

Clinical Terminologies

Jessie Tenenbaum, PhD

Jessie.Tenenbaum@duke.edu

@jessiet1023

Natural and Political
OBSERVATIONS

Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

1910
1665
245

BY
Capt. **JOHN GRANT**,
Fellow of the *Royal Society*.

With reference to the *Government, Religion, Trade, Growth, Air, Diseases*, and the several Changes of the said **CITY**.

— *Non, me ut miretur Turba, laboro,*
Contentus paucis Lectoribus. —

The Third EDITION,
much Enlarged.

LONDON,
Printed by *John Martyn*, and *James Allestry*,
Printers to the *Royal Society*, and are to be sold at the
sign of the *Bell* in *St. Pauls Church-yard*.
MDC LX V.

The Table of CASUALTIES.

The Years of our Lord	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664
Abortive and Stil-born	335	329	327	351	389	381	384	433	483	419	463	467	421	544	499	439	410	4
Aged	916	835	889	696	780	834	864	974	743	892	869	1176	909	1095	579	712	661	6
Ague and Fever	1260	884	751	970	1038	1212	282	1371	689	875	999	1800	2303	2148	956	1091	1115	11
Apoplex and Suddenly	68	74	64	74	106	111	118	86	92	102	113	138	91	67	22	36		
Bleach			1	3	7	2				1								
Blasted	4	1			6	6			4		5	5	3	8	13	8	10	
Bleeding	3	2	5	1	3	4	3	2	7	3	5	4	7	2	5	2	5	
Bloody Flux, Scouring and Flux	155	176	802	289	833	762	200	386	168	368	362	233	346	251	449	438	352	3
Burnt and Scalded	3	6	10	5	11	8	5	7	10	5	7	4	6	6	3	10	7	
Calenture	1			1		2	1	1			3							
Cancer, Gangrene and Fistula	26	29	31	19	31	53	36	37	73	31	24	35	63	52	20	14	23	
Wolf				8														
Canker, Sore-mouth and Thrush	66	28	54	42	68	51	53	72	44	81	19	27	73	68	6	4	4	
Child-bed	161	106	114	117	206	213	158	192	177	201	236	225	226	194	150	157	112	1
Chrisoms and Infants	1369	1254	1065	990	1237	1280	1050	1343	1089	1393	1162	1144	858	1123	2596	2378	2035	22
Colick and Wind	103	71	85	82	76	102	80	101	85	120	113	179	116	167	48	57		
Cold and Cough							41	36	21	58	30	31	33	24	10	58	51	
Consumption and Cough	2423	2200	2388	1988	2350	2410	2286	2868	2606	3184	2757	3610	2982	3414	1827	1910	1713	17
Convulsion	684	491	530	493	569	653	600	828	702	1027	807	841	742	1031	52	87	18	2
Cramp			1														1	
Cut of the Stone		2	1	3		1	1	2	4	1	3	5	6	4				
Dropie and Tympany	185	434	421	508	444	556	617	704	660	706	631	911	646	872	235	252	279	21
Drowned	47	40	30	27	49	50	53	30	43	49	63	60	57	48	43	33	29	
Excessive drinking			2															
Executed	8	17	29	43	24	12	19	21	19	22	20	18	7	18	19	13	12	
Fainted in a Bath					1													
Falling-Sickness	3	2	2	3		3	4	1	4	3	1		4	5	3	10	7	
Flox and small Pox	139	400	1190	184	525	1279	139	812	1294	823	835	409	1523	354	72	40	58	5
Found dead in the Streets	6	6	9	8	7	9	14	4	3	4	9	11	2	6	18	33	20	
French-Pox	18	29	15	18	21	20	20	20	29	23	25	53	51	31	17	12	12	
Frighted	4	4	1		3		2		1	1			9	1				

Uses of Medical Terminologies

- Data encoding – for storage, retrieval and presentation e.g. in EMR
- Data sharing e.g. HL7 messages
- Information indexing and retrieval e.g. MedLine, health information websites
- Natural language processing
- Knowledge representation e.g. gene annotation by Gene Ontology

More fun with standards: Clinical terminologies

- ICD9/10- diagnoses for billing
 - LOINC- labs
 - SNOMED- all things biomedical
 - RxNorm- drugs
-
- Unified Medical Language System (UMLS) & tools

“The Desiderata”

- **Desiderata for Controlled Medical Vocabularies in the Twenty-First Century**, JJ Cimino, *Methods Inf Med*. 1998 Nov;37(4-5):394-403.



The *Desiderata*

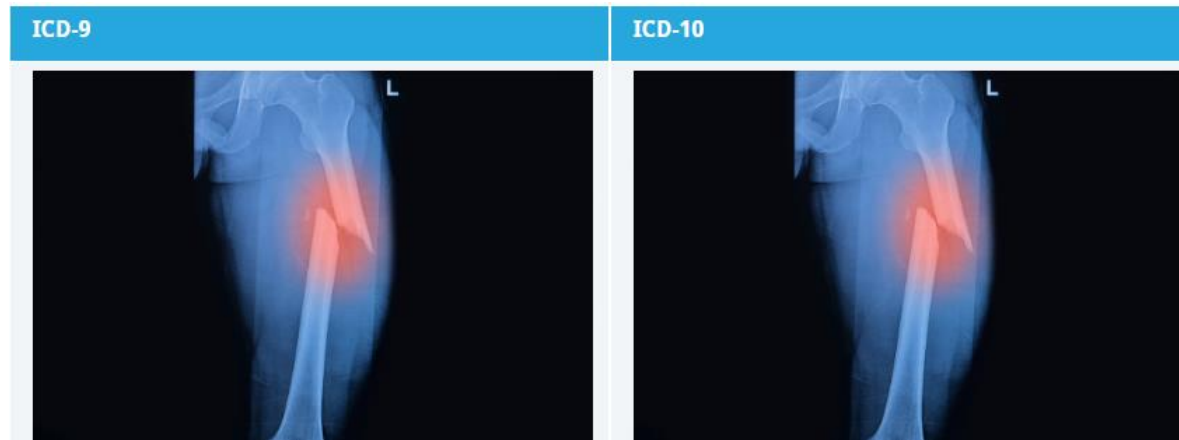
- **Coverage**
- **Concept Orientation**
 - Non-vague
 - Non-overlapping
 - Non-ambiguous
- **Concept permanence**
- **Unique, Stable, Dumb Identifiers**
- **Polyhierarchy**
- **Formal definitions**
- **Reject NEC!**
- **Evolve gracefully**
- **Recognize redundancy**
-

ICD9/10

- International Classification of Diseases, overseen by World Health Organization
- Translated into 43 languages
- For tracking morbidity & mortality
- Billing and reimbursement!
- Versions
 - ICD9 adopted in US in 1979, finally transitioned to ICD10 last year
 - Most other countries have been on ICD10 for years
 - ICD11 currently under development- to be completed in 2018

What's new in ICD10?

- Almost 4x as many diagnosis codes
- Documents laterality- right vs. left broken pinky toe
- Extensive severity parameters and combination codes



821.11 Open fracture of Shaft of Femur

All codes for femur fracture = 16

S72.352C Displaced comminuted fracture of shaft of left femur, initial encounter for open fracture type IIIA, IIIB, or IIIC

All codes for femur fracture = 1530

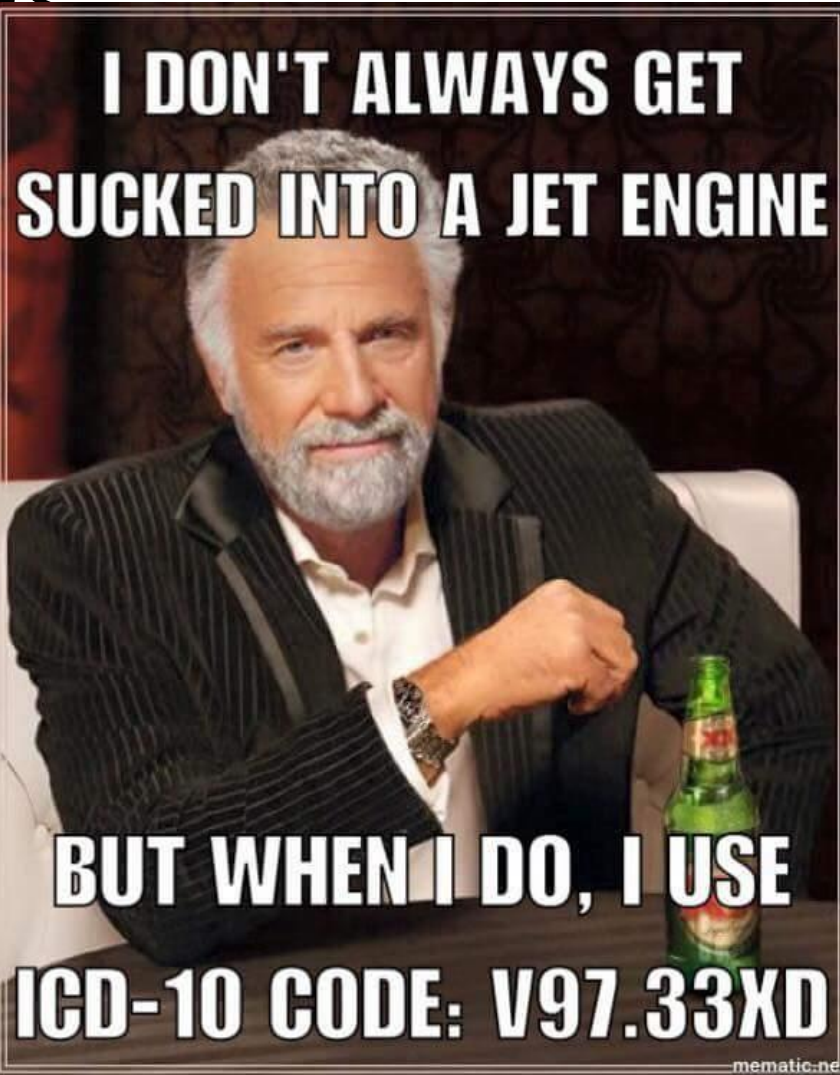
Why the changes?

