

Semantic annotations and querying SBML simulations



Michel Dumontier, Ph.D.

**Associate Professor of Bioinformatics, Department of Biology, School of Computer
Science, Institute of Biochemistry, Carleton University**

**Professeur Associé, Département d'informatique et de génie
logiciel, Université Laval**

Ottawa Institute of Systems Biology

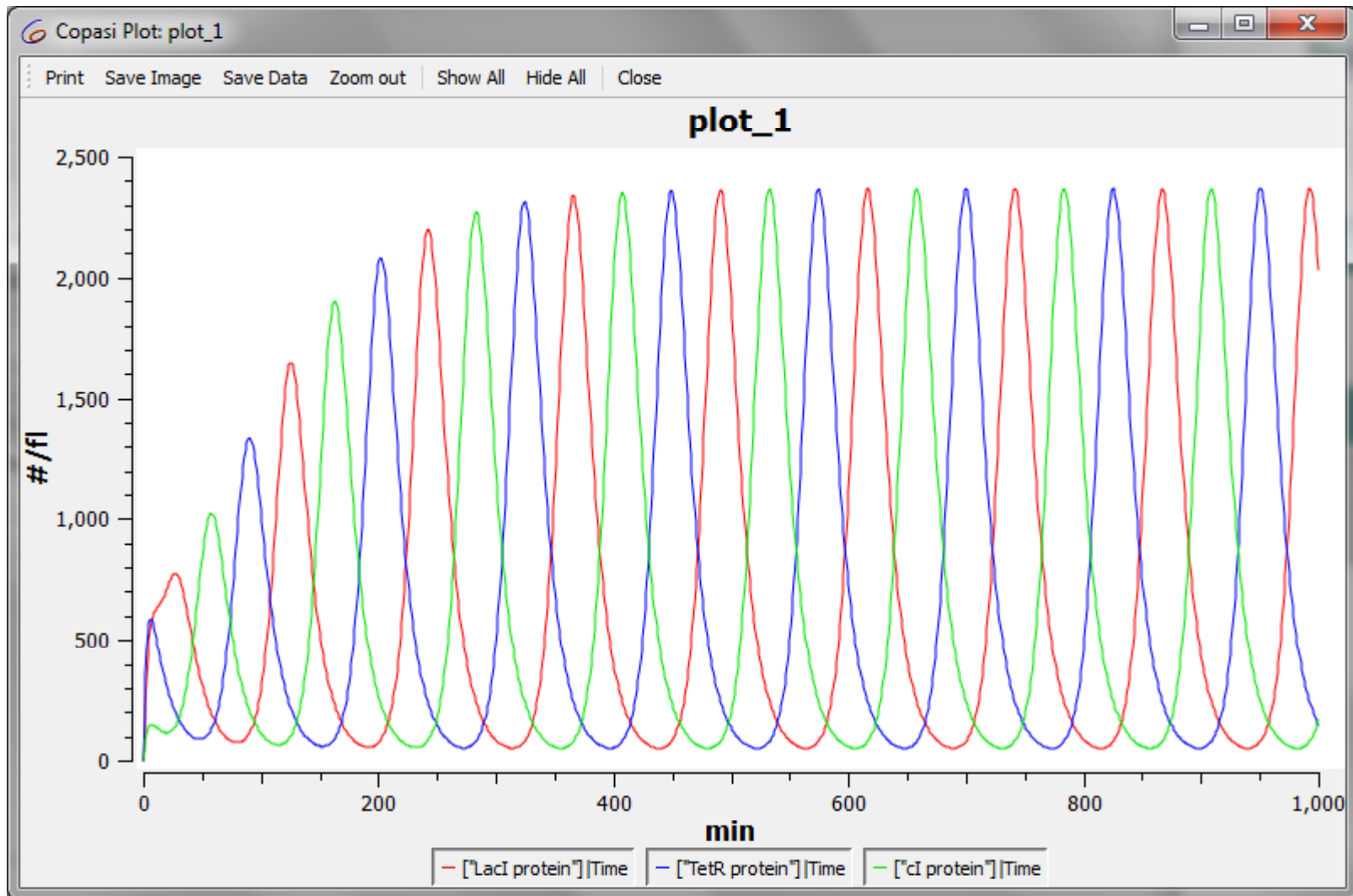
Ottawa-Carleton Institute of Biomedical Engineering

Objective

To answer questions about SBML models and the results of simulations

Approach

- We formally represent semantically annotated biomodels using the Web Ontology Language (OWL), such that it becomes possible to:
 - reason about the knowledge and check it's consistency
 - simultaneously query the results of simulations with what is known about the model components (species, compartments, reactions, mathematical expressions, parameters, etc).

