A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases

Document Navigation: Ontologies or Knowledge Organisation Systems

Simon Jupp - NETTAB 2007

BioHealth Informatics Group

University of Manchester, UK

Introduction

Bioinformatics relies heavily on web for IR.

Ontologies are being developed as background knowledge to drive the Semantic Web.

Message: Formal ontologies are not the only knowledge artefact needed, artefacts with weaker semantics have their role and are the best solution in some circumstances

COHSE

Navigation via Hypertext is a mainstay of WWW

Problem: Links are typically embedded to Web pages; hard-coding, format restrictions, ownership, legacy resources, maintenance, Unary targets etc.

Which model is best suited for Navigation?

Strict semantics like OWL or something with weaker semantics like SKOS.

Is ontological formality a help or a hindrances?

SeaLife use case - study of disease

NeLI: National electronic Library of Infection portal. Range of users but few links.

Given a document about Tuberculosis, where would users want to navigate to next?

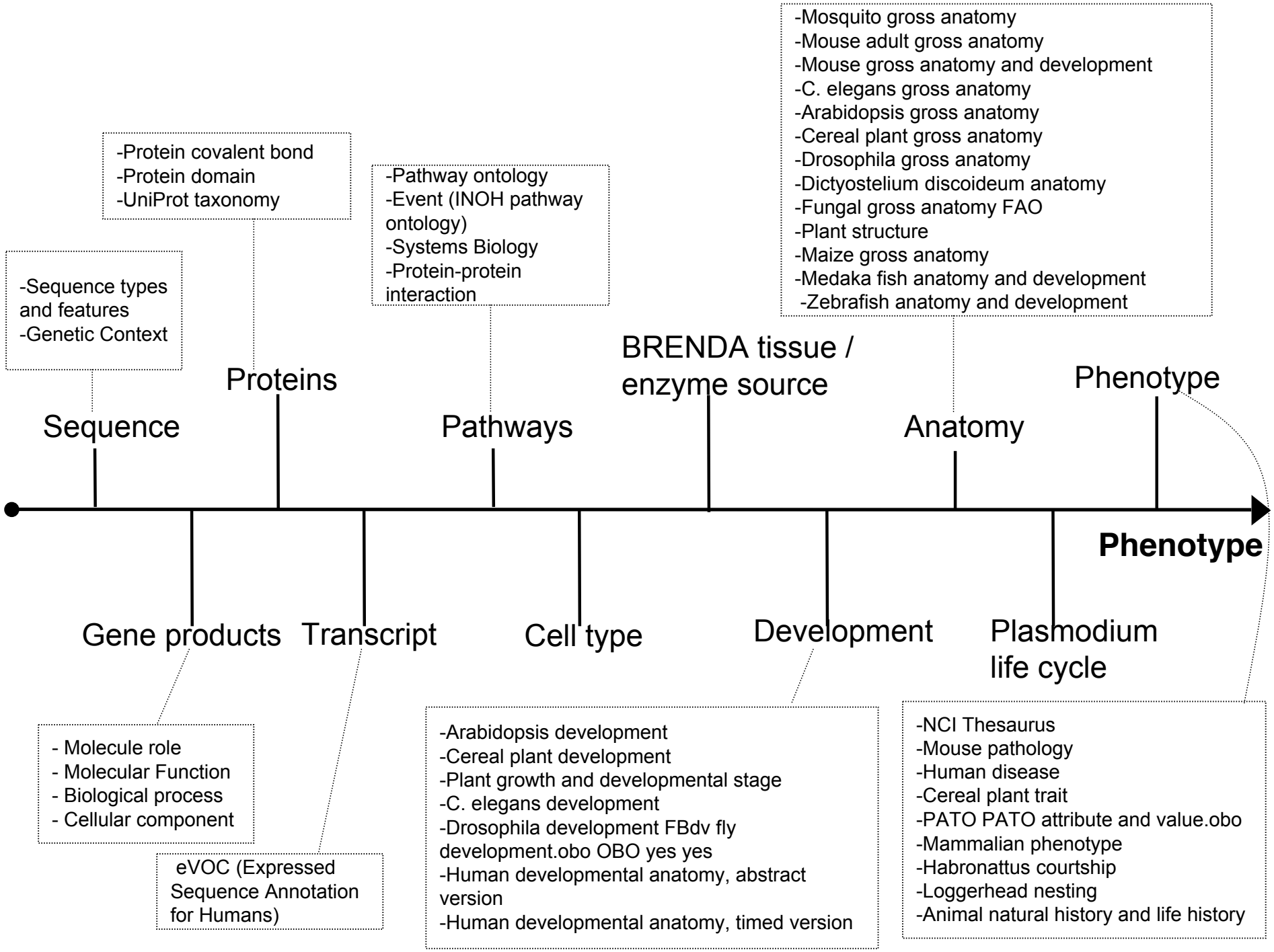
User Group	Question	Targets
Family Doctor (GP)	Tuberculosis drugs and side effects?	British National Formulary (BNF)
Clinicians	Tuberculosis treatments guidelines?	Public Health Observatories (PHO)
Molecular Biologists	Drug resistant tuberculosis species?	PubMed
General Public	What is tuberculosis?	Health Protection Agency (HPA) or the NHS direct online website.

