

# Composite Annotations

## Say what you mean with biomedical ontologies

Daniel L. Cook  
John H. Gennari  
Maxwell L. Neal  
Michal Galdicki  
Onard L. Mejino

Biomedical and Health Informatics  
Physiology & Biophysics, and Biological Structure  
University of Washington, Seattle

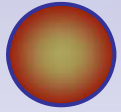
# Outline

- Overview and context
- Problem; solution requirements
- Current best practices
- Proposed solution: “composite annotations”
- Future challenges, future work

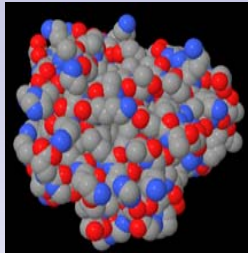
# Context

- **DARPA-Virtual Soldier Project (r.i.p.)**
  - integrate and encode patient-specific models (PSM) that predict outcomes of battlefield injuries
- **EU-Virtual Physiological Human (VPH; Feb '10)**
  - develop informatics infrastructure for managing biomedical knowledge required to develop PSMs for experimental and clinical investigation

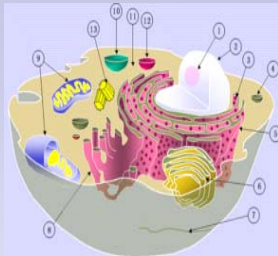
# Multiple structural scales — “multiscale”



> 100  
elements



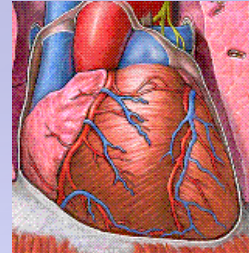
>> 100,000  
molecule types



>400  
cell-part types



>600  
cell types



63  
organ types

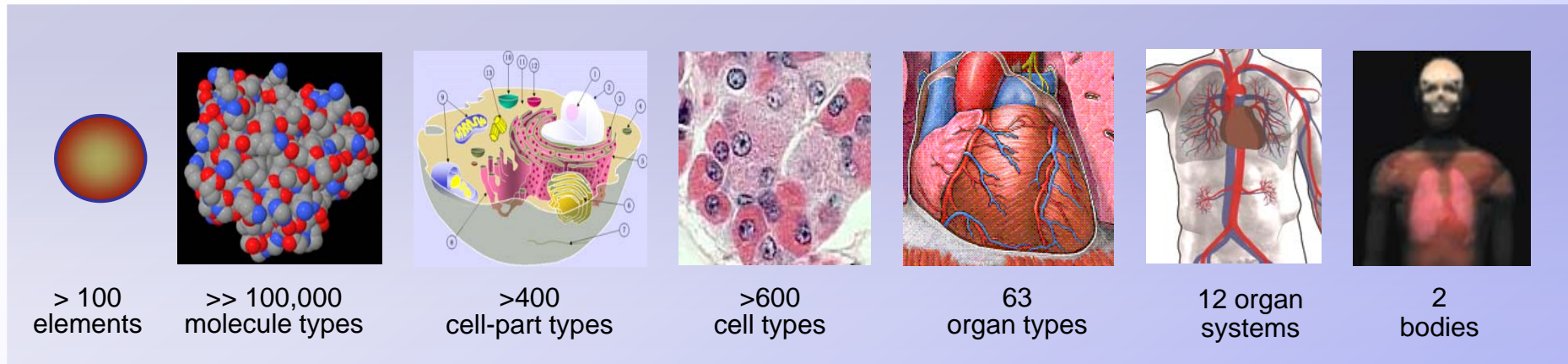


12 organ  
systems



2  
bodies

# Multiple structural scales — “multiscale”

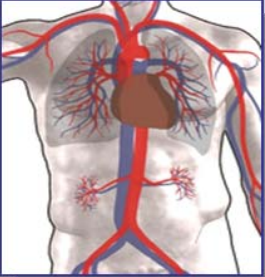


# Multiple process domains — “multidomain”

<u>Domain</u>	<u>Process</u>
fluids	axoplasm flowing, osmotic flowing...
solids	myocyte contracting, ...
chemical kinetics	metabolizing, gene expressing, cell signaling...
electrochemistry	ion pumping, action potential...
diffusion	intracellular calcium exchanging...
heat transfer	heat dissipating...

