

# Semantic Annotation with SBML and CellML Models



John Gennari & Max Neal  
(and team!)

# The John & Max show

## *Semantic Annotation & Composition of SBML and CellML models*

Motivation, methods,  
prior work



Details, examples,  
results

*What is it?*

## *What is it?*

- ◆ Model-level annotation
  - Describing the whole model
  - Author, date, publication, overview, etc.
- ◆ Code-level annotation
  - Descriptions of individual species, reactions, variables
  - Proteins, chemicals, measured values, parameters & variables

## *What is it?*

### ◆ Model-level annotation

- Describing the whole model
- Author, date, publication, overview, etc.

### ◆ Code-level annotation

- Descriptions of individual species, reactions, variables
- Proteins, chemicals, measured values, parameters & variables

## *Why do it?*

- ◆ Documentation
- ◆ Match data to models
- ◆ Merge and Reuse models

## *Why do it?*

- ◆ Documentation
- ◆ Match data to models
- ◆ Merge and Reuse models

# Story time

---

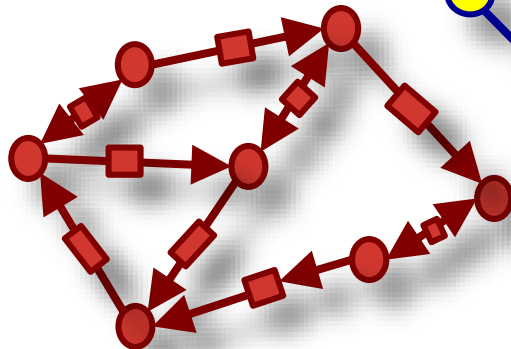




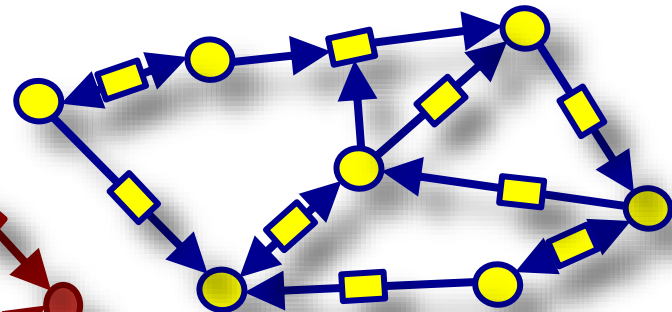
# Story time



Jane's model



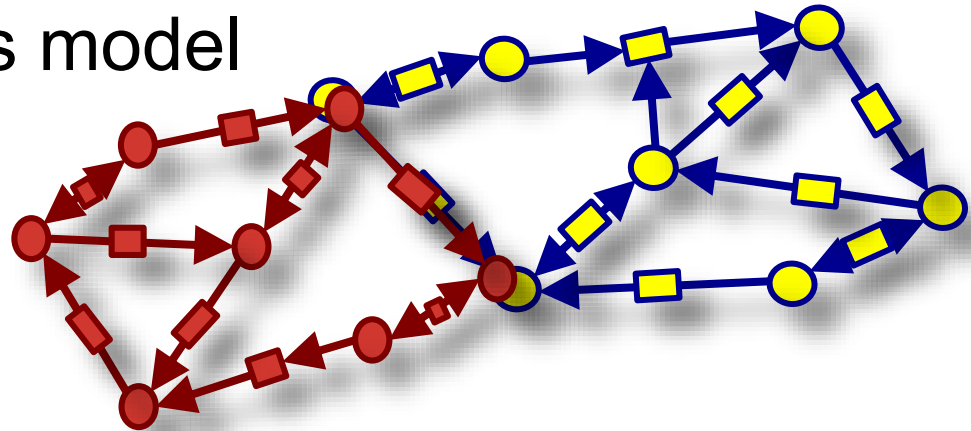
DeRown's model



# Story time



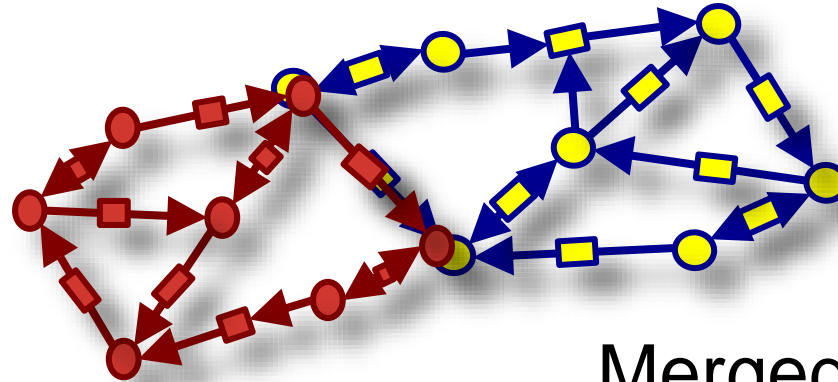
Jane's model



DeRow's model

Augmented and  
improved model!

