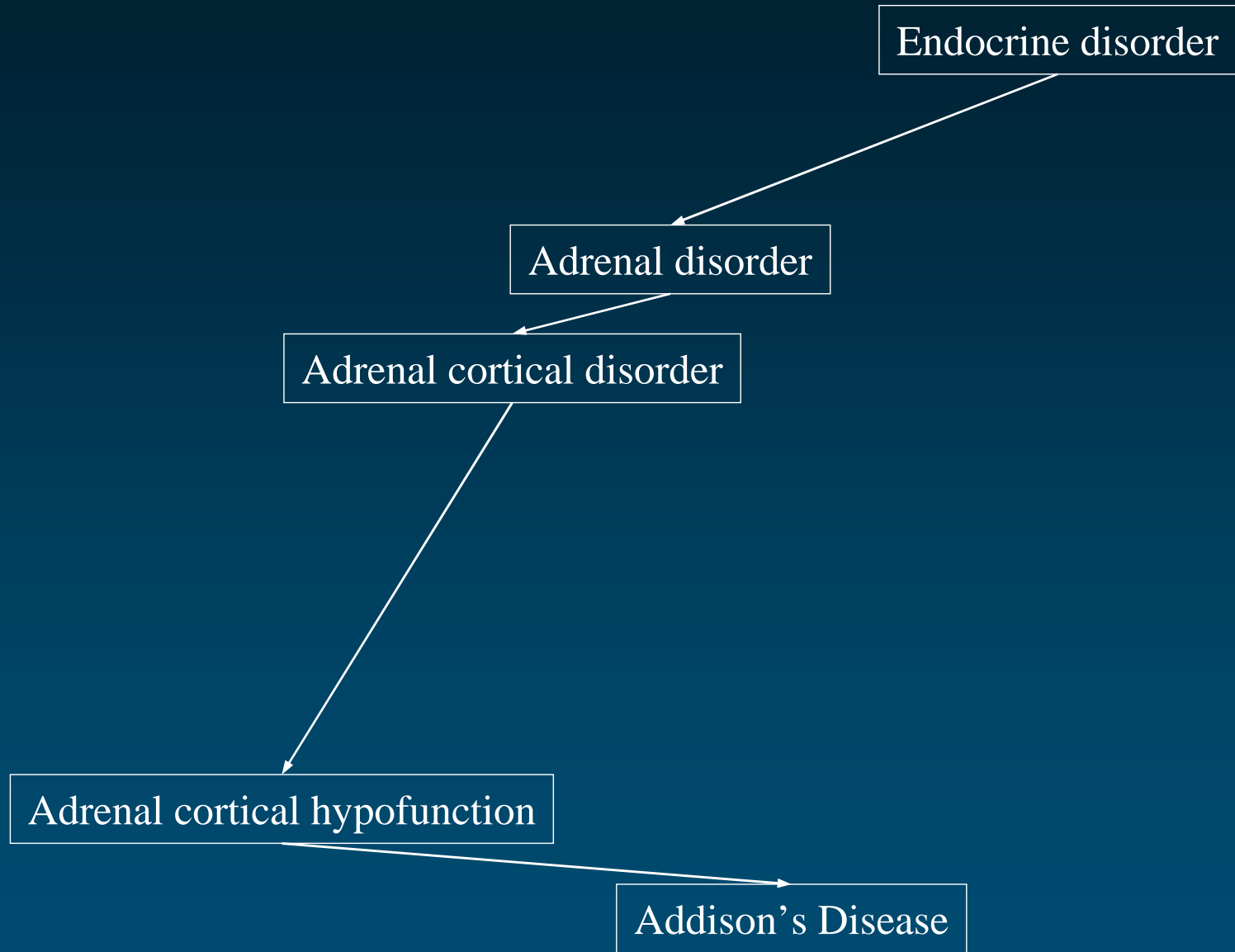
Addison's disease

SNOMED International

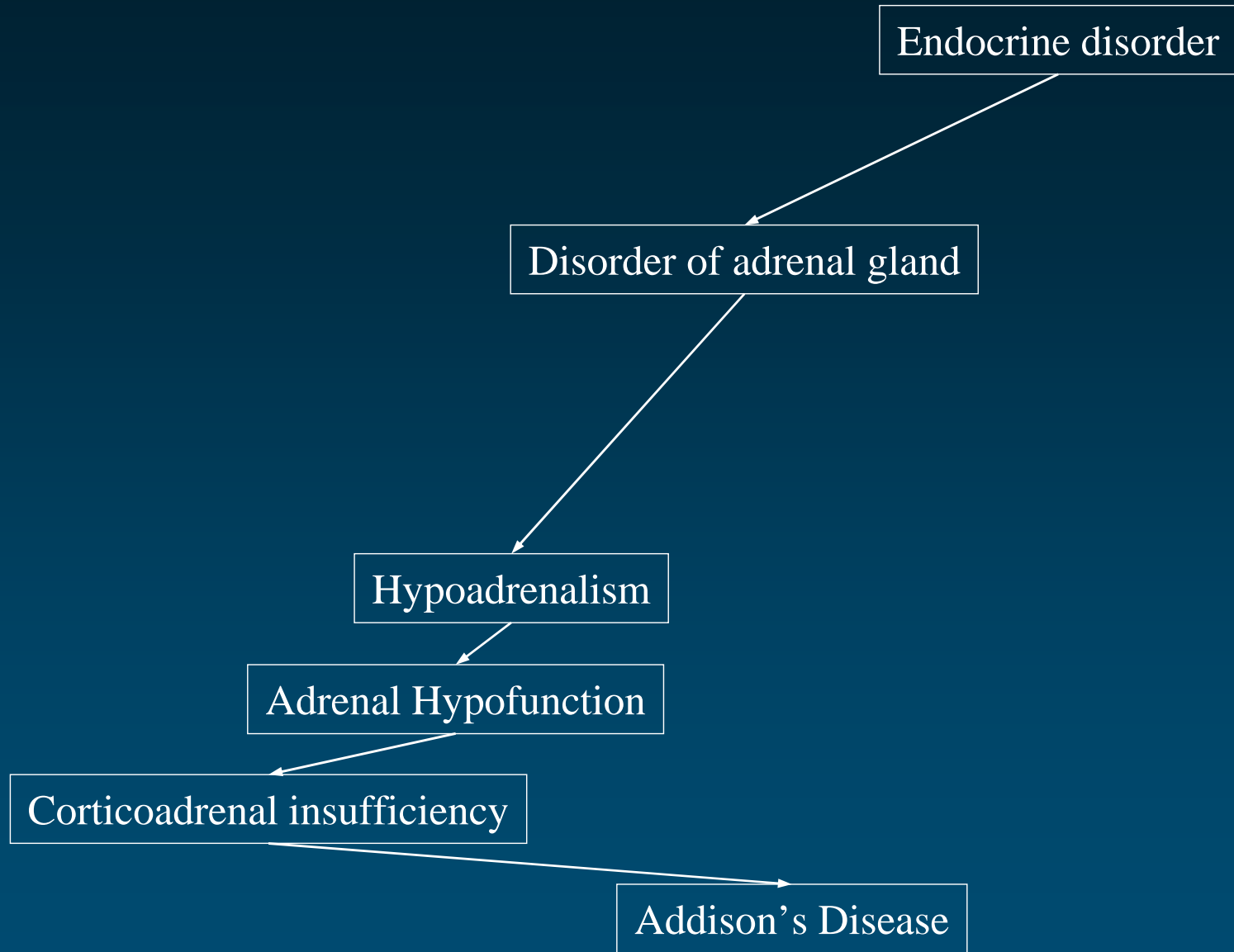




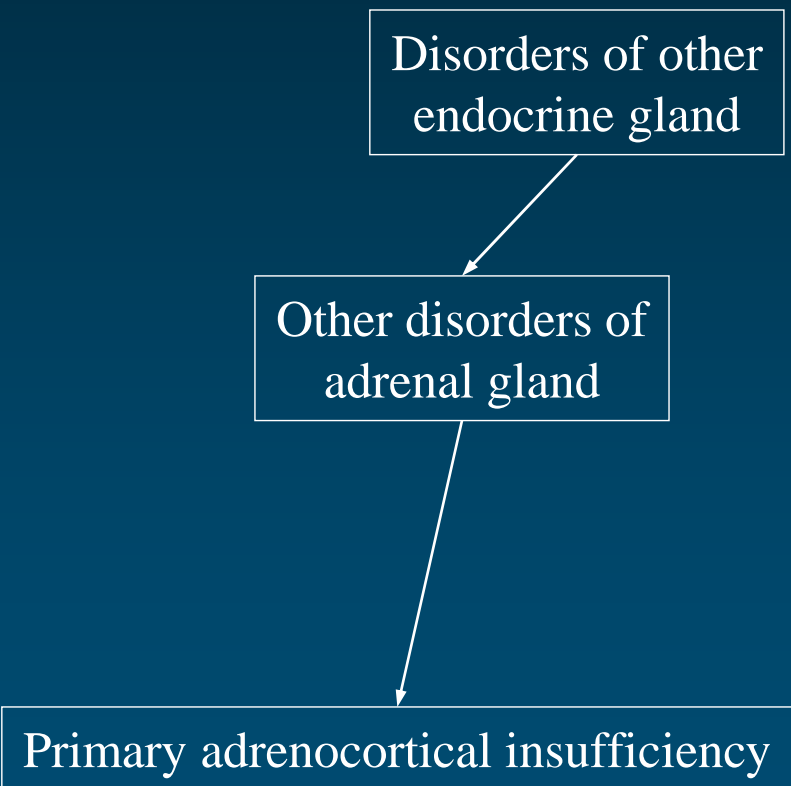
AOD



Read Codes

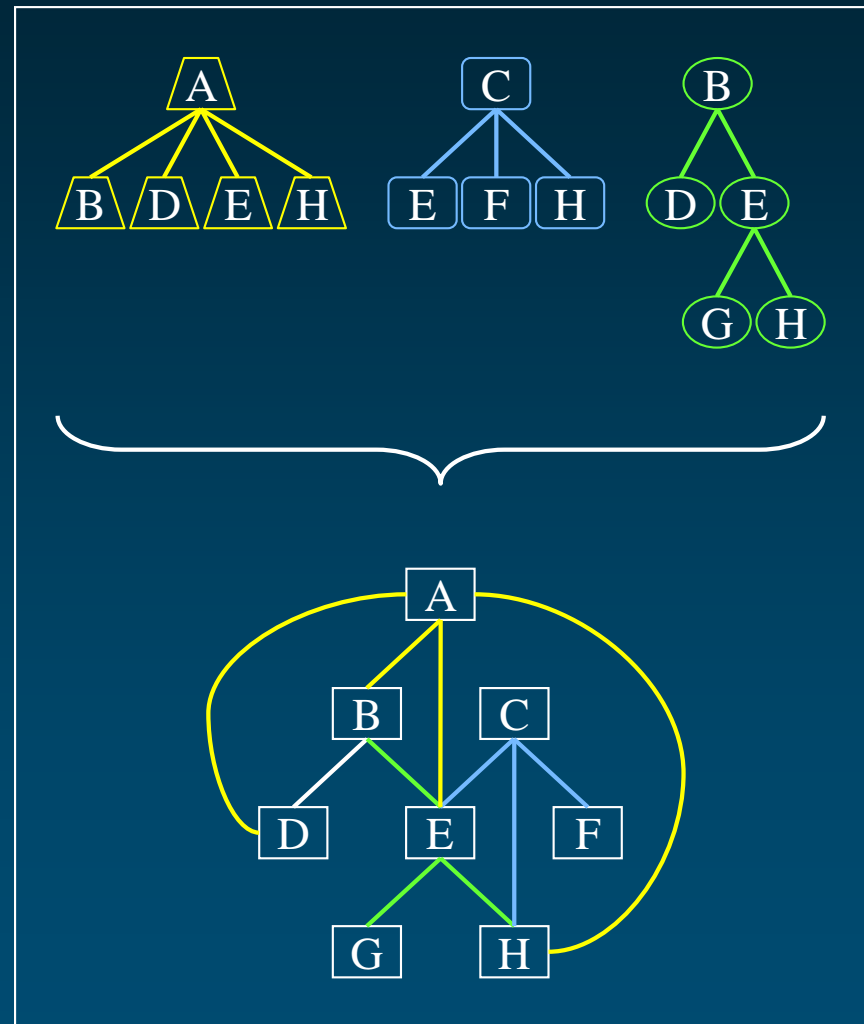


ICD-10



Organize concepts

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



organize concepts

