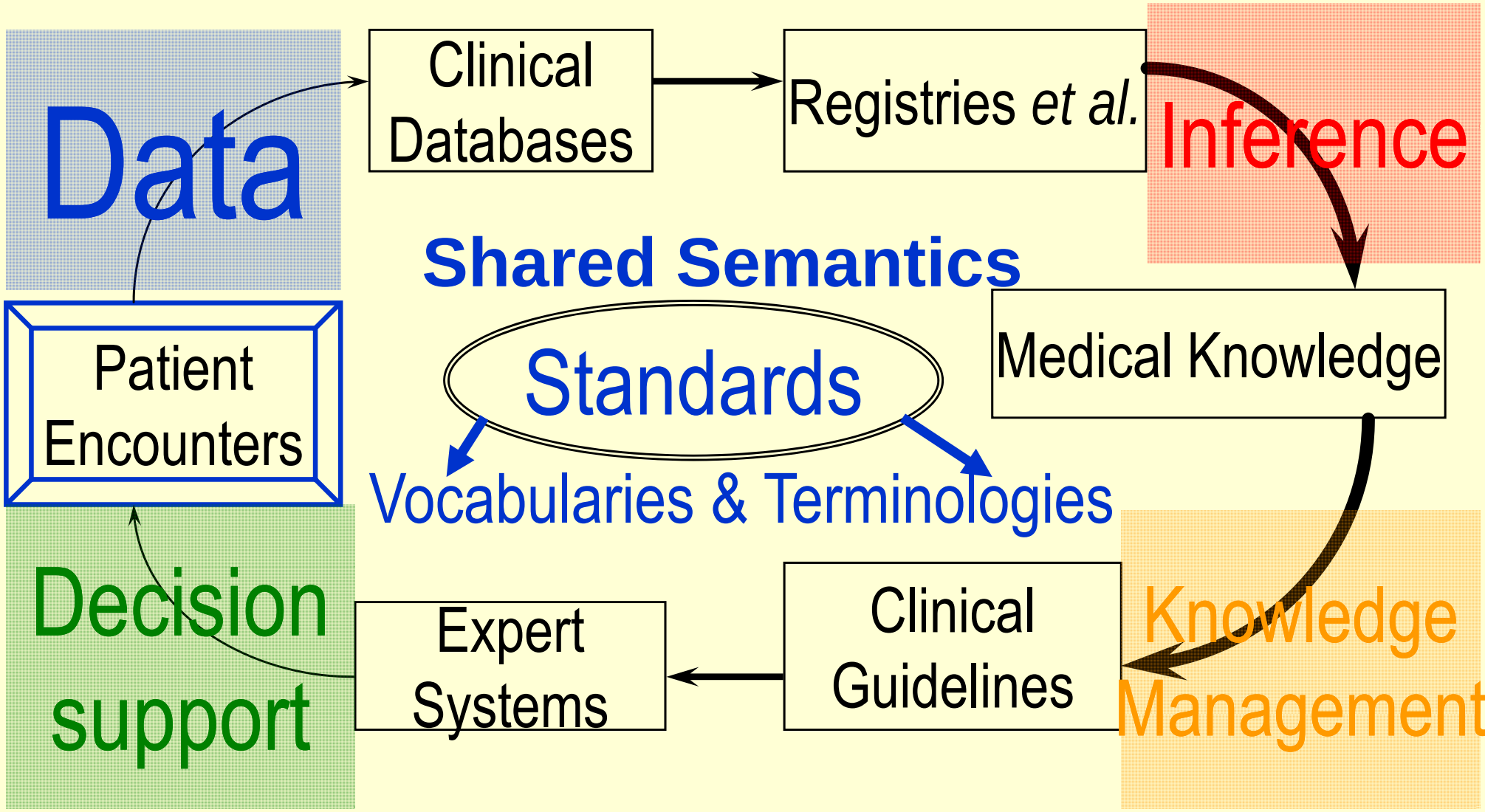
Relationships among Biomedical Ontologies and Classifications: The ICD11 Use-case

Christopher G. Chute, MD DrPH
Professor Biomedical Informatics
Vice-Chair, Data Governance
Mayo Clinic College of Medicine

Chair, International Classification of Disease Revision
Chair, ISO Technical Committee 215 on Health Informatics