# Use-case: multidomain biosimulations & data

structural knowledge

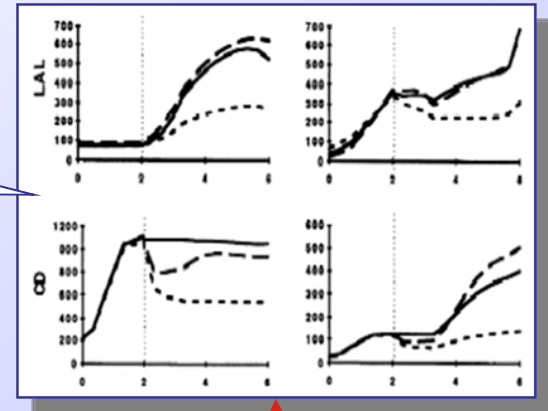


process knowledge

- fluids
- solids
- chemical kin
- electrochem
- diffusion
- heat transfer

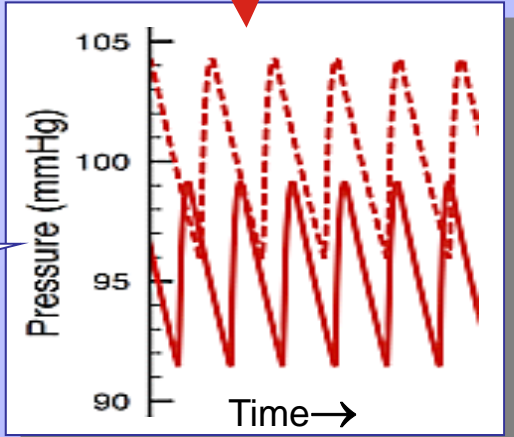


renal tubular model

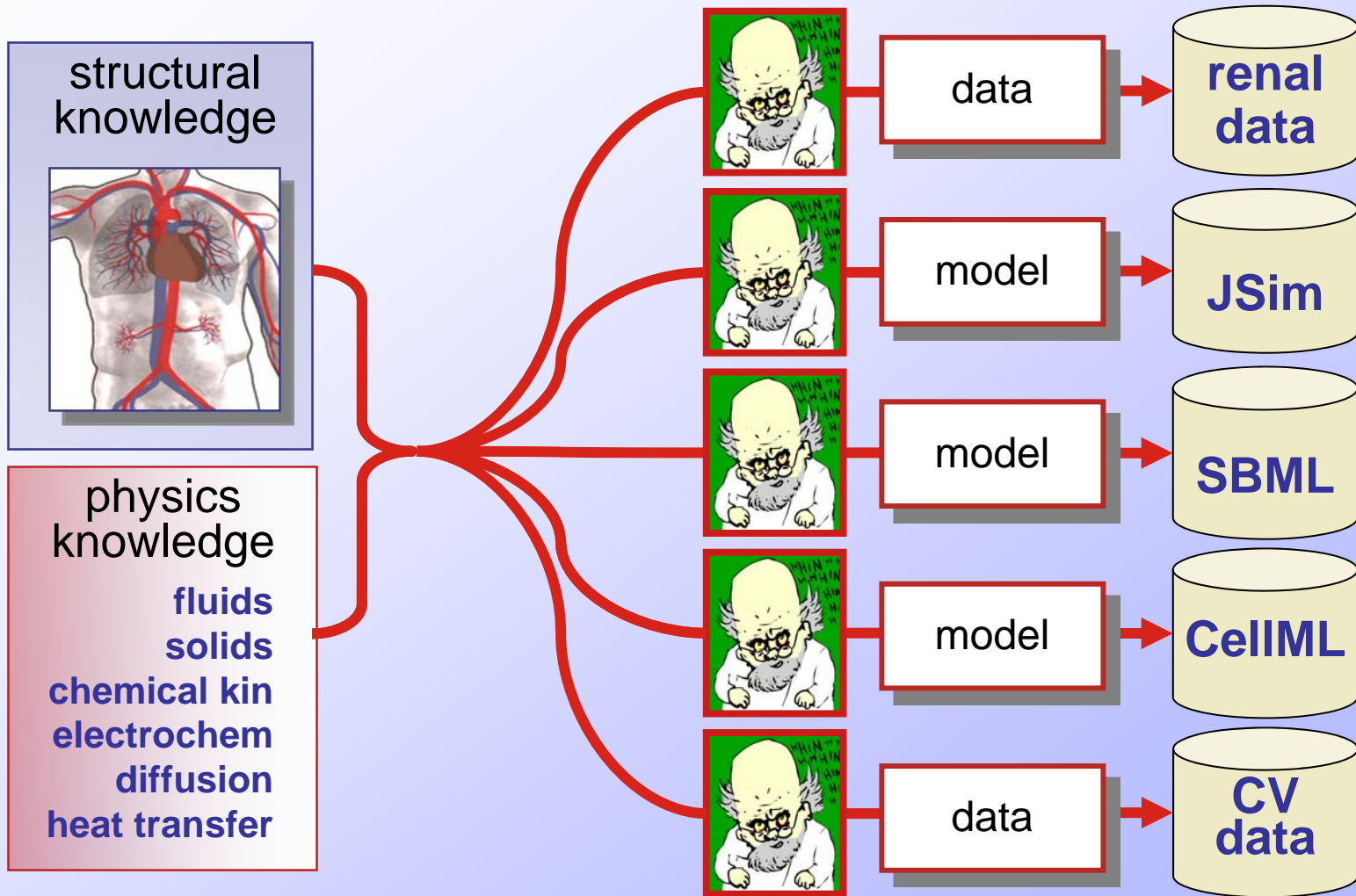


physics-based process biosimulation

cardio-vascular model



# Use-case: domain/language silos



# Problem — how to access biomedical resources

- Find models of calcium-calmodulin concentration dynamics in skeletal muscle.
- Find models that simulate aortic blood pressure.



# Solution — annotation requirements

- machine-readable
- independent of source language / schema
- preserve legacy resources
- specific — a single thing has a unique annotation
- general — multiscale and multidomain
- scaleable for large-scale, multicenter projects
- annotations are reusable

# Solution — state-of-art

- models: some human-readable, in-line annotation
- databases: no standards
- evolving in-line annotation standard: MIRIAM

# Solution — state-of-art example

- Find models of **calcium-calmodulin** concentration dynamics in skeletal muscle.