# How?

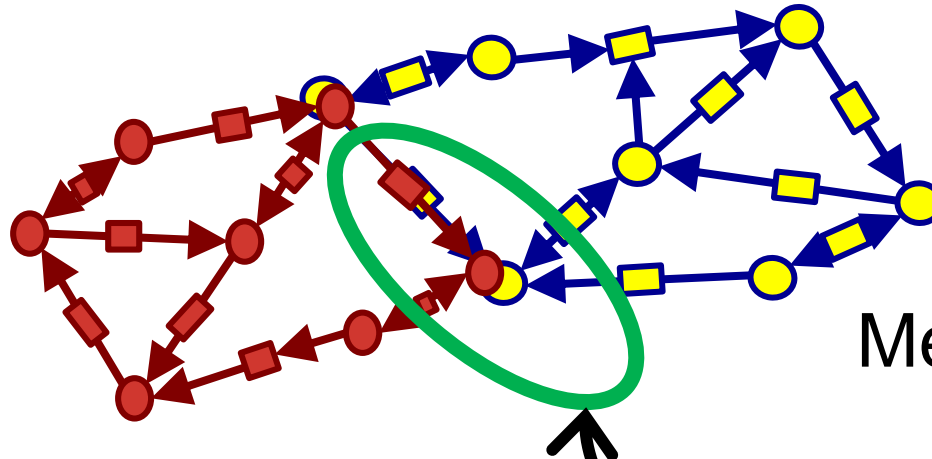


Merged model

- ◆ Rebuild DeRow's model from publication
- ◆ Retrieve's DeRow's code from GitHub
- ◆ Hire DeRow's co-author as Postdoc

**Success, but.....**

# (semi) Automatic merging

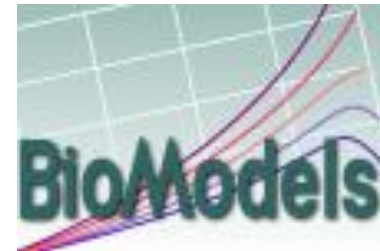
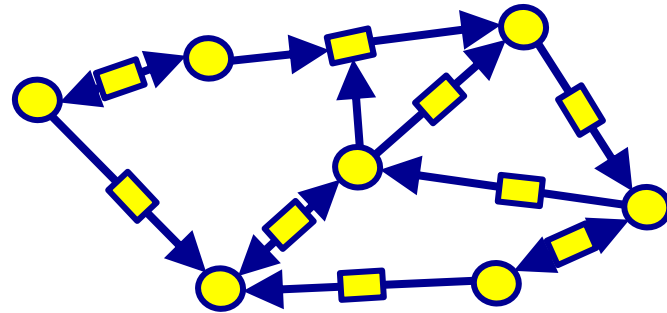
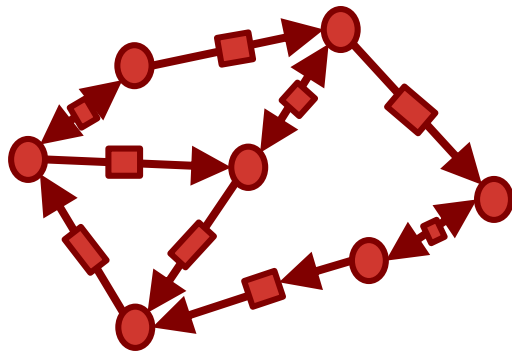


Merged model

Modeler may  
need to match  
these points

- ◆ SemGen does this (next talk)
- ◆ Never can be fully automatic

# Crossing the language barrier



# Annotation in the repositories



- ◆ No code-level annotation (until now)
- ◆ Recent work:  
25 fully annotated models
- ◆ SBML includes code-level annotation
- ◆ However...  
Standards alone are not enough

# Annotation in the repositories



- ◆ No code-level annotation (until now)
- ◆ Recent work:  
25 fully annotated models
- ◆ SBML includes code-level annotation
- ◆ However...  
Standards alone are not enough

*Policies and/or a culture of best practices for annotation*

# Problem 1: isVersionOf

Semantics of “isVersionOf” are weak:

- Similar to “is narrower than”.
- For merging, what do we do with this?