http://www.neli.org.uk

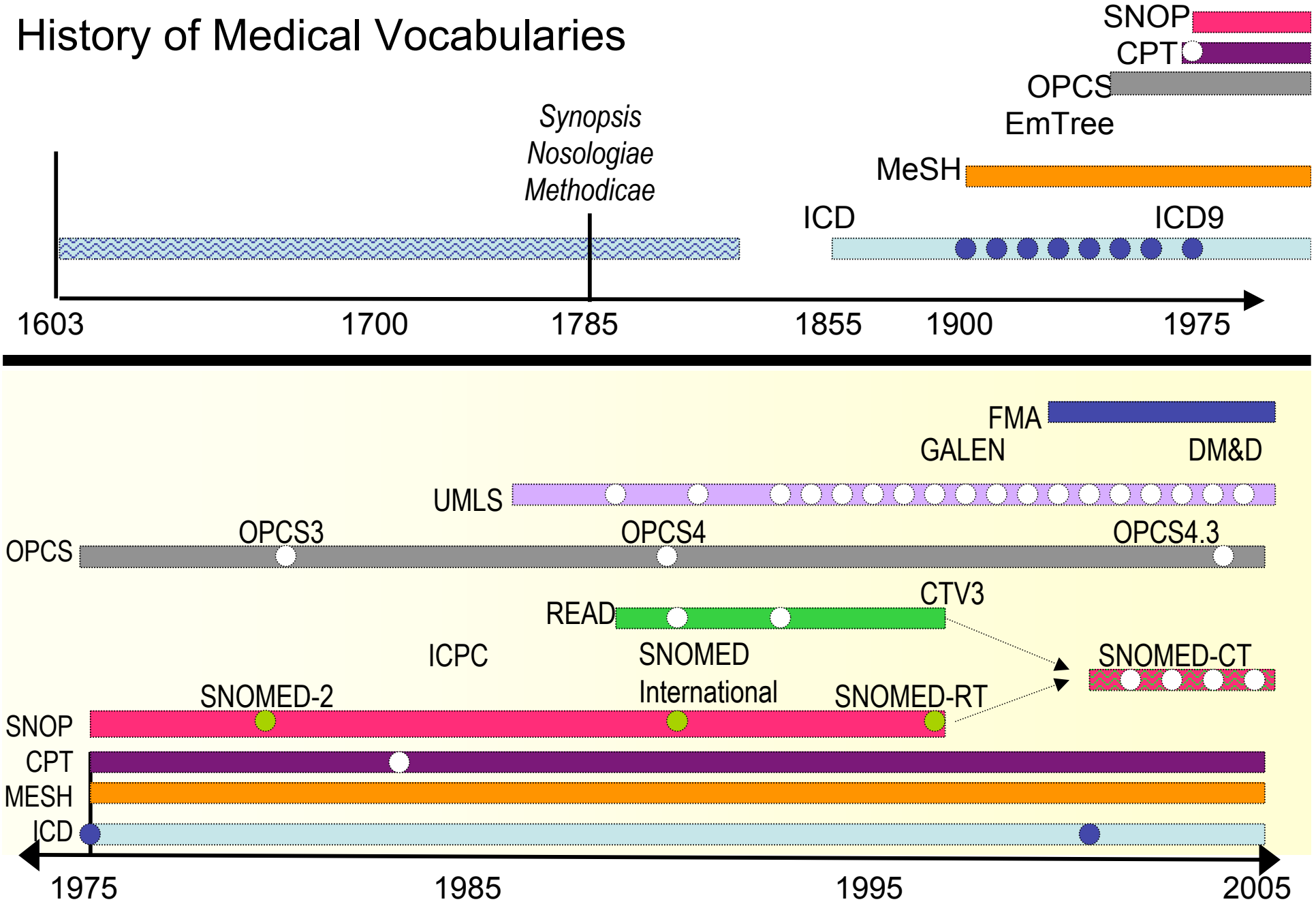
SeaLife background knowledge

To cover molecular biology through to medicine we need a large knowledge artefact to serve as background knowledge for SeaLife. This artefact must support sensible navigation between documents on the web.

Luckily…


-Mosquito gross anatomy
-Mouse adult gross anatomy
-Mouse gross anatomy and development
-C. elegans gross anatomy
-Arabidopsis gross anatomy
-Cereal plant gross anatomy
-Drosophila gross anatomy
-Dictyostelium discoideum anatomy
-Fungal gross anatomy FAO
-Plant structure
-Maize gross anatomy
-Medaka fish anatomy and development
-Zebrafish anatomy and development
-Protein covalent bond
-Protein domain
-UniProt taxonomy
-Pathway ontology
-Event (INOH pathway ontology)
-Systems Biology
-Protein-protein interaction
-Sequence types and features
-Genetic Context
Sequence
Proteins
Pathways
BRENDA tissue / enzyme source
Anatomy
Phenotype
Phenotype
Gene products
Transcript
Cell type
Development
Plasmodium life cycle
- Molecule role
- Molecular Function
- Biological process
- Cellular component
eVOC (Expressed Sequence Annotation for Humans)
-Arabidopsis development
-Cereal plant development
-Plant growth and developmental stage
-C. elegans development
-Drosophila development FBdv fly development.obo OBO yes yes
-Human developmental anatomy, abstract version
-Human developmental anatomy, timed version
-NCI Thesaurus
-Mouse pathology
-Human disease
-Cereal plant trait
-PATO PATO attribute and value.obo
