

Re-representing Biomedical Ontologies using the Web Ontology Language -A Checklist

Chintan O. Patel, MS

PhD Student

Yves A. Lussier, MD

- Director, Northeastern Biodefense Center

- Assistant Professor of Biomedical Informatics

and of Medicine, Columbia College of Surgeons and Physicians

- Clinical Director, Center for Advanced Technology



Department of Biomedical Informatics

Columbia University



Goals

- Review the issues and challenges in RE-representation of Biomedical ontologies using the Web Ontology Language
- Representing and interoperating non-molecular phenotypes across communities and different scales of Biology



Plan

- Web Ontology Language
 - Motivation
- Review on Re-representation issues with OWL
 - Default, Context, Multiple Interpretations,...
- Conclusion



OWL: Web Ontology Language

- W3C Recommendation Standard - Feb 2004
- Based on XML, RDF
- Strong industrial and academic backing
 - IBM, HP, Cerner, Nokia, Adobe,...
- Wide tool support
 - HTML, XML...



OWL: The Purpose

- A standard for representing data over Web to allow effective sharing of data across humans and computers
- A sound and decidable framework for reasoning
- Achieve a Web of meaningful information



Biomedical Ontologies

- Classify/organize Biomedical information
- Standardized terms and coding for a given domain
- Perform simple subsumption reasoning
- Efficient information retrieval



OWL fits Biomedical Ontologies

- Logic based formalism enforces consistency
 - Very important in case of large ontologies (~thousands of concepts)
- Collaborative representations across Web
 - Ontology linking (URIs)
- Efficient large scale reasoning implementations
 - FaCT, Racer



“socio-economic-political” benefits of OWL

- “Global” standard for Knowledge Representation
- Extensive tool support



Efforts to RE-represent Biomedical Ontologies in OWL

- **Gene Ontology** Next Generation (GONG)
- open**GALEN** – University of Manchester
- **UMLS** Semantic Network in OWL – NLM
- **Cancer Thesaurus** – National Cancer Institute
- **MeSH** - Catalogue and Index of French-speaking Medical Sites



The RE-Representation Checklist

- Defaults/Exceptions
- Context
- ISA/Part of/Instance of
- Multiple classification
- Multiple interpretations
- Inheritance blocking
- Polymorphic relationships
- Instance reasoning



Defaults/Exceptions

[openGALEN experience]

- By **default** the value of X is 1, but sometimes its -1 (singular exceptions)
- All arteries carry oxygenated blood **except** the pulmonary artery !

:Artery

```
rdfs:subclassof [owl:Restriction
                 owl:onProperty carries
                 owl:someValuesFrom :OxygenatedBlood]
```



...Defaults/Exceptions

- can also be said as... Aorta and its branches carry oxygenated blood.

:SystemicArtery

```
owl:unionOf (:Aorta [ a owl:Restriction;  
                    owl:onProperty isBranchOf  
                    owl:someValuesFrom :Aorta])  
rdfs:subclassof [owl:Restriction  
                owl:onProperty carries  
                owl:someValuesFrom :OxygenatedBlood]
```



Context

[openGALEN experience]

- Prostate
 - **Human**, 1 in number with 3 lobes
 - **Mice**, 5 in number without lobes

```
[owl:Restriction
 owl:onProperty isOfSpecies
 owl:someValuesFrom :Human]
[owl:Restriction
 owl:onProperty hasPart
 owl:cardinality 1
 owl:someValuesFrom :Prostrate]
```

```
[owl:Restriction
 owl:onProperty isOfSpecies
 owl:someValuesFrom :Mouse]
[owl:Restriction
 owl:onProperty hasPart
 owl:cardinality 5
 owl:someValuesFrom :Prostrate]
```



ISA/Part-Of/Instance-Of

- OBO Mouse anatomy use only *part-of*
- OBO Drosophila anatomy uses *isa*, *part-of* and *lineage*.
- *V Type ATPase Complex*
 - ① *ATPase V0 domain* (GO:0046610)
 - should be a Part-Of relationship
- OK for human usage but terrible for automated reasoning.



