

Information Technologies for Healthcare
Barriers to Implementation
NIST - August 1, 2002

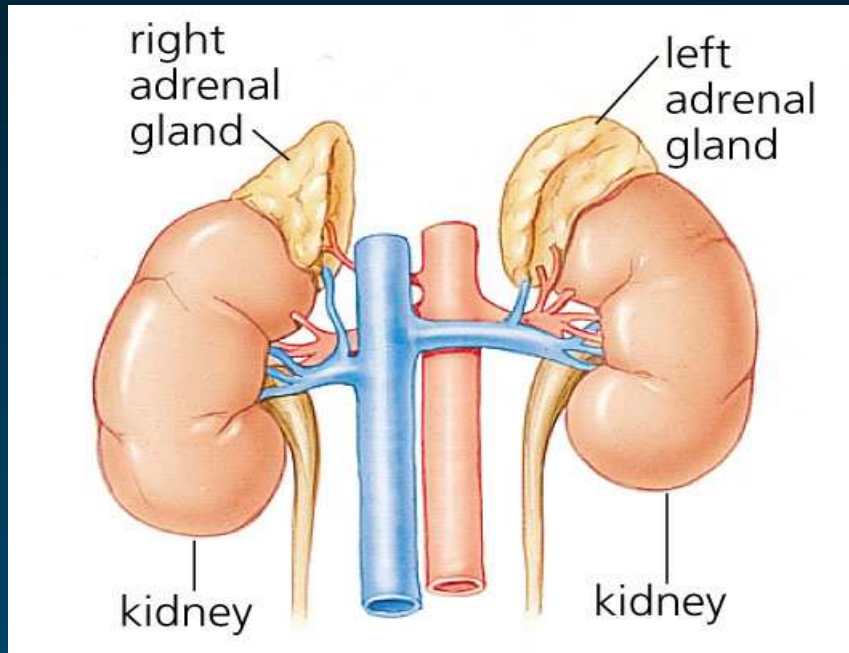
From Terminology Integration
to Interoperability



Olivier Bodenreider

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

The Problem Adrenal gland diseases



Adrenal Gland Diseases

C0001621

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

Outline

- ◆ Case study:
Unified Medical Language System (UMLS)
- ◆ History
- ◆ Overview
- ◆ Benefits
- ◆ Limitations



History

Motivation

[Lindberg & al., *Methods*, 1993]

[Humphreys & al., *JAMIA*, 1998]

- ◆ Started in 1986
- ◆ National Library of Medicine
- ◆ “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.»



UMLS research team

[Humphreys & al., *JAMIA*, 1998]

- ◆ Bill Hole
- ◆ L. Kingsland III
- ◆ Dan Masys
- ◆ Alexa McCray
- ◆ Stuart Nelson
- ◆ Roy Rada
- ◆ Rick Rodgers
- ◆ Peri Schuyler
- ◆ Brigham & Women's H.
- ◆ Carnegie-Mellon Univ.
- ◆ Columbia Univ.
- ◆ Lexical Technology, Inc.
- ◆ Massachusetts General H.
- ◆ UCSF
- ◆ Univ. of Pittsburgh
- ◆ Univ. of Utah
- ◆ [...]



UMLS chronology

- ◆ Definition of 3 knowledge sources (1986-88)
 - Metathesaurus
 - Semantic Network
 - Information Sources Map
- ◆ Building, distributing, and testing (1989-91)
 - Integration vs. *ad hoc* development
 - First release in 1990
- ◆ Development of applications (1992-94)



Overview

Biomedical terminologies

◆ Core vocabularies

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

◆ Several perspectives

- clinical terms (SNOMED, CTV3)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- standards (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)
- ◆ Knowledge bases (AI/Rheum, DXplain, QMR)