OntologySummit2011: panel session-9 "Grand Challenges"
Webcast, 24 March, 2011

From Practice-based Evidence to Evidence-based Practice



Whither Phenotype?
Spans spectrum from enzymes to disease

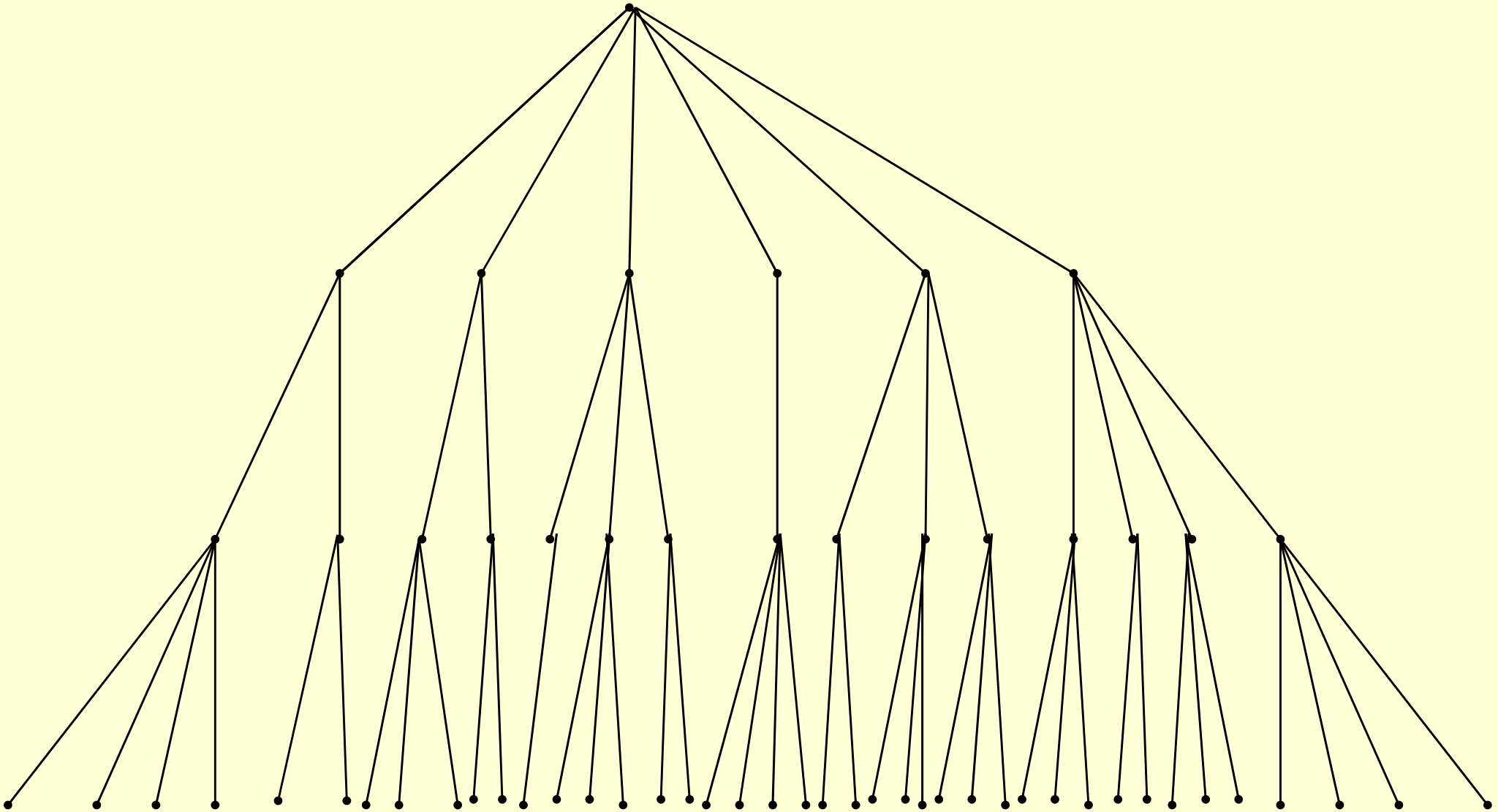
- Pharmacogenomics – enzyme functionality
- Physiologist – cellular function
- Systems biologist – pathway circuit flow
- Sub-specialist – organ functioning
- Patient/Clinician – disease manifestation
- Public Health – population characteristics

Highly specific to use-case context

ICD11 Use Cases

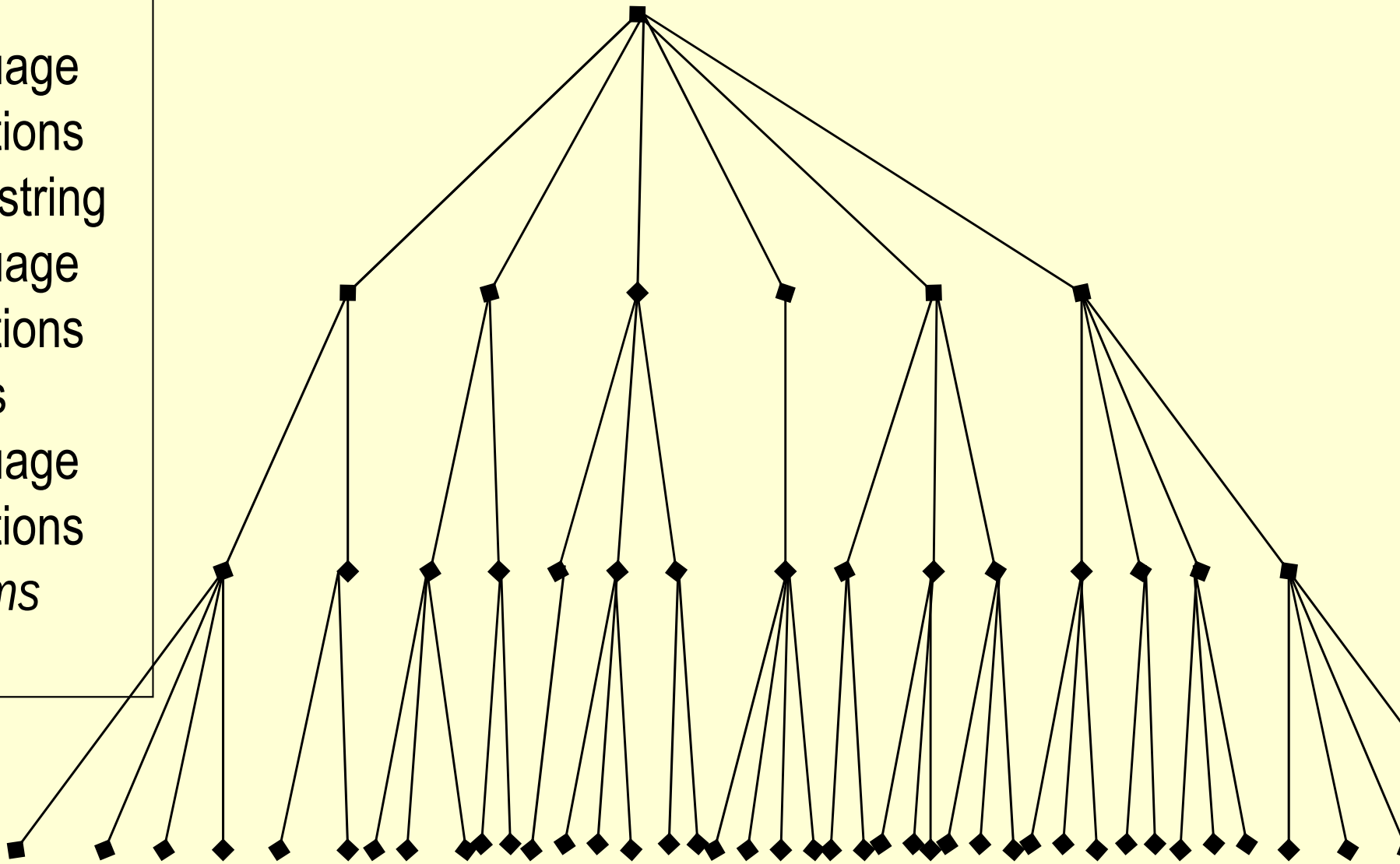
- Scientific consensus of *clinical phenotype*
- Public Health Surveillance
  - Mortality
  - Public Health Morbidity
- Clinical data aggregation
  - Metrics of clinical activity
  - Quality management
    - Patient Safety
  - Financial administration
    - Case mix
    - Resource allocation