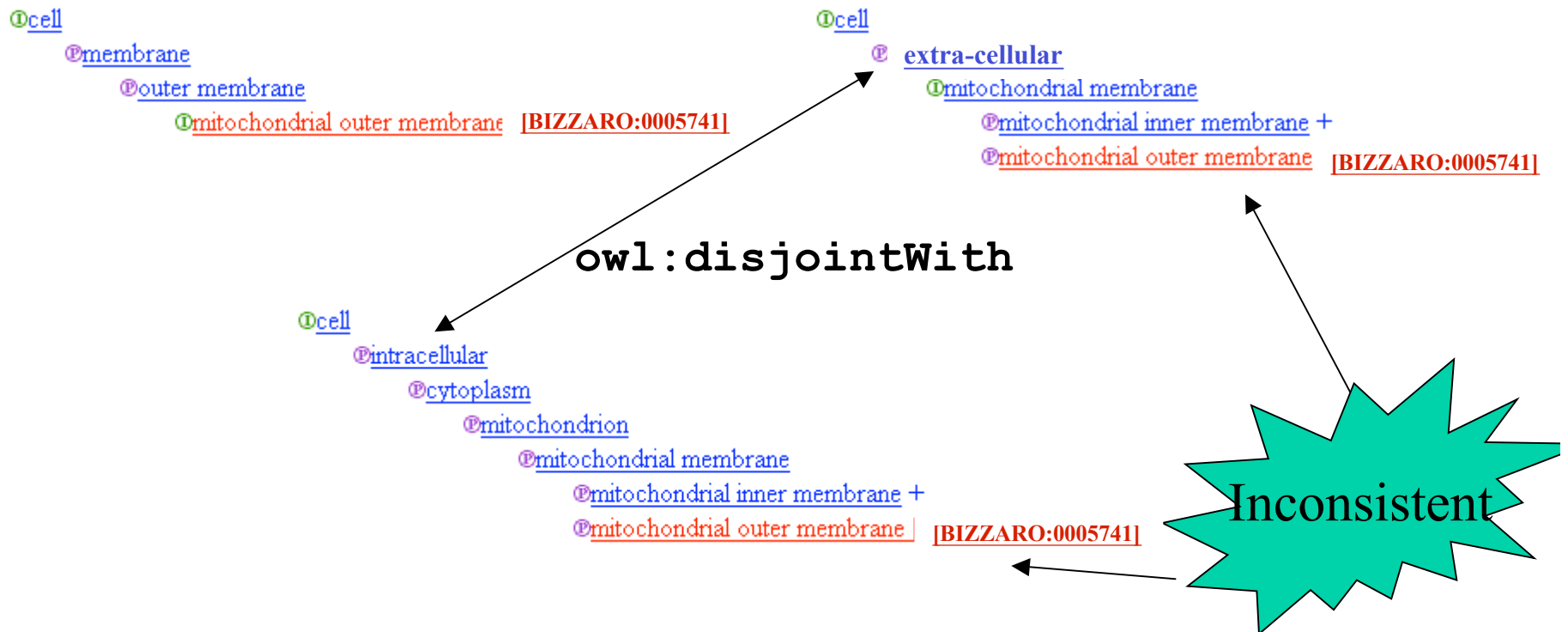
...ISA/Part-Of/Instance-Of

- Complete Partonomic reasoning is out of OWL-DL
- An OWL-Full model for part-of anatomical representation available [Aitken et al]



Multiple Classification

Mitochondrial outer membrane [BIZZARO:0005741]

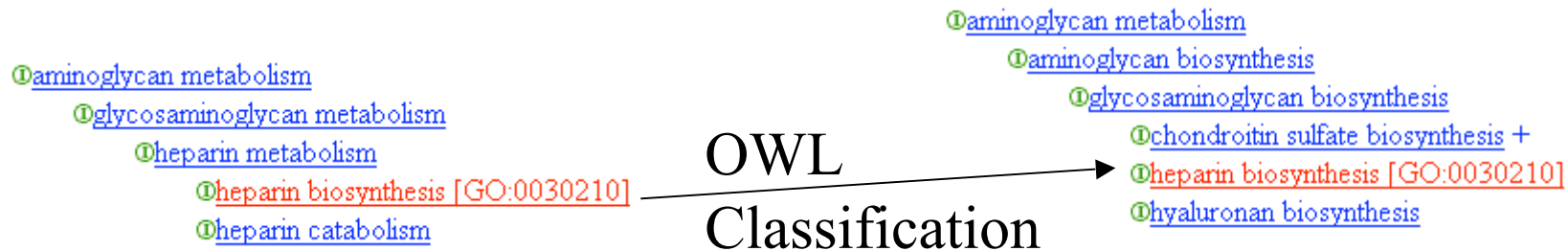




Classification

[GONG Experience]

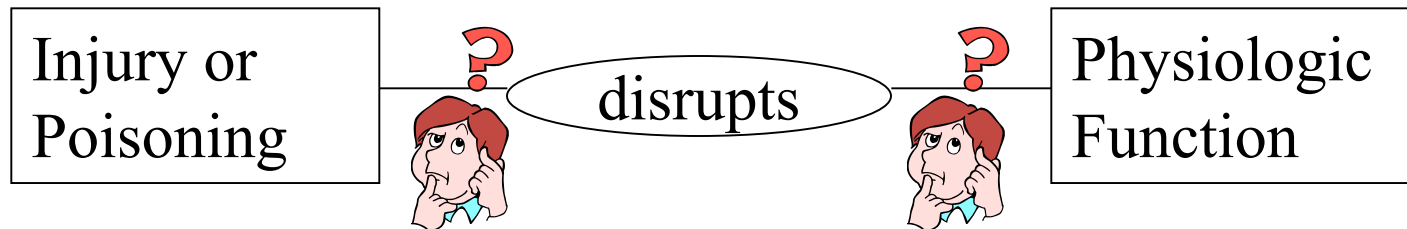
- Discover the right place in the subsumption hierarchy



heparin is a kind of glycosaminoglycan !



