

National Library of Medicine
Presentation to
National Committee on Vital and Health Statistics
January 10, 2018 Meeting, Washington, DC

UMLS Update



U.S. National Library of Medicine

Outline of Presentation

- Introduction – Jerry Sheehan
- UMLS Overview – Olivier Bodenreider
 - Brief History
 - Overview through an example (Addison's disease)
- UMLS Users and Use Cases – Patrick McLaughlin
 - UMLS Terminology Services (UTS)
 - Usage Statistics
 - Distribution and other Issues



UMLS Overview

Olivier Bodenreider, MD, PhD

Chief, Cognitive Science Branch

Lister Hill National Center for Biomedical Communications

National Library of Medicine

olivier@nlm.nih.gov



U.S. National Library of Medicine

What does UMLS stand for?

- Unified
- Medical
- Language
- System



Motivation

- 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.»



U.S. National Library of Medicine

Lindberg DA, Humphreys BL, McCray AT.
Methods Inf Med. 1993 Aug;32(4):281-91.

The UMLS in practice

- Database
 - Series of relational files
- Interfaces
 - UMLS Terminology Services (UTS)
 - Web-based UMLS browser
 - Controlled access
 - To files for download (UMLS, RxNorm, SNOMED CT)
 - To derived applications (VSAC, CDE repository)
 - Application programming interfaces (Restful API)
- Applications
 - lvg (lexical programs)
 - MetamorphoSys (installation and customization)

“middleware” rather than end-user application



U.S. National Library of Medicine

UMLS 3 components



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

- Organize terms
- Organize concepts
- Categorize concepts

- Support discovery of synonymy



U.S. National Library of Medicine

Source Vocabularies

(2017AB)

- 153 families of source vocabularies
 - Not counting 48 translations
- 25 languages (through translations)
- Broad coverage of biomedicine
 - 10.7M names (normalized)
 - ~3.6M concepts
 - >10M relations
- Common presentation

- SNOMED CT
 - English
 - Spanish
- RxNorm
 - FDB
 - Micromedex
 - Multum
 - CVX
- LOINC
- ICD10
- ICD10-CM
- CPT
- MedDRA
- HPO
- [...]



U.S. National Library of Medicine

Internal Classification of Diseases



IV Endocrine, nutritional and metabolic diseases

- ▶ E00-E07 Disorders of thyroid gland
- ▶ E10-E14 Diabetes mellitus
- ▶ E15-E16 Other disorders of glucose regulation and pancreatic internal secretion
- ▼ E20-E35 Disorders of other endocrine glands
 - ▶ E20 Hypoparathyroidism
 - ▶ E21 Hyperparathyroidism and other disorders of parathyroid gland
 - ▶ E22 Hyperfunction of pituitary gland
 - ▶ E23 Hypofunction and other disorders of pituitary gland
 - ▶ E24 Cushing syndrome
 - ▶ E25 Adrenogenital disorders
 - ▶ E26 Hyperaldosteronism
 - ▼ E27 Other disorders of adrenal gland
 - E27.0 Other adrenocortical overactivity
 - E27.1 Primary adrenocortical insufficiency
 - E27.2 Addisonian crisis
 - E27.3 Drug-induced adrenocortical insufficiency
 - E27.4 Other and unspecified adrenocortical insufficiency
 - E27.5 Adrenomedullary hyperfunction
 - E27.8 Other specified disorders of adrenal gland
 - E27.9 Disorder of adrenal gland, unspecified
 - ▶ E28 Ovarian dysfunction
 - ▶ E29 Testicular dysfunction
 - ▶ E30 Disorders of puberty, not elsewhere classified
 - ▶ E31 Polyglandular dysfunction
 - ▶ E32 Diseases of thymus
 - ▶ E34 Other endocrine disorders
 - ▶ E35 Disorders of endocrine glands in diseases classified elsewhere

E27 Other disorders of adrenal gland

E27.0 Other adrenocortical overactivity

Overproduction of ACTH, not associated with Cushing disease
Premature adrenarche

Excl.: Cushing syndrome ([E24.-](#))

E27.1 Primary adrenocortical insufficiency

Addison disease
Autoimmune adrenalitis

Excl.: amyloidosis ([E85.-](#))
tuberculous Addison disease ([A18.7](#))
Waterhouse-Friderichsen syndrome ([A39.1](#))

E27.2 Addisonian crisis

Adrenal crisis
Adrenocortical crisis

E27.3 Drug-induced adrenocortical insufficiency

Use additional external cause code (Chapter XX), if desired, to identify drug.

E27.4 Other and unspecified adrenocortical insufficiency

Adrenal:

- haemorrhage
- infarction

 Adrenocortical insufficiency NOS
Hypoadosteronism

Excl.: adrenoleukodystrophy [Addison-Schilder] ([E71.3](#))
Waterhouse-Friderichsen syndrome ([A39.1](#))

E27.5 Adrenomedullary hyperfunction

Adrenomedullary hyperplasia
Catecholamine hypersecretion

E27.8 Other specified disorders of adrenal gland

Abnormality of cortisol-binding globulin

E27.9 Disorder of adrenal gland, unspecified

Medical Subject Headings

[Search](#)[Tree View](#)[MeSH on Demand](#)**NEW**[MeSH 2017](#)[MeSH Suggestions](#)[About MeSH Browser](#)

Addison Disease MeSH Descriptor Data 2018

[Details](#)[Qualifiers](#)[MeSH Tree Structures](#)[Concepts](#)