- Specialties advocated capturing additional detail based on information used in delivering patient care
- Accurate analysis of health data will help improve the quality and efficiency of delivering patient care
- Great for people who want to analyze health data...
 - Less so for people who need to record it



ICD10 funnies

- V97.33XD: Sucked
- S10.87XA: Other
- Y93.D: V91.07XD:
- W22.02XD: V95.4



t encounter.
cified part of neck,
fire, subsequent
uring

LOINC[®] - Labs standard

from Regenstrief

Logical Observation Identifiers Names and Codes

The Problem

You want to exchange health data, like labs, vitals, but local systems have different ways of identifying the same test or measurement.

The Solution

Map your local test codes to a universal standard that every system can understand.

The Result

Systems that actually recognize and aggregate data from other institutions. [Semantic interoperability](#).

A rich trove of 83,000+ standardized variables

Genetics



Lifestyle



Lab and clinical



Environmental

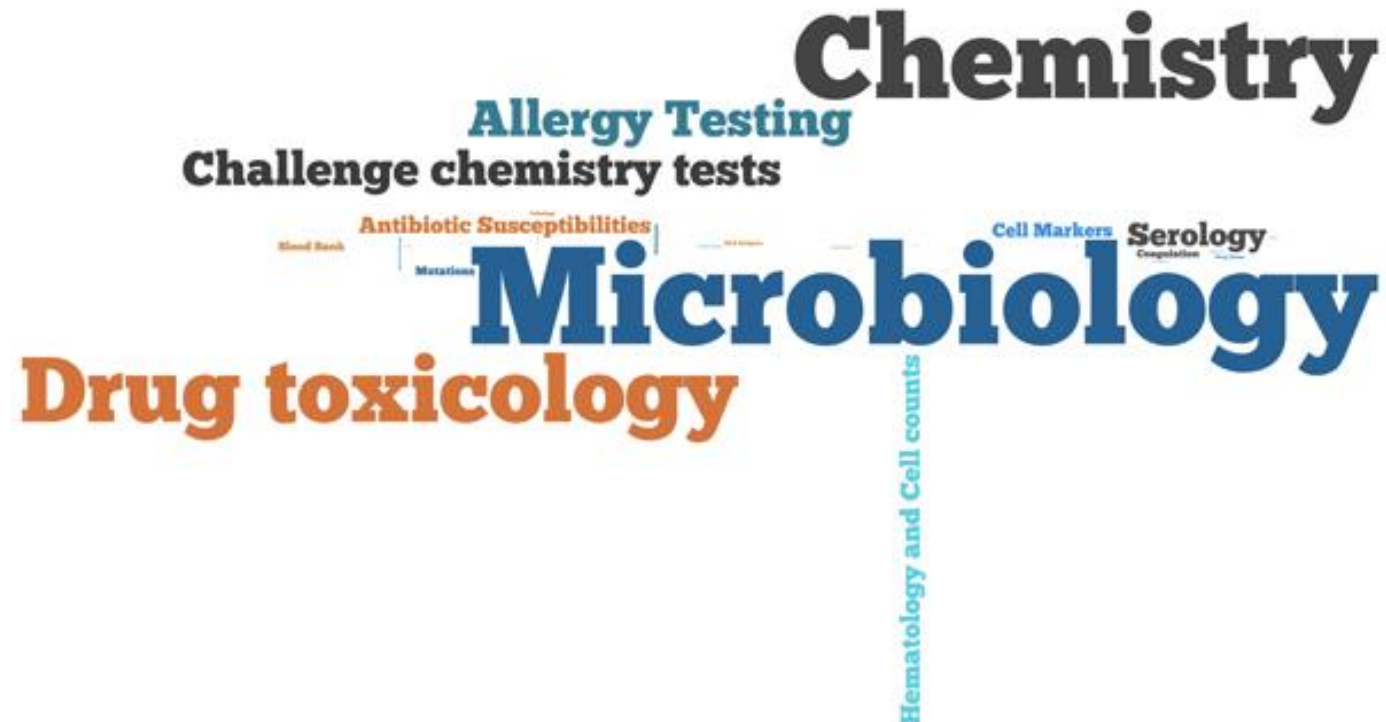


But what *exactly* does it do?

- *If an observation is a question and the observation value is an answer...*
- LOINC provides codes for questions
- Where needed, other vocabularies provide codes for answers

Laboratory

- Anything you can test, measure, or observe about a specimen.
- Chemistry, hematology, serology, microbiology, toxicology
- Cell counts, antibiotic susceptibilities, etc.



Clinical

- Everything except lab- anything that you can test, measure, or observe about a patient without removing a specimen.
- Vital signs, hemodynamics, intake/output, EKG, obstetric ultrasound, cardiac echo, urologic imaging, gastroendoscopic procedures,, radiology studies, selected survey instruments, etc.



LOINC Example :

Manual count of white blood cells in cerebral spinal fluid specimen



COMPONENT (ANALYTE)



PROPERTY



TIME



SYSTEM (SPECIMEN)



SCALE



METHOD

LOINC Term Components

- 1) **Component (analyte)** - e.g., potassium, hemoglobin, hepatitis C antigen.
- 2) **Property measured** - e.g., a mass concentration, enzyme activity (catalytic rate).
- 3) **Timing** - i.e., whether the measurement is an observation at a moment of time, or an observation integrated over an extended duration of time — e.g., 24-hour urine.
- 4) **The type of sample** - e.g., urine; blood.
- 5) **The type of scale** - e.g., whether the measurement is quantitative, ordinal, nominal, or narrative.
- 6) The **method** used to produce the result or other observation.

Data type of result (OBX-5) is a coded element

This code is from LOINC

This code is from SNOMED

OBX|CE|57131-5^Newborn conditions with pos markers^LN|7573000^PKU^SCT

Code identifying this observation
(what are these results?
Conditions identified by
newborn screening)

Code identifying the result
(Phenylketonuria)

Coded Result Values

CE (coded element)
means the answer
will be coded

Observation Identifier:
this is a blood culture

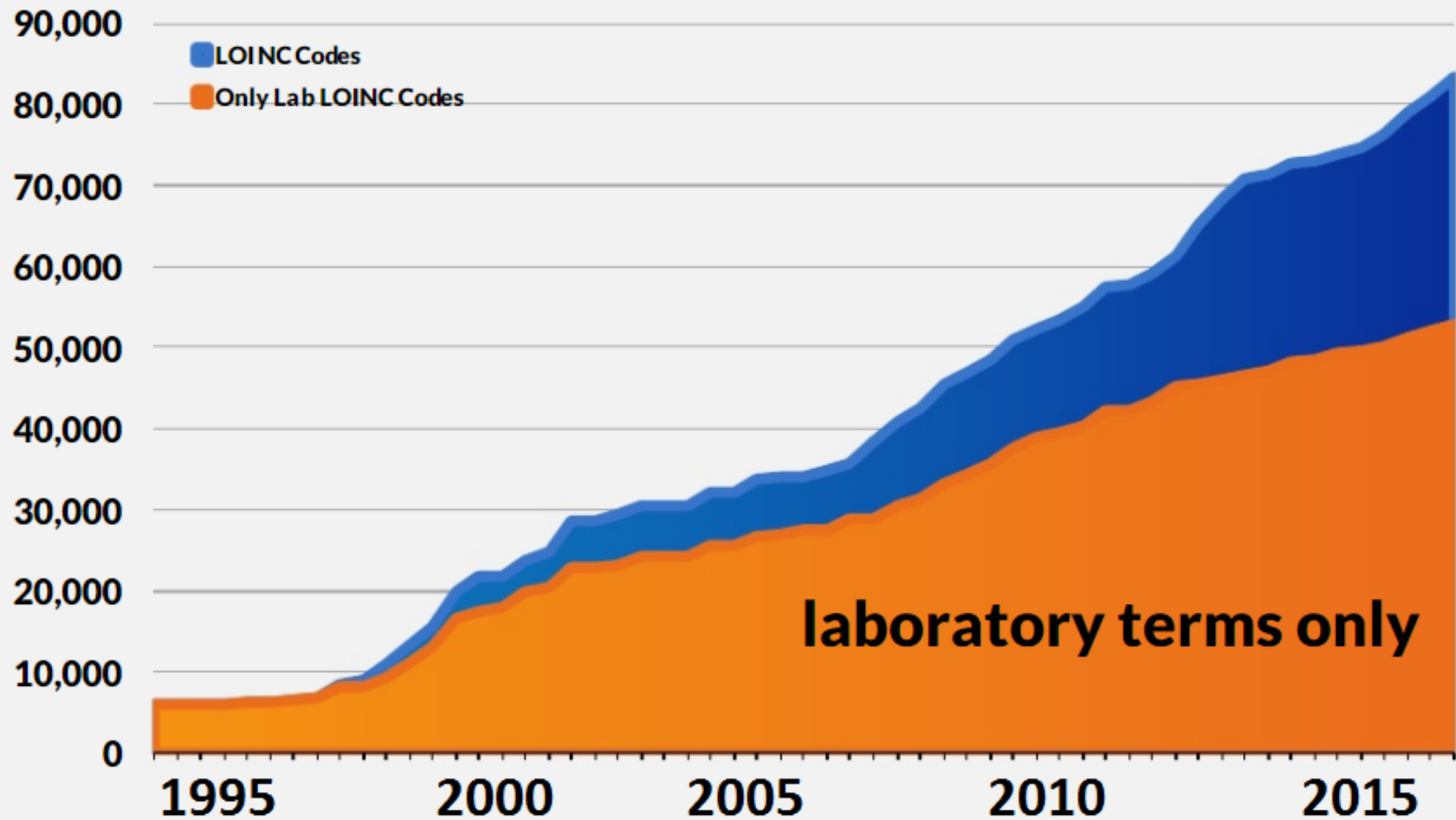
LN means this code
is from LOINC

```
OBX||CE|600-7^Bacteria identified in Blood by Culture^LN||  
17872004^Neisseria meningitidis^SCT
```

