OWL-Based Validation of Biomodels
a project in Semantic Systems Biology

Michel Dumontier, Ph.D.

Associate Professor of Bioinformatics, Department of Biology, School of Computer Science, Institute of Biochemistry, Carleton University
Professeur Associé, Université Laval

Ottawa Institute of Systems Biology
Ottawa-Carleton Institute of Biomedical Engineering
Semantic Systems Biology

An interdisciplinary approach that combines computational biology with formal semantics in order to accurately build, investigate and validate quantitative models of biological phenomena.

- michel dumontier
Bio-ontologies

• Provide rich **human and machine understandable descriptions** of the common terminology
• Generally used for **semantic annotation** of data, which when reused, facilitate integration, search and retrieval
• Facilitate **granular and cross-domain queries** on their contents
• Automated reasoning over formal ontologies can be used to generate **explanations for entailments**
The intent is to express that the species represents a substance composed of glucose molecules. We also know from the SBML model that this substance is located in the cytosol and with an initial concentration of 0.09765M.
By converting models into formal representations of knowledge we get to:

- **capture** the semantics of models and the biological systems they represent
- **leverage** knowledge explicit in linked terminologies
- **validate** the accuracy of the annotations / models
- **discover** biological implications inherent in the models
- **query** the results of simulations in the context of the biological knowledge
Have you heard of OWL?
OWL - Web Ontology Language – enables more accurate knowledge capture than RDF

**Enhanced vocabulary (strong axioms)** to express knowledge relating to classes, properties, individuals and data values

- quantifiers
  - existential, universal, cardinality restriction
- negation
- disjunction
- property characteristics
  - transitive, functional, inverse functional, symmetric, antisymmetric, reflexive, irreflexive
- complex classes in domain and range restrictions
- property chains
Reasoning over OWL ontologies

- **Consistency**: determines whether the ontology contains contradictions.

- **Satisfiability**: determines whether classes can have instances.

- **Subsumption**: is C2 implicitly a subclass of C1?

- **Classification**: repetitive application of subsumption to discover implicit subclass links between named classes

- **Realization**: find the most specific class that an individual belongs to.
Triples to axioms: Many possible formalizations – knowledge of logics and domain expertise comes in handy here!

We make an *ontological commitment* by converting RDF triples into OWL axioms that have a specific meaning.

**Triple in RDF:**

\(<C1 \text{ R } C2>\)

- If \(C1\) and \(C2\) are classes, and \(R\) a relation between 2 classes
- specify meaning:
  - \(C1\ SubClassOf: C2\)
  - \(C1\ SubClassOf: R\ some\ C2\)
  - \(C1\ SubClassOf: R\ only\ C2\)
  - \(C2\ SubClassOf: R\ some\ C1\)
  - \(C1\ SubClassOf: S\ some\ C2\)
  - \(C1\ \text{DisjointFrom: } C2\)
  - \(C1\ and\ C2\ SubClassOf: owl:\text{Nothing}\)
  - \(R\ some\ C1\ \text{DisjointFrom: } R\ some\ C2\)
  - \(C1\ EquivalentClasses\ C2\)
  - ...
- in general: \(P(C1, C2)\), where \(P\) is an OWL axiom (template)
Top-level ontologies can make additional commitment by enforcing *disjointness* among basic types.

Material object, Process, Function and Quality are mutually disjoint.
Relations impose additional constraints, such that inconsistencies arise when incorrectly used.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Domain</th>
<th>Range</th>
<th>Inverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>part-of</td>
<td>Entity</td>
<td>Entity</td>
<td>has-part</td>
</tr>
<tr>
<td>participates-in</td>
<td>Individual</td>
<td>Process</td>
<td>has-participant</td>
</tr>
<tr>
<td>function-of</td>
<td>Function</td>
<td>Function</td>
<td>has-function</td>
</tr>
<tr>
<td>realizes</td>
<td>Process</td>
<td>Physical object</td>
<td>realized-by</td>
</tr>
<tr>
<td>occurs-in</td>
<td>Quality</td>
<td>Physical object</td>
<td>has-process-occurring</td>
</tr>
<tr>
<td>quality-of</td>
<td>Physical object</td>
<td>Physical object</td>
<td>has-quality</td>
</tr>
<tr>
<td>input-of</td>
<td>Physical object</td>
<td>Process</td>
<td>has-input</td>
</tr>
<tr>
<td>output-of</td>
<td>Physical object</td>
<td>Process</td>
<td>has-output</td>
</tr>
<tr>
<td>modifier-of</td>
<td>Physical object</td>
<td>Process</td>
<td>has-modifier</td>
</tr>
<tr>
<td>represents</td>
<td>Model entity</td>
<td>Physical object</td>
<td></td>
</tr>
<tr>
<td>model-of</td>
<td>Model</td>
<td>Physical object</td>
<td></td>
</tr>
</tbody>
</table>
Conceptualization:
Models and their components represent physical entities (material entities, processes)

Formalization:
every element \( E \) of the SBML language represents a class \( \text{Rep}(E) \) and we assert that

\[ E \text{ subClassOf: represents some } \text{Rep}(E) \]
For each model annotation, we make a commitment to what it represents.