Traditional Hierarchical System
ICD-10 and family



Addition of structured attributes to concepts

- Concept name
  - Definition
    - Language translations
  - Preferred string
    - Language translations
  - Synonyms
    - Language translations
  - *Index Terms*



THE CONTENT MODEL

Any Category in ICD is represented by:

1. ICD Concept Title
   - 1.1. Fully Specified Name
   - 1.2 Preferred Name
   - 1.3 Synonyms
2. Classification Properties
   - 2.1. Parents
   - 2.2 Type
   - 2.3. Use and Linearization(s)
3. Textual Definition(s)
4. Terms
   - 4.1. Base Index Terms
   - 4.2. Inclusion Terms
   - 4.3. Exclusions
5. Body System/Structure
   - 5.1. Body System(s)
   - 5.2. Body Part(s) [Anatomical Site(s)]
   - 5.3. Morphological Properties
6. Manifestation Properties
   - 6.1. Signs & Symptoms
   - 6.2. Investigation findings
7. Causal Properties
   - 7.1. Etiology Type
   - 7.2. Causal Properties - Agents
   - 7.3. Causal Properties - Causal Mechanisms
   - 7.4. Genomic Linkages
   - 7.5. Risk Factors
8. Temporal Properties
   - 8.1. Age of Occurrence & Occurrence Frequency
   - 8.2. Development Course/Stage
9. Severity of Subtypes Properties
10. Functioning Properties
    - 10.1. Impact on Activities and Participation
    - 10.2. Contextual factors
    - 10.3. Body functions
11. Specific Condition Properties
    - 11.1 Biological Sex
    - 11.2. Life-Cycle Properties
12. Treatment Properties
13. Diagnostic Criteria

Definitional representation of Disease

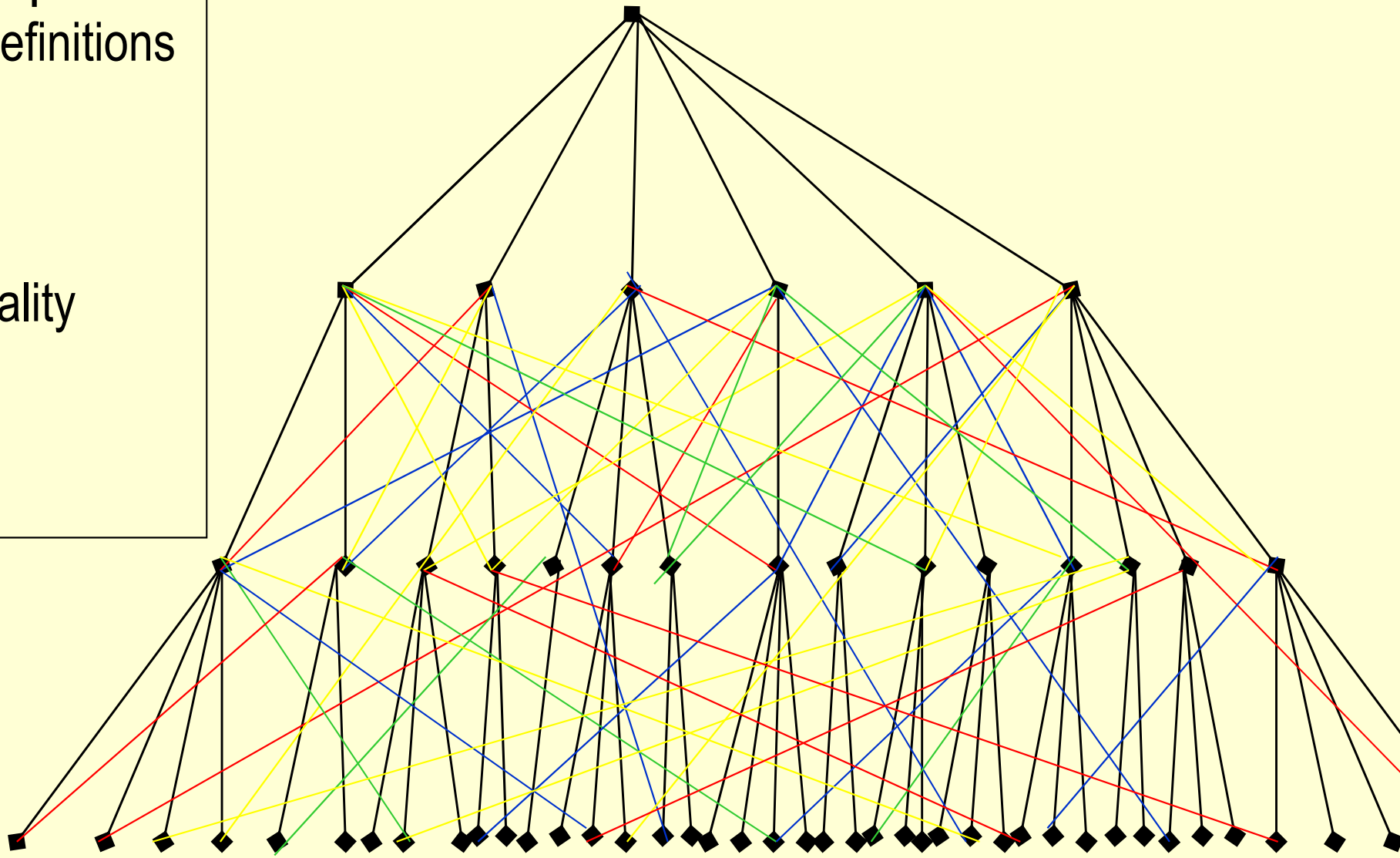
Lung Cancer

has_anatomy	Lung (all lobes), Bronchus
has_histology	Adenocarinoma, squamous, large cell, small cell, bronchioloalveolar
has_exclusion	mesothelioma
has_symptom	cough, hemoptysis, wheeze and stridor, dyspnea, and ostobstructive pneumonitis
has_finding	pleural effusion; radiologic nodules, cavitation destruction of tissue or bone