# Solution — state-of-art MIRIAM in SBML

- Find models of calcium-calmodulin concentration dynamics in skeletal muscle.

```
<species metaid="_120913" id="s277"
name="Ca_super_2_plus__endsuper_CaM"
compartment="c1" initialConcentration="0.3">
  <annotation>
    <rdf:Description rdf:about="#_120913">
      <bqbiol:hasVersion>
        <rdf:Bag>
          <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A29108"/>
          <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
          <rdf:li rdf:resource="urn:miriam:kegg.compound:C00076"/>
          <rdf:li rdf:resource="urn:miriam:kegg.compound:C00391"/>
        </rdf:Bag>
      </bqbiol:hasVersion>
    </rdf:Description>
  </annotation>
</species>
```

The diagram illustrates the mapping of biological entities to the SBML code. Callouts are shown as red-bordered boxes with lines pointing to specific parts of the code:

- A callout labeled **Ca<sup>2+</sup>-CaM** points to the `name="Ca_super_2_plus__endsuper_CaM"` attribute.
- A callout labeled **Ca<sup>2+</sup>** points to the `<rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A29108"/>` line.
- A callout labeled **CaM** points to the `<rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>` line.
- A callout labeled **CaM** points to the `<rdf:li rdf:resource="urn:miriam:kegg.compound:C00391"/>` line.

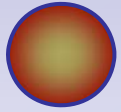
# Problem — annotation examples

- Find models of calcium-calmodulin concentration dynamics in skeletal muscle.
- Find models that simulate aortic blood pressure.

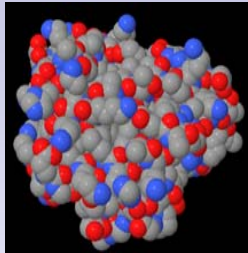


physical  
entities

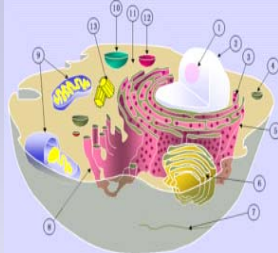
# Biomedical ontologies for annotations



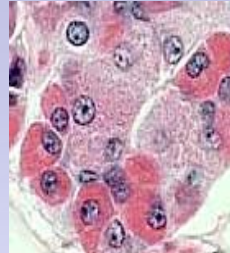
> 100  
elements



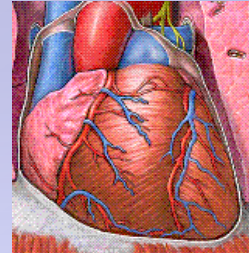
>> 100,000  
molecule types



>400  
cell-part types



>600  
cell types



63  
organ types



12 organ  
systems



2  
bodies

KEGG

Foundational Model of Anatomy (FMA)

UniProt

Cell Type

ChEBI

Gene Ontology

*FMA: Foundational Model of Anatomy*  
*CL: Cell Type Ontology*  
*GO: Gene Ontology*  
*ChEBI: Chemical Entities of Biological Interest*  
*KEGG: Kyoto Encyclopedia of Genes and Genomes*  
*UniProt: Unified Protein Resource (EBI)*

# Problem— annotation examples

- Find models of calcium-calmodulin concentration dynamics in skeletal muscle.
- Find models that simulate aortic blood pressure.

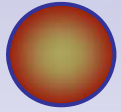


physical  
entities

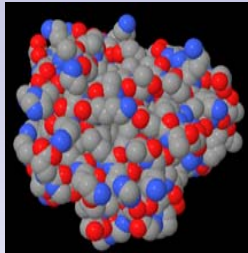


physical  
properties

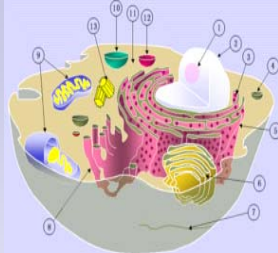
# Biomedical ontologies for annotations



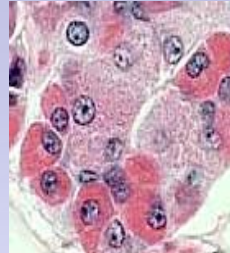
> 100  
elements



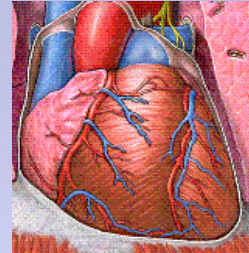
>> 100,000  
molecule types



>400  
cell-part types



>600  
cell types



63  
organ types



12 organ  
systems



2  
bodies

KEGG

Foundational Model of Anatomy (FMA)

UniProt

Cell Type

ChEBI

Gene Ontology

Ontology  
of Physics  
for Biology  
(OPB)

## Domain

**fluids**

**solids**

**chemical kinetics**

**electrochemistry**

**diffusion**

**heat transfer**

## Physical property

volume flow rate, fluid pressure

tensile force, shear strain...

concentration, chemical flow rate...

