

# Ontology-based data access

Robert Hoehndorf



King Abdullah University of Science and Technology

# Ontology-based data access

What do we gain when using ontologies to access our data?

# Ontology-based data access

Identify inconsistencies and incoherent descriptions of data

- 'prokaryotic cell' DisjointFrom: 'eukaryotic cell'
- 'fungal cell' SubclassOf: 'eukaryotic cell'
- *spore<sub>1</sub>* instanceOf: 'fungal cell'
- *spore<sub>1</sub>* instanceOf: 'prokaryotic cell'
  - knowledgebase becomes inconsistent

# Ontology-based data access

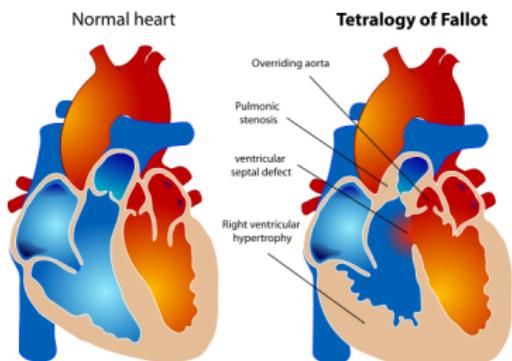
Identify inconsistencies and incoherent descriptions of data

- 'prokaryotic cell' DisjointFrom: 'eukaryotic cell'
- 'fungal cell' SubclassOf: 'eukaryotic cell'
- *spore<sub>1</sub>* instanceOf: 'fungal cell'
- *spore<sub>1</sub>* instanceOf: 'prokaryotic cell'
  - knowledgebase becomes inconsistent
- *spore* SubclassOf: 'fungal cell'
- *spore* SubclassOf: 'prokaryotic cell'
  - *spore* becomes unsatisfiable

# Ontology-based data access

Enrich possibly incomplete data with background knowledge

Retrieve all patients with *Tetralogy of Fallot*:

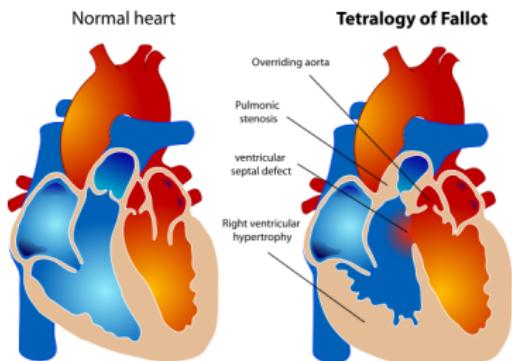


- P instanceOf has-symptom some 'overriding aorta'
- P instanceOf has-symptom some 'ventricular septal defect'
- P instanceOf has-symptom some 'pulmonic stenosis'
- P instanceOf has-symptom some 'right ventricular hypertrophy'

# Ontology-based data access

Enrich the data schema used to query data sources with additional information

Retrieve all patients with *ventricular septal defect*:



- P instanceOf has-symptom some 'ICD9:745.2 (tetralogy of fallot)'

## Ontology-based data access

Provide a uniform view over multi-modal data sources

>gi|568815590|105318690-105804539 Homo sapiens chromosome 8, GRCh38 Primary Assembly

TCCCTCCTCCCCCTCCCTCACTGTACACTCTGTGCCCGCTCTCTCTCATTTGCTTGC  
ATCTCGGAACGTGAATCCGGGCTCC  
GGGGCGGAGCTGGCCAGCGGGCGGG  
GCGGGCACCGGGAGGGCCAGGGCGC  
GGAGAGATGAAAGCTAATGTTTGCTGCTC

**Die Fallot'sche Tetralogie** ist die häufigste Form angeborener Herzfehler, die mit einer Blaufärbung der Haut (Zyanose) einhergehen, und der erste Herzfehler, der bereits 1954 erfolgreich und zufriedenstellend operativ korrigiert werden konnte. Kinder, die heute mit einer Fallot'schen Tetralogie geboren werden, dürfen in nahezu allen Fällen erwarten, die operative Korrektur ihres Herzfehlers zu überleben und zumindest die ersten drei bis vier Jahrzehnte ihres Lebens ein weitgehend normales Leben als Jugendliche und Erwachsene führen zu können.

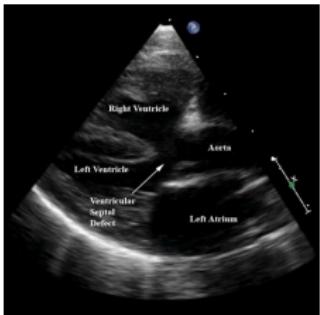
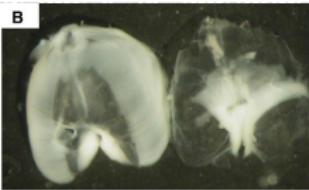
Tex Heart Inst J. 2014 Dec; 14(6):664-7. doi: 10.14503/THIJ-13-3674. eCollection 2014.

**Importance of absent ductus arteriosus in tetralogy of fallot with absent pulmonary valve syndrome.**

Qureshi MY, Burkhardt HM, Julianud P, Cetta F.

### **Abstract**

Tetralogy of Fallot without pulmonary valve syndrome is almost always associated with an absent ductus arteriosus. Patients with right aortic arch and retroesophageal left subclavian artery have a vascular ring if the left ductus arteriosus or its remnant and the Kommerell diverticulum are present. We report the cases of 2 infants in whom the role of an absent ductus arteriosus or its remnant is noteworthy. Both patients had a combination of tetralogy of Fallot with absent pulmonary valve syndrome and right aortic arch with retroesophageal left subclavian artery without a vascular ring. The absence of the ductus arteriosus has a role in the pathogenesis of tetralogy of Fallot with absent pulmonary valve syndrome. The absence of a ductus arteriosus in the right aortic arch with retroesophageal left subclavian artery precludes a vascular ring.



# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# Ontology-based data access

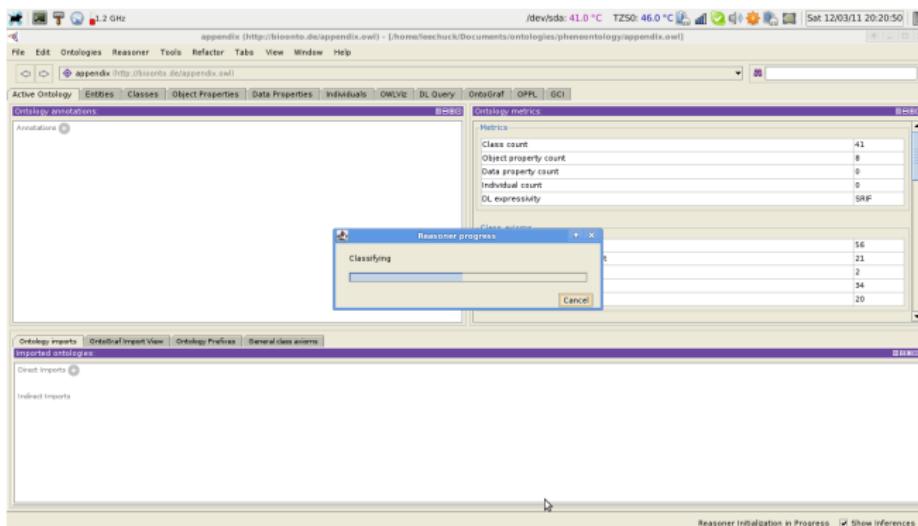
## Challenges for ontology-based access to biological data

- hundreds of ontologies
- millions of classes
- complex/expressive ontologies

Statistics	
Ontologies	371
Classes	5,872,930
Resources Indexed	39
Indexed Records	5,126,145
Direct Annotations	1,883,854,337
Direct Plus Expanded Annotations	24,828,631,205

# Web Ontology Language

- based on Description Logic *SROIQ*
- satisfiability is decidable  $\Rightarrow$  automated reasoning
- complexity for satisfiability in SROIQ is  
2-NEXPTIME-complete



# Modularization

- tractable subsets of OWL 2: EL, QL, RL
- problem: given an OWL ontology  $O$ , can we identify a maximal (EL, QL, RL)-module of  $O$ ?

