Medical Data,

Standard Vocabularies,

Communication Standards

6.872/HST950

Lecture #8

(with some material from Chris Cimino)

Harvard-MIT Division of Health Sciences and Technology
HST.950J: Medical Computing
Peter Szolovits, PHD

Recall Children's Clinicians Workstation Database

☐　　Demographics

☐　　Problems

☐　　Allergies

☐　　Medications

　　　　Immunizations

☐　　Lab Data

☐　　Clinical Measurements

　　　　Growth Charts

☐　　Visit History

☐　　Reports and Letters

Vocabularies and Terminology

☐ **Why?**

Surrogate for messy reality

Uses

☐ **How?**

Flat list

Taxonomy

Heterarchy

Combinatorial Language

  ☐ Derivation rules

  ☐ Inference

  ☐ knowledge representation

Definitions

☐	Word a set of characters including punctuation delimited by white space.

☐	Term one or more words used as a unit.

☐	Concept an idea, action, or thing.

☐	Synonym two terms for the same concept.

Vocabulary Uses

☐ Indexing Finding what you want

☐ Cataloging Putting away what you have

    E.g., WHO, DRGs

☐ Knowledge Representation

    Representing the facts

    Blurring the facts

    Creating new shades of meaning

Describe a term for aLaboratory Test.

☐ Where was it done?

☐ How was it done?

☐ Under what conditions was it done?

☐ How many minutes after eating carbohydrate was it measured?

Describe a Vocabulary for a Gene

☐   Whose gene?

☐   Gene fragment?

☐   Open Reading Frame?

☐   Promoter + all exons and introns

☐   Promoter + all exons + all introns + other binding sites affecting function?

☐   Final/draft/species/SNP/Alternative splicing?

Knowledge vs. Language

☐ Get two or more people to enumerate terms to describe the same set.

Do any terms match exactly?

Do terms differ by word order?

Do terms differ by word suffix or prefix?

Are there terms that some people think are synonyms that other people think are not?

History of 3 Vocabularies

☐ MeSH Index

☐ ICD Precoodinated

☐ SNOMED Post-coordinated

History

☐   The modern history of medical controlled vocabularies begins with the U.S. Army General Surgeon who petitioned Congress to fund a medical library.

☐   The position eventually became The US Surgeon General and the library the National Library of Medicine

http://www.nlm.nih.gov/

History

☐   Library collection was indexed with Index Medicus (created by NLM) which is published in book form.

☐   Index Medicus was extended to index medical

literature articles.

☐   Index Medicus was extended further to provide on-line indexing (1960). This became the Medical Subject Headings (MeSH).

MeSH

☐   Purpose is to index the medical literature.

☐   Content of MeSH is driven by publications.

☐   Who ¡°owns¡± MeSH?

☐   What impact do vocabulary changes have?

MeSH Structure

☐ MeSH is organized into a series of trees (e.g. physical findings, diseases, chemicals)

☐ A MeSH main heading is a concept. (e.g. Neurologic Disease, Epilepsy)

☐ Main Heading (MH) is often called a term. Try to avoid this.

MeSH Structure

☐   Each MH has a unique identifier.

☐   Each MH may have multiple synonyms.

☐   Each MH may have multiple locations in multiple trees. Each of these contexts has a unique tree address. The concept of context is synonymous with multiple inheritance.

MeSH Structure

☐ 	There is a small set of subheadings (50) that modify MH based on tree address. (e.g. diagnosis applies to MH in the Disease tree but not to the Chemical tree).

☐ 	There is a small set of tag terms (15) which exist unrelated to the rest of MeSH. (e.g., Review Article, Human, Animal)

MeSH Structure

☐	Every article is indexed with tag terms.

☐	Every article is indexed with MH terms for focus (main index term) and mention (minor index term).

☐	Every index term is checked for subheadings.

☐	This is all done by trained reviewers.

☐	The MeSH Vocabulary is revised annually.

International Classification of

Disease (ICD)

☐   Any agency that dispenses funds for health care needs a way to assess needs and effectiveness.

☐   The United Nations World Health Organization (WHO) funds health care prevention projects world wide and gathers statistics for member nations.

☐   Who owns ICD?

☐   What impact will changes have?

ICD Structure

☐   ICD is divided into categories based on a 5-digit numeric code. (e.g., 133.21)

☐   Usually round numbers are more general concepts (e.g., 100 subsumes 130 which subsumes 133)

☐   The fourth and fifth digit is called a modifier but it isn't really.