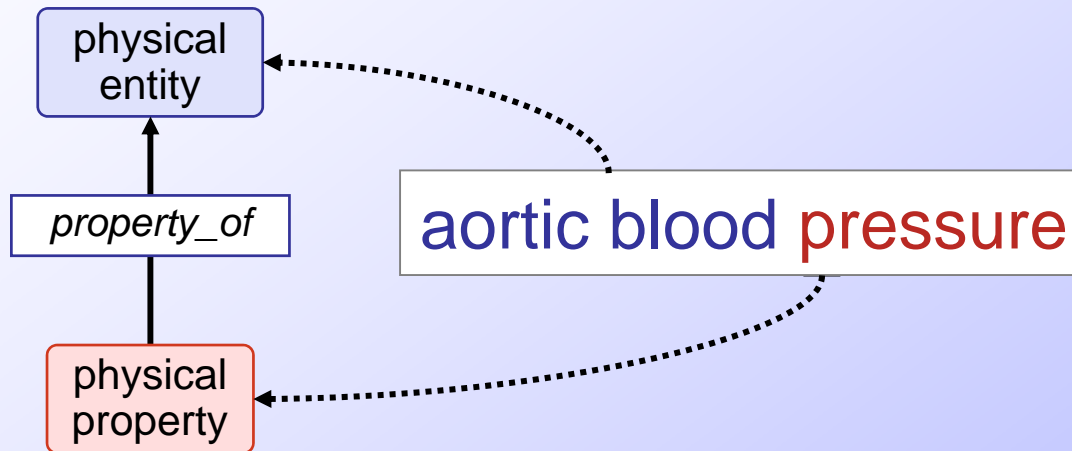
electrical potential, ionic current...

particle concentration, chemical potential

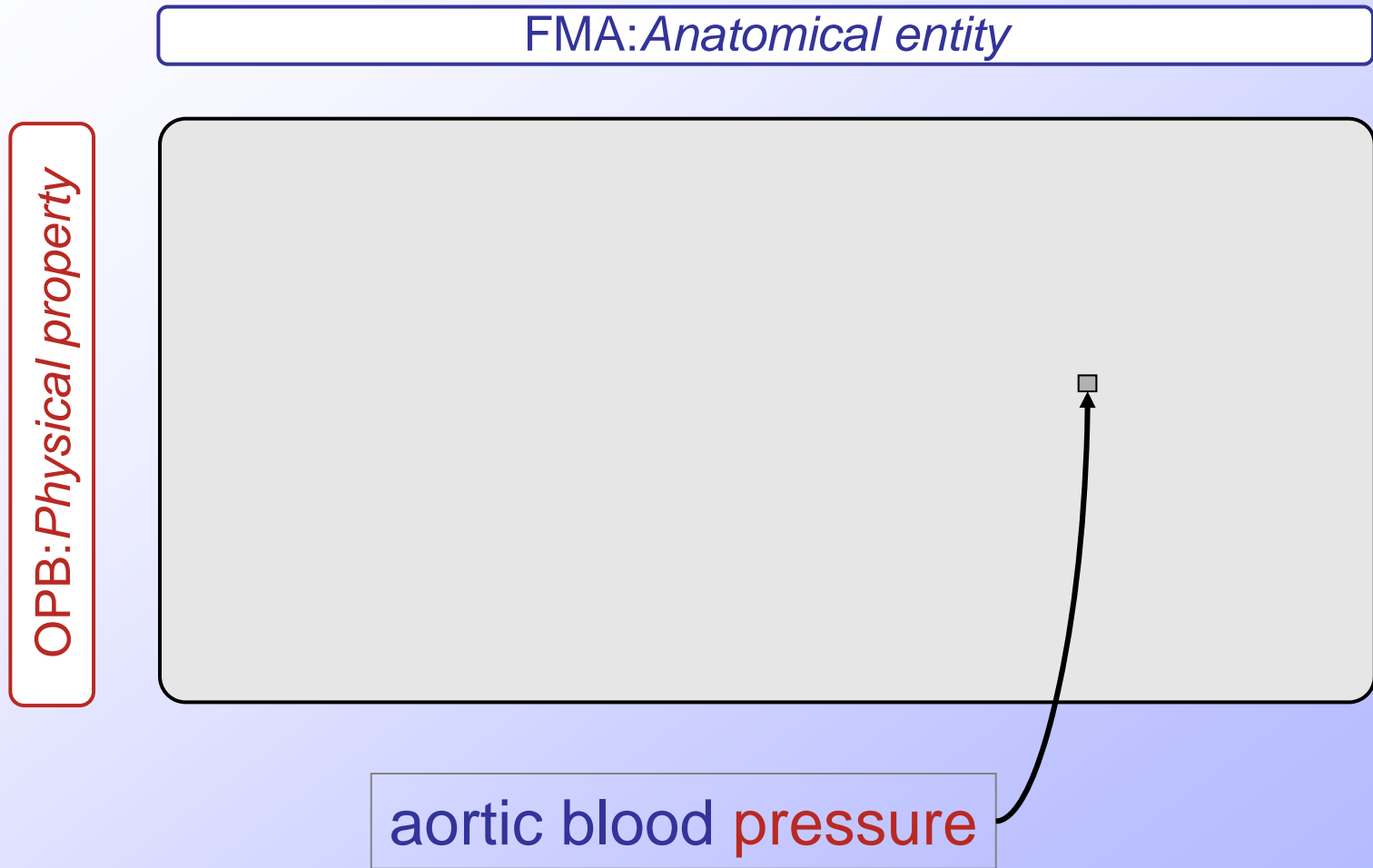
heat flow rate, temperature...