OWL Axiom:
M SubClassOf: represents some MaterialEntity

Conversion rule: a Model annotated with class C represents:

If C is a SubClassOf **MaterialEntity** then
M SubClassOf: represents some C

If C is a SubClassOf **Function** then
M SubClassOf: represents some (has-function some C)

If C is a SubClassOf **Process** then
M SubClassOf: represents some (has-function some (realized-by only C))
BIOMODEL 82: Converting Model

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

represents an object O that has the capability F that is realized in processes of the type **heterotrimeric G-protein complex cycle**

M SubClassOf: represents some O1
O1 SubClassOf: (has-function some (realized-by only GO:0031684))
Models and their components represent physical entities (material entities, processes)

Class
Individual
Datatype

---

SBML2OWL: Implementation

- libSBML to access model structure & extract RDF annotations
- Jena RDF API to parse RDF annotations
- OWLAPI to create OWL axioms & reason with top-level ontology

Application to BioModels repository yields:
- OWL ontology with more than 300,000 classes, 800,000 axioms
- includes all referenced ontologies
  - GO (functions, compartments, processes)
  - ChEBI (molecules)
  - Celltype (cell types)
  - FMA (anatomy)
  - PATO (qualities)
### Answering questions

<table>
<thead>
<tr>
<th>Query</th>
<th>Query string</th>
<th># results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradictory defined entities</td>
<td>Nothing</td>
<td>4,899</td>
</tr>
<tr>
<td>Models which represent a process involving sugar</td>
<td>model-of some (has-part some (has-function some (realized-by only (has-participant some sugar))))</td>
<td>54</td>
</tr>
<tr>
<td>Parts of BIOM000000000015 that represent processes involving sugar</td>
<td>part-of some BIOM000000000015 and represents some (has-function some (realized-by only (has-participant some sugar)))</td>
<td>29</td>
</tr>
<tr>
<td>Model entities that represent the cell cycle</td>
<td>represents some (has-part some (has-function some (realized-by only 'cell cycle')))</td>
<td>14</td>
</tr>
<tr>
<td>Model entities that represent mutagenic central nervous system drugs in the gastrointestinal systems</td>
<td>represents some (has-part some ('has role' some 'central nervous system drug' and 'has role' some mutagen and part-of some 'Gastrointestinal system'))</td>
<td>2</td>
</tr>
<tr>
<td>Model entities that represent catalytic activity involving sugar in the endocrine pancreas</td>
<td>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')))))</td>
<td>4</td>
</tr>
</tbody>
</table>
Model verification

After reasoning, we found 27 models to be inconsistent

reasons
1. **our representation** - functions sometimes found in the place of physical entities (e.g. entities that secrete insulin). better to constrain with appropriate relations
2. **SBML abused** – e.g. species used as a measure of time
3. **Incorrect annotations** - constraints in the ontologies themselves mean that the annotation is simply not possible
Finding inconsistencies with axiomatically enhanced ontologies

related work treated function as a process and described ATPase activity (GO:0004002) as a Catalytic activity that has Water and ATP as input, ADP and phosphate as output and is a part of an ATP catabolic process.

To this, we add:

- GO: ATP + Water the only inputs (universal quantification)
- ChEBI: Water, ATP, alpha-D-glucose 6-phosphate are all different (disjointness)

BIOMD0000000176 and BIOMD0000000177 models of anaerobic glycolysis in yeast.

- “ATP” input to “ATPase” reaction, which is annotated with ATPase activity. The species “ATP”, however, is mis-annotated with Alpha-D-glucose 6-phosphate (CHEBI:17665), not with ATP.
Species are further described with ‘modifiers’ in the context of a reaction

**essential activator**

```xml
<listOfModifiers>
  <modifierSpeciesReference sboTerm="SBO:0000461" species="X"/>
</listOfModifiers>
```

**partial inhibitor**

```xml
<listOfModifiers>
  <modifierSpeciesReference sboTerm="SBO:0000536" species="PX"/>
</listOfModifiers>
```
Roles are realized in the context of processes.
Semanticscience Integrated Ontology (SIO)

- OWL2 ontology
- 1000+ classes covering basic types (physical, processual, abstract, informational) with an emphasis on biological entities
- 184 basic relations (mereological, participatory, attribute/quality, spatial, temporal and representational)
- Axioms can be used by reasoners to compute inferences for consistency checking, classification and answering questions about life science knowledge
- Embodies emerging ontology design patterns
  - Specifies a data model
- Dereferenceable URIs – used in Bio2RDF and SADI
- Searchable in the NCBO bioportal
- Available at [http://semanticscience.org/ontology/sio.owl](http://semanticscience.org/ontology/sio.owl)
When running a simulation, certain quantities can vary with time.
Copasi output:

