



ICDR Conference

New Federal Applications of the ICF

Arlington, VA – July 11, 2007

Plenary Session V

ICF as a CHI-Endorsed Vocabulary: The Nuts and Bolts of ICF

ICF, UMLS, and NLM



Olivier Bodenreider

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

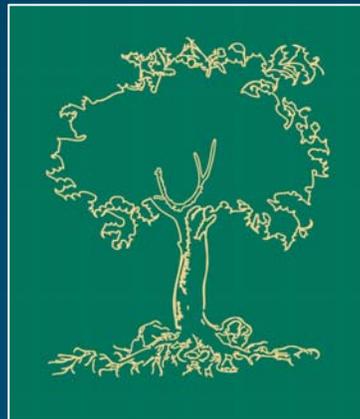
Outline

- ◆ Terminology integration
 - *The Unified Medical Language System*
- ◆ Experience with ICF
 - Term normalization
 - Semantic validation



Terminology integration in biomedicine

Unified Medical Language System



Motivation

- ◆ Started in 1986
- ◆ National Library of Medicine
- ◆ “Long-term R&D project”

«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- The first is **the variety of ways the same concepts are expressed** in different machine-readable sources and by different people.
- The second is the **distribution** of useful information among many disparate databases and systems.»





Unified Medical Language System

◆ SPECIALIST Lexicon

- 200,000 lexical items
- Part of speech and variant information

◆ Metathesaurus

- 5M names from over 100 terminologies
- 1.5M concepts
- 16M relations

◆ Semantic Network

- 135 high-level categories
- 7000 relations among them

Lexical
resources

Terminological
resources

Ontological
resources



Organize terms

- ◆ Synonymous terms clustered into a concept
- ◆ Preferred term
- ◆ Unique identifier (CUI)

Addison Disease	MeSH	D000224
Primary hypoadrenalism	MedDRA	10036696
Primary adrenocortical insufficiency	ICD-10	E27.1
Addison's disease (disorder)	SNOMED CT	363732003