Endocrine System Diseases [C19]

Adrenal Gland Diseases [C19.053]

Adrenal Insufficiency [C19.053.500]

Addison Disease [C19.053.500.263] -

Hypoadrenocorticism, Familial [C19.053.500.263.500]

Adrenoleukodystrophy [C19.053.500.270]

Hypoaldosteronism [C19.053.500.480]

Waterhouse-Friderichsen Syndrome [C19.053.500.740]

Immune System Diseases [C20]

Autoimmune Diseases [C20.111]

Addison Disease [C20.111.163]

Anemia, Hemolytic, Autoimmune [C20.111.175]

Anti-Glomerular Basement Membrane Disease [C20.111.190]

Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis [C20.111.193] +

Antiphospholipid Syndrome [C20.111.197]

[...]

SNOMED CT

SNOMED
International

Concept Details

Concept Details

SummaryDetailsDiagramExpressionRe

Parents

- Abdominal organ finding (finding)
- Disorder of abdomen (disorder)
- Disorder of endocrine system (disorder)
- Disorder of adrenal gland (disorder)
- Hypoadrenalism (disorder)
- Adrenal hypofunction (disorder)
- Disorder of adrenal gland (disorder)
- Disorder of adrenal cortex (disorder)
- Adrenal cortical hypofunction (disorder)

Disorder of endocrine system

Disorder of adrenal gland

Hypoadrenalism

Adrenal hypofunction

Disorder of adrenal cortex

Adrenal cortical hypofunction

Addison's Disease

StatedInferred

Addison's disease (disorder) ☆

SCTID: 363732003

363732003 | Addison's disease (disorder) |

Addison disease

Addison's disease

Addison's disease (disorder)

Finding site → Adrenal cortex structure

Children (4)

- Addison's disease due to autoimmunity (disorder)
- Addison's disease with adrenoleucodystrophy (disorder)
- Polyglandular autoimmune syndrome, type 1 (disorder)
- Tuberculous Addison's disease (disorder)

