

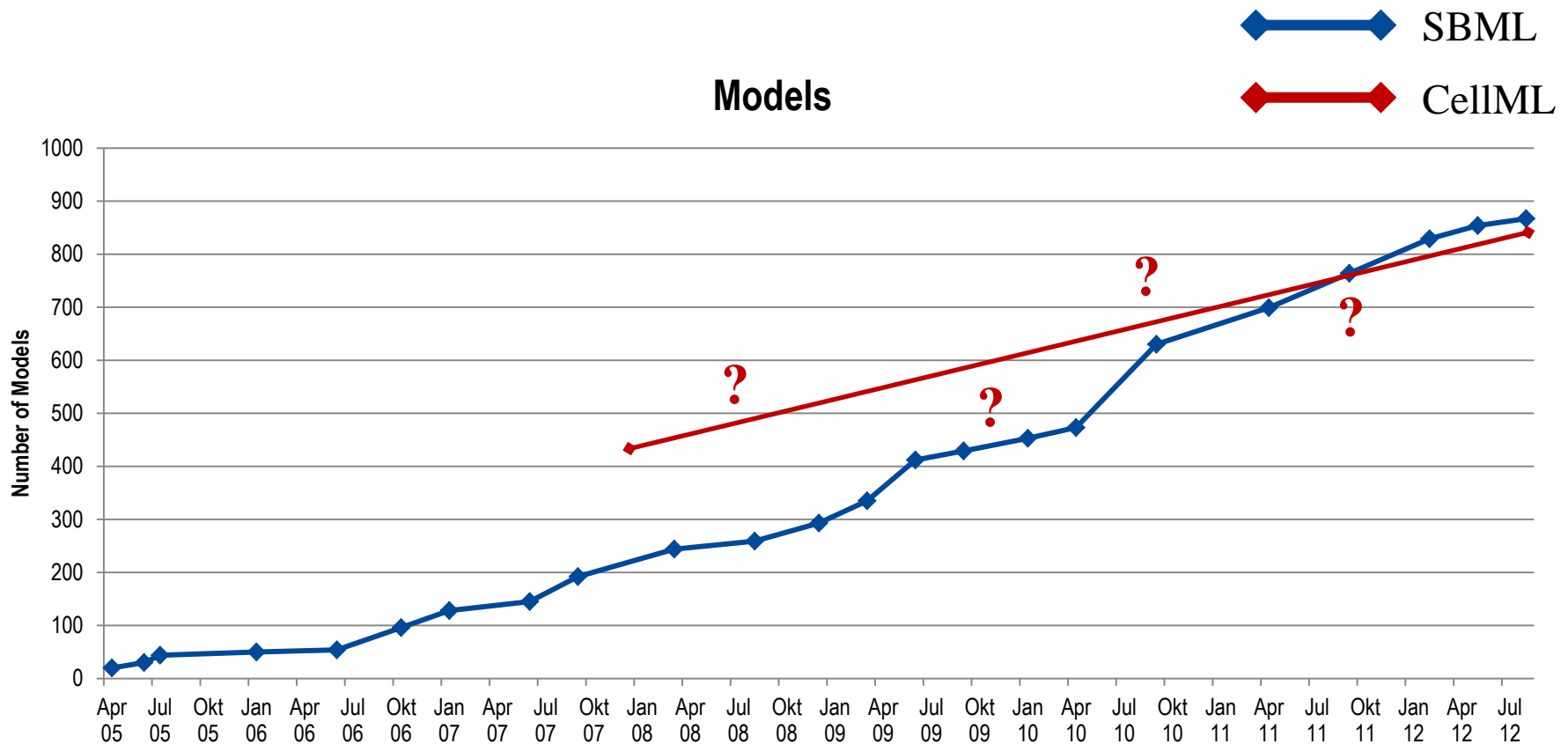
Strategies for integrated management of computational biology models and associated simulations

Ron Henkel

*Department of Systems Biology and Bioinformatics
The University of Rostock*

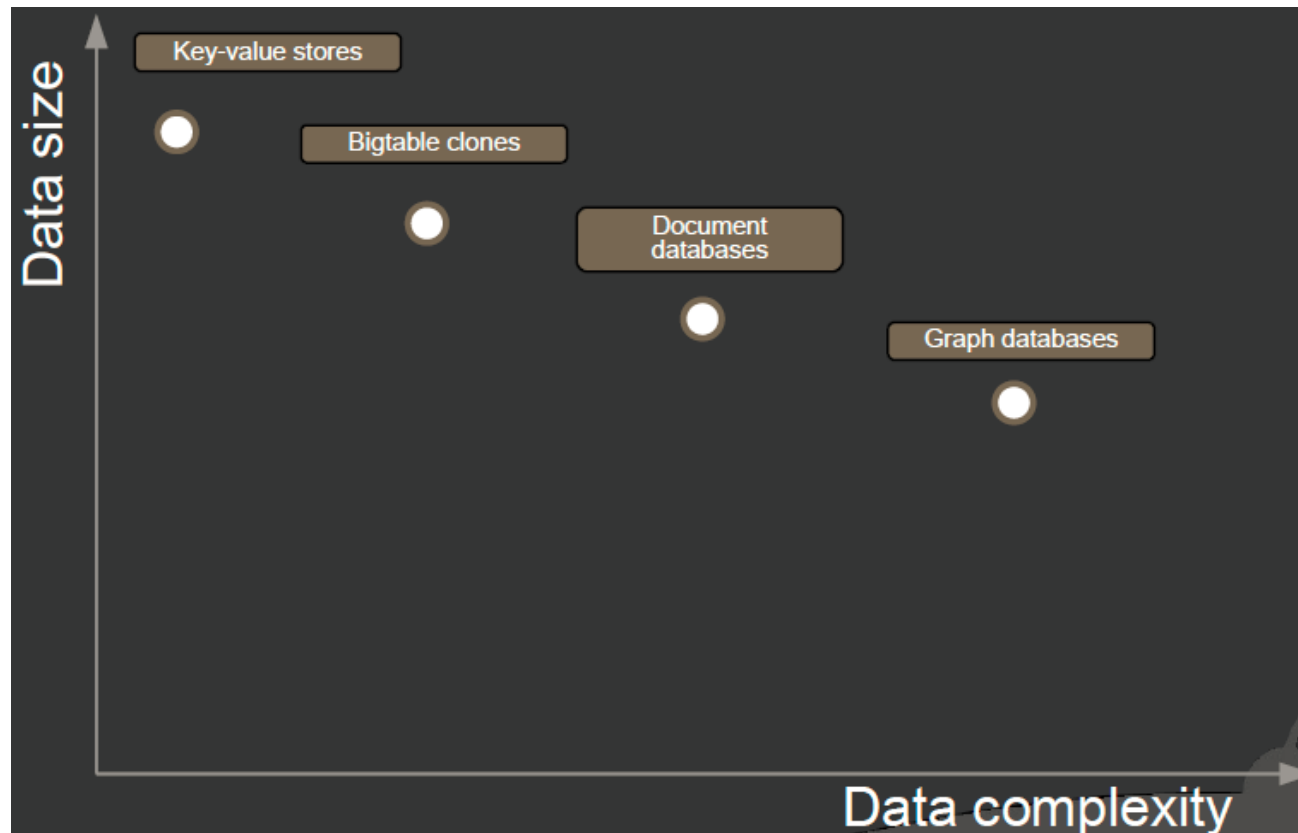


Motivation



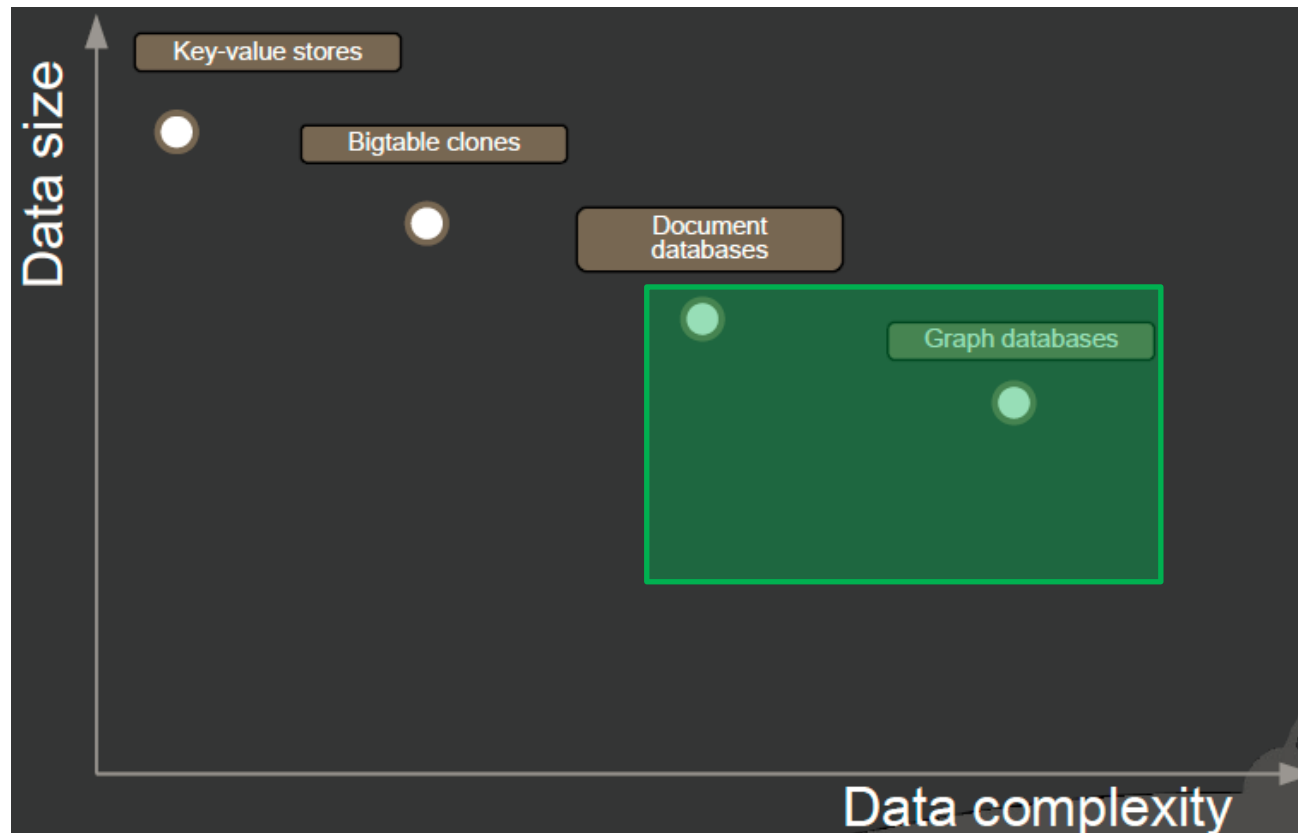
SBML Data from BioModels Database, CellML data from Tommy Yu

NoSQL Data Models



Taken from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.174.8830&rep=rep1&type=pdf>

NoSQL Data Models



Taken from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.174.8830&rep=rep1&type=pdf>

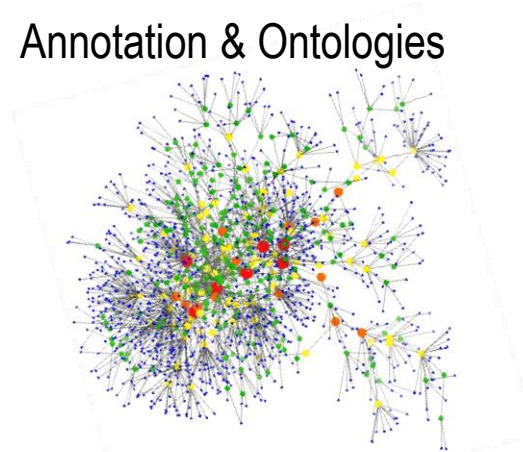
Available Data

Model and Simulation



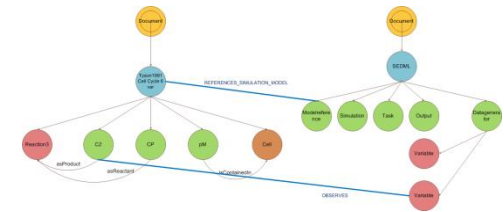
- Constituent names
- XML encoded representation

Annotation & Ontologies



- Biochemical background
- Synonyms

A model's network



- Model structure
- Aggregation values
- Model related data

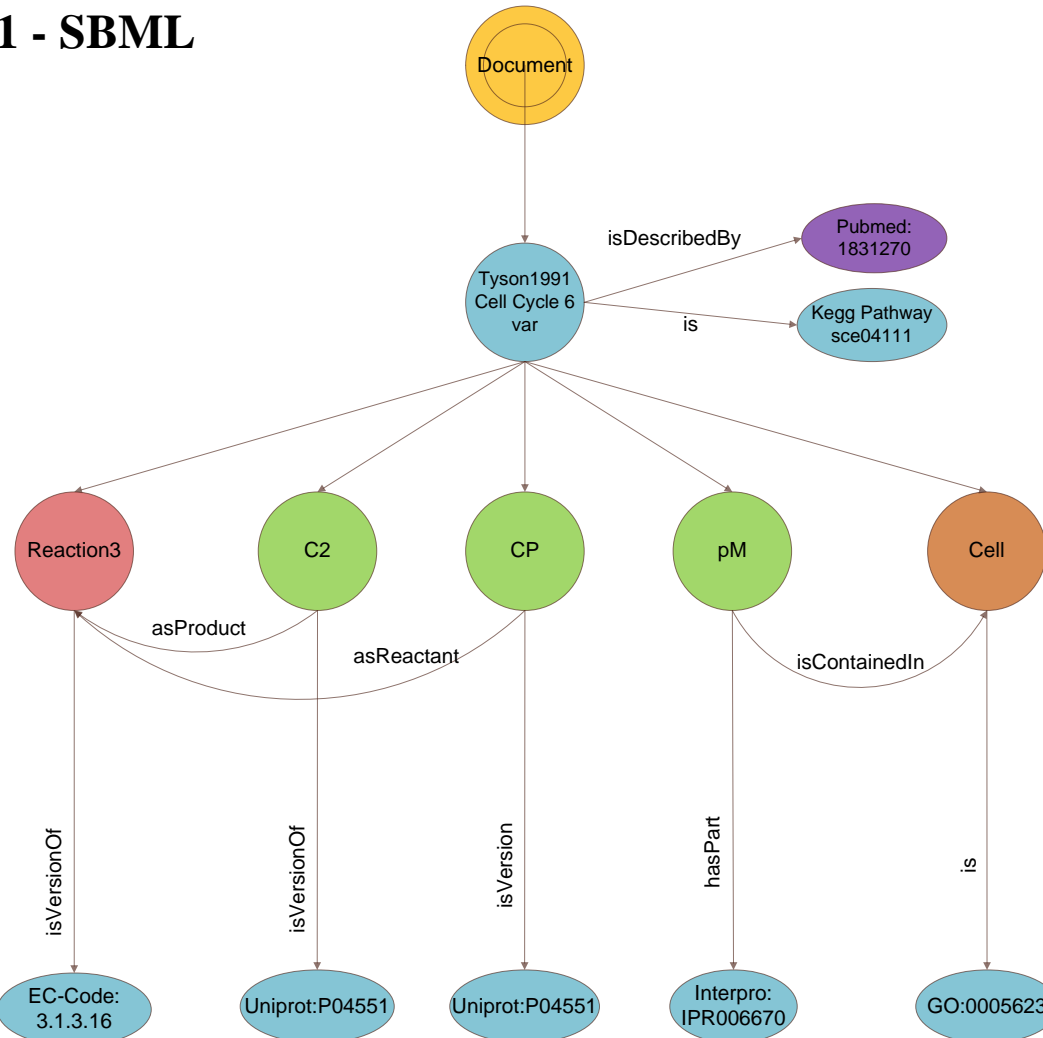
Tyson 1991 - SBML



Model:

Entities:

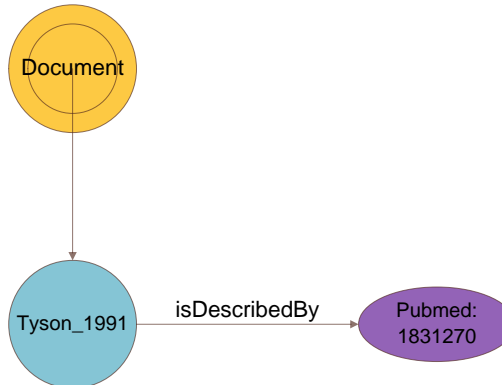
Annotation:



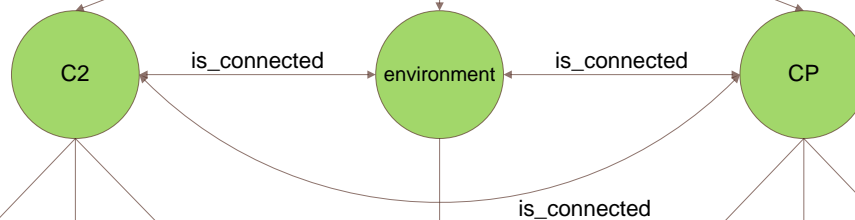
Tyson 1991 – CellML



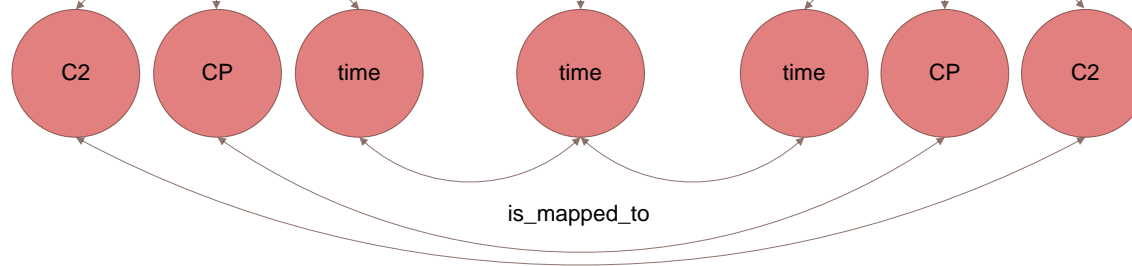
Model:



Component:



Variable:



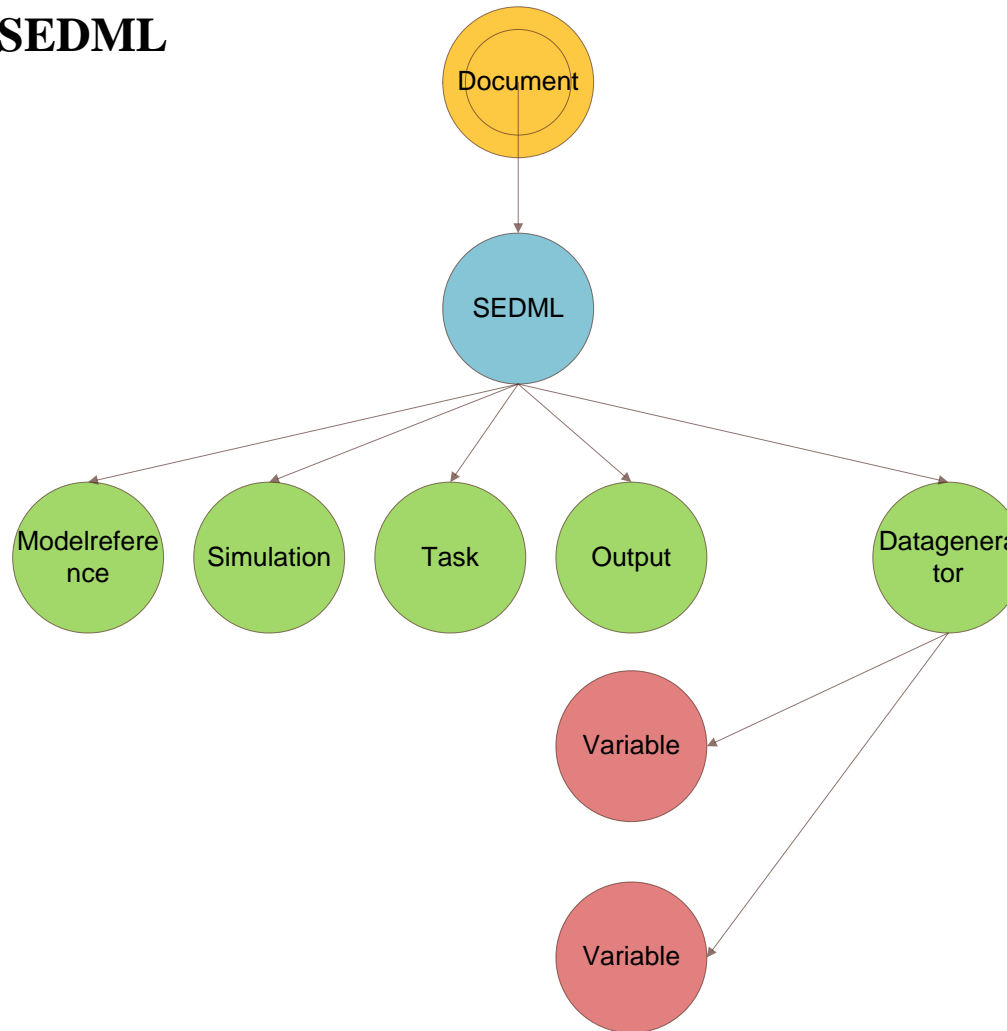
Tyson 1991 – SEDML



Simulation:

Entities:

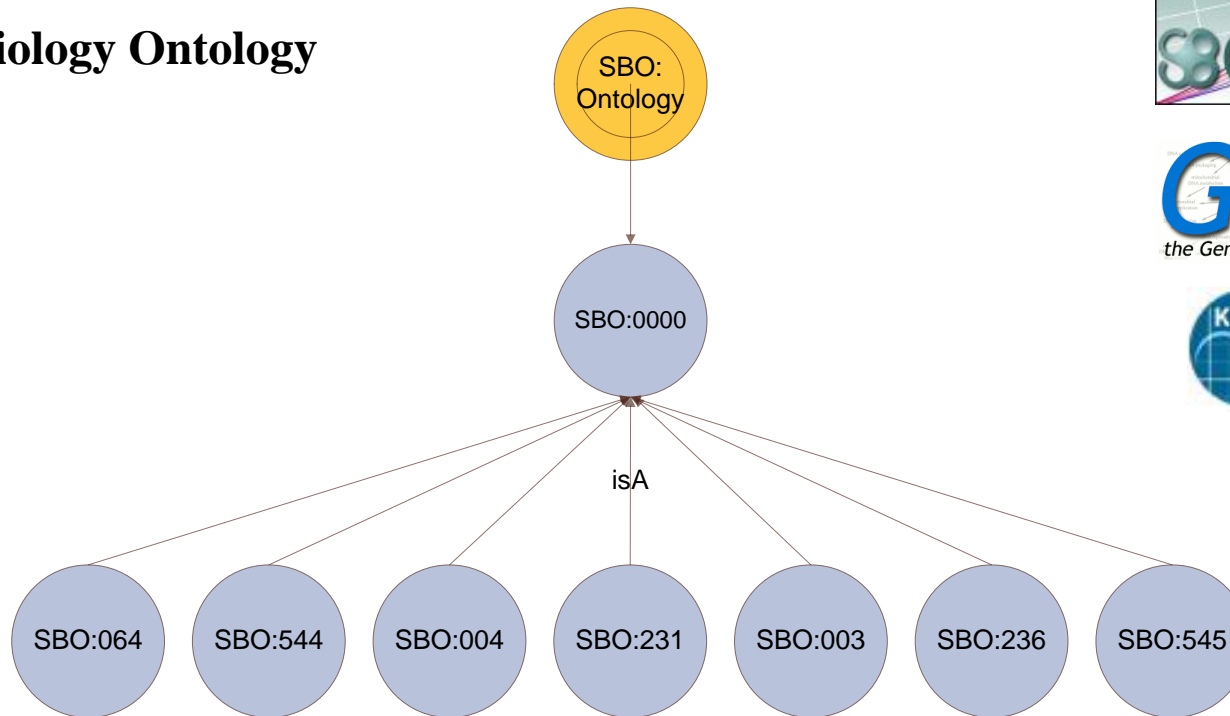
Variable:



Systems Biology Ontology



Root:



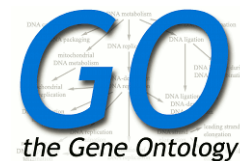
1. Level:

All Ontologies available in OWL can be added easily

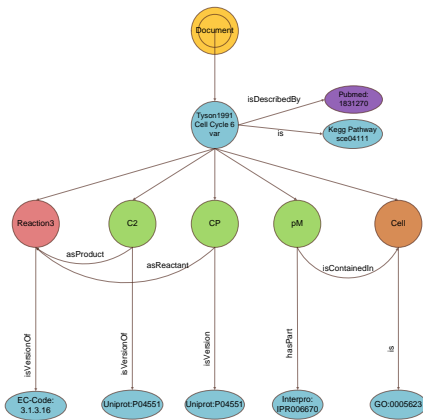
Models



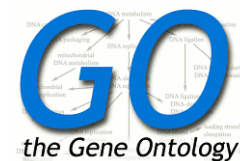
Model related data



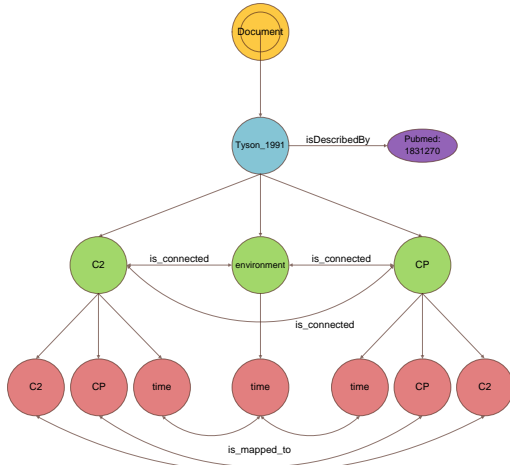
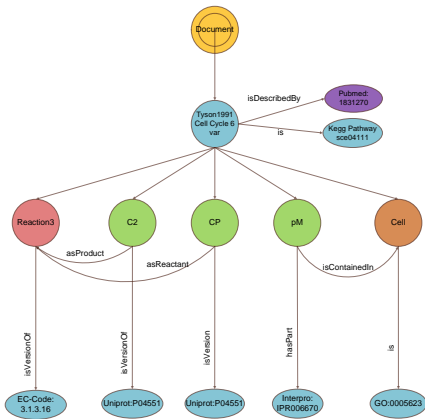
Models



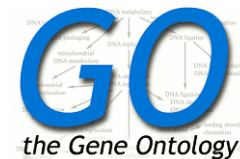
Model related data



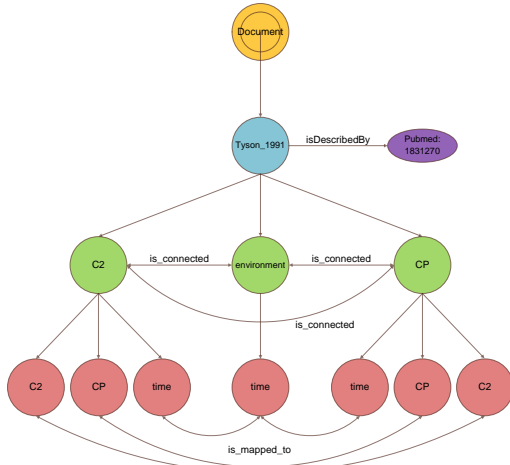
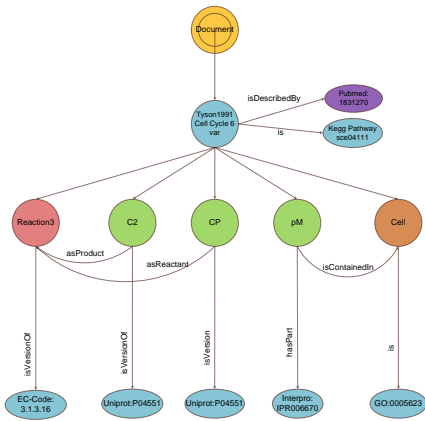
Models



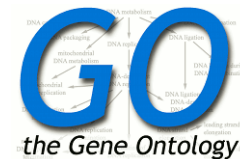
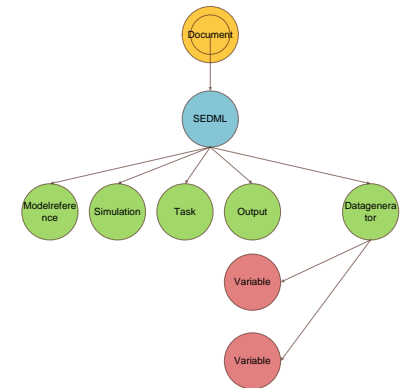
Model related data



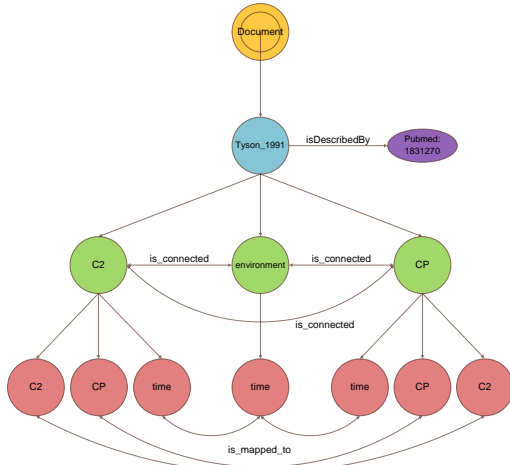
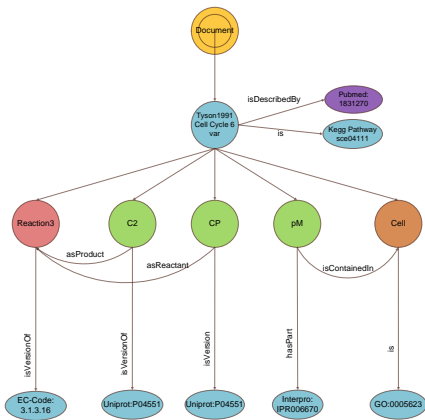
Models



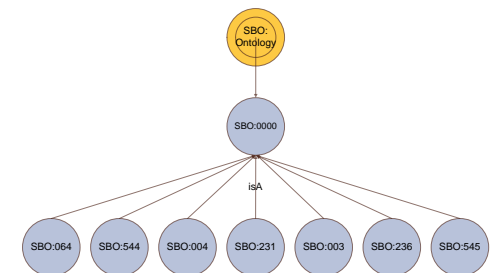
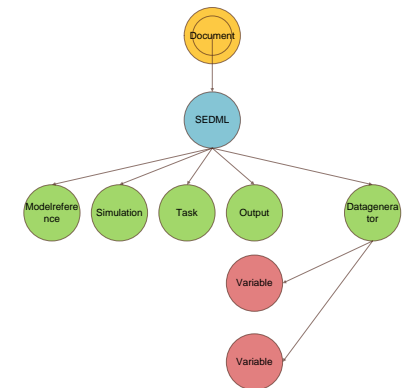
Model related data