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Addison's Disease

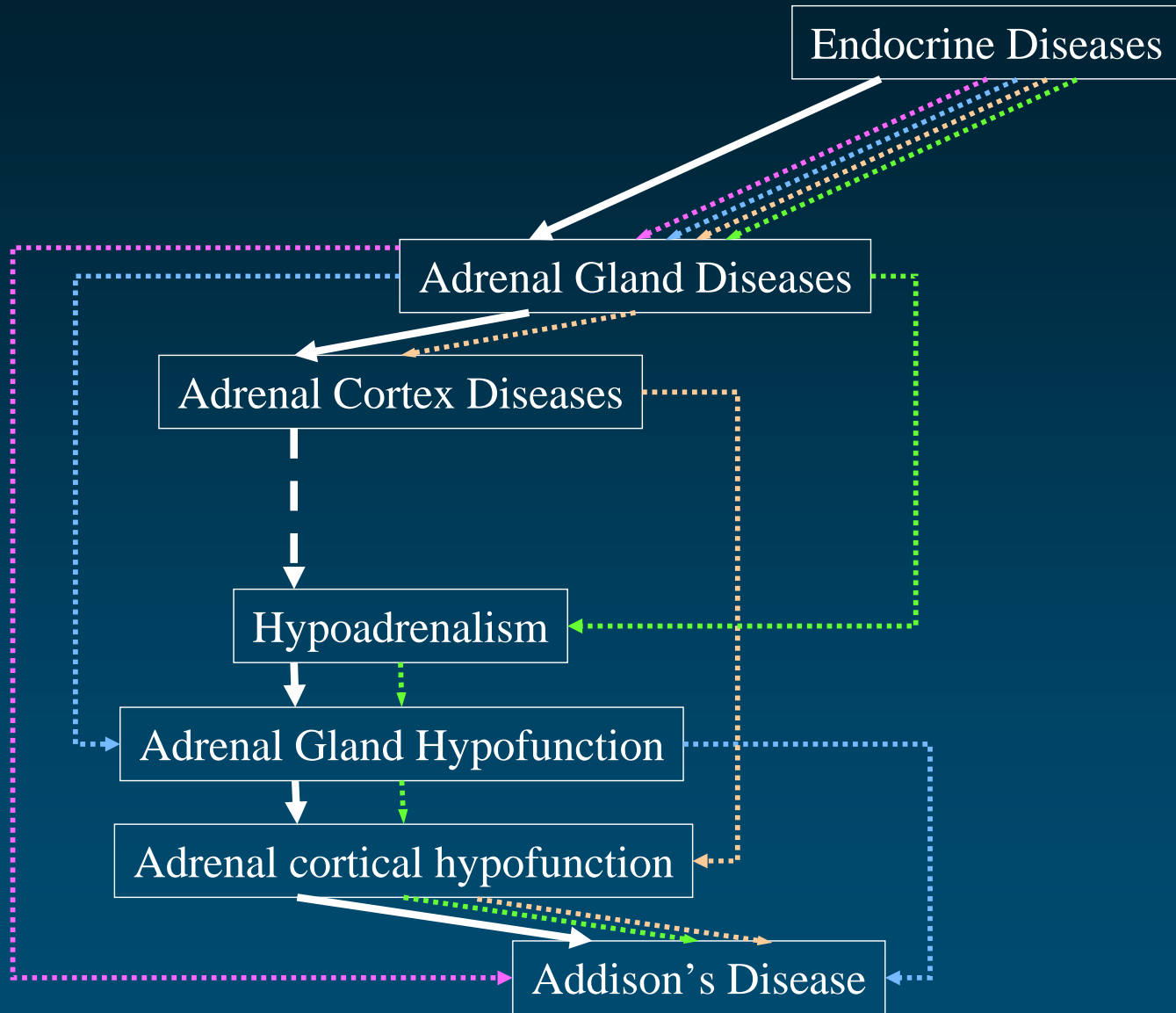
SNOMED

MeSH

AOD

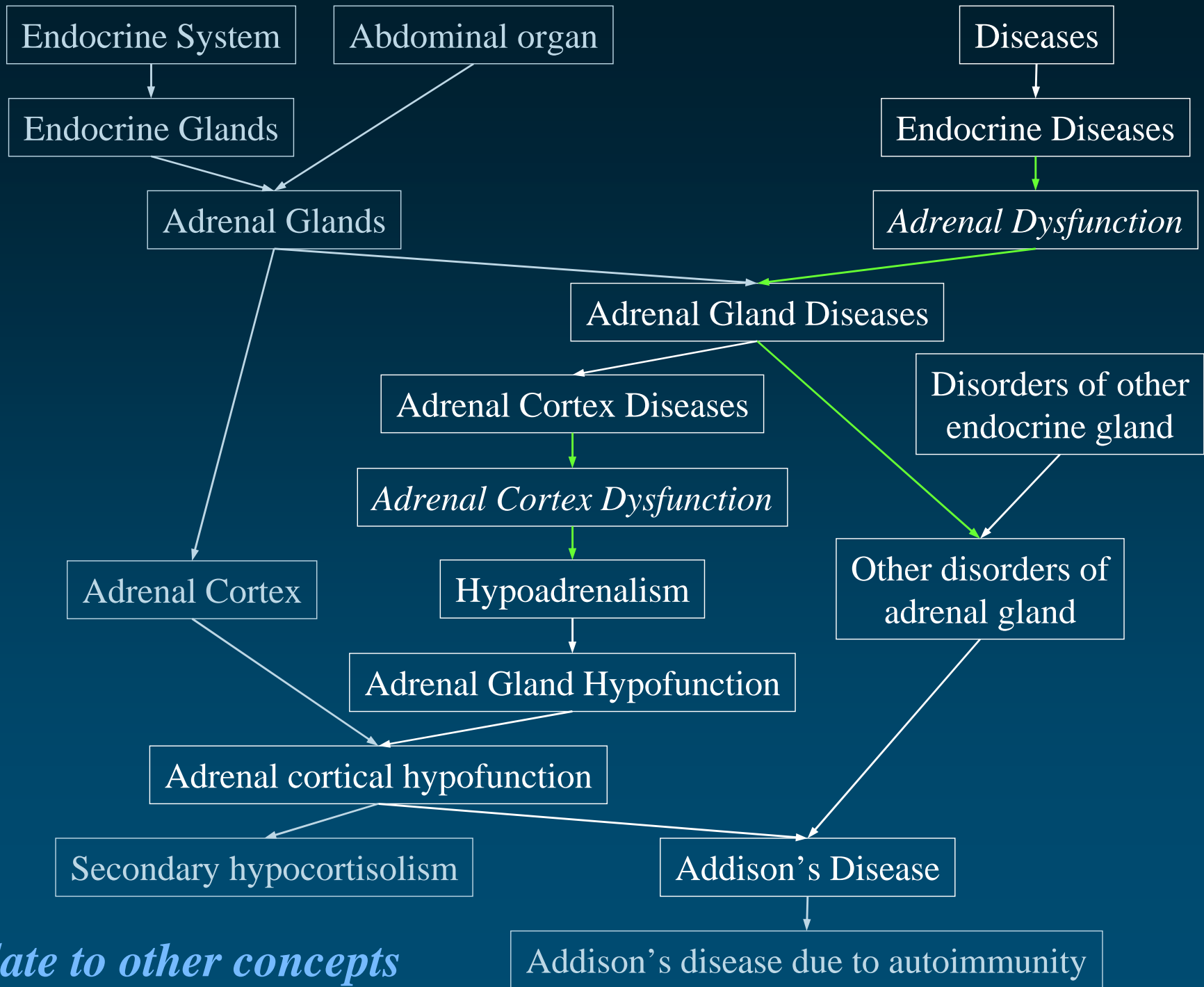
Read Codes

UMLS



Relate to other concepts

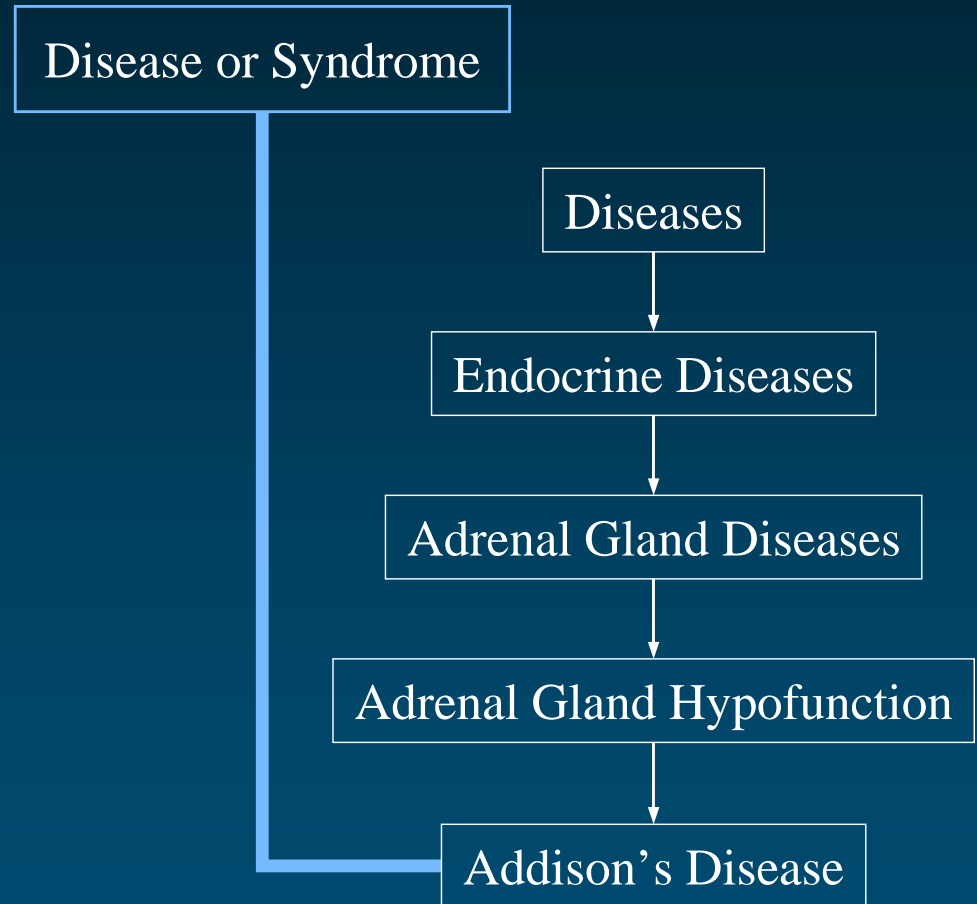
- ◆ Additional hierarchical relations
 - link to other trees
 - make relationships explicit
- ◆ Non-hierarchical relations
- ◆ Co-occurring concepts
- ◆ Mapping relations



relate to other concepts

Categorize concepts

- ◆ High-level categories (semantic types)
- ◆ Assigned by the Metathesaurus editors
- ◆ Independently of the hierarchies in which these concepts are located