# Modularization

- tractable subsets of OWL 2: EL, QL, RL
- problem: given an OWL ontology  $O$ , can we identify a maximal (EL, QL, RL)-module of  $O$ ?
- 'Abnormality of appendix' `EquivalentTo: not (has-part some (Appendix and Normal))` (
- 'Absent appendix' `EquivalentTo: not (has-part some Appendix)` (

# Modularization

- tractable subsets of OWL 2: EL, QL, RL
- problem: given an OWL ontology  $O$ , can we identify a maximal (EL, QL, RL)-module of  $O$ ?
- 'Abnormality of appendix' EquivalentTo: not (has-part some (Appendix and Normal)) (- 'Absent appendix' EquivalentTo: not (has-part some Appendix) (- Inference: 'Absent appendix' SubclassOf: 'Abnormality of appendix' (EL)

# Modularization

## Definition

Let  $T$  be a set of  $L_1$ -formulas over  $\Sigma$ , and  $L_2 \subseteq L_1$  a sub-language of  $L_1$ . The  $L_2$ -module of  $T$  is defined as  $T \cap Fm(L_2)$ .

- Question: is the  $EL$ -module of a  $SROIQ$ -theory  $T$  finitely axiomatizable?
- Given two sub-languages,  $L_1$  and  $L_2$ , of FOL (e.g., OWL 2 ( $SROIQ$ ) and OWL 2 EL),  $L_2 \subseteq L_1$ , and a theory  $T \subseteq Fm(L_1)$ , can we find a finite set  $Ax \subseteq Fm(L_2)$  such that  $Ax \vdash_{L_2} = T \vdash_{L_1} \cap Fm(L_2)$ ?

# Modularization

EL Vira



el-vira

Convert OWL ontologies into the EL, QL or RL subset of OWL

<http://el-vira.googlecode.com>

- ontology modularization
- identify EL, QL, RL axiom patterns in the *deductive closure* of an ontology ⇒ use of reasoner with full support of OWL
- retain signature of ontology
- maximality is open problem

# Modularization

## EL Module

- `'Abnormality of appendix' EquivalentTo: not (has-part some (Appendix and Normal))`
- `'Absent appendix' EquivalentTo: not (has-part some Appendix)`
- `'Absent appendix' SubclassOf: 'Abnormality of appendix'`

# Modularization

## EL Module

- '~~Abnormality of appendix~~' EquivalentTo: ~~not (has-part some (Appendix and Normal))~~
- '~~Absent appendix~~' EquivalentTo: ~~not (has-part some Appendix)~~
- 'Absent appendix' SubclassOf: 'Abnormality of appendix'

Next: collect all the relevant ontologies in the domain

# BioPortal

Welcome to the NCI BioPortal | NCI BioPortal - Chromium

Google Calendar > Monograph of Anter... Nachrichten-SMEC > Welcome to the NCI >

bioportal.bioontology.org

Robert [Sign In | Help | Feedback]

Current Release: 4.22 (November 2015)

BioPortal Browse Search Mappings Recommender Annotator Resource Index Projects

Welcome to BioPortal, the world's most comprehensive repository of biomedical ontologies.

For help using BioPortal, click on this icon:

Search all ontologies  Enter concepts, e.g. definition  Search Advanced Search

Find an ontology  Enter ontology name, e.g. NCI Thesaurus  Explore Browse Ontologies

Search resources  Enter a concept, e.g. Metabolite  Search Advanced Resource Search

Ontology Visits (October 2015)

General-Procedure Terminology (GPT)	38910
SNOMED (SNOMED)	17576
Stereotyped Nomenclature of Medicine - Clinical Terms (SNOMEDCT)	14071
Medical Dictionary for Regulatory Activities (MEDDRA)	13611
National Drug Data File (NDDF)	6447
Mmts	0

Statistics

Ontologies	472
Classes	6,444,217
Resources Indexed	40
Indexed Records	38,359,542
Direct Annotations	95,468,431,792
Direct Plus Expanded Annotations	144,798,362,932

Linked Data

Grade definition (CancerTerminology Criteria for Adverse Event)  
Last updated 2 months ago by bio2rdf  
Hi, we would like for users to be able to range down to the complete definition for each grade i...  
Grade definition (CancerTerminology Criteria for Adverse Event)  
Last updated 1 month ago by bio2rdf  
Hi, we would like for users to be able to range down to the complete definition for each grade i...  
bio2rdf.knowledge (PubMed Central)  
Last updated 2 months ago by bio2rdf  
<http://www.ncbi.nlm.nih.gov/pmc/articles/knowledge/>  
bio2rdf.knowledge (PubMed Central)  
Last updated 2 months ago by bio2rdf  
<http://www.ncbi.nlm.nih.gov/pmc/articles/knowledge/>  
Change Project Value Proposed: Summary proposed for use for "viral hepatitis/drug induced hepatitis" as a term in the ICD-10-CM/ICD-9-CM/ICD-10-AM/ICD-10-CA/ICD-10-PA/ICD-10-CM/ICD-10-AM/ICD-10-CA/ICD-10-PA  
Last updated 8 months ago by bio2rdf  
8 months ago by bio2rdf

Health Entities (HDO) Site of care (SNOMEDCT)  
HDO Mapping (4/20/2015) by page

Site of care (SNOMEDCT) Health Entities (HDO)  
HDO Mapping (4/20/2015) by page

Emergency Medical Services Criteria (HDO) Land ambulance (SNOMEDCT)  
HDO Mapping (4/20/2015) by page

Land ambulance (SNOMEDCT) Emergency Medical Services Entities (HDO)  
HDO Mapping (4/20/2015) by page

Emergency Care Criteria (HDO) Primary care sites (SNOMEDCT)  
HDO Mapping (4/20/2015) by page

THE NATIONAL CENTER FOR  
BIOMEDICAL ONTOLOGY

The National Center for Biomedical Ontology is one of the National Centers of Biomedical Computing supported by the NCIBI, the NIH, and the NCI Cancer Center Fund under grant U54 HG004328.  
Copyright © 2009-2015, The Board of Trustees of Leland Stanford Junior University. All rights reserved.  
NCIBI Website | Release Notes | Terms of Use | Privacy Policy | Site Map

bio2rdf.Bio2Rdf Plugins  
bio2rdf.Annotator  
bio2rdf.Semantic  
bio2rdf.Semantics

- SPARQL endpoint
- ontology recommender
- resource index

# OntoBee

OntoBee : Chirurgie

www.ontobee.org

Home | Introduction | Statistics | SPARQL | Tutorial | FAQs | References | Links | Contact | Acknowledge | News

Welcome to OntoBee!

Database: A [bio2rdf](#) server designed for ontologies. OntoBee is used to facilitate ontology data sharing, visualization, query, integration, and analysis. OntoBee generates [dashboards](#) and presents individual ontology term URLs to (i) HTML web pages for user-friendly web browsing and navigation, and to (ii) RDF source files for Semantic Web applications. OntoBee is the default linked data server for more [GOBID](#) linked library [catalogs](#). OntoBee has also been used for many non-OSBO ontologies.

Please select an ontology [optional]:  Search terms

Jump to <http://purl.obolibrary.org/obo/>

Currently OntoBee has been applied for the following ontologies:

No.	Ontology Prefix	Ontology Full Name	Last update
1	<a href="#">AO</a>	Anatomical Entity Ontology	  
2	<a href="#">AERD</a>	Adverse Event Reporting Ontology	  
3	<a href="#">APO</a>	Assonome phenotype ontology	  
4	<a href="#">APOLLO_SV</a>	Apollo Structured Vocabulary	  
5	<a href="#">ATD</a>	Antibiotic taxonomy	  
6	<a href="#">B2GIO</a>	Beta Cell Genetics Ontology	  
7	<a href="#">BCIO</a>	Biological Collections Ontology	  
8	<a href="#">BFO</a>	Basic Formal Ontology	  
9	<a href="#">BFO211</a>	Basic Formal Ontology (BFO) 1.1	  
10	<a href="#">BLA</a>	Bladder anatomy	  
11	<a href="#">BSPQ</a>	Biological Spatial Ontology	  
12	<a href="#">BTBD</a>	WikiPathways MeSH / Enzyme source	  
13	<a href="#">CABO</a>	Common Anatomy Reference Ontology	  
14	<a href="#">CDAO</a>	Comparative Data Analysis Ontology	  
15	<a href="#">CEPH</a>	Cephalopod Ontology	  
16	<a href="#">CBIO</a>	Chemical Entities of Biological Interest	  
17	<a href="#">CHEMME</a>	Chemical Information Ontology	  
18	<a href="#">CHMO</a>	Chemical Methods Ontology	  
19	<a href="#">CL</a>	Cell Ontology	  
20	<a href="#">CLO</a>	Cell Line Ontology	  
21	<a href="#">CMDO</a>	Clinical measurement ontology	  

