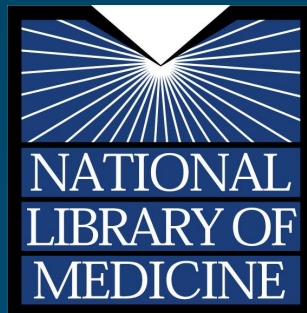


San Francisco
September 10, 2004



Large Formal Ontologies for Biomedicine

Panel

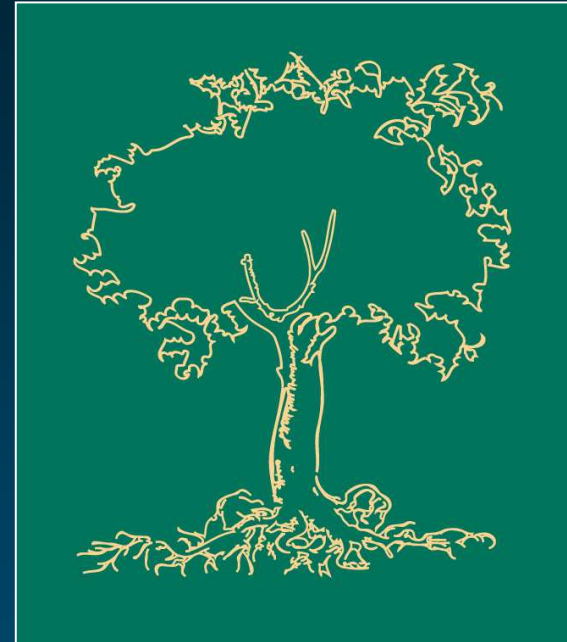


Olivier Bodenreider

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

What does UMLS stand for?

- ◆ 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.»

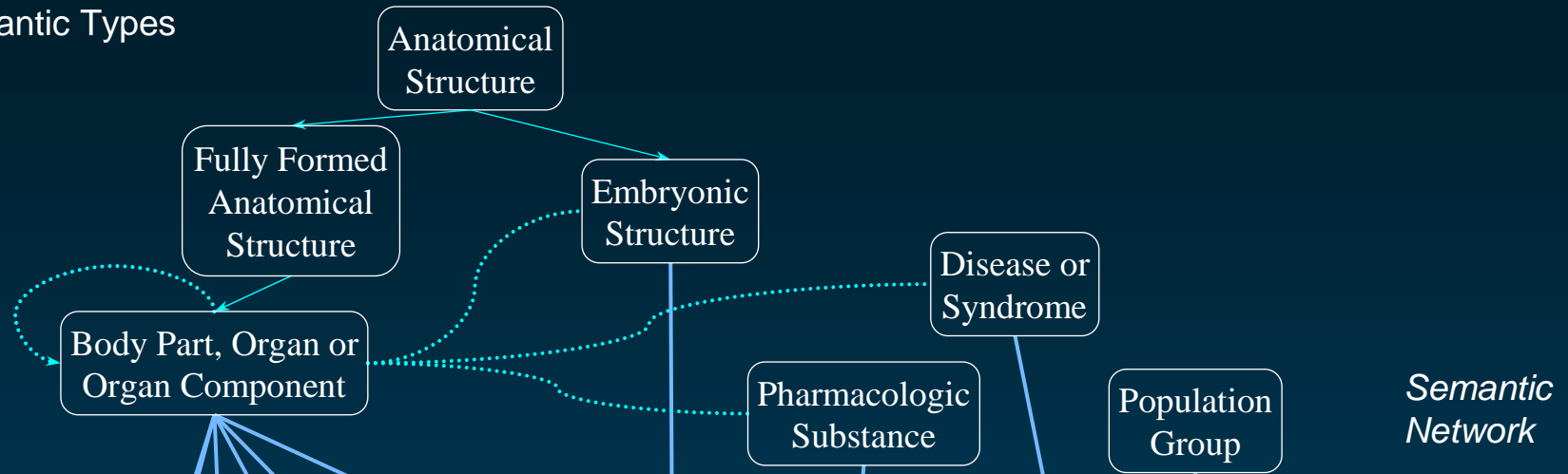


UMLS 3 components

- ◆ Metathesaurus
 - Concepts
 - Inter-concept relationships
- ◆ Semantic Network
 - Semantic types
 - Semantic network relationships
- ◆ Lexical resources
 - SPECIALIST Lexicon
 - Lexical tools



