U.S. National Library of Medicine
Lister Hill National Center for Biomedical Communications

Health Data Standards and Open Science Activities at the National Library of Medicine

Kin Wah Fung*, *MD, MS, MA, FRCSEd

Lister Hill National Center for Biomedical Communications
National Library of Medicine, National Institutes of Health

Slide credits: Liz Amos, Steve Emrick, Valerie Florance, Mike Huerta

Dental Public Health Informatics: Opportunities in a Changing Environment

Jan 19-20, 2017 San Antonio, TX

Role of the National Library of Medicine

- NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS).
- NLM works closely with the Office of the National Coordinator for Health Information Technology (ONC) to ensure NLM efforts are aligned with the goal of the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care.

NLM Health Data Standards Activities

- Develop, fund or provide access to clinical terminologies to be used in the electronic health record (EHR) - SNOMED CT, LOINC, RxNorm
- Develop tools to facilitate the use and re-use of terminologies
- Carry out or support research in areas such as clinical vocabulary standards, semantic interoperability and natural language processing
- Provide education and training for medical informatics professionals

What are medical terminologies?

- Terminology: A finite set of terms used to convey information unambiguously in a specified domain
- Terminologies enable computers to
  - “understand” the information entered by human, because free narrative text is too complicated
  - process information efficiently by attaching a symbolic handle (a code) to a unit of meaning (a concept)
- Analogy
  - computer programming needs a set of controlled words that the computer can understand e.g. GOTO, DO…WHILE, FOR…NEXT
- Medical terminologies help to reduce complexity by restricting content (what can be said) and syntax (how it is said)

Free text processing is hard

- Difficulties:
  - Same meaning, different words (redundancy) e.g. oral, per oral, by mouth, P.O, orally…
  - Same word, different meanings (ambiguity) e.g. gunshot in the atrium, COLD, MI, DM
  - Meaning-modifying context e.g. family history of breast cancer
  - Uncertainty and negation e.g. possible pneumonia, myocardial infarction excluded
  - Typos and lexical variants e.g. anemia vs. anaemia
- Natural Language Processing (NLP) – a lot of progress in recent decades, but far from perfect
- Digital medical information is much more useful if it is encoded in a terminology

Benefits of encoded data

- Retrieval - Indexing by codes
  - Fast, complete and accurate data retrieval
- Analysis - Groupings and hierarchies
  - Easy data analysis, aggregation and summarization
- Reasoning – Symbolic representation of meaning
  - Efficient information processing and inferencing
- Sharing - Unambiguous representation of meaning
  - Reliable information sharing

Data retrieval (1)

- Unstructured lab results (e.g. pages of lab reports in pdf):

- Q: Show me the fasting blood glucose results of Mr. X in the past 6 months
- A: ?????

Data retrieval (2)

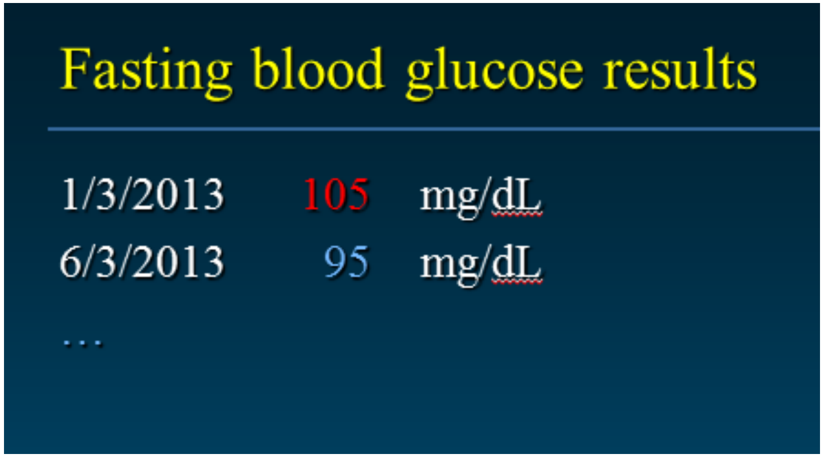
- Structured lab results with encoded information:

Date	Lab test name	Lab test code	result	units
1/3/2013	Blood glucose (fasting)	1558-6	105	mg/dL
4/5/2013	Serum sodium	2947-0	140	mmol/L
5/6/2013	Blood glucose (spot)	2339-0	121	mg/dL
6/3/2013	Blood glucose (fasting)	1558-6	95	mg/dL

LOINC codes

Data retrieval (3)

- Q: Show me the fasting blood glucose results of Mr. X in the past 6 months
- A:



Fasting blood glucose results

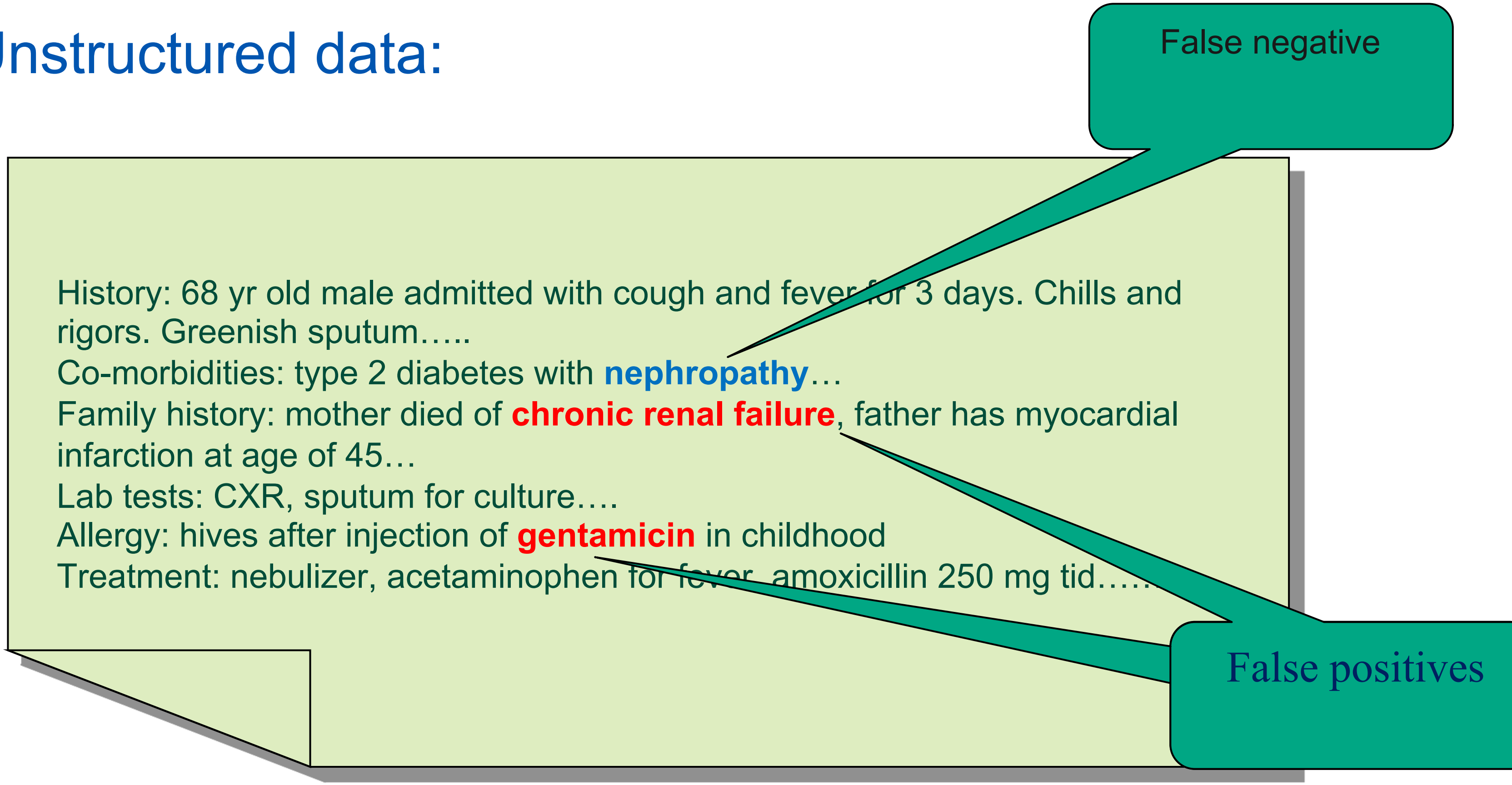
1/3/2013	105	mg/dL
6/3/2013	95	mg/dL
…		

Clinical decision support (1)