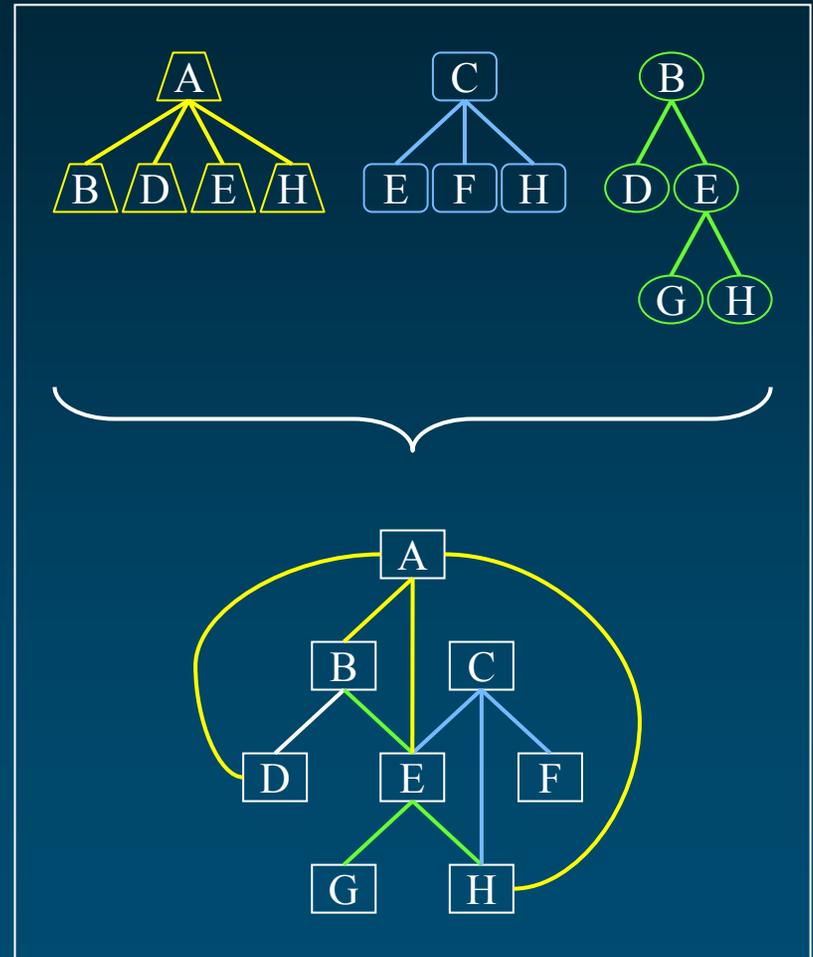
C0001403

Addison's disease



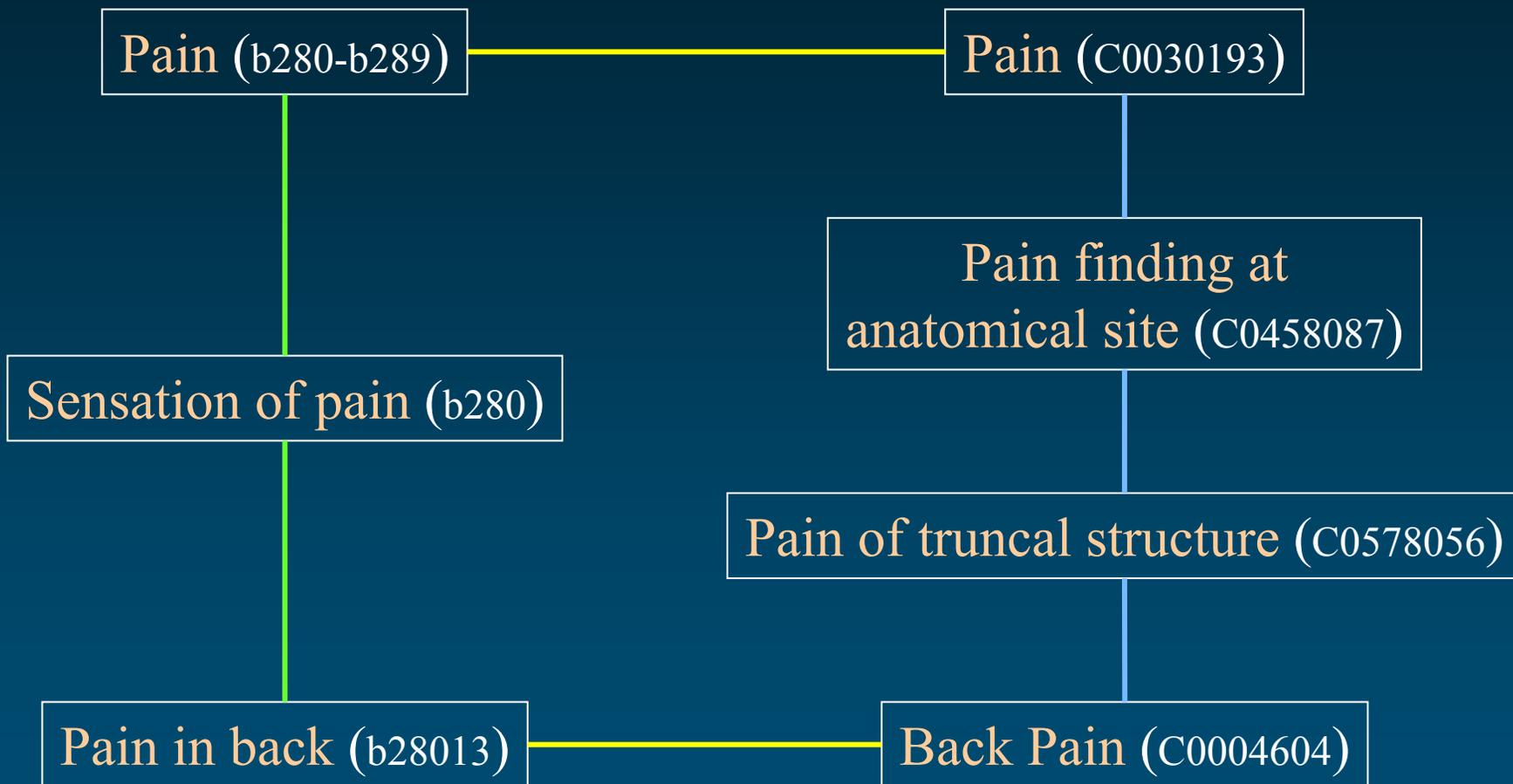
Organize concepts

- ◆ Inter-concept relationships: hierarchies from the source vocabularies
- ◆ Redundancy: multiple paths
- ◆ One graph instead of multiple trees (multiple inheritance)



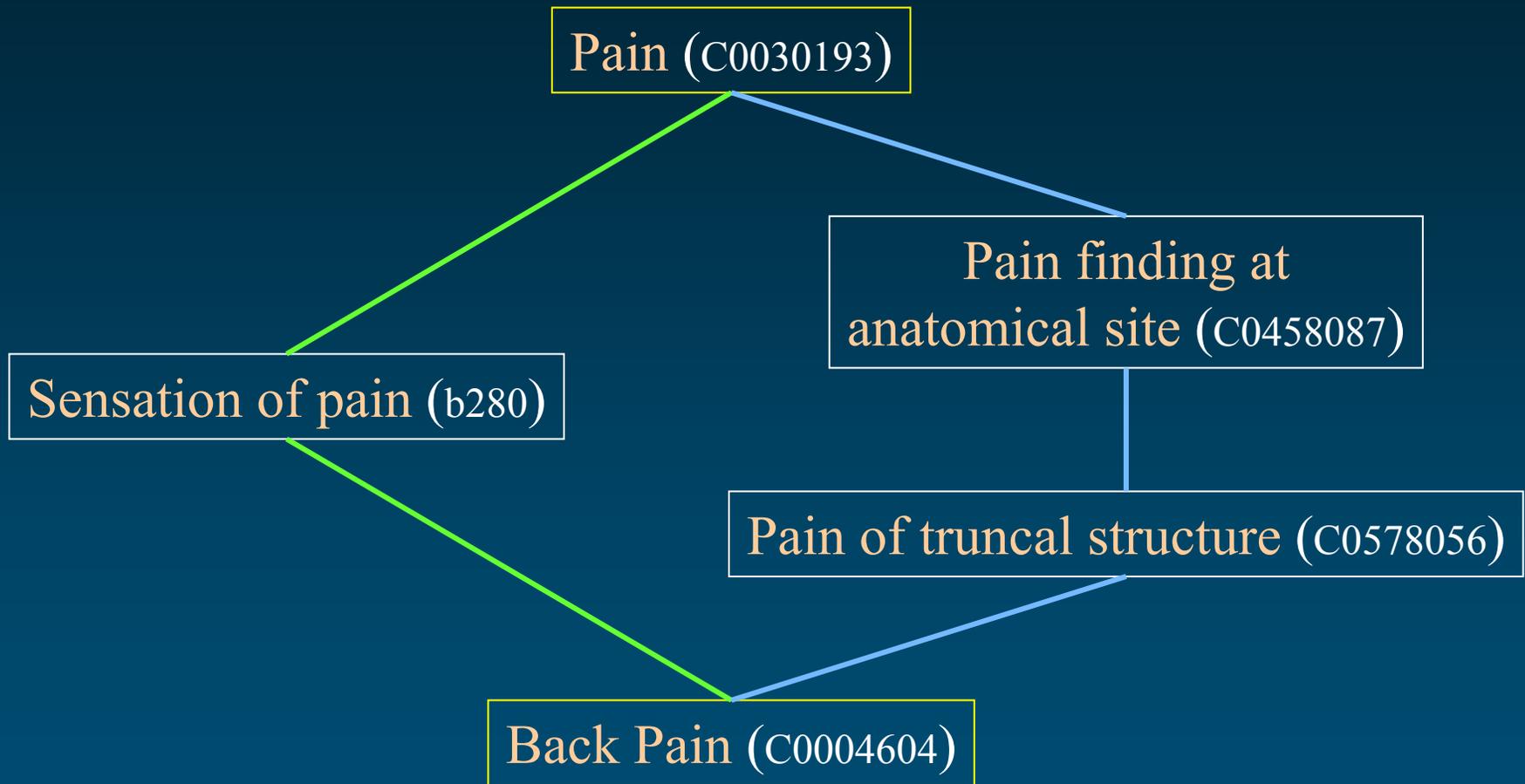


ICF relations in the Metathesaurus





ICF relations in the Metathesaurus





Source Vocabularies

(2007AB)