-Mammalian phenotype
-Habronattus courtship
-Loggerhead nesting
-Animal natural history and life history

History of Medical Vocabularies



What do we need for navigation?

The bio-medical domain is rich in vocabularies and ontologies.

Large lexical resource including textual definitions and synonyms

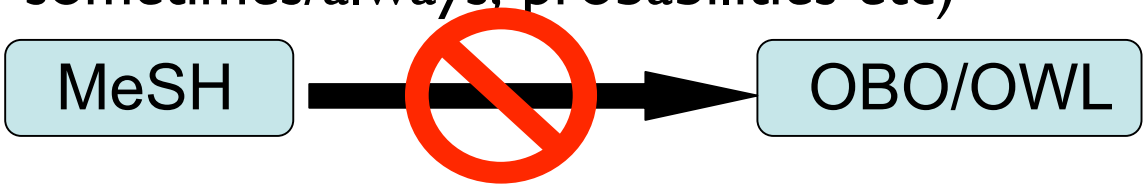
There is a varying degree of semantics, expressivity and formality in these vocabularies (e.g. MeSH) and ontologies (e.g FMA). Most include some form of hierarchy. Hierarchies are well suited for driving navigation.

Question: Do we want strict sub/super class relationships? Or, do we want looser notations such as broader/narrower?

Ontology or Vocabulary?

Initial approach to COHSE and SeaLife was to represent everything in OWL

The strict semantics of OWL do not always lend themselves to sensible navigation, conversion from vocabularies to OWL are difficult. It's hard to model some things in OWL (e.g. sometimes/always, probabilities etc)



Head <-- Ear

<-- Nose

Accident <-- Traffic Accident

<-- Accident Prevention

Nucleus *part_of* **Cell**

Cell *has_part* **Nucleus - Not always True**

PolioVirus *causes some* **PolioDisease**

SKOS (Simple Knowledge Organisation System)

Purpose: Subject Metadata and information retrieval

e.g. This document is about tuberculosis

Model for representing concept schemes, thesauri, classification system, taxonomies etc…

Importantly for us, the semantics are more suitable for document navigation

e.g. broader, narrower, related.