- Example: “Prompt physician to adjust dosage of nephrotoxic drugs (drugs that can harm the kidneys) in patients with impaired renal function”
- Computer needs to know
  - Is the patient’s kidney function normal?
    - Diagnosis related to abnormal renal function e.g. acute renal failure, chronic renal failure
    - Abnormal renal function test results
  - Is the doctor prescribing a nephrotoxic drug?
    - Gentamicin, tobramycin…

Clinical decision support (2)

- Unstructured data:



History: 68 yr old male admitted with cough and fever for 3 days. Chills and rigors. Greenish sputum…..
Co-morbidities: type 2 diabetes with **nephropathy**…
Family history: mother died of **chronic renal failure**, father has myocardial infarction at age of 45…
Lab tests: CXR, sputum for culture….
Allergy: hives after injection of **gentamicin** in childhood
Treatment: nebulizer, acetaminophen for fever, amoxicillin 250 mg tid…..

False negative

False positives

Clinical decision support (3)

- Structured encoded data:

EHR section	Item number	Textual entry	Code	Terminology	Test result	Ref. range
Problem list	1	Type 2 diabetes with nephropathy	420279001	SNOMED CT		
Family history (mother)	1	Chronic renal failure	90688005	SNOMED CT		
Family history (father)	2	Acute myocardial infarction	57054005	SNOMED CT		
Allergy list	1	Gentamicin	142438	RxNorm		
Lab results	1	Serum creatinine	2160-0	LOINC	1.8 mg/dL	0.7 – 1.3 mg/dL
prescriptions	1	Amoxicillin 250 mg capsule	308182	RxNorm		

SNOMED CT

- Systematized Nomenclature of Medicine - Clinical Terms
- The most comprehensive clinical healthcare terminology in the world
- An emerging international terminology standard owned by SNOMED International (previously known as IHTSDO – International Health Terminology Standards Development Organisation)
  - 30 member countries (from 9 in 2007) including: US, UK, Canada, Denmark, Spain, Netherlands, Sweden, India, Australia, New Zealand, Hong Kong, Singapore, Malaysia etc.
- SNOMED CT license free for
  - all member countries
  - 40 low income countries (defined by World Bank)
  - qualifying research/humanitarian/charitable projects

Meaningful Use and SNOMED CT

- Centers for Medicare & Medicaid Services (CMS) ‘Meaningful Use’ incentive program for EHR - certification criteria require SNOMED CT to be used for the following data elements:
    - Problems
    - Procedures
    - Smoking status
    - Some laboratory tests results
    - Family health history
    - Cancer registry

Benefits of SNOMED CT

- Comprehensive coverage
  - Over 300,000 concepts covering diseases, symptoms, procedures, body parts, micro-organisms, drugs etc.
  - Better and more granular coverage than ICD codes (ICD-9-CM 14,000 codes; ICD-10-CM 68,000 codes)
  - Coverage can be further extended by combining concepts in a standardized way (post-coordination)
- Clinician friendly terms
  - Rich collection of names and synonyms that are used in clinical discourse (unlike some ICD terms)
- Flexible data entry and retrieval
  - Description logic framework and concept definitions make SNOMED CT more computable than other terminologies

SNODENT

- Developed and maintained by the American Dental Association
- A subset of SNOMED CT
  - About 8,000 concepts
  - Across multiple SNOMED CT hierarchies including clinical finding, procedure, body structure, organism etc.
  - Same structure as SNOMED CT
    - Concepts - permanent unique SNODENT and SNOMED CT identifiers
    - Descriptions – fully-specified names, preferred names and synonyms
    - Relationships – hierarchical and defining relationships
  - Maps to ICD-9-CM, ICD-10-CM and CDT
- Semi-annual update (following each SNOMED CT release)
- Licensed by ADA, no charge

SNODENT and SNOMED CT

- Collaborative agreement between ADA and SNOMED International
  - Harmonization between SNODENT and SNOMED CT
  - Provision of subject matter experts through the International Dentistry Special Interest Group
  - Development of the General Dentistry Subset
    - About 250 concepts – disorders and findings only
    - Will be available with January 2017 release of SNOMED CT
  - Another subset for odontogram under development

Benefits of SNODENT

- Patients and healthcare providers
  - Consistent, accurate and complete information capture during office visits (beyond just procedures)
  - Point of care clinical decision support
  - Improved coordination of care e.g., follow-up reminder
  - Sharing of information between providers
- Public health
  - Electronic reporting and sharing of clinical information
  - High quality demographic and clinical data
  - Identify and monitor oral health issues
- Research and evidence-based care
  - Facilitate research by providing standardized terms for describing dental conditions and identifying patients
  - Analysis of patient care services, quality, cost-effectiveness and outcome

LOINC

- Logical Observation Identifiers Names and Codes
- International standard for tests, measurements and observations
- The Regenstrief Institute for Health Care developed LOINC (first release 1995) under the sponsorship of NLM and other government and private organizations
- Available at no cost
- Designated terminology for laboratory tests and other data elements in Meaningful Use
- Used heavily in communication of clinical data between independent computer systems using, for example, HL7 (Health level Seven) messages

The problem (before LOINC)

Lab test

Test result

- Site 1:

  OBX|1|CE|ABO^ABO GROUP||O^Type O|
- Site 2:

  OBX|1|CE|BLDTYP^ABO GROUP||TYPEO^Type O|
- Site 3:

  OBX|1|CE|ABOTYPE^ABO GROUP||OPOS^Type O|

You and I may know that these are similar results, but our computers will not.

With LOINC

LOINC codes

SNOMED CT codes

- Site 1:

  OBX|1|CE|883-9^ABO GROUP||58460004^Blood Group O|

- Site 2:

  OBX|1|CE|883-9^ABO GROUP||58460004^Blood Group O|

- Site 3:

  OBX|1|CE|883-9^ABO GROUP||58460004^Blood Group O|

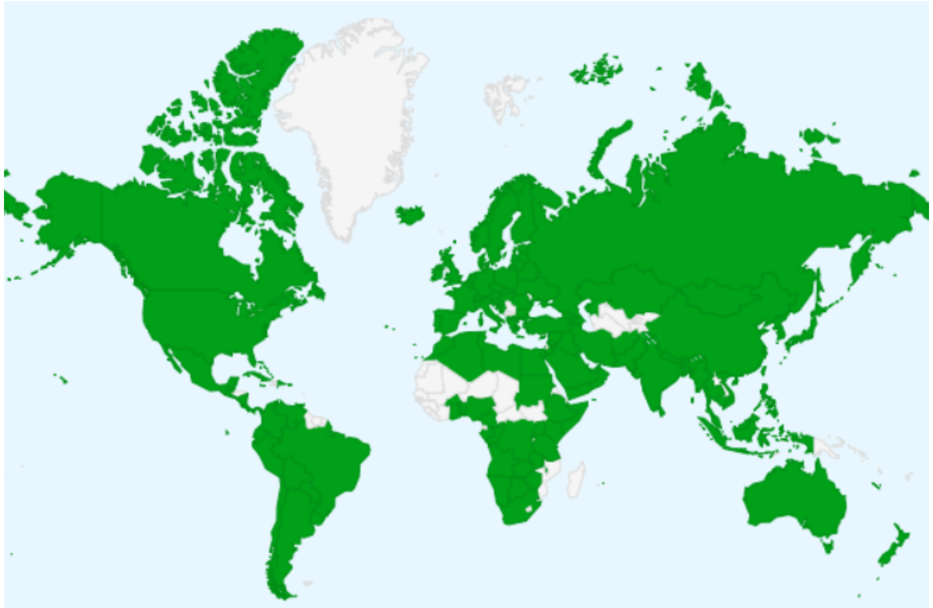
Use LOINC as a universal coding system for clinical observations