[1] Organize terms

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

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

C0001403

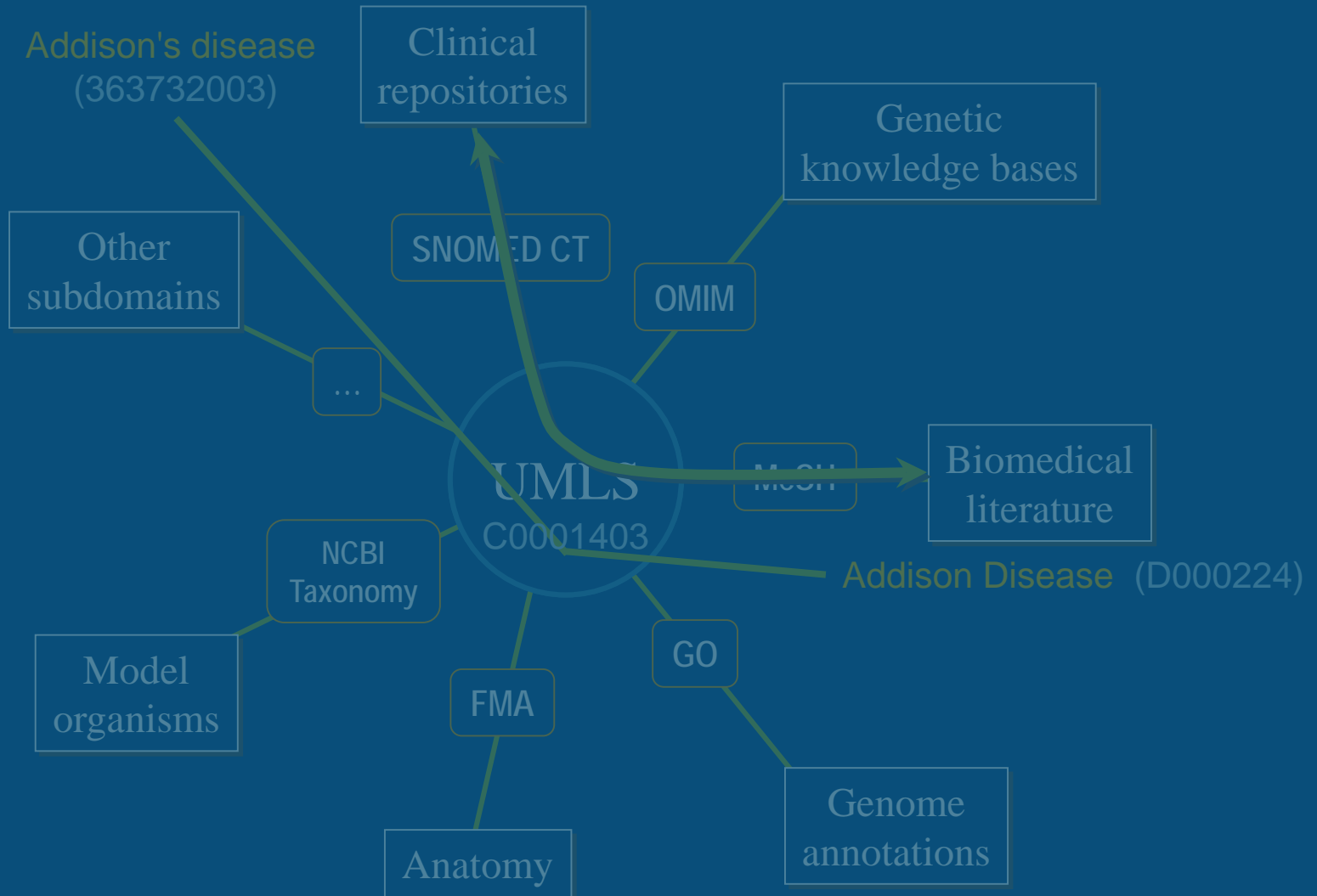
Addison's disease

