



2014 Protein Ontology Workshop
Washington, DC
June 18, 2014

Ontologies at the NLM



Olivier Bodenreider

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



U.S. National Library of Medicine

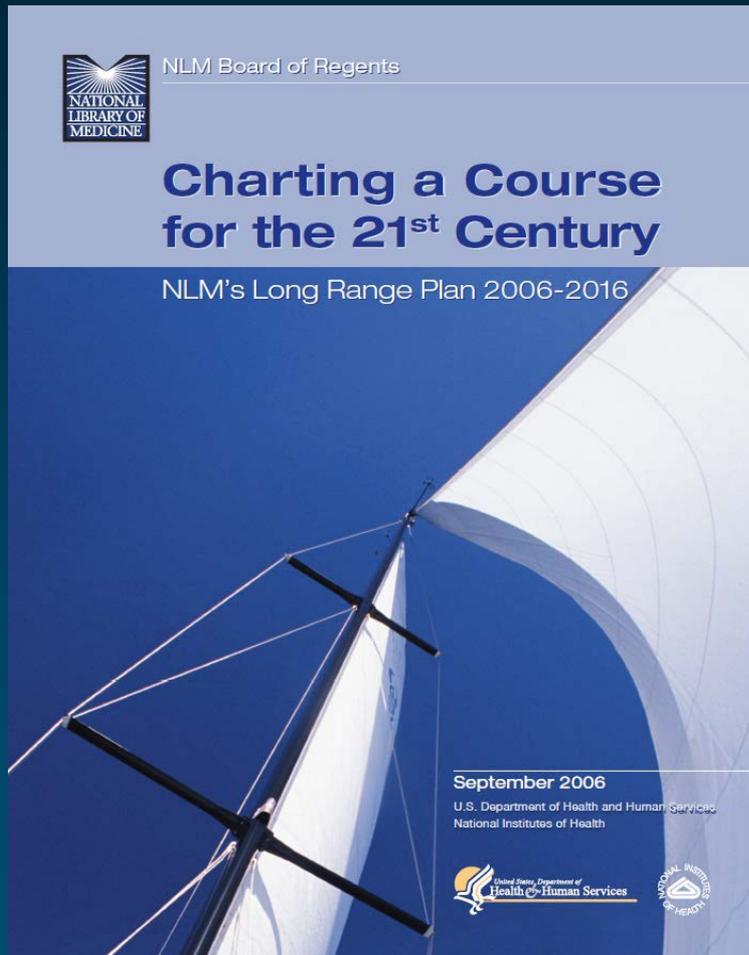


Disclaimer

The views and opinions expressed do not necessarily state or reflect those of the U.S. Government, and they may not be used for advertising or product endorsement purposes.



NLM Portfolio



Goal 1. Seamless, Uninterrupted Access to Expanding Collections of Biomedical Data, Medical Knowledge, and Health Information

Goal 2. Trusted Information Services that Promote Health Literacy and the Reduction of Health Disparities Worldwide

Goal 3. Integrated Biomedical, Clinical, and Public Health Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice

Goal 4. A Strong and Diverse Workforce for Biomedical Informatics Research, Systems Development, and Innovative Service Delivery



Ontologies at the NLM

Goal 3. Integrated Biomedical, Clinical, and Public Health Information Systems that Promote Scientific Discovery and Speed the Translation of Research into Practice

Recommendation 3.3. *Promote development and use of advanced electronic representations of biomedical knowledge in conjunction with electronic health records.*



Overview

- ◆ Terminology development
 - MeSH
 - RxNorm
 - SNOMED CT
 - Support for LOINC, FMA
- ◆ Terminology derivatives
 - CORE problem list subset
 - Interface terminology
 - Mappings
- ◆ Terminology integration & distribution
 - UMLS
 - APIs
- ◆ Terminology binding
 - Value sets (VSAC)
- ◆ Terminology research
 - Medical Ontology Research
 - Applied Medical Terminology Research
- ◆ Terminology in action
 - Natural language processing
 - Information retrieval