ICD Structure

☐   The code is both the concept and the unique identifier. Multiple terms are linked to the same code.

☐   Every patient is coded with as many terms as possible.

☐   Terms should be the most specific one to describe a particular problem.

ICD Structure

☐   Coding scheme limits the size of the vocabulary.

☐   Obsolete codes must be reused.

☐   Base ten results in limited flexibility and the need for other, NOS, and NOC terms.

ICD Structure

☐　　　Lack of multiple contexts or multiple inheritance results in duplicate terms.

☐　　　Lack of overall organization results inambiguous terms.

ICD Structure

☐	ICD has been adopted by most insurance companies as a method for controlling billing and payment.

☐	Economic forces drive how the vocabulary is used which drives how ICD is modified which drives changes in reimbursement which drives how the vocabulary is used

☐	Who owns ICD?

☐	The Vocabulary is revised sporadically.

SNOMED Structure

☐   Developed by the American College of Pathologists to overcome the faults of ICD.

☐   Really describes 6 [now 12] different vocabularies, one for each axis of a concept (e.g., anatomy, environment, history).

☐   Every concept is built up from a term from each axis (e.g., surgery of blue nevus of left forearm).

SNOMED Structure

☐　　　　There is some overlap of the axes so it is possible to form two different versions for the same concept (e.g., blue nevus nevus colored blue).

☐　　　　There are few rules for how to combine axesterms so it is possible to form valid nonsense terms (e.g., nevus of left esophagus).

☐　　　　Who owns SNOMED?

SNOMED Axes

- D Diseases
- C Drugs
- F Function
- L Living Organisms
- X Manufacturers
- G Modifiers

- M Morphology
- J Occupations
- A Physical Agents
- P Procedures
- S Social Context
- T -- Topography

Postcoordination

- D5-46210 Acute Appendicitis

- G-A231 Acute, D5-46100 Appendicitis NOS

- G-A231 Acute, M-40000 Inflammation, T-59200 Appendix

- M-41000 Acute Inflammation, T-59200 Appendix

- T-59200 Appendix, M-41000 Acute Inflammation

History

☐	For everyone who wants to own a medical vocabulary, there is a set of terms which are likely to overlap but be inconsistent with every other vocabulary.

☐	Read, CPT, COSTART, ChemAbstracts,

☐	In theory they are all describing agreed upon concepts. A single standard vocabulary would improve the automated flow of medical information.

Ideal Vocabulary

?

Ideal Vocabulary

☐ Boundary

☐ Organization

☐ Completeness

☐ Absence of ambiguity

☐ Growth

☐ Aging

Ideal Vocabulary Definitions

☐ String a unique sequence of characters. The same set of characters may represent different concepts.

☐ Lexical variants synonyms with minor differences. Word order, capitalization, and punctuation are usually included. Suffixes (plural) and prefixes may be included.

☐ One man's lexical variant is another's synonym.

Ideal Vocabulary Definitions

❑	Related terms distinct terms whose concepts overlap in some way. The most used relations are broader and narrower (e.g., Neurologic Disease includes but is broader than Epilepsy.)

❑	One man' s related term is another's synonym.

Ideal Vocabulary Definitions

☐ Controlled versus free text

   Freedom of expression

   Automatic indexing accuracy

☐ Atomic versus enumerated (Pre vs Post)

   Handle the unexpected

   Predict what to expect

☐ Definitions

   Free text versus semantic

Unified Medical Language

☐	The Unified Medical Language System (UMLS) started as an NLM collaborative program with 7 centers around the country.

☐	Proceeded in 3 3 year phases.

Explore ideas (1986)

One winner selected and developed (1988)

Usage Testing (1991)

UMLS Structure

□ Three components

Metathesaurus (META)

Semantic Network

Information Sources Map (ISM)

□ Recently dropped

Specialist Lexicon

□ Recently added

META Structure

☐	NOT a controlled vocabulary.

☐	Database of information about other controlled vocabularies.

☐	Contains sufficient info to recreate most of the component vocabularies. (Who owns META? Who owns the components?)

☐	Basic unit is the concept. A concept is linked to multiple strings from multiple vocabularies.

META - Structure

☐   Each concept-string pair is either a preferred term, synonym, or lexical variant.

☐   The same string may be linked to multiple concepts but a term, synonym, or lexical variant will only link to one concept each.

☐   Other links exist based solely on the existence of those links in a source vocabulary.