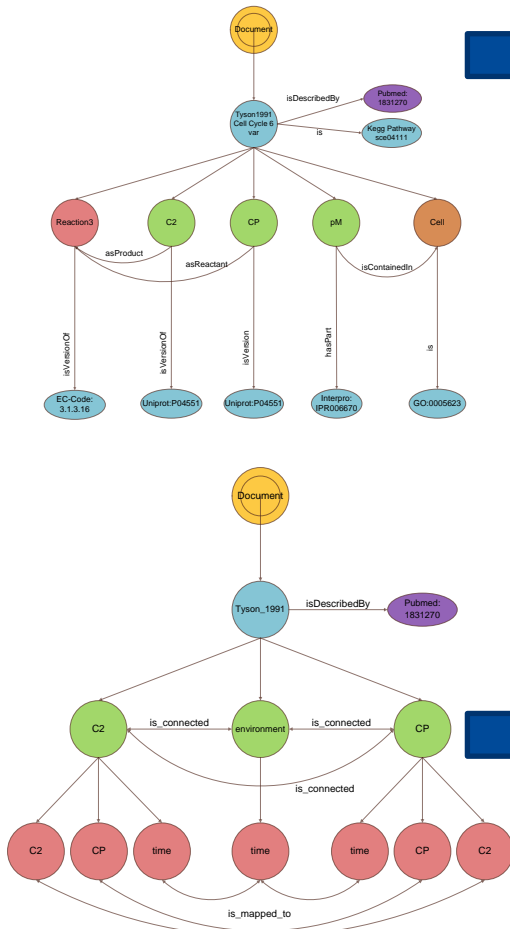
Models



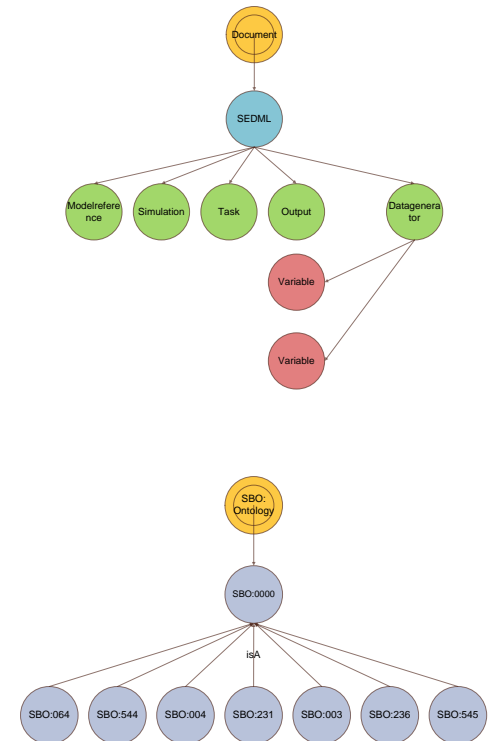
Model related data



Models



Model related data

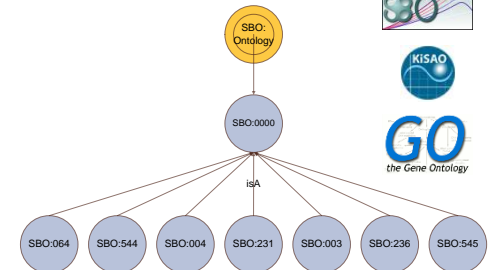
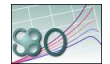
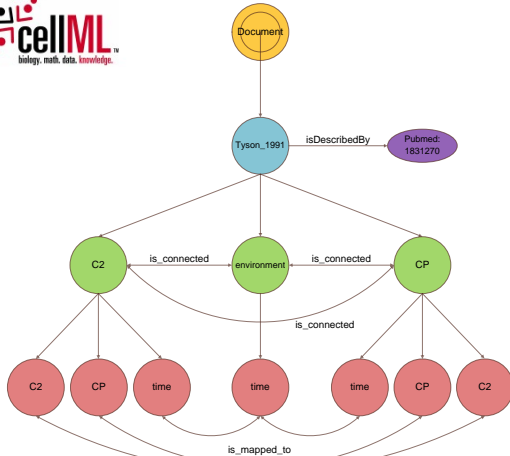
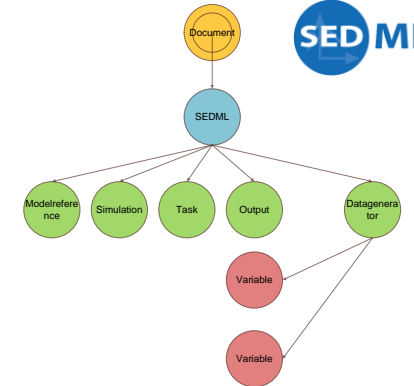
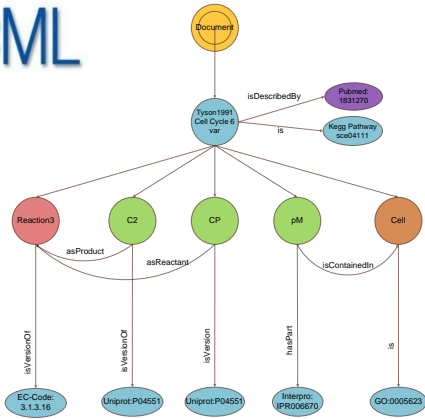


Neo4j
the world's leading
graph database

Models

Index

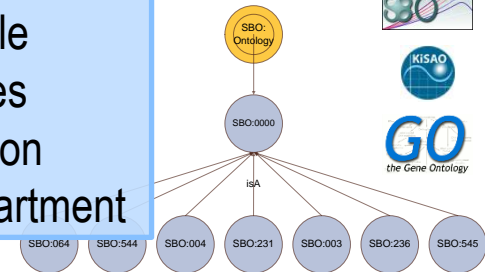
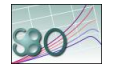
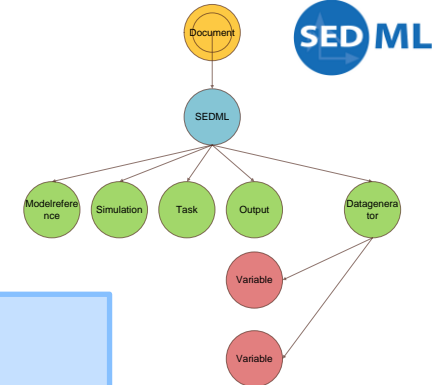
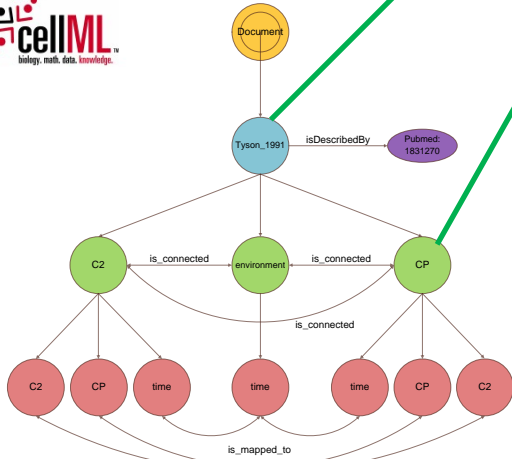
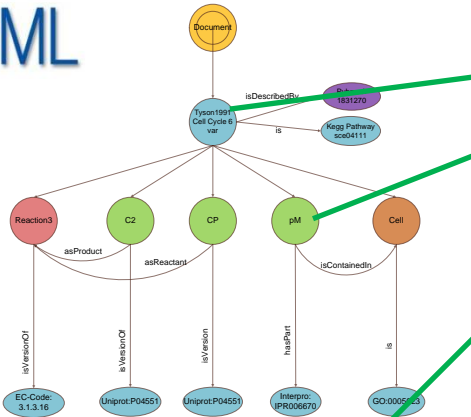
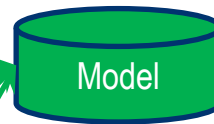
Model related data



Models

Index

Model related data

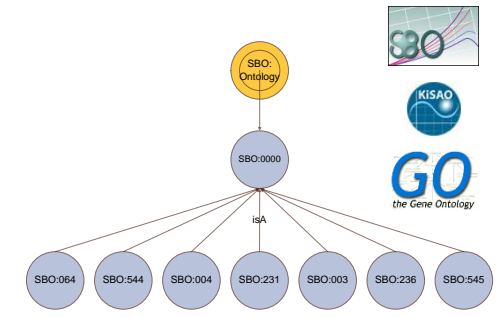
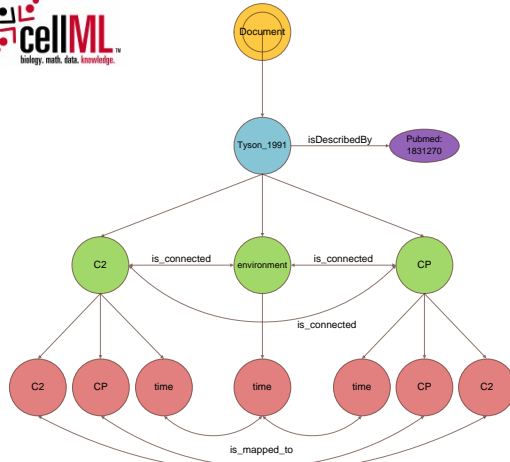
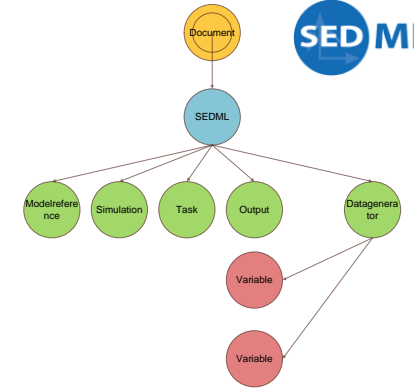
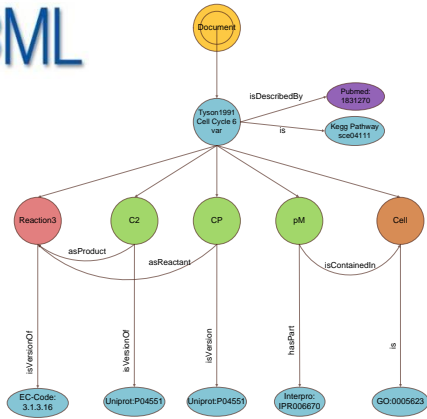
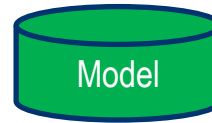


- Id
- Name
- Component
- Variable
- Species
- Reaction
- Compartment

Models

Index

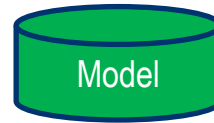
Model related data



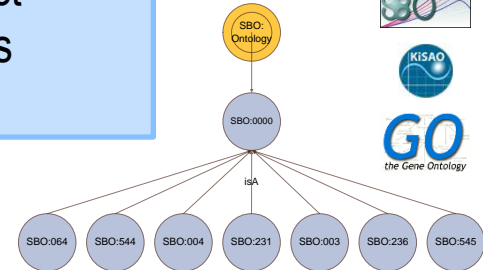
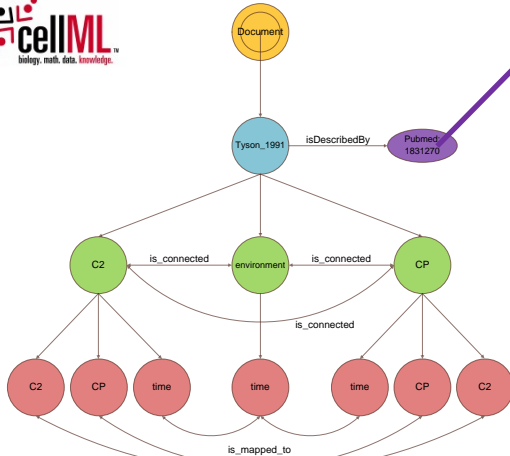
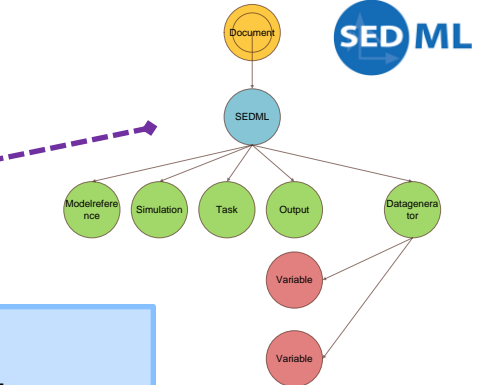
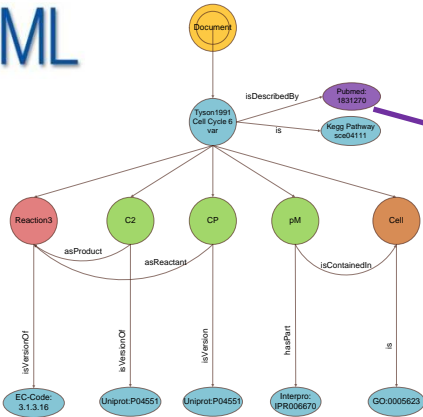
Models

Index

Model related data



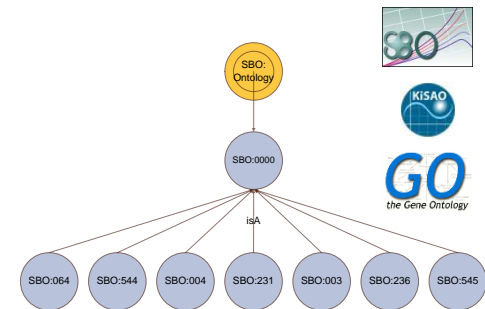
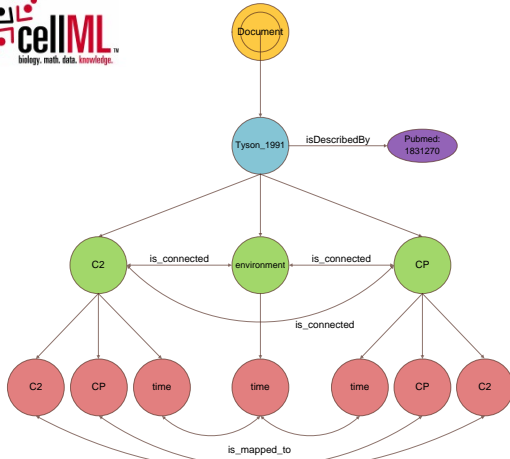
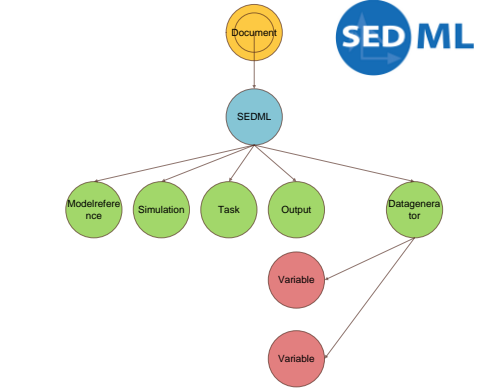
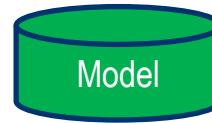
- Title
- Journal
- Abstract
- Authors
- ...