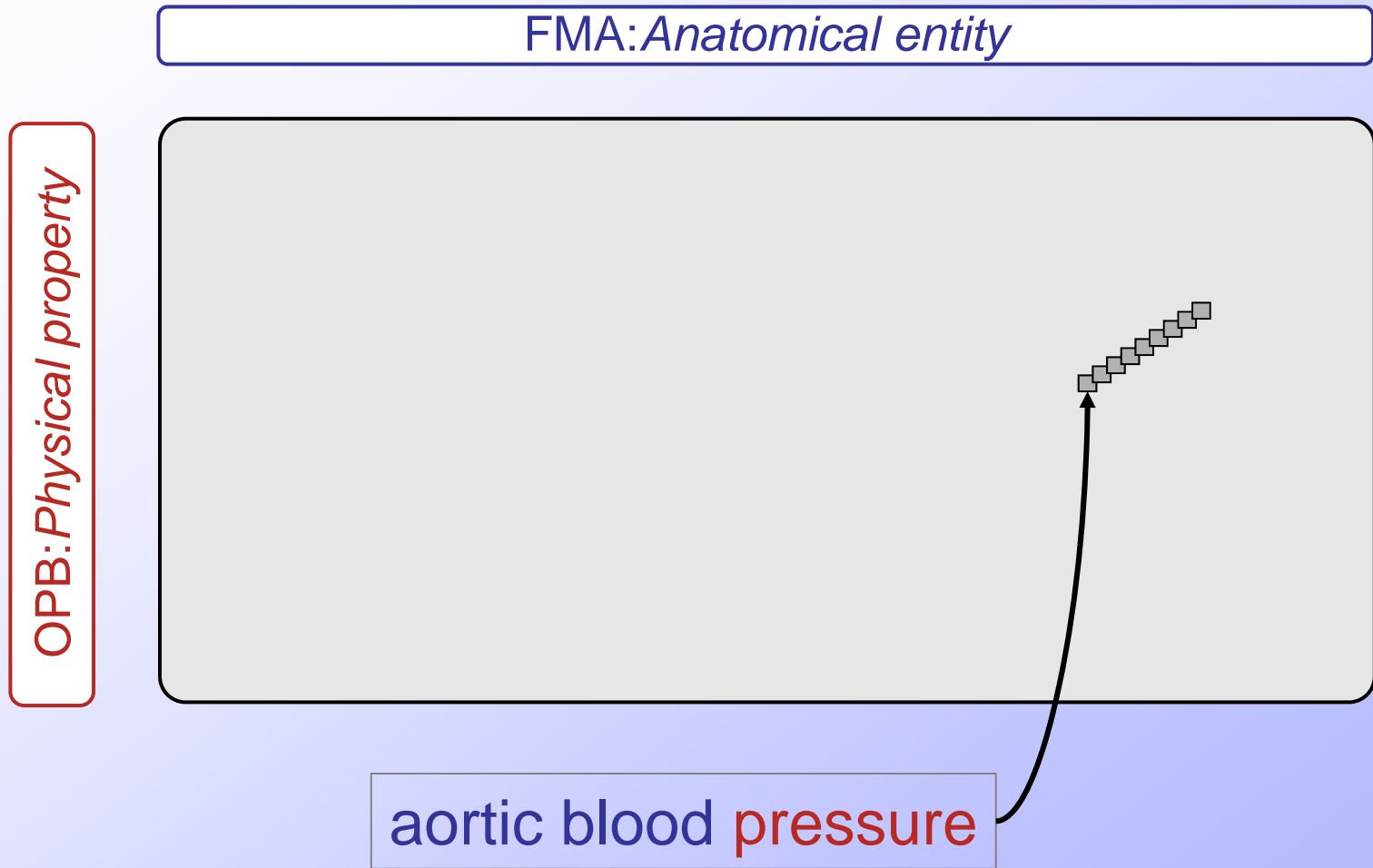
# Solution: precoordinate classes



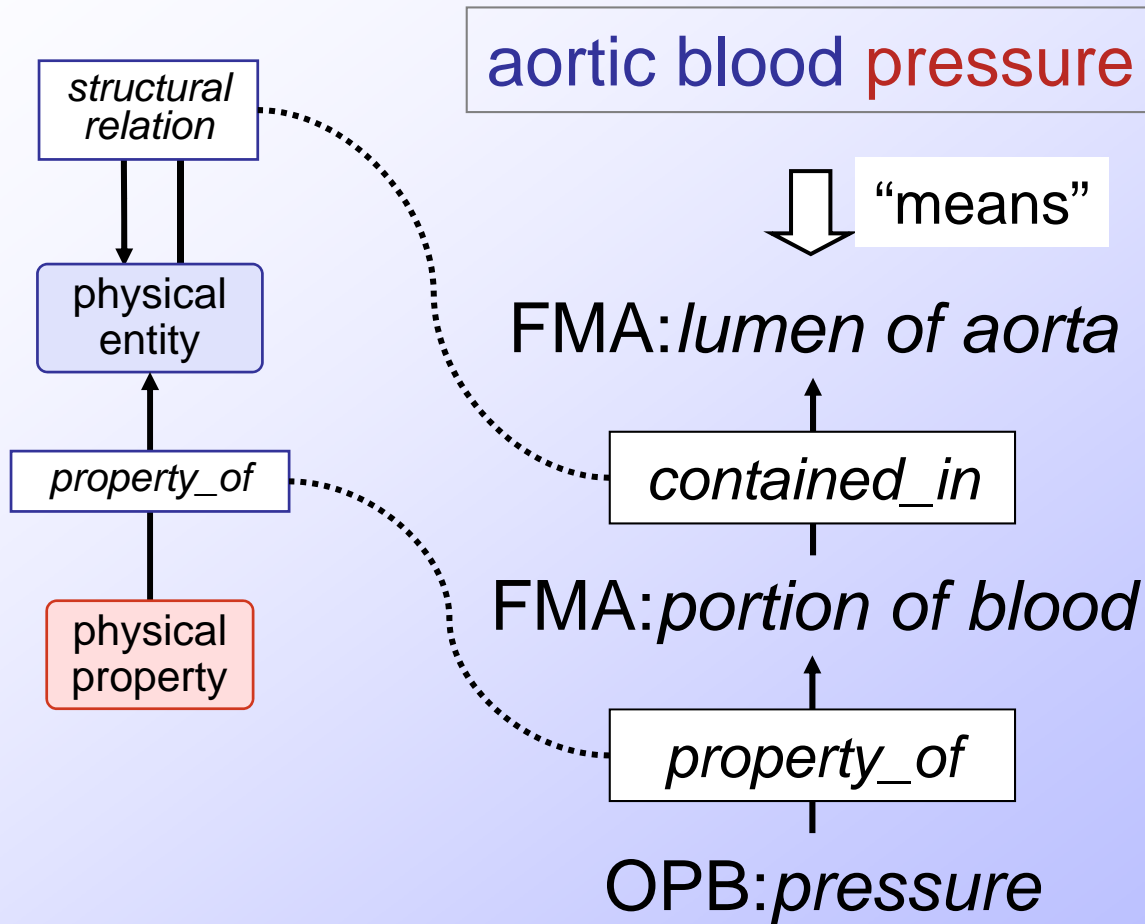
# Solution: generate ontology cross-products



# Problems: combinatorics, irrelevant classes...



# Solution: Composite annotation schema



# Composite annotations in SemGen

The screenshot shows the SemGen application window. The top menu bar includes File, Edit, Annotate, Merge, Extract, Encode, View, and Help. Below the menu is a toolbar with a CV icon and a 'Close tab' button. The main text area contains the following code:

```
// Equations for systemic circulation