Terminology development

Standard vocabularies in the era of Meaningful Use

- ◆ Diagnoses / Diseases / Conditions
 - International classification of diseases (ICD)
 - **SNOMED CT**
- ◆ Procedures
 - Current Procedural terminology (CPT)
 - ICD10-PCS
 - **SNOMED CT**
- ◆ Drugs
 - **RxNorm**
- ◆ Laboratory tests
 - **LOINC**



NLM role

- ◆ In-house development
 - Medical Subject Headings (MeSH)
 - RxNorm
- ◆ National release center for SNOMED CT
 - Development of the U.S. extension
- ◆ Support for standards development
 - LOINC
 - Foundational Model of Anatomy (FMA)
 - For the Visible Human





Welcome to the U.S. SNOMED CT® Content Request System (USCRS). This system allows users to request basic changes to SNOMED CT.

Log into the USCRS with a UMLS® Terminology Services (UTS) account. Visit the UTS site to sign up for an account if you do not already have one.

Users select from 10 online templates to enter requests to add, change or retire information in SNOMED CT:

- New Concept
- New Synonym
- Add Parent
- Change Description
- Change Parent
- Change Relationship
- New Relationship
- Retire Concept
- Retire Description
- Retire Relationship

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

Users submit online requests separately or grouped together to form a batch. Save draft requests for later completion, validation, and submission.

Users may also download the USCRS batch request template to submit multiple requests prepared offline. The template, in Excel® spreadsheet format, includes instructions for use, and a separate worksheet for submission of each of the request types. Upload completed spreadsheets into the USCRS to generate tasks to review, edit, validate and submit.

User may view and search requests and track their progress through evaluation and eventual inclusion into either the International Release of SNOMED CT (if accepted) or the US Extension to SNOMED CT. Search and view requests submitted by other users for similar or identical requests.

Each page of the USCRS contains a help link to context-sensitive help documentation.

An understanding of SNOMED CT content and structure is required, and each request must include a justification, or practical use case. Every request must also include an identifier that links to the current version of either SNOMED CT International or the US Extension to SNOMED CT. Both SNOMED CT International and the US Extension are available from the Downloads menu of the UTS. Users may search for SNOMED CT terms, ConceptIDs, and DescriptionIDs in the NLM SNOMED CT Browser which is available from the Applications menu of the UTS.

The U.S. SNOMED CT Content Request System is an important new tool in the NLM effort to support the development, enhancement, and distribution of clinically specific vocabularies to facilitate the exchange of clinical data and improve retrieval of health information. SNOMED CT is one of a suite of designated standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information and is also a required standard in interoperability specifications of the U.S. Healthcare Information Technology Standards Panel.

NLM welcomes your suggestions to improve this new system. Send questions and suggestions on the U.S. SNOMED CT Content Request System to NLM Customer Service (custserv@nlm.nih.gov) with the subject line "U.S. SNOMED CT Content Request System."

USCRS Version: [2.2.4.RELEASE](#)

Doc Version: 2.2.4.RELEASE

Last Reviewed: 20 January 2012

Last Updated: 20 January 2012

First Published: 20 January 2012

Terminology derivatives

CORE problem list subset

- ◆ CORE - Clinical Observations Recording and Encoding
- ◆ Problem List Subset of SNOMED CT
- ◆ Derived from empirical problem frequencies reported by several large clinical institutions
- ◆ Uses
 - Problem lists in EHR systems
 - Resource prioritization (mapping, translation, quality assurance)
 - Linking to patient education materials through NLM's MedlinePlus Connect

http://www.nlm.nih.gov/research/umls/Snomed/core_subset.html



CORE problem list subset

SNOMED CT ID	Name	UMLS CUI	# inst.	Frequency
38341003	Hypertensive disorder, systemic arterial (disorder)	C0020538	8	3.2242
55822004	Hyperlipidemia (disorder)	C0020473	8	2.1369
35489007	Depressive disorder (disorder)	C0011581	8	1.5077
235595009	Gastroesophageal reflux disease (disorder)	C0017168	8	1.3691
268565007	Adult health examination (procedure)	C0420151	5	1.1992
44054006	Diabetes mellitus type 2 (disorder)	C0011860	8	1.0432
59621000	Essential hypertension (disorder)	C0085580	7	0.9291
414916001	Obesity (disorder)	C0028754	8	0.9252
73211009	Diabetes mellitus (disorder)	C0011849	8	0.9239
195967001	Asthma (disorder)	C0004096	8	0.8856



