Institute of Medical Information and Library
Chinese Academy of Medical Sciences
Beijing, China
August 31, 2017

Standard terminology
research and development
at the U.S. National Library of Medicine

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.
## LHN CBC Research Areas

### Clinical Data Standards & Electronic Medical Records
- Through R&D in standard clinical terminologies and interoperability across clinical information systems and NLM resources, LHN CBC advances user-tailored information retrieval.
- Clinical Vocabulary Standards
- InfoBot
- Medical Ontology Research

### Collaboration Technologies & Mobile Health Applications
- This LHN CBC R&D enables remote collaboration, education, training, and access to NLM information resources and disaster aids anytime, anywhere, and from devices like smart phones.
- People Locator for Disasters
- Remote Virtual Dialog System
- Virtual Microscope

### Document Processing
- LHN CBC conducts R&D in text and data mining, machine learning, electronic preservation and online access for multimedia, print-only, and centuries-old biomedical documents.
- Medical Article Record System
- Interactive Publications
- Turning The Pages

### Health Information Resources
- R&D staff at LHN CBC are developing and enhancing large, complex information systems to meet new needs in health information, biomedical research, and historical preservation.
- Consumer Health Question Answering
- Genetics Home Reference
- Open-i

### Health Information Resources
- Health Information Resources
- R&D staff at LHN CBC are developing and enhancing large, complex information systems to meet new needs in health information, biomedical research, and historical preservation.

### Health Information Resources
- Health Information Resources
- R&D staff at LHN CBC are developing and enhancing large, complex information systems to meet new needs in health information, biomedical research, and historical preservation.

### Health Information Resources
- Health Information Resources
- R&D staff at LHN CBC are developing and enhancing large, complex information systems to meet new needs in health information, biomedical research, and historical preservation.

### Image Processing & Visualization
- For use in biomedical education and the diagnosis and treatment of diseases, LHN CBC conducts R&D in the analysis, presentation, and retrieval of images and the creation of visualizations.
- Computer-aided TB Screening on Chest X-rays
- Imaging Tools for Cancer Research
- Visible Human Project

### Natural Language Processing
- LHN CBC's NLP R&D improves search and retrieval and facilitates discovery through advances in analyzing biomedical texts, graphical presentation of results, and multi-language search.
- Lexical Systems & Tools
- Automated Indexing Research
- Semantic Knowledge Representation

---

### Medical Informatics Training Program

---
Outline

◆ NLM involvement with standard terminologies
◆ Terminology development at NLM – RxNorm
  ● Introduction to RxNorm
  ● RxNorm development
◆ Terminology research at NLM
  ● Examples of terminology-related projects
NLM involvement with standard terminologies
Timeline

- 1960 – First release of MeSH
- 1990 – First release of the UMLS Metathesaurus
- 1999 – NLM begins funding LOINC
- 2002 – First release of RxNorm
- 2003 – NLM negotiates US-wide license for SNOMED CT
- 2007 – The U.S. join the IHTSDO as a founding member
- 2011 – First release of the U.S. Extension of SNOMED CT (U.S. National Release Center)
- 2012 – NLM releases the Value Set Authority Center
Focus

- Controlled vocabularies for indexing and retrieval (PubMed/MEDLINE)
  - Medical Subject Headings (MeSH)
    - Now available in RDF (Semantic Web technologies)
- Standard clinical vocabularies (Meaningful Use incentive program)
  - SNOMED CT
  - LOINC
  - RxNorm
- Derivatives
  - Mapping between SNOMED CT and ICD10-CM
  - Terminology integration: UMLS Metathesaurus
  - Value sets (Value Set Authority Center)
Support

◆ In-house development (data + services)
  ● MeSH
  ● RxNorm
  ● U.S. extension of SNOMED CT
  ● Unified Medical Language System (UMLS)
  ● SNOMED CT—ICD10-CM mapping
  ● Value Set Authority Center

◆ Funding support
  ● SNOMED CT (international release)
  ● LOINC
Terminology development at NLM

Using RxNorm as an example
Introduction to RxNorm
RxNorm

- **Terminology integration system**
  - Structured Product Labels, First DataBank, Micromedex, Multum, MeSH, SNOMED CT, NDF-RT, ATC, ...