UMLS Knowledge Sources

UMLS 3 components

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

UMLS Metathesaurus

Metathesaurus Basic organization

◆ Concepts

- Synonymous terms are clustered into a concept
- Properties are attached to concepts, e.g.,
 - Unique identifier
 - Definition

◆ Relations

- Concepts are related to other concepts
- Properties are attached to relations, e.g.,
 - Type of relationship
 - Source

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 (file format + Unicode)



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)

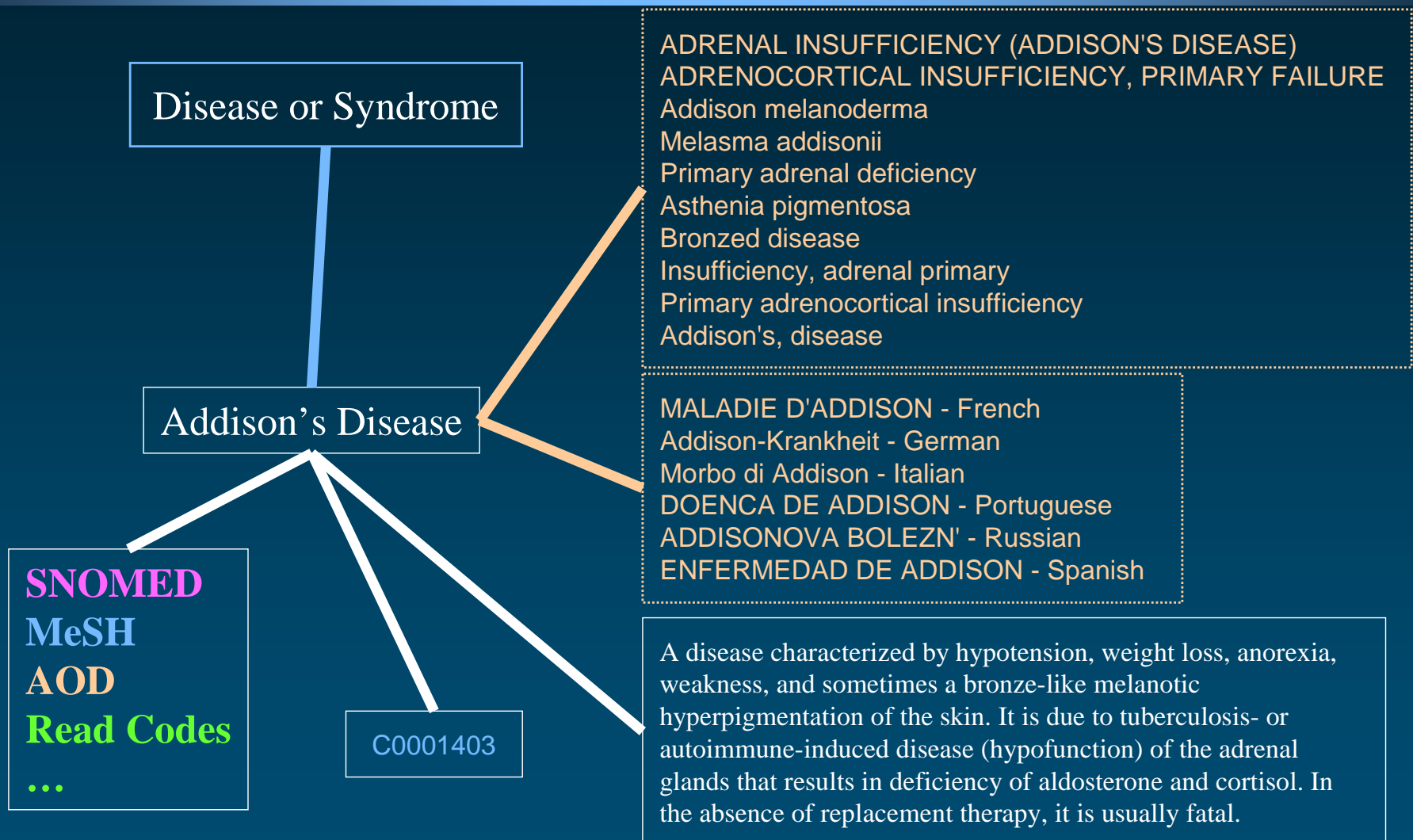
Biomedical terminologies (cont'd)

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



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

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



Cluster of synonymous terms

Concept
C0001403

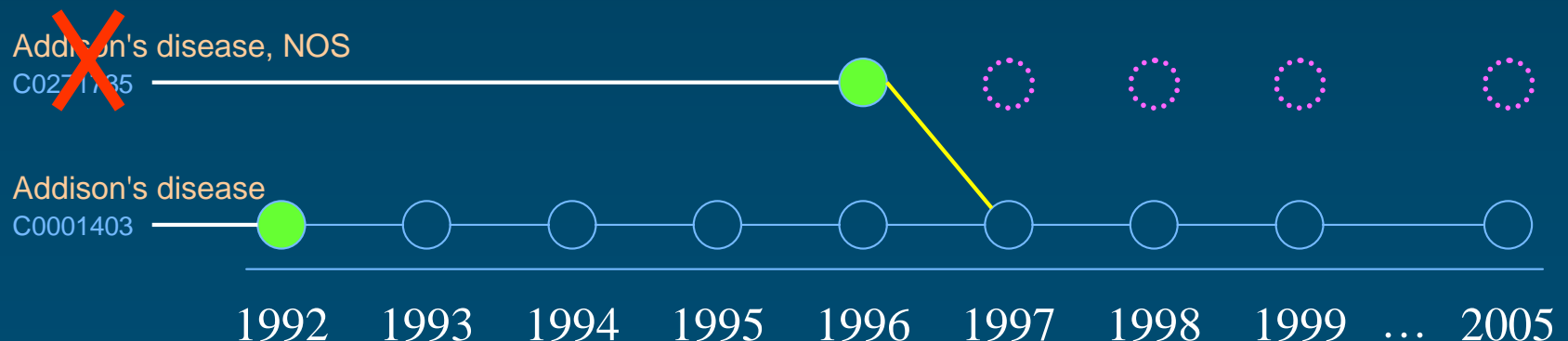
| | | |
|------------------|--|-------|
| Term L0001403 | <p>S0354372 <i>Addison's disease</i></p> <p>S0010792 Addison Disease</p> <p>S0010794 Addison's Disease</p> <p>S0010796 Addisons Disease</p> <p>S0033587 Disease, Addison</p> <p>S0352253 ADDISON'S DISEASE</p> | [...] |
| Term L2799243 | S3341310 <i>Addison's disease (disorder)</i> | |
| Term L0494940 | <p>S5907336 <i>Primary Adrenocortical Insufficiency</i></p> <p>S5901878 Insufficiencies, Primary Adrenocortical</p> | [...] |
| Term L0494851 | <p>S5907334 <i>Primary Adrenal Insufficiency</i></p> <p>S4094828 adrenal; insufficiency, primary</p> | [...] |
| Term L0585243 | <p>S5907343 <i>Primary Hypoadrenalism</i></p> <p>S5901432 Hypoadrenalism, Primary</p> | [...] |
| Term L1229627 | S1471573 <i>Addison-Krankheit</i> | GER |
| Term L5345155 | S6107160 <i>Maladie d'Addison</i> | FRE |

[...]



