

Integrating systems biology and biomedical ontologies

Robert Hoehndorf

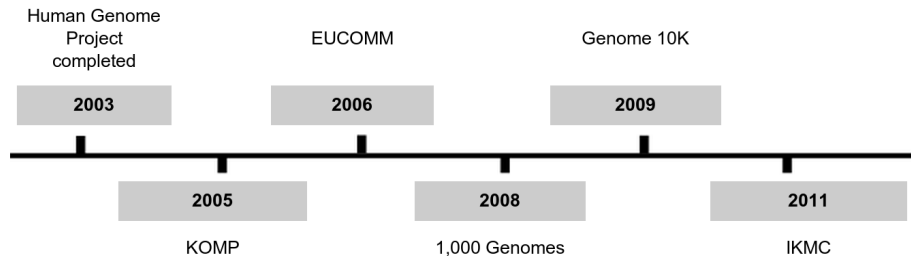
Department of Genetics
University of Cambridge

COMBINE 2011

Work by...

- George Gkoutos (Cambridge)
- Michel Dumontier (Ottawa)
- Paul Schofield (Cambridge)
- Dan Cook (Seattle)
- Bernard de Bono (Hinxton, EBI)
- John Gennari (Seattle)
- Heinrich Herre (Leipzig)
- Janet Kelso (Leipzig)
- Anika Oellrich (Hinxton, EBI)
- Dietrich Rebholz-Schuhmann (Hinxton, EBI)
- Sarala Wimalaratne (Hinxton, EBI)

Motivation

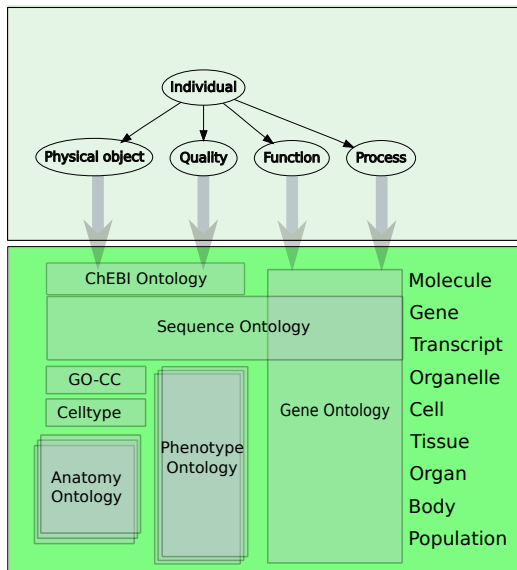


Ontology

- ontology (philosophy) studies the nature of existence and categories of being
- an ontology (computer science) is the “explicit specification of a conceptualization of a domain” [Gruber, 1993]
- ontologies specify the *meaning* of terms in a vocabulary
- *formalized* ontologies can be used by computers and automated systems

Applied ontology is the branch of knowledge representation that focuses on *how to represent the content*.

Biomedical ontology



Systems biology

Systems biology...is about putting together rather than taking apart, integration rather than reduction. [Denis Noble]

- multi-scale data integration
 - domains and levels of granularity
 - species
 - kinds of data
- integration of *in silico*, *in vitro* and *in vivo* research
 - focus on emergent properties
- simulation of biological systems
 - predict and simulate systems' behavior

Systems biology

Challenges (Kitano, 2002)

- data integration
- validation
- standard languages
 - specification
 - exchange
 - results

Systems biology

Challenges (Kitano, 2002)

- data integration
- validation
- standard languages
 - specification
 - exchange
 - results

Can we use ontologies to address some of these problems?

Data integration on the *knowledge level*

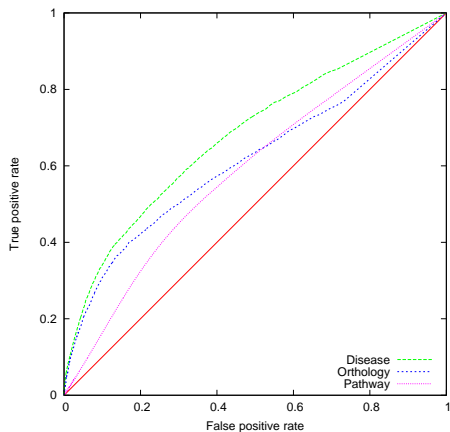
- make the ontological commitment of a biological database, language, representation format, etc. explicit
- integrate with biomedical ontologies
 - utilize ontology-based annotations
- access through automated reasoning (deductive proofs)
- connect ontologies to establish new connections between data

Connecting phenotypes across species

- integrate yeast, fly, worm, fish, mouse, human phenotype databases
- integration of anatomy and phenotype ontologies
 - exploit through OWL reasoning
 - more than 300,000 classes and 1,000,000 axioms
- semantic similarity
- quantitative evaluation based on predicting orthology, pathway, disease

H, R., Schofield, P. N., and Gkoutos, G. V. (2011). PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Research*.