Answer Identifier:
meningococcus (trouble)

SCT means this
code is from
SNOMED CT

LOINC Codes Over Time by Release



Clinical Pharmacogenetics Implementation Consortium (CPIC)

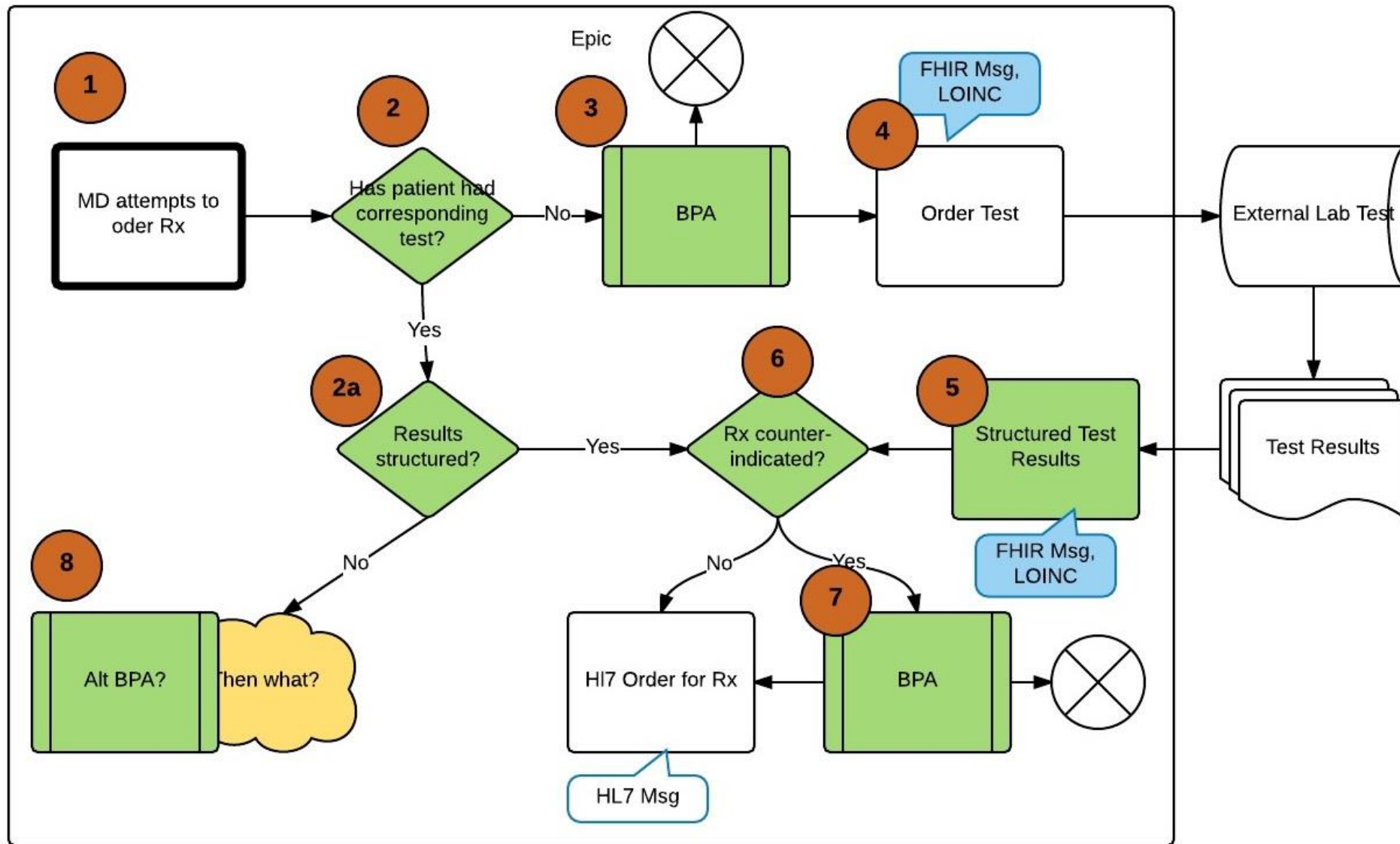
- Goal: address barriers to implementation of PGX tests into clinical practice
- Collaboration between PharmGKB and Pharmacogenomics Research Network
- CPIC guidelines are peer-reviewed, published
 - Guide for how to use genetic test results, not WHETHER tests should be ordered
 - Key assumption is that clinical high-throughput genotyping will become more widespread



DIGITizE: Displaying and Integrating Genetic Information Through the EHR

- National Academy of Medicine (formerly IOM) Roundtable Initiative
- Issue: EHRs are not good at storing structured genomic data
- Pilot project: implement minimal infrastructure needed for clinical decision support for 2 drug/genetic test combinations
 - Abacavir requires HLA B 5701 test
 - Azathioprine requires TMPT test
- NOT storing genetic data directly
 - only phenotypic results of test- e.g. “has mutation”





LOINC and SNOMED enable clinical decision support

LOINC CD	Component	Long Common Name		
50956-2	HLA-B*57:01	HLA-B*57:01 [Presence]		
Part Definition/Description(s)				
Part of HLA-B57 allele family that is associated with Abacavir hypersensitivity reaction (AHSR)				
Answer List*				
	Seq #	Answer	LOINC Answer ID	SNOMED Code
	1	Positive	LA6576-8	10828004
	2	Negative	LA6577-6	260385009

SNOMED CT

The global
language of
healthcare

- Created 1999 by merger of 2 existing terminologies- one focused on specialties, one on general practice
- Owned and distributed by The International Health Terminology Standards Development Organization (IHTSDO)
- “The most comprehensive and precise clinical health terminology product in the world”- IHTSDO
- Not just diagnoses- all different types of terms

SNOMED CT DESIGN

ROOT CONCEPT

Concepts

Hierarchies

Attributes

Identifiers

Descriptions

Relationships

SNOMED CT

SNOMED CT HIERARCHIES

Concepts are organized into top-level hierarchies

- Body structure
- Clinical finding
- Environment or geographical location
- Event
- Linkage concept
- Observable entity
- Organism
- Pharmaceutical / biologic product
- Physical force
- Physical object
- Procedure
- Qualifier value
- Record artifact
- Situation with explicit context
- Social context
- Special concept
- Specimen
- Staging and scales
- Substance

Is a

Low granularity

Finding by site

Is a

Musculoskeletal finding

Is a

Joint finding

Is a

Arthropathy

Is a

Arthropathy of knee joint

Is a

Arthritis of knee

High granularity

Heart failure (disorder)

84114007

Weak heart

Cardiac failure

HF – Heart failure

Myocardial failure

RELATIONSHIPS

Is a relationships connect concepts in a hierarchy

Arthropathy

Is a

Joint finding

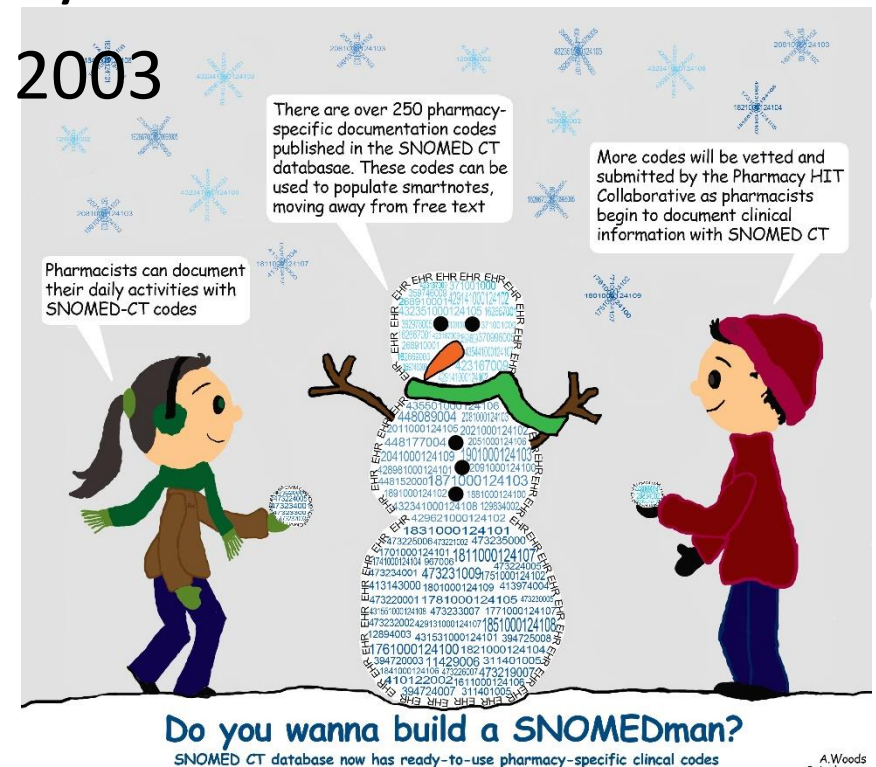
Appendicitis

Associated morphology

Inflammation

SNOMED-CT Cont.