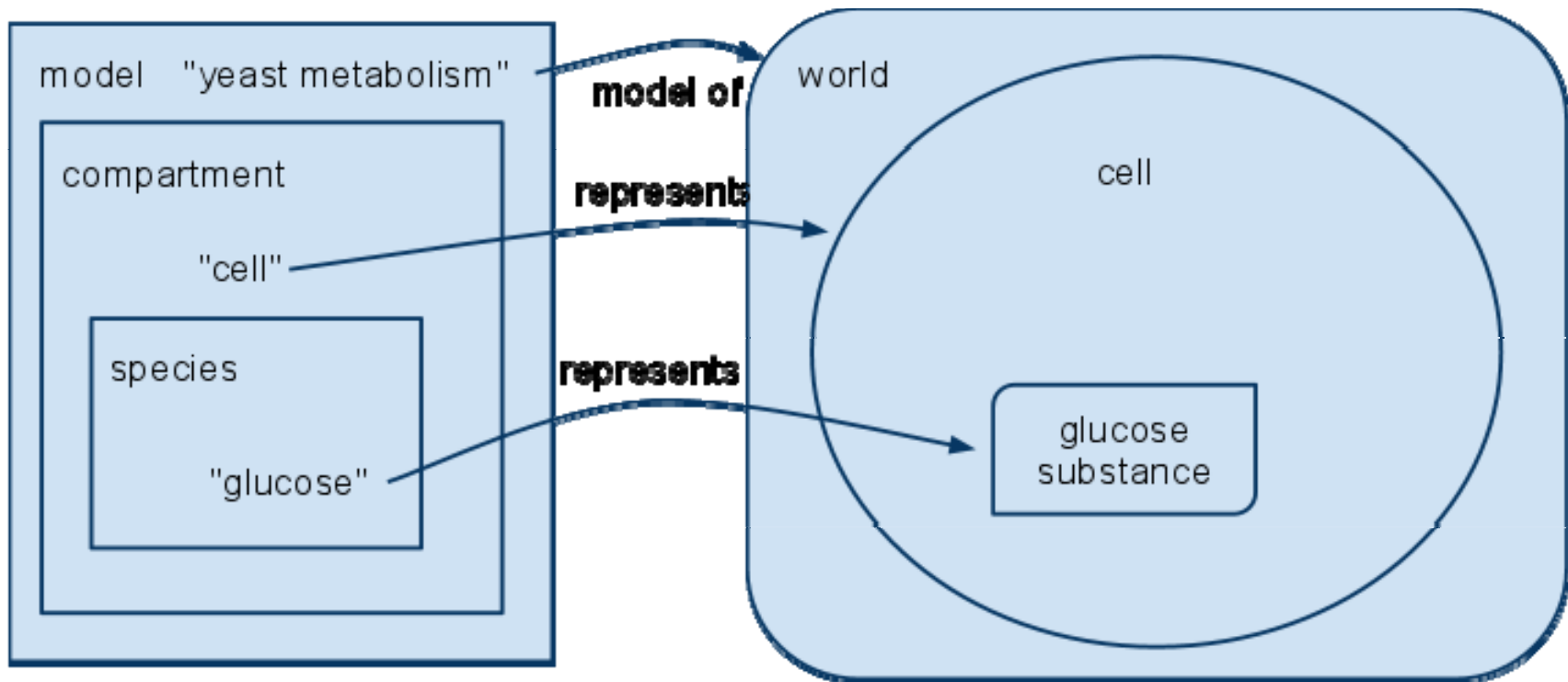


Annotated SBML Biomodel

```
<sbml xmlns="http://www.sbml.org/sbml/level2/version3" metaid="_153818" level="2" version="3">
  <model metaid="_000001" id="repressilator" name="Elowitz2000_Repressilator">
    <annotation>
      <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms=
"http://purl.org/dc/terms/" xmlns:vCard="http://www.w3.org/2001/vcard-rdf/3.0#" xmlns:bqbiol=
"http://biomodels.net/biology-qualifiers/" xmlns:bqmodel="http://biomodels.net/model-qualifiers/">
        <rdf:Description rdf:about="#_000001">
          <dc:creator rdf:parseType="Resource">
            <dcterms:created rdf:parseType="Resource">
              <dcterms:modified rdf:parseType="Resource">
                <bqmodel:is>
                  <rdf:Bag>
                    <rdf:li rdf:resource="urn:miriam:biomodels.db:BIOMD0000000012"/>
                  </rdf:Bag>
                </bqmodel:is>
                <bqmodel:is>
                  <bqmodel:isDescribedBy>
                    <bqbiol:isVersionOf>
                      <rdf:Bag>
                        <rdf:li rdf:resource="urn:miriam:obo.go:GO%3A0040029"/>
                      </rdf:Bag>
                    </bqbiol:isVersionOf>
                    <bqbiol:is>
                      <rdf:Bag>
                        <rdf:li rdf:resource="urn:miriam:taxonomy:562"/>
                      </rdf:Bag>
                    </bqbiol:is>
                  </rdf:Description>
                </rdf:RDF>
              </annotation>
            <listOfUnitDefinitions>
              <listOfCompartments>
                <listOfSpecies>
                  <listOfParameters>
                    <listOfRules>
                      <listOfReactions>
                    </model>
                  </sbml>
```

Conceptualization:

Model entities (models and model components) are distinguished from the entities they represent