Connecting phenotypes across species



Drugs and phenotypes

- formalization of drug databases
- integration with biomedical ontologies (GO, DO, ATC, ChEBI, ...)
- quantitative evaluation based on predicting drug–disease associations
- <https://code.google.com/p/phenomeblast/wiki/PhenomeDrug>

H, R., Oellrich, A., Rebholz-Schuhmann, D., Schofield, P., and Gkoutos, G. (2011) Linking PharmGKB to phenotype studies and animal models of disease for drug repurposing. *Under review*.

H, R., Dumontier, M., and Gkoutos, G. Ontology-based integration of knowledge in pharmacogenomics (2011). *Under review*.

Further applications

- Pharmacogenomics: <http://code.google.com/p/pharmgkb-owl/>
- Sequence data (GFF3):
<https://code.google.com/p/gff3-to-owl/>
- Pathway data: (Michel's talk)
- Phenotypes: <http://code.google.com/p/phenomeblast/>
- Literature annotations: <http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/home/wiki.cgi?page=Relation+annotation>

work by Michel Dumontier, George Gkoutos, Axel Ngonga, Sampo Pyysalo, RH

MIRIAM annotations

Annotation of SBML

```

xmlns:bqmodel="http://biomodels.net/model-qualifiers/">
  <rdf:Description rdf:about="#_525530">
    <bqbiol:is>
      <rdf:Bag>
        <rdf:li rdf:resource="urn:miriam:obo_chebi:CHEBI%3A4167"/>
        <rdf:li rdf:resource="urn:miriam:kegg.compound:C00031"/>
      </rdf:Bag>
    </bqbiol:is>
  </rdf:Description>
</rdf:RDF>
</annotation>

```

MIRIAM annotations

Annotation of SBML

- MIRIAM provides *annotation* of SBML entities
- ontologies are treated as *meta-data*
 - search
 - semantic similarity
 - documentation
- no integration with modelling language

MIRIAM annotations

Information flow hypothesis

Integration of SBML and ontologies could lead to *information flow* between models and ontologies.

Information flow enables the use of ontologies for

- verification,
- access to data,
- integration and combination of models *on the knowledge level*.

MIRIAM annotations

Reference Publication	
Publication ID: 2904647	Mol Pharmacol 1988 Dec;34(6):814-22. Inhibition of adenylate cyclase is mediated by the high affinity conformation of the al Thomsen WJ, Jacquez JA, Neubig RR. Department of Pharmacology, University of Michigan, Ann Arbor 48109-0626. [more]
Model	
Original Model: BIOMD0000000082.xml.origin	set #1 bqbiol:isVersionOf Gene Ontology negative regulation of adenylate cyclase Gene Ontology heterotrimeric G-protein complex cycle
Submitter: Enuo He	
Submission ID: MODEL1166069080	
Submission Date: 27 Nov 2006 22:16:10 UTC	
Last Modification Date: 21 Aug 2008 12:08:11 UTC	
Creation Date: 26 Sep 2006 18:47:38 UTC	

Ontological commitment

Rule 1: models

Model M annotated with $A1$:

- M represents an object O_1
- O_1 can have functions
- O_1 's functions can be realized by processes
- model components represent parts of O_1
- `M SubClassOf: represents some A1`
- `M SubClassOf: represents some (has-function some A1)`
- `M SubClassOf: represents some (has-function some (realized-by only A1))`

Ontological commitment

BioModel 82

annotated with *heterotrimeric G-protein complex cycle* (GO:0031684):

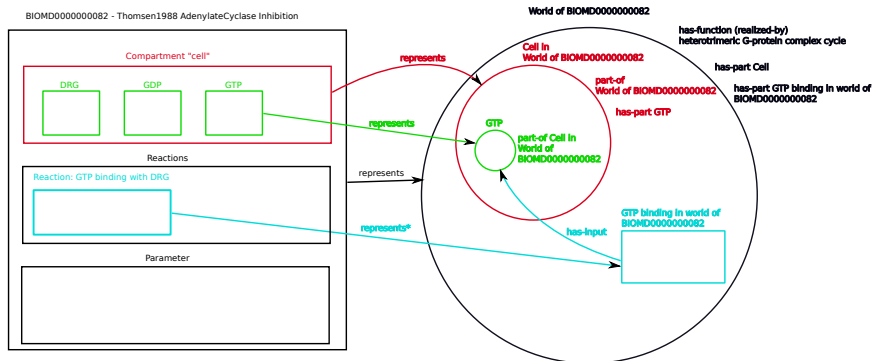
- represents an object O_1
- O_1 has a function F_1
- F_1 is realized by processes of the type *heterotrimeric G-protein complex cycle*
- M SubClassOf: represents some O1
- O1 SubClassOf: (has-function some (realized-by only GO:0031684))