- Includes concept, description, relationships, plus reference sets (groupings)
- Particularly useful for research, secondary analysis
- Available for free for research in the US since 2003
- [Browse SNOMED](#)



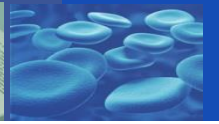
- >100,000 concepts for clinical care, translational and basic research, and public information and administrative activities.
- Focused more on research than SNOMED-CT (more clinical)
- Much smaller than SNOMED-CT
- Cancer-centric, but not exclusive
- Different from NCI Metathesaurus, a subset of UMLS Metathesaurus (more on that later...)

RxNorm- Drugs



- Normalized names for clinical drugs
- Links to many drug vocabularies commonly used in pharmacy management and drug interaction software
 - Can mediate messages between systems not using the same software and vocabulary
- Now includes NDF-RT from the Veterans Health Administration.
 - National Drug File - Reference Terminology
 - Used to code clinical drug properties, including mechanism of action, physiologic effect, and therapeutic category


RxNav Demo



Project: Converting free text medication data to structured terms



Participant reported medications



For office use only
Participant ID: _____
Reviewer Staff Initials: ____
Site ID where reviewed: _____
Date Reviewed: ____/____/____
mm dd yyyy

Concomitant Medication Tracking Form

Please list any pharmaceutical and/or natural medications (including vitamins) that you are currently taking.

Not currently taking any pharmaceutical or natural medications and/or vitamins.

Name of Medication	Reason for use (For Office Use Only)
<i>Prozac</i>	<i>depression</i>
<i>Multi-vitamin</i>	<i>couldn't hurt</i>





Drug mapping objective

- Map free text to coded terminology to enable querying by medication
- **Query by drug name (handle variations and misspellings)**
- **Query by category (e.g. all SSRIs)**





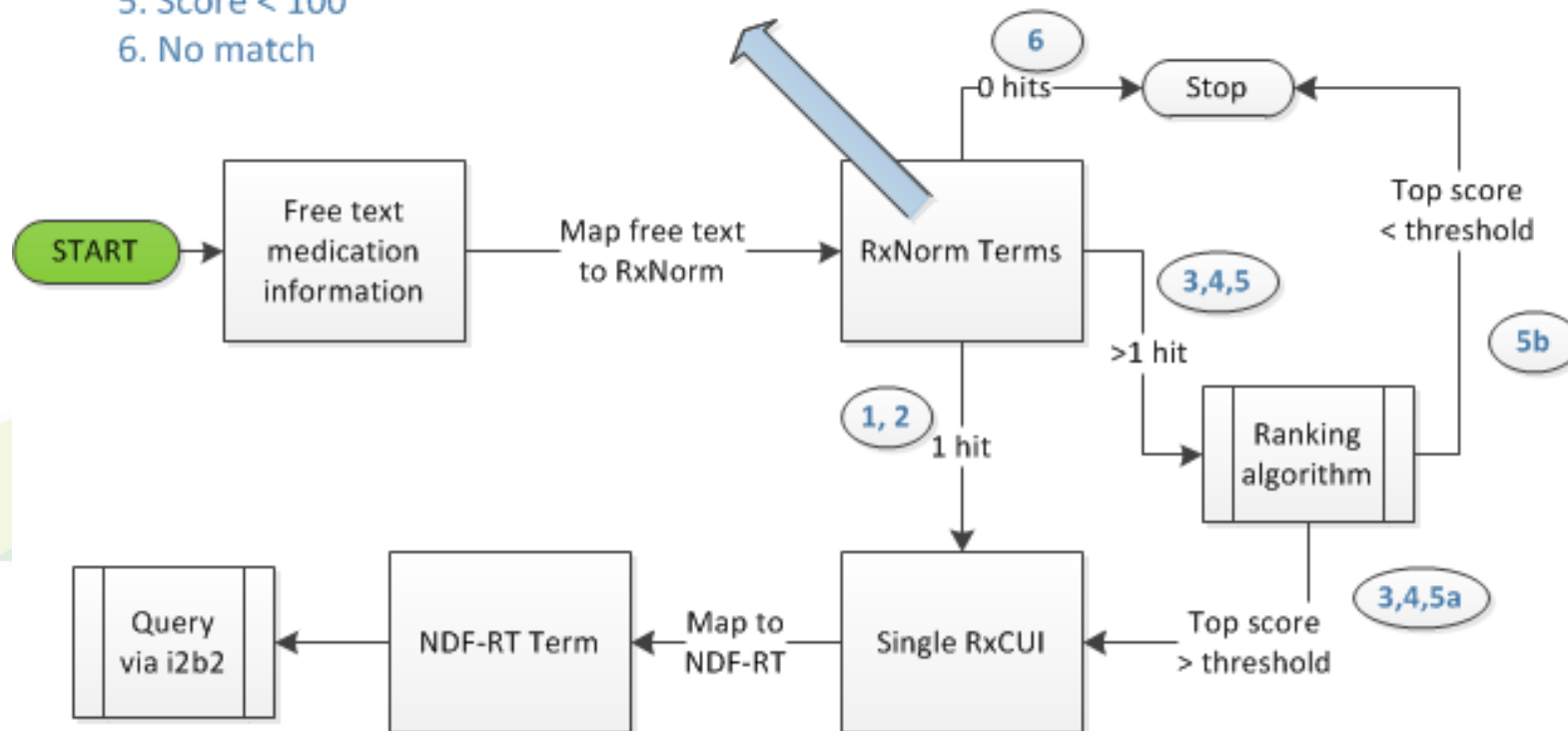
NLM Tools

- RxNav
- RxNorm APIs
 - <https://rxnav.nlm.nih.gov/REST/approximateTerm?term=prozak>
- RxMix

Text → RxNorm mapping workflow

5 outcomes:

1. Perfect match- no score given
2. Exactly 1 hit with score of 100
3. N hits with score of 100, top hit is non-proprietary
4. Proprietary-only hits with score of 100
5. Score < 100
6. No match

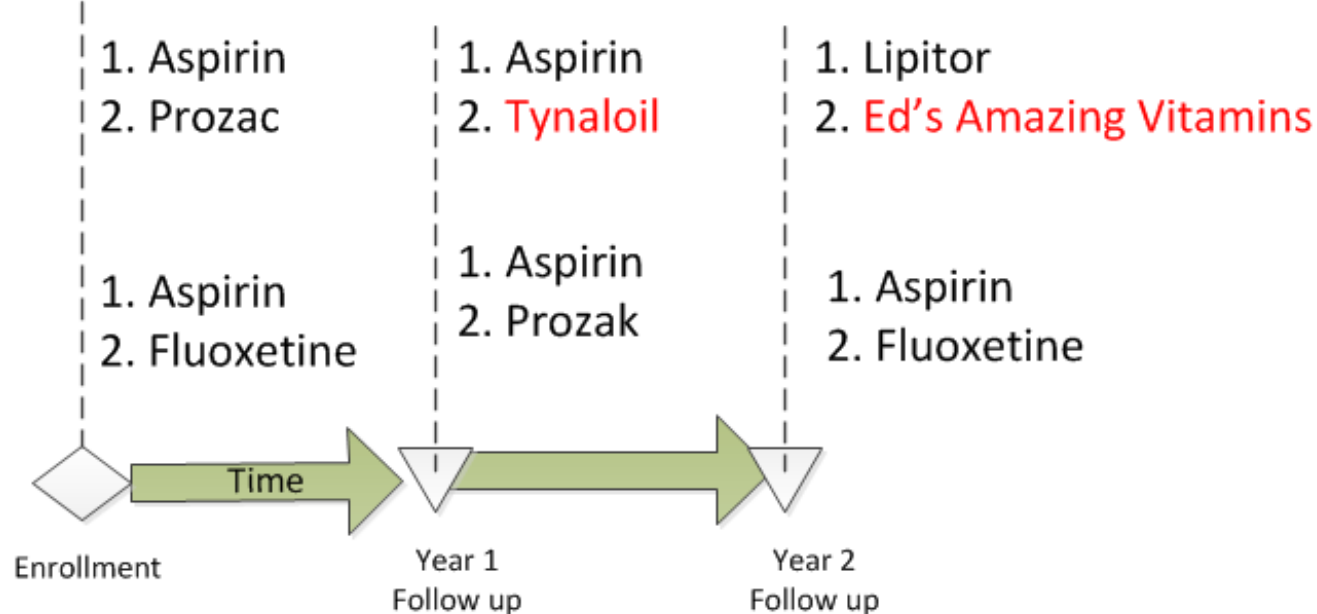




Person 1



Person 2



Total:

- Aspirin- RXCUI 1191
- Aspirin- RXCUI 1191
- Aspirin- RXCUI 1191
- Aspirin- RXCUI 1191
- Aspirin- RXCUI 1191
- Lipitor- RXCUI 153165
- Prozac- RXCUI 58827
- Prozac- RXCUI 58827
- Fluoxetine- RXCUI 4493
- Fluoxetine- RXCUI 4493
- Ed's Amazing Vitamins- ???
- Tynaloil- RXCUI 584198 (Tygacil)

10/12 mapped correctly
 1 mapped incorrectly
 1 not mapped

Unique:

- Aspirin- RXCUI 1191
- Prozac- RXCUI 58827
- Prozac- RXCUI 58827
- Fluoxetine- RXCUI 4493
- Lipitor- RXCUI 153165
- Ed's Amazing Vitamins- ???
- Tynaloil- RXCUI 584198 (Tygacil)

5/7 mapped correctly
 1 mapped incorrectly
 1 not mapped

Accuracy

At the given threshold:

Q: What % of total entries did we code correctly?