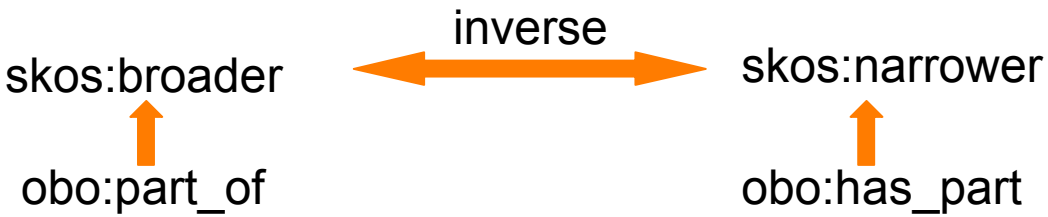
RDF/XML representation - Semantic Web friendly.

http://www.w3.org/2004/02/skos/

Conversion to SKOS

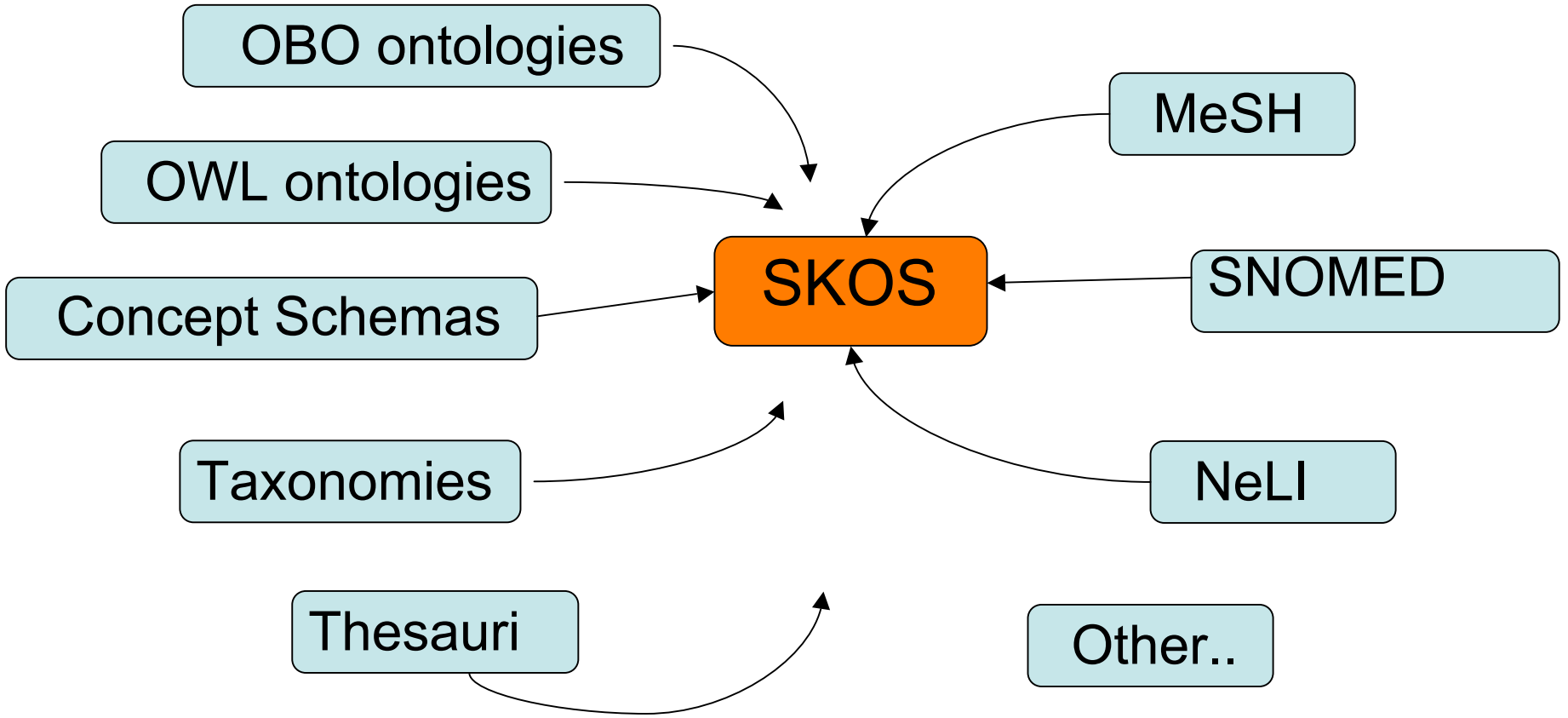
- relationship:part_of --> skos:broader (e.g. finger part_of hand)
- relationship:contains --> skos:narrower (e.g. skull contains brain)
- relationship:causes --> skos:related (e.g. PolioVirus causes PolioDisease)