[http://co.mbine.org/standards/qualifiers: isVersionOf](http://co.mbine.org/standards/qualifiers:isVersionOf):

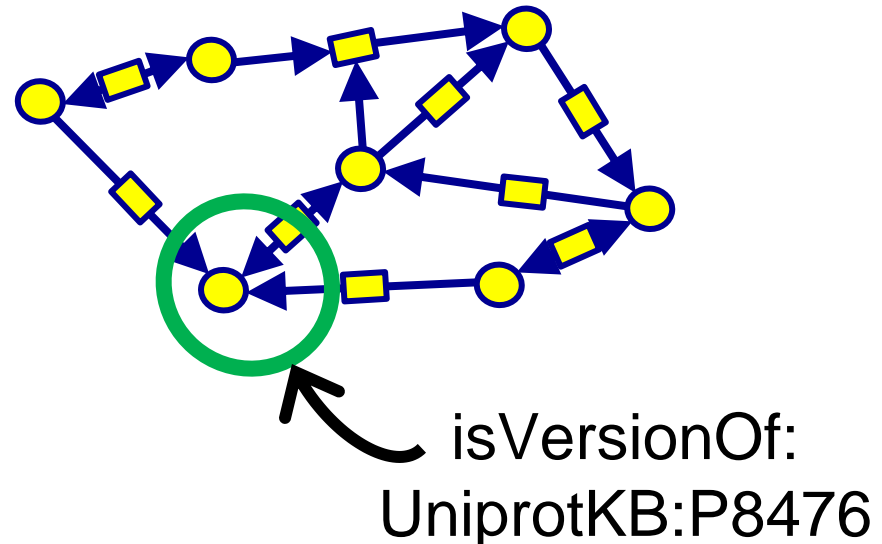
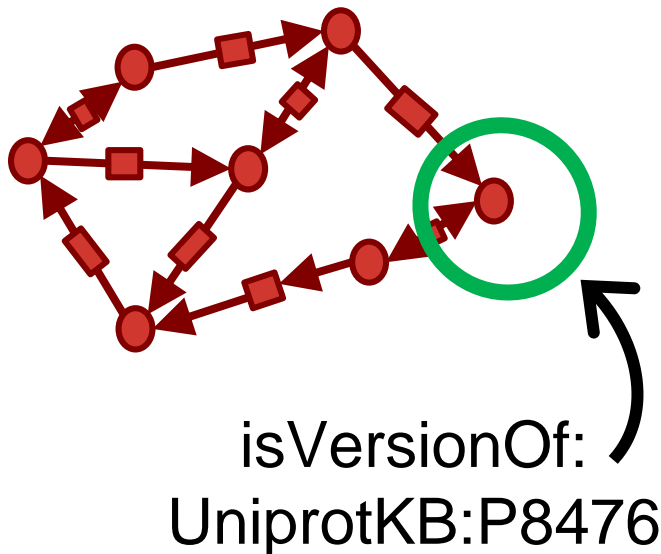
*The biological entity represented by the model element is a version or an instance of the subject of the referenced resource (biological entity B). This relation may be used to represent, for example, the 'superclass' or 'parent' form of a particular biological entity.*



# Problem 1: isVersionOf

Semantics of “isVersionOf” are weak:

- Similar to “is narrower than”.
- For merging, what do we do with this?



# Problem 2: Ontology choices

*A set of appropriate, compatible, orthogonal  
ontologies (credit OBO)*

Non orthogonal example: Kegg v. Uniprot  
(v. Reactome, v. InterPro, etc.)

# Problem 2: Ontology choices

*A set of appropriate, compatible, orthogonal ontologies (credit OBO)*

Non orthogonal example: Kegg v. Uniprot  
(v. Reactome, v. InterPro, etc.)

```
<bqbiol:isVersionOf>  
  <rdf:Bag>  
    <rdf:li rdf:resource="/uniprot/P35568"/>  
    <rdf:li rdf:resource="/kegg.compound/C00562"/>  
  </rdf:Bag>  
</bqbiol:isVersionOf>
```

# Problem 2: Ontology choices

*A set of appropriate, compatible, orthogonal  
ontologies (credit OBO)*

Non-appropriate (?) example: UniProt

# Problem 2: Ontology choices

*A set of appropriate, compatible, orthogonal ontologies (credit OBO)*

Non-appropriate (?) example: UniProt

- ◆ Protein Ontology entry: PRO #000016502  
“based on” three UniProt entries
  - Human: UniProtKB:P63316
  - Chicken: UniProtKB:P09860
  - Mouse: UniProtKB:P19123
- ◆ Also “isoform-1”, which has different uniprotIDs...

