

Overview of Data Types and Terminologies

Overview

- Terminologies
- Codified data
 - ICD 9 and 10 billing codes
 - Current Procedural Terminology (CPT)
- Semi-codified data
 - Laboratory test results
 - Vitals
 - Medication records
 - Problem Lists
- Unstructured data
 - Clinical notes

Terminology Basics

SLIDES ADAPTED FROM JOSH DENNY

A solid orange horizontal bar spanning the width of the slide at the bottom.

Why Use Terminologies?

Without a terminology to constrain data, computer systems are unable to answer questions with certainty

- Unable to share data with partners except as lumps of text that must be individually read
- Unable to process decision support rules and prevent adverse events
- Unable to perform quality improvement or research between partners

First, an overview:

What are Controlled Terminologies

Word strings called **terms**

Convert “boundless chaos of living speech” into sharable, reusable, computable “concepts” for a number of uses

Examples:

- SNOMED-CT (Systematized Nomenclature of Medicine -- Clinical Terms)
- LOINC (Logical Observation Identifiers Names and Codes)
- Medical Subject Headings (MeSH)

UMLS aggregates many terminologies with common concepts

Types of terminologies

Rosenbloom et al. 2006;13:277-288

Administrative terminologies

- ICD9, CPT

Reference terminologies

- SNOMED-CT, MeSH, RxNorm
- Designed to contain “gold standard” concepts and relations between them
- Often contains “assertional knowledge” – “chest pain” can have “laterality”, “severity”, and “can radiate”

Interface terminologies

- Quill/CHISL
- Often designed to be coordinated with a reference terminology
- Encoding clinical narratives into structured forms or reviewing previously structured information

Reference Terminology

IS a terminology where each term has a formal definition supporting data aggregation and retrieval.

IS NOT simply a terminology we “refer” to

What is a Formal Definition?

Ontology – defines the kinds of things that can exist in the application domain

Without ontology, terms and symbols are ill-defined, confused, and confusing

What is a Formal Definition?

Logic – provides formal structure and the rules of inference

Without logic, knowledge representation is vague, without criteria for determining whether statements are redundant or contradictory

Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations

Desiderata of Terminologies

Jim Cimino. Methods Inf Med. 1998 Nov;37(4-5):394-403, >500 citations

Terms should be grouped into concepts

Concepts should be polyhierarchical

- “jugular venous pulse” can be a child of “neck exam” or “cardiovascular exam”

Aim for domain completeness

... which means you should specify the domain for which the terminology is designed

Integrate with other terminologies

Terminology ‘Desiderata’

Graceful Evolution

The content and structure of controlled medical terminologies must change over time. Desirable changes include additions, refinements, pre-coordination, disambiguation, obsolescence, discovered redundancy, and minor name changes. Bad reasons for change include major named changes, code reuse, and changed codes. Clear, detailed descriptions of the changes are necessary.

- “042” shouldn’t be “HIV” one year and “CHF” another...
- Also means the concept unique identifiers (UMLS term - CUIs) should not really have meaning in and of themselves but just be numbers/alphanumerics

Terminology ‘Desiderata’