[Google Analytics](#)

## Shortcomings

- SPARQL endpoints: different semantics
- OWL reasoning not supported

# Aber-OWL Ontology Repository

<http://aber-owl.net>

- contains a repository of biomedical ontologies (including all OBO ontologies)
- reasoning with the ELK reasoner
  - OWL-EL
  - parallel and incremental reasoning
- web and JSON interface
- query single or multiple ontologies in the repository
- query ontologies on the web

# Aber-OWL Ontology Repository

*"merely using ontologies [...] does not reduce heterogeneity: it just raises heterogeneity problems to a higher level" [Euzenat, 2007]*

- 471 ontologies
- 6,450,141 classes
- 67,786,617 logical axioms
- 3 inconsistent ontologies
- identifies many unsatisfiable classes
  - even using only OWL EL!
  - often due to changing imports

# Aber-OWL Ontology Repository

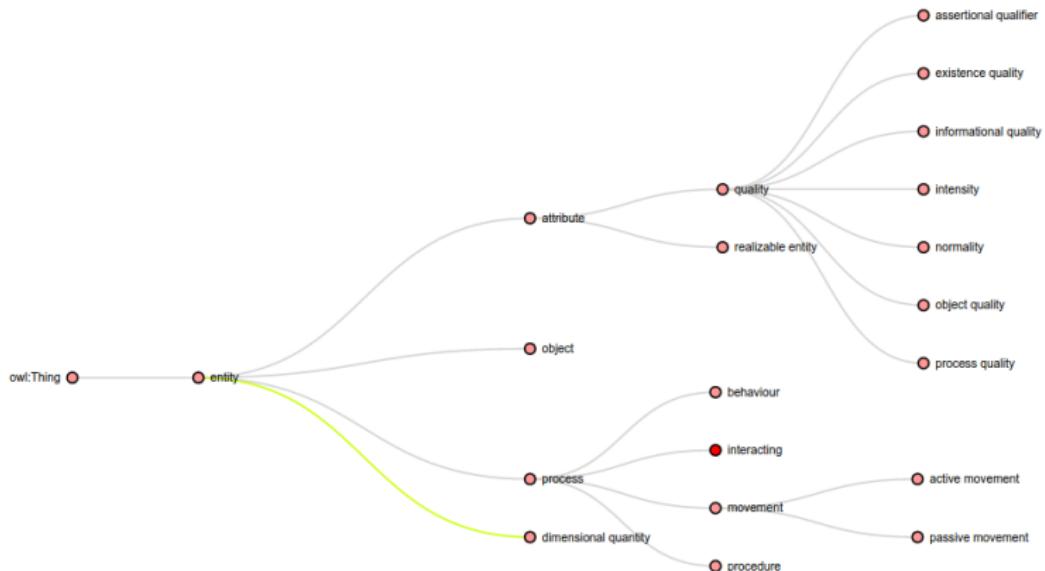
Labels are not just rdfs:label...

- Labels:
  - rdfs:label
  - <http://www.w3.org/2004/02/skos/core#prefLabel>
  - [http://purl.obolibrary.org/obo/IAO\\_0000111](http://purl.obolibrary.org/obo/IAO_0000111)
- Definitions:
  - <http://www.w3.org/2004/02/skos/core#definition>
  - <http://purl.org/dc/elements/1.1/description>
  - <http://www.geneontology.org/formats/-oboInOwl#hasDefinition>
  - ...

## Average query times

- 10ms for a simple query (named classes)
- 240ms for query of complex concept descriptions

# Ontologies and graphs



# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# Aber-OWL: SPARQL

<http://aber-owl.net/aber-owl/sparql/>

Retrieve all proteins annotated to a part of apoptosis:

```
PREFIX GO: <http://purl.uniprot.org/go/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT ?pname ?protein ?label ?ontid WHERE {
  VALUES ?ontid {
    OWL subclass <http://aber-owl.net/aber-owl/service/> <>
      { part_of some 'apoptotic process' }
  } .
  ?protein a up:Protein .
  ?protein up:organism taxon:9606 .
  ?protein up:mnemonic ?pname .
  ?protein up:classifiedWith ?ontid .
  ?ontid skos:prefLabel ?label .
}
```

# Aber-OWL: SPARQL

<http://aber-owl.net/aber-owl/sparql/>

Retrieve all proteins annotated to a part of apoptosis:

```
PREFIX GO: <http://purl.uniprot.org/go/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT DISTINCT ?pname ?protein ?label ?ontid WHERE {
  VALUES ?ontid {
    GO:0071550 GO:0006309 GO:0042771 GO:0070059 GO:0039650 GO:0043154
    GO:0097345 GO:0097297 GO:0008637 GO:0001836 GO:1902109 ...
  } .
  ?protein a up:Protein .
  ?protein up:organism taxon:9606 .
  ?protein up:mnemonic ?pname .
  ?protein up:classifiedWith ?ontid .
  ?ontid skos:prefLabel ?label .
}
```

# Aber-OWL: SPARQL

E5RG18_HUMAN	<a href="http://purl.uniprot.org/uniprot/E5RG18">http://purl.uniprot.org/uniprot/E5RG18</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	extrinsic apoptotic signaling pathway via death domain receptors
Q8IUB7_HUMAN	<a href="http://purl.uniprot.org/uniprot/Q8IUB7">http://purl.uniprot.org/uniprot/Q8IUB7</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	extrinsic apoptotic signaling pathway via death domain receptors
E5RJY0_HUMAN	<a href="http://purl.uniprot.org/uniprot/E5RJY0">http://purl.uniprot.org/uniprot/E5RJY0</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	extrinsic apoptotic signaling pathway via death domain receptors
J7GXU7_HUMAN	<a href="http://purl.uniprot.org/uniprot/J7GXU7">http://purl.uniprot.org/uniprot/J7GXU7</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	extrinsic apoptotic signaling pathway via death domain receptors
V9GY27_HUMAN	<a href="http://purl.uniprot.org/uniprot/V9GY27">http://purl.uniprot.org/uniprot/V9GY27</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	extrinsic apoptotic signaling pathway via death domain receptors
LAMP1_HUMAN	<a href="http://purl.uniprot.org/uniprot/P11279">http://purl.uniprot.org/uniprot/P11279</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	granzyme-mediated apoptotic signaling pathway
BNIP3_HUMAN	<a href="http://purl.uniprot.org/uniprot/Q12983">http://purl.uniprot.org/uniprot/Q12983</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	granzyme-mediated apoptotic signaling pathway
SRGN_HUMAN	<a href="http://purl.uniprot.org/uniprot/P10124">http://purl.uniprot.org/uniprot/P10124</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	granzyme-mediated apoptotic signaling pathway
UBE4B_HUMAN	<a href="http://purl.uniprot.org/uniprot/Q95155">http://purl.uniprot.org/uniprot/Q95155</a>	<a href="http://purl.uniprot.org/go/0008625">http://purl.uniprot.org/go/0008625</a>	granzyme-mediated apoptotic signaling pathway
CRF_HUMAN	<a href="http://purl.uniprot.org/uniprot/P06850">http://purl.uniprot.org/uniprot/P06850</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
PTHY_HUMAN	<a href="http://purl.uniprot.org/uniprot/P01270">http://purl.uniprot.org/uniprot/P01270</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
Q7LCB3_HUMAN	<a href="http://purl.uniprot.org/uniprot/Q7LCB3">http://purl.uniprot.org/uniprot/Q7LCB3</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
ESR2_HUMAN	<a href="http://purl.uniprot.org/uniprot/Q92731">http://purl.uniprot.org/uniprot/Q92731</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
SMS_HUMAN	<a href="http://purl.uniprot.org/uniprot/P61278">http://purl.uniprot.org/uniprot/P61278</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
SSR3_HUMAN	<a href="http://purl.uniprot.org/uniprot/P32745">http://purl.uniprot.org/uniprot/P32745</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway
G3V5S2_HUMAN	<a href="http://purl.uniprot.org/uniprot/G3V5S2">http://purl.uniprot.org/uniprot/G3V5S2</a>	<a href="http://purl.uniprot.org/go/0008628">http://purl.uniprot.org/go/0008628</a>	hormone-mediated apoptotic signaling pathway