# Problem 3: Names

---

◆ What are things? What are processes?

*Sometimes, PHILOSOPHY really does matter*

# Problem 3: Names

## ◆ What are things? What are processes?

*Sometimes, PHILOSOPHY really does matter*

## ◆ Example: Ion channels

- A port?
- A flow of ions? A flow of electricity?
- A complex? (A set of proteins bound together)

*Wait.... How do we name a complex?*

# Semantics of Biological Processes

To share, understand, reuse, and re-purpose biosimulation models and modules, we must..

- ◆ Share a common grounding: Biology & Physics
- ◆ Annotate models using that grounding:  
Semantic annotation of biological processes
- ◆ Use tools that leverage these annotations and ontologies



# Ontology of Physics for Biology (OPB)

## Physical continuant (things)

- Physical entity
- Physical property
- Physical dependency

a heart, a portion of blood,  
a portion of chemical

solid mass, fluid volume,  
amount of chemical

Hooke's law, Ohm's law,  
law of mass action

## Physical processural entity (processes)

- Physical process

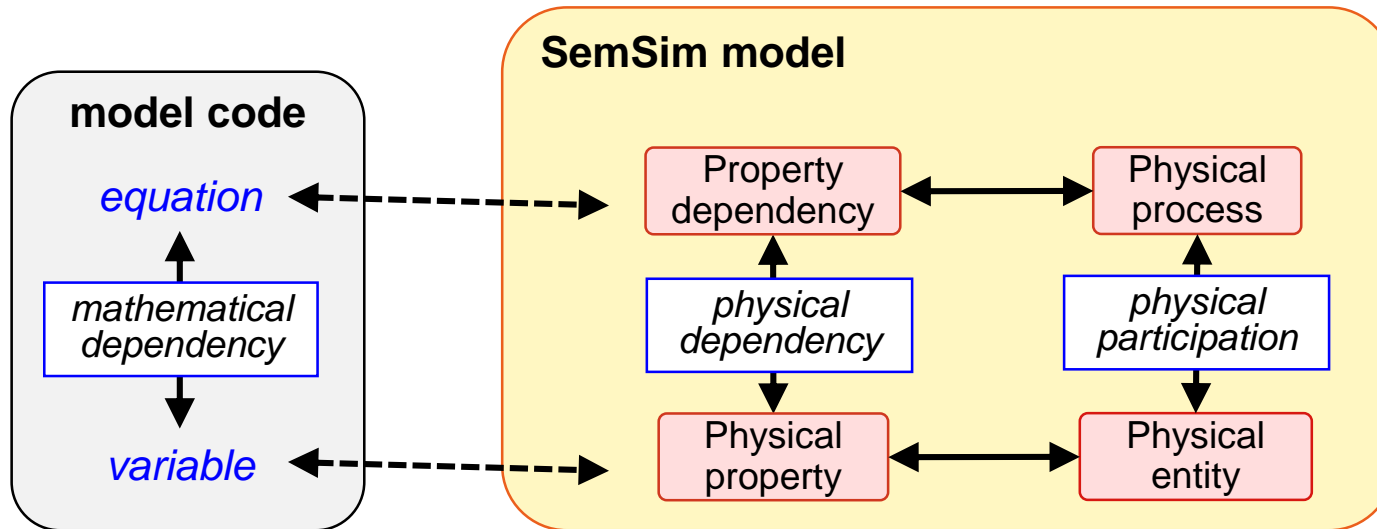
contracting, fluid flowing,  
chemicals reacting

# Past deeds / methods

- ◆ Merging Pandit with Hinch with Niederer:
  - Pandit: a model of cardiac electrophysiology
  - Hinch: a model of intracellular calcium dynamics
  - Niederer: a model of tension development
- ◆ How?
  - A “flat” approach to Module definitions (SAIM)
  - SemGen software
  - Composite annotations