Paorta = ((VSysArtCaps-VrestSysArtCaps)/CSysArtCaps); // Law of elastance
FSysArtCaps = (Paorta-PSysVeins)/Rartcap; // Ohm's Law
PSysVeins = ((VSysVeins-VrestSysVeins)/CSysVeins); // Law of elastance
FSysVeins = (PSysVeins-PRA)/RsysVeins; // Ohm's Law

// Equations for pulmonary circulation
PPulArtCaps = ((VPulArtCaps-VrestPulArtCaps)/CPulArtCaps); // Law of elastance
```

Below the code is a table of variables and their properties:

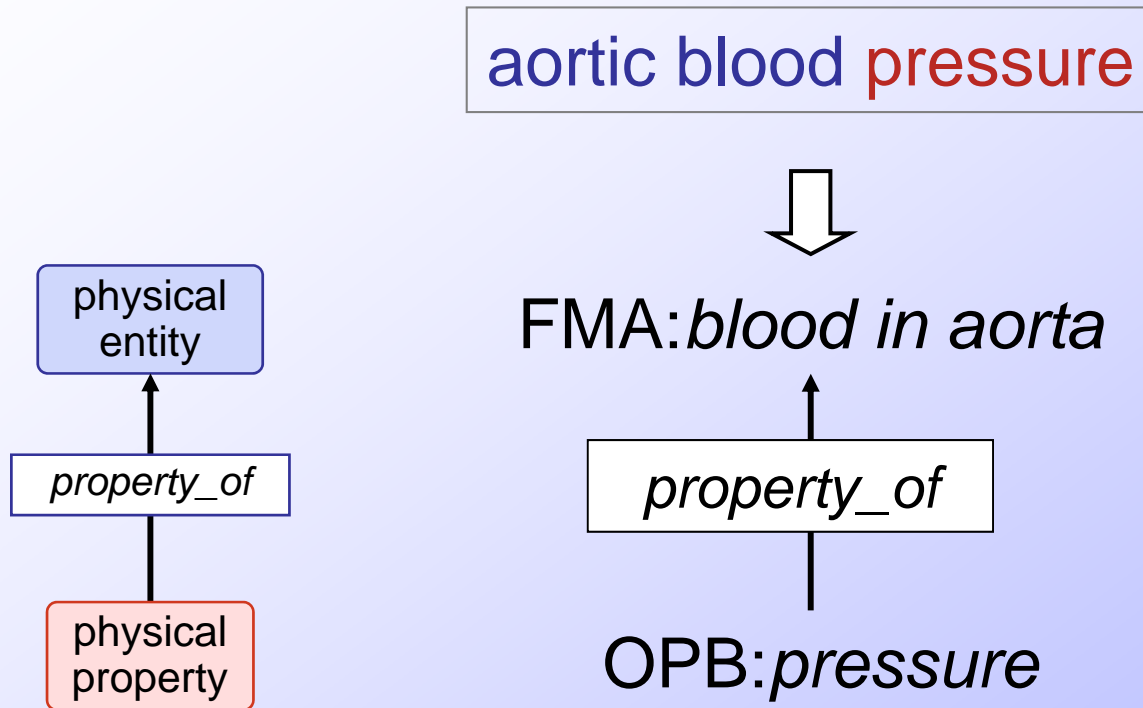
FSysVeins	P+E	_
FTricuspidValve	P+E	---
HR	-	N
Paorta	P+E	---
PLA	P+E	---
PLV	P+E	---
PPulArtCaps	P+E	---
PPulVeins	P+E	---