No curation of the sources by NLM

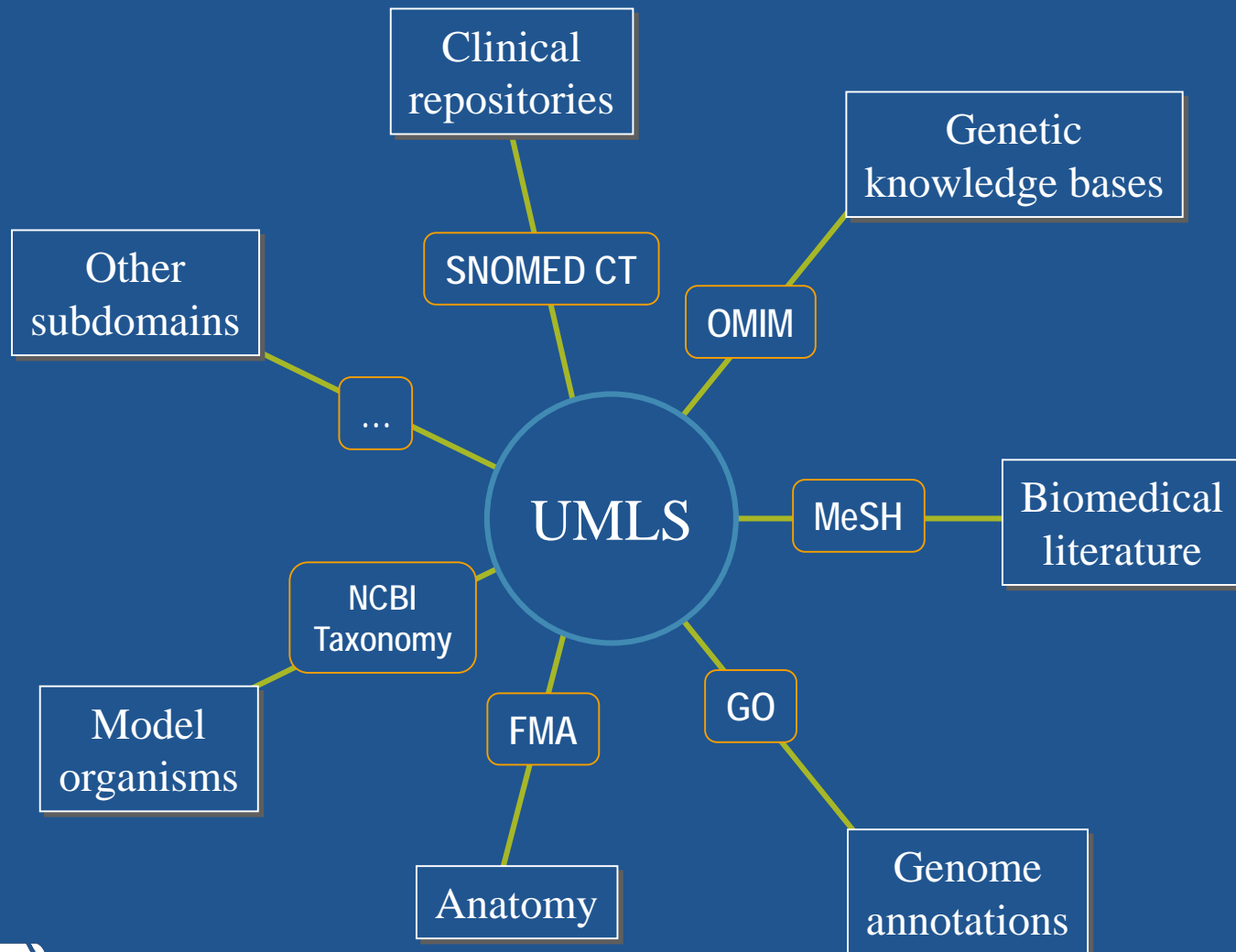


U.S. National Library of Medicine

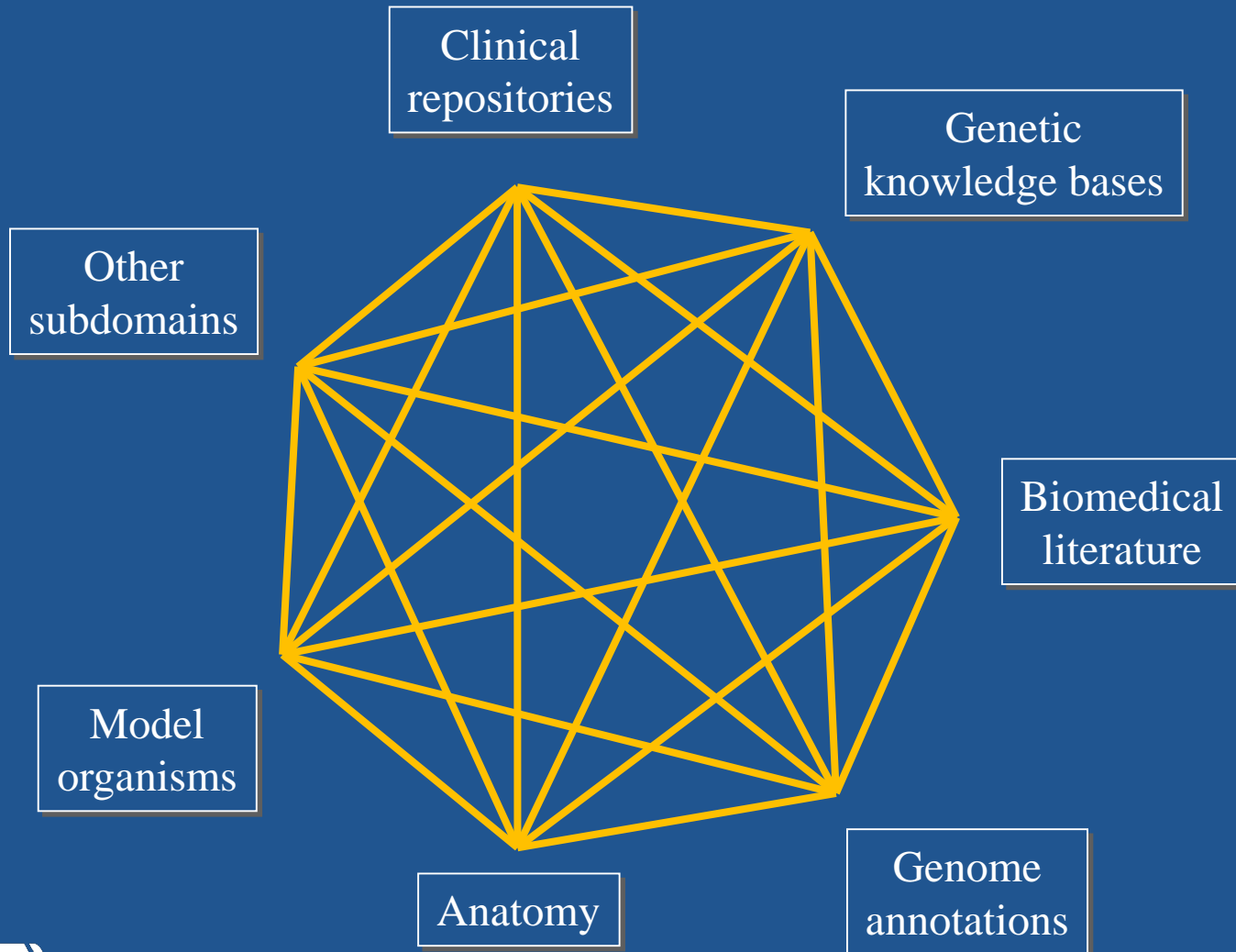
Terminology integration



Integrating subdomains