organize concepts

Endocrine Diseases

Adrenal Gland Diseases

Adrenal Cortex Diseases

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Addison's Disease

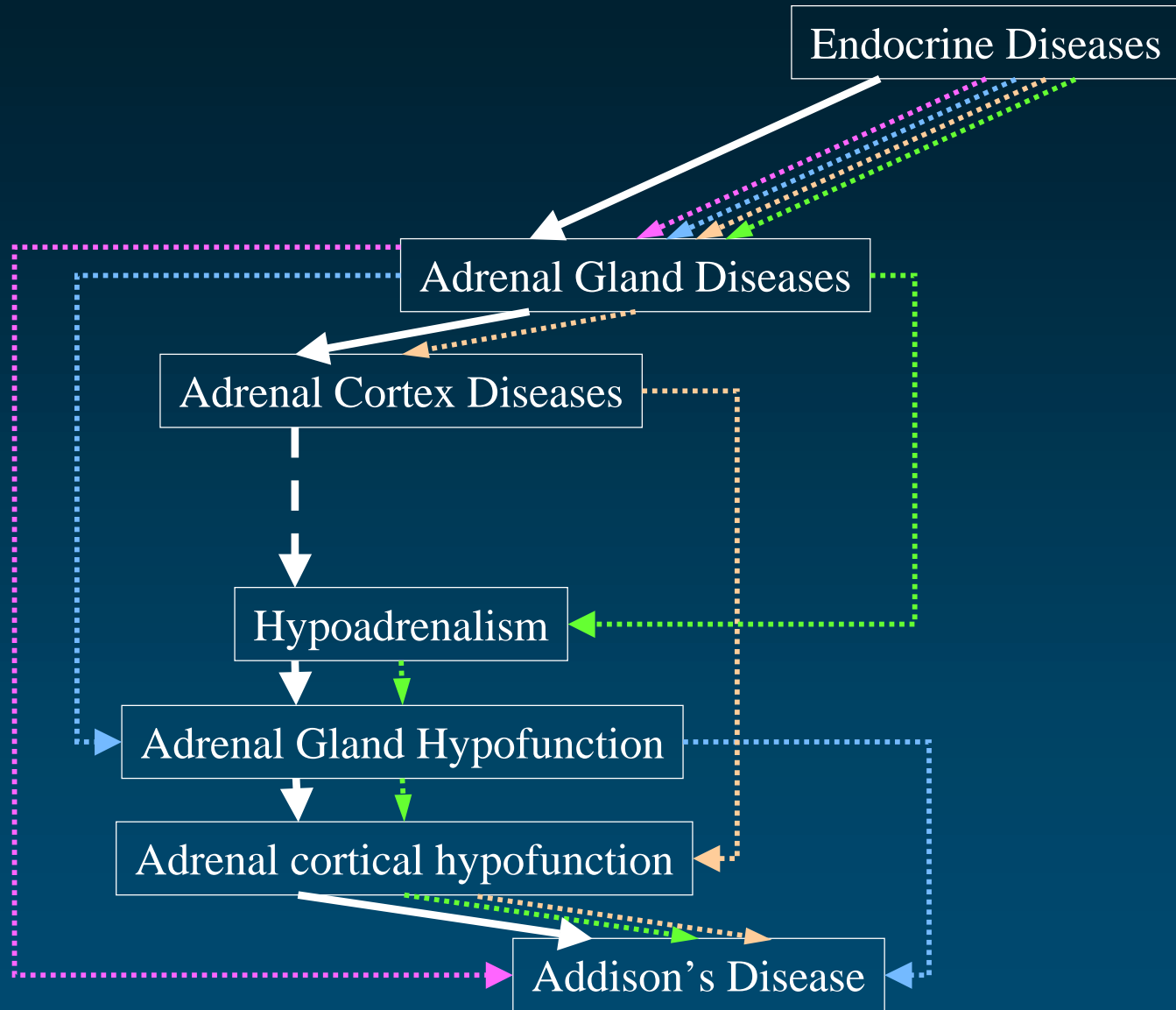
SNOMED

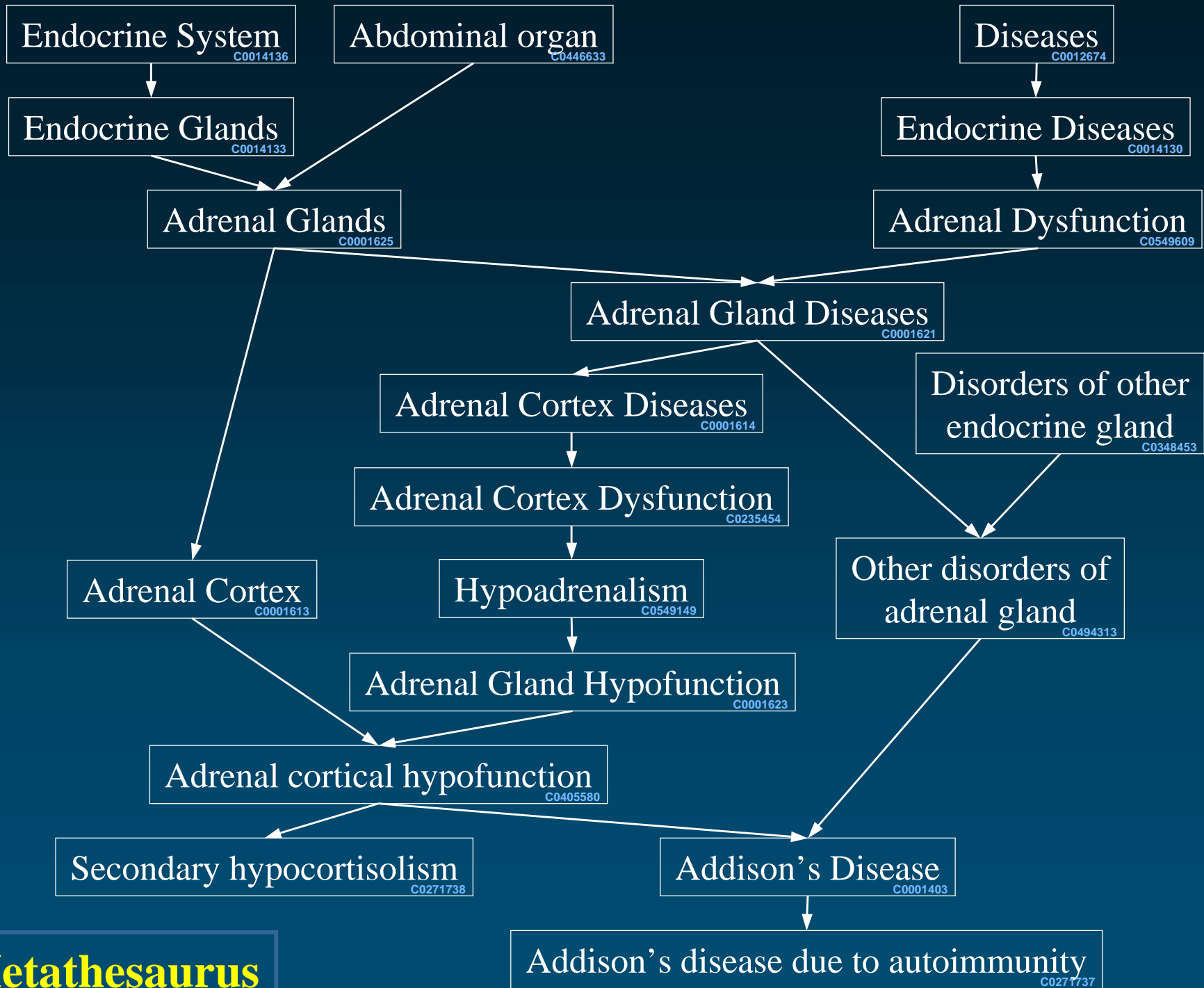
MeSH

AOD

Read Codes