Models

Index

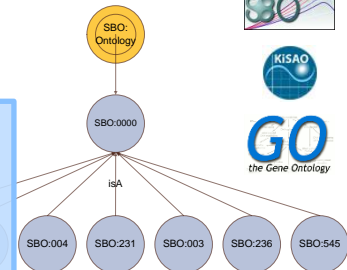
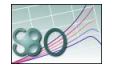
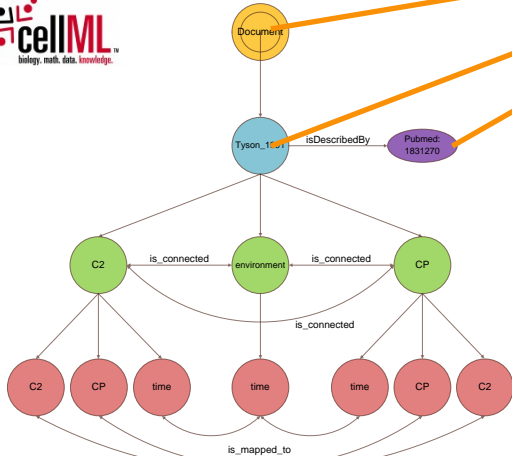
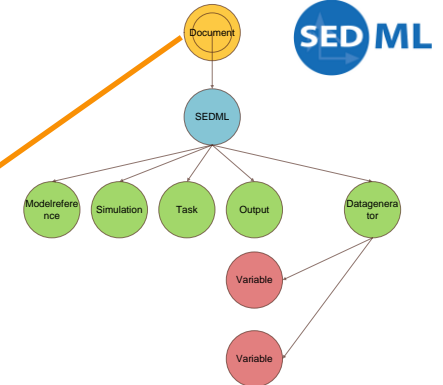
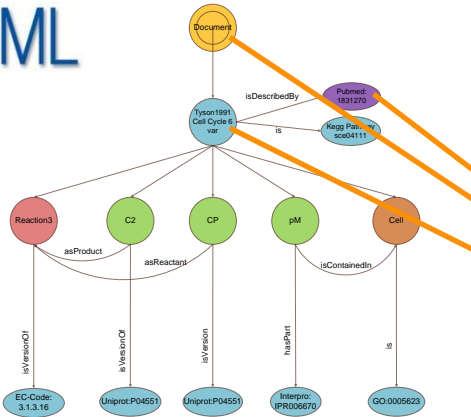
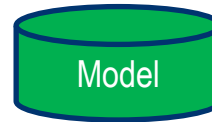
Model related data



Models

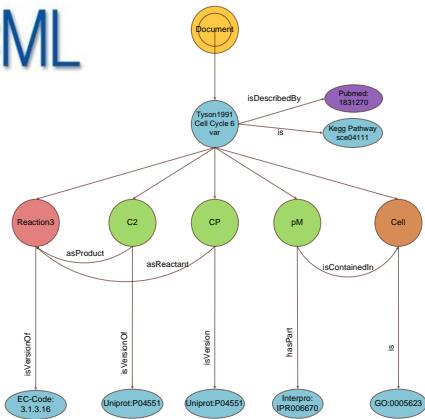
Index

Model related data

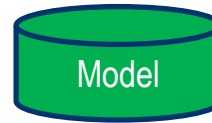


- First name
- Last name
- Organization
- Email

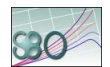
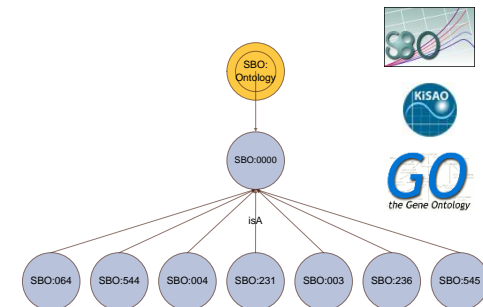
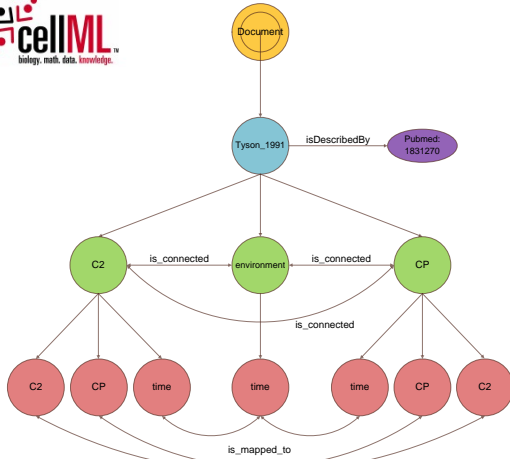
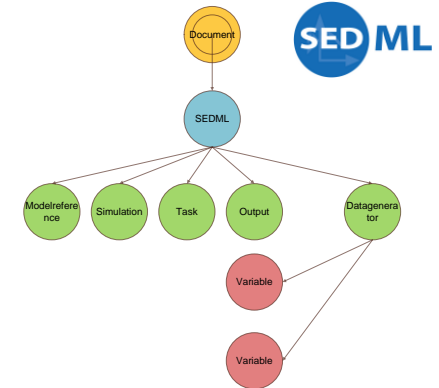
Models



Index



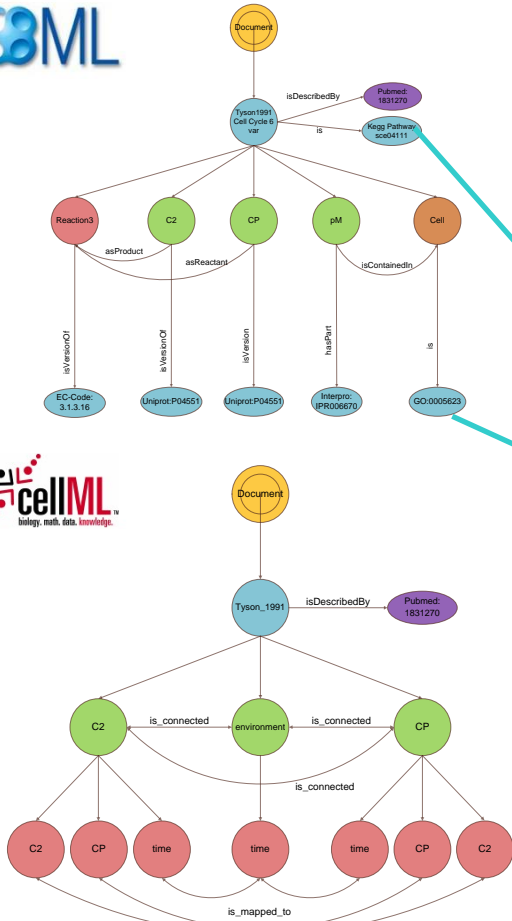
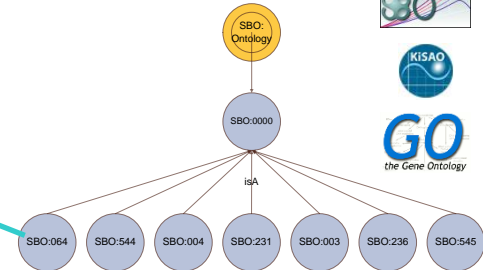
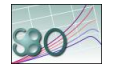
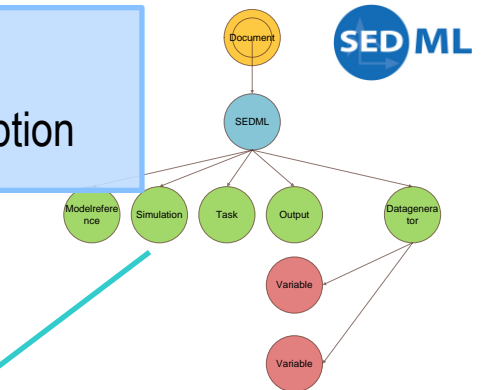
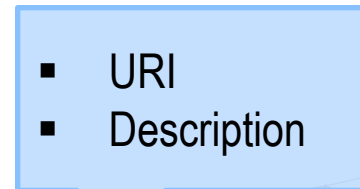
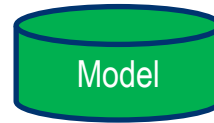
Model related data



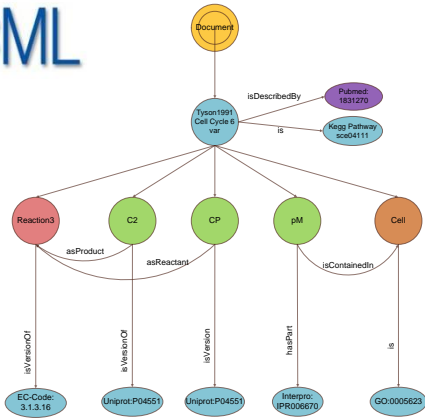
Models

Index

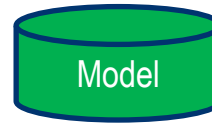
Model related data



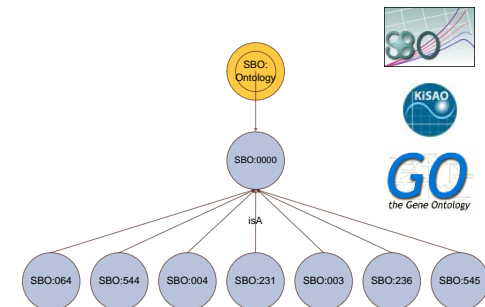
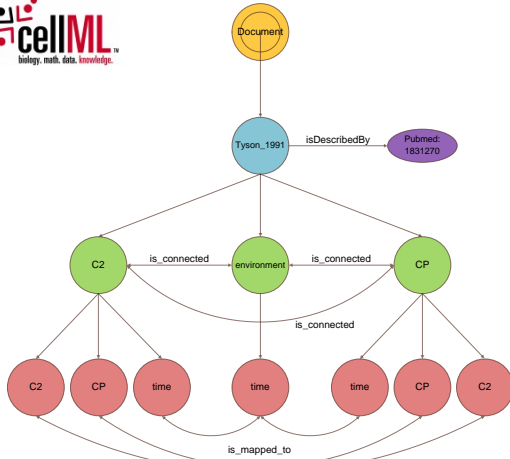
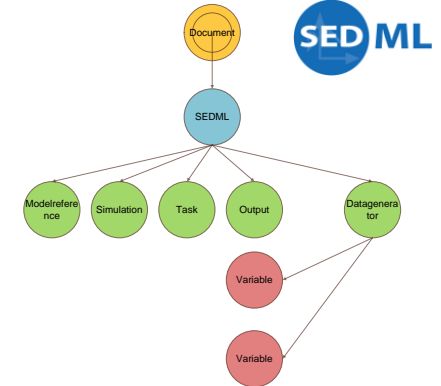
Models



Index



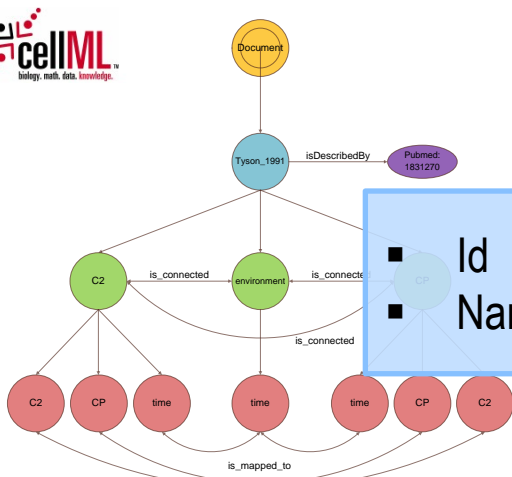
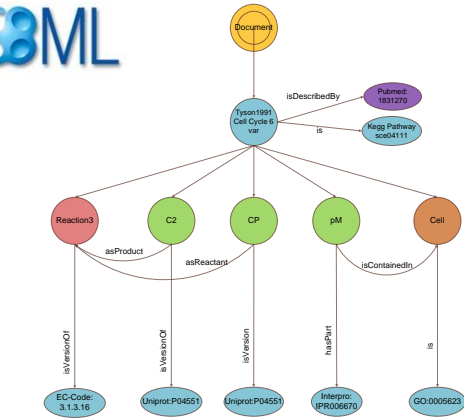
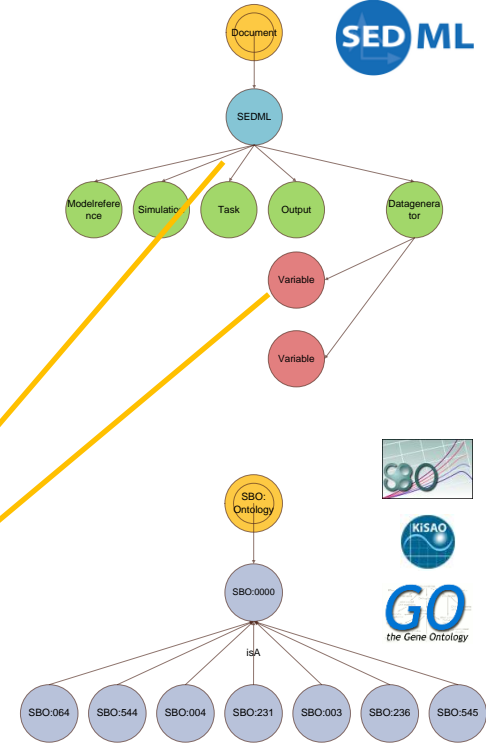
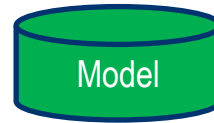
Model related data