On the right side of the table, there is a detailed view for the selected 'Paorta' variable:

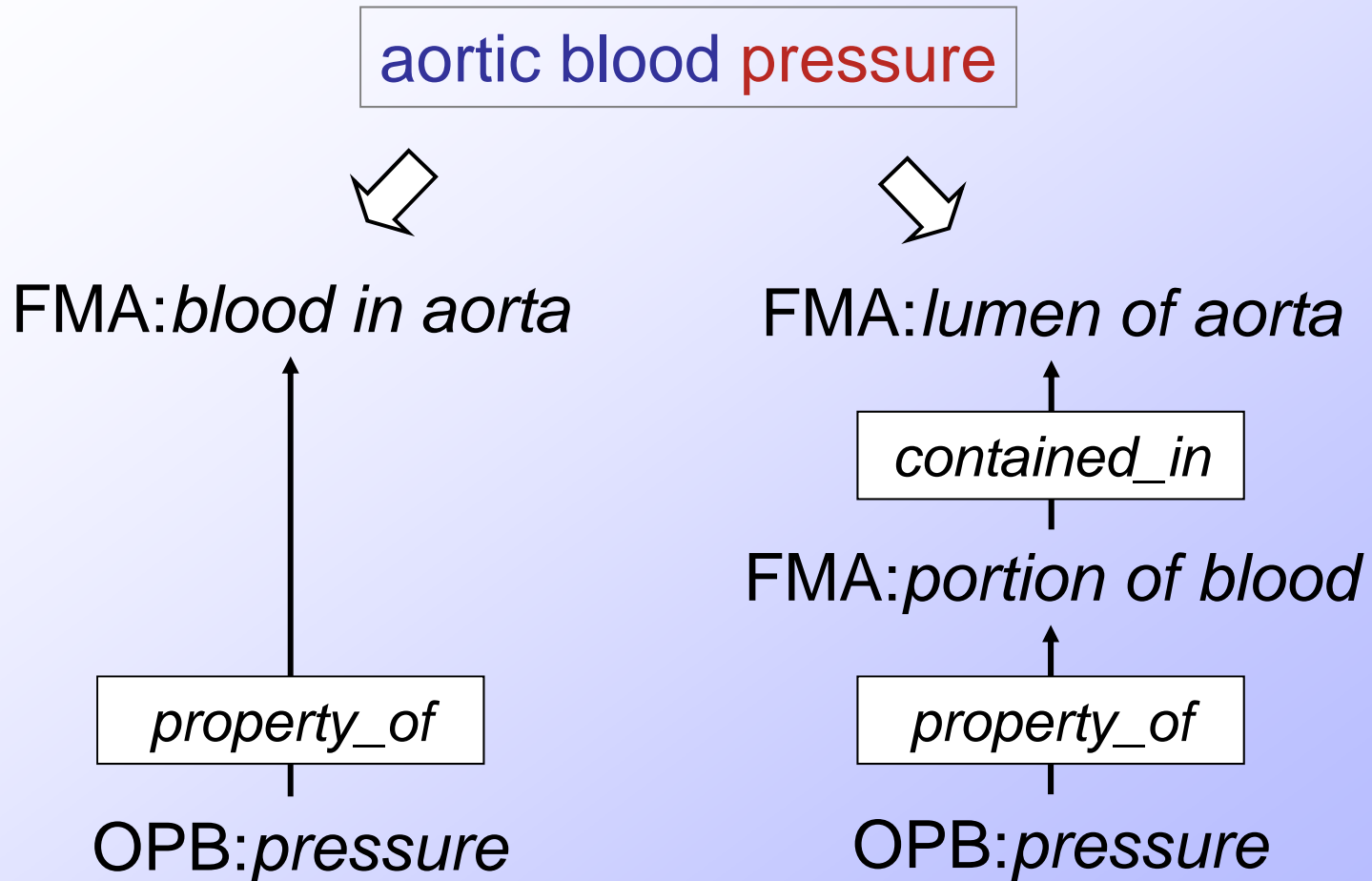
- Composite annotation** (with an 'edit' button): Fluid pressure <propertyOf> Portion of blood <contained in> Lumen of aorta
- Non-composite annotation** (with an 'edit' button and a close 'X' button): [unspecified]
- Human-readable definition** (with an 'edit' button and a close 'X' button): [unspecified]

Red boxes highlight the 'Paorta' variable in the code, the 'Paorta' row in the table, and the 'Composite annotation' box. A red arrow labeled 'means' points from the 'Paorta' in the code to the 'Paorta' in the table, and another red arrow points from the 'Paorta' in the table to the 'Composite annotation' box.

