Ontologies at the NLM

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

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.

Users 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.”
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

## CORE problem list subset

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

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

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
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Other subdomains
- SNOMED CT
- OMIM
- UMLS
- MeSH
- Biomedical literature
- GO
- Genome annotations
- FMA
- Anatomy
- NCBI Taxonomy
- Model organisms
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains
Terminology integration

Addison's disease (363732003)

Clinical repositories

SNOMED CT

UMLS C0001403

OMIM

Biomedical literature

Addison Disease (D000224)

Genetic knowledge bases

NCBI Taxonomy

Model organisms

FMA

Anatomy

Genome annotations

MeSH

Biomedical literature

Addison Disease (D000224)

Genetic knowledge bases

NCBI Taxonomy

Model organisms

FMA

Anatomy

Genome annotations

MeSH

Biomedical literature
Welcome to the UTS

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.

http://rxnav.nlm.nih.gov/
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
Hemoglobin A1c Test for Pediatric Patients

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

= 

# diabetic patients [age 5-17]
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
- Requires date of birth

- Excludes gestational diabetes
Hemoglobin A1c Test for Pediatric Patients

- Type 1 or Type 2 diabetes
- Excludes gestational diabetes
- Requires date of birth

# diabetic patients [age 5-17] \( \text{tested for HbA1c} \)

\[ = \]

# diabetic patients [age 5-17]

List of LOINC codes

List of SNOMED CT or ICD 10 codes

Tests for HbA1c

Data element
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 (2.16.840.1.113883.3.464.1003.198.12.1012)"

Value set = List of LOINC codes for HbA1c tests
**ASSOCIATED VALUE SET**

**Metadata**
- **Name:** HbA1c Laboratory Test
- **Type:** Grouping
- **Note:**
- **OID:** 2.16.840.1.113883.3.464.1003.198.12.1013
- **Developer:** National Committee for Quality Assurance

**Value Set Members**

<table>
<thead>
<tr>
<th>Code</th>
<th>Descriptor</th>
<th>Code System</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>17855-8</td>
<td>Hemoglobin A1c/Hemoglobin.total in Blood by calculation</td>
<td>LOINC</td>
<td>2.40</td>
</tr>
<tr>
<td>17856-6</td>
<td>Hemoglobin A1c/Hemoglobin.total in Blood by HPLC</td>
<td>LOINC</td>
<td>2.40</td>
</tr>
<tr>
<td>4548-4</td>
<td>Hemoglobin A1c/Hemoglobin.total in Blood</td>
<td>LOINC</td>
<td>2.40</td>
</tr>
<tr>
<td>Name</td>
<td>Type</td>
<td>Code System</td>
<td>Developer</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>birth date</td>
<td>Extensonal</td>
<td>LOINC</td>
<td>NQF</td>
</tr>
<tr>
<td>Carotid Intervention</td>
<td>Grouping</td>
<td>ICD10PCS</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Discharge To Another Hospital</td>
<td>Extensonal</td>
<td>SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Discharged to Health Care Facility for Hospice Care</td>
<td>Extensonal</td>
<td>SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Discharged to Home for Hospice Care</td>
<td>Extensonal</td>
<td>SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Discharged to Rehabilitation Facility</td>
<td>Extensonal</td>
<td>SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Emergency Department Visit</td>
<td>Grouping</td>
<td>SNOMEDCT</td>
<td>Lantana</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Extensonal</td>
<td>CDCREC</td>
<td>CDC NCHS</td>
</tr>
<tr>
<td>Hemorrhagic Stroke</td>
<td>Grouping</td>
<td>ICD10CM ICD9CM SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
<tr>
<td>Inpatient Encounter</td>
<td>Extensonal</td>
<td>SNOMEDCT</td>
<td>Joint Commission</td>
</tr>
</tbody>
</table>
Common data elements at NIH


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.

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