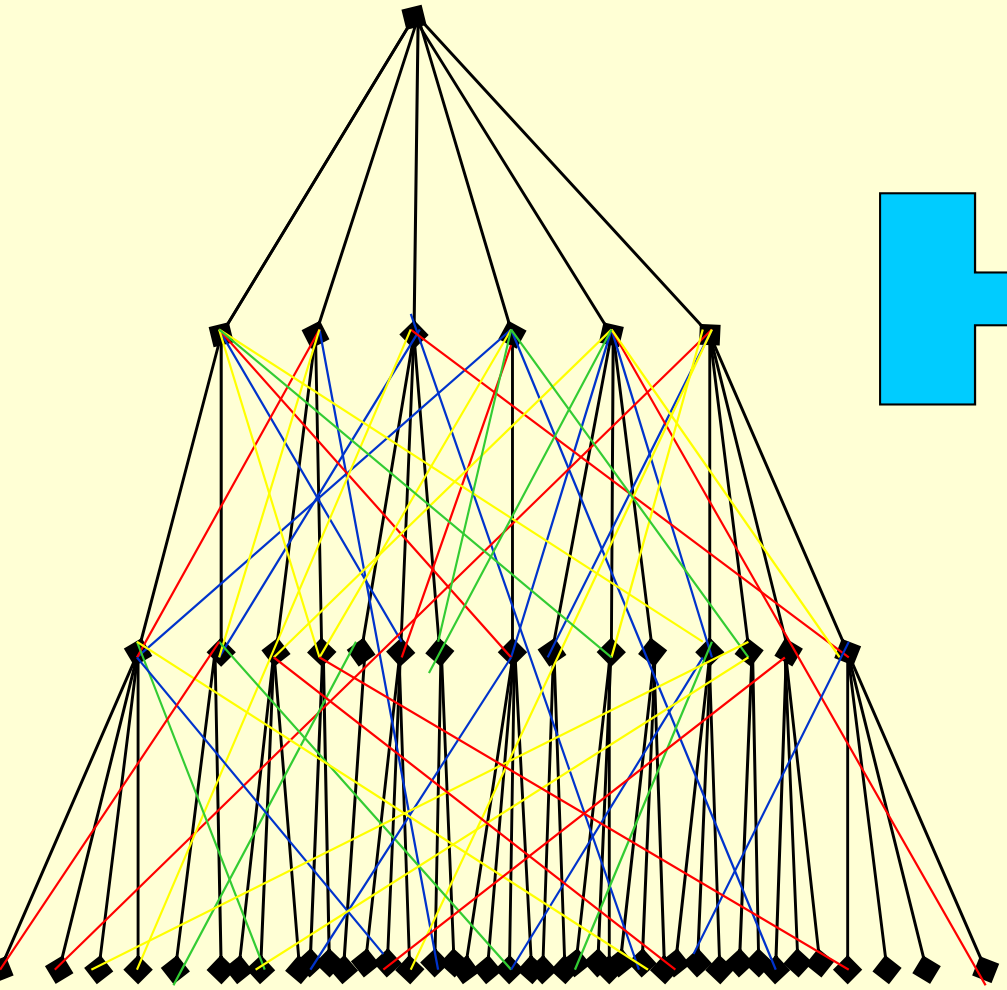
Addition of semantic arcs - Ontology

Relationships

- Logical Definitions
- Etiology
- Genomic
- Location
  - Laterality
- Histology
- Severity
- Acuity