Sub properties:



Leaves us open to migration back to OWL

Conversion to SKOS



Advantage of this approach

For a given concept e.g. "Polio Virus", we can query multiple resources and bring related concepts together.

Source	Terms found	SKOS relation to "Poliovirus"
MeSH	Brunhilde Virus	skos:altTerm
Disease Ontology	Spinal cord disease	skos:broaderThan
	Postpoliomyelitis Syndrome	skos:narrowerThan
SNOMED	Microorganism	skos:broaderThan
	Enterovirus	skos:broaderThan

- Rapid (and cheap!) generation of knowledge artefact
- Take advantage efforts in multiple biomedical communities
- We don't have to make any strong ontological distinctions

Disadvantage of this approach

Trade off: Lose the inferential power when querying a knowledge resource

Unwanted concepts & relationship - especially from OWL conversion e.g. 'Physical Entity', 'Continuant' etc….

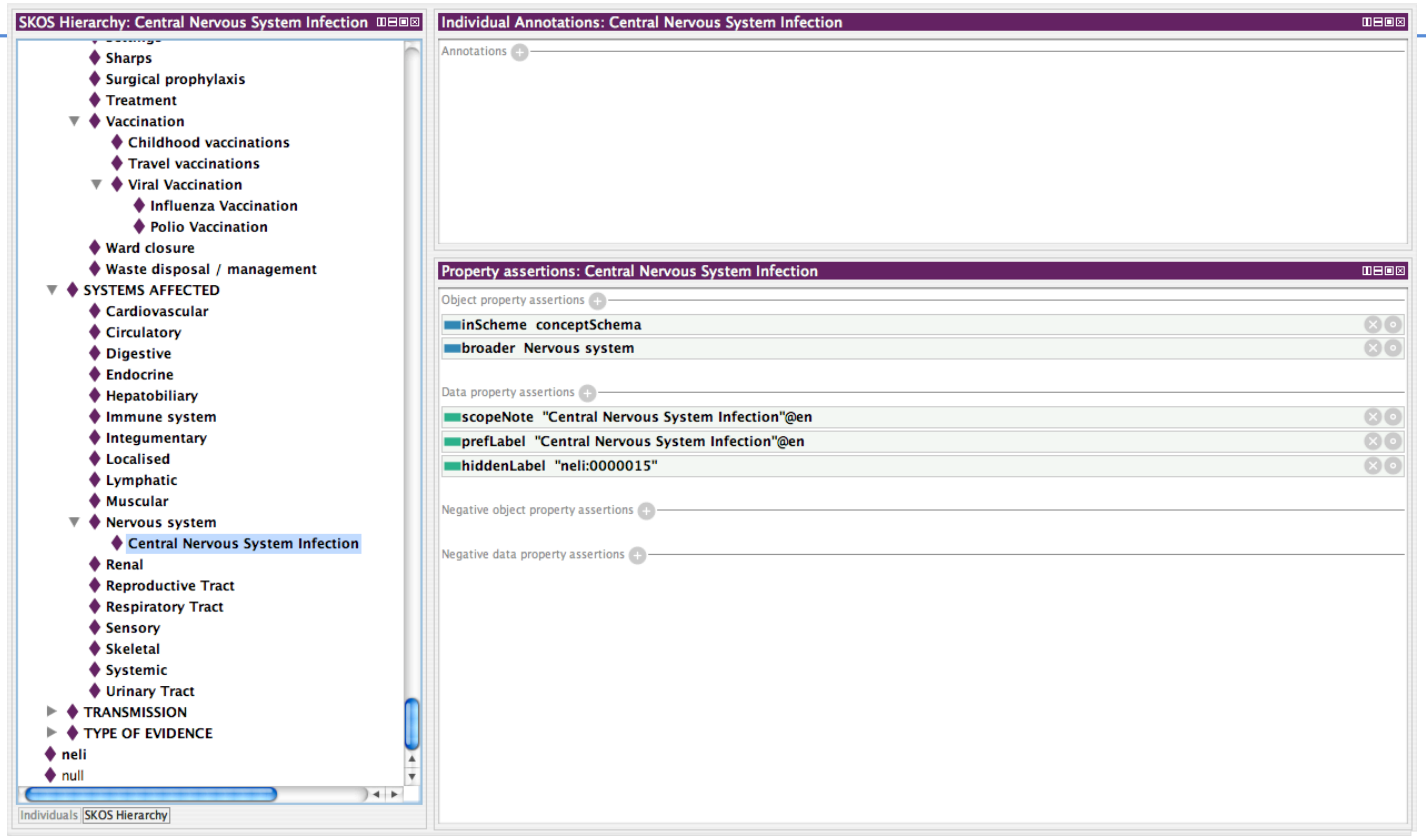
Linking overload!