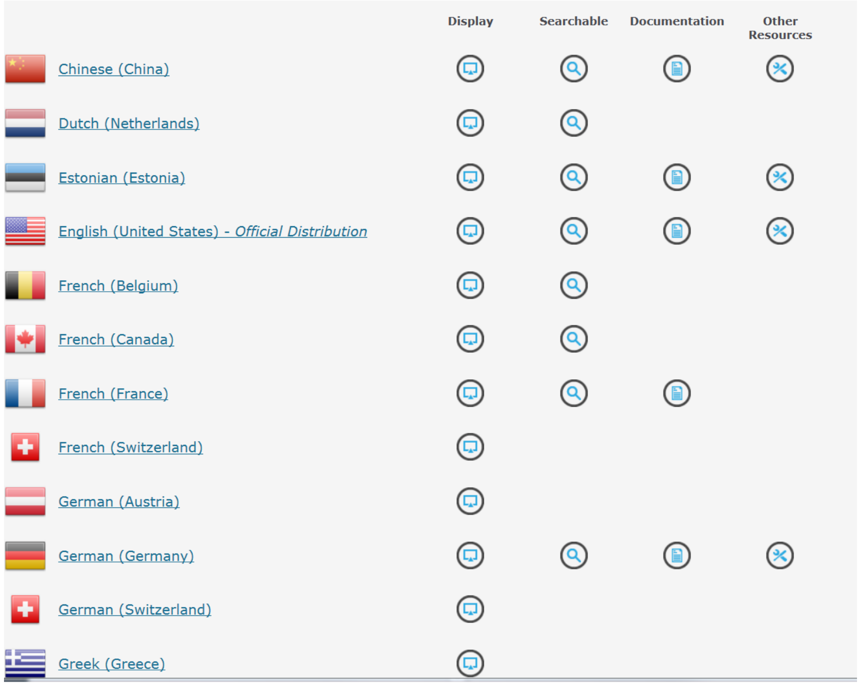
LOINC as question, SNOMED CT as answer (for categorical results)

International use of LOINC

- More than 47,000 users in 175 countries

- Translated to many languages including Spanish, French, German, Russian, Dutch, Chinese etc.





	Display	Searchable	Documentation	Other Resources
Chinese (China)	✓	✓	✓	✓
Dutch (Netherlands)	✓	✓		
Estonian (Estonia)	✓	✓	✓	✓
English (United States) - *Official Distribution*	✓	✓	✓	✓
French (Belgium)	✓	✓		
French (Canada)	✓	✓		
French (France)	✓	✓	✓	
French (Switzerland)	✓			
German (Austria)	✓			
German (Germany)	✓	✓	✓	✓
German (Switzerland)	✓			
Greek (Greece)	✓			

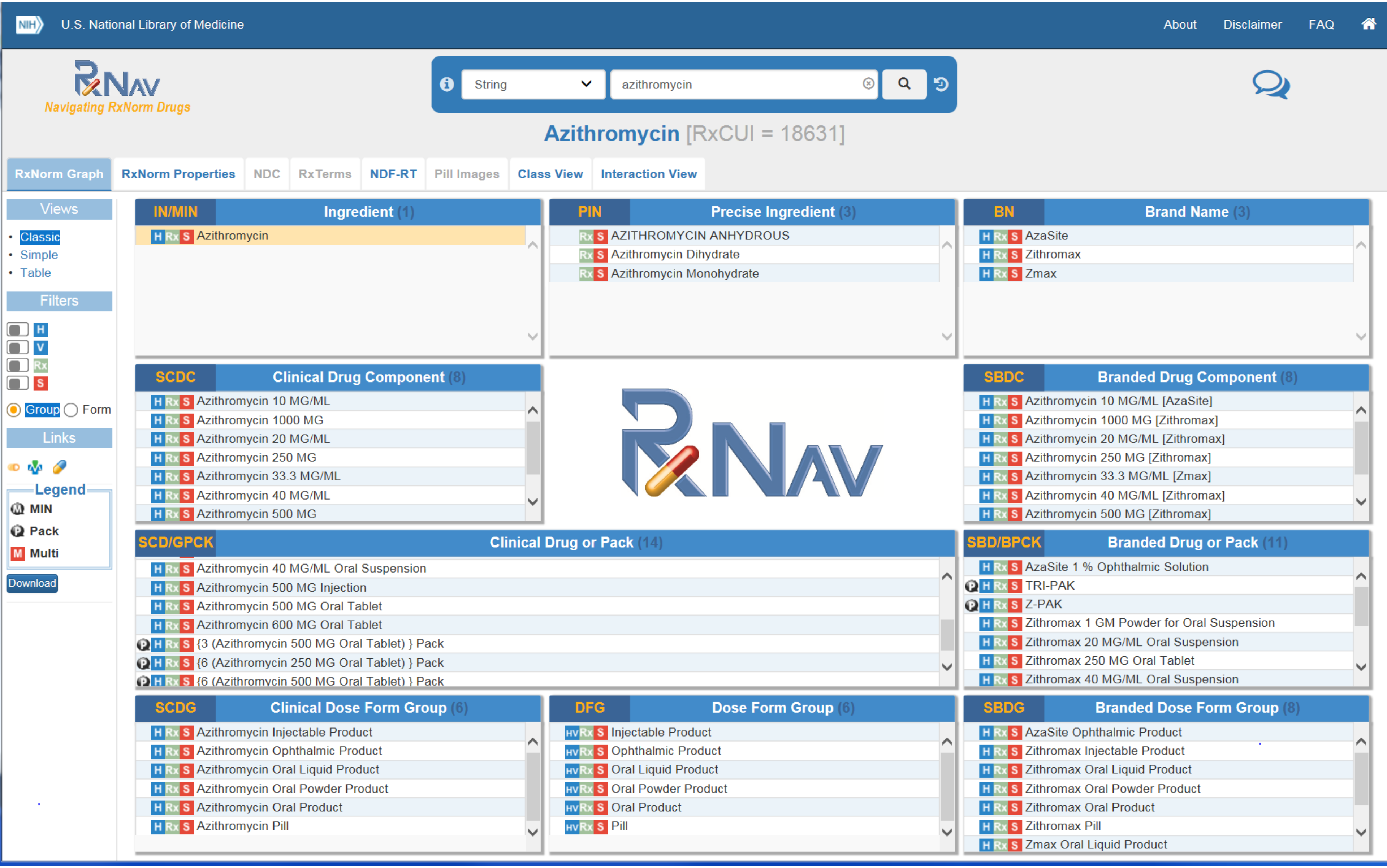
RxNorm

- Created and maintained by NLM
- Two goals
  - Provide normalized naming system for drugs
  - Support semantic interoperation
- Scope
  - Clinical drugs – pharmaceutical products given to a patient with therapeutic or diagnostic intent
  - Excluded: radiopharmaceuticals, contrast media, food, dietary supplements, medical devices