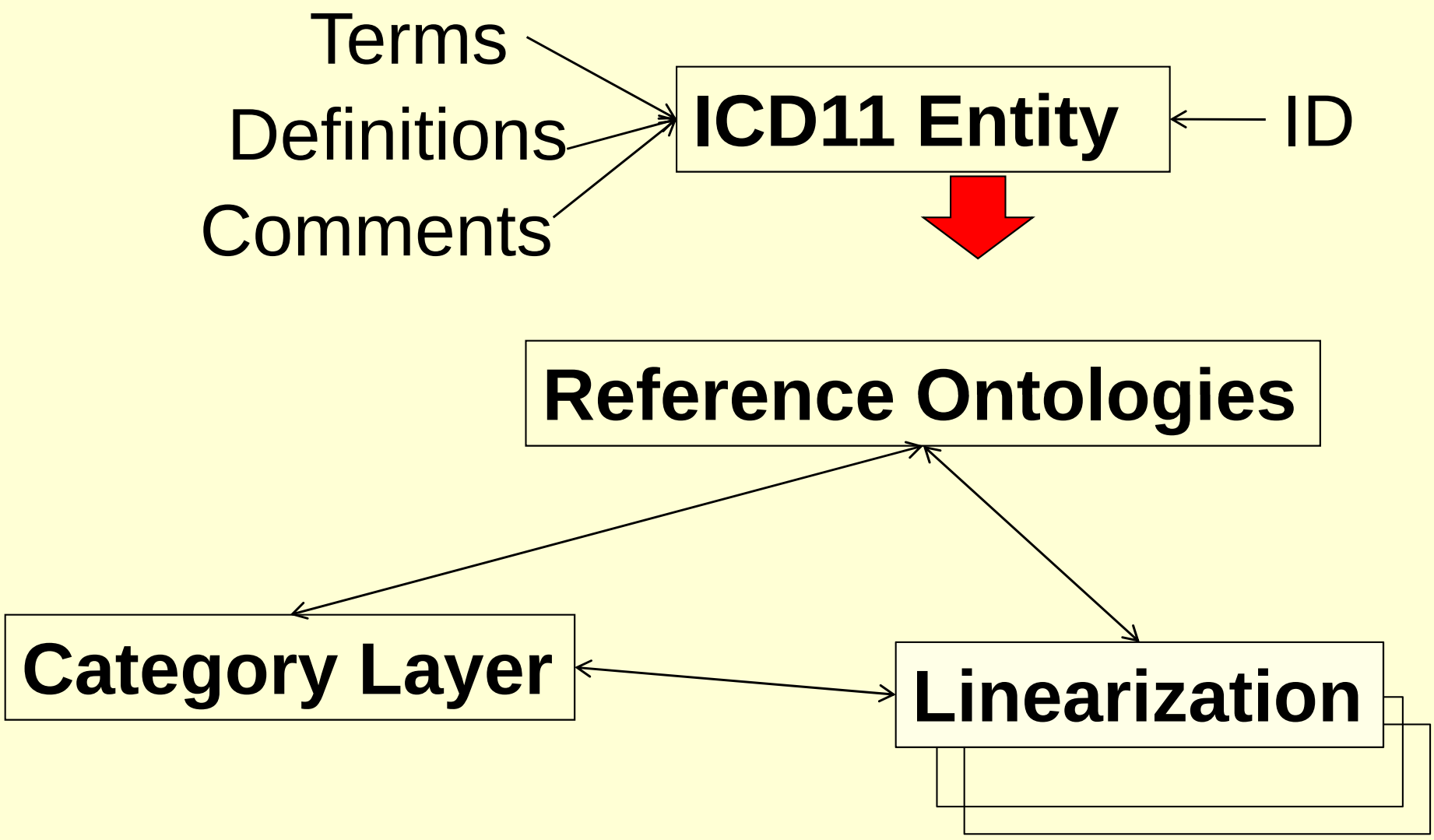


Serialization of "the cloud" Algorithmic Derivation

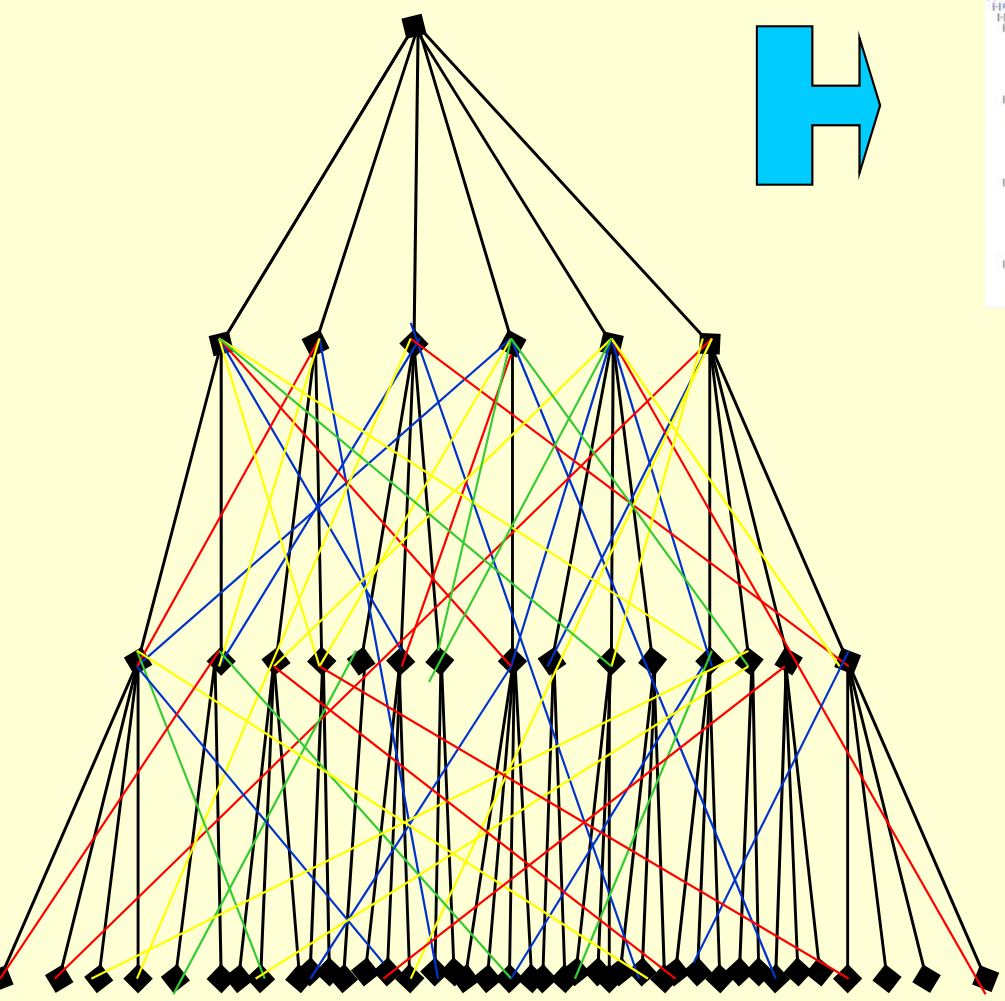


[–] ICD10WHO
[–] ICD10 Certain conditions originating in the perinatal period (CHAPTER XVI)(P00-P96)
[–] ICD10 Birth trauma(P10-P15)
[–] ICD10 Birth injury to peripheral nervous system(P14)
[+] ICD10 Birth injuries to other parts of peripheral nervous system(P14.8)
[+] ICD10 Birth injury to peripheral nervous system unspecified(P14.9)
[+] ICD10 Erb's paralysis due to birth injury(P14.0)
[+] ICD10 Klumpke's paralysis due to birth injury(P14.1)
[+] ICD10 Other brachial plexus birth injuries(P14.3)
[+] ICD10 Phrenic nerve paralysis due to birth injury(P14.2)
[–] ICD10 Birth injury to scalp(P12)
[+] ICD10 Birth injury to scalp unspecified(P12.9)
[+] ICD10 Bruising of scalp due to birth injury(P12.3)
[+] ICD10 Cephalhaematoma due to birth injury(P12.0)
[+] ICD10 Chignon due to birth injury(P12.1)
[+] ICD10 Epicranial subaponeurotic haemorrhage due to birth injury(P12.2)
[+] ICD10 Monitoring injury of scalp of newborn(P12.4)
[+] ICD10 Other birth injuries to scalp(P12.8)
[–] ICD10 Birth injury to skeleton(P13)
[+] ICD10 Birth injuries to other parts of skeleton(P13.8)
[+] ICD10 Birth injury to femur(P13.2)
[+] ICD10 Birth injury to other long bones(P13.3)
[+] ICD10 Birth injury to skeleton unspecified(P13.9)
[+] ICD10 Fracture of clavicle due to birth injury(P13.4)
[+] ICD10 Fracture of skull due to birth injury(P13.0)
[+] ICD10 Other birth injuries to skull(P13.1)
[–] ICD10 Intracranial laceration and haemorrhage due to birth injury(P10)
[+] ICD10 Cerebral haemorrhage due to birth injury(P10.1)
[+] ICD10 Intraventricular haemorrhage due to birth injury(P10.2)
[+] ICD10 Other intracranial lacerations and haemorrhages due to birth injury(P10.8)
[+] ICD10 Subarachnoid haemorrhage due to birth injury(P10.3)