- **Scope**
  - Drug names and codes
  - Drugs available on the U.S. market

- **Developer: National Library of Medicine**

- **Publicly available**

- **Monthly updates**

- **Size:** > 10k ingredients; 19k clinical drugs

- **Uses:** e-prescription, information exchange, analytics

https://www.nlm.nih.gov/research/umls/rxnorm/
## Normalization  Lexical level

<table>
<thead>
<tr>
<th>Source</th>
<th>Code</th>
<th>String</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSL</td>
<td>5977</td>
<td>azithromycin 250 mg oral tablet</td>
</tr>
<tr>
<td>RXNORM</td>
<td>308460</td>
<td>Azithromycin 250 MG Oral Tablet</td>
</tr>
<tr>
<td>MTHSPL</td>
<td>54868-5478</td>
<td>AZITHROMYCIN 250 mg ORAL TABLET, FILM COATED</td>
</tr>
<tr>
<td>MMX</td>
<td>124911</td>
<td>Azithromycin 250 MG Oral Tablet_#1</td>
</tr>
<tr>
<td>NDDF</td>
<td>26721</td>
<td>AZITHROMYCIN 250 mg ORAL TABLET_#1</td>
</tr>
<tr>
<td>GS</td>
<td>22681</td>
<td>Azithromycin 250mg Oral tablet_#2</td>
</tr>
<tr>
<td>NDFRT</td>
<td>N0000158080</td>
<td>AZITHROMYCIN 250MG TAB</td>
</tr>
<tr>
<td>MTHSPL</td>
<td>21695-012</td>
<td>AZITHROMYCIN ANHYDROUS 250 mg ORAL TABLET, FILM COATED</td>
</tr>
<tr>
<td>MTHSPL</td>
<td>60505-2581</td>
<td>AZITHROMYCIN DIHYDRATE 250 mg ORAL TABLET, FILM COATED [Azithromycin Dihydrate]</td>
</tr>
<tr>
<td>SNOMEDCT_US</td>
<td>375555002</td>
<td>Azithromycin dihydrate 250mg tablet</td>
</tr>
<tr>
<td>MTHSPL</td>
<td>66116-418</td>
<td>AZITHROMYCIN MONOHYDRATE 250 mg ORAL TABLET, FILM COATED</td>
</tr>
<tr>
<td>MTHSPL</td>
<td>0093-7146</td>
<td>AZITHROMYCIN MONOHYDRATE 250 mg ORAL TABLET, FILM COATED_#1</td>
</tr>
</tbody>
</table>

...  

**308460**  Azithromycin 250 MG Oral Tablet
Normalized form

Strength | Ingredient | Dose form
250 MG | Azithromycin | Oral Tablet

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug
Applications

◆ **RxNav**
  - Drug-centric browser
  - Links among drug entities (graph)
  - Links to other sources of information
    - Drug classes
    - Drug-drug interactions from DrugBank

◆ **RxClass**
  - Drug class-centric browser
    - ATC, NDF-RT, DailyMed (SPL), MeSH
  - All classes for a given drug
  - All drug members for a given class
  - Class-class similarity

https://rxnav.nlm.nih.gov/
Application Programming Interfaces (APIs)

◆ RxNorm
  • Map drug names and codes to RxNorm
    ▪ Including approximate matches and spelling suggestions
  • Navigate among drug entities (e.g., brand to generic)

◆ RxClass
  • Map drug class names and codes to classification systems
    ▪ ATC, NDF-RT, DailyMed (SPL), MeSH
  • Link between drug classes and their drug members
  • Similarity between drug classes

◆ Related APIs
  • RxTerms, NDF-RT, Interactions

◆ Usage
  • 30,000 unique users per month
  • 1B calls in 2015
RxNorm development
Personnel