# Solution: post-coordinated composite annotation



# Challenge: non-unique composite annotations



# Problem: complicated annotation example

OPB

ChEBI

must be explicit that  $\text{Ca}^{2+}$  is in  
FMA:cytosol of endoplasmic  
reticulum

**Concentration** of calcium ions in the endoplasmic  
reticulum of a vascular smooth muscle cell in the  
wall of a systemic arteriole.



# Problem: complicated annotation example

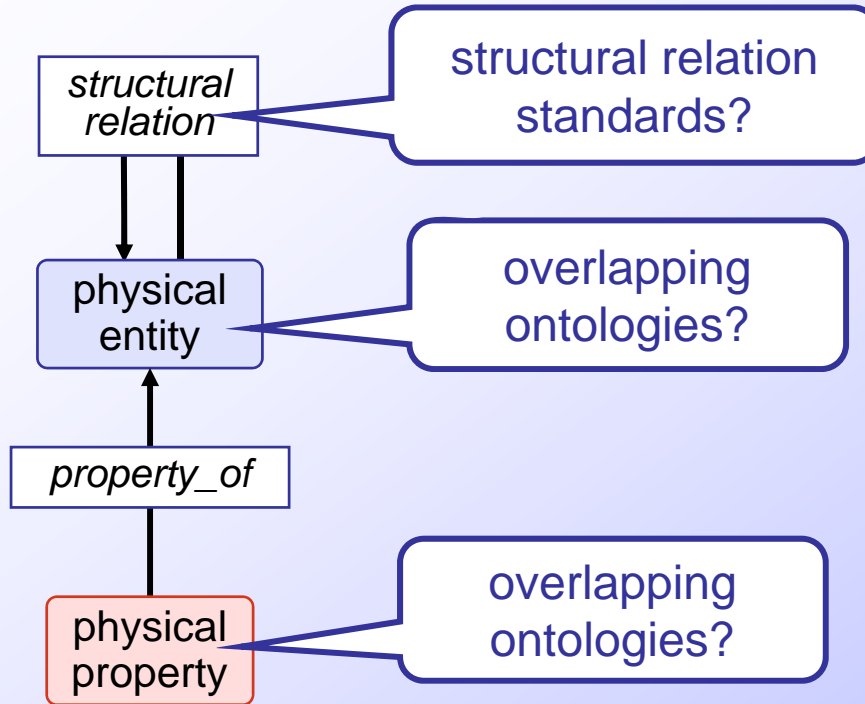
OPB

FMA: *portion  
of blood*

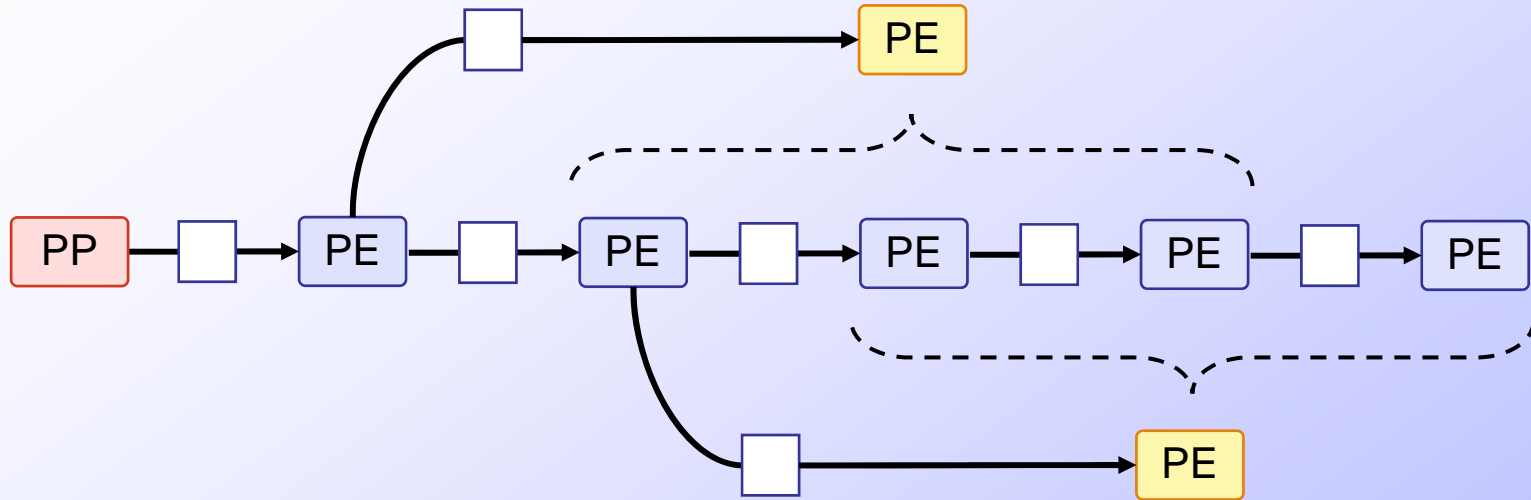
Fluid pressure of blood in atrial septal defect

a pathological entity not in FMA  
(or any other ontology)

# Challenges for composite annotations



# Challenges for reusing composite annotations



# Outline

- Overview and context
- Problem; solution requirements
- Current best practices
- Proposed solution: “composite annotations”
- Future challenges, future work