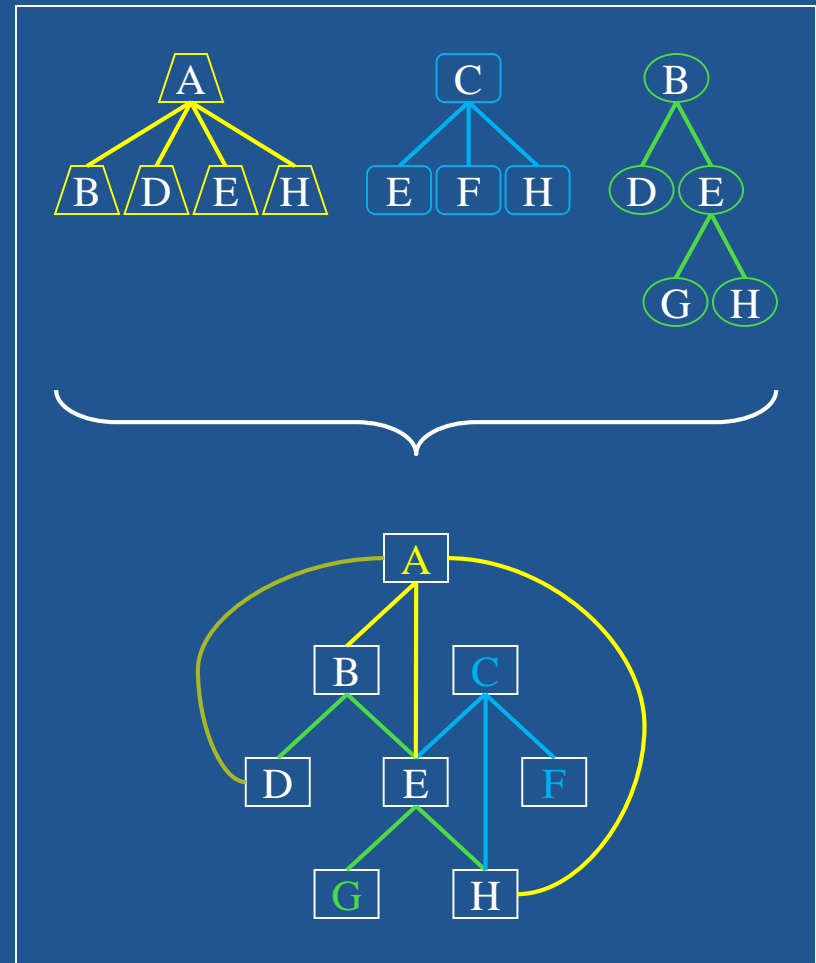


Integrating subdomains

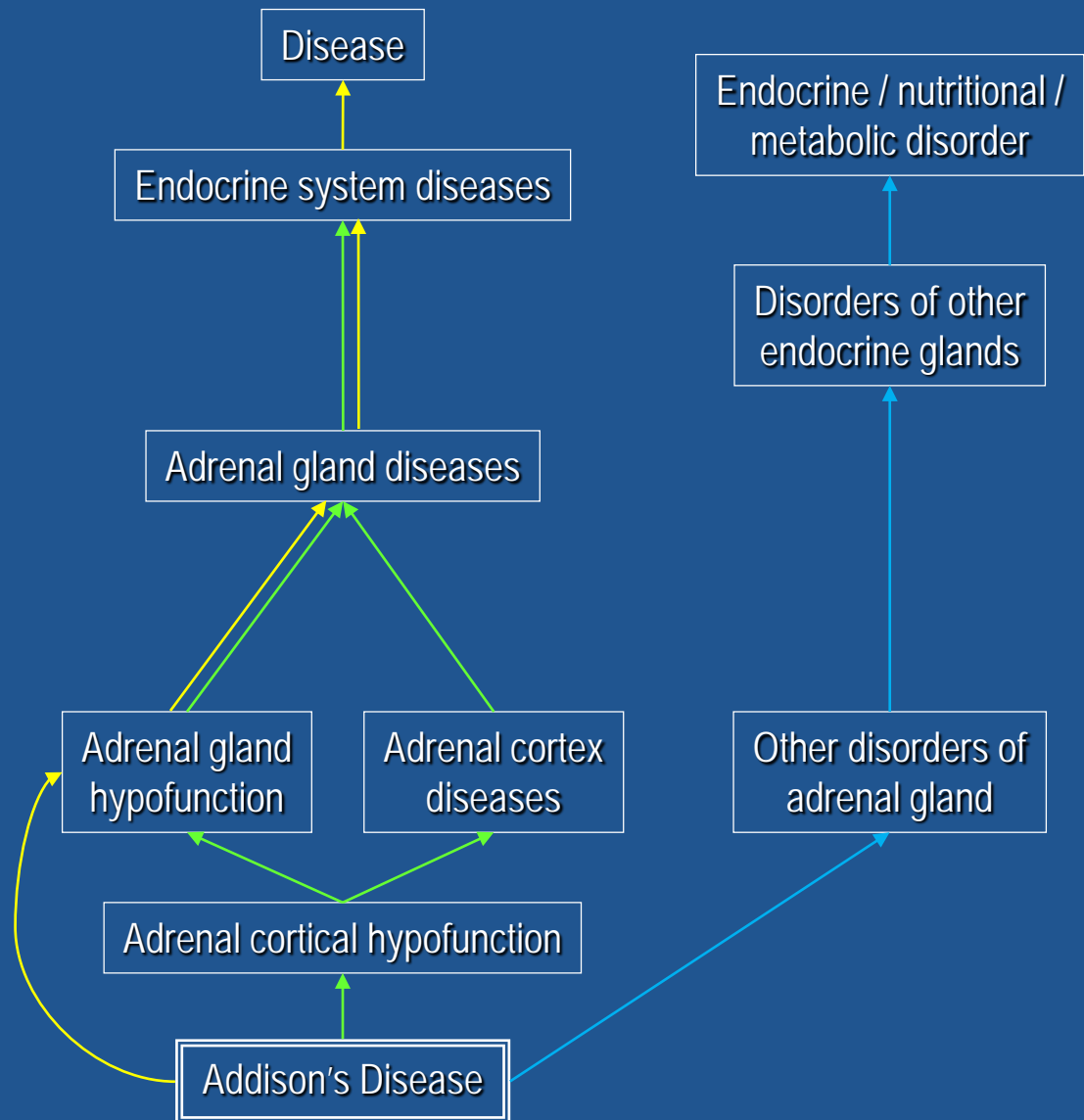


[2] Organize concepts

- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)
- *No curation of the relations by NLM*