Levels of drug representation in RxNorm

RxCUI	Term Type	Example	Count
18631	Ingredient	Azithromycin	11,221
308460	Semantic clinical drug	Azithromycin 250 MG Oral Tablet	18,824
212446	Semantic branded drug	Azithromycin 250 MG Oral Tablet [Zithromax]	10,566
749783	Generic pack	{6 (Azithromycin 250 MG Oral Tablet) } Pack	373
750149	Branded pack	{6 (Azithromycin 250 MG Oral Tablet [Zithromax]) } Pack [Z-PAK]	430

https://mor.nlm.nih.gov/RxNav/



NIH U.S. National Library of Medicine — About — Disclaimer — FAQ

RxNav *Navigating RxNorm Drugs*

String — azithromycin

Azithromycin [RxCUI = 18631]

RxNorm Graph | RxNorm Properties | NDC | RxTerms | NDF-RT | Pill Images | Class View | Interaction View

Views

- Classic
- Simple
- Table

Filters

H
V
Rx
S

● Group ○ Form

Links

Legend

M MIN
P Pack
M Multi

Download

IN/MIN — Ingredient (1)

H Rx S Azithromycin

PIN — Precise Ingredient (3)

Rx S AZITHROMYCIN ANHYDROUS
Rx S Azithromycin Dihydrate
Rx S Azithromycin Monohydrate

BN — Brand Name (3)

H Rx S AzaSite
H Rx S Zithromax
H Rx S Zmax

SCDC — Clinical Drug Component (8)

H Rx S Azithromycin 10 MG/ML
H Rx S Azithromycin 1000 MG
H Rx S Azithromycin 20 MG/ML
H Rx S Azithromycin 250 MG
H Rx S Azithromycin 33.3 MG/ML
H Rx S Azithromycin 40 MG/ML
H Rx S Azithromycin 500 MG

SBDC — Branded Drug Component (8)

H Rx S Azithromycin 10 MG/ML [AzaSite]
H Rx S Azithromycin 1000 MG [Zithromax]
H Rx S Azithromycin 20 MG/ML [Zithromax]
H Rx S Azithromycin 250 MG [Zithromax]
H Rx S Azithromycin 33.3 MG/ML [Zmax]
H Rx S Azithromycin 40 MG/ML [Zithromax]
H Rx S Azithromycin 500 MG [Zithromax]

SCD/GPCK — Clinical Drug or Pack (14)

H Rx S Azithromycin 40 MG/ML Oral Suspension
H Rx S Azithromycin 500 MG Injection
H Rx S Azithromycin 500 MG Oral Tablet
H Rx S Azithromycin 600 MG Oral Tablet
P H Rx S {3 (Azithromycin 500 MG Oral Tablet) } Pack
P H Rx S {6 (Azithromycin 250 MG Oral Tablet) } Pack
P H Rx S {6 (Azithromycin 500 MG Oral Tablet) } Pack

SBD/BPCK — Branded Drug or Pack (11)

H Rx S AzaSite 1 % Ophthalmic Solution
P H Rx S TRI-PAK
P H Rx S Z-PAK
H Rx S Zithromax 1 GM Powder for Oral Suspension
H Rx S Zithromax 20 MG/ML Oral Suspension
H Rx S Zithromax 250 MG Oral Tablet
H Rx S Zithromax 40 MG/ML Oral Suspension

SCDG — Clinical Dose Form Group (6)

H Rx S Azithromycin Injectable Product
H Rx S Azithromycin Ophthalmic Product
H Rx S Azithromycin Oral Liquid Product
H Rx S Azithromycin Oral Powder Product
H Rx S Azithromycin Oral Product
H Rx S Azithromycin Pill

DFG — Dose Form Group (6)

HV Rx S Injectable Product
HV Rx S Ophthalmic Product
HV Rx S Oral Liquid Product
HV Rx S Oral Powder Product
HV Rx S Oral Product
HV Rx S Pill

SBDG — Branded Dose Form Group (8)

H Rx S AzaSite Ophthalmic Product
H Rx S Zithromax Injectable Product
H Rx S Zithromax Oral Liquid Product
H Rx S Zithromax Oral Powder Product
H Rx S Zithromax Oral Product
H Rx S Zithromax Pill
H Rx S Zmax Oral Liquid Product

Use of RxNorm

- Center for Medicare and Medicaid Services (CMS)
  - Formulary Reference File
  - Clinical Quality Measures (CQMs), Meaningful Use Value Sets
- Office of the National Coordinator (ONC)
  - ‘Meaningful Use’ certification requirement
  - Standard for medications and medication allergies
- National Council for Prescription Drug Programs (NCPDP)
  - SCRIPT e-prescribing standard
  - Formulary and Benefit standard
- Veterans Affairs – Department of Defense (VA-DoD)
  - For integration of drug data among systems

Terminologies are great……BUT

- There are so many of them:
  - Diagnosis and findings - ICD9CM, ICD10, ICD10CM, ICD10AM, ICD-O, ICPC, ICF, SNOMED CT, Read Codes, MedDRA, CTCAE, WHOART, MEDCIN, DSM
  - Procedures - CPT, CDT, HCPCS, OCPS, SNOMED CT, ICD9CM, ICD10-PCS
  - Nursing - NANDA, NIC, NOC, OMS, HHC
  - Diagnostic tests - LOINC, UltraSTAR
  - Drugs - VANDF, NDC, RxNorm, NDDF
  - Medical devices - UMDNS, GMDN, SPN
  - Genomics - GO, HUGO, NCBI Taxonomy
  - …

Why so many?

- New need for information encoding → urge to create a new terminology
- Reasons why existing terminologies are not re-used:
  - Geographical, language
  - Laziness, ignorance
  - Historical/political; ‘not-invented-here’
- Given the diversity of needs, universal terminology unlikely to satisfy all needs

Unified Medical Language System (UMLS)

- NLM terminology resource to provide links between biomedical terminologies by identifying terms that are synonymous in clinical meaning
- Project started in 1986, first release 1990
- Contains over 150 vocabularies in 21 languages
- Over 3 million concepts, 12 million names
- Common uses of the UMLS
  - Terminology research and teaching
  - Access to various terminologies – facilitated by a common data structure
  - Natural language processing – e.g., MetaMap
  - Mapping between terminologies

Metathesaurus: clusters terms by meaning

- Synonymous terms grouped into UMLS concepts
- Preferred term is chosen (default can be changed)
- Unique identifier (CUI) is assigned