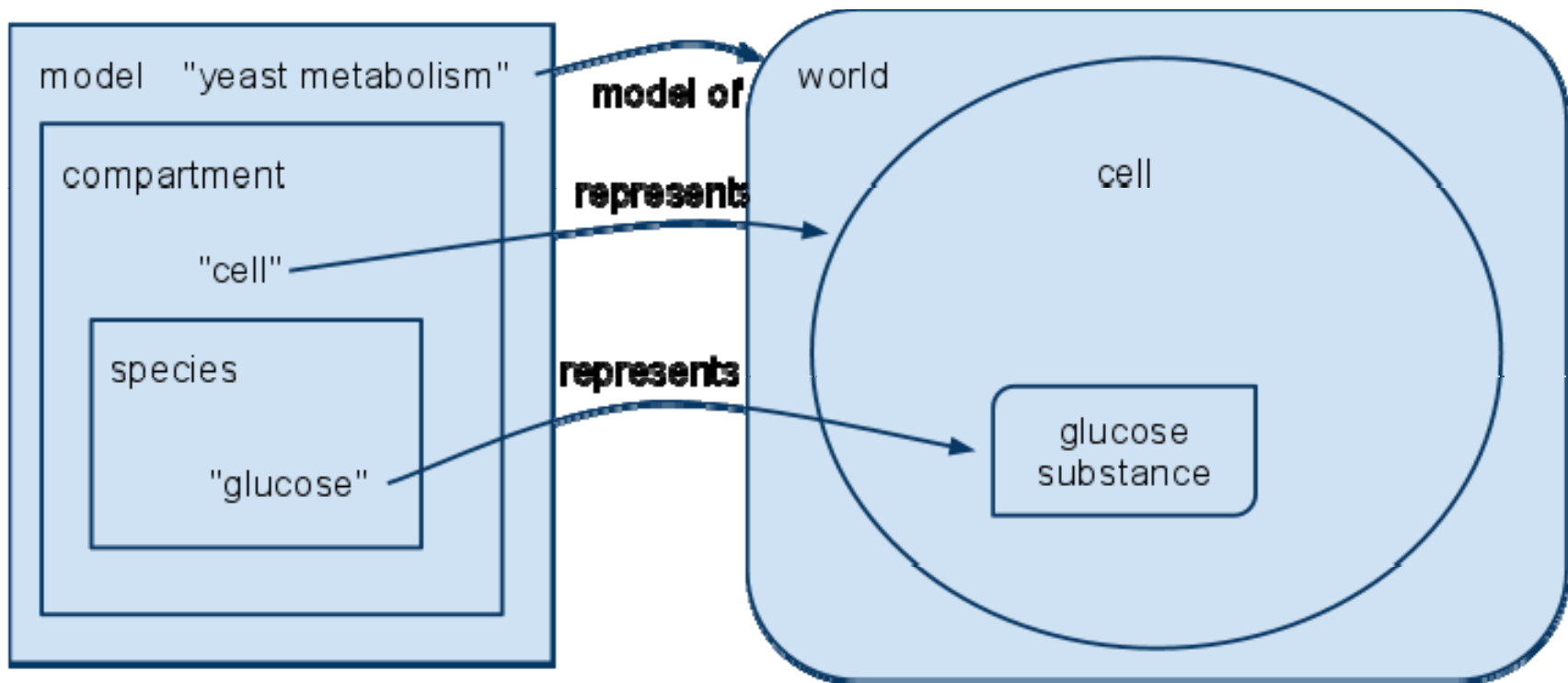


The Web Ontology Language (OWL) Has Explicit Semantics

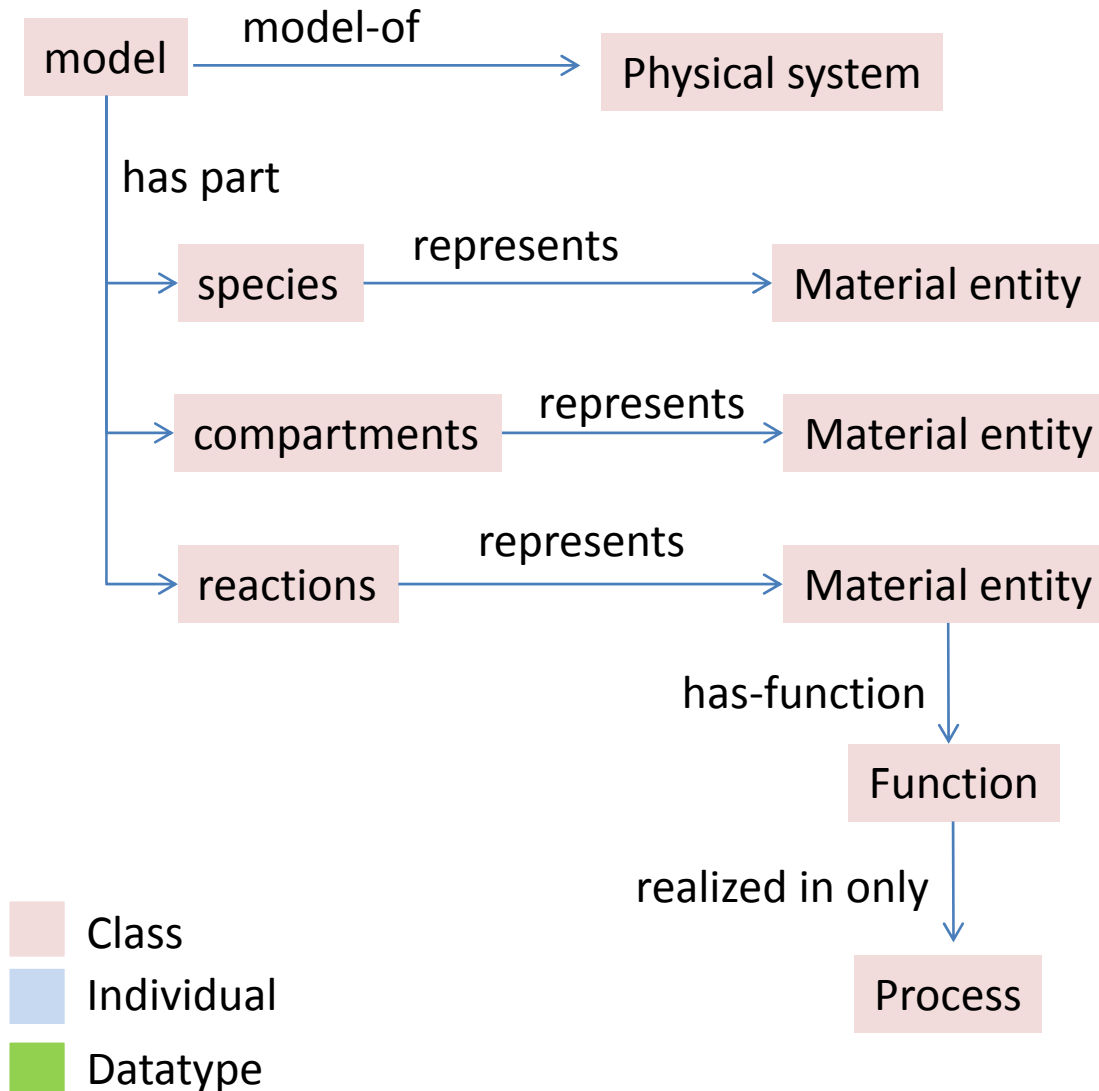


**Can therefore be used to capture knowledge in a
machine understandable way**

Formalization: every element E of the SBML language represents a class Rep(E) and we assert that E subClassOf: represents some Rep(E)



Models and their components represent physical entities (material entities, processes)



SBMLHarvester

Robert Hoehndorf, Michel Dumontier, John H Gennari, Sarala Wimalaratne, Bernard de Bono, Daniel L Cook and Georgios V Gkoutos. [Integrating systems biology models and biomedical ontologies](#). BMC Systems Biology 2011, 5:124.

Species are further described with 'modifiers' *in the context of a reaction*