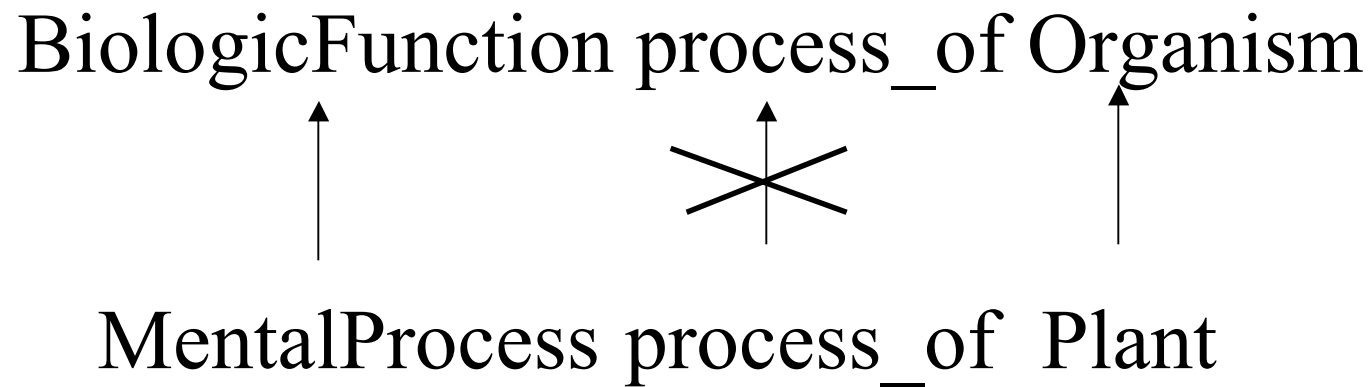
Multiple Interpretations [UMLS-SN OWL, Kashyap et al]



- **All** Injury disrupts {**each/only/some**}
Physiologic Function
- **Some** Injury distrupts {**all/some**}
Physiologic Function
-



Inheritance blocking



$\text{BiologicFunction} \cup \neg\text{MentalProcess} \subset \text{domain}(\text{process_of})$
 $\text{Organism} \cup \neg\text{Plant} \subset \text{range}(\text{process_of})$



Polymorphic relationships

BodyLocationorRegion **location_of** BiologicFunction

BodySpaceorJunction **location_of** BodySubstance

$\text{domain}(\text{location_of}) \subset \text{BodyLocationorRegion} \cup \text{BodySpaceorJunction}$

$\text{range}(\text{location_of}) \subset \text{BiologicFunction} \cup \text{BodySubstance}$



Instance Reasoning

- Reasoning over large number of instances
- InstanceStore [Horrocks et al]
 - Database support for reasoning
 - Tested with 653,762 individuals (SWISS-PROT) and 13,929 Gene Ontology concepts.



Summary

	openGALEN	GONG	UMLS Semantic Network-OWL	NCI Cancer Thesaurus- OWL
Default/Context	✓			
Multiple Classification	✓	✓	✓	✓
Inheritance Blocking			✓	
Polymorphic Relationship			✓	
Instance Reasoning	✓	✓		



...to-do list

- Reasoning with Temporal Representations in OWL
- Large scale T-Box reasoning
 - Memory and architectural limitations
- More on OWL...
 - W3C Semantic Web Best Practices and Deployment working group
<http://www.w3.org/2001/sw/BestPractices/>



Conclusion

- A concise checklist of issues related to RE-representation of Biomedical ontologies based on existing projects
- Some fundamental questions
 - Large number of primitive classes
 - Basic differences in modeling style (Frame-based)
- Need for tools that would facilitate the conversion process (OBO2OWL ?)



References

- [openGALEN] <http://www.opengalen.org/>
- [Gene Ontology]
 - C.J. Wroe, R. Stevens, C.A. Goble, M. Ashburner, *A Methodology to Migrate the Gene Ontology to a Description Logic Environment Using DAML+OIL* Pacific Symposium on Biocomputing 8:624-635(2003).
 - Smith B, Williams J, Schulze-Kremer S. , *The ontology of the gene ontology*, AMIA Annu Symp Proc. 2003:609-13.
- Vipul Kashyap, Alexander Borgida: *Representing the UMLS Semantic Network Using OWL: (Or "What's in a Semantic Web Link?")*. International Semantic Web Conference 2003: 1-16
- Aitken JS, Webber BL, Bard JB, *Part-of relations in anatomy ontologies: a proposal for RDFS and OWL formalizations*, Pac Symp Biocomput. 2004:166-77.
- Ian Horrocks, Lei Li, Daniele Turi, Sean Bechhofer: *The Instance Store: DL Reasoning with Large Numbers of Individuals*. Description Logics 2004



Acknowledgement

This work was partially funded by

- NLM 1K22 LM008308-01 – Lussier YA (PI)
Semantic Approaches to Phenotypic Database Analysis
- NIAID 5U54 AI057158-02 Lipkin WI (PI)
RCE for Biodefense and Emerging Infections Diseases Research