High Level Structure – Core Model



Linear views may serve multiple use-cases
Morbidity, Mortality, Quality, …



iCAT ICD Collaborative Authoring Tool

ICD Categories

Create | Watch Branch | Search: I21 'Acute myocardial infarcti

- 09 IX Diseases of the circulatory system 38
- 10 X Diseases of the respiratory system 30
- 11 XI Diseases of the digestive system 1046
- 12 XII Diseases of the skin 3 2294
  - Draft for ICD-10 Revision update 1 108
  - Needing a decision to be made 1 16
  - To be retired 1 91
  - LA Infections and infestations affecting the skin 2 467
  - LB Inflammatory dermatoses 1 264
    - LB0 Dermatitis and eczema 1 48
    - LB1 Papulosquamous dermatoses 1 66
    - LB2 Urticaria and angio-oedema 1 17
    - LB3 Inflammatory erythemas and other reactive inflammato
    - LB4 Immunobullous diseases of the skin 1 36
      - LB40 Pemphigus 2 16
        - LB400 Pemphigus vulgaris 1 3
          - LB4000 Oral pemphigus 1
          - LB4004 Pemphigus vegetans 2
        - LB401 Pemphigus foliaceus 1
        - LB402 Endemic pemphigus foliaceus 2
        - LB404 Paraneoplastic pemphigus 2

Details for LB4 Immunobullous diseases of the skin

Title & Definition | Classification Properties | Terms | Clinical Description | Manifestation Properties

Causal Properties | Temporal Properties | Severity Properties | Functioning Properties

Specific Condition Properties | Treatment | Diagnostic Criteria | ICD 10 Notes and Hints

ICD 10 Code

Sorting label LB4

ICD Title Immunobullous diseases of the skin

Fully Specified Name

Short Definition

Text

A group of disorders characterized by the presence of circulating auto-antibodies directed against specific skin or mucous membrane antigens and resulting in blisters or erosions.