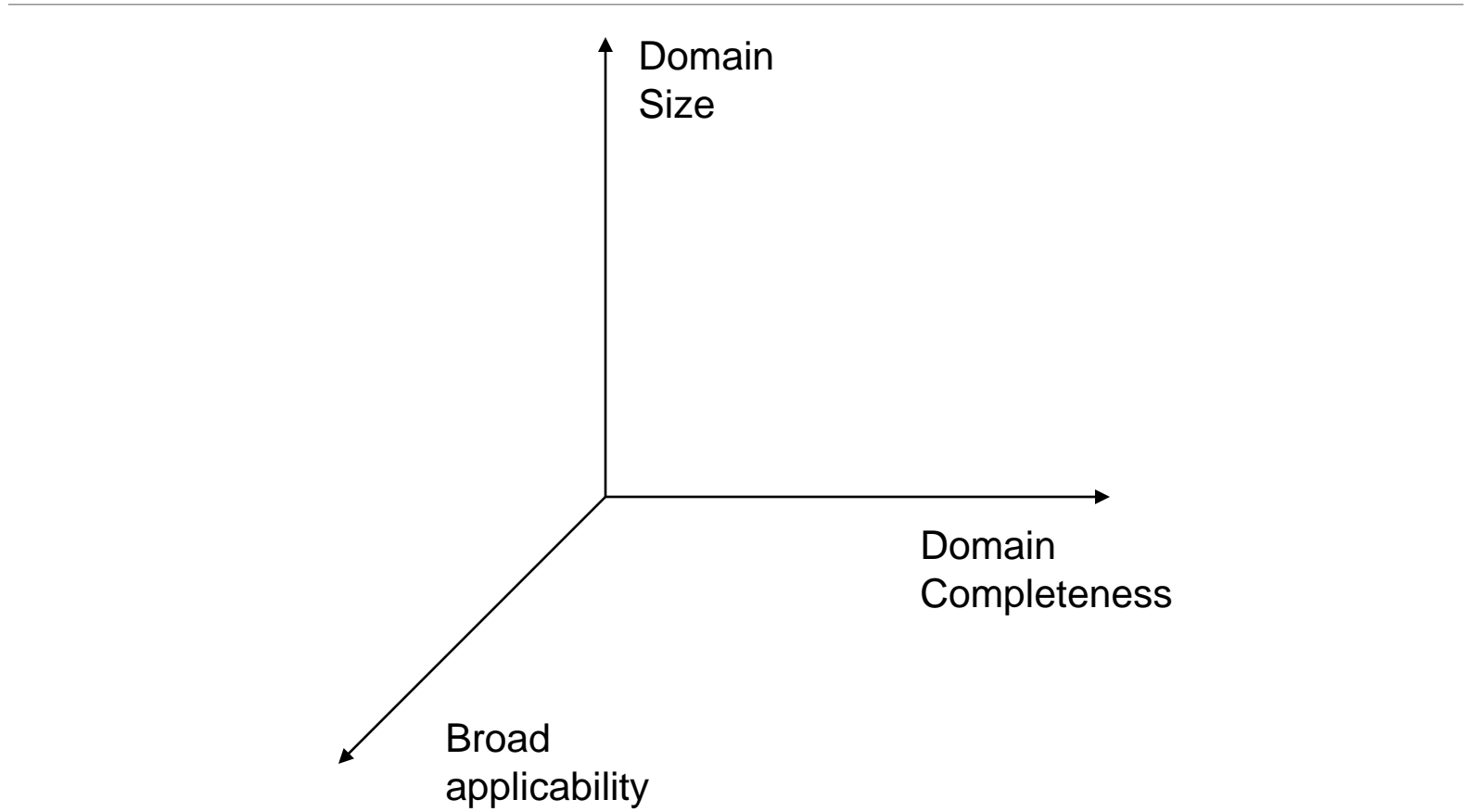
Reject “Not Elsewhere Classified”

Catch-all terms can only be defined by exclusion. As a terminology evolves, the meaning of “NEC” changes. Additionally, NEC can never have a formal definition. Therefore, NEC cannot be considered a valid term.

Terminology ‘Desiderata’

Multiple Granularities

The granularity of a term is a measure of its specificity and refinement. For example, Diabetes Mellitus is more coarsely granular than Diabetes Mellitus Type II. Multiple granularities are needed for multi-purpose terminologies.



Coordination of terms

Post-coordination: users of the terminology can combine terms to create/refine new meanings.

Pre-coordination: The terminology defines what mixtures of terms are available.

Many trade offs between the usability and descriptive power of terminologies!

Common Terminologies

LOINC (Logical Observation Identifiers Names and Codes)

- Covers labs great

SNOMED-CT (Systematized Nomenclature of Medicine)

- Covers diseases, s/sx well but not meds

ICD9-CM, ICD9, ICD10, ICD10-CM

- Diseases, s/sx, procedures

CPT (Current Procedural Terminology)

- procedures

RxNorm

Unified Medical Language System (UMLS) includes all of these and many more...

What is OWL?

- OWL – Web Ontology Language
 - Actually a family of languages that builds upon Resource Description Framework Schema RDF(S)
- OWL allows the description of classes, individuals, and properties.
- OWL formal semantics can be applied against OWL ontologies to reason out facts that may not be explicitly defined.


Class Descriptions: Enumeration

An enumerated list of the individuals (or instances) are used to define the class.

Utilizes the built in *oneOf* property

The below example identifies the class of the parts of a car engine.

```
<owl:Class>
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#Block">
    <owl:Thing rdf:about="#Cylinder">
    <owl:Thing rdf:about="#Piston">
    <owl:Thing rdf:about="#Ignition System">
    <owl:Thing rdf:about="#Fuel System">
    ....
  </owl:oneOf>
</owl:Class>
```



Statements in a level are order independent

Tools for creating/editing ontologies

Apelon TDE

Protégé

Add some others?