- ◆ 143 source vocabularies
 - 17 languages
- ◆ Broad coverage of biomedicine
 - 5.9M names
 - 1.4M concepts
 - 16M relations
- ◆ Common presentation

Biomedical terminologies

◆ General vocabularies

- anatomy (UWDA, Neuronames)
- drugs (RxNorm, First DataBank, Micromedex, ...)
- medical devices (UMD, SPN)

◆ Several perspectives

- clinical terms (SNOMED CT)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- data exchange terminologies (HL7, LOINC)

Biomedical terminologies (cont'd)

- ◆ Specialized vocabularies
 - nursing (NIC, NOC, NANDA, Omaha, PCDS)
 - dentistry (CDT)
 - oncology (NCI Thesaurus, PDQ)
 - psychiatry (DSM, APA)
 - adverse reactions (COSTART, WHO ART, MedDRA)
 - primary care (ICPC)
 - genomics (Gene Ontology, HUGO, OMIM)
- ◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)

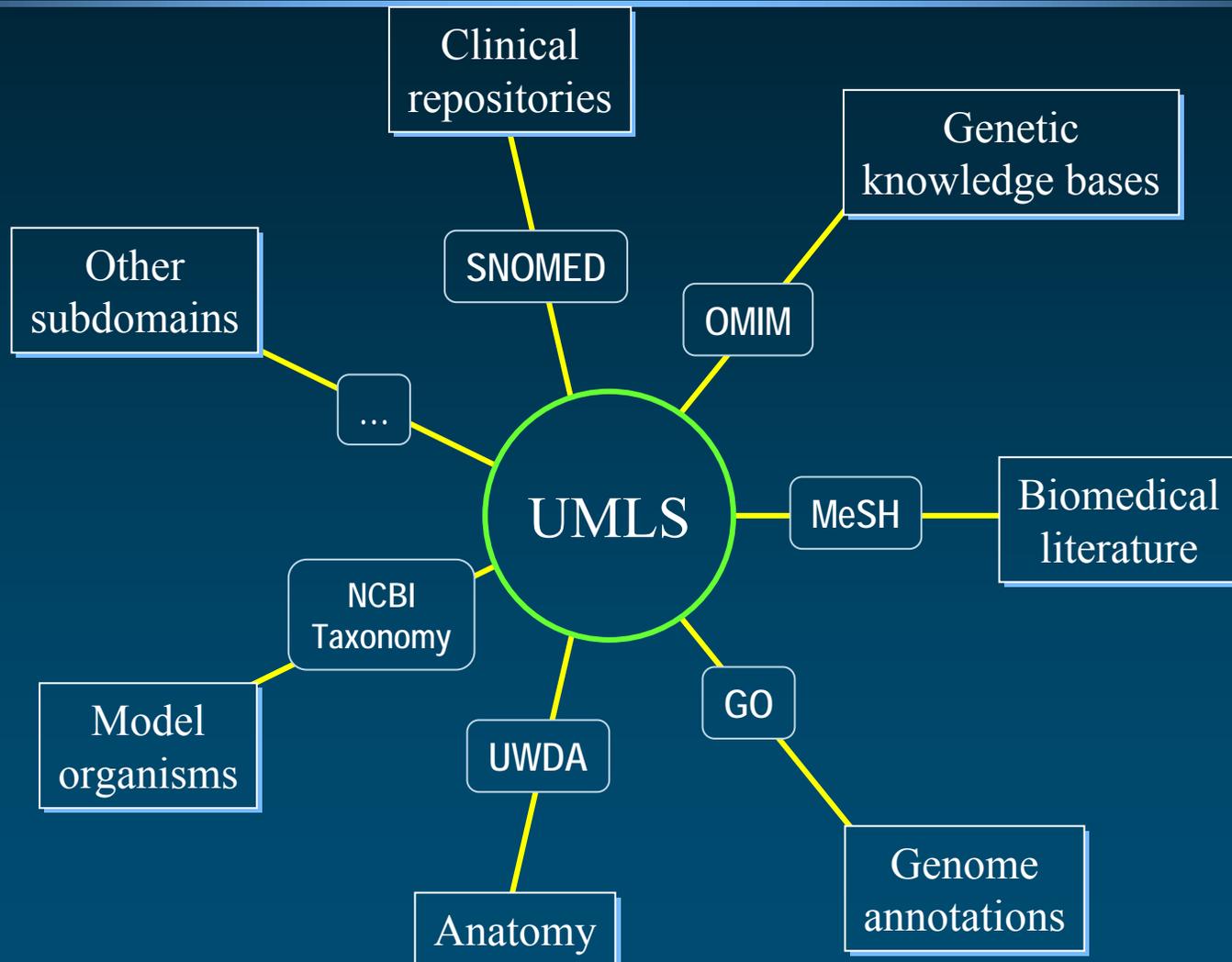
CHI standards in UMLS

(2007AB)