not machine understandable

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H mRN</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LacI</td>
<td>prote</td>
<td>TetR</td>
<td>mRNA</td>
<td>cl</td>
<td>mRNA</td>
<td>Compartn Values</td>
<td>be</td>
<td>Values</td>
<td>al</td>
<td>Values</td>
<td>al</td>
<td>Values</td>
<td>trc</td>
<td>Values</td>
<td>n</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>81.4405</td>
<td>188.382</td>
<td>42.6413</td>
<td>19.9034</td>
<td>30.6156</td>
<td>7.491</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>2</td>
<td>218.539</td>
<td>358.027</td>
<td>84.5888</td>
<td>21.2335</td>
<td>23.6089</td>
<td>5.86627</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>3</td>
<td>337.623</td>
<td>469.627</td>
<td>113.055</td>
<td>18.5441</td>
<td>17.2349</td>
<td>4.40364</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>4</td>
<td>428.939</td>
<td>536.664</td>
<td>131.035</td>
<td>15.5998</td>
<td>12.4821</td>
<td>3.29582</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>5</td>
<td>495.497</td>
<td>572.019</td>
<td>141.441</td>
<td>13.018</td>
<td>9.03773</td>
<td>2.48567</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>6</td>
<td>542.466</td>
<td>585.398</td>
<td>146.532</td>
<td>11.0337</td>
<td>6.56428</td>
<td>1.90266</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>7</td>
<td>574.778</td>
<td>583.832</td>
<td>147.977</td>
<td>9.56151</td>
<td>4.79541</td>
<td>1.48837</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>8</td>
<td>596.57</td>
<td>572.335</td>
<td>146.993</td>
<td>8.51888</td>
<td>3.53331</td>
<td>1.19835</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>9</td>
<td>611.14</td>
<td>554.454</td>
<td>144.458</td>
<td>7.82324</td>
<td>2.63412</td>
<td>0.999828</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>10</td>
<td>621.044</td>
<td>532.674</td>
<td>141.004</td>
<td>7.40287</td>
<td>1.99408</td>
<td>0.859004</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>11</td>
<td>628.231</td>
<td>508.728</td>
<td>137.085</td>
<td>7.19849</td>
<td>1.5387</td>
<td>0.78821</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>12</td>
<td>634.16</td>
<td>483.805</td>
<td>133.028</td>
<td>7.16163</td>
<td>1.21462</td>
<td>0.747164</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>13</td>
<td>639.897</td>
<td>458.714</td>
<td>129.068</td>
<td>7.25233</td>
<td>0.983648</td>
<td>0.735535</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>14</td>
<td>646.193</td>
<td>433.993</td>
<td>125.381</td>
<td>7.43676</td>
<td>0.818491</td>
<td>0.748075</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>15</td>
<td>653.535</td>
<td>409.987</td>
<td>122.095</td>
<td>7.68527</td>
<td>0.699579</td>
<td>0.780855</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>16</td>
<td>662.19</td>
<td>386.909</td>
<td>119.309</td>
<td>7.97095</td>
<td>0.61336</td>
<td>0.83136</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>17</td>
<td>672.232</td>
<td>364.881</td>
<td>117.104</td>
<td>8.26865</td>
<td>0.549728</td>
<td>0.89813</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>18</td>
<td>683.569</td>
<td>343.958</td>
<td>115.547</td>
<td>8.55459</td>
<td>0.501895</td>
<td>0.980494</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
<tr>
<td>19</td>
<td>695.958</td>
<td>324.155</td>
<td>114.698</td>
<td>8.80638</td>
<td>0.465075</td>
<td>1.07639</td>
<td>1</td>
<td>0.2</td>
<td>0.216404</td>
<td>216.188</td>
<td>20</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>10</td>
<td>2.88539</td>
</tr>
</tbody>
</table>
Query Answering over RDF/OWL

Find those concentration measurements for species that represent molecular entities that contain ribonucleotide residues

‘concentration’
and (‘measured at’ some double[>20.0, <40.0])
and ‘is attribute of’ some ( ‘species’
and ‘represents’ some (‘has part’ some ‘ribonucleotide residue’) )
Curve Analysis: Elements of a Plot

Curves Contain
- Points
- Line Segments
- Curve Segments

Curves Annotated As
- Monotonic/Non-Monotonic
- Overall Increasing/Decreasing

Curve Segments
- Contain line segments
- Same properties as curve

Line Segments
- Contain points
- Increasing/Decreasing

Points
- Have attributes and values
- Can be local minima/maxima
- Can be inflection points
Get non-monotonic behaviour for entities that bind to DNA

‘non-monotonic curve’
and ‘has part’ some (‘concentration’
and ‘is attribute of’ some (‘species’
and ‘has function’ some ‘dna binding’) )
Conclusion

The SBML-derived ontologies can be

i) checked for their consistency, thereby uncovering erroneous curation

ii) infer attributes and relations of the substances, compartments and reactions beyond what was originally described in the models

iii) answer sophisticated questions across a model knowledge base

iv) extended with modifiers, mathematical expressions and parameters, simulation Results (from tab files) to answer questions about simulation results with reference to the semantic annotations (GO) in biomodels, UniProt
Acknowledgements

Leonid Chepelev

Thank You

Michel Dumontier
michel_dumontier@carleton.ca

Publications: http://dumontierlab.com
Presentations: http://slideshare.com/micheldumontier