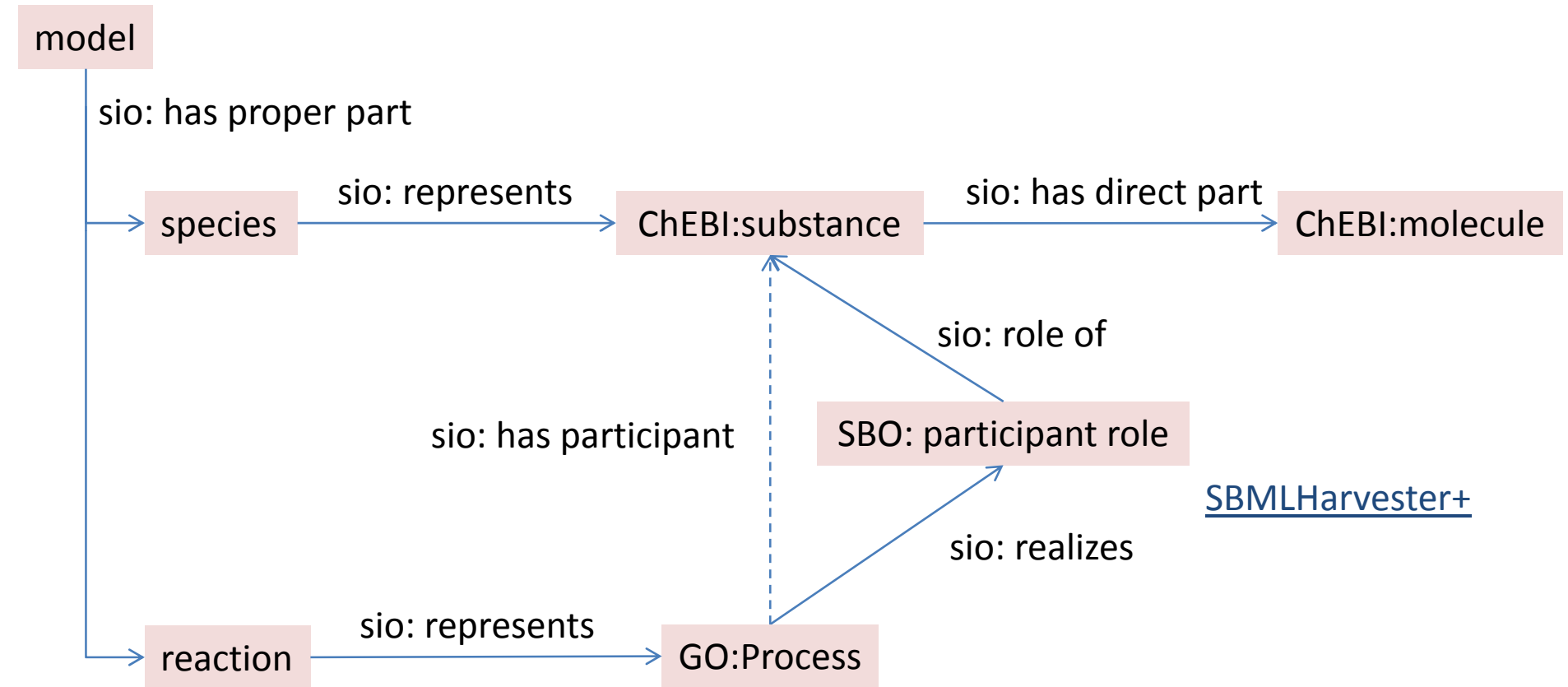
essential activator

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

partial inhibitor

```
<listOfModifiers>  
  <modifierSpeciesReference sboTerm="SBO:0000536" species="PX"/>  
</listOfModifiers>
```

Roles are realized in the context of processes by material entities



SBMLHarvester+

- Class
- Individual
- Datatype

role chain: realizes o role of -> has participant

Semanticscience Integrated Ontology (SIO)

- OWL2 ontology
- 100+ classes covering basic types (physical, processual, abstract, informational) with an emphasis on biological entities
- 183 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
- searchable in the NCBO bioportal
- Available at <http://semanticscience.org/ontology/sio.owl>