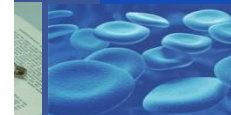
A: 83% total terms (10/12)

Q: What % of unique entries did we code correctly?

A: 71% unique terms (5/7)

Q: What % of people did we get perfectly correct?

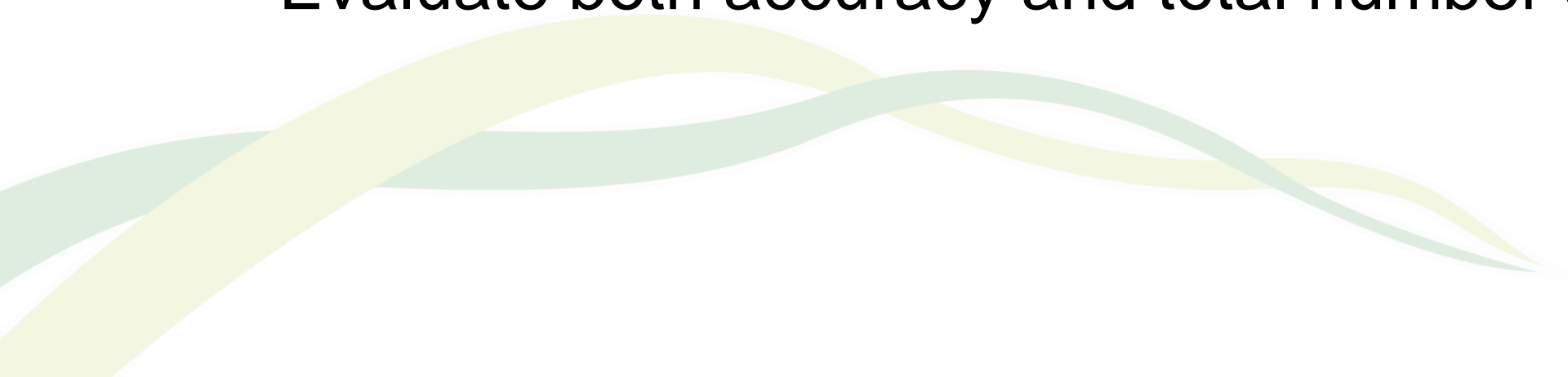
A: 50% of participants (1/2)

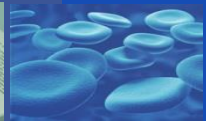




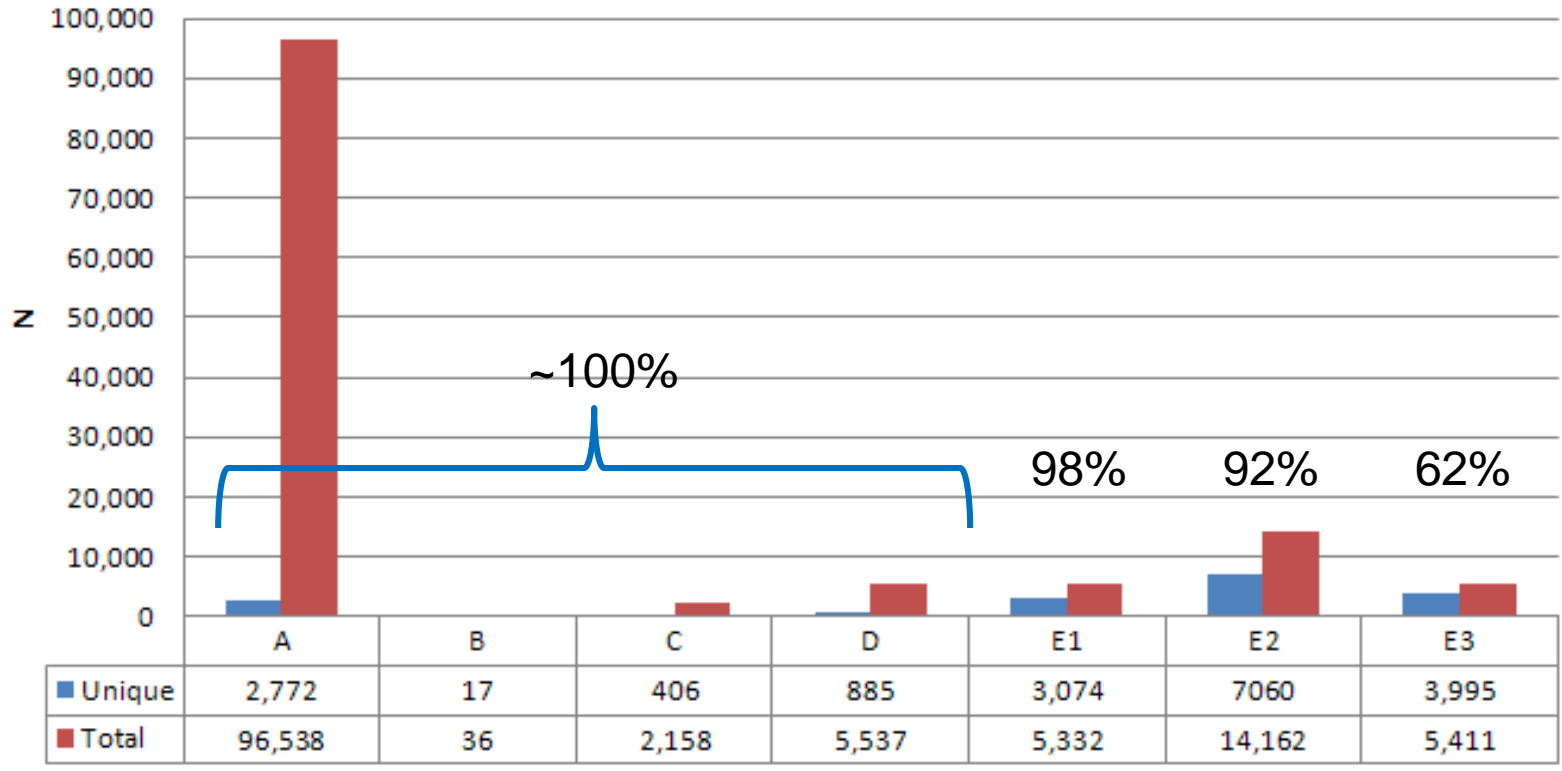
Choosing a threshold

- Break down results into classes based on score
- Determine accuracy of different classes
 - Choose 100 at random
 - Analyst with clinical expertise rates match
- Evaluate both accuracy and total number of terms





Breakdown of RxNorm mapping results



- A: Perfect matches
- B,C,D: Near perfect matches
- E1, E2, E3: Scores >75, >50, <50 respectively