## Aber-OWL: SPARQL

- SPARQL query expansion
- integration of DL queries in SPARQL
- one problem: IRIs difference between SPARQL endpoints

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - concept recognition, entity recognition
  - image processing

# The ontology-based data access paradigm

- Conceptual layer
  - conceptual model
  - domain model/ontologies
- Data layer
  - single or federated *structured databases*
  - text
  - images
  - custom files (sequences, genetic variation)
- Mapping between both
  - explicit facts in a knowledge base
  - direct, manual assertions/annotations
  - **concept recognition, entity recognition**
  - image processing

# Aber-OWL: Pubmed

## Concept recognition

- find mentions of ontology classes in text
- common approaches:
  - use labels and synonyms
  - stemming, text segmentation, POS tagging, machine learning

## Aber-OWL: Pubmed

### Aber-OWL: PubMed:

- Lucene index of Medline and PubMed Central
- identify labels and synonyms of class *and subclasses*
- stop words, stemming, case removal
- combine text mining and reasoning over ontologies: find mentions of classes satisfying an OWL queries
- “lightweight” concept recognition, but high speed over large volumes of text (query time < 100ms)

# Aber-OWL: PubMed

Retrieve documents mentioning 'ventricular septal defect'

## [Staged repair of tetralogy of Fallot with pulmonary atresia and major aortopulmonary collateral arteries.](#)

To assess the results of a staged surgical approach for **tetralogy of Fallot with pulmonary atresia**, hypoplastic or absent pulmonary arteries, and major aortopulmonary collateral arteries. We retrospectively reviewed a consecutive series of these patients from a single institution. From July 1993 to April 2001, 46 consecutive patients with **tetralogy of Fallot**, **pulmonary atresia**, and **major aortopulmonary collateral arteries** were treated with staged surgical repair. The operative sequence usually began with a central aortopulmonary shunt followed by unifocalization of

## Aber-OWL: PubMed

- concept recognition incorporates automated reasoning (ontology hierarchy, axiom patterns)
- recognition of complex concepts (class expressions)
- useful for difficult concepts (gene functions, processes, phenotypes)

## Aber-OWL: PubMed

<http://aber-owl.net/aber-owl/diseasephenotypes/>

- find phenotypes (signs and symptoms) associated with common diseases
  - no resource available for comparison
- pattern-based mining of literature with Aber-OWL: PubMed
- evaluation (of genetically based disease phenotypes) with experimentally validated disease genes

# Aber-OWL: PubMed

<http://aber-owl.net/aber-owl/diseasephenotypes/>

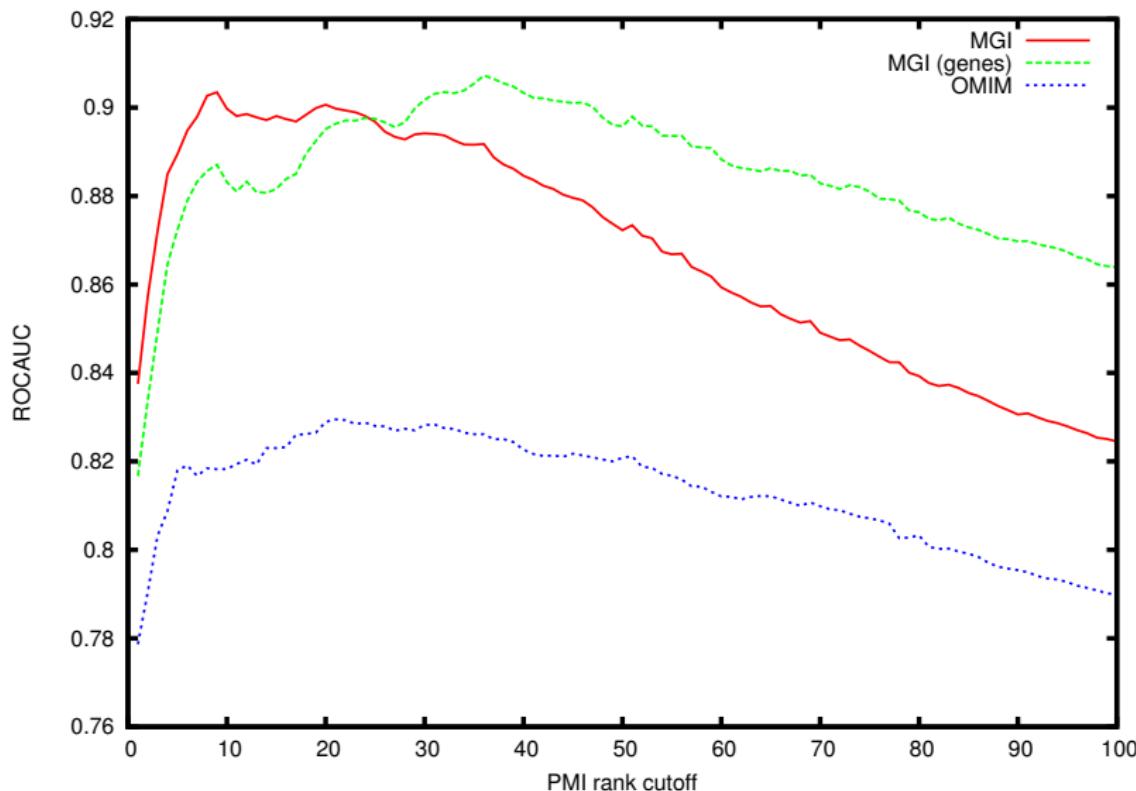
## Phenotypes for 'bubonic plague'

- [Mouse models](#)
- [Network](#)

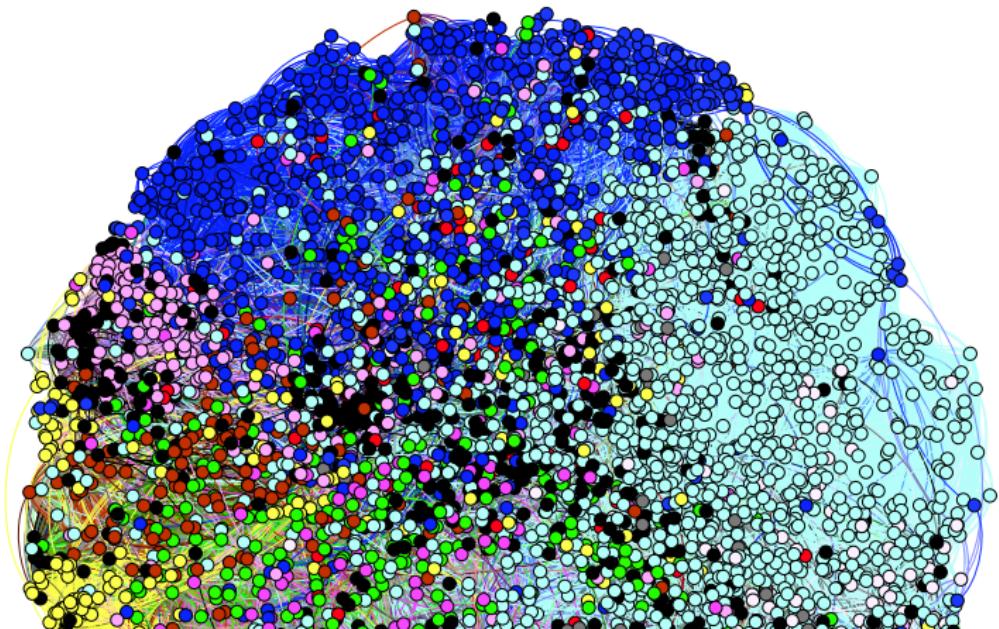
Show	100	entries
Phenotype	ID	Pointwise Mutual Information
death	MP:0001641	0.22837393990989352
Increased IgA level	HP:0003261	0.21745157416550895
eye lesions	MP:0013146	0.21110960811807034
Lymphadenitis	HP:0002840	0.20726444631165528
lymph node inflammation	MP:0003865	0.2070583004125834
Mediastinal lymphadenopathy	HP:0100721	0.16667331911036262
Abnormality of the lymph nodes	HP:0002733	0.14851361059698084
decreased susceptibility to infection	MP:0002409	0.1442019284611426
Pustule	HP:0200039	0.10968881998477503
tachypnea	MP:0005426	0.10850084511186676
Atrioventricular canal defect	HP:0006695	0.10364210532782134
Tachypnea	HP:0002789	0.10273022349597967
increased pulmonary respiratory rate	MP:0005573	0.09848893732226834

# Aber-OWL: PubMed

<http://aber-owl.net/aber-owl/diseasephenotypes/>



# Aber-OWL: PubMed



# Diseases and phenotypes

## Phenotypes

Phenotypes are the observable characteristics of an organism arising from its *genotype* and its response to the *environment*.