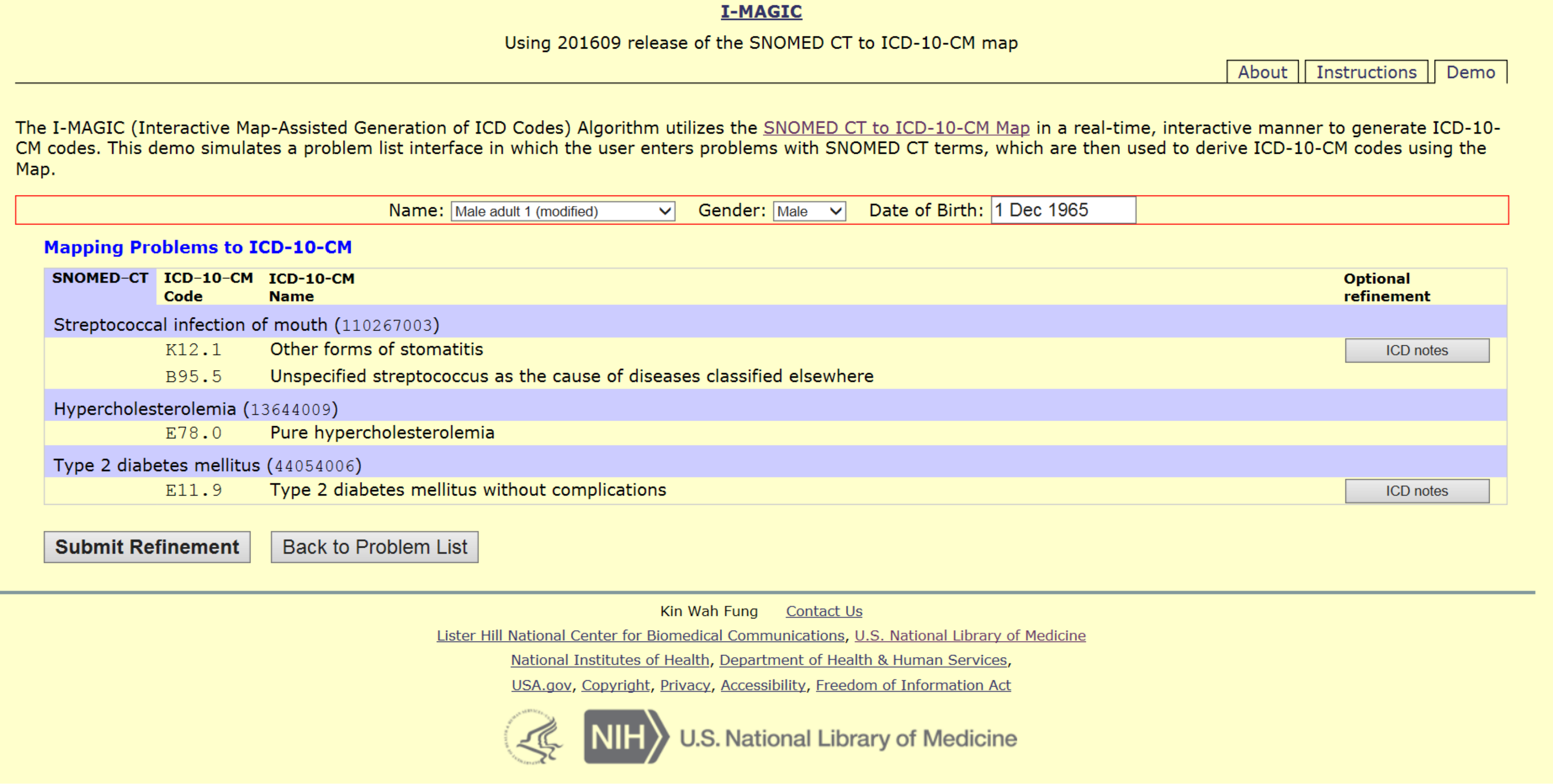
term	source	term type	source ID
Addison disease	Metathesaurus	PN	
Addison’s disease	SNOMED CT	PT	363732003
Addison’s Disease	MedlinePlus	PT	T1233
Addison Disease	MeSH	PT	D000224
Bronzed disease	SNOMED Intl	SY	DB-70620
Primary Adrenal Insufficiency	MeSH	EN	D000224
Primary hypoadrenalism syndrome, Addison	MedDRA	LT	10036696

C0001403 Addison’s disease

Mapping between terminologies

- Purpose of mapping
  - Re-use of encoded data e.g., generation of administrative codes based on clinical data
  - Data integration e.g., data analytics combining clinical and claims data sets
- Examples of maps
  - SNOMED CT to ICD-9-CM – by SNOMED International
  - SNOMED CT to ICD-10-CM – by NLM
  - ICD-9-CM to SNOMED CT - by NLM
  - SNOMED CT to CPT – by American Medical Association

Interactive Map-Assisted Generation of ICD Codes
https://imagic.nlm.nih.gov/



I-MAGIC

Using 201609 release of the SNOMED CT to ICD-10-CM map

The I-MAGIC (Interactive Map-Assisted Generation of ICD Codes) Algorithm utilizes the SNOMED CT to ICD-10-CM Map in a real-time, interactive manner to generate ICD-10-CM codes. This demo simulates a problem list interface in which the user enters problems with SNOMED CT terms, which are then used to derive ICD-10-CM codes using the Map.

Name: Male adult 1 (modified) Gender: Male Date of Birth: 1 Dec 1965

Mapping Problems to ICD-10-CM

SNOMED-CT	ICD-10-CM Code	ICD-10-CM Name	Optional refinement
Streptococcal infection of mouth (110267003)			
	K12.1	Other forms of stomatitis	ICD notes
	B95.5	Unspecified streptococcus as the cause of diseases classified elsewhere	
Hypercholesterolemia (13644009)			
	E78.0	Pure hypercholesterolemia	
Type 2 diabetes mellitus (44054006)			
	E11.9	Type 2 diabetes mellitus without complications	ICD notes

Kin Wah Fung Contact Us

Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine

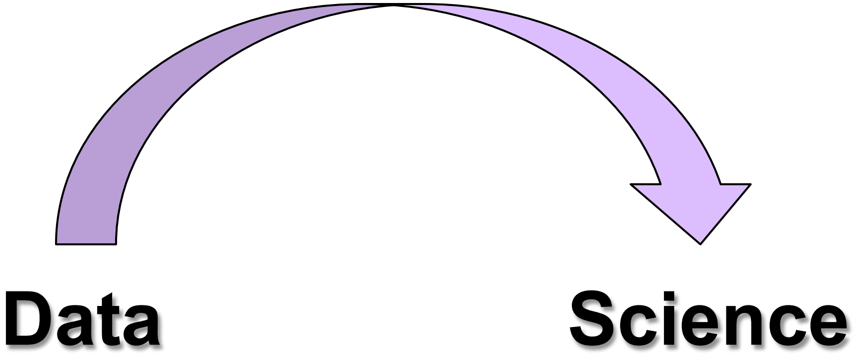
National Institutes of Health, Department of Health & Human Services,

USA.gov, Copyright, Privacy, Accessibility, Freedom of Information Act

Education and Training

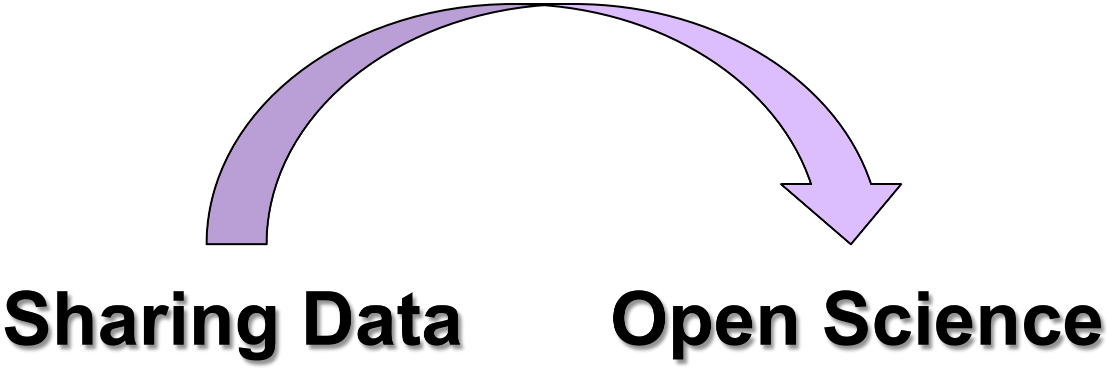
- NLM Georgia Biomedical Informatics Course
  - In partnership with Augusta University
  - Week-long immersive experience
- NLM Medical Informatics Training Program
  - Provides medical informatics and clinical informatics training and research opportunities at various stages of career development
  - Summer, semester training and clinical electives for graduate and medical/dental students
  - Postdoctoral fellowships
  - Visiting scientists
- Research grants
  - Among the 24 NIH institutes and centers that award grants, NLM is the only institute whose grants focus solely on biomedical informatics/information science research