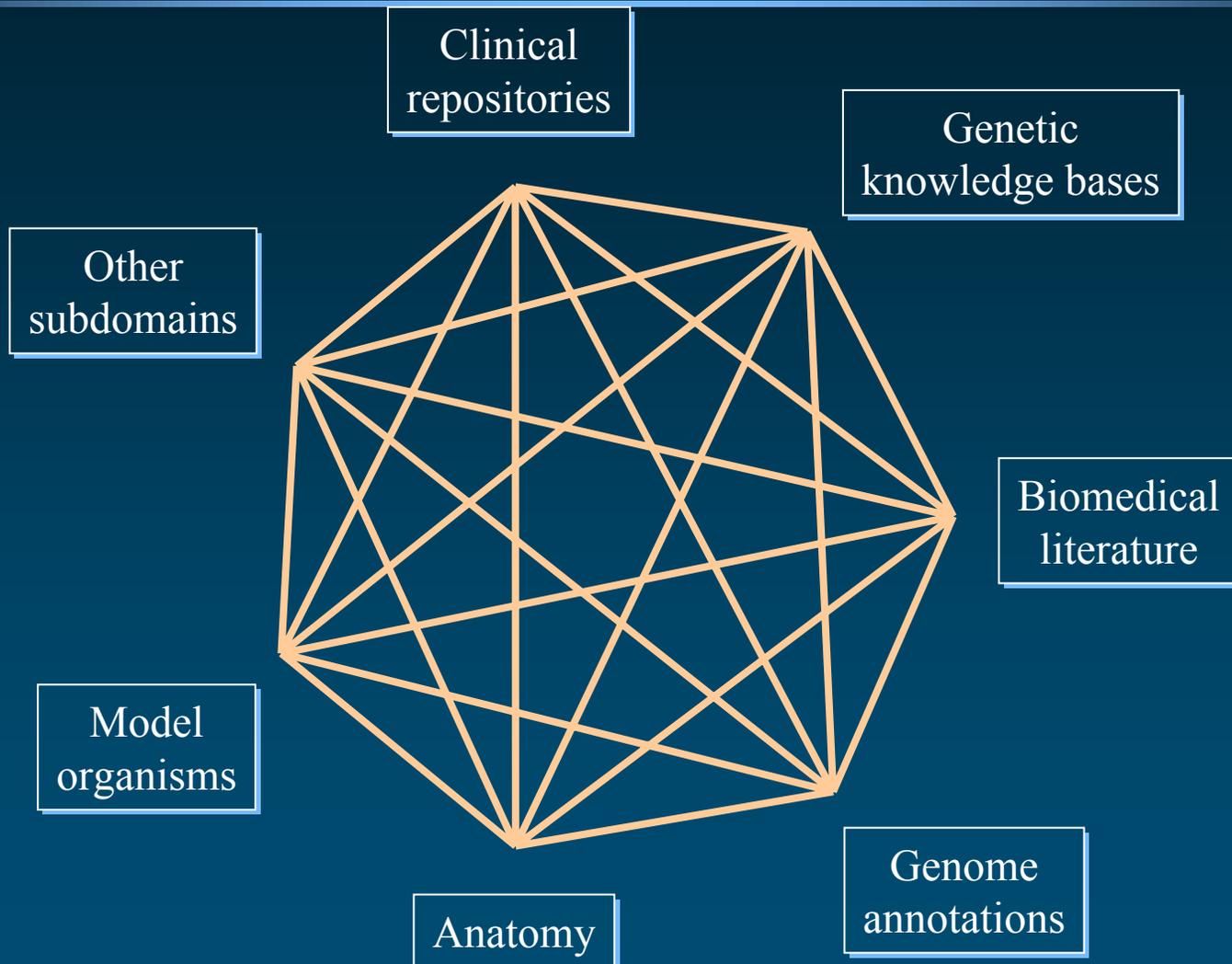
- ◆ CPT-4
 - Current Procedural Terminology
- ◆ HCPCS
 - Healthcare Common Procedure Coding System
- ◆ ICD-9-CM
 - International Classification of Diseases / Clinical Modifications
- ◆ LOINC
 - Logical Observation Identifier Names and Codes
- ◆ SNOMED Clinical Terms
- ◆ Others: NDF-RT, RxNorm, NCI Thesaurus, HUGO, ...



Integrating subdomains



Integrating subdomains



Experience with ICF

Acknowledgments



- ◆ Marcy Harris
- ◆ Guergana Savova

Materials

- ◆ ICF: 1495 terms
 - 478 terms filtered out
 - 218 terms with *other specified*
 - 217 terms with *unspecified*
 - 37 terms with *other specified* and *unspecified*
 - 2 terms with *specified* (alone)
 - 1 term with *other specified* (alone)
 - 1017 terms remaining
- ◆ UMLS: version 2004AA

Mapping to UMLS Metathesaurus

◆ Methods

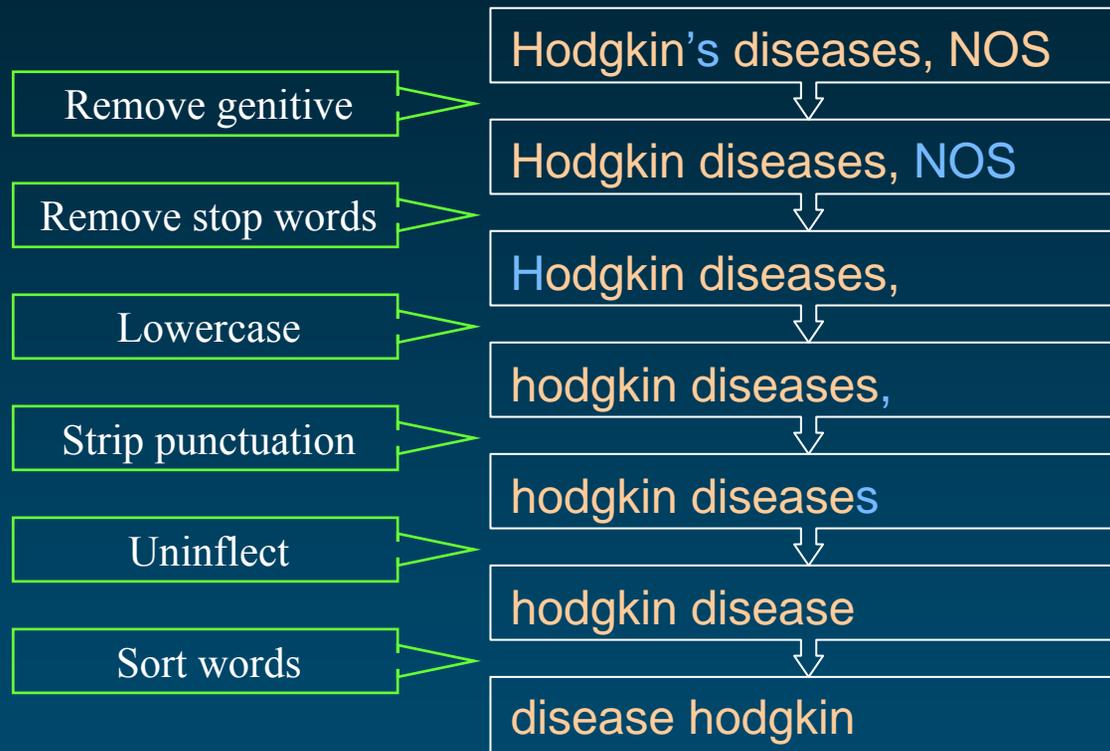
- Exact match first
- Normalized match, if necessary