Interface terminology – RxTerms

- ◆ Derived from RxNorm
- ◆ To support efficient data entry for e-prescribing
- ◆ Used in an assessment tool by the Centers for Medicare and Medicaid Services (CMS)
- ◆ Incorporated in EHRs and other medication-related applications

<http://wwwcf.nlm.nih.gov/umlslicense/rxtermApp/rxTerm.cfm>



Interface terminology – RxTerms

◆ RxNorm

- 259255 atorvastatin 80 MG Oral Tablet

◆ RxTerms

- Atorvastatin (Oral Pill)
 - Oral Pill
 - Tab
 - 80 mg



Mappings between terminologies

- ◆ Main objective
 - To facilitate the adoption of SNOMED CT (for clinical documentation) by deriving billing codes automatically
- ◆ Point-to-point vs. rule-based mapping
 - When additional information is required

http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html



SNOMED CT to ICD-10-CM

◆ Point-to-point

- 41893002 Left ventricular-right atrial communication (disorder)
- Q21.0 Ventricular septal defect

◆ Rule-based

- 335002 Pylorospasm (disorder)
 - If age of onset < 29 days: Q40.0 Congenital hypertrophic pyloric stenosis
 - Otherwise: K31.3 Pylorospasm, not elsewhere classified



Terminology integration & distribution

Terminology integration

- ◆ Unified Medical Language System (UMLS)
 - Metathesaurus
- ◆ 139 families of source vocabularies
 - Not counting translations
- ◆ 21 languages
- ◆ Broad coverage of biomedicine
 - 8.6M names (normalized)
 - ~3M concepts
 - >10M relations
- ◆ Common presentation