NLM and Open Science



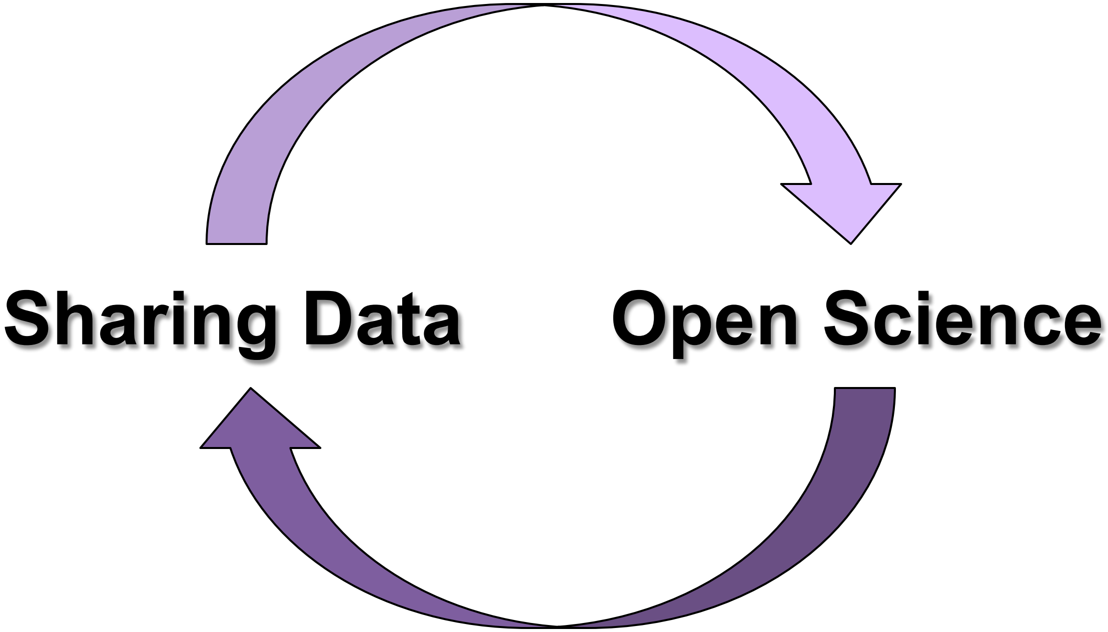
Data

Science



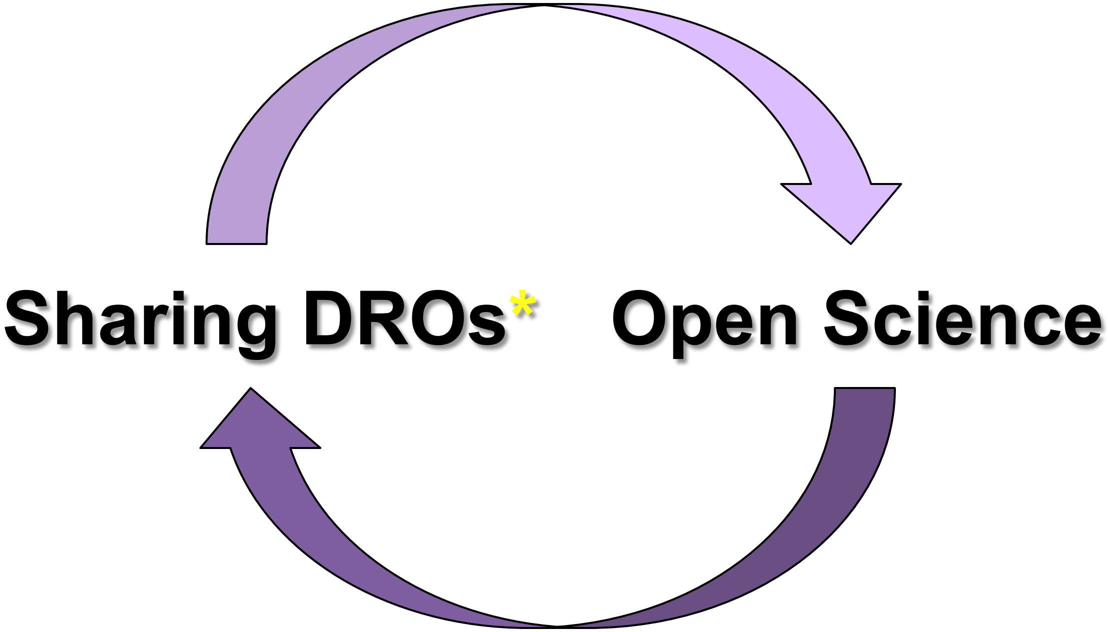
Sharing Data

Open Science



Sharing Data

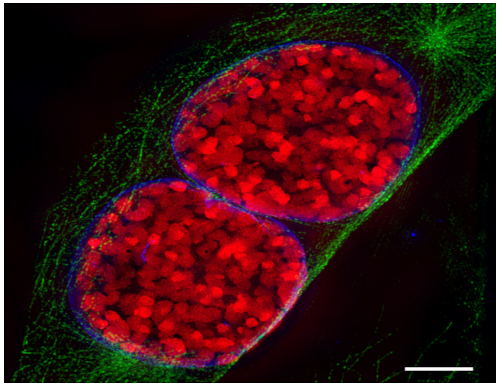
Open Science

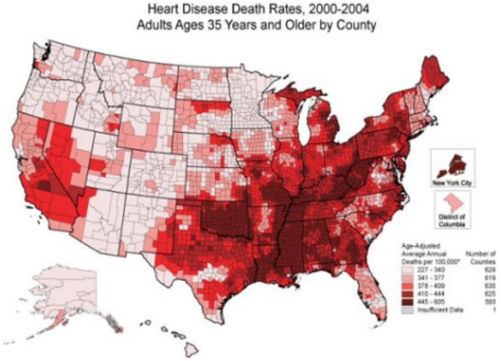


Sharing DROs* Open Science

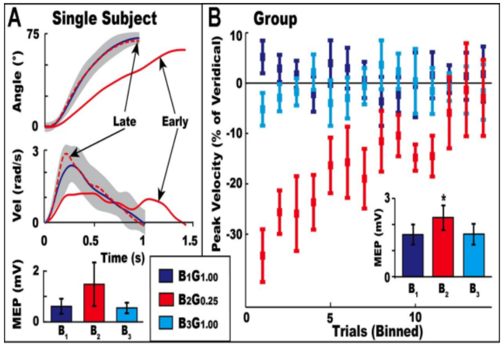
*Digital Research Objects

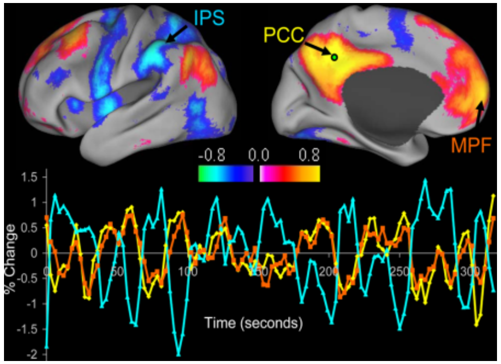
- Data
- Software
- Publications
- Workflows, algorithms etc.





Most domains of biomedicine (except genomics) are **not data-centric nor open**





For these **non-data-centric** domains, the major public products of research are **scientific papers** that describe the authors' **ideas about** the data...

…but the underlying **data are never seen**.

NATIONAL
LIBRARY OF
MEDICINE

…but the underlying data are never seen.