Analysis of phenotypes should reveal information about

- genotype,
- environment,
- mechanisms and processes that determine phenotype from genotype.

# Phenotypes

In biodiversity and ecology



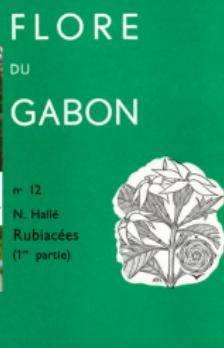
MALAYSIAN FOREST RECORDS NO. 51

Revision of Malaysian Species of  
Boletales s. l. (Basidiomycota)  
Described by E.J.H. Corner  
(1972, 1974)

E. Horak



Flora Malesiana



FLORA ZAMBESIACA

EDITED BY  
J.R. THOMAS & S.A. MARTINS  
ON BEHALF OF THE EDITORIAL BOARD

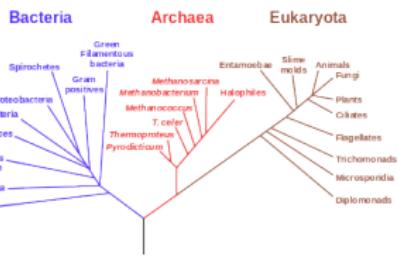
ROYAL BOTANIC GARDEN, KENYA  
FOR FLORA ZAMBESIACA MANAGING COMMITTEE

# Phenotypes

## In evolutionary biology

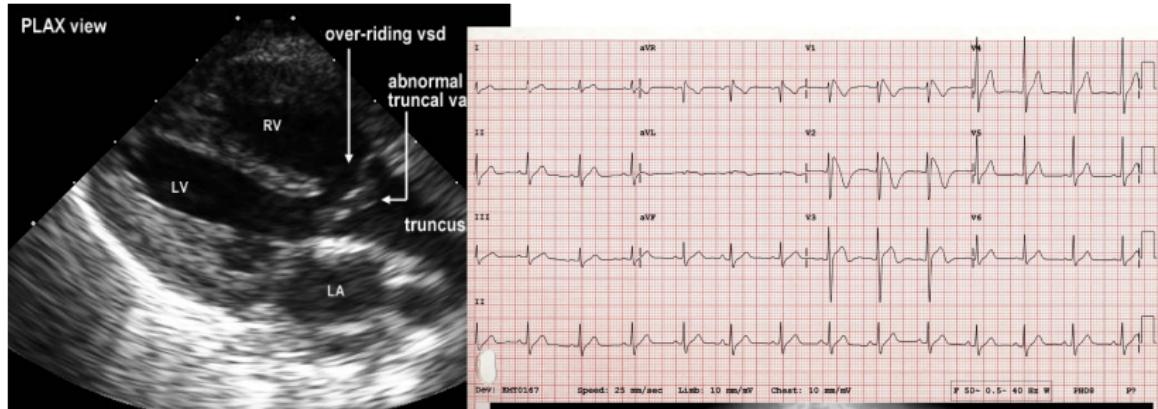
Taxon \ Character	1 Inflorescence position	2 Inflorescence emerges	3 Stamen abscission	4 Type of stamen spine?	5 Lateral staminal dev.	6 Staminal spine types?
1 homhai	Separated	?	No	?	absent	?
2 etlico	Separated	?	No	?	absent	?
3 alipgal	Not separated	Terminal	No	?	present	teeth-like
4 etfiela	Separated	?	No	?	absent	?
5 homspi	Separated	?	No	?	absent	?
6 globvel	Not separated	Terminal	yes	anther crest	present	petal-like
7 boespul	Not separated	Between Sheath	No	?	present	petal-like
8 elebsur	Separated	?	yes	anther crest	absent	?
9 globeshi	Not separated	Terminal	yes	?	present	petal-like
10 plagaga	Not separated	From Basal Sheat	No	?	present	teeth-like
11 burbst	Not separated	Terminal	No	?	present	petal-like
12 plagcam	Not separated	From Basal Sheat	No	?	present	teeth-like
13 globatr	Not separated	Terminal	yes	4 triangular	present	petal-like
14 etlifilm	Separated	?	yes	anther crest	absent	?
15 hedicyl	Not separated	Terminal	No	?	present	petal-like

## Phylogenetic Tree of Life



# Phenotypes

## In the clinic



### Complete Blood Count:

Patient	Value	Normal Range
		2 years - 6 years

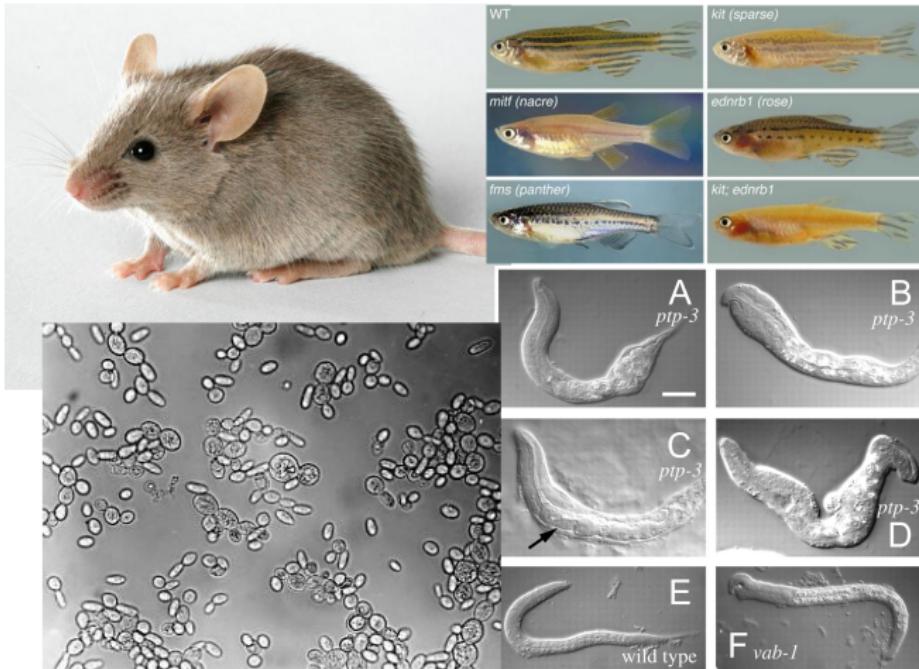
WBC	$8.4 \times 10^9 / L$	( 5.0 – 17.0 )
RBC	$2.77 \times 10^{12} / L$	( 3.90 – 5.30 )
Hgb	7.5 g/dl	( 11.5 – 13.5 )
Hct	21.8 %	( 34.0 – 40.0 )
MCV	78.6 fl	( 75.0 – 87.0 )
MCH	26.9 pg	( 25.0 – 31.0 )
MCHC	34.2 gm/dl	( 31.0 – 36.0 )
RDW	17.3 %	( 11.5 – 15.0 )
PLT	$192 \times 10^9 / L$	( 150 – 450 )

### Differential:

	Absolute	Normal Range
	Number	2 years - 6
Neutrophils	43 %	( 3.61 )
Bands	6 %	( 0.00 – 1.00 )

# Phenotypes

In the lab



# Phenotypes

## Approach

We want to:

- ① describe phenotypes formally
  - morphology
  - function
  - ⇒ formal ontology

# Phenotypes

## Approach

We want to:

- ① describe phenotypes formally
  - morphology
  - function
  - ⇒ formal ontology
- ② integrate/compare phenotypes (within and between species)
  - homologous organ structures
  - related/identical function
  - ⇒ ontologies and automated reasoning

# Phenotypes

## Approach

We want to:

- ① describe phenotypes formally
  - morphology
  - function
  - ⇒ formal ontology
- ② integrate/compare phenotypes (within and between species)
  - homologous organ structures
  - related/identical function
  - ⇒ ontologies and automated reasoning
- ③ understand genotype-phenotype relations
  - use morphological or functional similarity
  - similarity in attribute values
  - ⇒ semantic similarity

# Phenotype representation

## EQ formalism

Entity + Quality = Phenotype

A phenotype can be described by an entity and its quality.