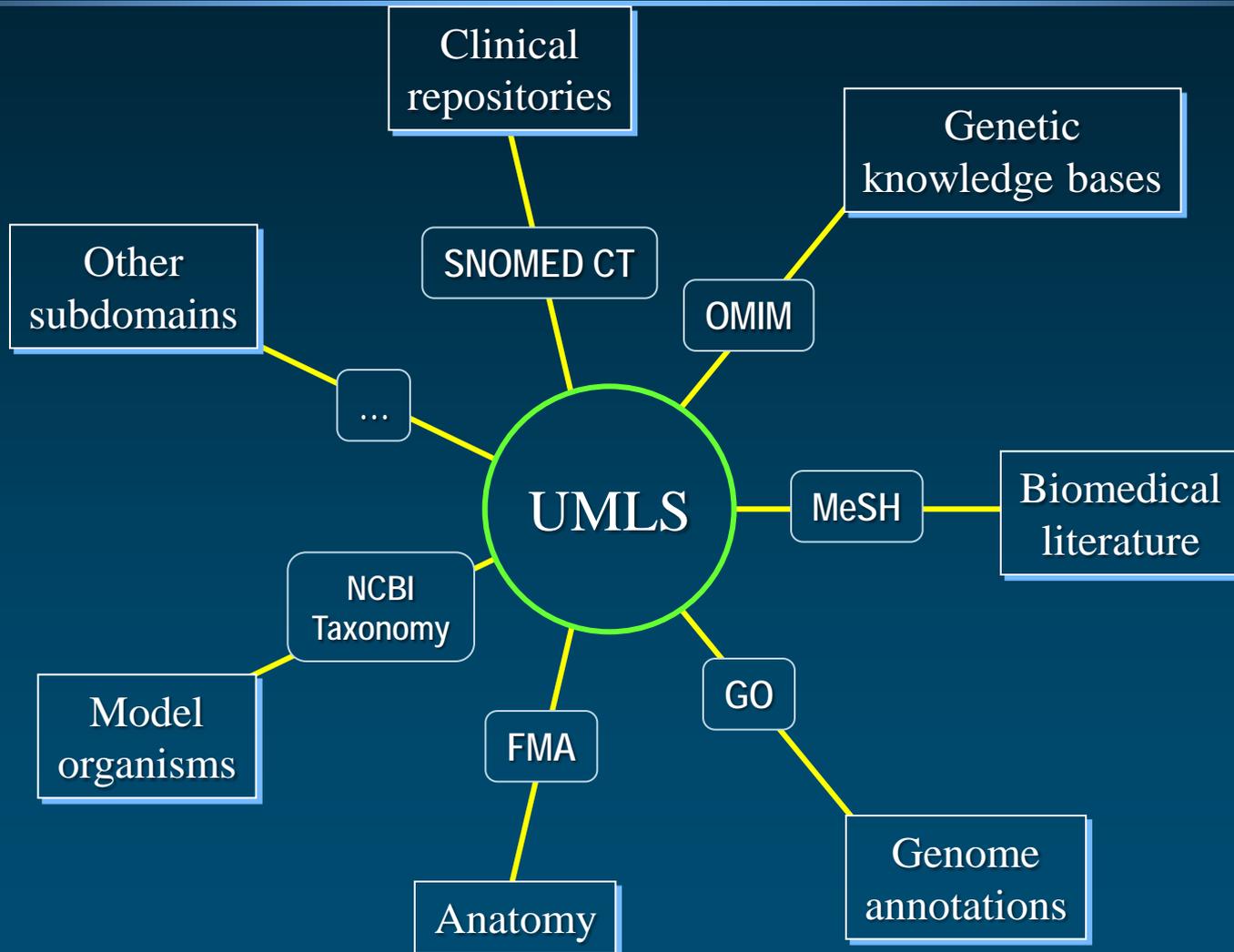
Addison's disease – Different names

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

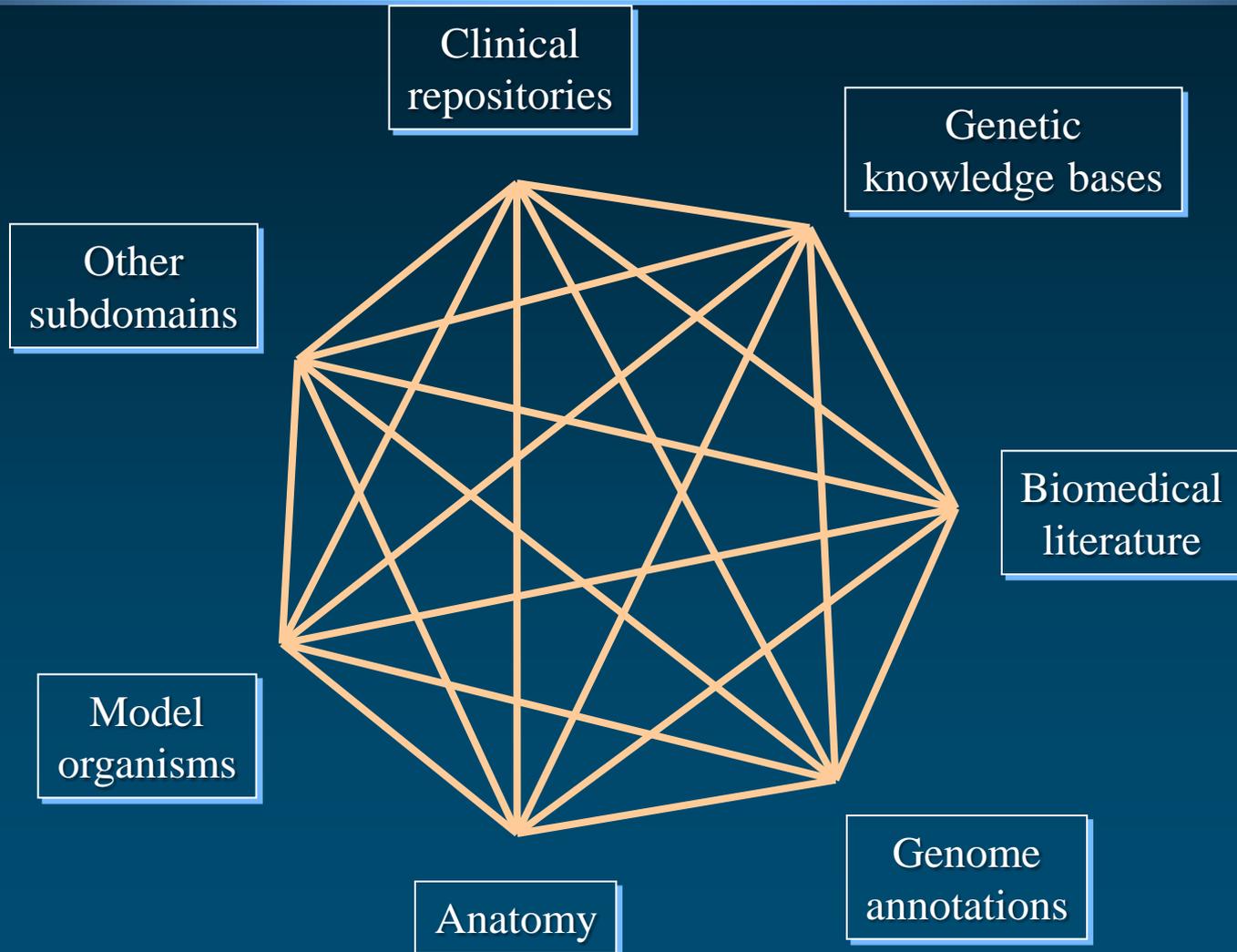
Addison's disease



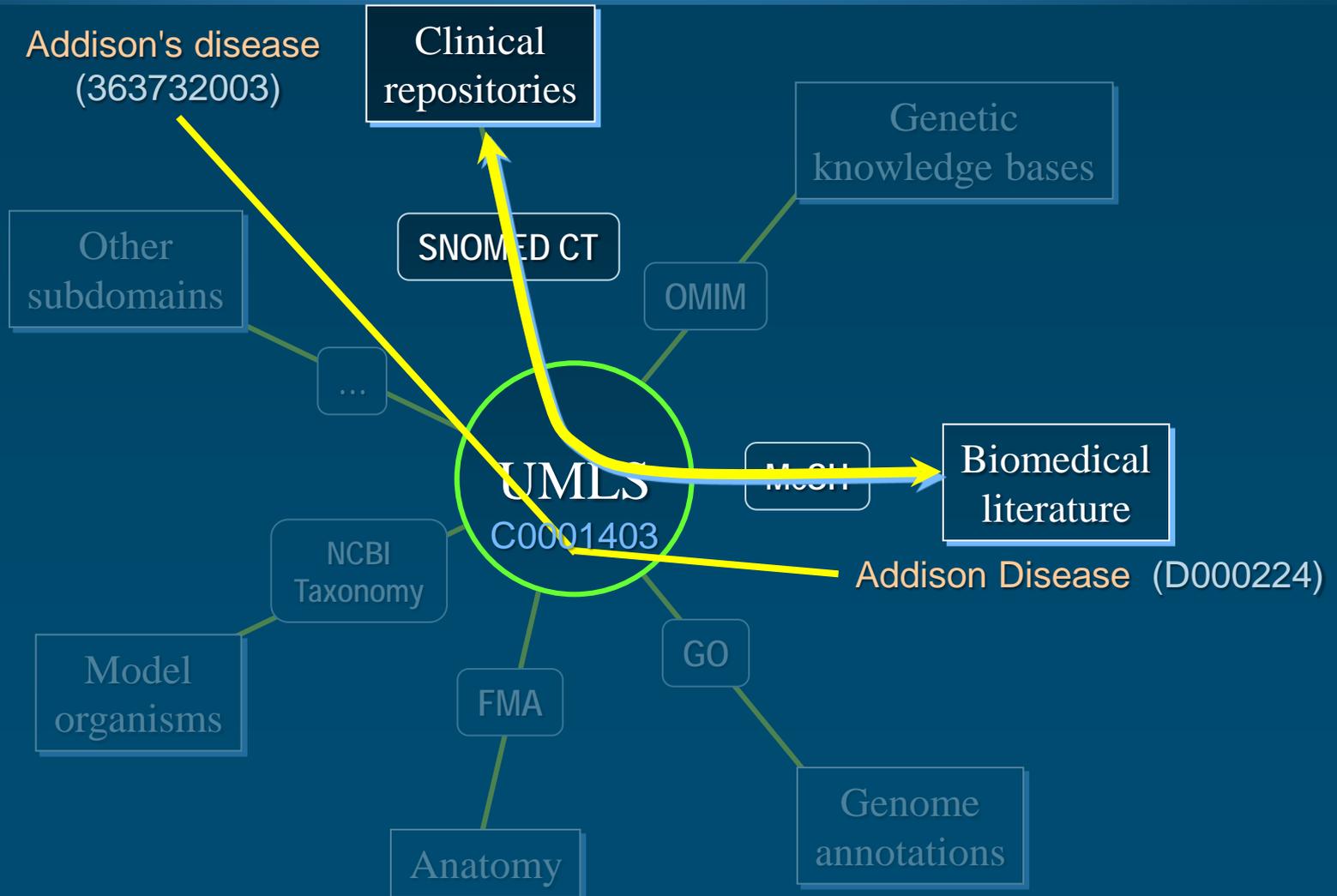
Integrating subdomains



Integrating subdomains



Terminology integration



Unified Medical Language System

A service of the U.S. National Library of Medicine | National Institutes of Health

[Sign In](#) | [Sign Up](#) | [Contact](#)



Unified Medical
Language System®

UMLS Terminology Services

[UTS Home](#) [Applications](#) [SNOMED CT](#) [Resources](#) [Downloads](#) [Documentation](#) [UMLS Home](#) ↗

Welcome to the UTS

[Welcome](#) »

[Request a License](#)

[News/Announcements](#)

[Training](#)

[Supported Browsers](#)