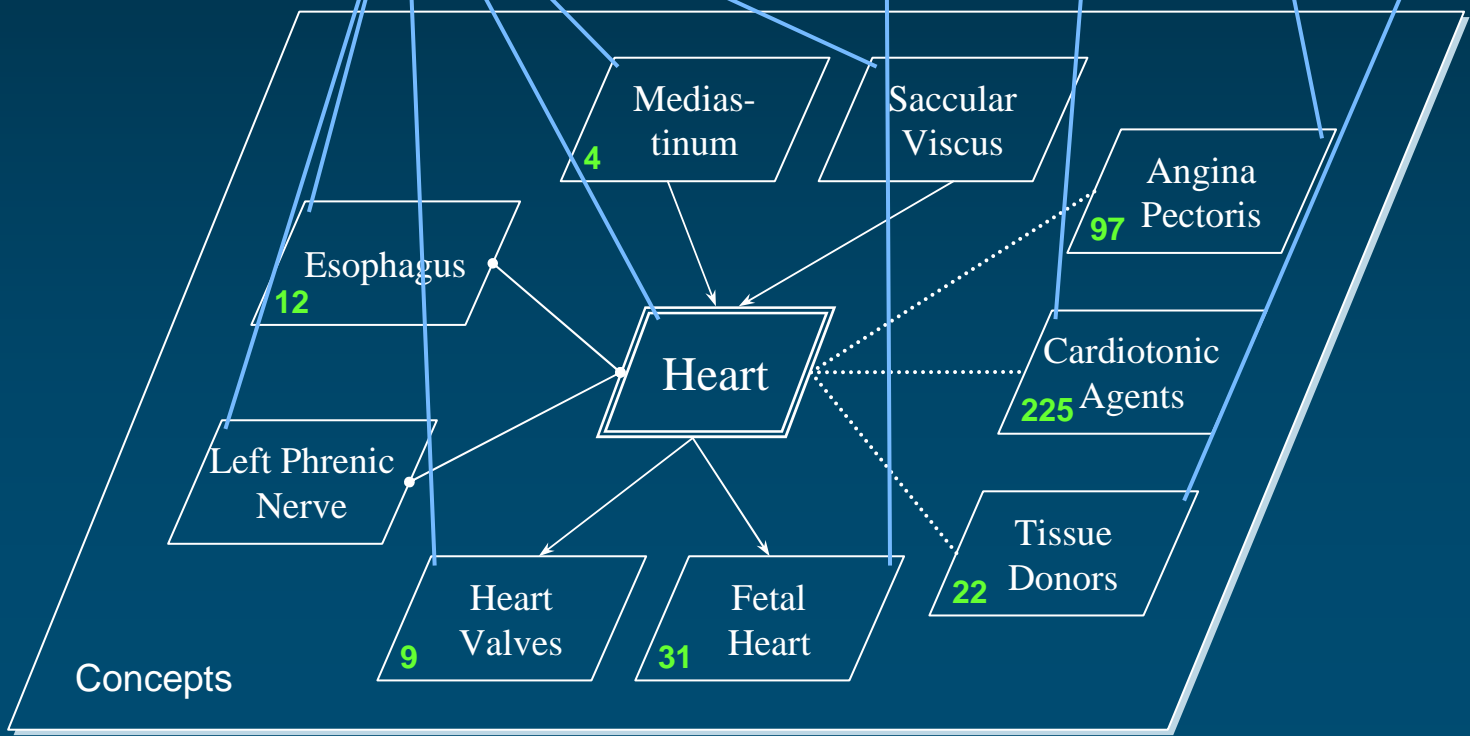
Semantic Types



Semantic Network

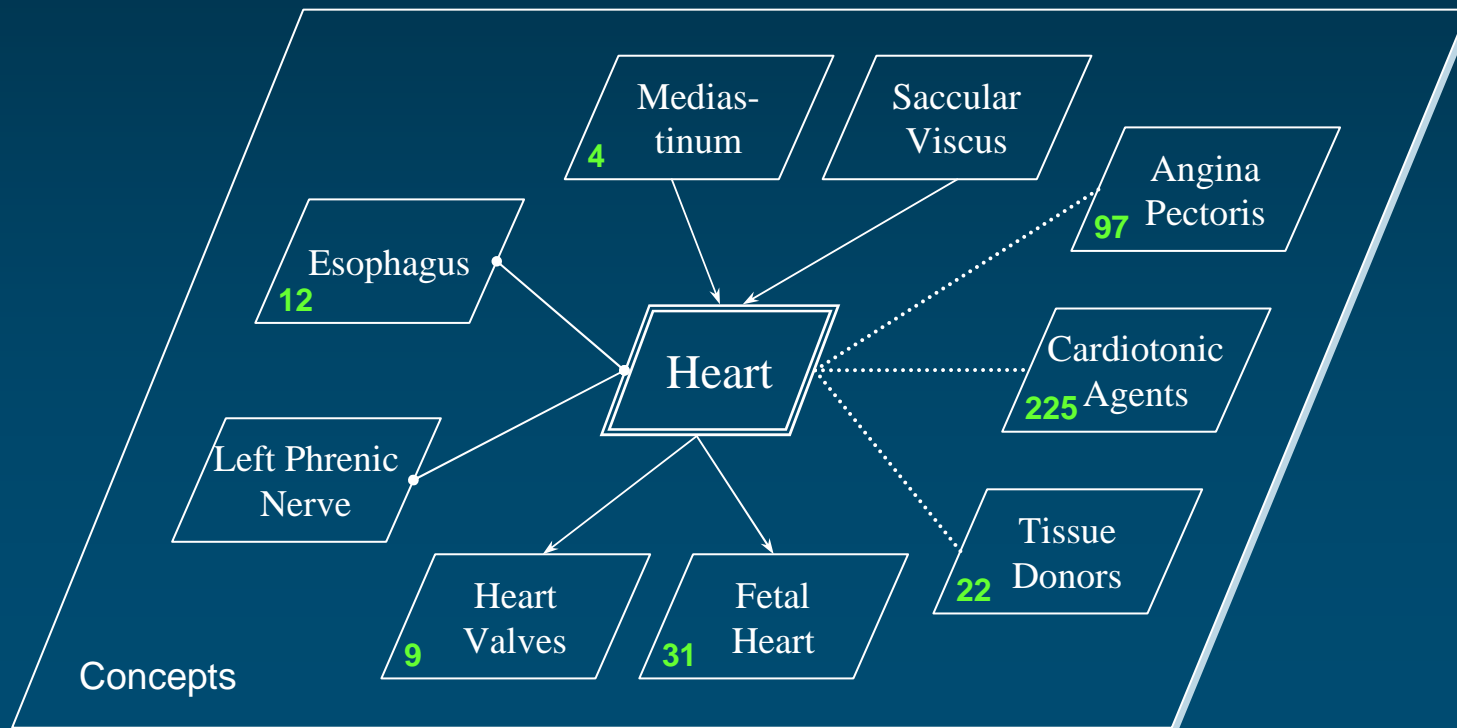


Metathesaurus



Concepts

UMLS Metathesaurus



Metathesaurus

Metathesaurus Large

(2004AB)

- ◆ 134 source vocabularies
 - 73 families of vocabularies
- ◆ 1M concepts
 - 4.3M distinct concept names
- ◆ 15M relations