organize concepts

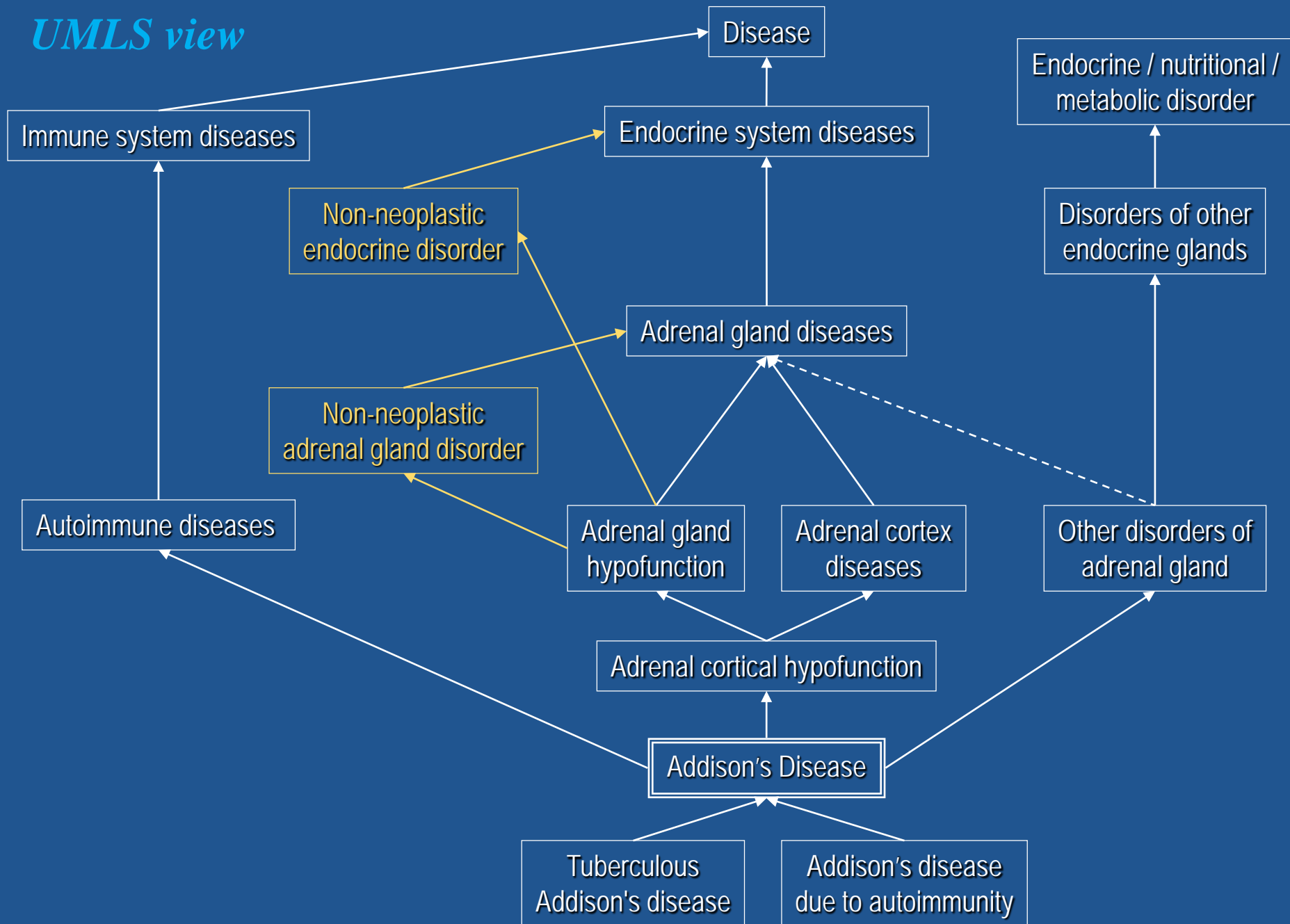


SNOMED CT

MeSH

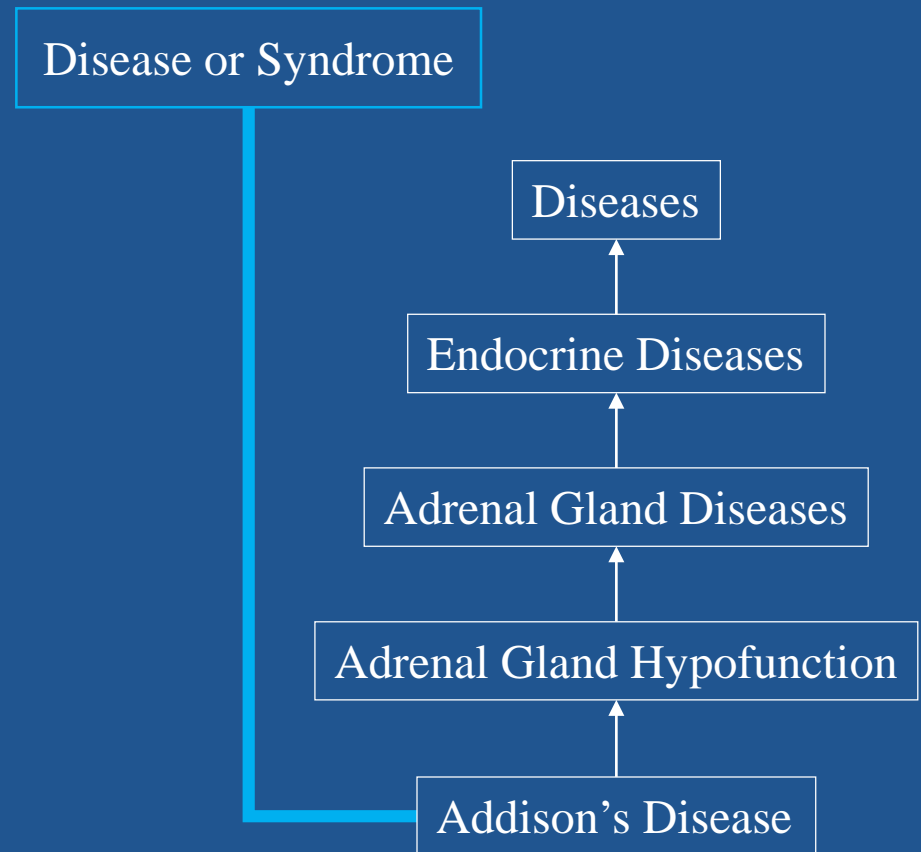
ICD-10

UMLS view

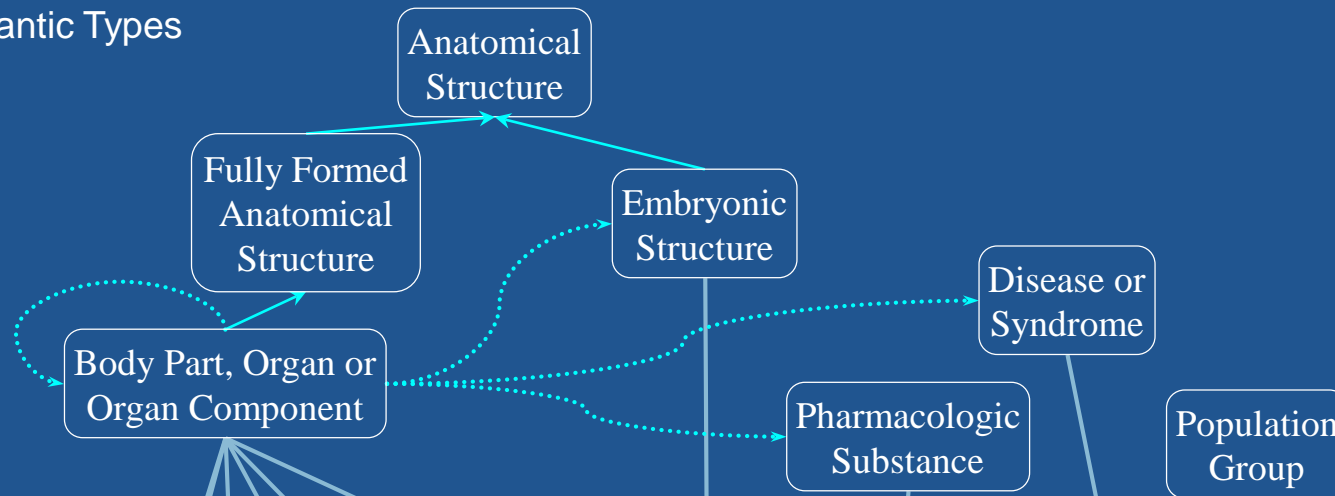


[3] Categorize concepts

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



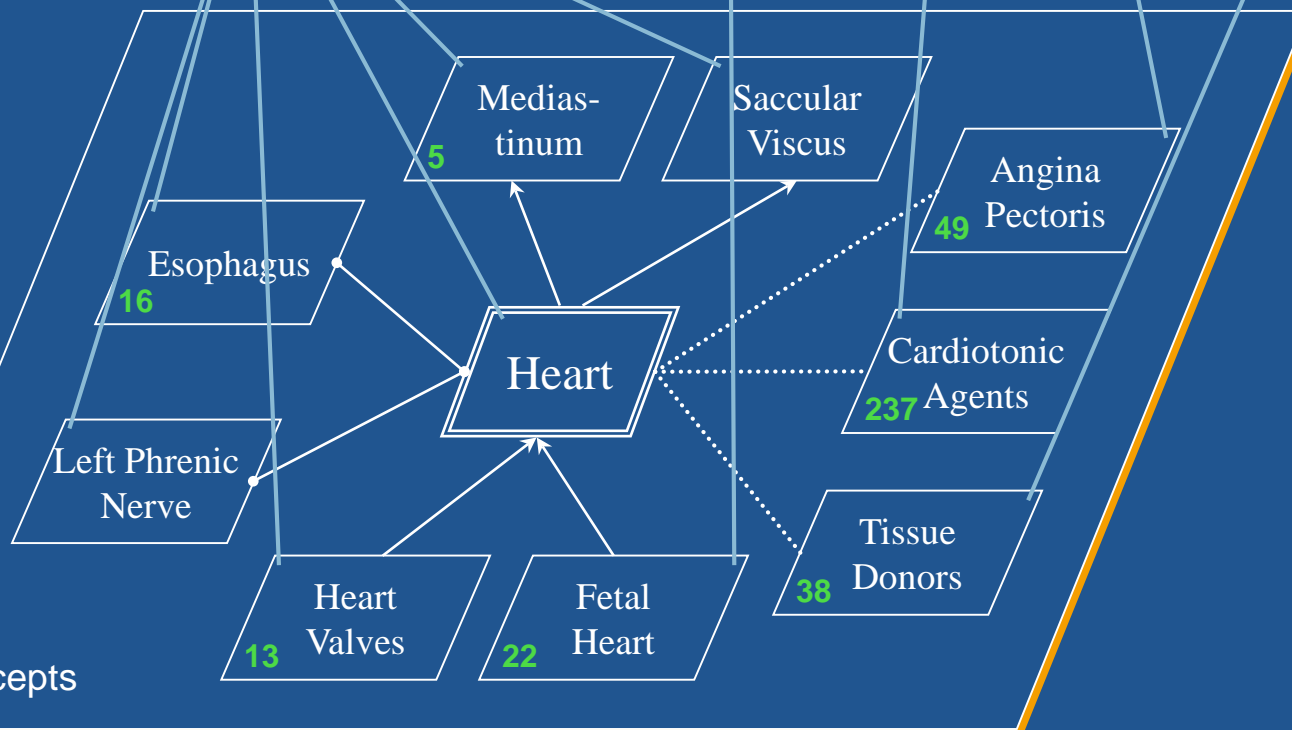
Semantic Types



Semantic Network

Metathesaurus

Concepts



UMLS Uses and Use Cases

Patrick McLaughlin

Head, Terminology QA and Customer Services

National Library of Medicine

Patrick.mclaughlin@nih.gov



U.S. National Library of Medicine

UMLS Terminology Services (UTS)