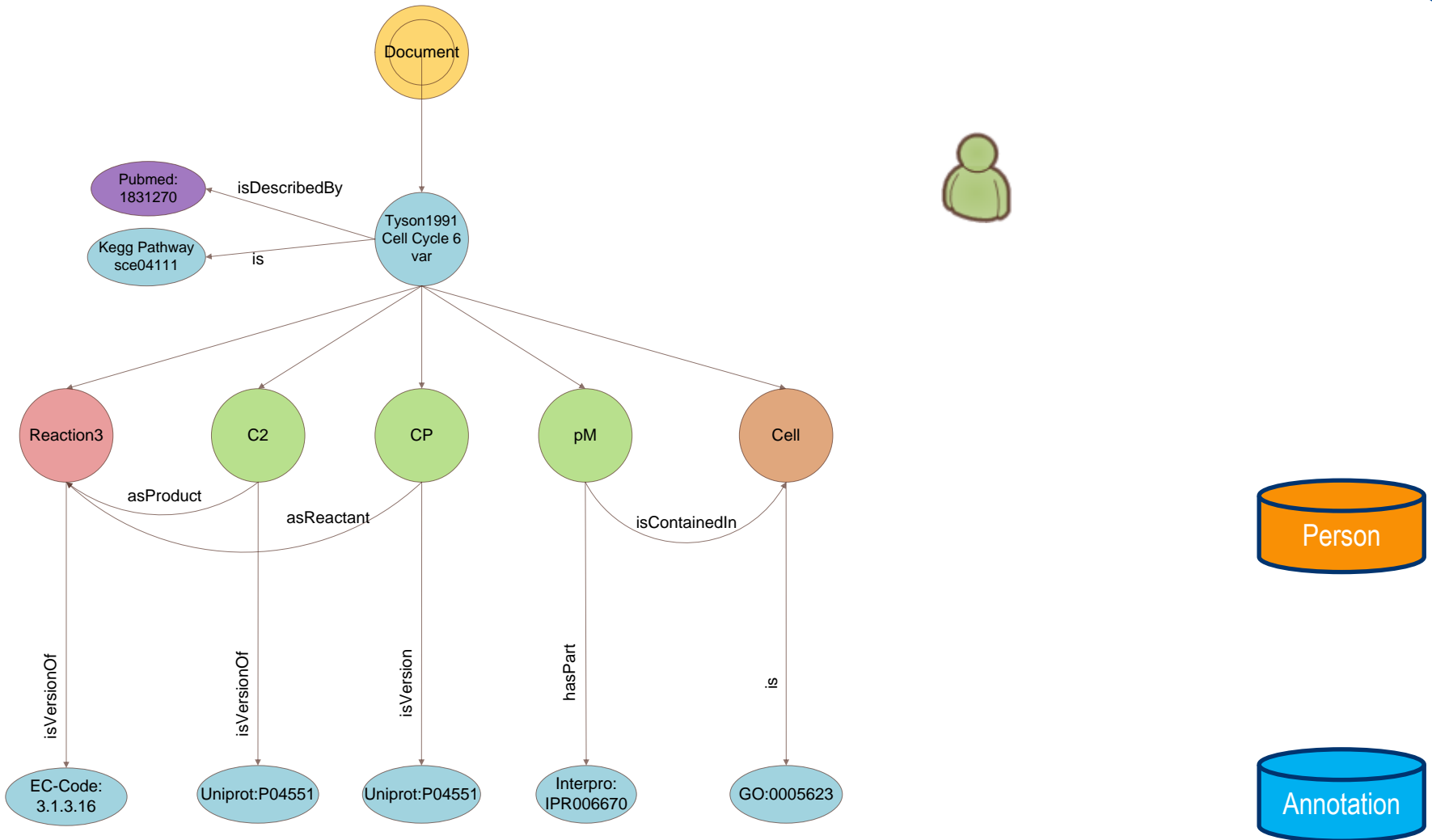
Models

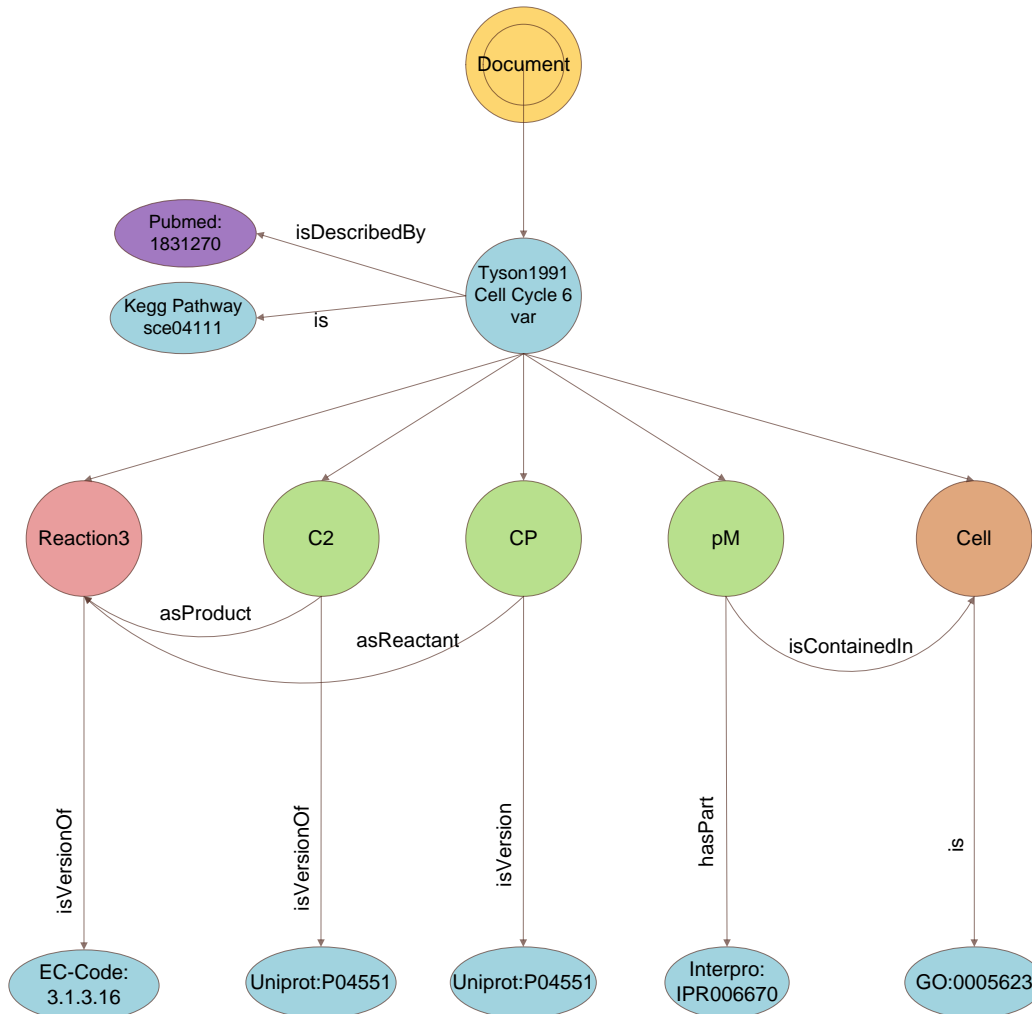
Index

Model related data



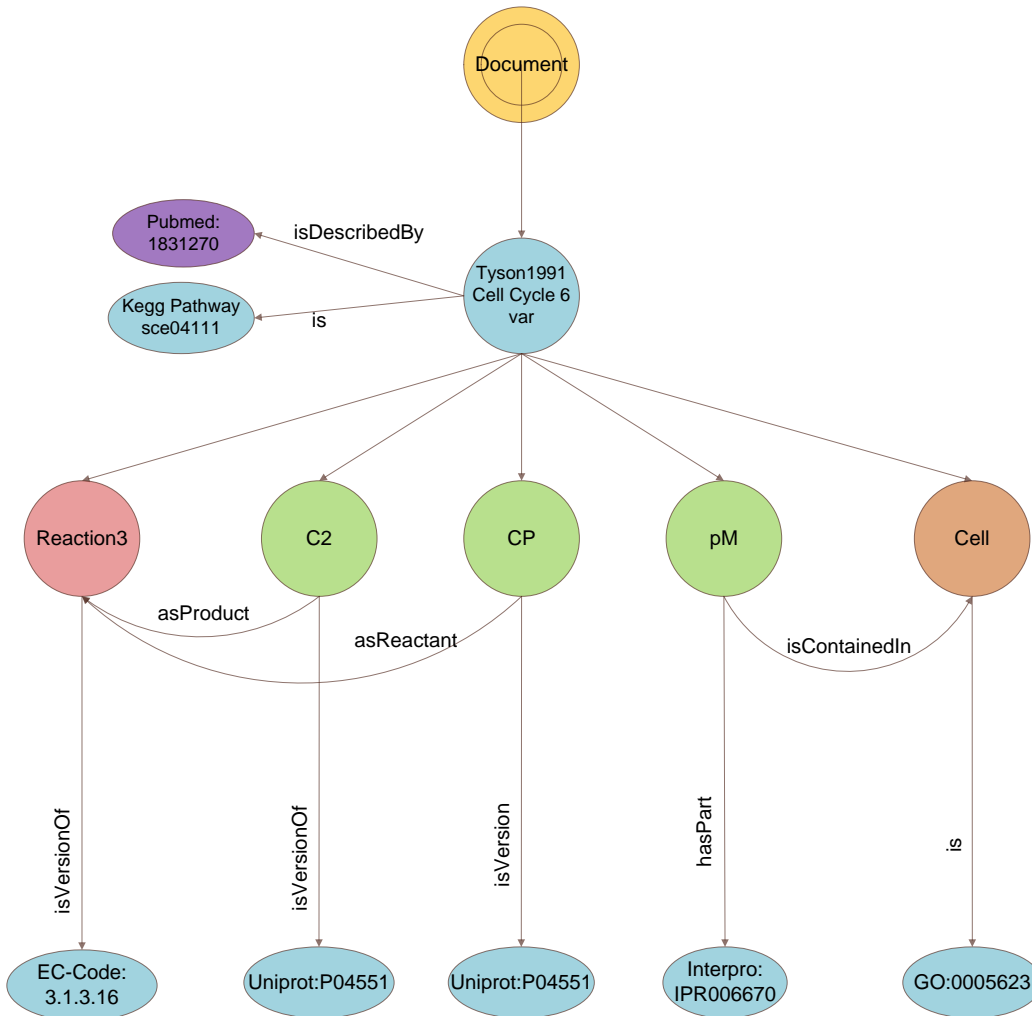
Id Name





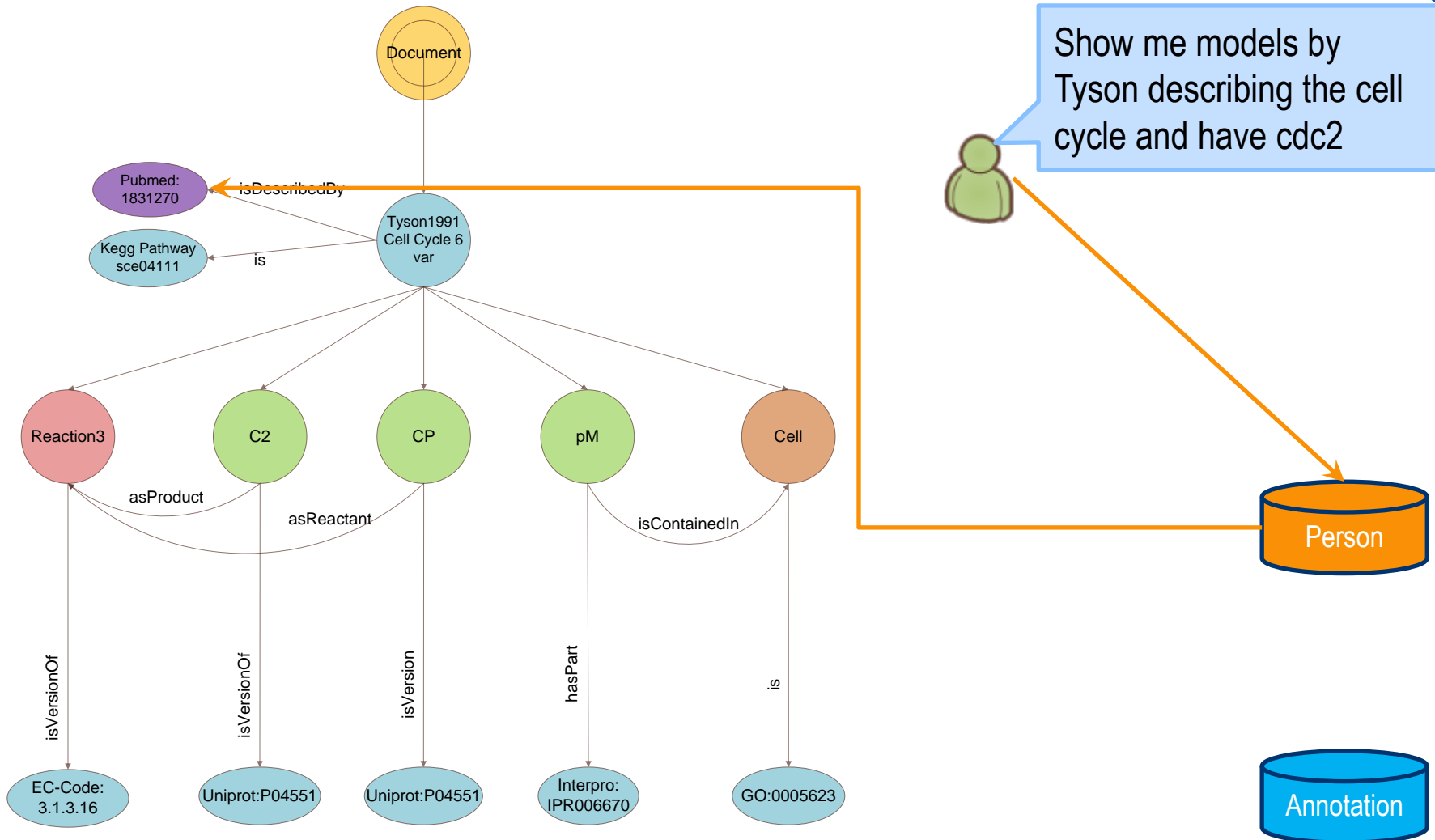
Show me models by Tyson describing the cell cycle and have cdc2