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 Relationships

- ◆ 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

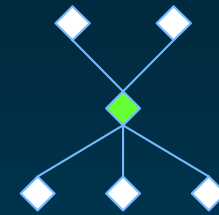
Symbolic relationships Type

◆ Hierarchical

- Parent / Child
- Broader / Narrower than

PAR/CHD

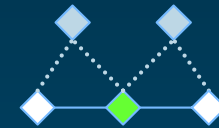
RB/RN



◆ Derived from hierarchies

- Siblings (children of parents)

SIB



◆ Associative

- Other

RO



◆ Various flavors of near-synonymy

- Similar
- Source asserted synonymy
- Possible synonymy

RL

SY

RQ



Symbolic relationships Attribute

- ◆ Hierarchical
 - isa (is-a-kind-of)
 - part-of
- ◆ Associative
 - location-of
 - caused-by
 - treats
 - ...
- ◆ Cross-references (mapping)

Semantic Types

Anatomical Structure

Fully Formed Anatomical Structure

Embryonic Structure

Disease or Syndrome

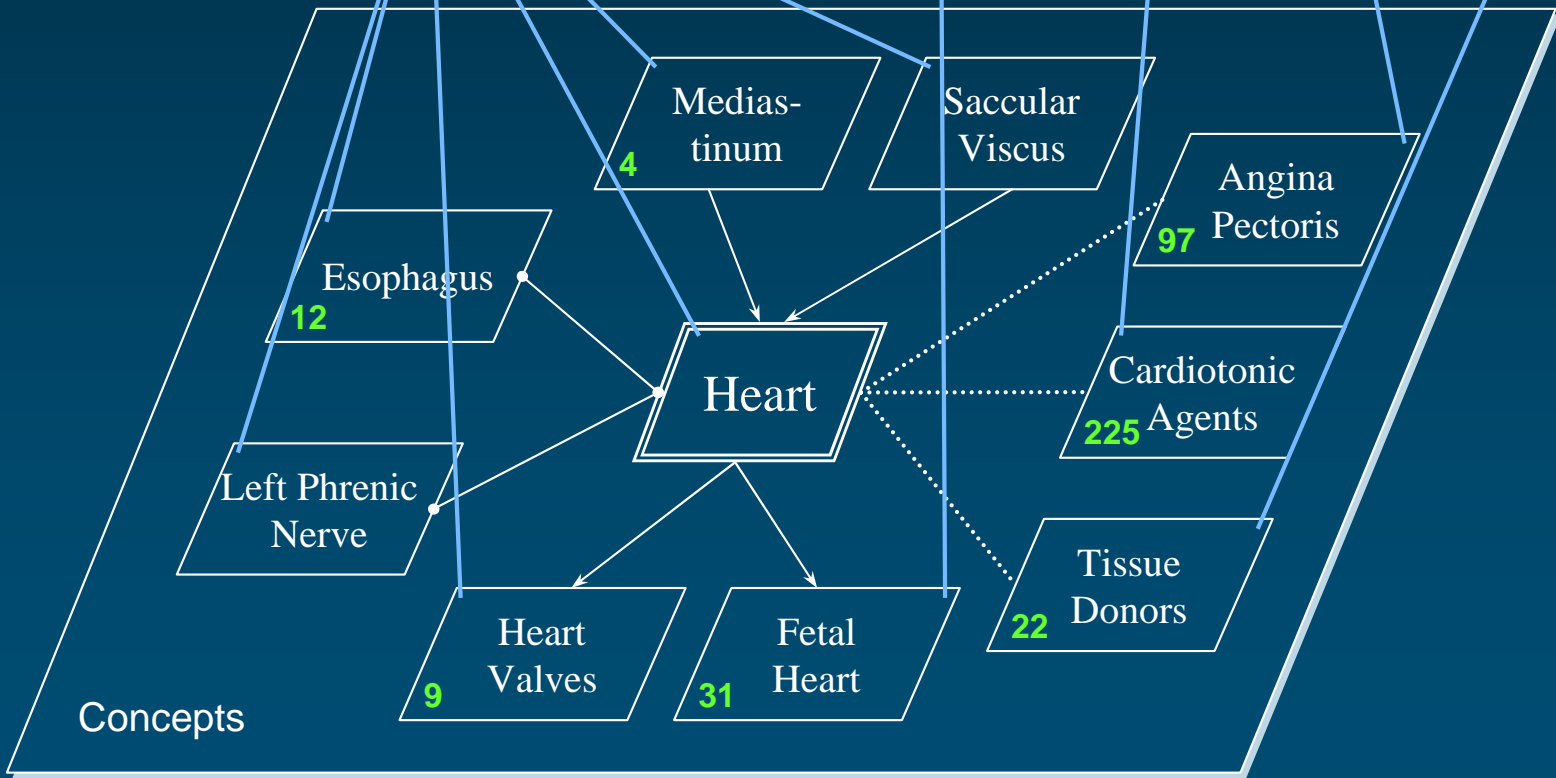
Body Part, Organ or Organ Component

Pharmacologic Substance

Population Group

Semantic Network

Metathesaurus



Concepts

UMLS Semantic Network

Semantic Network

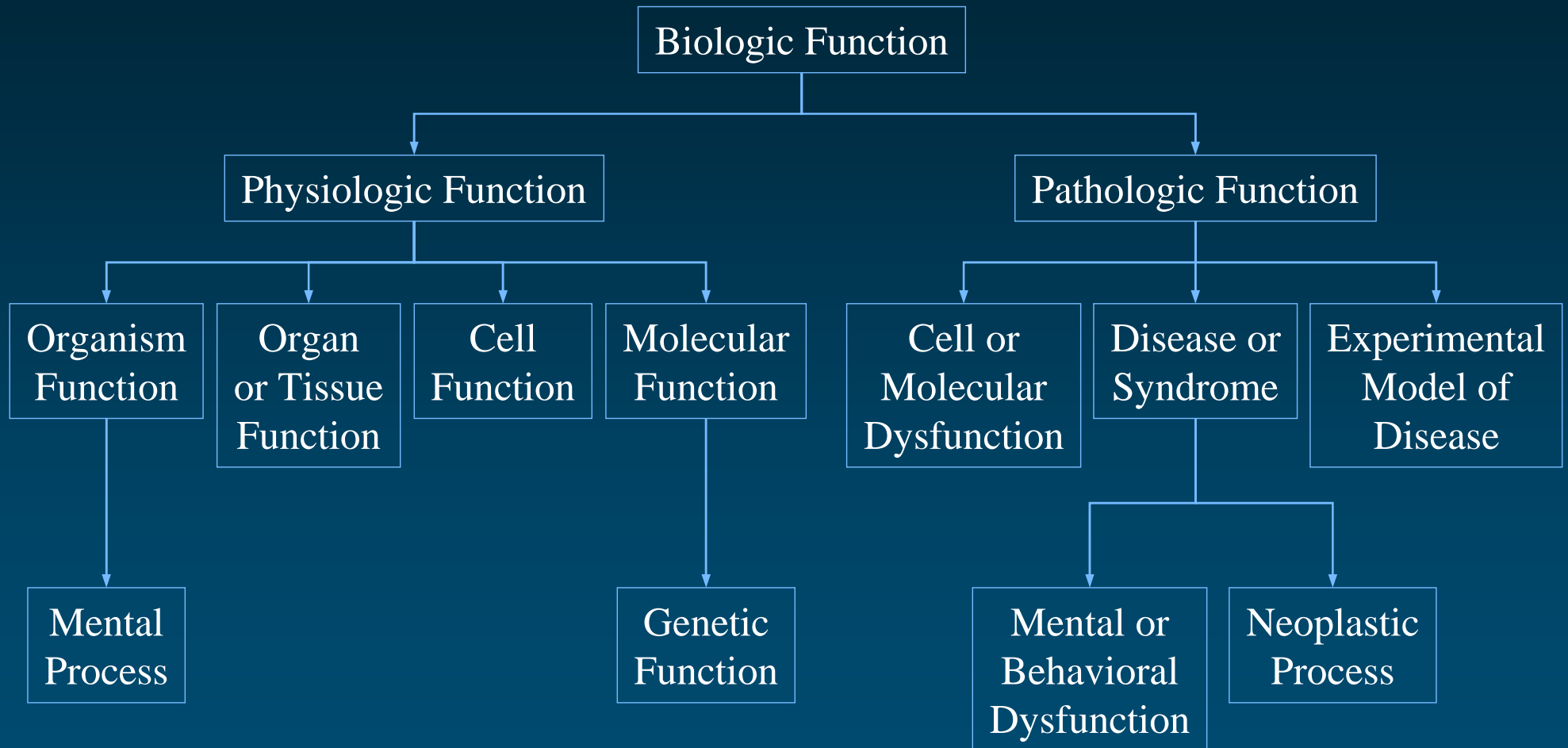
◆ Semantic types (135)

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

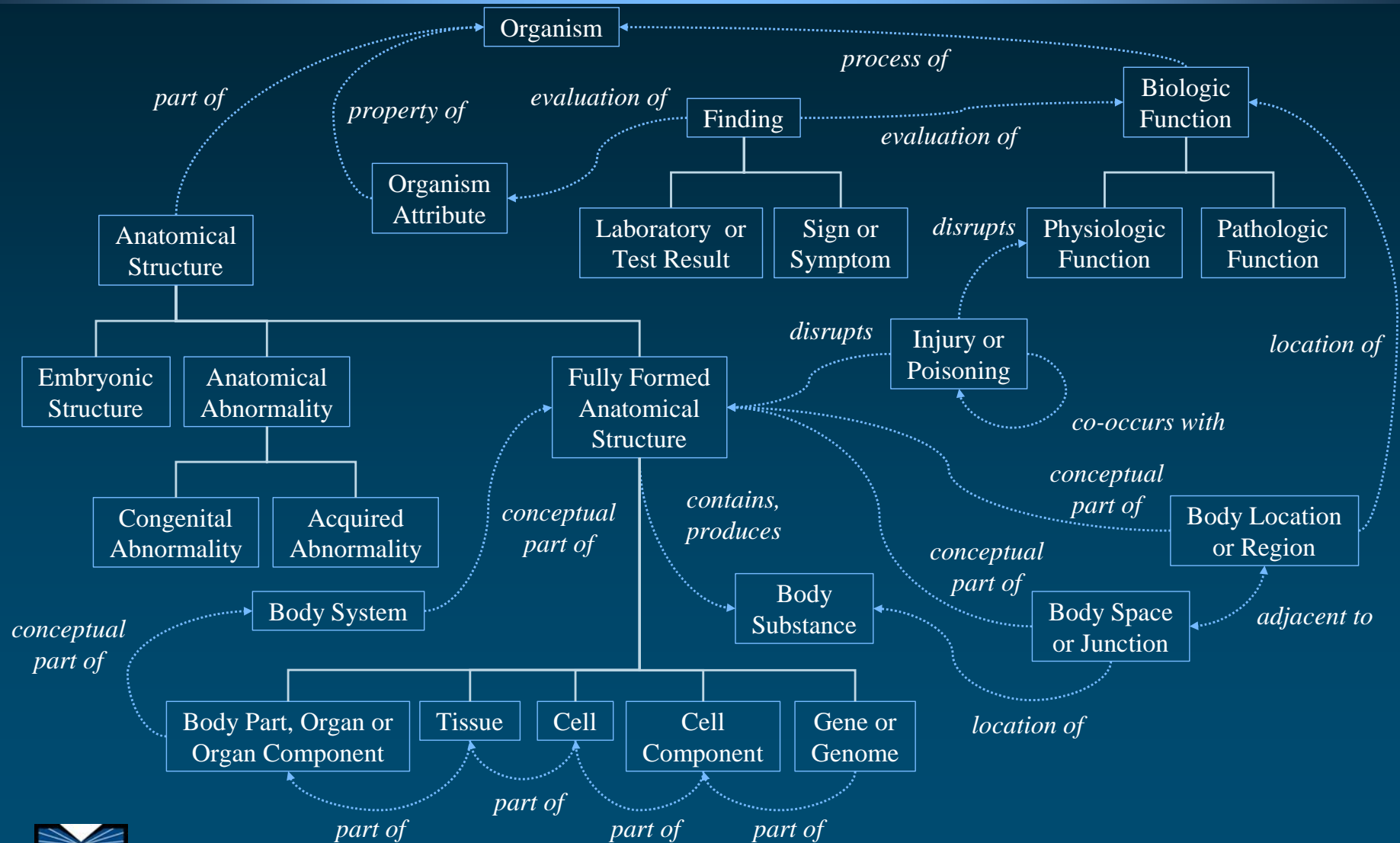
Semantic Network

- ◆ 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*

“Biologic Function” hierarchy (isa)