<https://uts.nlm.nih.gov/>

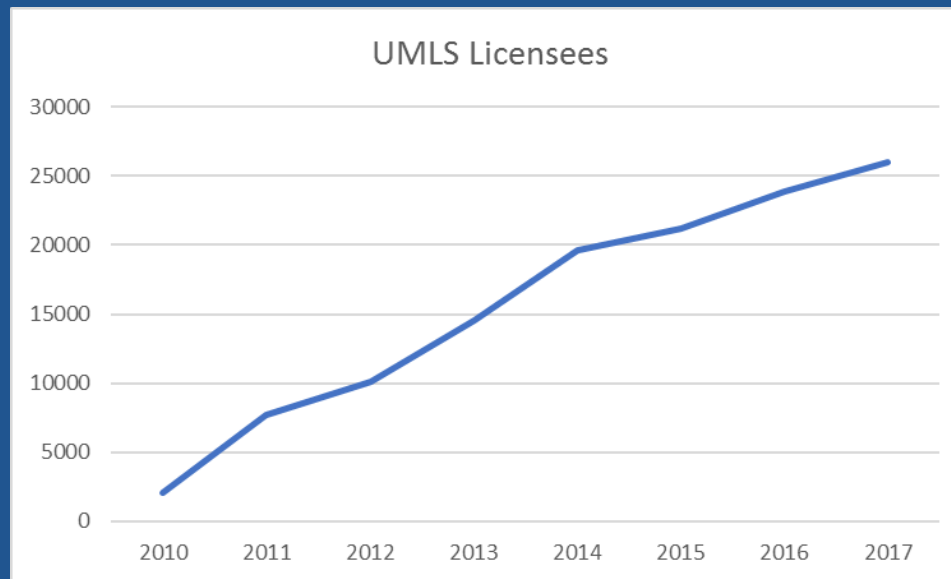
- Browsing/searching of Metathesaurus and SNOMED CT
- API access to Metathesaurus
- Distribution point for terminology products
 - UMLS, RxNorm, SNOMED CT
- Licensing and user authentication system
 - Designed to protect the IP of source vocabularies – provide access for research purposes and provide contact information for secondary licensing where necessary (e.g. CPT, nursing terminologies)
 - User authentication used for internal (VSAC, MetaMap, CDE Repository, etc.) and external applications
- Annual usage reporting



U.S. National Library of Medicine

Licensees

- 26,000 licensees currently
 - 78% of licensees based in the US (134 total countries represented)



Annual Report - Statistics

Results from the CY2016 Annual Usage Report

- Total Licensees: 23,000
- Number of respondents: 11,000
 - UMLS: 4500
 - RxNorm: 3600
 - SNOMED CT International Release: 3600
 - US Edition of SNOMED CT: 3100
 - LOINC: 3000



U.S. National Library of Medicine

Annual Report - Affiliations

| | |
|--|-----|
| Academic Institution | 33% |
| For-profit entity | 23% |
| Not-for-profit entity | 18% |
| Individual Use | 17% |
| Government Institution - US Federal Government - not NLM | 2% |
| Government Institution - US State or local government | 2% |
| Government Institution - Government outside the US | 1% |
| Other | 4% |



Annual Report - Organizations

- IBM
- Elsevier
- HCA
- Cerner
- Allscripts
- Partners Healthcare
- Kaiser Permanente
- Athenahealth
- Intermountain Healthcare
- Mayo Clinic
- MITRE
- VA
- FDA
- NIH
- Columbia University
- Harvard Medical School
- University of Michigan
- George Mason University
- Indiana University
- University of Minnesota
- UCLA
- MIT
- Georgia Institute of Technology
- University of Utah
- Stanford University
- University of Pittsburgh
- OHSU
- etc.



U.S. National Library of Medicine

Annual Report – Use Cases

- Utilize specific terminologies from the Metathesaurus
- Mapping between terminologies
- Terminology research
- Processing of texts to extract concepts, relationships or knowledge
- Information indexing/annotations and retrieval
- Creation and maintenance of local terminology
- Concept-oriented synonymy
- Support of a terminology server or service



U.S. National Library of Medicine

UMLS Uses at NLM

- Semi-automatic indexing of MEDLINE
- Information retrieval in PubMed and MedGen
- Consumer health information exchange with EHRs/PHRs from MedlinePlus Connect
- Production of SNOMED CT subsets:
 - CORE Problem List of SNOMED CT
 - Nursing Problem List Subset of SNOMED CT



U.S. National Library of Medicine

Annual Report – Example Uses

- Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES)
- VisualDx – tool for diagnosis of skin conditions and disorders
- CrowdTruth – framework for crowdsourcing the collection of annotation data on text, images and videos
- Observational Health Data Sciences and Informatics (OHDSI) – vocabulary resources and OMOP data model
- Indian Health Service – Resource and Patient Management System (RPMS)
- PatientsLikeMe – patient health information site



Annual Report – Feedback

- At the request of the user community, NLM now produces the Metathesaurus only twice a year (May and November)
- Frequent requests for more non-English content, more vocabularies
- Increasing pressure from various parts of the user community saying the license is a burden
 - Restriction on usage of content
 - Need for authentication to access data



Questions?



U.S. National Library of Medicine