Examining Mathematical Expressions

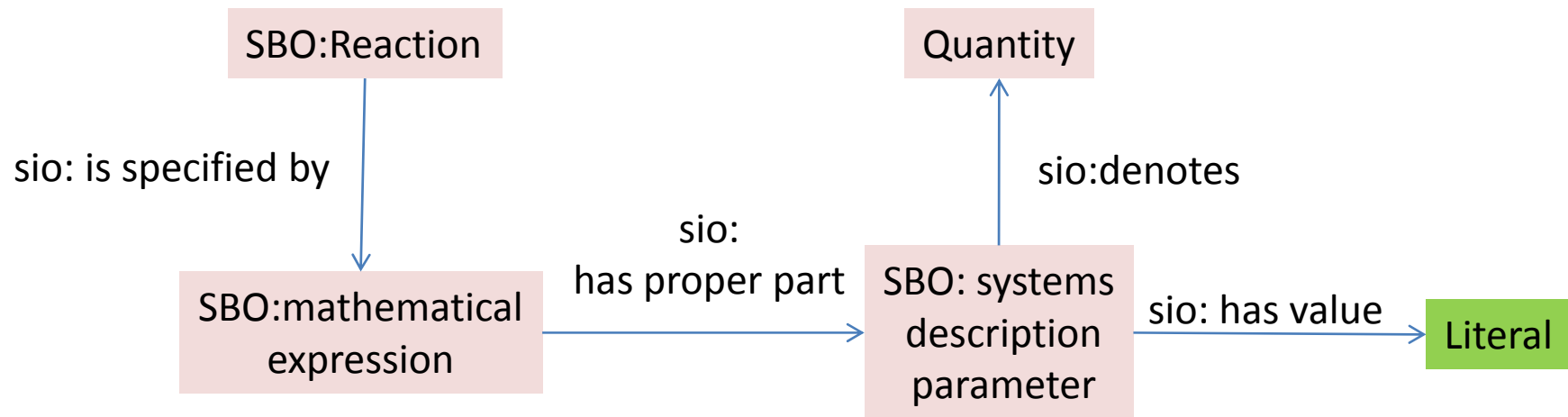
```
<kineticLaw sboTerm="SBO:0000049">
  <math>
    <apply>
      <times/>
        <ci> k_tl </ci>
        <ci> X </ci>
      </apply>
    </math>
  </kineticLaw>
```

```
<assignmentRule metaid="metaid_0400235"
  variable="k_tl">
  <math>
    <apply>
      <divide/>
        <ci> eff </ci>
        <ci> t_ave </ci>
      </apply>
    </math>
  </assignmentRule>
```

```
<parameter metaid="metaid_0000233" id="k_tl" name="k_tl" constant="false"
  sboTerm="SBO:0000016">  (unimolecular rate constant)
  <notes> <p xmlns="http://www.w3.org/1999/xhtml">Translation rate constant</p> </notes>
</parameter>
```

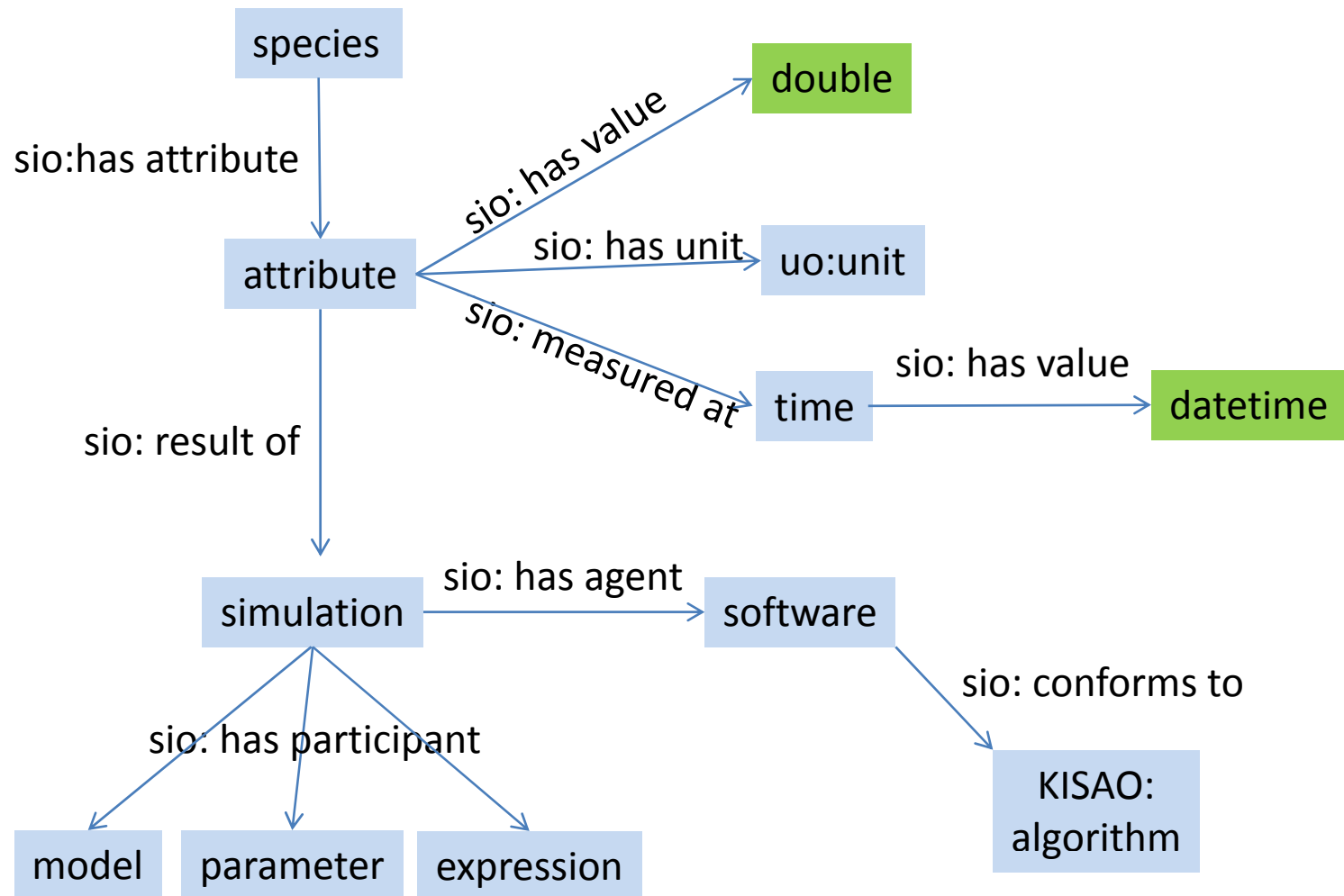
```
<parameter metaid="metaid_0000025" id="eff" name="translation efficiency" value="20">
  <notes> <p xmlns="http://www.w3.org/1999/xhtml">Average number of proteins per
transcript</p> </notes>
</parameter>
```