The UMLS Terminology Services (UTS) allows you to:

- Request a UMLS Metathesaurus License and create a UTS account
- Search and display content from UTS Applications including:
 - Metathesaurus Browser
 - Semantic Network Browser
 - SNOMED CT Browser
- Download data files including:
 - UMLS Knowledge Sources
 - RxNorm weekly and monthly updates
 - SNOMED CT
 - CORE Problem List and Route of Administration Subsets of SNOMED CT
- Query data remotely via Web Services (see API Documentation)
- Complete UMLS Annual Report and SNOMED CT® Affiliate Reports

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

Terminology distribution

◆ Various mechanisms

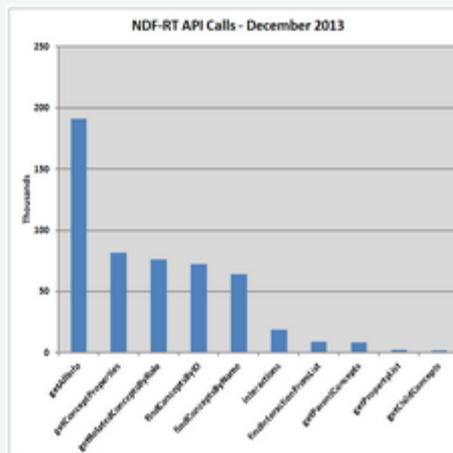
- Set of relational files / XML dataset
 - UMLS
 - MeSH
 - RxNorm
- Coming soon: RDF datasets and SPARQL endpoints
 - MeSH, PubChem
- Application programming interfaces (APIs)
 - UMLS
 - RxNorm



APIs

Application Program Interfaces (APIs) are available to users to retrieve data from several drug information sources, including RxNorm, NDF-RT and RxTerms.

Currently there are four APIs available - RxNorm, Prescribable RxNorm, NDF-RT and RxTerms



RxNav

APIs

RxMix

<http://rxnav.nlm.nih.gov/>

News

Statistics

RxNav brand names

Recent brand names added

Alocane Plus	Dalvance
Izba	Omtryg
Optimend Corneal Repair	Optivet Allergy Relief
Optivet Eye Irrigating	

NDF-RT Drug Interactions

Interactions to be removed from NDF-RT

The NDF-RT data provider has announced that the drug interactions will be removed from the data set as early as the July release. There is currently no replacement for this interactions data.