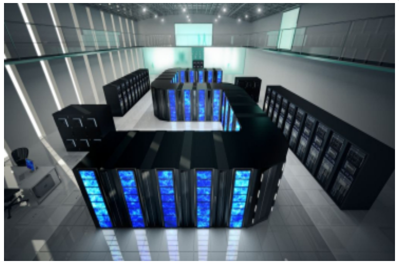
Much less shared

But this is about to change…

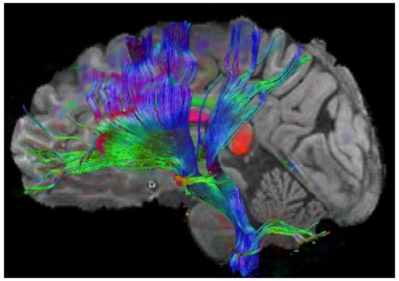
Societal expectations

Policy directives

All Domains will be more Data-centric and Open



Technical capabilities



Scientific opportunities

Sharing Data – Benefits

- Within a study
  - Deep understanding of the scientific paper
  - Reanalysis of data → additional insights
- Across studies
  - Rigorous comparison/analysis across studies
  - Aggregation of data across studies
  - Data mining/big data analysis of heterogeneous data
  - Promotes collaboration
- Efficiency
  - Negative results exposed to community
  - Prevent unnecessary duplication of effort
- Accountability
  - Necessary for assessing reproducibility/replicability
  - Maximize the return on investment

Sharing Data – Objections

- Sharing data costs time and effort
- Sharing data too soon → scooped
- Intellectual property concerns
- Patient privacy & confidentiality
- Scientific credit derives from papers, not data
- Data will not be understood
- Data will be misused

Sharing Data – Solutions

- Sharing data costs time and effort
  - **Funders can support**
- Sharing data too soon → scooped
  - **Policies can include "embargo periods"**
- Intellectual property concerns
  - **Policies can address this**
- Patient privacy & confidentiality
  - **De-identification, consent & other policies**
- Scientific credit derives from papers, not data
  - **Change the incentive structure of science**
- Data will not be understood
  - **Metadata**
- Data will be misused – **Well, OK, bad people will still exist**

How do we share data?

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Scientific Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18.
PMID: 26978244 Free PMC Article

The FAIR principle

- Findable
  - Data and metadata are assigned unique identifiers
  - Data are described with rich metadata
  - Data and metadata in a searchable resource
- Accessible
  - Data and metadata retrievable by standardized protocol
- Interoperable
  - Data and metadata use shared language for knowledge representation
  - Standard terminologies
- Reusable
  - Data described with detailed provenance and clear usage licenses

PubMed Central

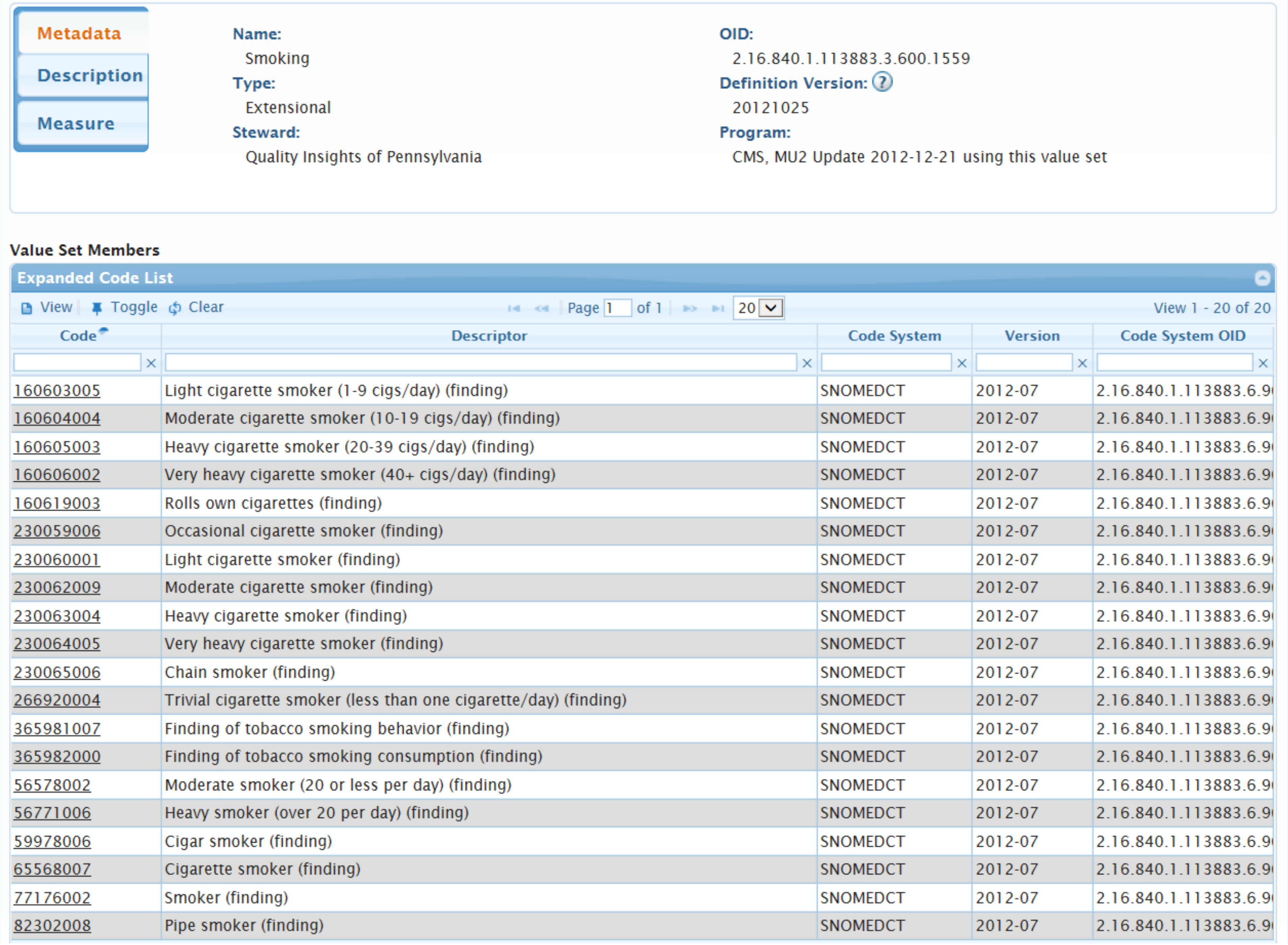
- Provides free access to full-text biomedical and life sciences journal articles
- Started in 2000
- Over 3 million articles
- Linked to PubMed
- Since 2008 - investigators required to submit papers reporting NIH-funded research to PubMed Central within 12 months of publication

ClinicalTrials.gov

- A registry and results database of publicly and privately supported clinical studies of human participants conducted around the world
- First made available to public in 2000
- Results database was first released in 2008
- Currently lists over 200,000 studies (40,000 recruiting) in all 50 states and 195 countries
- Main users: patients, their families and caregivers, health care professionals, clinical researchers, and study record managers
- Over 200 million page views per month

Value Set Authority Center (VSAC)

- Value sets – code sets drawn from standard terminologies to define some clinical characteristics (e.g., smokers, diabetic) to support specific use cases (e.g., cohort definition, clinical quality measure)
- VSAC provides:
  - One-stop shop for value sets supporting various purposes
  - Searchable and downloadable
  - Accessible by API (application programming interface)
  - Authoring and collaboration environment for value set authors



Name:
Smoking

Type:
Extensional

Steward:
Quality Insights of Pennsylvania

OID:
2.16.840.1.113883.3.600.1559

Definition Version:
20121025

Program:
CMS, MU2 Update 2012-12-21 using this value set

Value Set Members

Expanded Code List

Code	Descriptor	Code System	Version	Code System OID
160603005	Light cigarette smoker (1-9 cigs/day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
160604004	Moderate cigarette smoker (10-19 cigs/day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
160605003	Heavy cigarette smoker (20-39 cigs/day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
160606002	Very heavy cigarette smoker (40+ cigs/day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
160619003	Rolls own cigarettes (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230059006	Occasional cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230060001	Light cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230062009	Moderate cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230063004	Heavy cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230064005	Very heavy cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
230065006	Chain smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
266920004	Trivial cigarette smoker (less than one cigarette/day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
365981007	Finding of tobacco smoking behavior (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
365982000	Finding of tobacco smoking consumption (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
56578002	Moderate smoker (20 or less per day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
56771006	Heavy smoker (over 20 per day) (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
59978006	Cigar smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
65568007	Cigarette smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
77176002	Smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9
82302008	Pipe smoker (finding)	SNOMEDCT	2012-07	2.16.840.1.113883.6.9