◆ Content development
  • 1 senior editor
  • 5 editors

◆ Information technology
  • 1 lead developer
  • 3 developers
Editorial principles

◆ Scope
  ◆ Limited to normalized names and codes, mostly for prescription drugs (U.S. market)
  ◆ Over-the-counter (OTC) drugs often included
  ◆ Excluding drug classes, supplies

◆ Collaborative development
  ◆ Collaboration with the developers of drug compendia (First DataBank, Micromedex, Multum, etc.)
  ◆ Mutually beneficial (quality assurance)
Distribution

◆ Relational database files
  ● Similar to UMLS format
  ● Requires UMLS license for download

◆ Applications
  ● RxNav – drug-centric browser
  ● RxClass – class-centric browser

◆ Application programming interfaces (APIs)
  ● For integrating RxNorm in applications
  ● No license required; no proprietary data returned
Releases

◆ 2 types of releases
  ● Monthly release – full release
    ■ New drug added
    ■ Obsolete drugs removed
  ● Weekly release – addition of recently marketed drugs

◆ Fixed dates
  ● First Monday of the month (monthly release)
  ● Each Wednesday (weekly release)

◆ Consistent file names

◆ Synchronization with UMLS
  ● Twice a year
Evolution

◆ Editorial guidelines
  ● Extended identity criteria for a drug
    ◦ Quantity factor
      – 24 HR Nicotine 0.292 MG/HR Transdermal System
    ◦ Quality distinction
      – Sugar-Free Cholestyramine Resin 4000 MG Powder for Oral Suspension
  ● “Prescribable names”
  ● Harmonization with international standards
    ◦ Identification of Medicinal Products (IDMP)
◆ Support for drug classes (through the API)
◆ Support for analytics (through the API)
Challenges

◆ Finding reliable information sources
  ◆ Inconsistencies in the Structured Product Labels
  ◆ Inconsistencies with drug compendia
    ▪ Opportunity for quality assurance at both ends

◆ Supporting multiple use cases
  ◆ E-prescribing
    ▪ Only current drugs, no obsolete drugs
  ◆ Health analytics
    ▪ All drugs, including currently obsolete drugs

◆ Supporting new requirements
  ◆ Abuse-deterrent opioid drugs
  ◆ Sugar-free drug formulations (for diabetic patients)
Terminology research at NLM

Examples of terminology-related projects
Quality assurance in SNOMED CT

Non-lattice subgraph

Suggested remediation

Duodenal ulcer with perforation AND obstruction → Chronic duodenal ulcer with perforation AND obstruction
Coverage of phenotypes

- UMLS
- SNOMED CT
- Consumer Health Vocabulary
- MedDRA
- MeSH
- NCI thesaurus
- ICD-10-CM
- ICD-9-CM
- ICD-10
- OMIM
- MedlinePlus

Legend:
- % HPO concepts covered
- % HPO concepts with Cross-references
Identifying terms for Fetal Medicine

- meta-analysis
- genome
- retrospective study
- cohort study
- risk factors
- invasive testing
- guidelines
- p-values
- obstetricians
- confidence interval
- significant difference
- statistic
- singleton pregnancies
- systematic review
- prenatal testing
- institutional review board
- maternal-fetal medicine
- maternal plasma
- fetal DNA
- recommendation
- clinical practice
- decision making
- fetal aneuploidy
- genetic testing
Suitability of drug classification systems
Evolution of the UMLS Metathesaurus

- Disorders
- Living Beings
- Chemical & Drugs
- Procedures

Categories:
- Physiology
- Phenomena
- Organizations
- Occupations
- Objects
- Geographic Areas
- Genes & Molecular Sequences
- Disorders
- Devices
- Concepts & Ideas
- Living Beings
- Chemicals & Drugs
- Procedures
- Anatomy
- Activities & Behaviors

2002 to 2015
Medical Ontology Research

Contact: olivier@nlm.nih.gov

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