APIs - What's New

Drug Class Members - the RxNorm API function `getClassMembers` (REST: `/members`) retrieves the members of a specified drug class from ATC, MeSH or NDF-RT. [Read more](#)

Drug Classes. A new function `getAllClasses` (REST: `/classes`) allows users to get drug class names from ATC, MeSH or NDF-RT.

Terminology binding

Terminology binding

◆ Binding between

- Information model artifact
 - Question in a form
 - Field in a survey instrument
- Terminology artifact
 - Set of values in from a given code system
 - Extensional
 - Intensional

◆ Example

- Clinical quality measures



CLINICAL QUALITY MEASURE (simplified)

Hemoglobin A1c Test for Pediatric Patients

diabetic patients [age 5-17] *tested for HbA1c*

=

diabetic patients [age 5-17]

CLINICAL QUALITY MEASURE (details)

Hemoglobin A1c Test for Pediatric Patients

Tests for HbA1c

diabetic patients [age 5-17] *tested for HbA1c*

=

diabetic patients [age 5-17]

- Type 1 or Type 2 diabetes
- Excludes gestational diabetes

- Requires date of birth

CLINICAL QUALITY MEASURE (implementation)

Hemoglobin A1c Test for Pediatric Patients

List of LOINC codes

Tests for HbA1c

diabetic patients [age 5-17] *tested for HbA1c*

=

diabetic patients [age 5-17]

- Type 1 or Type 2 diabetes
- Excludes gestational diabetes

- Requires date of birth

Data element

List of SNOMED CT or ICD 10 codes

ANATOMY OF A CLINICAL QUALITY MEASURE

Population criteria

- **Initial Patient Population =**
 - AND: "Patient Characteristic Birthdate: birth date" >= 5 year(s) starts before start of "Measurement Period"
 - AND: "Patient Characteristic Birthdate: birth date" <= 17 year(s) starts before start of "Measurement Period"
 - AND: "Diagnosis, Active: Diabetes" starts before or during (MOST RECENT : "Occurrence A of Encounter, Performed: Diabetes Visit" during "Measurement Period")
 - AND: "Encounter, Performed: Diabetes Visit" >= 12 month(s) starts before start of "Occurrence A of Encounter, Performed: Diabetes Visit"
- **Denominator =**
 - AND: "Initial Patient Population"
- **Denominator Exclusions =**
 - AND NOT: "Occurrence A of Diagnosis, Active: Gestational Diabetes" ends before start of "Measurement Period"
 - AND: "Occurrence A of Diagnosis, Active: Gestational Diabetes" starts before or during "Measurement Period"
- **Numerator =**
 - AND: "Laboratory Test, Result: HbA1c Laboratory Test (result)" during "Measurement Period"
- **Denominator Exceptions =**
 - None

Data criteria (QDM Data Elements)

- "Diagnosis, Active: Diabetes" using "Diabetes Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1001)"
- "Diagnosis, Active: Gestational Diabetes" using "Gestational Diabetes Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1010)"
- "Encounter, Performed: Diabetes Visit" using "Diabetes Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1012)"
- "Laboratory Test, Result: HbA1c Laboratory Test" using "HbA1c Laboratory Test Grouping Value Set (2.16.840.1.113883.3.464.1003.198.12.1013)"
- "Patient Characteristic Birthdate: birth date" using "birth date LOINC Value Set (2.16.840.1.113883.3.555.1.1.1)"

Value set = List of
LOINC codes for
HbA1c tests

ASSOCIATED VALUE SET

Metadata

Measure

Grouping

Name:

HbA1c Laboratory Test

OID:

2.16.840.1.113883.3.464.1003.198.12.1013

Type:

Grouping

Developer:

National Committee for Quality Assurance

Note:

Value Set Members Expansion

Expanded Code List

View Toggle Clear

Page 1 of 1 20

View 1 - 3 of 3

Code	Descriptor	Code System	Version
17855-8	Hemoglobin A1c/Hemoglobin.total in Blood by calculation	LOINC	2.40
17856-6	Hemoglobin A1c/Hemoglobin.total in Blood by HPLC	LOINC	2.40
4548-4	Hemoglobin A1c/Hemoglobin.total in Blood	LOINC	2.40





Welcome

Search Value Sets

Help

[Apply Filters](#)