Standard Terminologies?

“The good thing about standards is that there are so many of them” - Ed Hammond

Standards Selection Processes

- CHI
- NCVHS
- HITSP
- ONC EHR Certification (“meaningful use”)

Codified Data

Generally speaking, it's the easiest to work with.

Tends to be limited in descriptive power.

ICD codes

International Classification of Disease (ICD)

The US now uses ICD-10 Clinical Modification (CM), much of the rest of world uses ICD-10

- US continued to use ICD-9 CM through 2015

Diagnostic codes:

- ICD-9-CM: ~13,500
- ICD-10: ~14,000
- ICD-10-CM: ~68,000
 - Some new biology, but most differences not clinical/biologic (which side is the lesion, 1st visit or repeat, etc)
- ICD-11 in development

ICD9 codes

3-digit codes (000-999): diagnoses, signs, symptoms

2-digit codes (00-99): procedures

V-codes and E-codes

Grouping	Examples	Count
Chapter	390-459.99 DISEASES OF THE CIRCULATORY SYSTEM	20
Section	401-405.99 HYPERTENSIVE DISEASE 390-392.99 ACUTE RHEUMATIC FEVER	120
Category (3-digit)	401 Essential Hypertension 402 Hypertensive heart disease	900+
Fully-specified (3-5 digits)	401.9 Benign essential hypertension 402.11 Benign hypertensive heart disease with heart failure	~13,500

Example: ICD9

401-405.99

HYPERTENSIVE DISEASE

401	Essential hypertension	404.01	Hypertensive heart and renal disease, malignant, with heart failure
401.0	Malignant essential hypertension	404.02	Hypertensive heart and renal disease, malignant, with renal failure
401.1	Benign essential hypertension	404.03	Hypertensive heart and renal disease, malignant, with heart failure and renal failure
401.9	Unspecified essential hypertension	404.1	Hypertensive heart and renal disease, benign
402	Hypertensive heart disease	404.10	Hypertensive heart and renal disease, benign, without mention of heart failure or renal failure
402.0	Malignant hypertensive heart disease	404.11	Hypertensive heart and renal disease, benign, with heart failure
402.00	Malignant hypertensive heart disease without heart failure	404.12	Hypertensive heart and renal disease, benign, with renal failure
402.01	Malignant hypertensive heart disease with heart failure	404.13	Hypertensive heart and renal disease, benign, with heart failure and renal failure
402.1	Benign hypertensive heart disease	404.9	Hypertensive heart and renal disease, unspecified
402.10	Benign hypertensive heart disease without heart failure	404.90	Hypertensive heart and renal disease, unspecified, without mention of heart failure or renal failure
402.11	Benign hypertensive heart disease with heart failure	404.91	Hypertensive heart and renal disease, unspecified, with heart failure
402.9	Unspecified hypertensive heart disease	404.92	Hypertensive heart and renal disease, unspecified, with renal failure
402.90	Unspecified hypertensive heart disease without heart failure	404.93	Hypertensive heart and renal disease, unspecified, with heart failure and renal failure
402.91	Unspecified hypertensive heart disease with heart failure	405	Secondary hypertension
403	Hypertensive kidney disease	405.0	Malignant secondary hypertension
403.0	Hypertensive renal disease, malignant	405.01	Malignant renovascular hypertension
403.00	Hypertensive renal disease, malignant, without mention of renal failure	405.09	Other malignant secondary hypertension
403.01	Hypertensive renal disease, malignant, with renal failure	405.1	Benign secondary hypertension
403.1	Hypertensive renal disease, benign	405.11	Benign renovascular hypertension
403.10	Hypertensive renal disease, benign, without mention of renal failure	405.19	Other benign secondary hypertension
403.11	Hypertensive renal disease, benign, with renal failure	405.9	Unspecified secondary hypertension
403.9	Hypertensive renal disease, unspecified	405.91	Unspecified renovascular hypertension
403.90	Hypertensive renal disease, unspecified, without mention of renal failure	405.99	Other unspecified secondary hypertension
403.91	Hypertensive renal disease, unspecified, with renal failure		
404	Hypertensive heart and kidney disease		
404.0	Hypertensive heart and renal disease, malignant		
404.00	Hypertensive heart and renal disease, malignant, without mention of heart failure or renal failure		