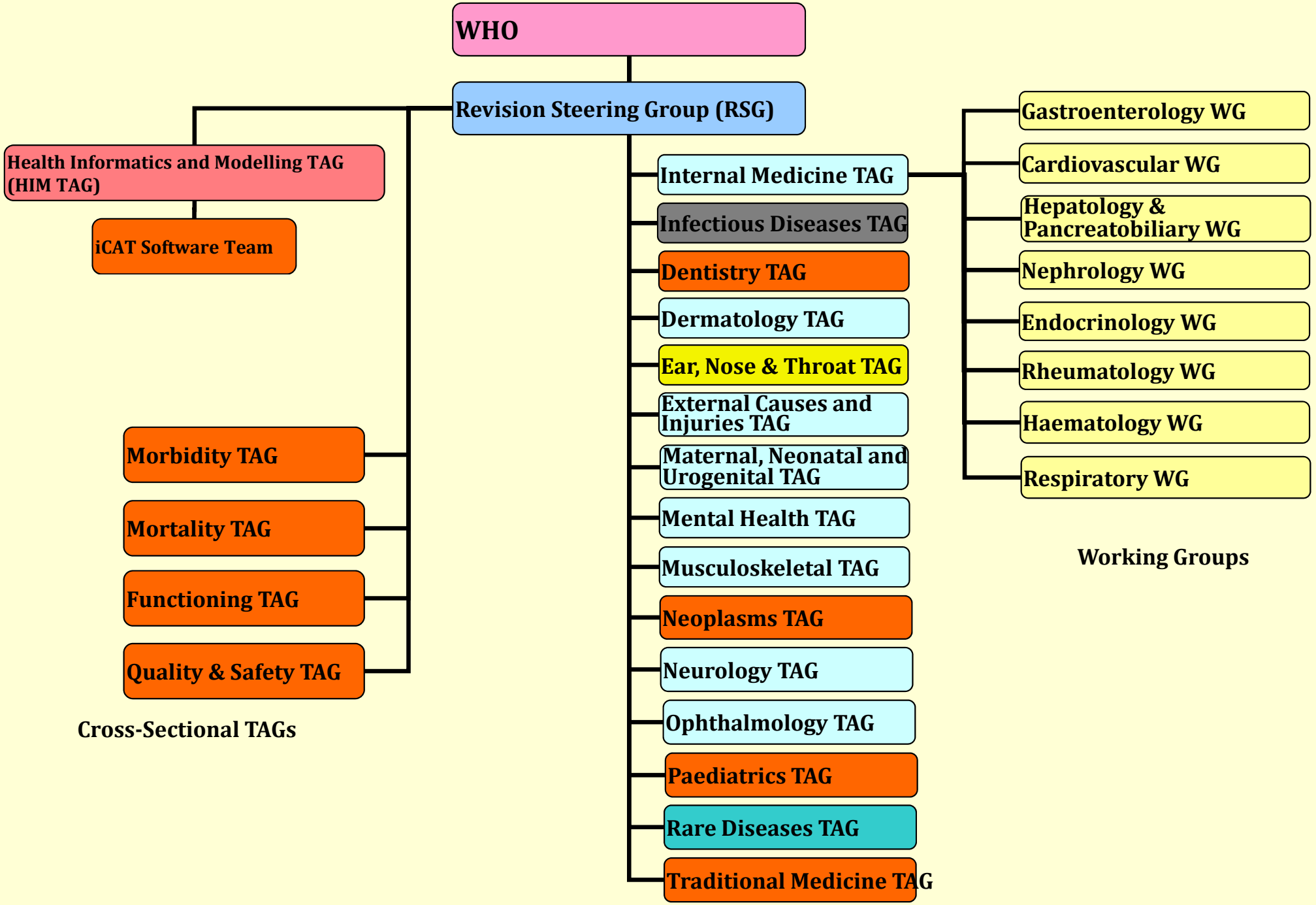
Detailed Definition

Text

Add new value

MAYO CLINIC
Biomedical Informatics

ICD Revision Organizational Structure



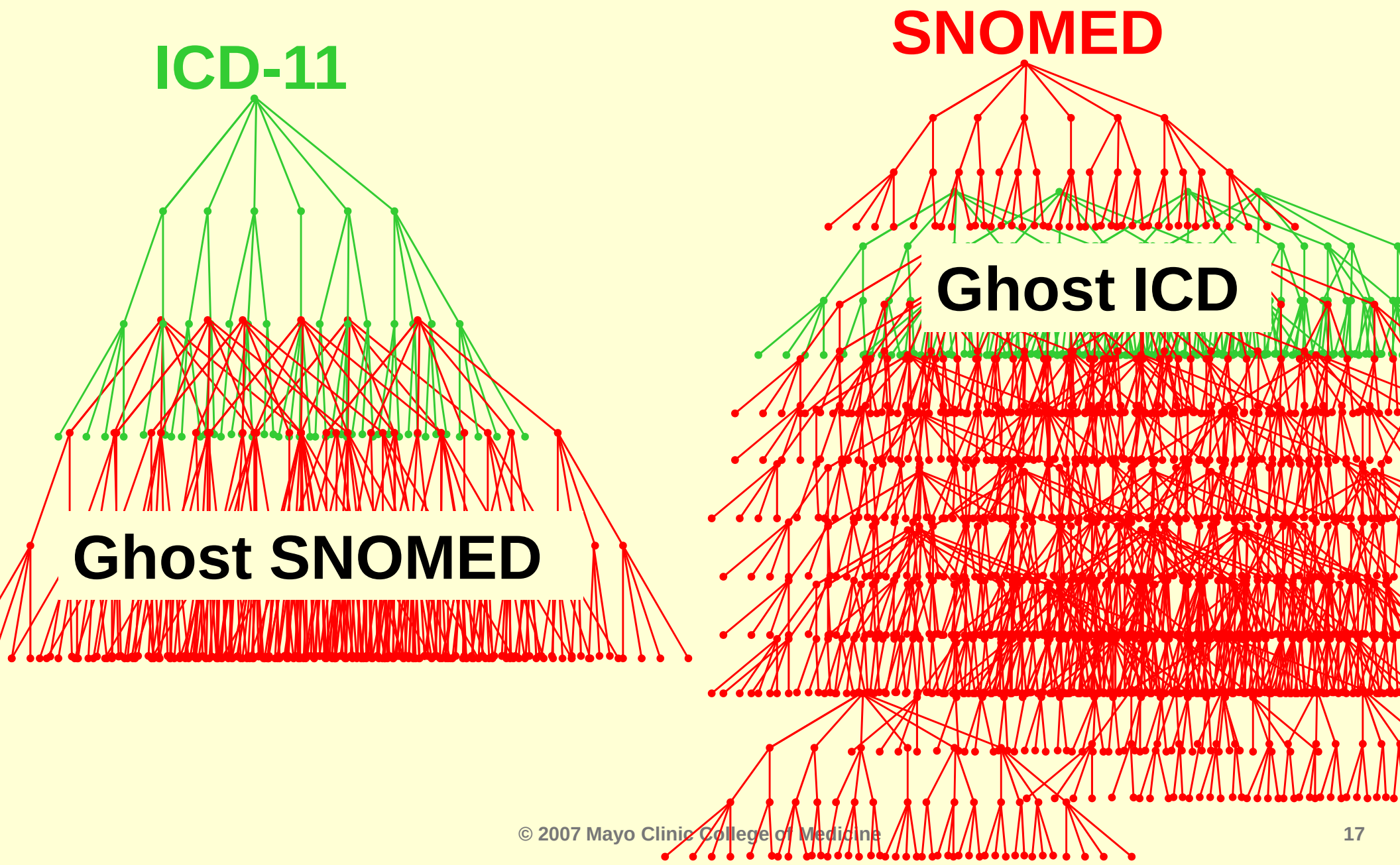
Wise Crowd-Sourcing
Wiki science?

- Acknowledge that the “full” ICD11 exceeds our resources
- Engage a hierarchy of expert talent to “wiki” proposals
  - Fill in details of models, extend depth
- Pre-process to manageable input
- Use TAGs as editorial boards
- Anticipate prototypes by fall 2011

Relationship with IHTSDO SNOMED content

- IHT (SNOMED) will require high-level nodes that aggregate more granular data
  - Use-cases include mutually exclusive, exhaustive,…
  - Sounds a lot like ICD
- ICD-11 will require lower level terminology for value sets which populate content model
  - Detailed terminological underpinning
  - Sounds a lot like SNOMED
- Memorandum of Agreement – July 2010!
  - WHO right to use for authoring and interpretation

Potential Future States

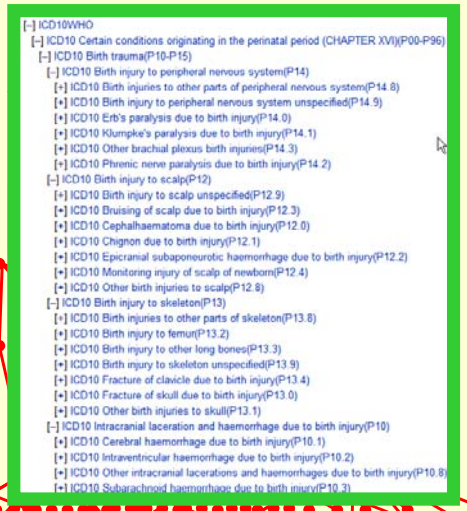


Joint Alternate Future Effort

ICD-11

ICD-IHTSDO

SNO



How should a Disease Taxonomy be Constructed?

- Phenomenological – symptomatic, syndromic?
- Logical assertions – pathophysiological facts?
- Etiological – causes and predispositions
  - At what level of granularity?
- Anatomically structured – what goes wrong where?
  - At what level of anatomy?
- Descriptive – random observations?
- Clinical – disease course and management?
- Genomic – disease polymorphisms?