UMLS



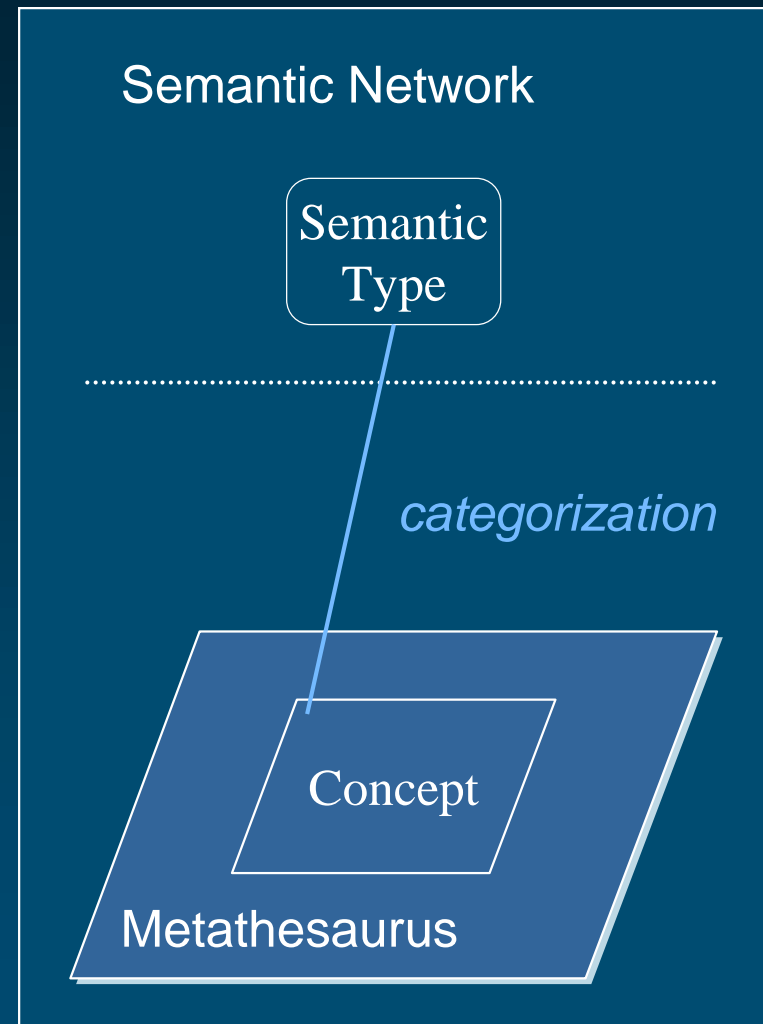


Metathesaurus

UMLS

◆ Two-level structure

- Semantic Network
 - 134 Semantic Types (STs)
 - 54 types of relationships among STs
- Metathesaurus
 - 800,000 concepts
 - ~10 M inter-concept relationships
- Link = categorization