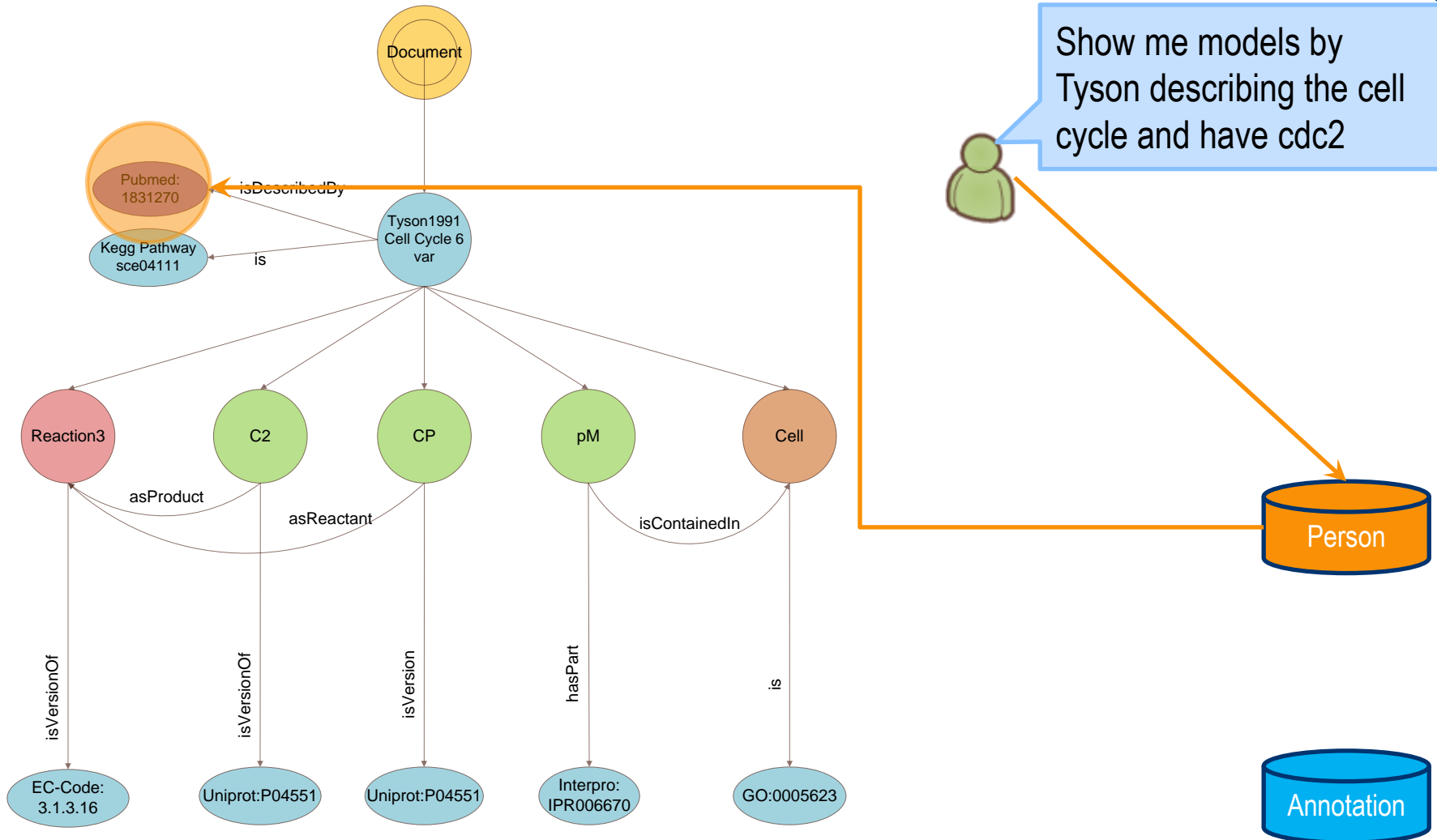




Show me models by Tyson describing the cell cycle and have cdc2



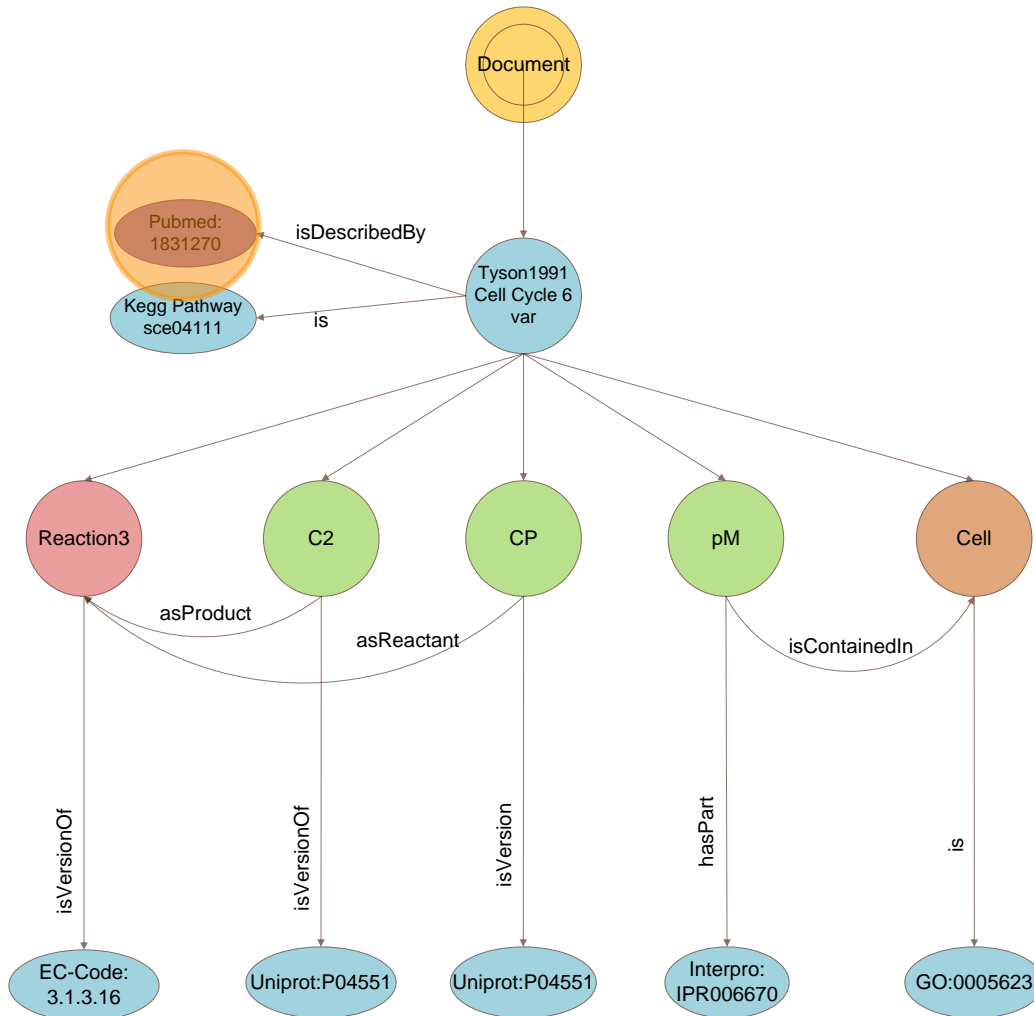




Show me models by
Tyson describing the cell
cycle and have cdc2

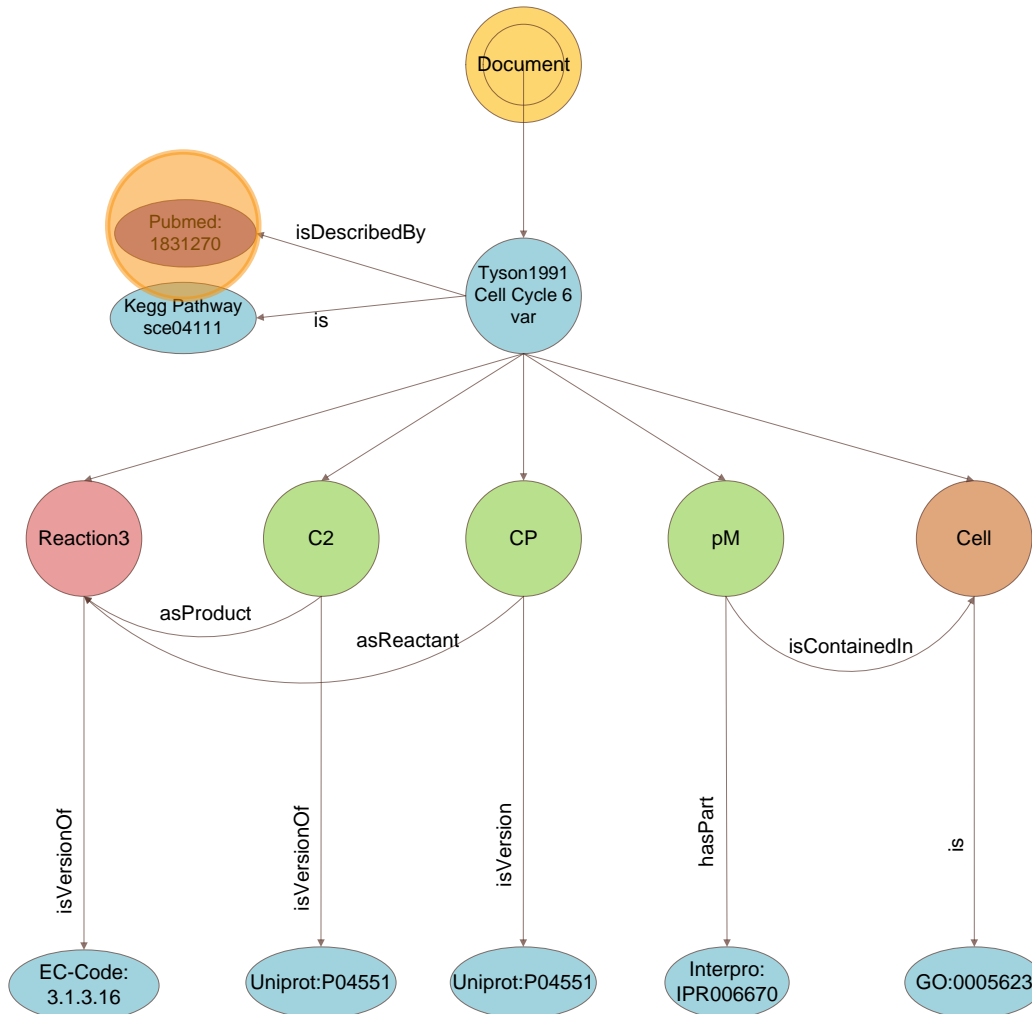
Person

Annotation



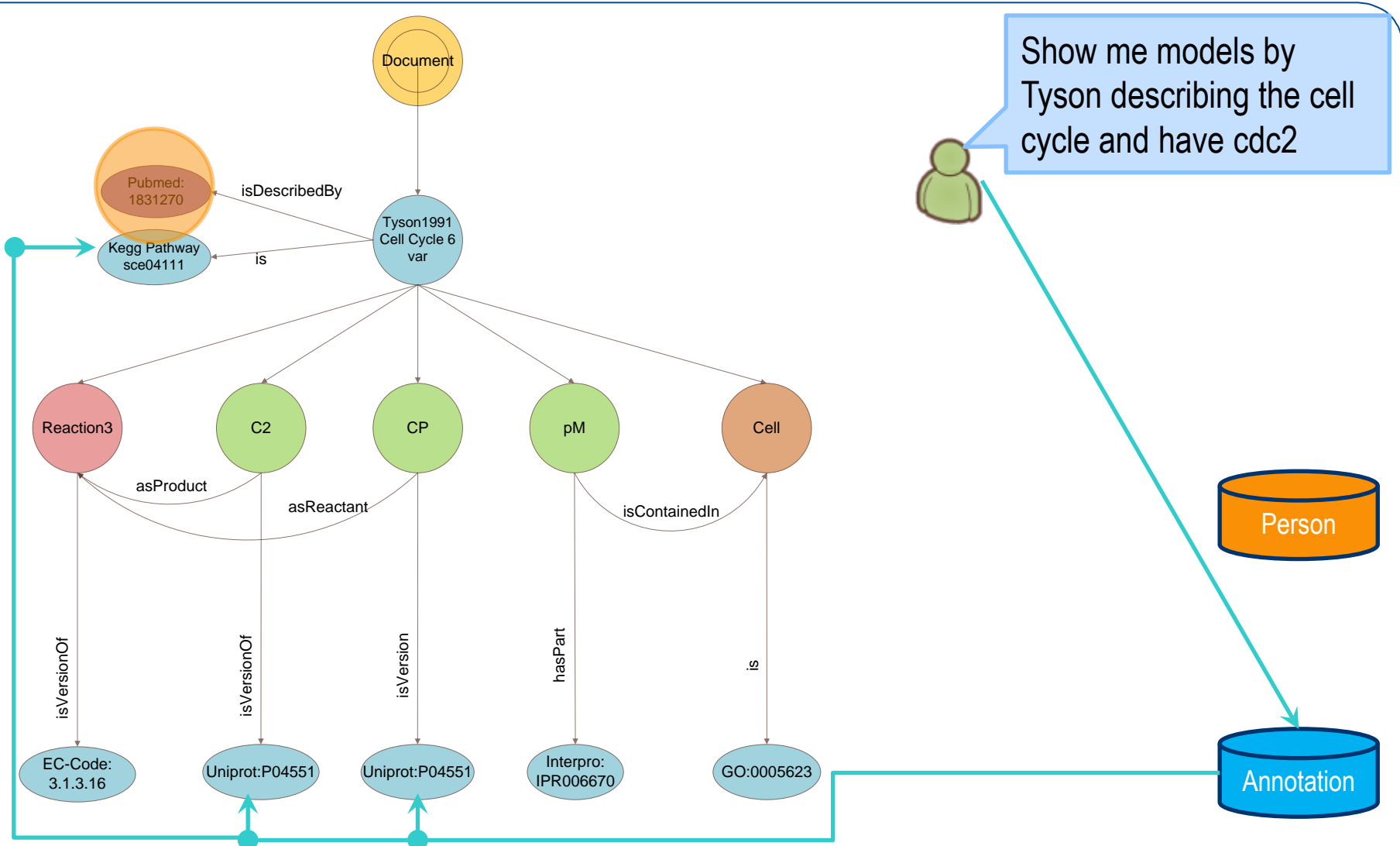
Show me models by Tyson describing the cell cycle and have cdc2