[Clear Filters](#)

Search the NLM Value Set Repository

Query:

Narrow search results by selecting from pull-down menus below:

CMS eMeasure (NQF Number)

Quality Data Model Category

Value Set Developer

Meaningful Use Measures

Search Results

Value Set Details

[Export Search Results \(Excel\)](#)

Matched Value Sets					
Download		View		Page 1 of 2	
				View 1 - 20 of 21	
<input type="checkbox"/>	Name	Type	Code System	Developer	OID
<input type="checkbox"/>	birth date	Extensional	LOINC	NQF	2.16.840.1.113883.3.560.100.4
<input type="checkbox"/>	Carotid Intervention	Grouping	ICD10PCS ICD9CM SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.204
<input type="checkbox"/>	Discharge To Another Hospital	Extensional	SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.87
<input type="checkbox"/>	Discharged to Health Care Facility for Hospice Care	Extensional	SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.207
<input type="checkbox"/>	Discharged to Home for Hospice Care	Extensional	SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.209
<input type="checkbox"/>	Discharged to Rehabilitation Facility	Extensional	SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.132
<input type="checkbox"/>	Emergency Department Visit	Grouping	SNOMEDCT	Lantana	2.16.840.1.113883.3.117.1.7.1.293
<input type="checkbox"/>	Ethnicity	Extensional	CDCREC	CDC NCHS	2.16.840.1.114222.4.11.837
<input type="checkbox"/>	Hemorrhagic Stroke	Grouping	ICD10CM ICD9CM SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.212
<input type="checkbox"/>	Inpatient Encounter	Extensional	SNOMEDCT	Joint Commission	2.16.840.1.113883.3.117.1.7.1.23

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

Common data elements at NIH



- Databases
- Find, Read, Learn
- Explore NLM
- Research at NLM
- NLM for You

Contact NLM

NIH Common Data Element (CDE) Resource Portal

Home | Resource Summaries | Glossary

<http://www.nlm.nih.gov/cde/>

Home

Home

NIH encourages the use of common data elements (CDEs) in clinical research, patient registries, and other human subject research in order to improve data quality and opportunities for comparison and combination of data from multiple studies and with electronic health records. This portal provides access to NIH-supported CDE initiatives and other tools and resources that can assist investigators developing protocols for data collection. [What is a CDE?](#)

NIH CDE Initiatives

Collections of CDEs that have been identified for use in particular NIH-supported research projects or registries after a formal evaluation and selection processes.

Summary Table

Subject Areas

NIH CDE Tools and Resources

Databases and repositories of data elements and case report forms that may assist investigators in identifying and selecting data elements for use in their projects.

Summary Table

Subject Areas

The CDE Resource Portal also includes [Other CDE Resources](#) and [Relevant Standards](#). Descriptions of all four groups can be found in the

Terminology research

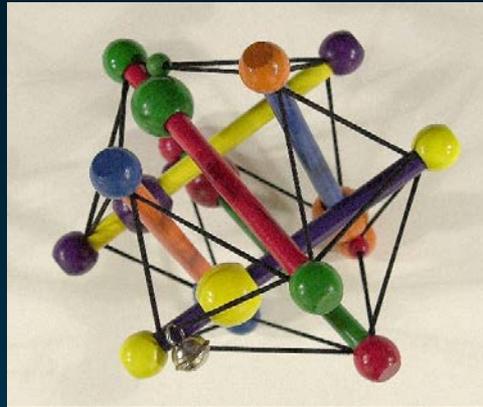
Medical Ontology Research

- ◆ Olivier Bodenreider
- ◆ Ontology alignment
- ◆ Quality assurance of ontologies
- ◆ APIs for drug information sources
- ◆ Interoperability between drug terminologies
 - Drug classes

Applied Medical Terminology Research

- ◆ Kin Wah Fung
- ◆ Terminology derivatives
 - CORE Problem list subset
 - RxTerms
 - Mapping between SNOMED CT and ICD
- ◆ Applications of (drug) terminology
 - Drug-drug interactions
 - Medication history for ER physicians





Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: <http://mor.nlm.nih.gov>



Olivier Bodenreider

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