Value Set Development – until very recently

NIH VSAC *Collaboration*

As an authoring platform, VSAC provides

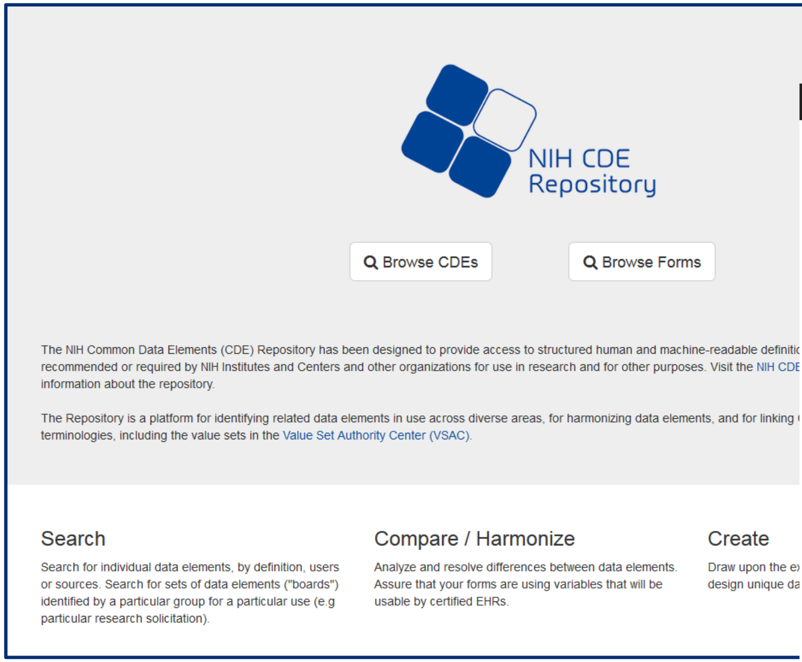
- Terminology support – check for obsolete codes and help to find replacements
- Centralized collaboration environment – identify overlap between value sets, minimize duplication of effort, improve consistency and quality

NIH Common Data Element Repository

- Common data element (CDE) is a representation of a variable (usually a question) common to multiple studies e.g., how often do you consume alcohol?
- Usually the response is a fixed list of values e.g., not at all, some days, everyday
- In the repository, CDEs are defined unambiguously in human and machine readable terms
- Sets of CDEs can be combined into more complex questionnaires, survey instruments, and case report forms

The NIH CDE Repository is a tool to search across CDE initiatives, harmonize differences and create new CDEs

https://cde.nlm.nih.gov



The NIH Common Data Elements (CDE) Repository has been designed to provide access to structured human and machine-readable definitic recommended or required by NIH Institutes and Centers and other organizations for use in research and for other purposes. Visit the NIH CDE information about the repository.

The Repository is a platform for identifying related data elements in use across diverse areas, for harmonizing data elements, and for linking terminologies, including the value sets in the Value Set Authority Center (VSAC).

Search

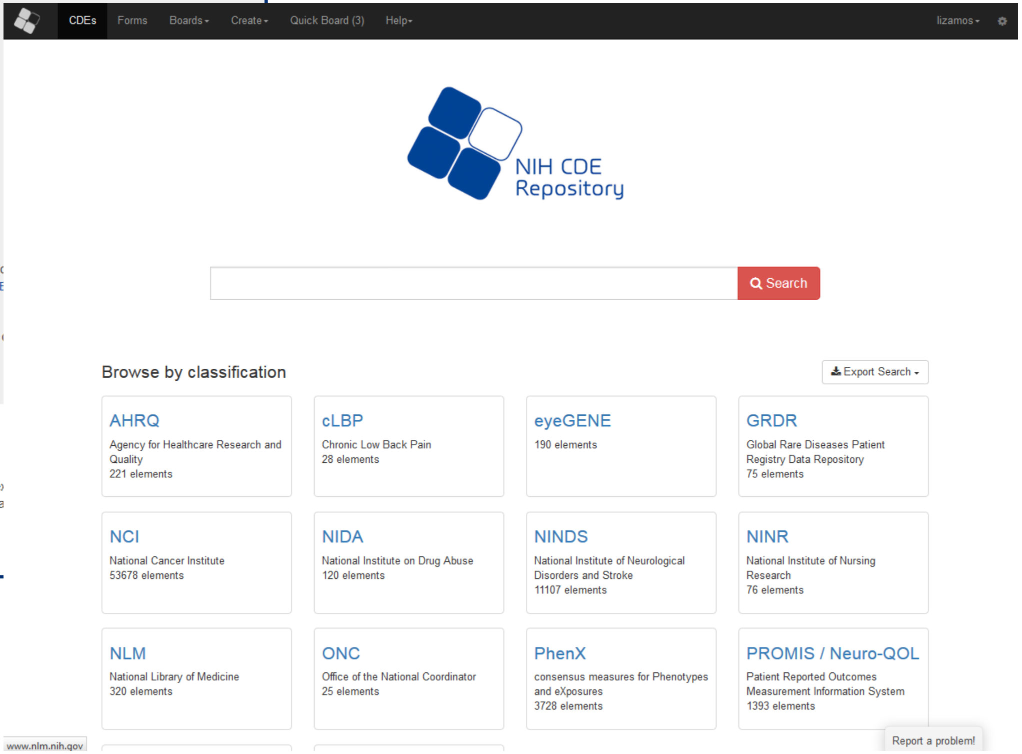
Search for individual data elements, by definition, users or sources. Search for sets of data elements ("boards") identified by a particular group for a particular use (e.g particular research solicitation).

Compare / Harmonize

Analyze and resolve differences between data elements. Assure that your forms are using variables that will be usable by certified EHRs.

Create

Draw upon the ex design unique da



Browse by classification

AHRQ Agency for Healthcare Research and Quality 221 elements	cLBP Chronic Low Back Pain 28 elements	eyeGENE 190 elements	GRDR Global Rare Diseases Patient Registry Data Repository 75 elements
NCI National Cancer Institute 53678 elements	NIDA National Institute on Drug Abuse 120 elements	NINDS National Institute of Neurological Disorders and Stroke 11107 elements	NINR National Institute of Nursing Research 76 elements
NLM National Library of Medicine 320 elements	ONC Office of the National Coordinator 25 elements	PhenX consensus measures for Phenotypes and eXposures 3728 elements	PROMIS / Neuro-QOL Patient Reported Outcomes Measurement Information System 1393 elements

Benefits of re-using CDEs

- Consistent data collection of core set of variables from different sources (sites, projects, initiatives) to allow:
  - Aggregation of data to increase statistical power
  - Rigorous comparison of data & results
- Can be used to promote research:
  - Efficiency – off-the-shelf data elements
  - Quality – validated instruments & measures
  - Clarity – unambiguously defined data elements
  - Reproducibility – from rigorous comparison

NLM’s role in Open Science and Data Science

- NIH Advisory Committee to the Director NLM Working Group report 2015 - NLM should
  - be the intellectual and programmatic epicenter for data science at NIH and stimulate its advancement throughout biomedical research and application
  - lead efforts to support and catalyze open science, data sharing, and research reproducibility
- NLM Director Patricia Brennan appointed as NIH Interim Associate Director for Data Science
- NLM Strategic Planning Activity (Sept 2016 – Dec 2017)
  - 4 themes
    - Advancing data science, open science, & informatics
    - Advancing biomedical discovery & translation
    - Supporting the public’s health: clinical systems, public health systems and services, & personal health
    - Building 21st Century collections for discovery & health

U.S. PUBLIC HEALTH SERVICE
1798
Center Drive
Library of Medicine

Thank you!