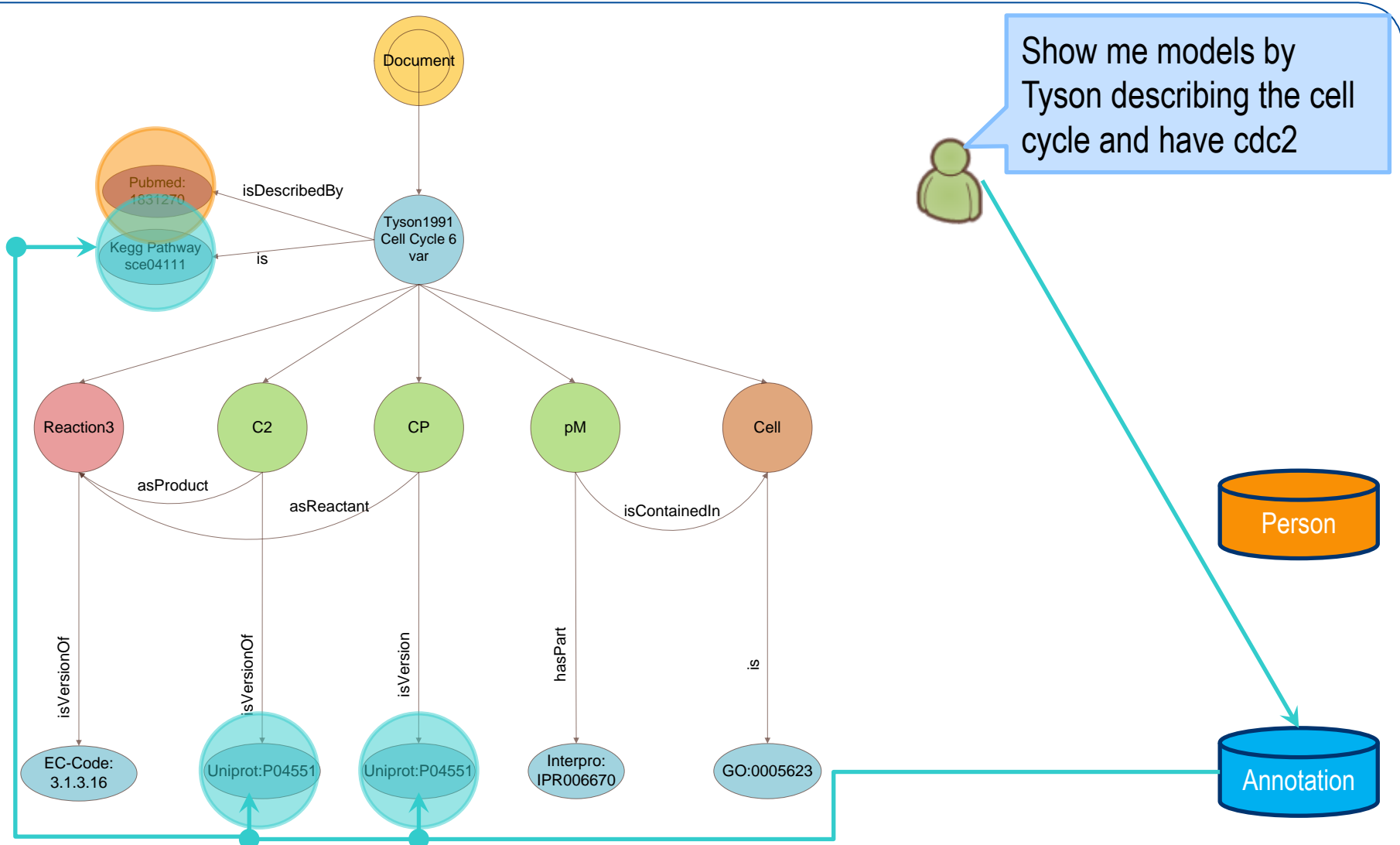


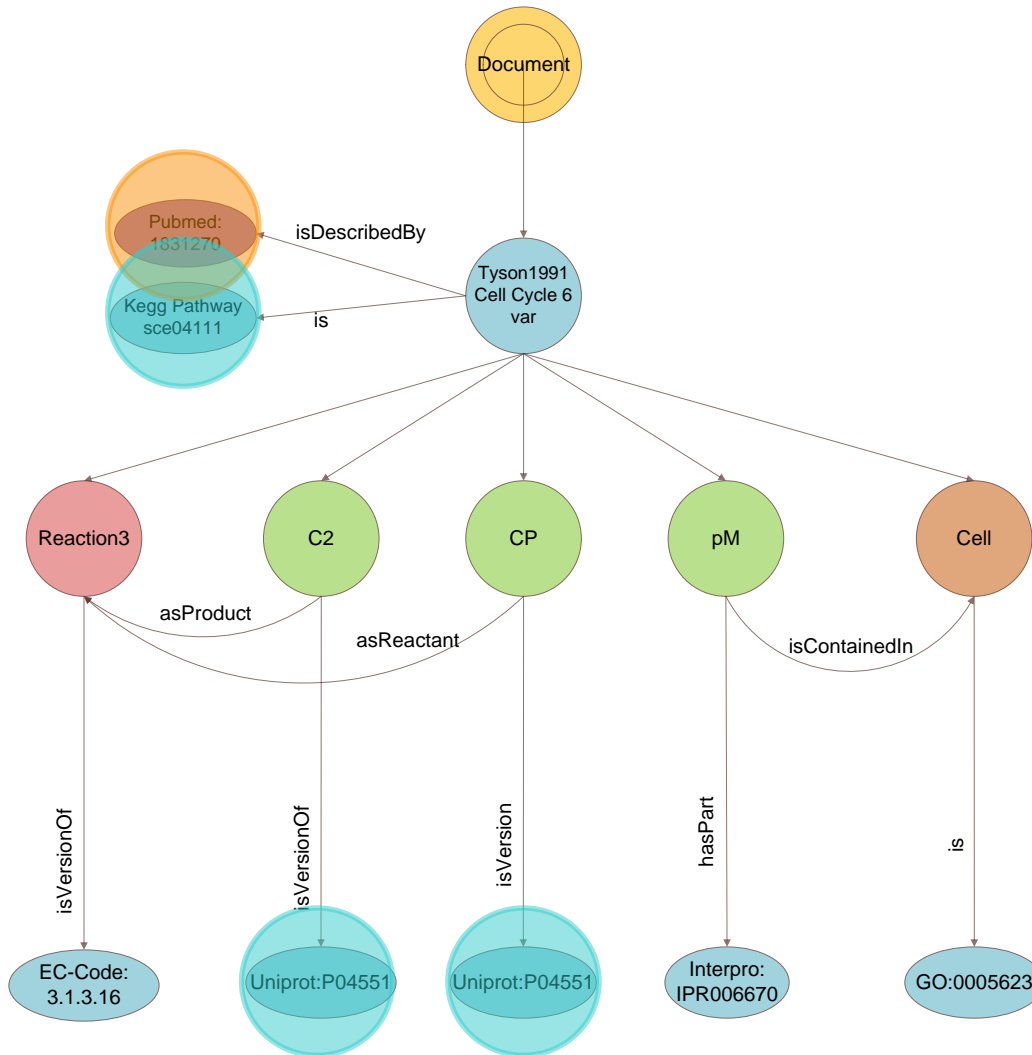


Show me models by
Tyson describing the cell
cycle and have cdc2



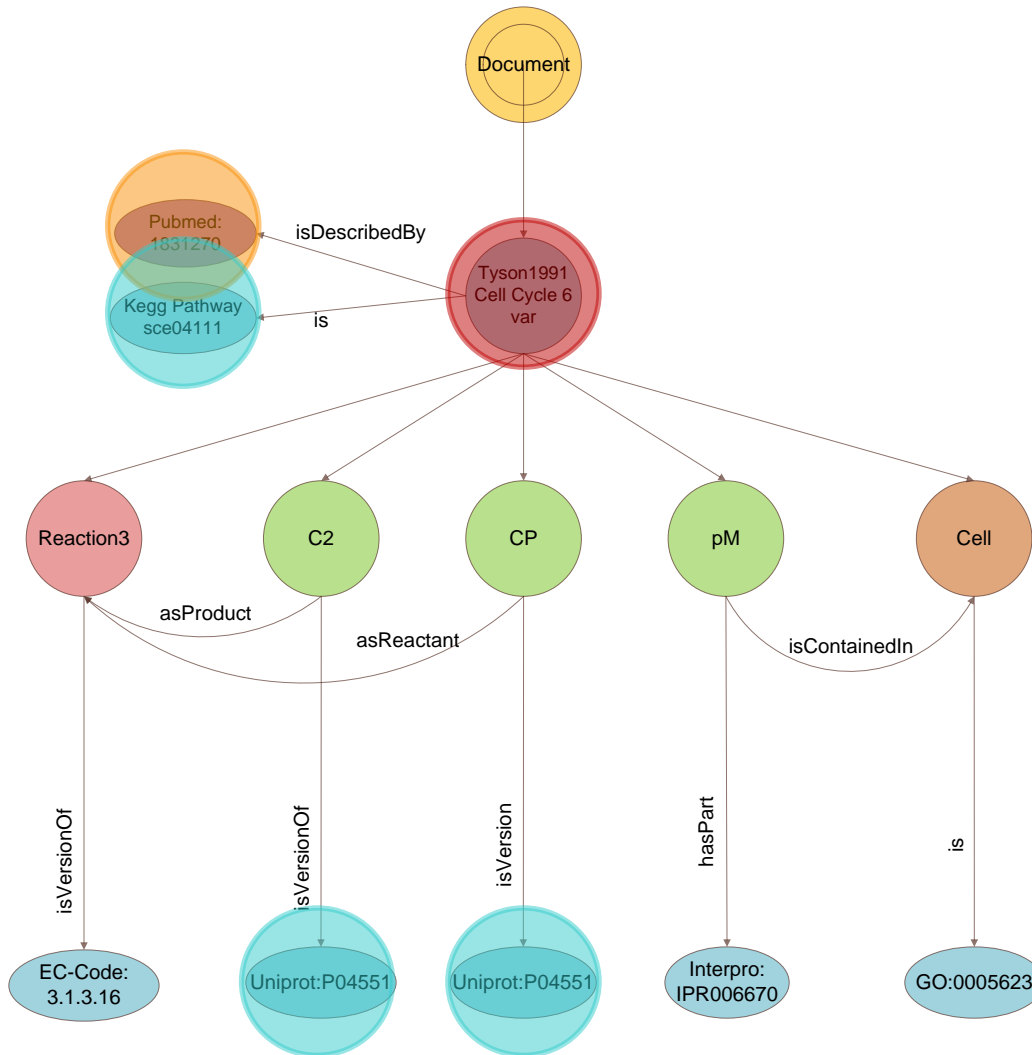






Show me models by
Tyson describing the cell
cycle and have cdc2

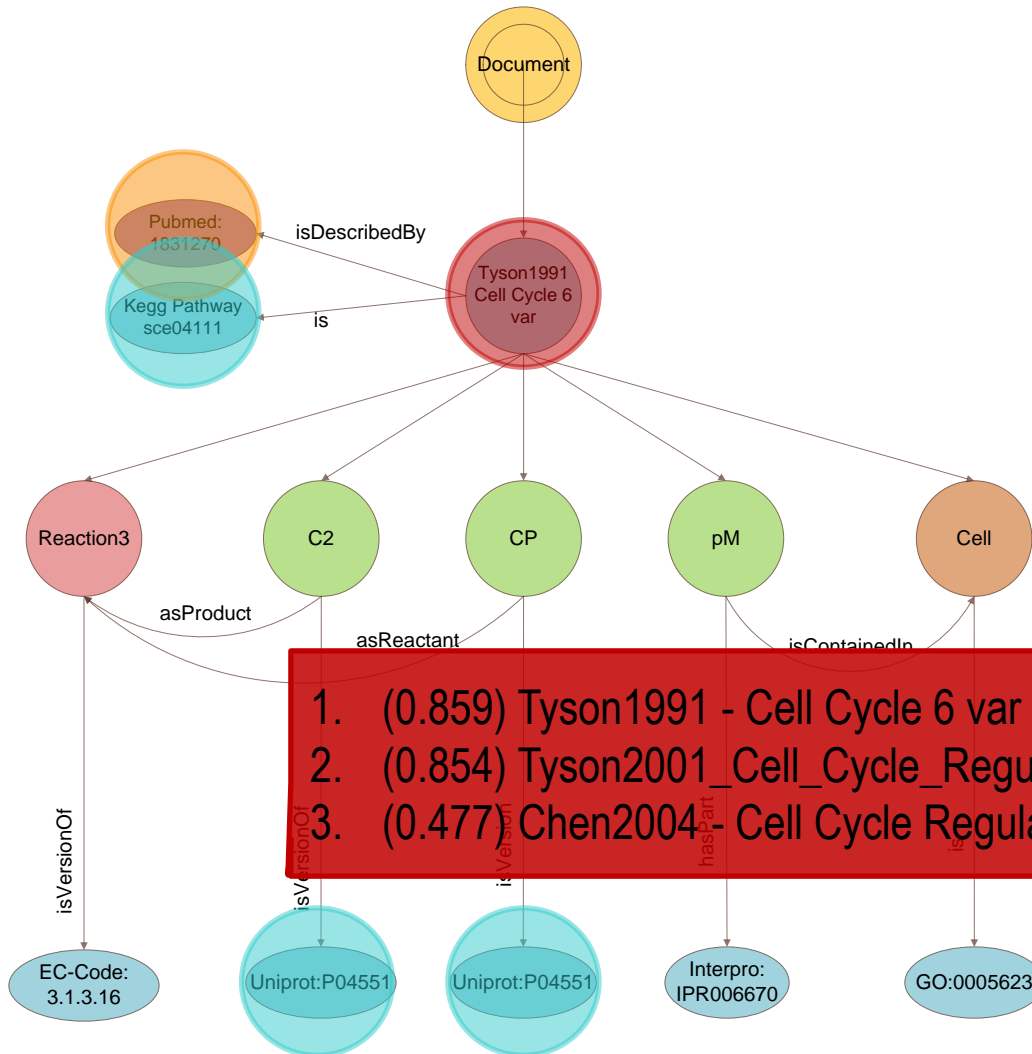




Show me models by Tyson describing the cell cycle and have cdc2

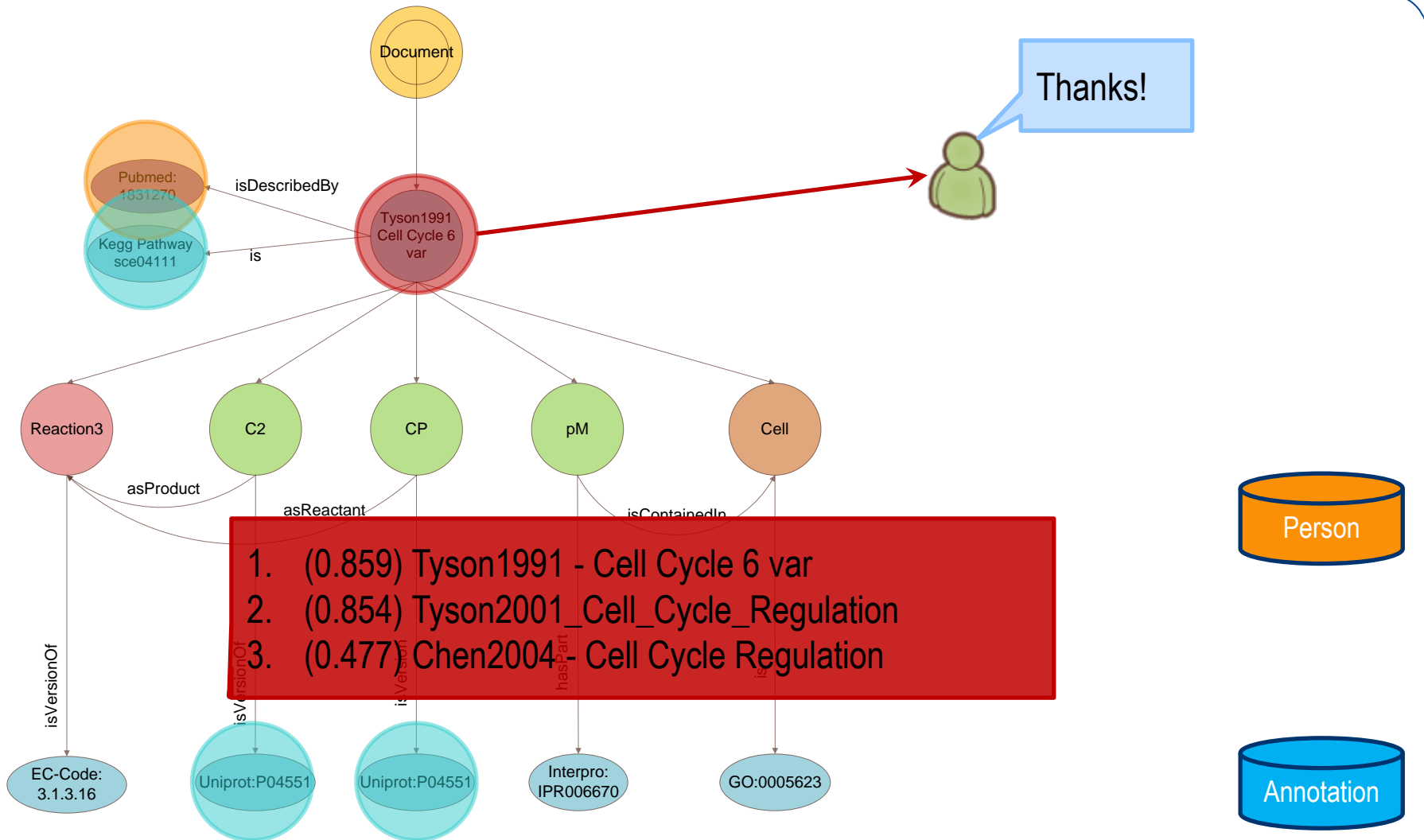


Show me models by Tyson describing the cell cycle and have cdc2



1. (0.859) Tyson1991 - Cell Cycle 6 var
2. (0.854) Tyson2001_Cell_Cycle_Regulation
3. (0.477) Chen2004- Cell Cycle Regulation





Preliminary Results

- Dataset 1:
 - curated models available in Biomedels DB
 - 841 exposed CellML Models from Physiome Model Repository
 - 38 SED-ML files (from our Cell-Cycle side project - Uni Rostock)

- Webpage explaining the web service:
 - <http://sems.uni-rostock.de/projects/morre/>

- Pmr2 staging machine already implements the web service:
 - http://staging.physiomeproject.org/morre_pmr2_search
 - CellML only!

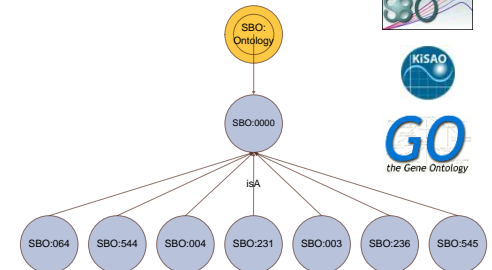
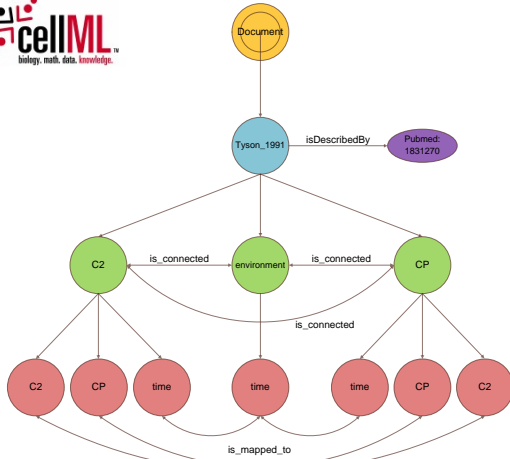
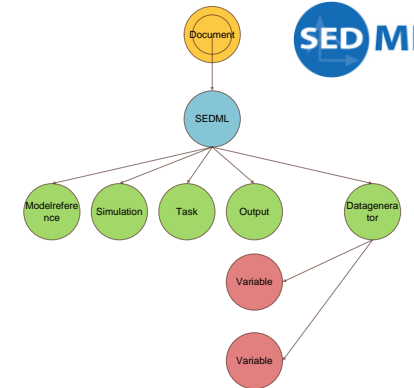
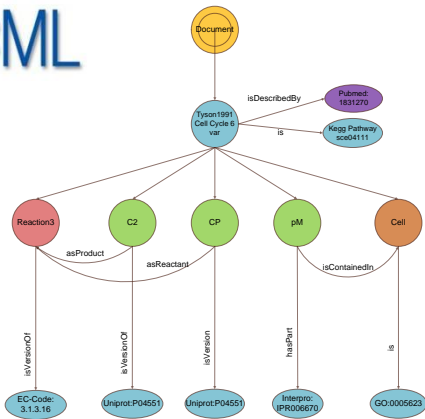
Preliminary Results

- Dataset 2:
 - 140.811 models from path2models project
- Database statistics:
 - 45,534,192 nodes
 - 31,784,961 properties
 - 492,296,146 relationships
 - 81GB database disk usage
- Currently no annotation index...
 - Downloading the content of 4,583,240 unique URIs is beyond the capability of my desktop or laptop machine

Models

Links

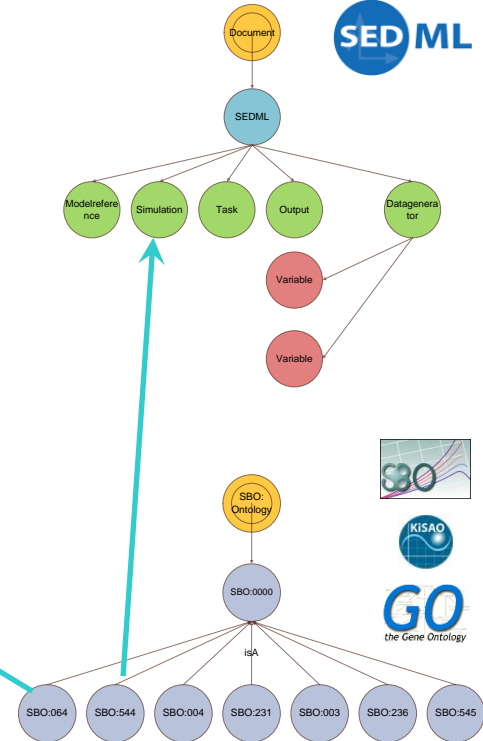
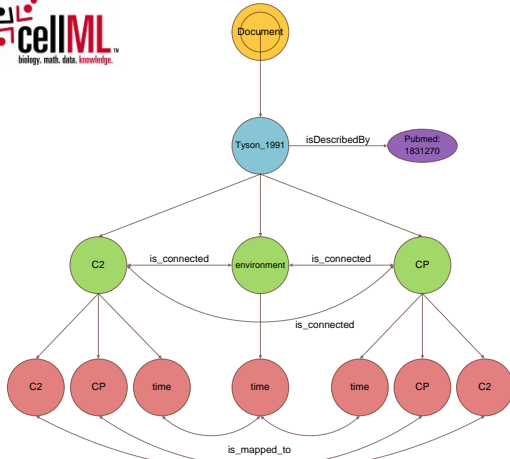
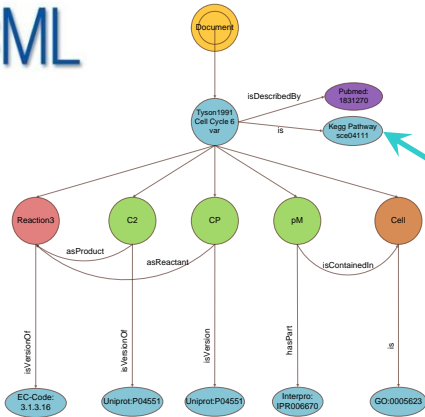
Model related data