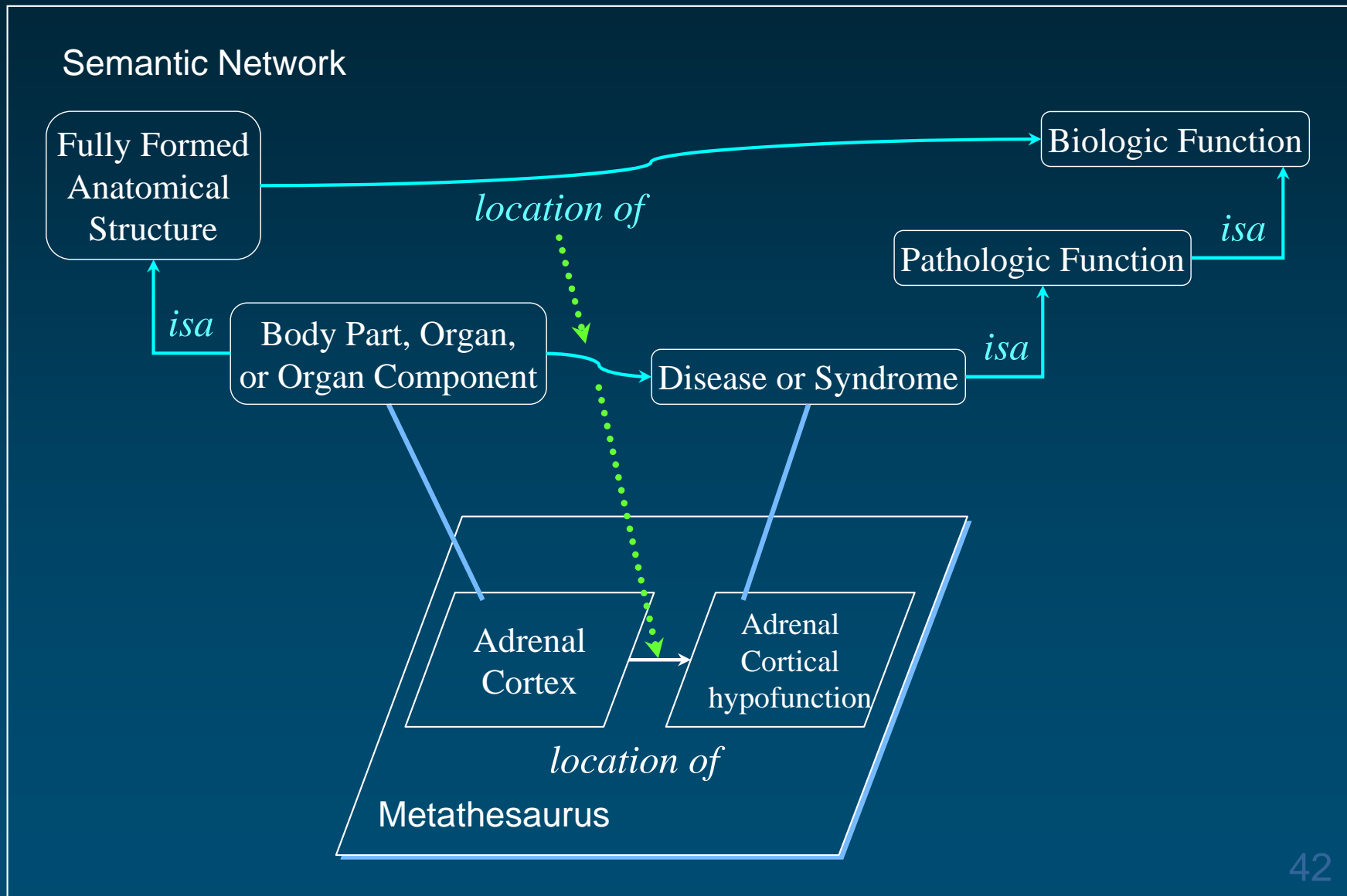
Associative (non-isa) relationships



Why a semantic network?

- ◆ Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently of their position in a hierarchy*
- ◆ A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs
 - The relationship may or may not hold at the concept level
 - Other relationships may apply at the concept level