Semantic Types

Anatomical Structure

Fully Formed Anatomical Structure

Embryonic Structure

Body Part, Organ or Organ Component

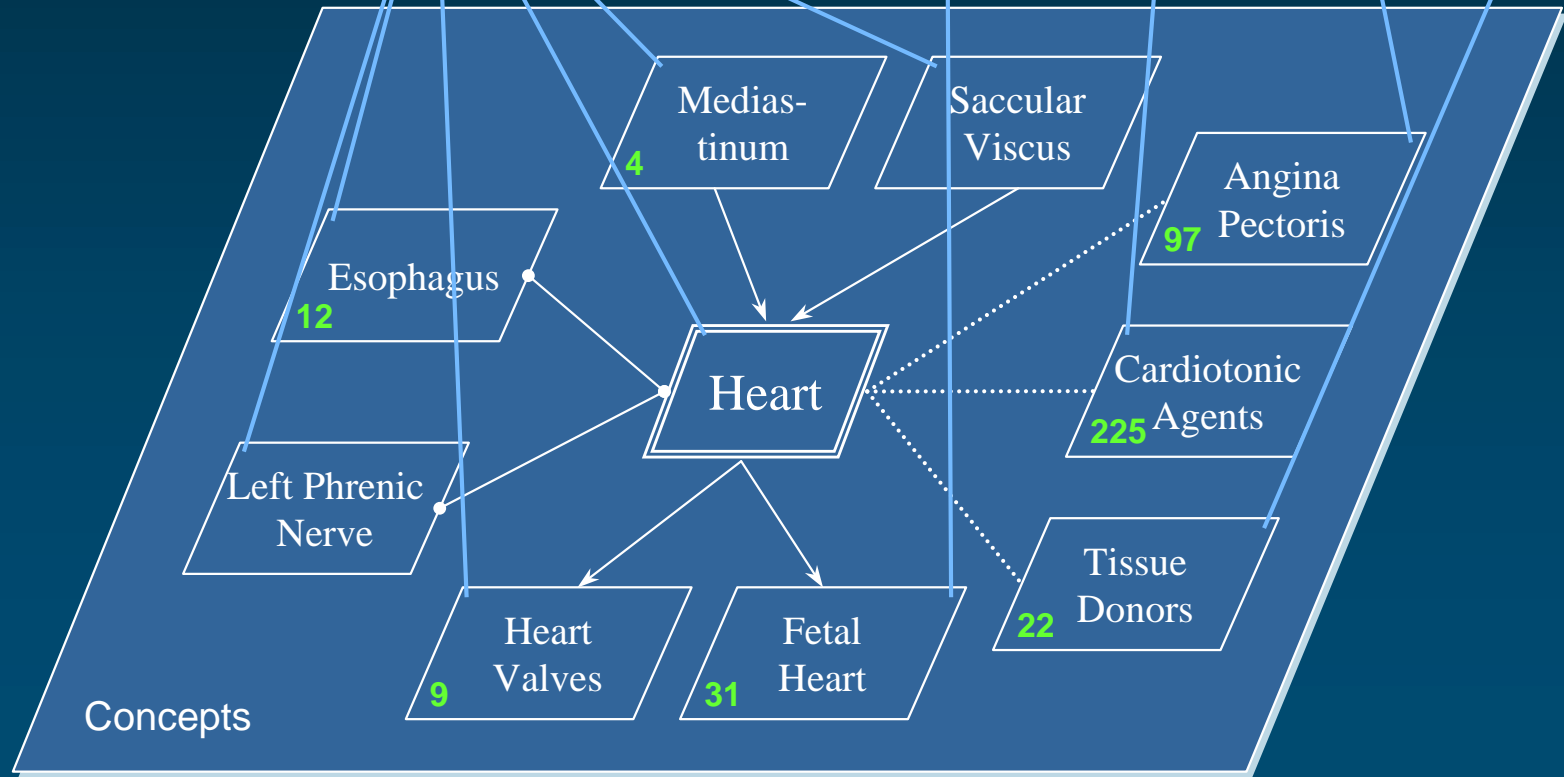
Disease or Syndrome

Pharmacologic Substance

Population Group

Semantic Network

Metathesaurus



Concepts

Esophagus
12

Mediastinum
4

Saccular Viscus

Heart

Left Phrenic Nerve

Heart Valves
9

Fetal Heart
31

Angina Pectoris
97

Cardiotonic Agents
225

Tissue Donors
22

Benefits

UMLS compared to individual vocabularies

- ◆ Broader scope
- ◆ Extended coverage
- ◆ Finer granularity
- ◆ Unique identifier
- ◆ Synonym terms clustered into concepts
- ◆ Additional synonyms
- ◆ Additional hierarchical relationships
- ◆ Semantic categorization



Limitations

Limitations

- ◆ Licensing mechanism
- ◆ Too much information
- ◆ Not enough information



Licensing mechanism

- ◆ Free UMLS registration
- ◆ 4 levels of restriction
 - L0 (~55%) must acknowledge NLM, no redistribution
 - L1 (~6%) must negotiate for translation
 - L2 (~.1%) must negotiate for creating health data
 - L3 (~39%) must negotiate for *any* production use