ICD10 codes

Organized into 21 chapters

Starts with a letter and then 2 digits, with more specificity after the decimal

Chapters typically are one letter each, but not always

I10-I15 Hypertensive diseases, eg, with I10 as Essential (primary) hypertension

The problem with billing codes

Billing codes only 50-80% accurate

False positives

- Diagnoses evolve over time -- physicians may initially bill for suspected diagnoses that later are determined to be incorrect
- Wrong code entered (easier to find or remember)
- Physicians may bill for a different condition if it pays for a given treatment
 - psoriatic arthritis and rheumatoid arthritis

False negatives:

- Outpatient billing limited to 4 diagnoses/visit
- Outpatient billing done by physicians (e.g., takes too long to find the unknown ICD9)
- Inpatient billing done by professional coders:
 - omit codes that don't pay well
 - can only code problems actually explicitly mentioned in documentation

Procedures

Clinical Procedural Terminology (CPT) codes are used in the US. The terms are copyrighted by the American Medical Association which makes them challenging to use and talk about.

Other coding sets exist, including extensions of ICD.

Can be helpful as they document events that happen, some of which can be very accurate, eg, an appendectomy code may be more specific than an appendicitis code.

This also means they can be missed if an individual was outside of the health system for the event.

Semi-structured data

Typically includes things like laboratory test results, vitals, medication records, and problem lists.

Varies in degree of structure.

Frequently easier to use in local contexts, but can be more challenging to share.

Also can be challenging over time as methods and more change over time.

Laboratory Logical Observation Identifier Name Codes (LOINC)

46,812 terms (12/06)

Uses: Lab test order names, lab result names, drug label section headers

Available free in Access format

- <http://www.regenstrief.org/medinformatics/loinc/downloads/>

Very specific: can be a challenge to map post-hoc

LOINC Parts



COMPONENT (ANALYTE)

The substance or entity being measured or observed.



PROPERTY

The characteristic or attribute of the analyte.



TIME

The interval of time over which an observation was made.



SYSTEM (SPECIMEN)

The specimen or thing upon which the observation was made.



SCALE

How the observation value is quantified or expressed: quantitative, ordinal, nominal.



METHOD

OPTIONAL A high-level classification of how the observation was made. Only needed when the technique affects the clinical interpretation of the results.

Vitals and anthropometrics

Blood pressure, pulse, respiratory rate, height, weight, etc, can all be meaningful measures.

While typically structured, there can also be variation and noise:

- High blood pressure/pulse if the patient was just rushing in
- Weights copied forward from prior visits
- Unit differences in some places (eg, inches/cm)