# Phenotype representation

## EQ formalism

Entity + Quality = Phenotype

A phenotype can be described by an entity and its quality.



Flower: red

# Phenotype representation

## EQ formalism

Entity + Quality = Phenotype

A phenotype can be described by an entity and its quality.



Flower: red

Flower: PO:0009046 (Plant Ontology)

# Phenotype representation

## EQ formalism

Entity + Quality = Phenotype

A phenotype can be described by an entity and its quality.



Flower: red

Flower: PO:0009046 (Plant Ontology)

Red: PATO:0000322 (Phenotype And Trait Ontology)

# Phenotype representation

## EQ formalism

Entity 1 + Quality + Entity 2 = Phenotype

Some qualities involve a second entity.



# Phenotype representation

## EQ formalism

Entity 1 + Quality + Entity 2 = Phenotype

Some qualities involve a second entity.



Leaf: PO:0025034 (Plant Ontology)

# Phenotype representation

## EQ formalism

Entity 1 + Quality + Entity 2 = Phenotype

Some qualities involve a second entity.



Leaf: PO:0025034 (Plant Ontology)

responsive to: PATO:0000487 (Phenotype And Trait Ontology)

# Phenotype representation

## EQ formalism

Entity 1 + Quality + Entity 2 = Phenotype

Some qualities involve a second entity.



Leaf: PO:0025034 (Plant Ontology)

responsive to: PATO:0000487 (Phenotype And Trait Ontology)

detection of mechanical stimulus: GO:0050982 (Gene Ontology)

# Phenotype representation

## EQ formalism

Entity + Quality = Phenotype

Entity 1 + Quality + Entity 2 = Phenotype

- Entity from reference ontology (Gene Ontology, anatomy ontology)
- Quality from PATO ontology

## Phenotype representation

A phenotype is a quality that inheres in its bearer [Mungall, 2010].

- Red flower:  $\text{Red}(x) \wedge \exists y(\text{inheresIn}(x, y) \wedge \text{Flower}(y))$
- $\text{Red} \sqcap \exists \text{inheresIn}.\text{Flower}$
- Red: reuse an ontology of qualities
- Flower: reuse ontology of plant anatomy

# Phenotype representation

EQ formalism: semantics

Absent tail:

- $\text{AbsentTail} \equiv \text{Absent} \sqcap \exists \text{inheresIn}. \text{Tail}$

# Phenotype representation

EQ formalism: semantics

Absent tail:

- $\text{AbsentTail} \equiv \text{Absent} \sqcap \exists \text{inheresIn.Tail}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards.Tail} \sqcap \exists \text{inheresIn.MouseBody}$

# Phenotype representation

EQ formalism: semantics

Absent tail:

- $\text{AbsentTail} \equiv \text{Absent} \sqcap \exists \text{inheresIn}. \text{Tail}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \text{Tail} \sqcap \exists \text{inheresIn}. \text{MouseBody}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \{ \text{Tail} \} \sqcap \exists \text{inheresIn}. \text{MouseBody}$   
(Mungall, 2007)

# Phenotype representation

EQ formalism: semantics

Absent tail:

- $\text{AbsentTail} \equiv \text{Absent} \sqcap \exists \text{inheresIn}. \text{Tail}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \text{Tail} \sqcap \exists \text{inheresIn}. \text{MouseBody}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \{ \text{Tail} \} \sqcap \exists \text{inheresIn}. \text{MouseBody}$   
(Mungall, 2007)
- no use of **part-of** or **has-part** relations
- no reuse of **part-of** in anatomy ontology

# Phenotype representation

EQ formalism: semantics

Absent tail:

- $\text{AbsentTail} \equiv \text{Absent} \sqcap \exists \text{inheresIn}. \text{Tail}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \text{Tail} \sqcap \exists \text{inheresIn}. \text{MouseBody}$
- $\text{AbsentTail} \equiv$   
 $\text{LacksParts} \sqcap \exists \text{towards}. \{ \text{Tail} \} \sqcap \exists \text{inheresIn}. \text{MouseBody}$   
(Mungall, 2007)
- no use of **part-of** or **has-part** relations
- no reuse of **part-of** in anatomy ontology
- Observation: having or lacking parts (functions, dispositions, processes, ...) are not *primarily* qualities.

# Phenotypic properties

## Definition

A *phenotypic property* (phene) is the basic observable characteristic possessed by an organism.

- **phene-of** and **has-phene** relations
- examples: having weight, being red, not having an appendix as part, participating in mating behaviour
- based on a *defining class Y*

Definition pattern: X EquivalentTo: pheneof some Y

# Properties

## Taxonomy of properties

Top-level ontology of properties based on defining class  $Y$ :

- Structural: having parts, lacking parts, being part of, not being part of
  - $\exists pheneOf.(Mouse \sqcap \exists hasPart.Tail)$
- Qualitative: having qualities, lacking qualities, having quality-values, not having quality-values
  - $\exists pheneOf.(Flower \sqcap \exists hasQuality.Red)$
- Functional: being capable to do  $X$ , not being capable to do  $X$ , being dysfunctional w.r.t.  $F$ 
  - $\exists pheneOf.(Heart \sqcap \exists hasFunction.PB)$
- Participatory: pumping blood, being a catalyst
  - $\exists pheneOf.(Heart \sqcap \exists participatesIn.PB)$

# Phenotype representation

- *phene-of* is functional:

$$\forall x, y, z (\text{pheneOf}(x, y) \wedge \text{pheneOf}(x, z) \rightarrow y = z)$$

- alternatively:  $\forall x, y (\text{pheneOf}(x, y) \rightarrow \text{inheritsIn}(x, y))$ , inherit functionality, NMP from GFO

- Absent tail:

$$\text{AbsentTail} \sqsubseteq \exists \text{pheneOf}.(\text{Organism} \sqcap \neg \exists \text{hasPart}. \text{Tail})$$

- Dysfunctional heart:

$$DH \sqsubseteq \exists \text{pheneOf}.(\text{Heart} \sqcap \neg \exists \text{hasFunction}. \text{PumpingBlood})$$

# Absence

Absent tail



- $\text{AbsentTail} \sqsubseteq \exists \text{pheneOf}.(\text{Mouse} \sqcap \neg \exists \text{hasPart}. \text{Tail})$
- MA:  $\text{Mouse} \sqsubseteq \exists \text{hasPart}. \text{Tail}$
- $\text{Mouse}(\text{Jerry}), \text{AbsentTail}(x), \text{hasPhene}(\text{Jerry}, x)$

## Absence

Absent appendix

- Removal of conflicting axioms (has-part/part-of in anatomy)

# Absence

## Absent appendix

- Removal of conflicting axioms (has-part/part-of in anatomy)
- Contextualize anatomy:
  - $\text{Normal} \sqcap \text{Mouse} \sqsubseteq \exists \text{hasPart}.(\text{Normal} \sqcap \text{Tail})$
  - Needs one *Normal* for each axiom

# Absence

## Absent appendix

- Removal of conflicting axioms (has-part/part-of in anatomy)
- Contextualize anatomy:
  - $\text{Normal} \sqcap \text{Mouse} \sqsubseteq \exists \text{hasPart}.(\text{Normal} \sqcap \text{Tail})$
  - Needs one *Normal* for each axiom
- Rewrite anatomy using non-monotonic reasoning:
  - *Normally*:  $\text{Mouse} \sqsubseteq \exists \text{hasPart}. \text{Tail}$
  - Circumscription of  $\neg \text{Normal}$
  - Better: use default logic and answer set programming
  - $\text{Mouse}(x) : \text{hasPart}(x, y) \wedge \text{Tail}(y) / \text{hasPart}(x, y) \wedge \text{Tail}(y)$
  - Implementation in dlvhex
  - IC-has-part(X,Y) :- ind(X), class(Y), inst(X,Z),  
CC-normally-has-part(Z,Y), not  
IC-lacks-has-part(X,Y), class(Z).