Inability to do inconsistency checking

Potentially large redundancy in our knowledge base

Maintenance and scalability (>1000000 concepts) - especially for dynamic hyper-linking.

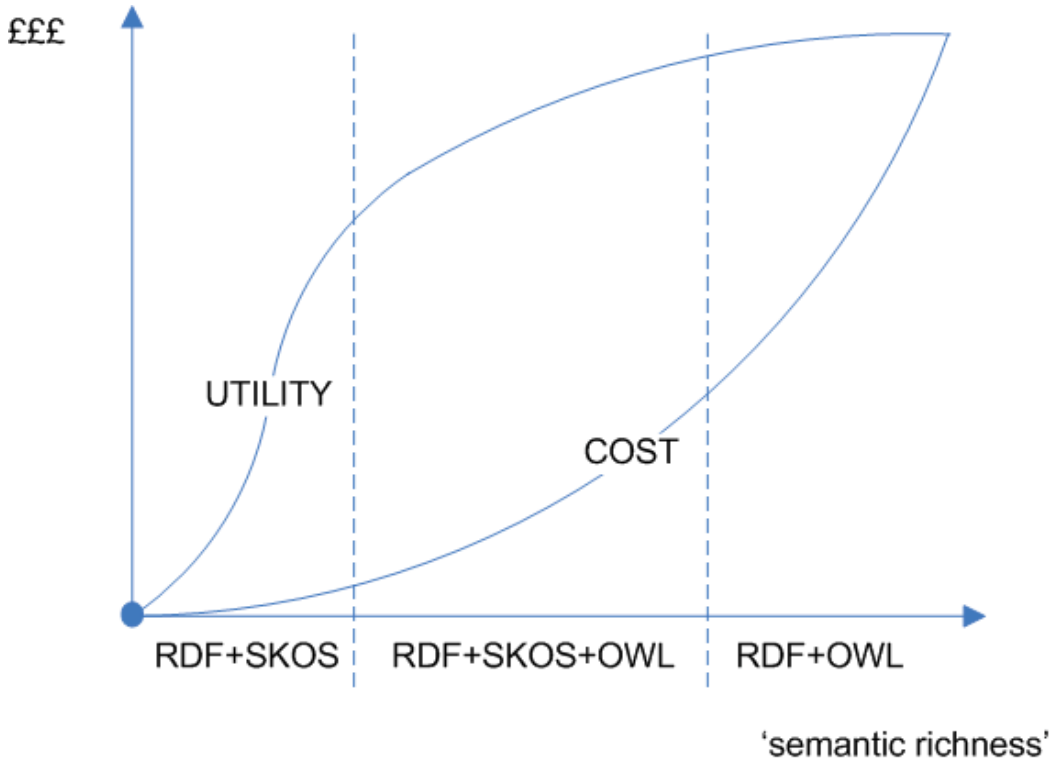
Plug for Manchester's SKOS plug-in - Protégé 4



- Instance hierarchy viewer
- OBO or OWL --> SKOS wizards
- Various rendering options

Conclusion

What does your semantic web application need?



Taken from Alistair Miles, XMLUK: "Ontologies and XML" 2005, slide

Acknowledgments

Manchester

Robert Stevens

Sean Bechhofer

Yeliz Yesilada

NeLI

Patty Kostkova

Other

COHSE developers

SeaLife project

Thank you.