Ontological commitment

- rules for compartments, species, reactions
- rules reflect the structure of models
- integration with bio-ontologies based on MIRIAM annotations
- OWL-based formal semantics for SBML

Ontological commitment

Reaction GTP-binding in BioModel 82



Ontological commitment

BioModels Result

Ontologies:

- FMA
- ChEBI
- GO
- Celltype
- PATO
- (KEGG, Reactome)

Result on BioModels (18):

- more than 300,000 classes
- more than 800,000 axioms
- 90,000 complex model annotations

<http://sbmlharvester.googlecode.com>

Inconsistency

Compartments/species annotated with functions or processes

BIOMD0000000008 - Gardner1998_CellCycle_Goldbeter

Download SBML | Other formats (auto-generated) | Actions

Model	Overview	Math
Cell	Spatial dimensions: 3 Compartment size: 1.0 (<i>Units: volume</i>)	
cyclin	Initial amount: 0.0 (<i>Units: substance</i>)	
	<i>Compartment: Cell</i>	
protease	Initial amount: 0.0 (<i>Units: substance</i>)	
	<i>Compartment: Cell</i>	
Annotations:	set #1 bqbiol:isVersionOf	Gene Ontology peptidase activity

Inconsistency

Biological inconsistency: Biomodel 176

⊟ ATPase	[ATP] → [ADP] ;
<i>Math:</i>	cyto × Katpase_14 × ATP (Details: )
<i>Annotations:</i>	set #1 bqbiol:isVersionOf Gene Ontology ATPase activity

⊟ ATP	Initial concentration: 2.52512746499271
<i>Compartment: cytoplasm</i>	
<i>Annotations:</i>	set #1 bqbiol:is ChEBI alpha-D-glucose 6-phosphate KEGG Compound C00668

Inconsistency

Biological inconsistency: Biomodel 176

[Term]

id: GO:0016887

name: ATPase activity

is_a: GO:0017111 ! nucleoside-triphosphatase activity

intersection_of: GO:0003824 ! catalytic activity

intersection_of: has_input CHEBI:15377 ! water

intersection_of: has_input CHEBI:15422 ! ATP

intersection_of: has_output CHEBI:16761 ! ADP

intersection_of: has_output CHEBI:26020 ! phosphates

Knowledge retrieval

Query	Query string	# results
Contradictory defined entities	Nothing	4,899
Models which represent a process involving sugar	model-of some (has-part some (has-function some (realized-by only (has-participant some sugar))))	54
Parts of BIOMD0000000015 that represent processes involving sugar	part-of some BIOMD0000000015 and represents some (has-function some (realized-by only (has-participant some sugar)))	29
Model entities that represent the cell cycle	represents some (has-part some (has-function some (realized-by only 'cell cycle')))	14
Model entities that represent mutagenic central nervous system drugs in the gastrointestinal systems	represents some (has-part some ('has role' some 'central nervous system drug' and 'has role' some mutagen and part-of some 'Gastrointestinal system'))	2
Model entities that represent catalytic activity involving sugar in the endocrine pancreas	represents some (has-function some (realized-by only (realizes some 'catalytic activity' and has-participant some (sugar and contained-in some (part-of some 'Endocrine pancreas')))))	4

Summary

- SBML Harvester makes (a part of) SBML's semantics *explicit*
- enables *verification* and *integration*
- contains a minimal upper-level ontology for biosimulation models
 - can enable integration of different modelling frameworks
- implementation and ontologies freely available

Future research

- extension to physiology modelling (CellML, FieldML, ...)
- integration of other resources:
 - YeastNet
 - Reactome
 - ...
- extension to simulation and results
- SBML Planter
- use of richer languages
 - Process Specification Language: <http://www.mel.nist.gov/psl/>
 - based on FOL
 - support for multiple theories of time

Thank you!

Biomodels, YeastNet in OWL:

<http://sbmlharvester.googlecode.com>

H, R., Dumontier, M., Gennari, J. H., Wimalaratne, S., de Bono, B., Cook, D. L., and Gkoutos, G. V. (2011). Integrating systems biology models and biomedical ontologies. *BMC Systems Biology*, 5(1), 124+.