Classification Structure

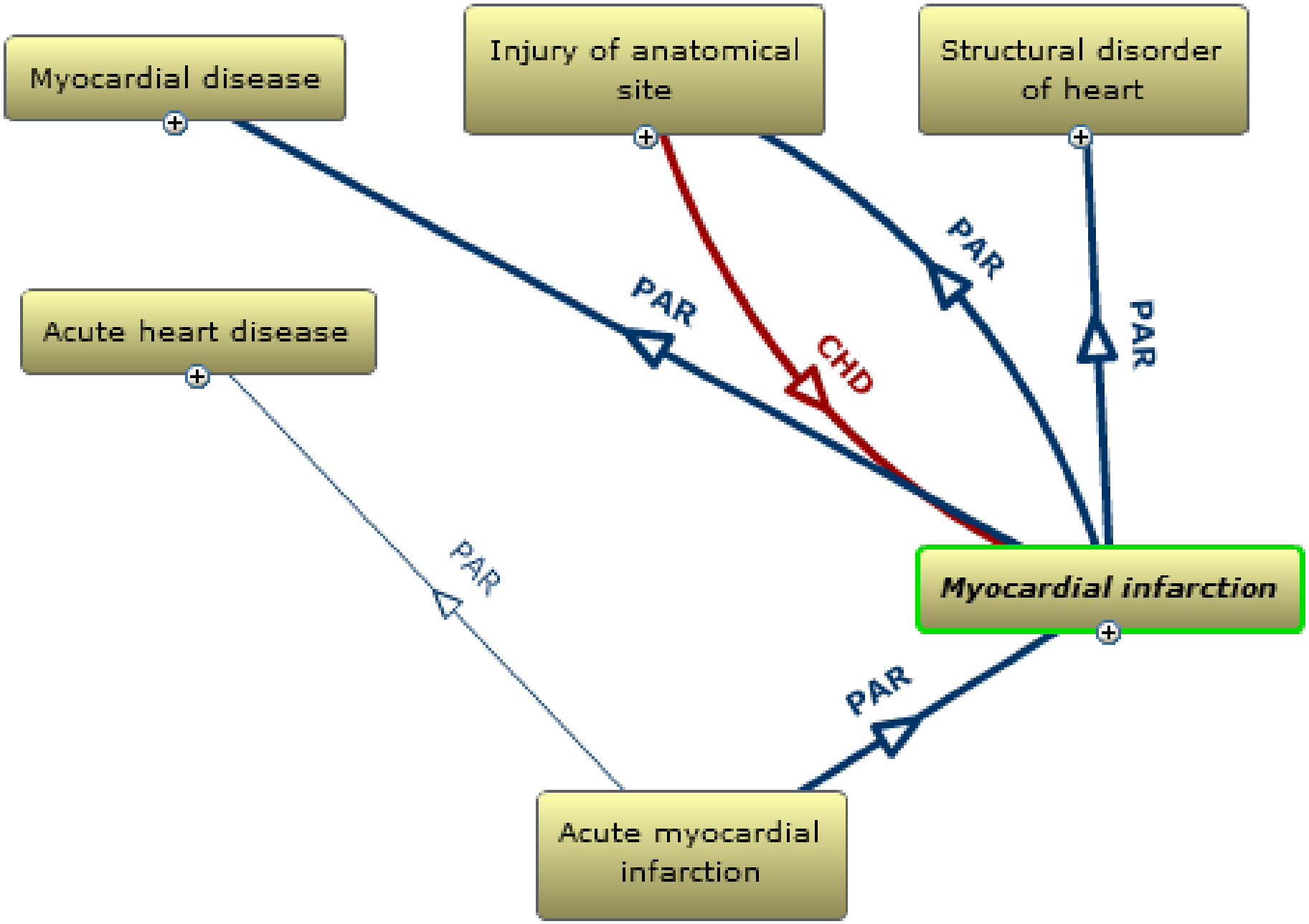
- Random list
- Structured Hierarchy
- Frame-structures
- Description Logics

Description Logics

- Basis for Modern Terminology Authoring
- A Computational Subset of Predicate Logic
  - Subset of First Order Predicate Logic (FOPR)
  - FOPR is *Not* Computationally Tractable
- Derived from "Frame Language" Heritage
- Always Balance Computational Tractability with Expressiveness
- Notation for DL ≠ FOPR
- OWL (Web Ontology Language) most popular dialect
  - Semantic evolution: XML → RDF → OWL

Logic assertions cannot handle probability

NCBO Browser View MI in SNOMED



Standardized diagnostic algorithm

WHO-ARIC-Olmsted Co

	Cardiac pain present		Cardiac pain absent	
Biomarker	**ECG**	**Diagnosis**	**ECG**	**Diagnosis**
Abnormal	Evolving Diag	Definite MI	Evolving Diag	Definite MI
	Diagnostic	Definite MI	Diagnostic	Definite MI
	Evolving ST-T	Definite MI	Evolving ST-T	Probable MI
	Equivocal	Definite MI	Equivocal	Suspect MI
	Absent	Probable MI	Absent	Suspect MI
Equivocal	Evolving Diag	Definite MI	Evolving Diag	Definite MI
	Diagnostic	Probable MI	Diagnostic	Suspect MI
	Evolving ST-T	Probable MI	Evolving ST-T	Suspect MI
	Equivocal	Suspect MI	Equivocal	Suspect MI
	Absent	Suspect MI	Absent	No MI
Incomplete	Evolving Diag	Definite MI	Evolving Diag	Definite MI
	Diagnostic	Suspect MI	Diagnostic	No MI
	Evolving ST-T	Suspect MI	Evolving ST-T	No MI
	Equivocal	No MI	Equivocal	No MI
	Absent	No MI	Absent	No MI
Normal	Evolving Diag	Definite MI	Evolving Diag	Definite MI
	Diagnostic	No MI	Diagnostic	No MI
	Evolving ST-T	No MI	Evolving ST-T	No MI
	Equivocal	No MI	Equivocal	No MI
	Absent	No MI	Absent	No MI

Gillum AHJ 1983, White Am J Epi 1989

electronic MEdical Records and GEnomics NHGRI eMERGE (U01) Goals

- GWAS – Genome Wide Association Study
  - 600k Affy chip
- High-throughput Phenotyping
  - Disease algorithm scans across EMRs
  - “catch up” with high throughput genomics
- Generalize Phenotypes across the Consortium
  - Measure reproducibility of algorithms among members
  - Vanderbilt, Northwestern, Marshfield, Group Health Seattle, Mayo

SHARP: Area 4: Secondary Use of EHR Data
A $15M National Consortium

Themes			Projects	Players
Data Normalization	Phenotype Recognition	Data Quality and Evaluation Frameworks	Clinical Data Normalization	IBM, Mayo, Utah, Agilex
			Natural Language Processing (NLP)	Harvard, Group Health, IBM, Utah ,Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado
			High-Throughput Phenotyping	CDISC, Centerphase, Mayo, Utah
			UIMA and Scaling Capacity	IBM, Mayo
			Data Quality	Mayo, Utah
			Evaluation Framework	Agilex, MN HIE, Mayo, Utah

ICD11 Use Cases

- Scientific consensus of *clinical phenotype*
- Public Health Surveillance
  - Mortality
  - Public Health Morbidity
- Clinical data aggregation
  - Metrics of clinical activity
  - Quality management
    - Patient Safety
  - Financial administration
    - Case mix
    - Resource allocation

Where is This Going?

- ICD is embracing a rich information model
  - Frame-based, hybrid assertions
- Explicit linkage to related ontologies is core to populating ICD11's Foundation Component
- Phenomenological description of phenotype should be undertaken with ontological elements
  - Raises questions about aggregation logics
  - May not be ideally suited to description logics
  - Must account for probabilistic information