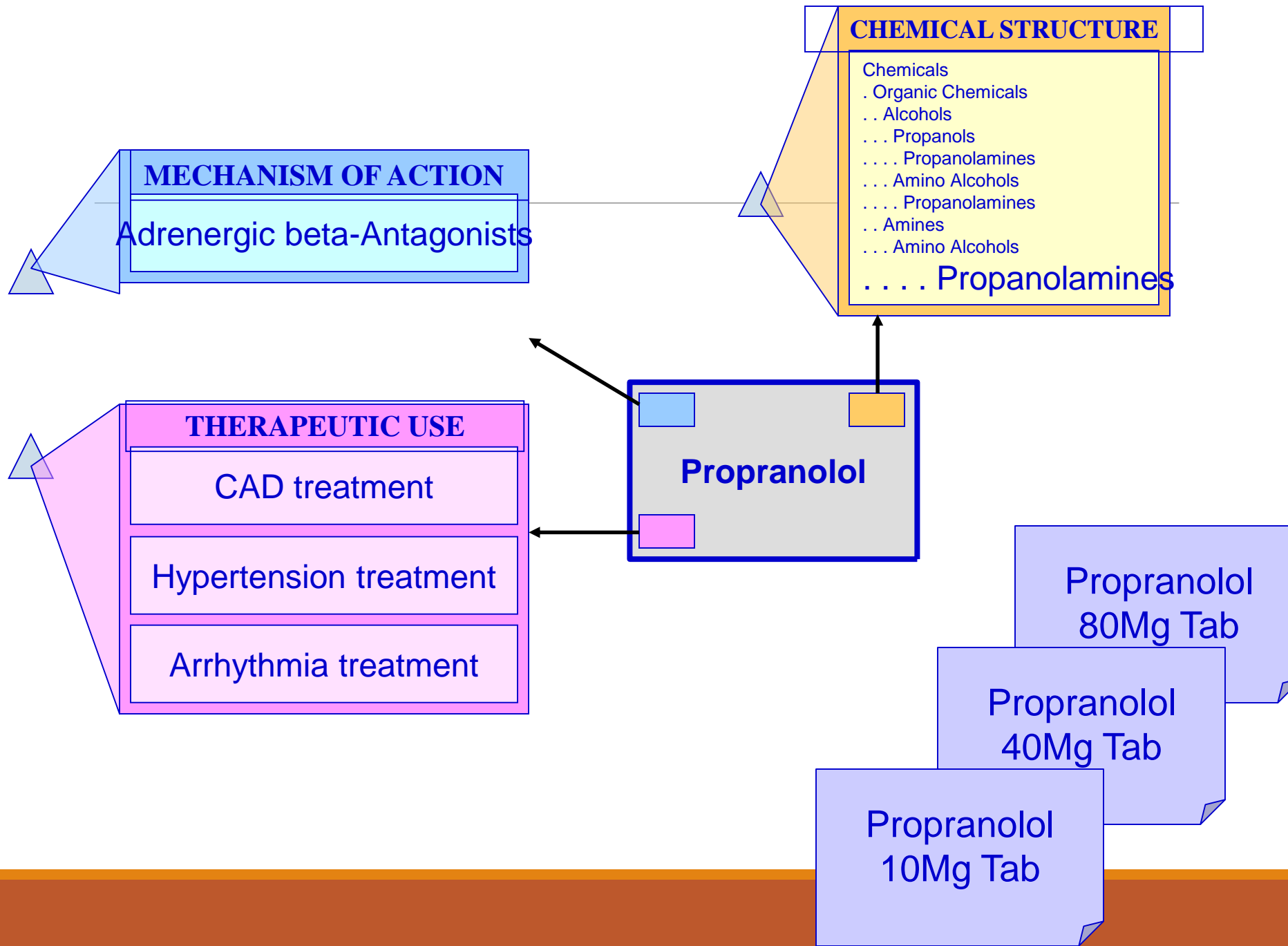
Medication Records

Wide variability in how structured this data can be

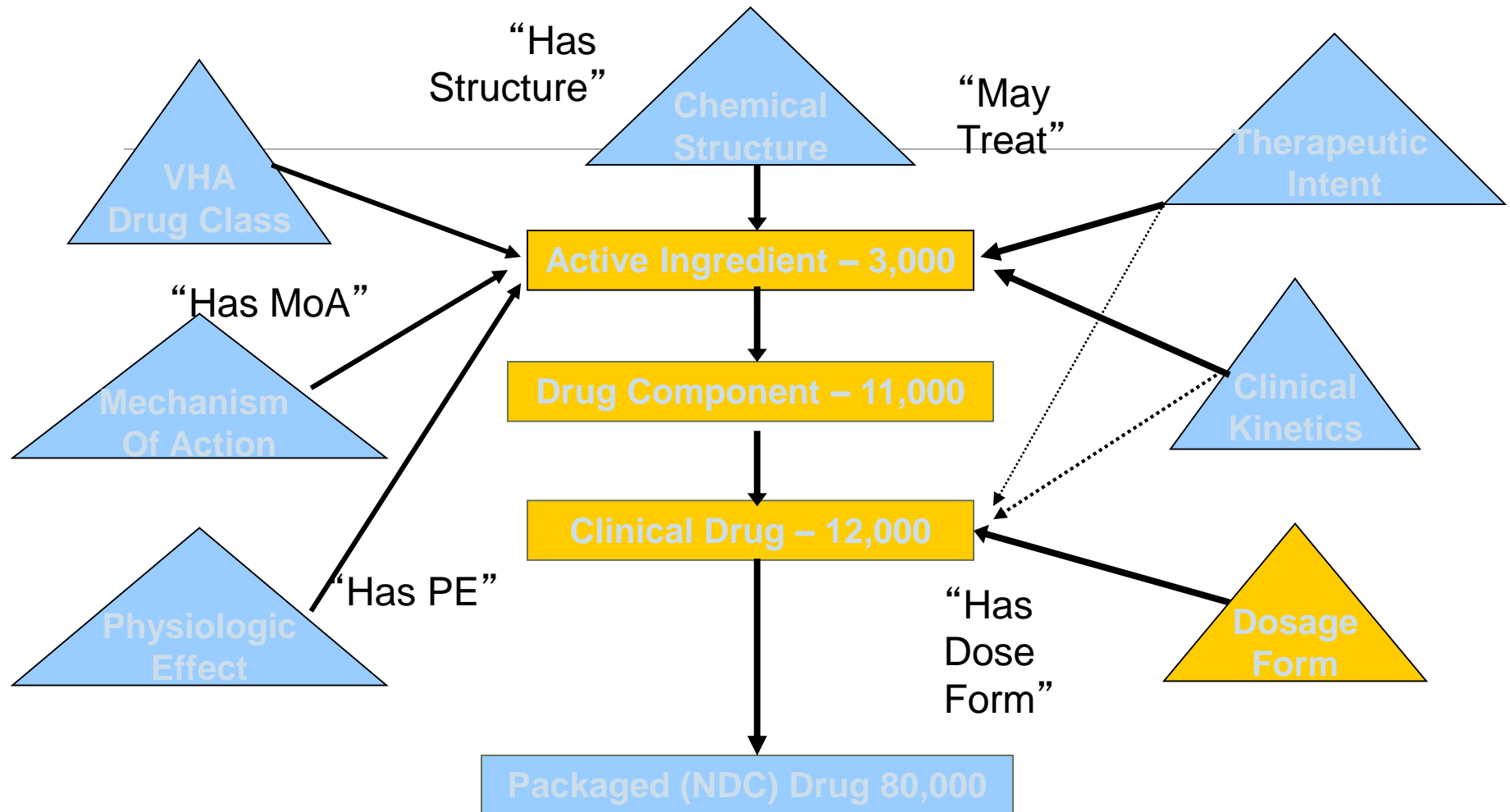
- Electronic ordering systems may have 100% specific codes
- Free text prescription may have no structure at all