◆ Results

- 359 ICF terms mapped (35%)
- 658 terms without mapping



Normalization





Integrating terms Examples

◆ Exact match (including to synonyms)

- Original term: Pain in back (b28013)
- Concept mapped to: Back Pain (C0003862)

Pain in back present in the Metathesaurus (from the Read Codes)

◆ Match after normalization

- Original term: Pain in joints (b28016)
- Normalized term: joint pain
- Concept mapped to: Arthralgia (C0003862)

Joint pain is a synonym for Arthralgia



Integrating terms Examples

◆ No match found

- Radiating pain in body part (b2801) *Too general*
- Radiating pain in a dermatome (b2803) *Too specific*
- Pain in stomach or abdomen (b28012) *Coordination*

e215	<u>Population</u>	(→ C0032659)
e2150	Demographic change	(→ C0681668)
e2151	Population density	(→ C0032665)
e2158	<u>Population</u> , other specified	
e2159	<u>Population</u> , unspecified	





Integrating terms Examples

◆ Multiple matches

- Impulse control (b1304)
 - Impulse control (C0150632)
 - Impulse control training (C0262701)
 - Ability to control impulses (C0517616)

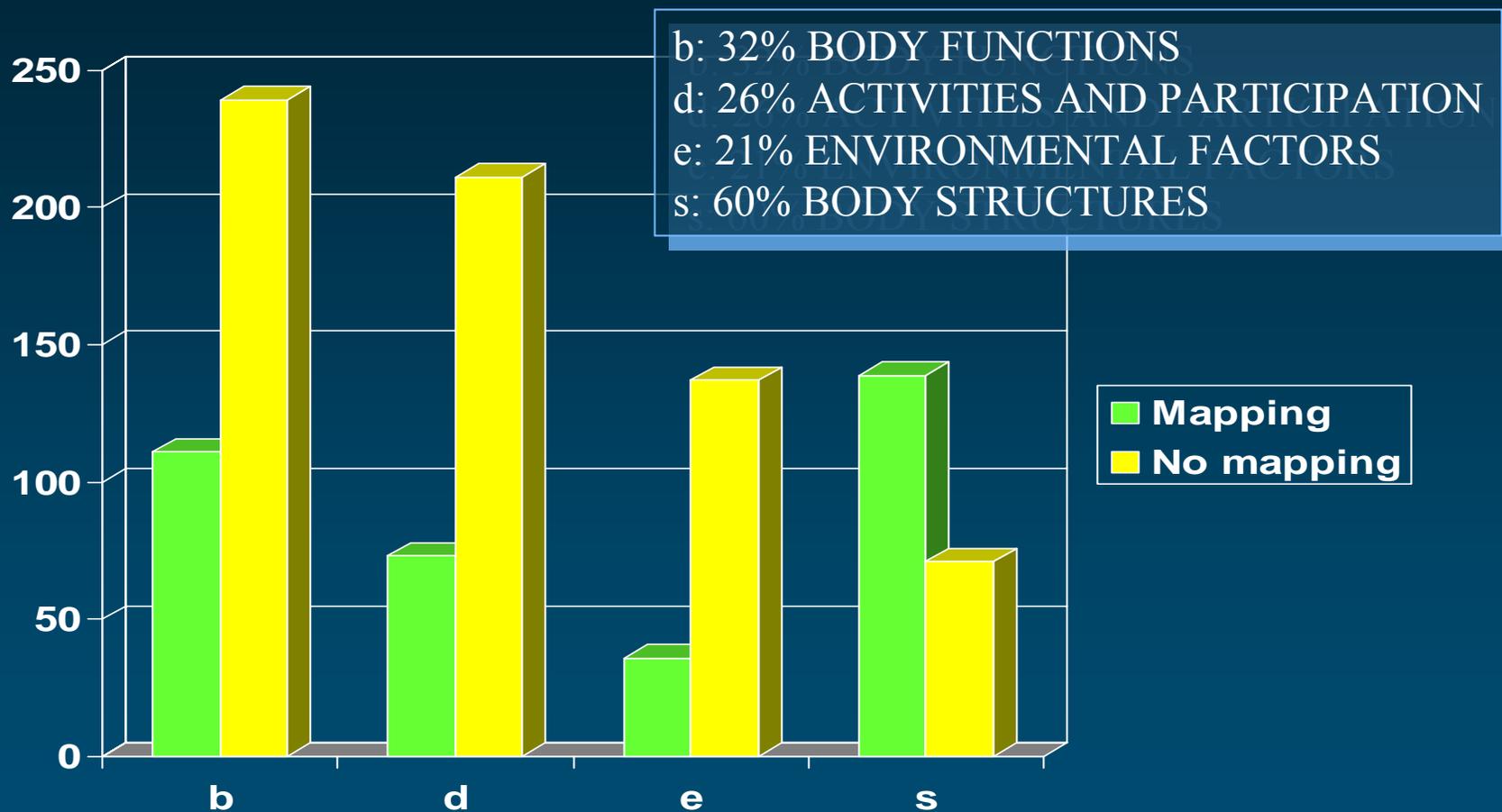
- Frontal lobe (s11000)
 - frontal lobe (C0016733)
 - Entire frontal lobe (C1268977) } *SNOMED CT distinction*

- Bites (b5101)
 - Biting (C0005658)
 - 2-(4-ethoxybenzyl)-1-diethylaminoethyl-5-isothiocyanatobenzimidazole (C0045724)

synonym for BIT alkylating agent



Mapping by category



Issues with mapping

- ◆ Phenomena preventing the terms from being mapped:
 - coordination with *and* alone: 147
 - Education and training policies (e5852)
 - coordination with *or* alone: 7
 - Pain in stomach or abdomen (b28012)
 - coordination with both *and* and *or*: 2
 - Assistive products and technology for the practice of religion or spirituality (e1451)