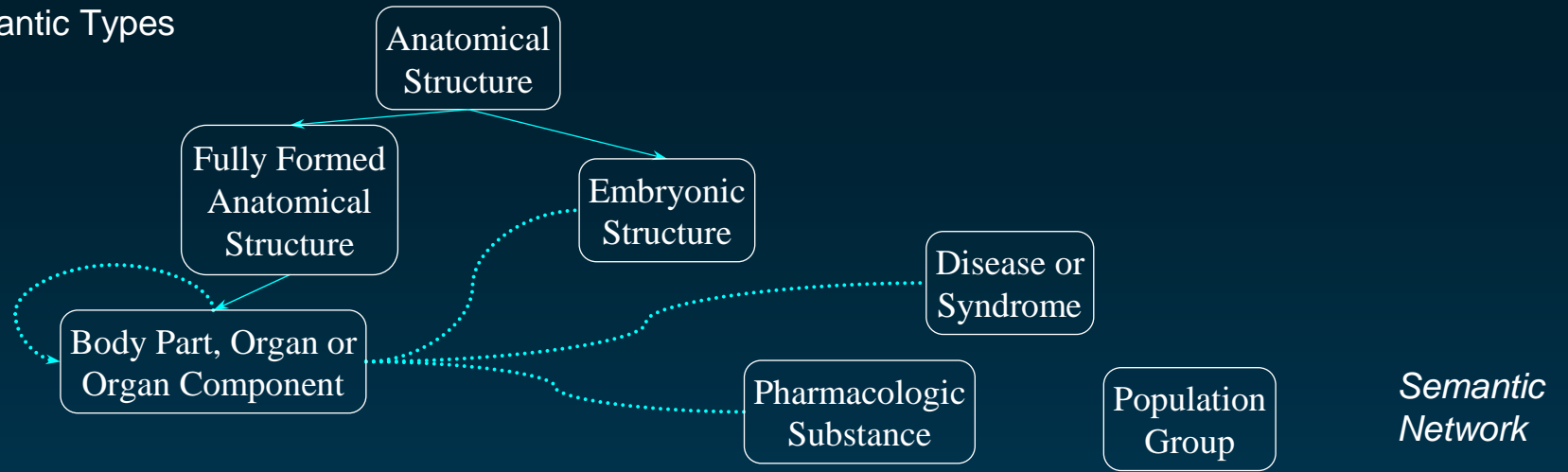


Metathesaurus Essentially informal

- ◆ By design
 - Terminology integration project
- ◆ By mandate
 - Vehicle for what is asserted in the source vocabularies
- ◆ Because of its purpose
 - Influenced by information retrieval
- ◆ No formalism
- ◆ Manual curation, but insufficient guidelines



Semantic Types



UMLS Semantic Network

Semantic Network More formal

- ◆ Manually created
 - Regardless of the position of the corresponding terms in existing vocabularies
- ◆ No particular formalism
 - Triples $\langle ST_1, rel, ST_2 \rangle$
- ◆ Quantification: some value from
 - $\langle Drug, treats, Disease\ or\ Syndrome \rangle$
 - some *Drug* *treats* some *Disease or Syndrome*



Semantic Network But small

- ◆ 135 semantic types
- ◆ 54 relationships
- ◆ 558 relations



Can the Metathesaurus be more formal?

- ◆ Probably not given its design
- ◆ Trade-offs
 - Ingredient/drug synonymy
 - Clinical synonymy/linguistic synonymy
- ◆ But some recent changes in its representation may help
 - Source transparency: relations are no longer recorded at the concept level, but at the atom level



Couldn't the Metathesaurus use another formalism (e.g., OWL)?

- ◆ Simply changing the formalism would not help
- ◆ What is needed are more precise, better defined relationships (isa vs. parent/child)
 - A trivial change of formalism would (wrongly) assume isa for each parent/child relationship
- ◆ Transforming portions of the UMLS in DL is resource intensive
- ◆ The OWLization of other terminology systems has not produced convincing results



Towards more consistency

- ◆ Many studies have investigated consistency issues in the Metathesaurus
- ◆ Some have developed algorithms, e.g.,
 - Identify and remove circular hierarchical relations
 - Use Semantic Network relations to assess the validity of Metathesaurus relations
 - ...

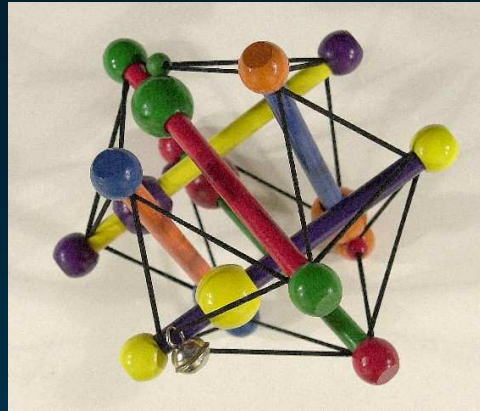


Conclusions

◆ Metathesaurus

- Large and informal
- But could be used as the basis for deriving an ontology of biomedicine
 - Manual curation
 - Lexical methods for acquiring ontological relations from terminology
 - Would still require an upper-level ontology
 - And a formalism

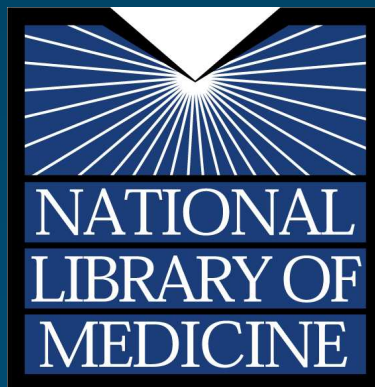




Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov



Olivier Bodenreider

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