Many terminologies for classification and representation of medication information.

- Anatomical Therapeutic Chemical (ATC) Classification System
 - Controlled by the World Health Organization
 - Drugs can have several codes associated with them if they have different indications
- RxNorm
 - US National Library of Medicine manages it
 - Attempts to normalize drugs to single concepts in a hierarchy
 - Helps with generics/brand names
 - Splits up dosing, but also has parent terms.
- National Drug Codes (NDCs), National Drug File Reference Terminology, and more



VHA NDF RT – Semantic Model



Roles/Properties such as “May Treat” are semantic (read: meaningful) linkages that assert knowledge by connecting ontologic “things”

Formal Definition of Atenolol

ATENOLOL PREPARATION

- Primitive
- Kind: DRUG_KIND
- Code: C11316
- Id: 11317
- [CV100] BETA-BLOCKERS/RELATED
 - some has_Ingredient: Atenolol
 - some has_MoA: Adrenergic beta1-Antagonists
 - some has_PE: Negative Chronotropy
 - some has_PE: Negative Inotropy
 - some has_PK: Renal Excretion
 - some may_prevent: Migraine
 - some may_treat: Angina Pectoris
 - some may_treat: Atrial Fibrillation
 - some may_treat: Atrial Flutter
 - some may_treat: Heart Failure, Congestive
 - some may_treat: Hypertension
 - some may_treat: Myocardial Infarction
 - some may_treat: Tachycardia, Supraventricular
 - some may_treat: Tachycardia, Ventricular
 - some may_treat: Ventricular Fibrillation
- Last_Reviewed: 2/21/2003
- Level: Ingredient
- Pharmacologic_Action: Sympatholytics
- Reviewed_By: CC | 2/21/2003
- Search_Name: ATENOLOL PREPARATION
- Status: Reviewed

Problem Lists

A growing number of EHRs are attempting to structure “problem list” data

The approach is to empower clinicians to map single line item problems to codified terms

Benefits:

- Clarity of communication
- Improved decision support
- Reuse of data

Free text data

The most challenging data to work with

Difficult to standardize

Requires some type of knowledge extraction: keyword identification, regular expression parsing, natural language processing, etc

Hard even for people to understand sometimes!

SNOMED CT

~350,000 Concepts

~1.1 Million Synonyms

Formal Definitions

Uses: Diagnosis and problems, Laboratory result contents, Non-laboratory interventions and procedures, Anatomy Nursing, Allergic reactions

Available for use in the US at no charge via UMLS

IHTSDO established on 23 March 2007 and acquired the SNOMED CT intellectual property on 26 April 2007

Browsable at:

http://bioportal.nci.nih.gov/ncbo/faces/pages/ontology_list.xhtml