Semantic validation

◆ Method

- Correspondence between ICF chapters and UMLS semantic types/groups

◆ Issues

- Correspondence difficult to establish for some subgroups in ENVIRONMENTAL FACTORS
 - PRODUCTS AND TECHNOLOGY
 - ATTITUDES



Semantic areas in ICF

- ◆ b BODY FUNCTIONS
 - **Physiology**
 - Sign or Symptom
 - Finding
 - Biologic Function
 - Individual Behavior
- ◆ s BODY STRUCTURES
 - **Anatomy**
- ◆ d ACTIVITIES AND PARTICIPATION
 - **Physiology**
 - **Activities & Behaviors**
 - Machine Activity
 - Sign or Symptom
 - Finding
 - Educational Activity

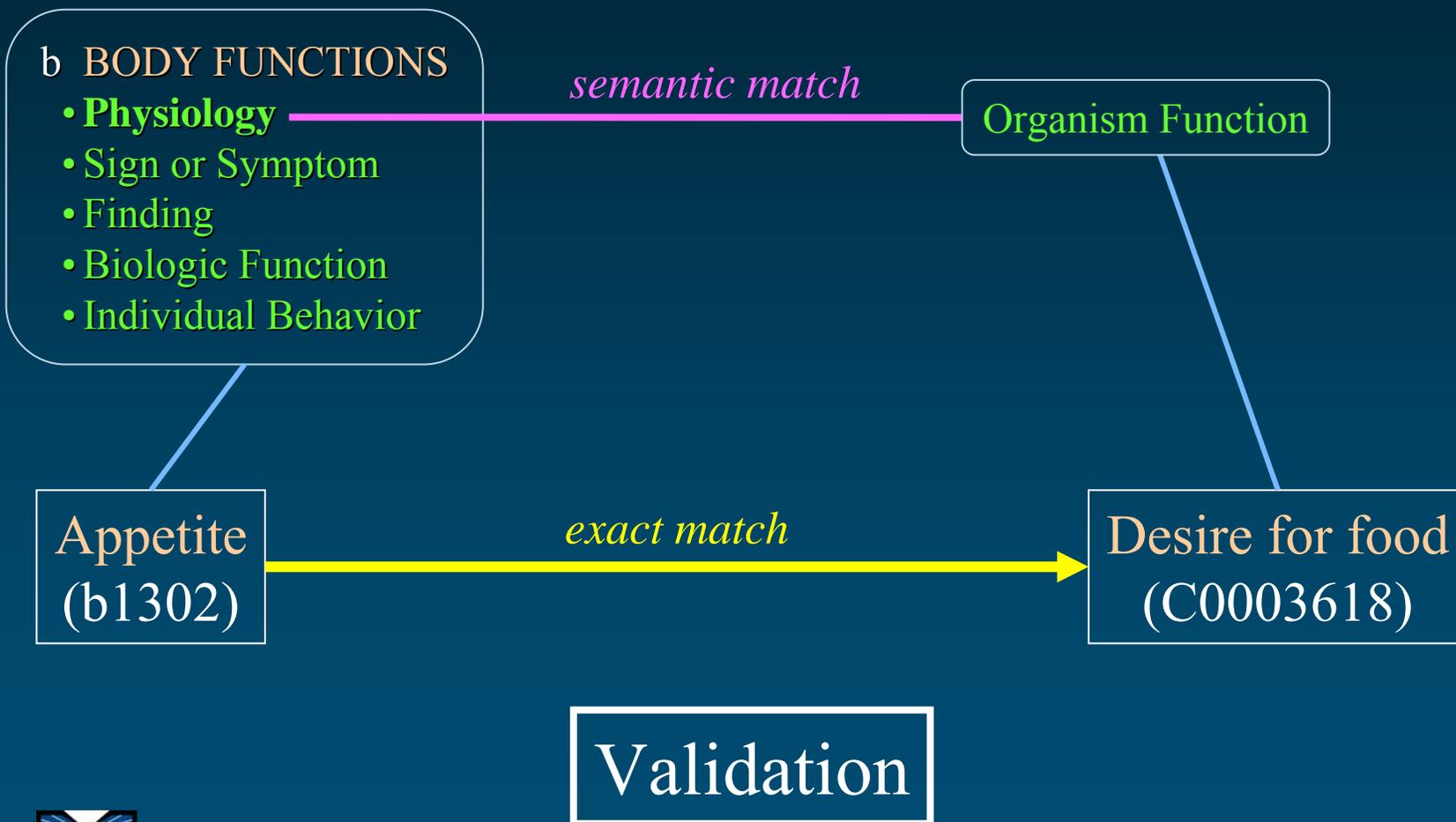


Semantic areas in ICF

- ◆ e1 PRODUCTS AND TECHNOLOGY
 - ????
- ◆ e2 NATURAL ENVIRONMENT AND HUMAN-MADE CHANGES TO ENVIRONMENT
 - **Phenomena**
- ◆ e3 SUPPORT AND RELATIONSHIPS
 - Family Group
 - Population Group
 - Professional or Occupational Group
- ◆ e4 ATTITUDES
 - ????
- ◆ e5 SERVICES, SYSTEMS AND POLICIES
 - Governmental or Regulatory Activity
 - Regulation or Law