Relationships can inherit semantics

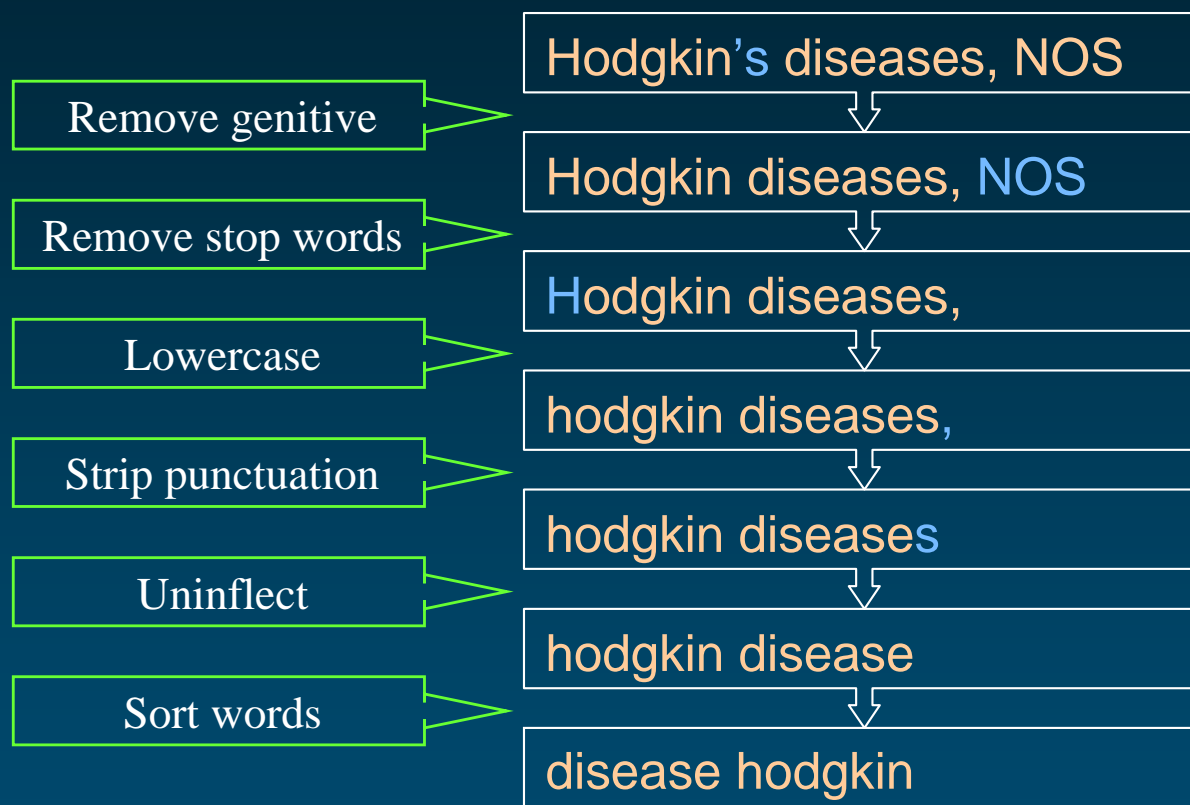


SPECIALIST Lexicon and lexical tools

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

Normalization



Normalization: Example

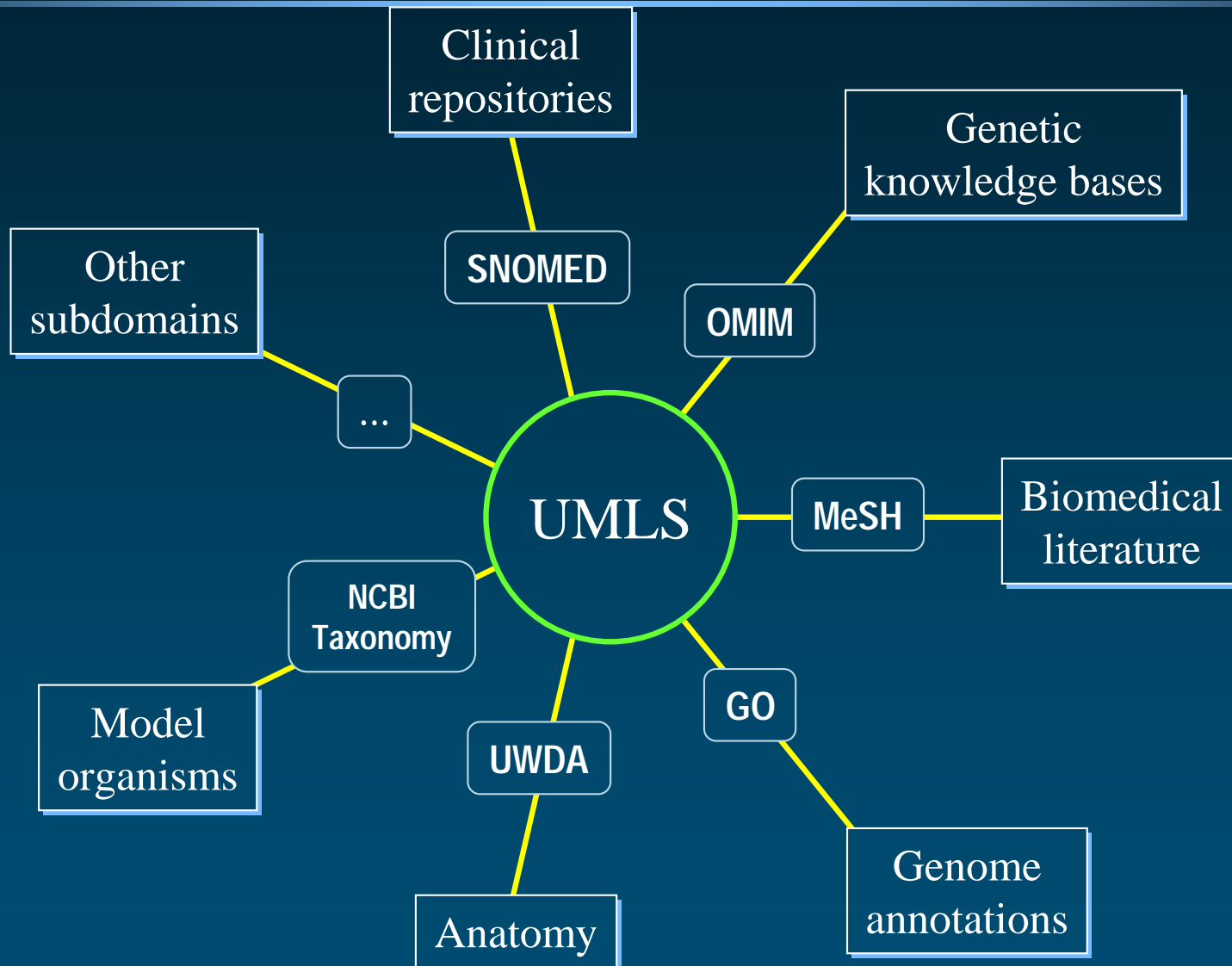
Hodgkin Disease
HODGKINS DISEASE
Hodgkin's Disease
Disease, Hodgkin's
Hodgkin's, disease
HODGKIN'S DISEASE
Hodgkin's disease
Hodgkins Disease
Hodgkin's disease NOS
Hodgkin's disease, NOS
Disease, Hodgkins
Diseases, Hodgkins
Hodgkins Diseases
Hodgkins disease
hodgkin's disease
Disease, Hodgkin

normalize

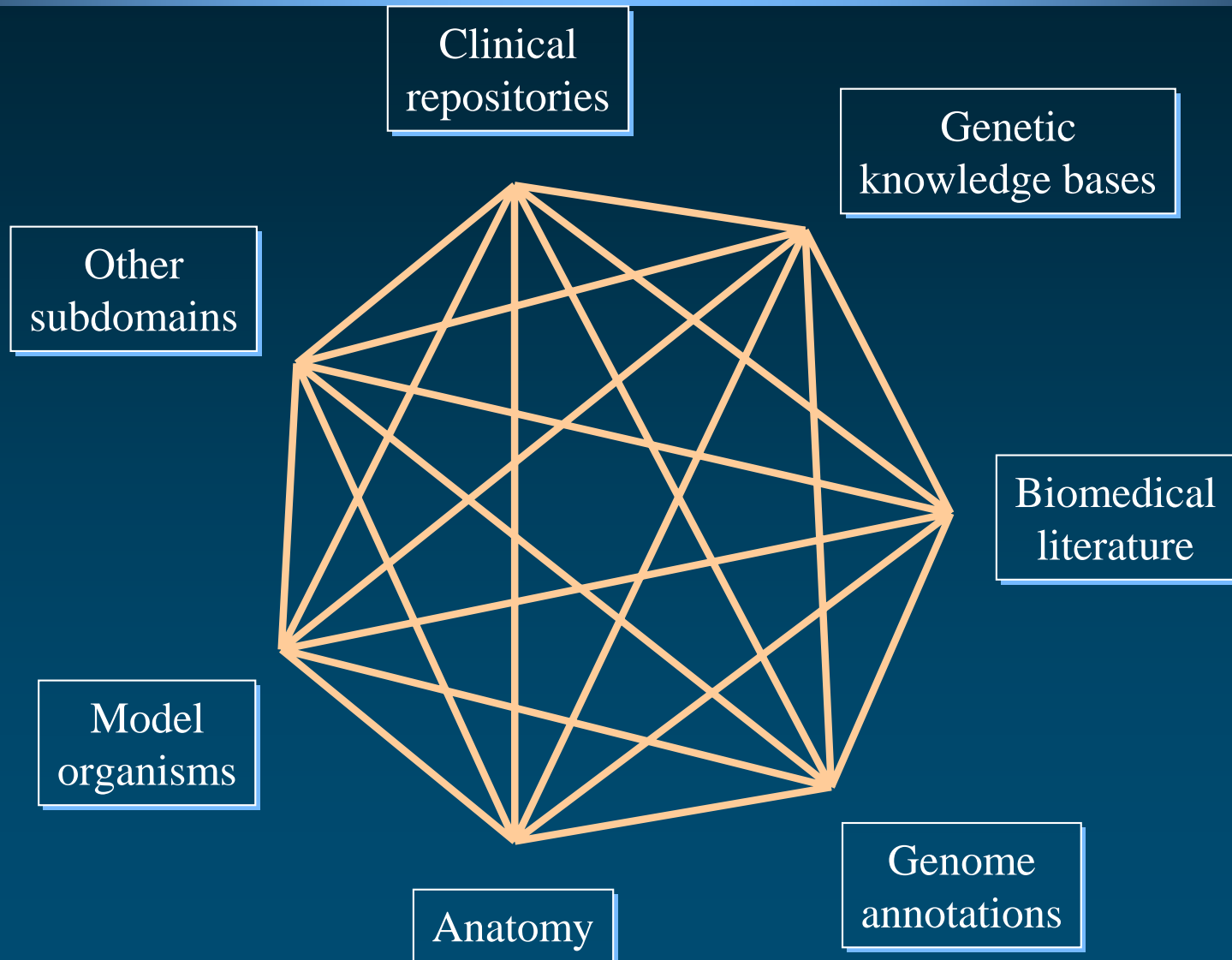
disease hodgkin

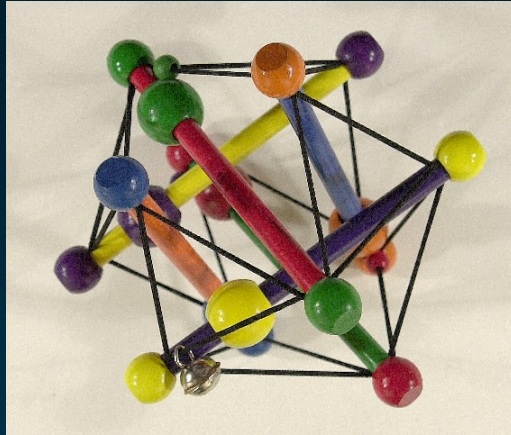
Conclusions

Integrating subdomains



Integrating subdomains





Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

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

References

- ◆ UMLS

umlsinfo.nlm.nih.gov

- ◆ UMLS browsers

(free, but UMLS license required)

- Knowledge Source Server: 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.