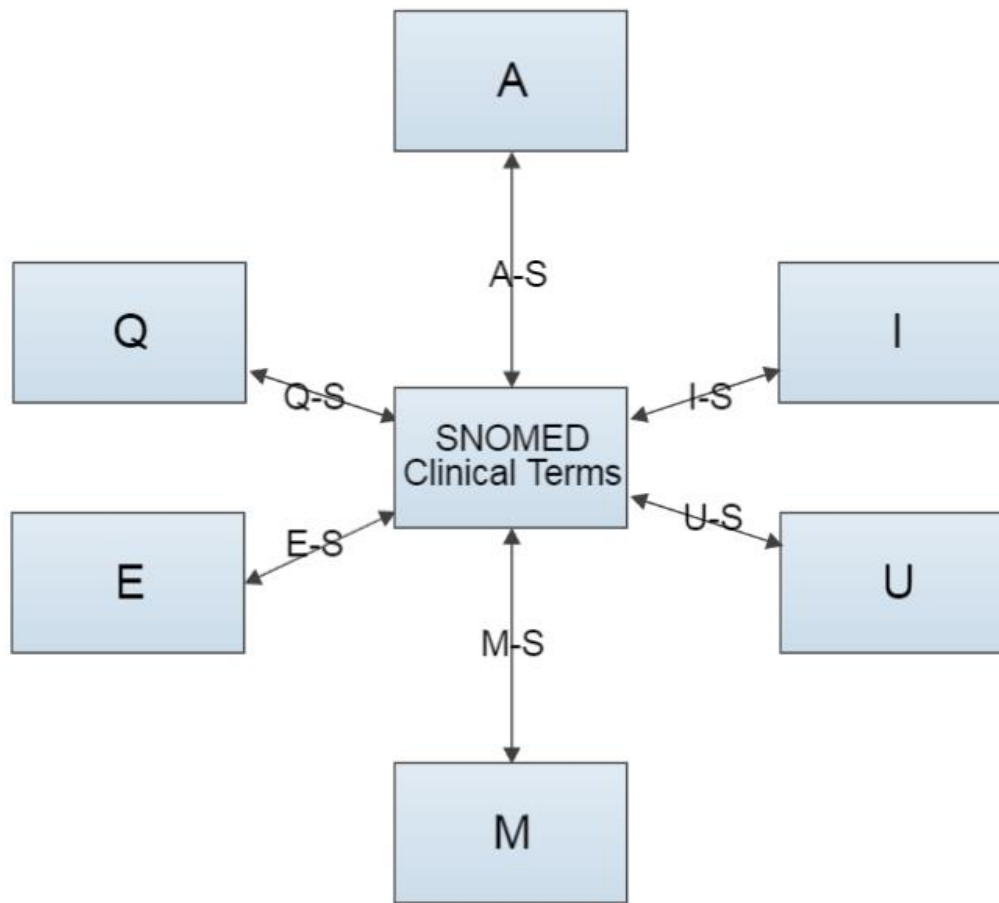
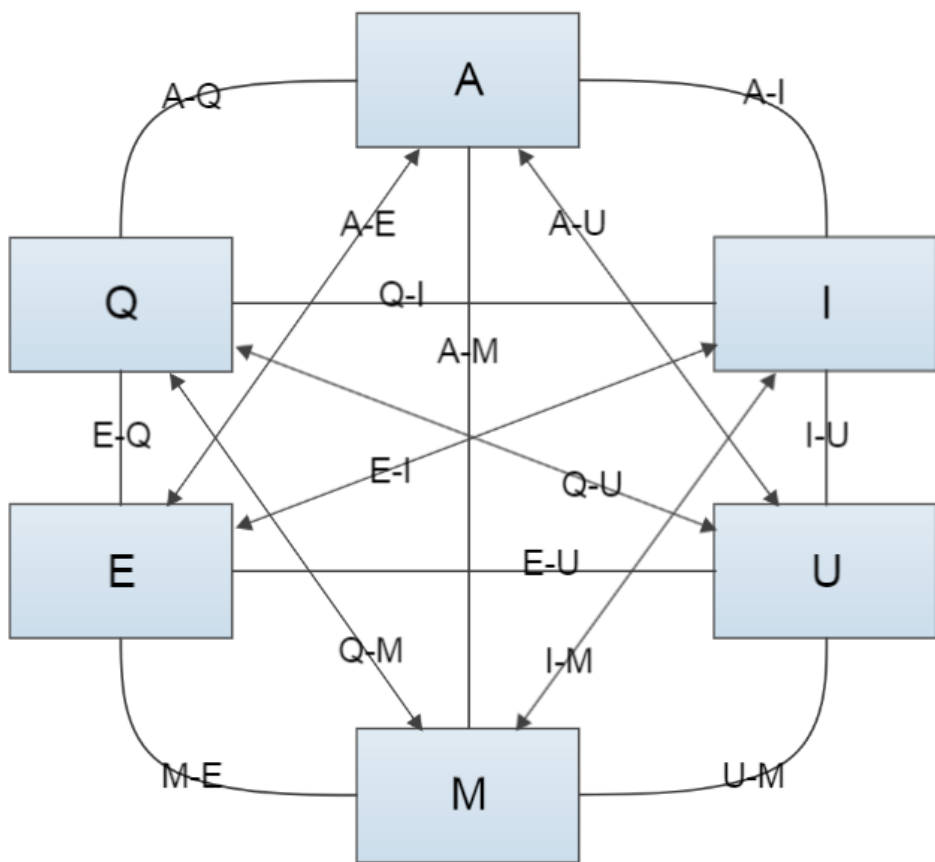


Conclusions

- Given threshold of ≥ 50 :
 - **Correctly mapped ~94% (122,523) of total terms**
 - **5% unmapped (6,510)**
 - **"false positive rate" (i.e. mapped, but incorrectly) of ~1% (1,240 entries)**
- Poor mapping happens largely with OTC, e.g. vitamins, minerals, supplements
- Important to consider use cases in deciding where to set cutoff
- Changes to RxNorm were unanticipated. Caveat emptor.



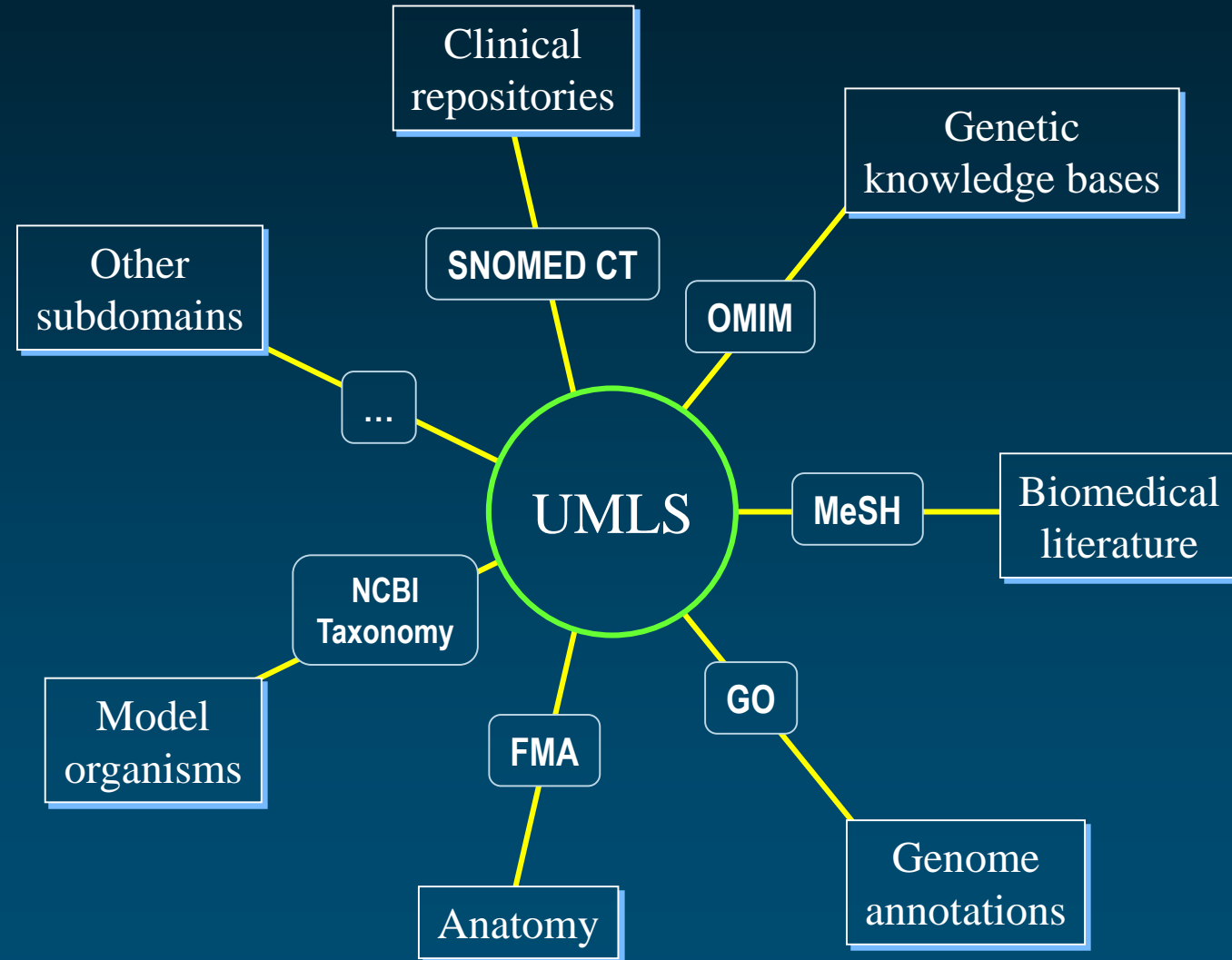
Reminder: why bother?