SBML Reactions may be *specified by* mathematical expressions, which contain quantitative variables that *denote* quantities

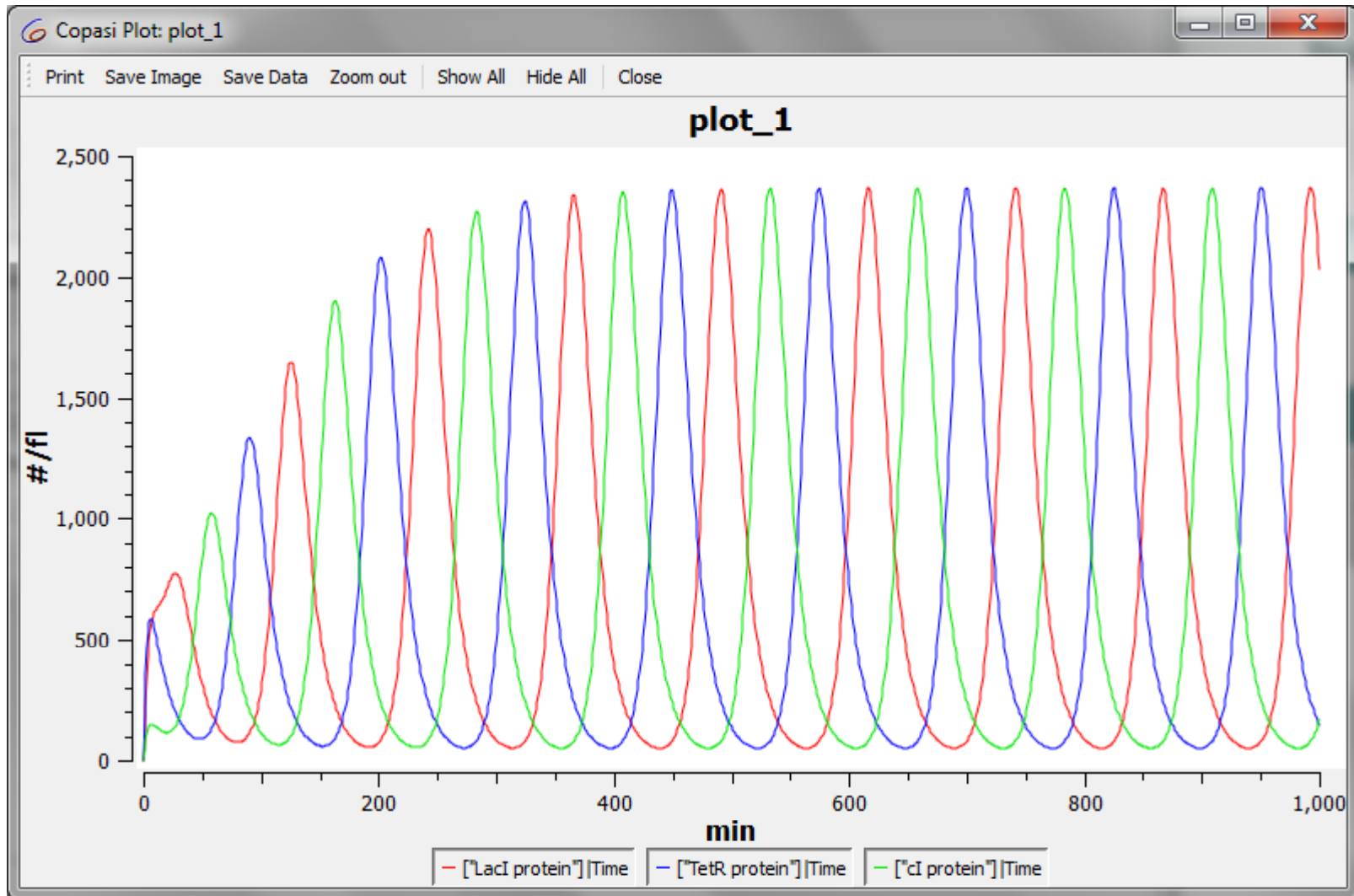


- Class
- Individual
- Datatype

When running a simulation, some attributes change with time

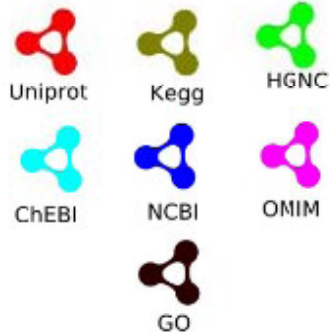
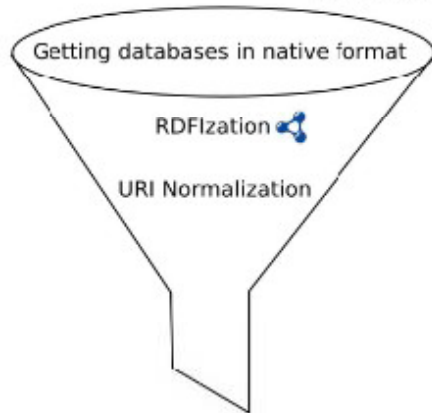