## Phenotype representation

Ontologies can be used to *formally* describe phenotypes and integrate phenotype descriptions with ontologies of anatomy and functions. We applied this method to

- Arabidopsis Information Resource
- Gramene
- WormBase
- FlyBase
- Saccharomyces Genome Database
- Mouse Genome Informatics database
- Zebrafish Model Organism Database
- OMIM
- OrphaNet
- ...

## Phenotype representation

Ontologies can be used to *formally* describe phenotypes and integrate phenotype descriptions with ontologies of anatomy and functions. We applied this method to

- Arabidopsis Information Resource
- Gramene
- WormBase
- FlyBase
- Saccharomyces Genome Database
- Mouse Genome Informatics database
- Zebrafish Model Organism Database
- OMIM
- OrphaNet
- ...

... but all of them have different anatomy and physiology.

## Crossspecies integration

Can we use this framework for data integration *across* species?

# Crossspecies integration

- UBERON ontology of homologous organ structures
  - *Heart (human)* homologous to *Heart (mouse)*

# Crossspecies integration

- UBERON ontology of homologous organ structures
  - *Heart (human)* homologous to *Heart (mouse)*
  - *Tail (mouse)* homologous to... ?

# Crossspecies integration

- UBERON ontology of homologous organ structures
  - *Heart (human)* homologous to *Heart (mouse)*
  - *Tail (mouse)* homologous to... ?
  - *Tail (mouse)* SubClassOf: part-of some *Trunk (mouse)*
  - *Trunk (mouse)* homologous to *Trunk (human)*

What if we treat homologous organ structures as *equivalent* (for this purpose)?

# Crossspecies integration

- Human absent appendix:  
 $\exists pheneOf.(Human \sqcap \neg \exists hasPart.HumanAppendix)$
- Mouse absent appendix:  
 $\exists pheneOf.(Mouse \sqcap \neg \exists hasPart.MouseAppendix)$
- Mouse homologous to (equivalent to) Human
- MouseAppendix homologous to (equivalent to)  
HumanAppendix
- $\Rightarrow$  Human absent appendix equivalent to Mouse absent appendix

## Crossspecies integration

- Mouse absent tail:  
 $\exists pheneOf.(Mouse \sqcap \neg \exists hasPart.MouseTail)$
- MouseTail homologous to (equivalent to) ???

## Crossspecies integration

- Mouse absent tail:  
 $\exists pheneOf.(Mouse \sqcap \neg \exists hasPart.MouseTail)$
- MouseTail homologous to (equivalent to) ???
- Infer using mouse anatomy

# Crossspecies integration

Starting with a pair of Entity and Quality, generate the following classes:

- $E\text{Phenotype} \equiv \exists pheneOf.(\exists partOf.E \sqcap \exists hasQuality.\top)$
- ' $E Q$ '  $\equiv \exists pheneOf.(E \sqcap \exists hasQuality.Q)$ 
  - or any of the other structural forms, depending on  $Q$
- assert equivalence between  $E$ s in different species (based on homology)
- then, include anatomy ontologies

# Crossspecies integration

Starting with a pair of Entity and Quality, generate the following classes:

- $E\text{Phenotype} \equiv \exists pheneOf.(\exists partOf.E \sqcap \exists hasQuality.\top)$
- ' $E Q$ '  $\equiv \exists pheneOf.(E \sqcap \exists hasQuality.Q)$ 
  - or any of the other structural forms, depending on  $Q$
- assert equivalence between  $E$ s in different species (based on homology)
- then, include anatomy ontologies
- AbsentTail (mouse) will become a subclass of TrunkPhenotype (mouse, human)

# Phenotype representation

- Arabidopsis Information Resource
- Gramene
- WormBase
- FlyBase
- Saccharomyces Genome Database
- Mouse Genome Informatics database
- Zebrafish Model Organism Database
- OMIM
- OrphaNet

# Phenotype representation

- Arabidopsis Information Resource
- Gramene
- WormBase
- FlyBase
- Saccharomyces Genome Database
- Mouse Genome Informatics database
- Zebrafish Model Organism Database
- OMIM
- OrphaNet
- more than 250,000 formal phenotype descriptions
- > 2,000,000 axioms
- ⇒ Elvira modularization method

## Crossspecies integration

- Classify the resulting ontology using a OWL (EL) reasoner

## Crossspecies integration

- Classify the resulting ontology using a OWL (EL) reasoner
- 70% of classes are unsatisfiable... why?

## Crossspecies integration

- Classify the resulting ontology using a OWL (EL) reasoner
- 70% of classes are unsatisfiable... why?
- it's not so easy to combine different anatomy ontologies; different conceptualizations!
  - Anus (human) is an orifice which is a kind of *immaterial anatomical entity*; Anus (mouse) is a *material entity*
  - *immaterial anatomical entity* and *material anatomical entity* are disjoint in human anatomy
- a (lossy) solution: get rid of all the disjointness axioms

# Crossspecies integration

We can now

- formally describe phenotypes
- integrate phenotype with anatomy and physiology ontologies
- integrate phenotypes across species (with some losses)
- integrate **disease and model organism phenotypes**

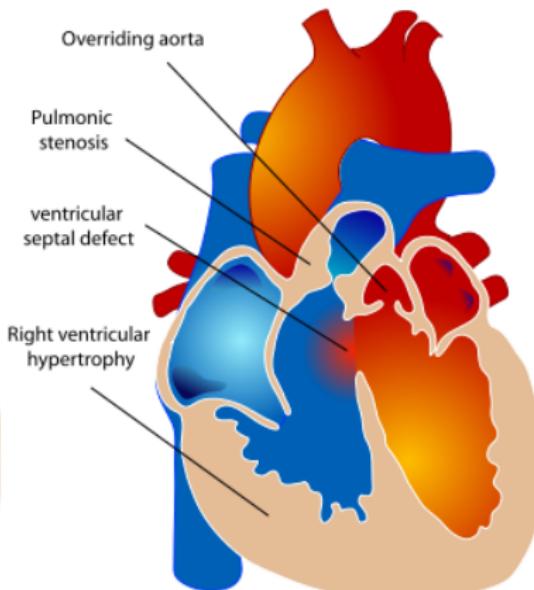
We can now compare *human* phenotypes (in diseases, drug effects)  
with *animal model* phenotypes!

# Integration

Normal heart



Tetralogy of Fallot



# Integration

## Human phenotypes

- Overriding aorta (HP:0002623)
- Ventricular septal defect (HP:0001629)
- Pulmonic stenosis (HP:0001642)
- Right ventricular hypertrophy (HP:0001667)

# Application

## Comparison of phenotypes

phenotype of mutations subclass of disease phenotype allows inference of gene-disease association if

- disease phenotypes *sufficient* for having the disease
- mutation phenotypes *necessary* for having a specific genotype

# Analyzing phenotypes

Phc1 knockout mice

Affected Systems	Genotypes:	
	hm1	hm2
<b>cardiovascular system</b>		✓
pulmonary trunk hypoplasia		✓
abnormal cardiovascular development		✓
abnormal heart looping		✓
abnormal bulbus cordis morphology		✓
abnormal outflow tract development		✓
abnormal heart morphology		✓
overriding aorta		✓
ventricular septal defect		✓
heart right ventricle hypertrophy		✓
abnormal semilunar valve morphology		✓
aortic valve stenosis		✓
pulmonary valve stenosis		✓
abnormal heart right ventricle outflow tract morphology		✓
dilated heart ventricle		✓
thin ventricular wall		✓

# Analyzing phenotypes

Integration of phenotype ontologies enables identification of disease phenotypes in mice.

Affected Systems	Genotypes:	hm1	hm2
<b>cardiovascular system</b>	▼	✓	
pulmonary trunk hypoplasia		✓	
abnormal cardiovascular development		✓	
abnormal heart looping		✓	
abnormal bulbus cordis morphology		✓	
abnormal outflow tract development		✓	
abnormal heart morphology		✓	
overriding aorta		✓	
ventricular septal defect		✓	
heart right ventricle hypertrophy		✓	
abnormal semilunar valve morphology		✓	
aortic valve stenosis		✓	
pulmonary valve stenosis		✓	
abnormal heart right ventricle outflow tract morphology		✓	
dilated heart ventricle		✓	
thin ventricular wall		✓	

# Analyzing phenotypes

- Overriding aorta (MP:0000273)
- Ventricular septal defect (MP:0010402)
- Pulmonary valve stenosis (MP:0006128)
- Heart right ventricle hypertrophy (MP:0000276)
- ...