- ◆ Possible license fees for certain vocabularies
- ◆ [MetamorphoSys](#) helps subset by source



Too much information

- ◆ Huge
 - 1.5 M unique English strings
 - 775,000 concepts
 - Over 10 M interconcept relationships
- ◆ Complex two-level structure
 - Metathesaurus
 - Semantic Network
- ◆ Steep learning curve



Not enough information

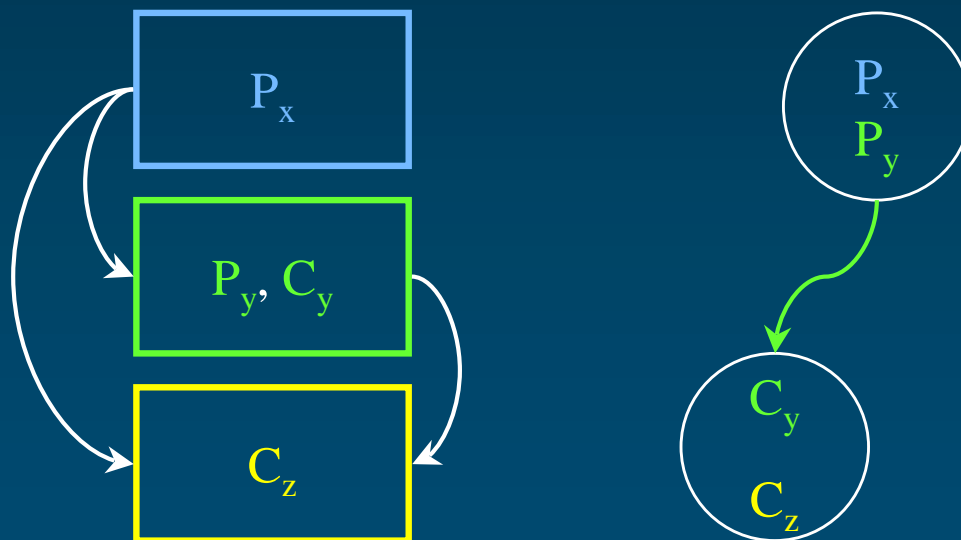
- ◆ Update
 - Frequency
 - Mechanism
- ◆ Lack of coverage
 - Major sources
 - Major subdomains
- ◆ Terminology vs. Ontology



Conclusions

Conclusions The up side

- ◆ Terminology integration is a step towards interoperability
 - Clusters of synonyms from different sources
 - Paths between terms from different sources



Conclusions The down side

- ◆ However, interoperability requires more than loosely aligned terminologies
- ◆ The UMLS does not claim to be an ontology
- ◆ The UMLS is, however, a resource for acquiring biomedical ontologies

Medical Ontology Research Project



Contact: olivier@nlm.nih.gov



Olivier Bodenreider

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