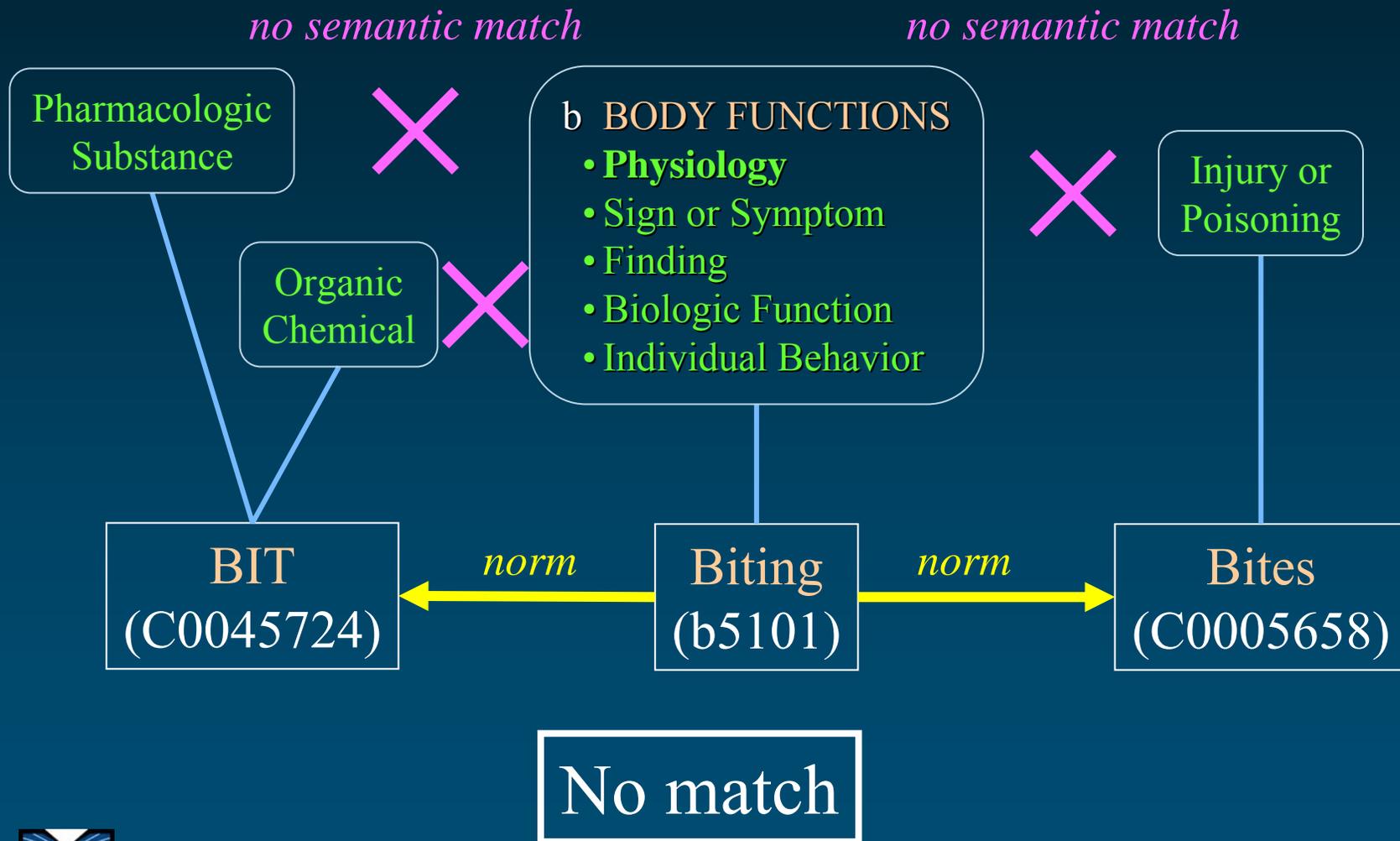


Semantic pre-processing Examples



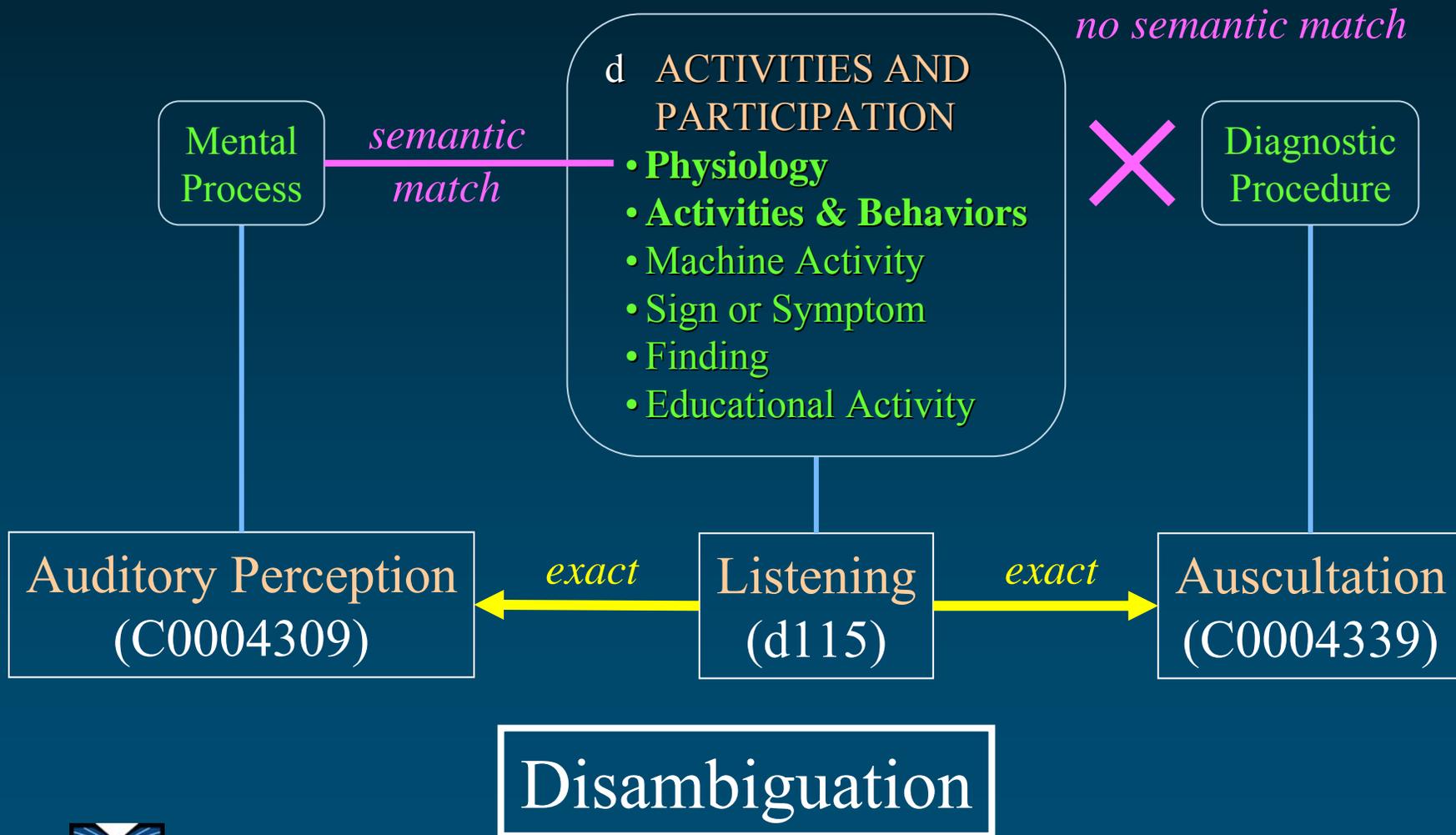


Semantic pre-processing Examples

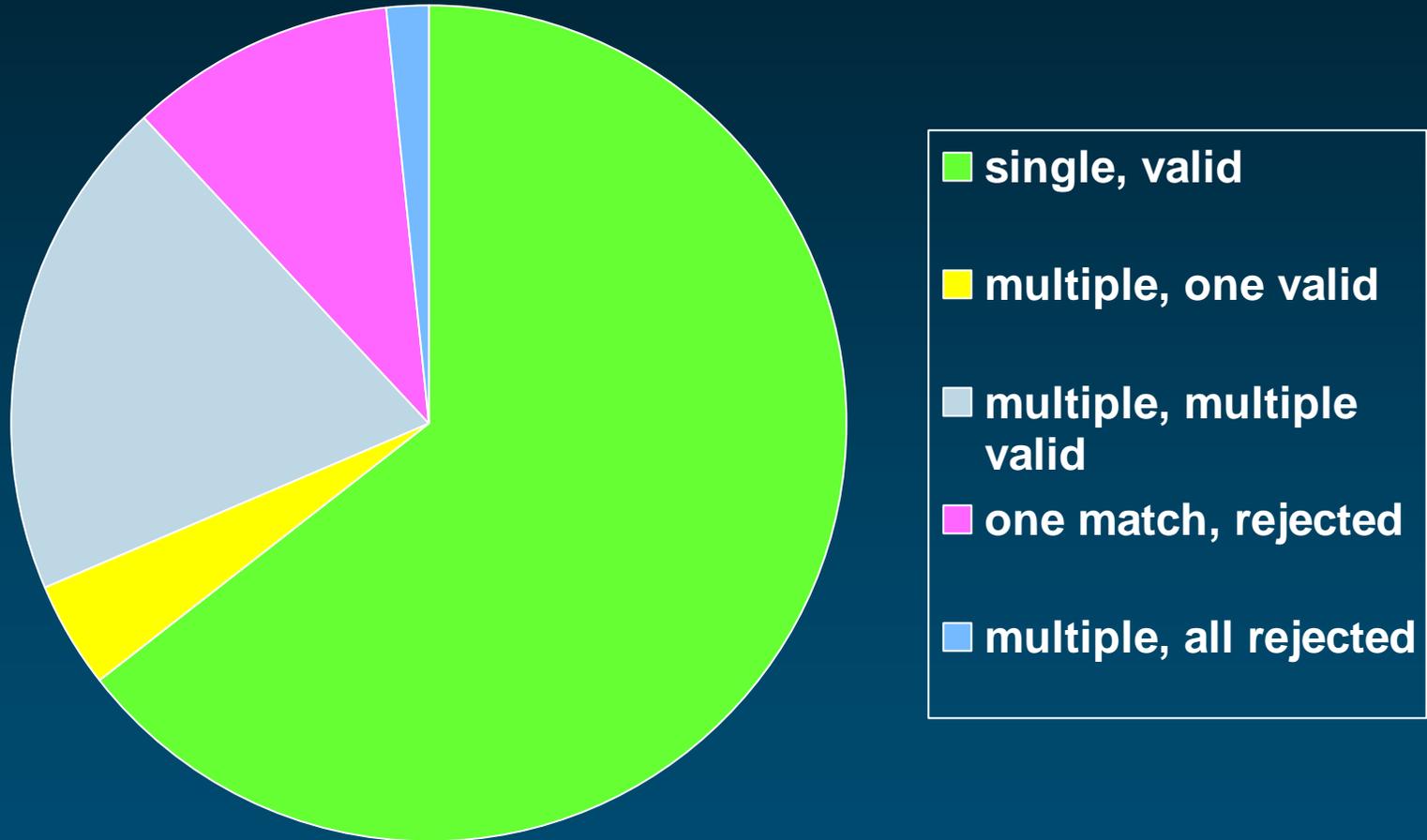




Semantic pre-processing Examples



Semantic validation Results



Issues with semantic validation

- ◆ Multiple “valid” matches must be reviewed by experts and disambiguated
- ◆ Rejected mappings
 - Semantically invalid UMLS concepts
 - or
 - Missing correspondence (ICF chapter/UMLS ST-SG)

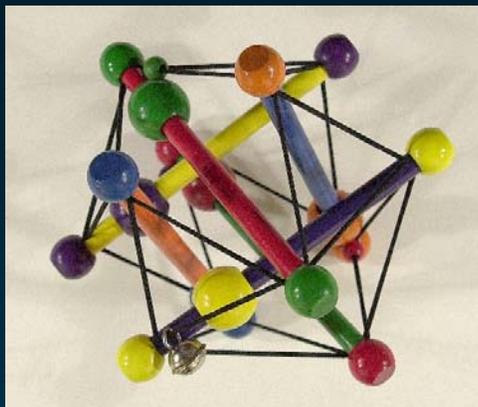
Conclusions

Conclusions (1)

- ◆ Integrating ICF into the UMLS
 - Should not be too difficult
 - Relatively small
 - Many concepts already present in UMLS
 - Challenges
 - Underspecified terms
 - Coordination
 - Specific perspective

Conclusions (2)

- ◆ Integrating ICF into the UMLS
 - Benefit for ICF
 - Links to other vocabularies
 - Facilitate downward extension
 - Benefit for UMLS
 - Adds specific perspective



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

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