# Analyzing phenotypes

4,000 genetic diseases in OMIM, 6,000 in OrphaNet, have an unknown molecular basis



Online Mendelian Inheritance in Man®

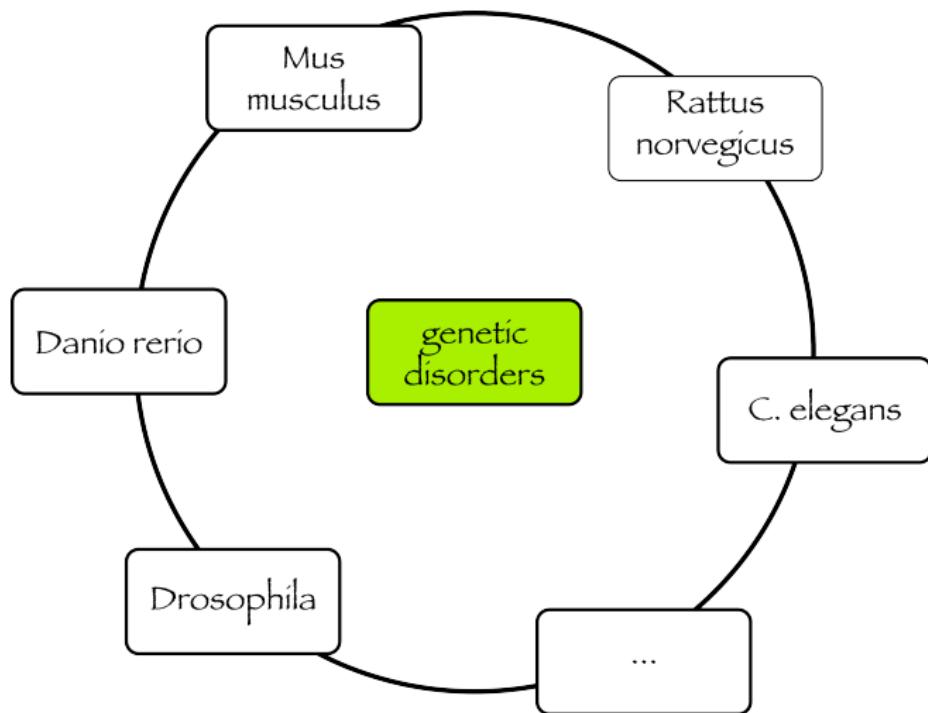
An Online Catalog of Human Genes and Genetic Disorders

Number of Entries:

Prefix	Autosomal	X Linked	Y Linked	Mitochondrial	Totals
* Gene description	12,750	627	48	35	13,460
+ Gene and phenotype, combined	250	14	0	2	266
# Phenotype description, molecular basis known	2,836	240	4	28	3,108
% Phenotype description or locus, molecular basis unknown	1,628	135	5	0	1,768
Other, mainly phenotypes with suspected mendelian basis	1,819	130	2	0	1,951
Totals	19,283	1,146	59	65	20,553



# Analyzing phenotypes



# Analyzing phenotypes

Semantic similarity over phenotype ontologies measures phenotypic similarity

- use methods from IR
- semantic similarity: similarity measure based on information contained in the axioms/structure of an ontology
  - anatomy: front limb – hind limb vs. front limb – eye
  - function: detection of salty taste – detection of sweet taste vs. detection of salty taste – apoptosis
  - quality: red – orange vs. red – green vs. red – round
- ⇒ phenotypic similarity combines similarity between anatomy, function, and quality

# Analyzing phenotypes

Information content of phenotype:

$$IC(x) = -\log(p(x))$$

Phenotype similarity:

$$sim(P, D) = \frac{\sum_{x \in CI(P) \cap CI(D)} IC(x)}{\sum_{y \in CI(P) \cup CI(D)} IC(y)}$$

# Analyzing phenotypes

Information content of phenotype:

$$IC(x) = -\log(p(x))$$

Phenotype similarity:

$$sim(P, D) = \frac{\sum_{x \in CI(P) \cap CI(D)} IC(x)}{\sum_{y \in CI(P) \cup CI(D)} IC(y)}$$

⇒ systematic, pairwise comparison of disease and model organism phenotypes

# Analyzing phenotypes

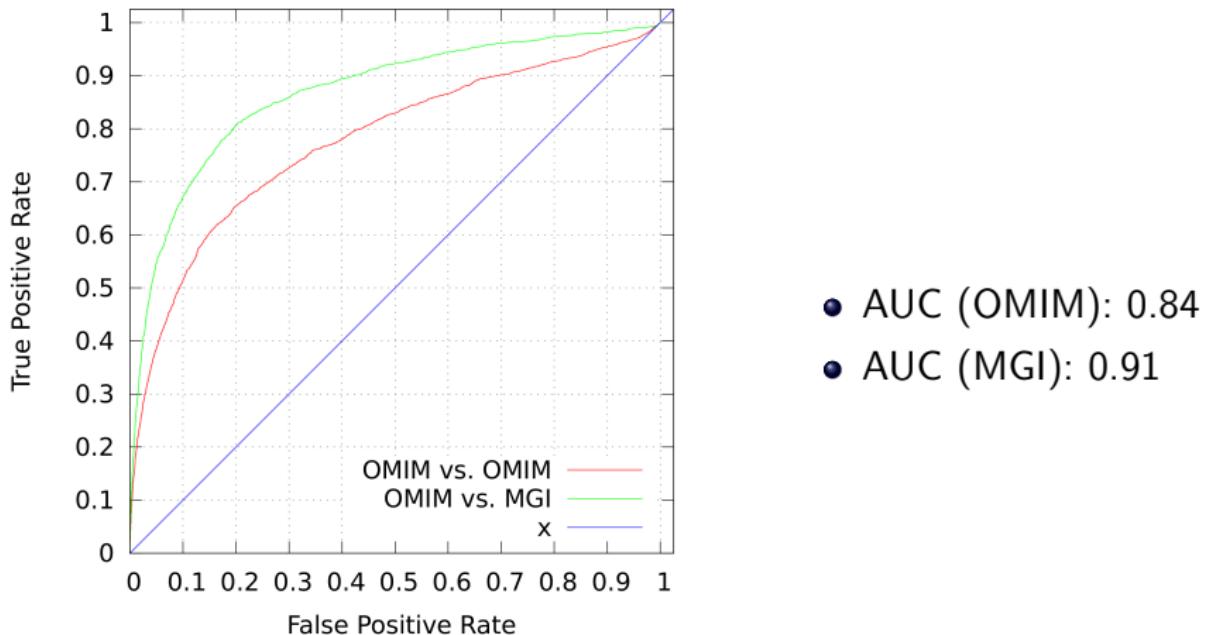
disease id	rank 1	rank 2	rank 3	...
604290	rd14	nmf242	nmf127	...
100050	frg	Spry4<tm1Ayos>	Ryk<tm1Stac>	...
100070	Atp7a<Mo-to>	Slc2a10<SI50F>	Pdgfrb<tm7Sor>	...

# Analyzing phenotypes

disease id	rank 1	rank 2	rank 3	...
604290	rd14	nmf242	nmf127	...
100050	frg	Spry4<tm1Ayos>	Ryk<tm1Stac>	...
100070	Atp7a<Mo-to>	Slc2a10<SI50F>	Pdgfrb<tm7Sor>	...

How well does this approach recover known disease genes?

# Analyzing phenotypes



# Analyzing phenotypes

- *Adam19* and *Fgf15* in mice and (mammalian homologs of) *Cx36.7* and *Nkx2.5* in zebrafish are candidates for Tetralogy of Fallot
- Gene disease associations for orphan diseases
  - *Slc34a1* (MGI:1345284) and Fanconi renotubular syndrome 1 (OMIM:134600)
  - *Hip1* and Bassoe syndrome
- Disease pathways
  - Cytokine-cytokine receptor interaction pathway (ko04060) is significantly correlated with Tetralogy of Fallot ( $p = 5 \cdot 10^{-7}$ , Wilcoxon signed-rank test)