- Class
- Individual
- Datatype

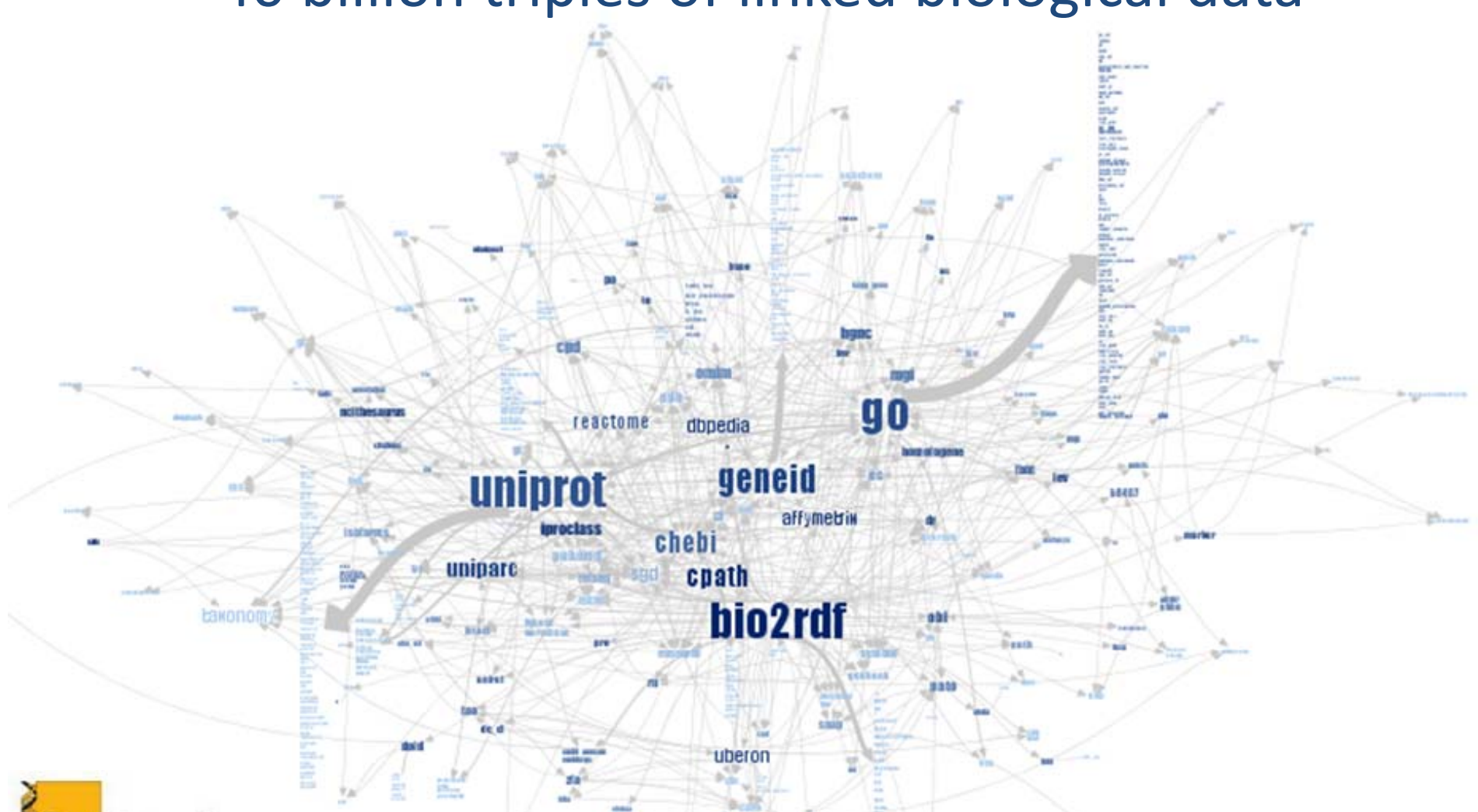


Copasi output: not machine understandable

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
# Time	LacI prote	TetR prote	cl protein	LacI mRNA	TetR mRN	cl mRNA	Compartm	Values[be	Values[al	Values[al	Values[tra	Values[n]	Values[KN	Values[m]	Values[pr	Values[av
0	0	0	0	0	20	0	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
1	81.4405	188.382	42.6413	19.9034	30.6156	7.491	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
2	218.539	358.027	84.5888	21.2335	23.6089	5.86627	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
3	337.623	469.627	113.055	18.5441	17.2349	4.40364	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
4	428.939	536.664	131.035	15.5598	12.4821	3.29582	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
5	495.497	572.019	141.441	13.018	9.03773	2.48567	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
6	542.466	585.398	146.532	11.0337	6.56428	1.90266	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
7	574.778	583.832	147.977	9.56151	4.79541	1.48837	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
8	596.57	572.335	146.993	8.51888	3.53331	1.19835	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
9	611.14	554.454	144.458	7.82324	2.63412	0.999828	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
10	621.044	532.674	141.004	7.40287	1.99408	0.869004	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
11	628.231	508.728	137.085	7.19849	1.5387	0.788821	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
12	634.16	483.805	133.028	7.16163	1.21462	0.747164	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
13	639.897	458.714	129.068	7.25233	0.983648	0.735535	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
14	646.193	433.993	125.381	7.43676	0.818491	0.748075	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
15	653.535	409.987	122.095	7.68527	0.699679	0.780855	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
16	662.19	386.909	119.309	7.97095	0.61336	0.83136	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
17	672.232	364.881	117.104	8.26865	0.549728	0.89813	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
18	683.569	343.958	115.547	8.55459	0.501895	0.980494	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
19	695.958	324.155	114.698	8.80638	0.465075	1.07839	1	0.2	0.216404	216.188	20	2	40	2	10	2.88539
..