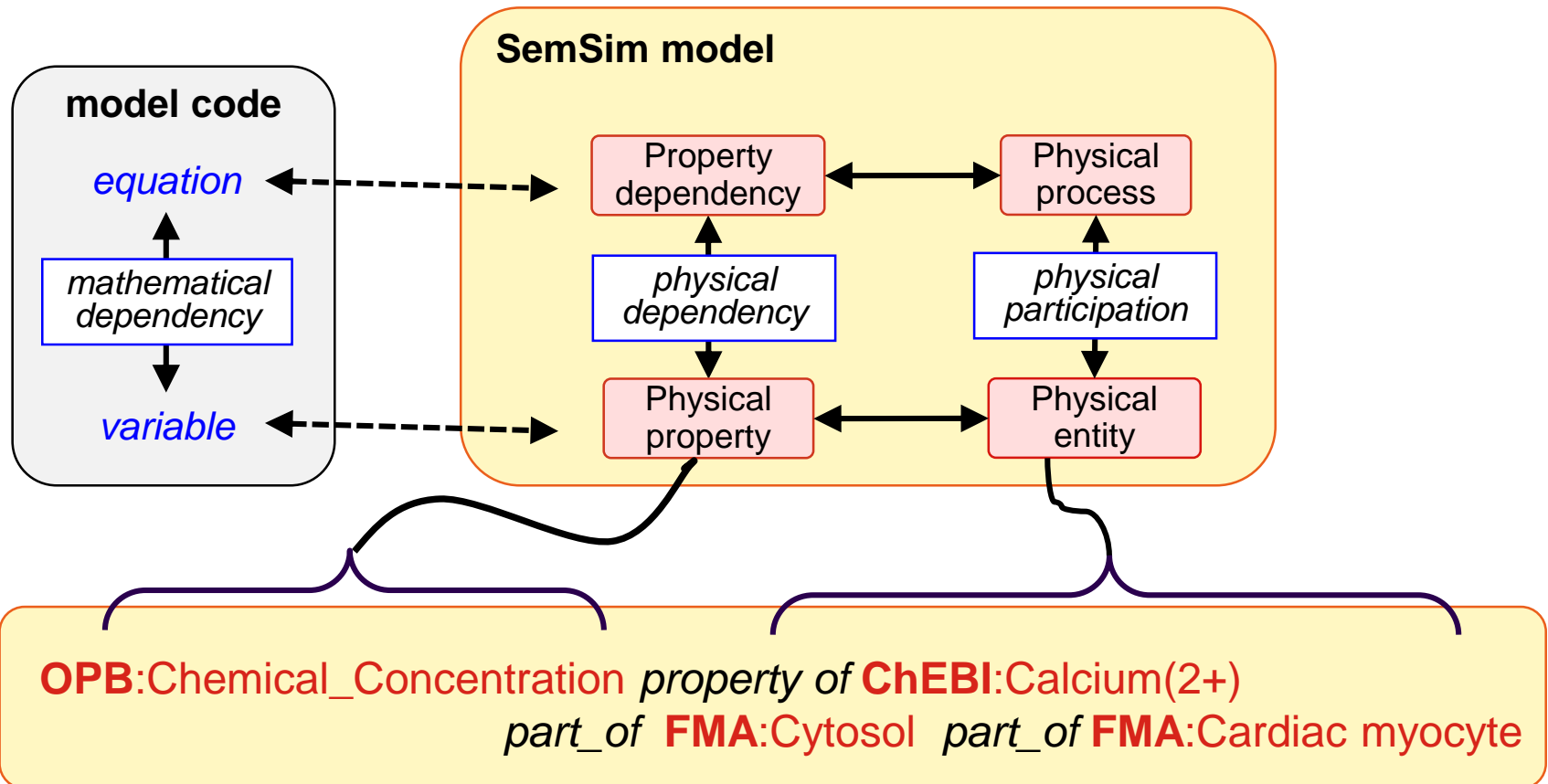
The **Chemical\_Concentration** of **Calcium(2+)**  
in the **cytosol** of the **Cardiac myocyte**

# Composite annotations



The **Chemical\_Concentration** of **Calcium(2+)**  
in the **cytosol** of the **Cardiac myocyte**

# Composite annotations



# Current results



With



Niederer:  
Intracellular tension

Abel 2011:  
Calcium signaling  
(bioMD#00355)

◆ Next talk!

# Thank you!

## Questions / Suggestions?

---

### Acknowledgements:

- Team: Brian Carlson, Dan Cook, Ryan James, Graham Kim, and Chris Thompson
- Funding: NIH R01LM011969