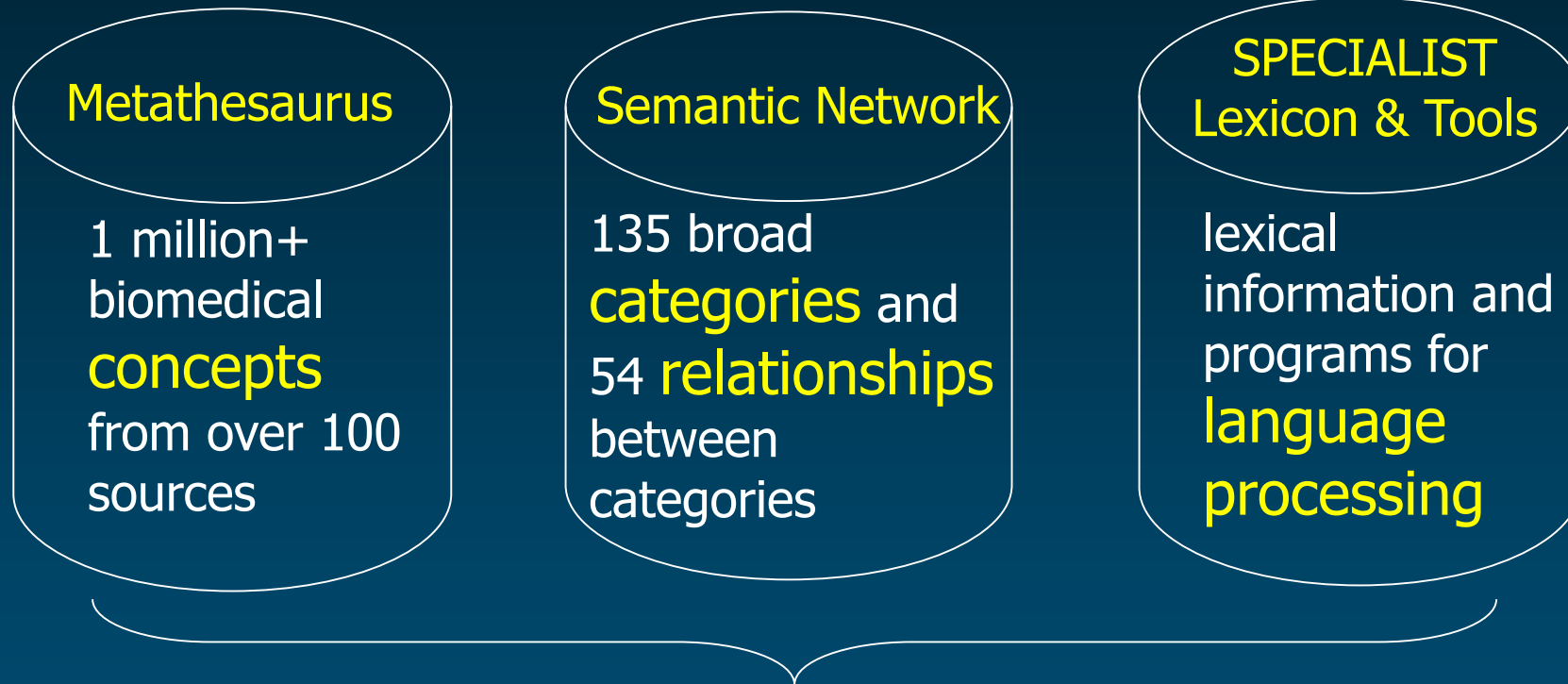
UMLS- Unified Medical Language System[®]

- NOT a vocabulary or terminology
- Collection of *many* terminologies, mappings between them, semantic tags, various tools
- Enable researchers to do useful things, especially in natural language processing

Integrating subdomains



The UMLS consists of



3 Knowledge Sources
used separately or together

Metathesaurus

- The Metathesaurus is a **large**, vocabulary database that contains biomedical and health related **concepts**, their names, and the **relationships** among them.
- The Metathesaurus contains over five million terms, or names, organized into concepts and assigned a unique identifier.
- The Metathesaurus is **not a vocabulary**. It contains many vocabularies and helps mappings between these vocabularies.

Metathesaurus cont.

Term	Source Vocabulary
Atrial fibrillation	ICD-9-CM
AF	NCI Thesaurus
AFib	MedDRA
Atrial fibrillation (disorder)	SNOMED Clinical Terms
atrium; fibrillation	ICPC2-ICD10 Thesaurus

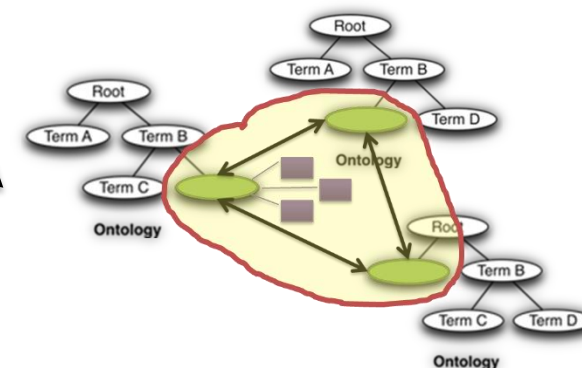
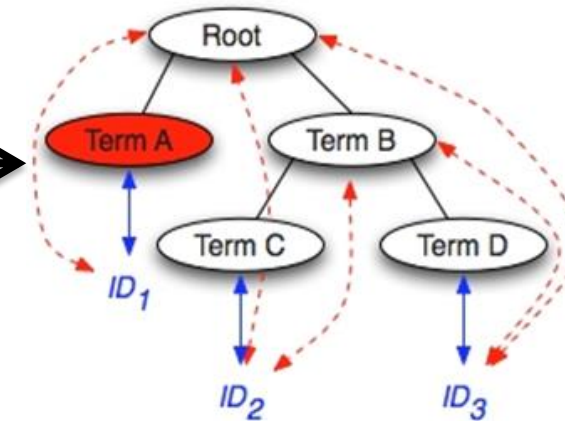
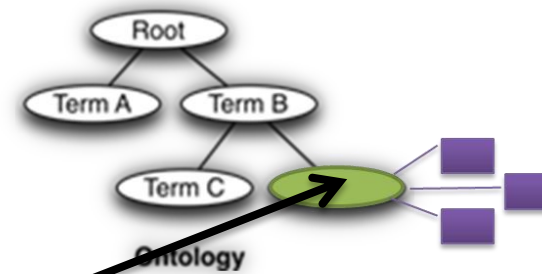
Preferred Terms

Collection of terms in the concept Hodgkin's Disease:

disease; Hodgkin
Hodgkins disease
Hodgkin Disease
Hodgkin's disease, unspecified
Hodgkin's disease, unspecified type
Hodgkin's disease (clinical)
Hodgkin's disease NOS, unspecified site
Hodgkin's disease NOS (disorder)
Hodgkin's sarcoma (clinical)
Hodgkin's sarcoma NOS
Hodgkin's sarcoma of unspecified site
Hodgkin's sarcoma of unspecified site (disorder)
Hodgkin's sarcoma-unspec. site
Hodgkin lymphoma
Lymphogranuloma, Malignant
Lymphogranulomatosis
Lymphogranulomatosis, malignant
Lymphomas Hodgkin's disease

The data files

File Name	Contents
MRCONSO	Names, Synonyms, Terms, Term Types, Codes
MRREL	Relationships
MRHIER	Hierarchies
MRSAT	Attributes
MRDEF	Definitions
MRMAP	Mappings
MRSMAP	Simplified Mappings
MRSTY	Semantic Types

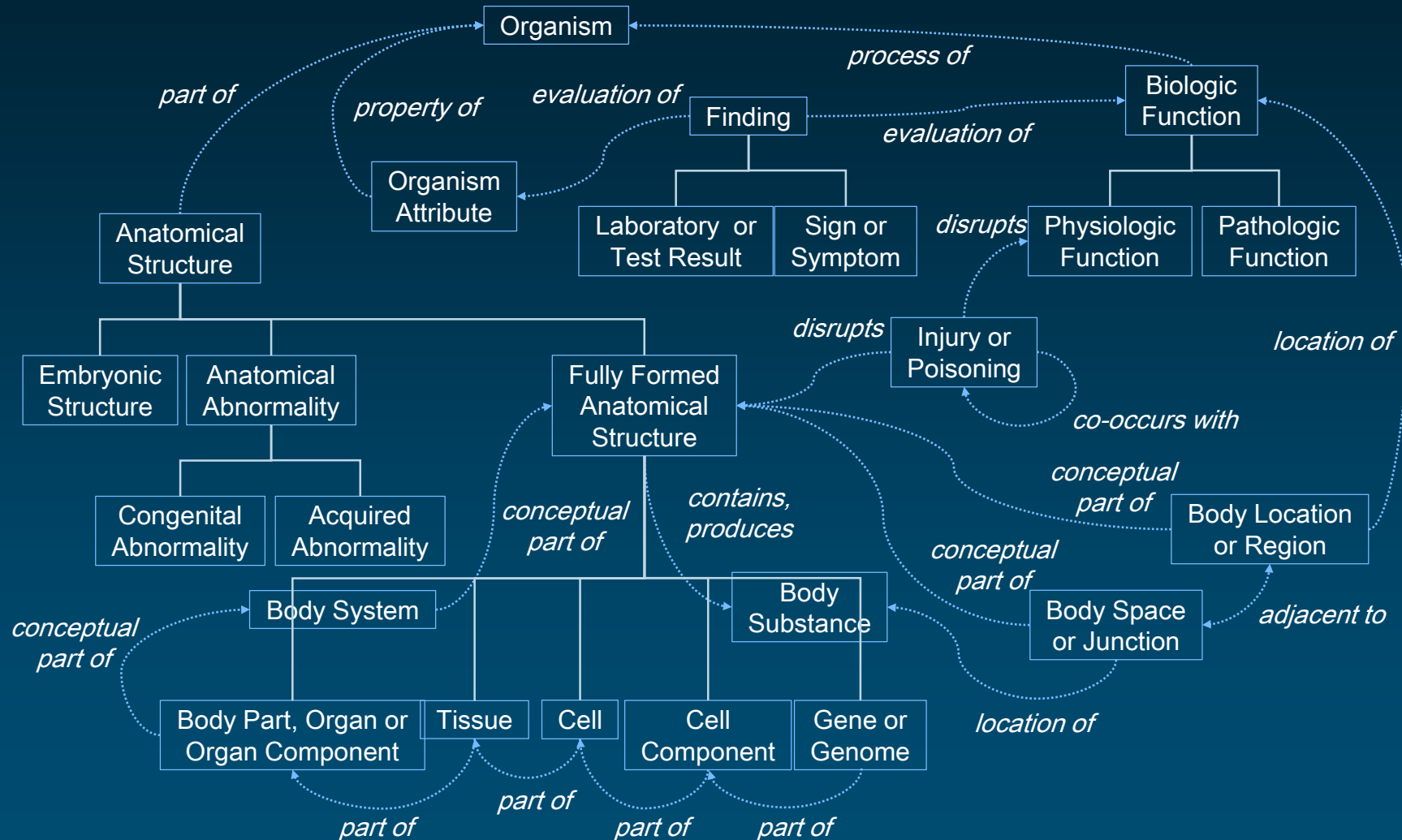


4 levels of specification

- Concept Unique Identifiers (CUI) – 2,612,024
- Lexical (term) Unique Identifiers (LUI) – 7,734,809
- String Unique Identifiers (SUI) – 8,677,735
- Atom Unique Identifiers (AUI) – 10,506,764