Bio2RDF now serving over 40 billion triples of linked biological data



**Centre de
Recherche**
du CHUL (CHUQ)



QUT Queensland University of Technology
Microsoft QUT eResearch Centre



POWERED BY
VIRTUOSO

LINKING OPENDATA
RISC W3C Community Project

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'

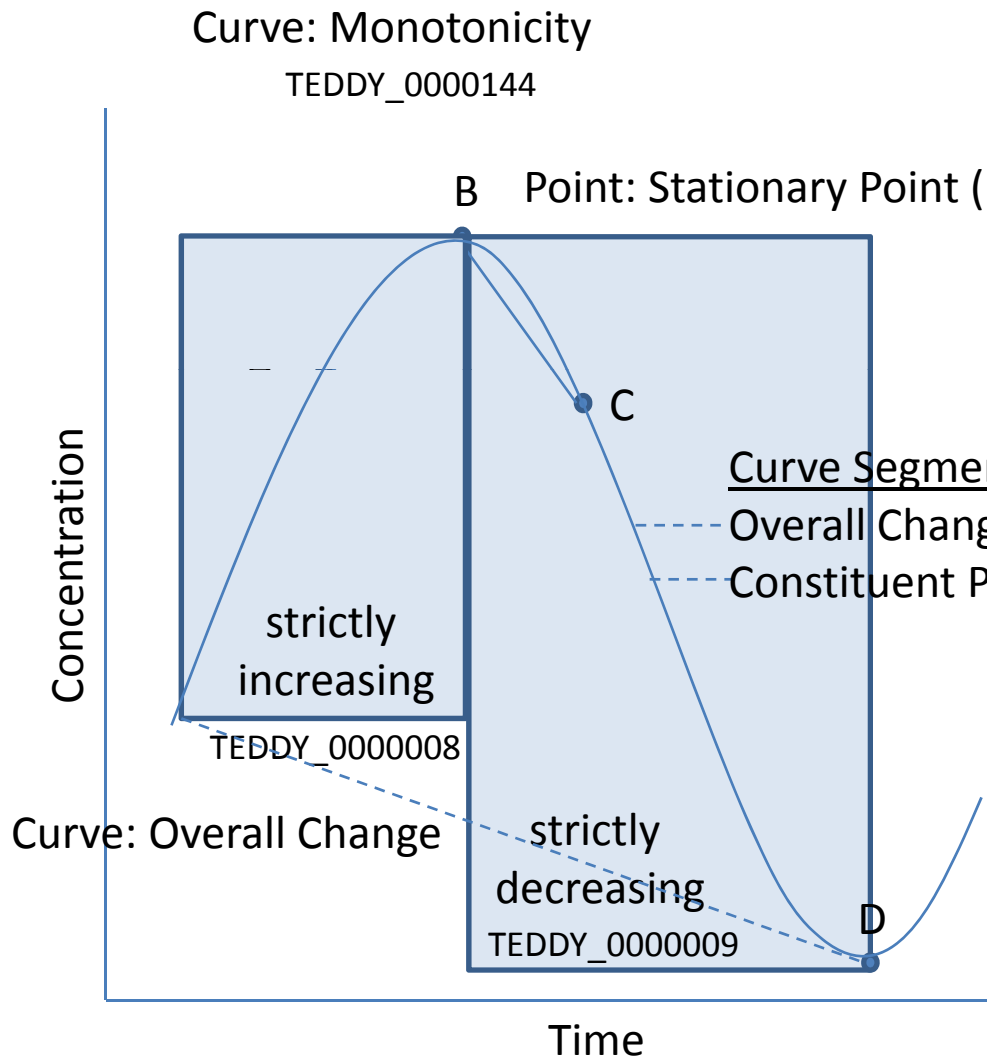
and 'represents' some ('has part' some 'ribonucleotide residue')

)

ChEBI ontology



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

Queries

'local maximum'
and 'is attribute of' some (
species
and represents some (
 'has function' some 'dna binding'
))

Biomodel +
UniProt + GO



Get the non-monotonic curves for protein species

'non-monotonic curve'
and 'has part' some (
 'concentration'
and 'is attribute of' some (
 'species'
 and 'represents' some 'protein'))
)

Conclusion

- We extended our OWL ontology based representation to include
 - Modifiers
 - Mathematical Expressions and Parameters
 - Simulation Results (from tab files)
- We could answer questions about simulation results with reference to the semantic annotations (GO) in biomodels, UniProt
- Further develop and release software.

Acknowledgements

Leonid Chepelev

Robert Hoehndorf





Michel Dumontier

michel_dumontier@carleton.ca

Publications: <http://dumontierlab.com>

Presentations: <http://slideshare.com/micheldumontier>