Models

Links

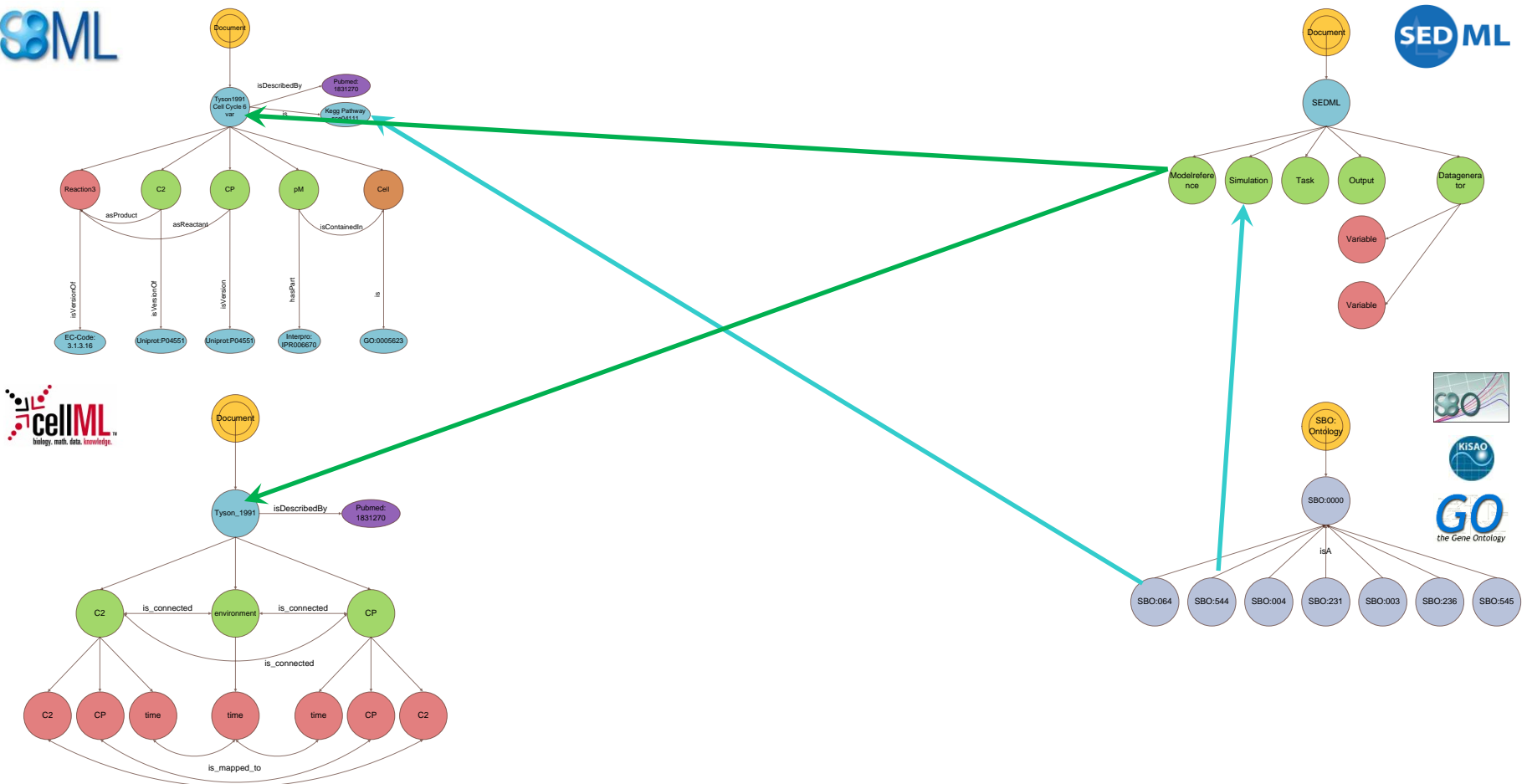
Model related data



Models

Links

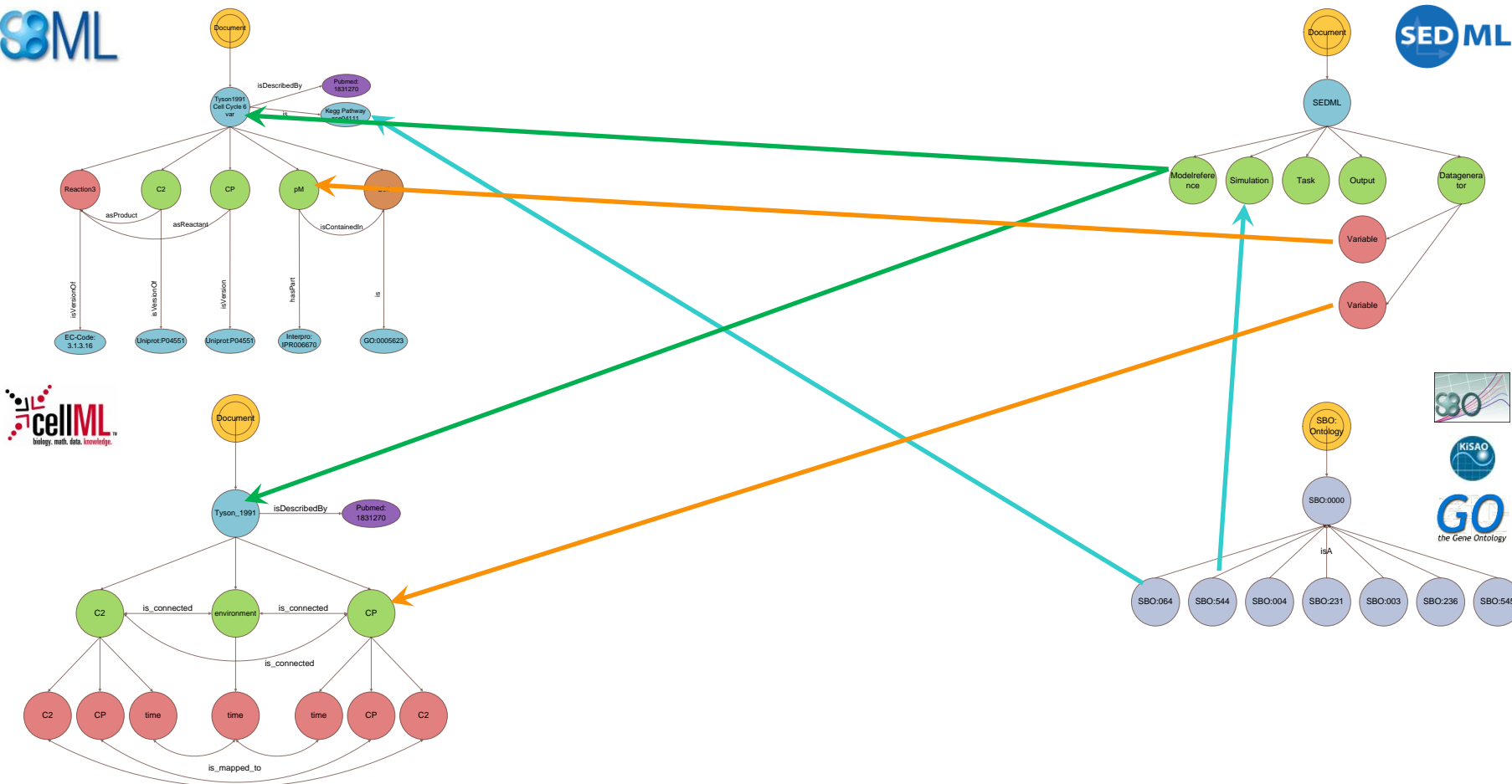
Model related data



Models

Links

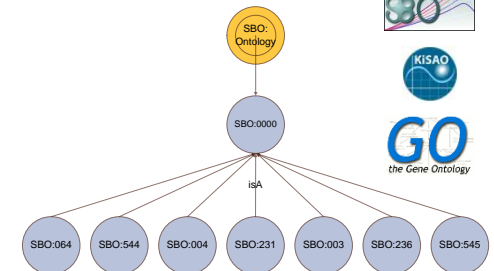
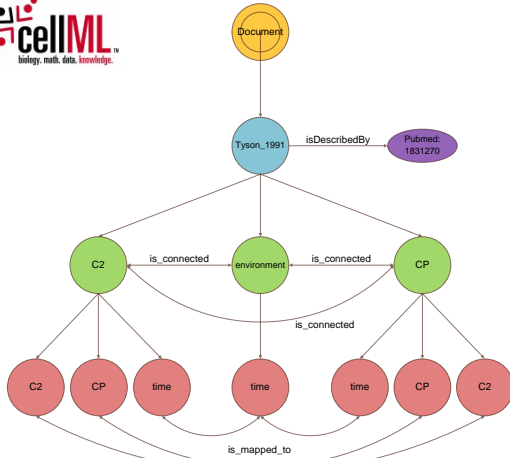
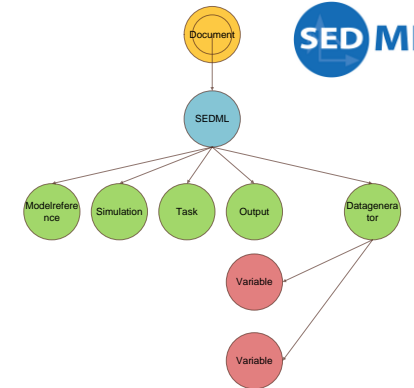
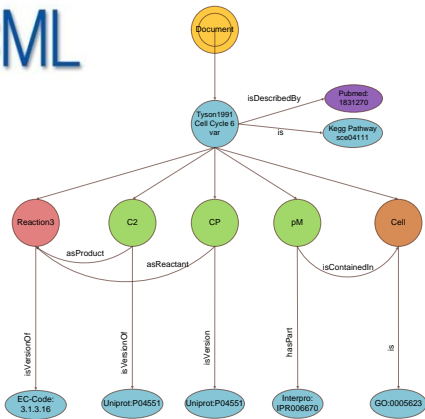
Model related data



Models

Links

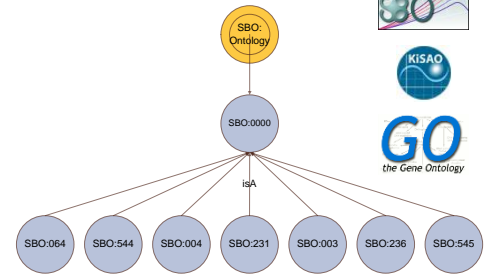
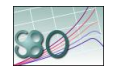
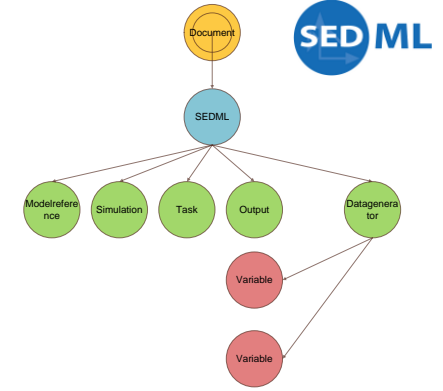
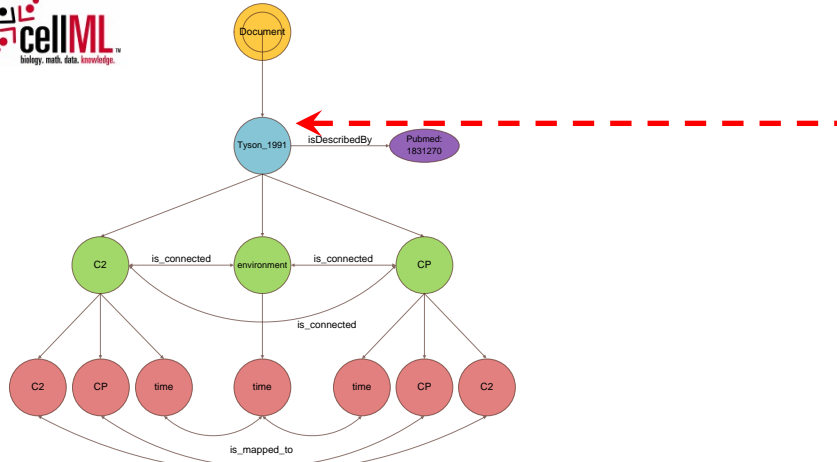
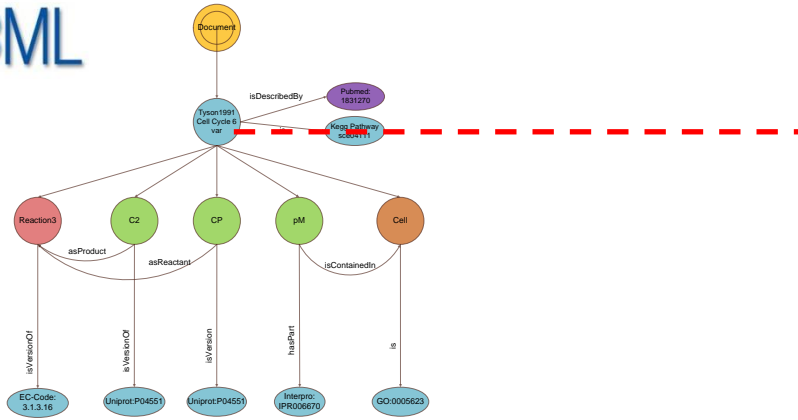
Model related data



Models

Links

Model related data



Making use of the Graph Database

```
MATCH o:GOOntology  
RETURN count(o);
```

```
MATCH o:GOOntology  
WHERE o<-[:IS_ONTOLOGY_ENTRY]-()  
RETURN count(o);
```

Making use of the Graph Database

```
MATCH o:GOOntology  
RETURN count(o);
```

==> 39787

```
MATCH o:GOOntology  
WHERE o<-[:IS_ONTOLOGY_ENTRY]-()  
RETURN count(o);
```


Making use of the Graph Database

```
MATCH o:GOOntology  
RETURN count(o);
```

==> 39787

```
MATCH o:GOOntology  
WHERE o<-[:IS_ONTOLOGY_ENTRY]-()  
RETURN count(o);
```

==> 1097

Making use of the Graph Database

Retrieve the three most used annotations:

```
MATCH r:RESOURCE, r-[rel:BELONGS_TO]->()
```

```
WITH r, count(rel) AS numOfRelations
```

```
ORDER BY numOfRelations DESC LIMIT 3
```

```
RETURN r.URI, numOfRelations;
```

Making use of the Graph Database

Retrieve the three most used annotations:

MATCH r:RESOURCE, r-[rel:BELONGS_TO]->()

WITH r, count(rel) **AS** numOfRelations

ORDER BY numOfRelations **DESC LIMIT** 3

RETURN r.URI, numOfRelations;

r.URI	numOfRelations
urn:miriam:biomodels.sbo:SBO:0000009	1127
urn:miriam:biomodels.sbo:SBO:0000252	509
http://identifiers.org/obo.go/GO:0043241	484

Tools

- All java based:
 - Neo4J – a graph database
 - Lucene – an index and search tool
 - Jetty – a web server
 - A parser for each format
 - CellML Api (java bindings)
 - jLibSBML
 - jSedML

Take Home Message

- Models and related data grow in size and complexity
- Ranked retrieval is a necessary feature for model databases.
- Linking models, simulation and other model related data allows to query the spanned network



Thanks for your attention.

Questions?

ron.henkel@uni-rostock.de