Semantic Types and Relationships

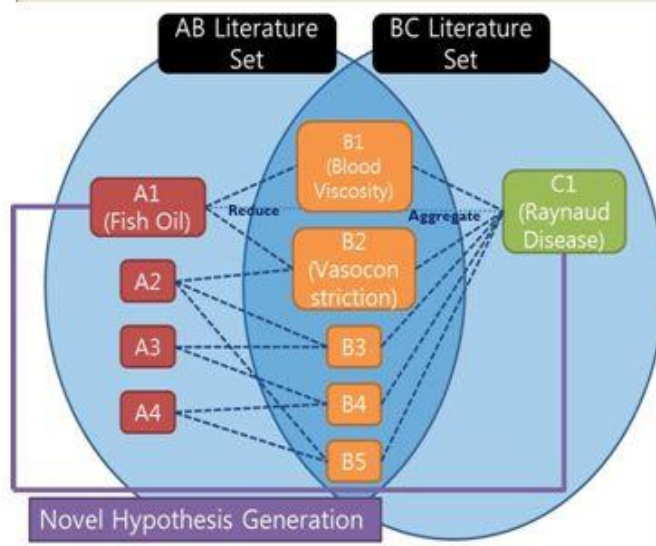


SPECIALIST Lexicon and Lexical Tools

- Treat, treats, treated, treating
- The SPECIALIST Lexicon is an English lexicon containing many words from the biomedical domain.
 - The majority of the words are nouns.
- The lexical tools are a collection of java programs that process natural language words and terms.
 - Include a normalizer, a word index generator, and a lexical variant generator.

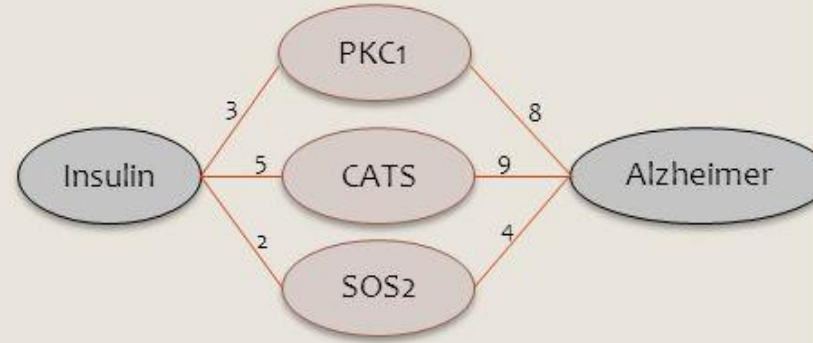
Literature Based Discovery (LBD)

- Conceptual Biology?



Swanson's ABC model

Drug repositioning



0706-0702 Nucleic Acids Research, 2008, Vol. 37, Database issue
doi:10.1093/nar/gln500

Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks

Allan Peter Davis, Cynthia G. Murphy, Cynthia A. Saracino, Michael C. Rosenstein, Thomas C. Wiegiers and Carolyn M. C. Harris

Received June 7, 2008; Revised August 26, 2008; Accepted August 27, 2008

ABSTRACT
The Comparative Toxicogenomics Database (CTD) is a curated database that promotes understanding

chemicals and their interactions with genes and proteins. To understand the molecular mechanisms of toxicity, it is essential to identify the genes and proteins that are affected by chemicals and to understand the interactions between these genes and proteins.

OPEN ACCESS Freely available online
PLOS COMPUTATIONAL BIOLOGY

Literature Mining for the Discovery of Hidden Connections between Drugs, Genes and Diseases

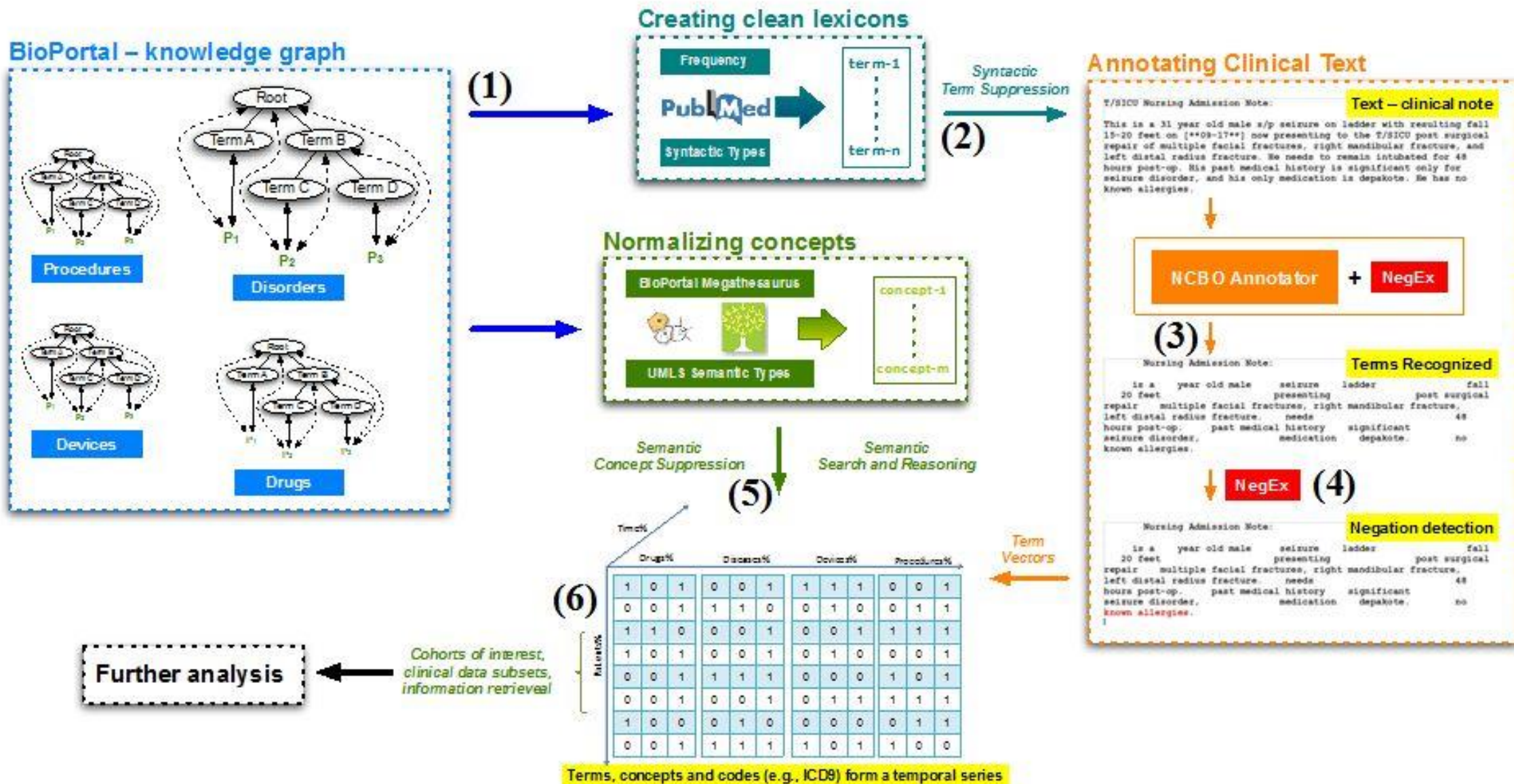
Raoul Frijters¹, Marianne van Vugt², Ruben Smeets², Renie van Schaik³, Jacob de Vlieg^{1,3}, Wynand Alkema^{1*}

¹ Computational Drug Discovery (CDD), Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, ² Department of Immune Therapeutics, Schering-Plough, Oss, The Netherlands, ³ Department of Molecular Design & Informatics, Schering-Plough, Oss, The Netherlands

Abstract

The scientific literature represents a rich source for retrieval of knowledge on associations between biomedical concepts such as genes, diseases and cellular processes. A commonly used method to establish relationships between biomedical concepts from literature is co-occurrence. Apart from its use in knowledge retrieval, the co-occurrence method is also well-suited to discover new, hidden relationships between biomedical concepts following a simple ABC principle, in which A and C have no direct relationship, but are connected via observed B intermediaries. In this paper we describe C-Phish, a literature mining tool that discovers hidden connections between biomedical concepts.

Cole et al. *Pediatric Rheumatology* 2013



Back to BioPortal

- Search for terms
- Search for ontologies
- Annotator- <http://bioportal.bioontology.org/annotator>
- Recommender-
<https://bioportal.bioontology.org/recommender>

Thanks!

Jessie.Tenenbaum@duke.edu