# Harmonizing semantic annotations for
# computational models in biology

*John Gennari*

Semantic Annotations:
Why? What?

Recommendations

Technical details & Next steps

# Semantic Annotations:
# ~~Why?~~ What?

# Recommendations

# Technical details & Next steps

# Semantic Annotations

- Model-level annotation
  - Describing the whole model
  - Author, date, publication, overview, etc.
- Code-level annotation
  - Refers to standards / ontologies
  - Describes individual species, (proteins, chemicals, anatomic entities), reactions, and parameters & variables

# Semantic Annotations

- Model-level annotation
  - Describing the whole model
  - Author, date, publication, overview, etc.
- Code-level annotation
  - Refers to standards / ontologies
  - Describes individual species, (proteins, chemicals, anatomic entities), reactions, and parameters & variables

# Examples

- Left ventricular volume
  - FMA term
  - Uberon term
  - OPB term

- Intracellular calcium+2 concentration
  - Chebi ID
  - Kegg ID

# Recommendations (3 of 7)

- Store annotations in a separate file
  - Connect to code file via COMBINE archives
  - Allows for cross language annotations
  - Allows for different annotations for one model
- Use RDF
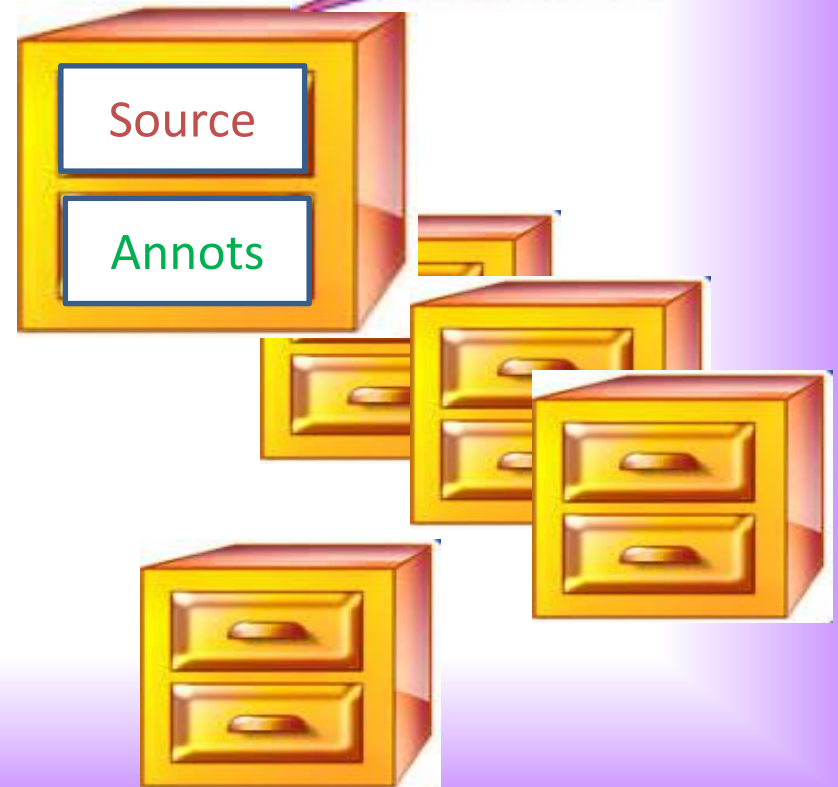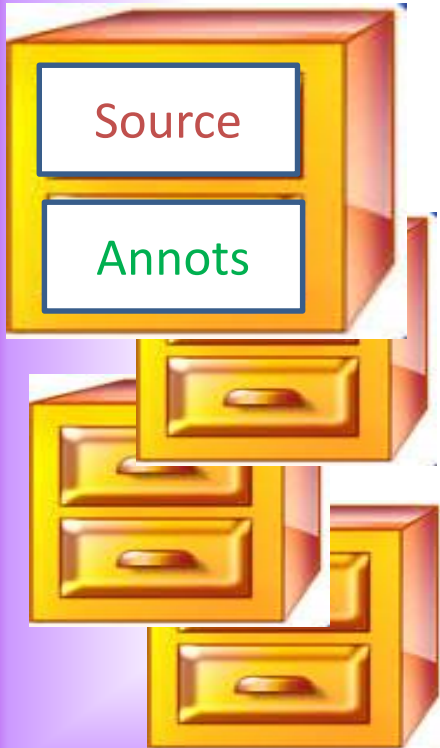- Develop software library for working with annotations
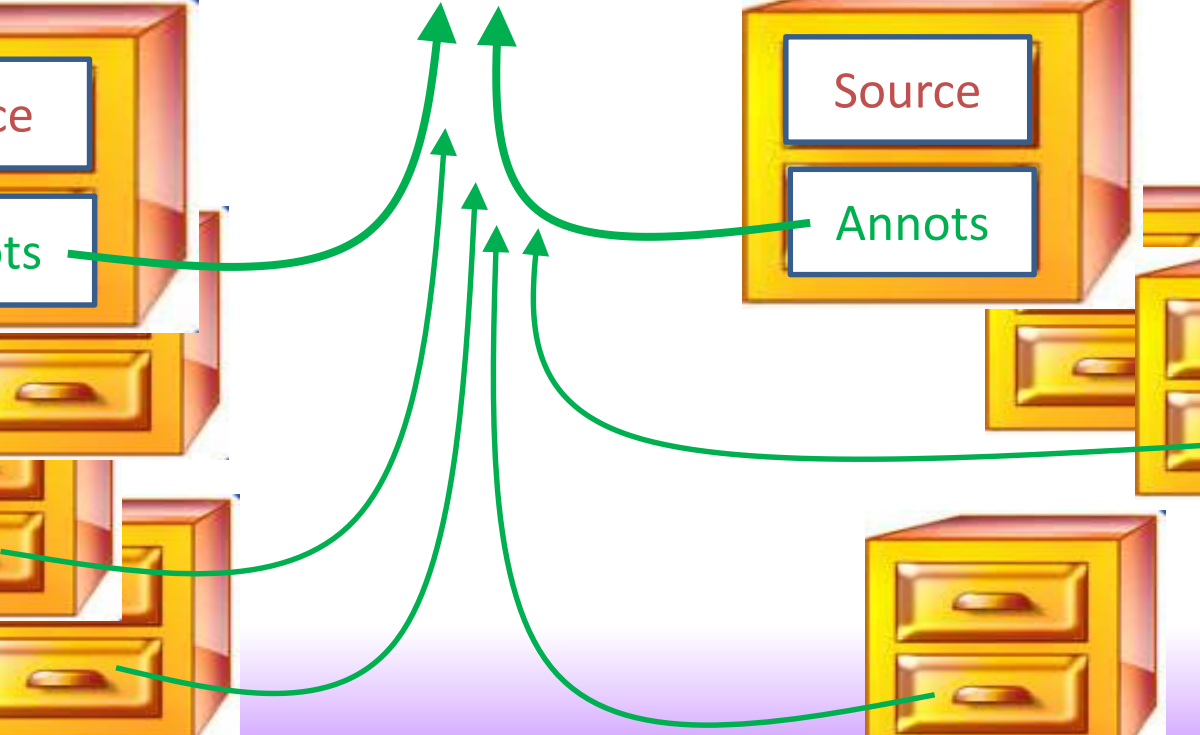
# Combine Archive
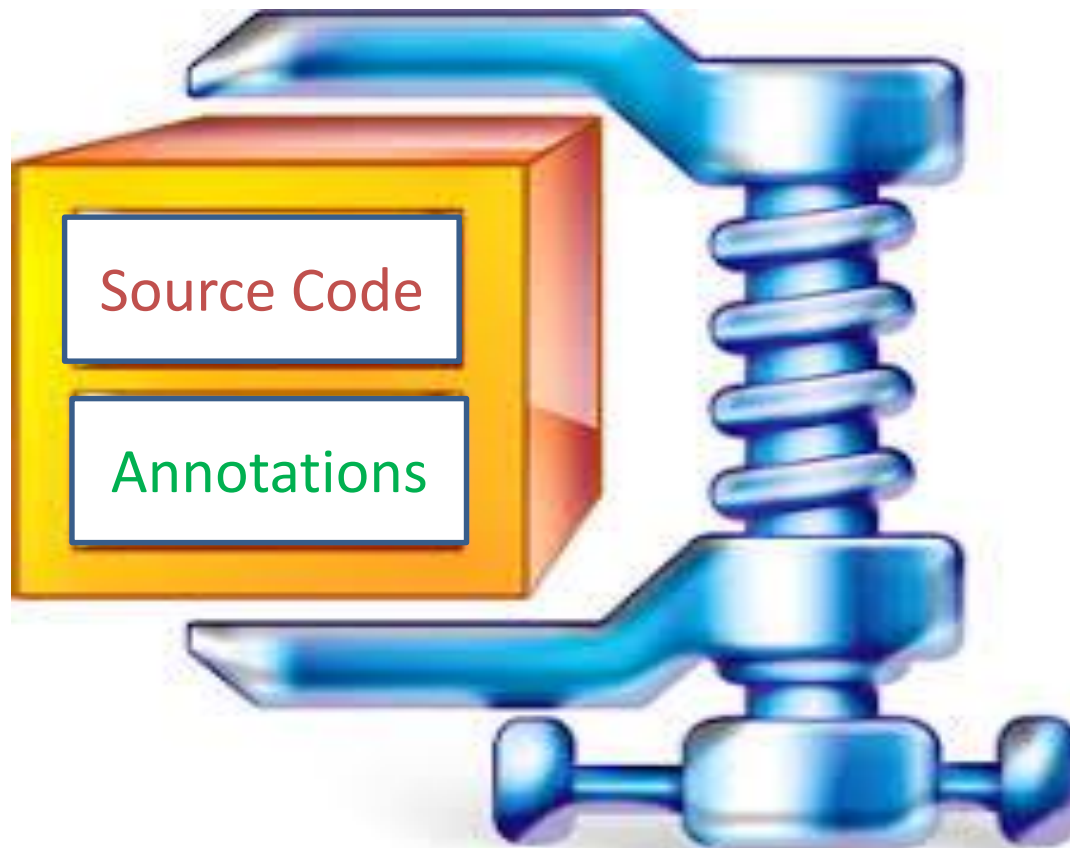
Semantic search
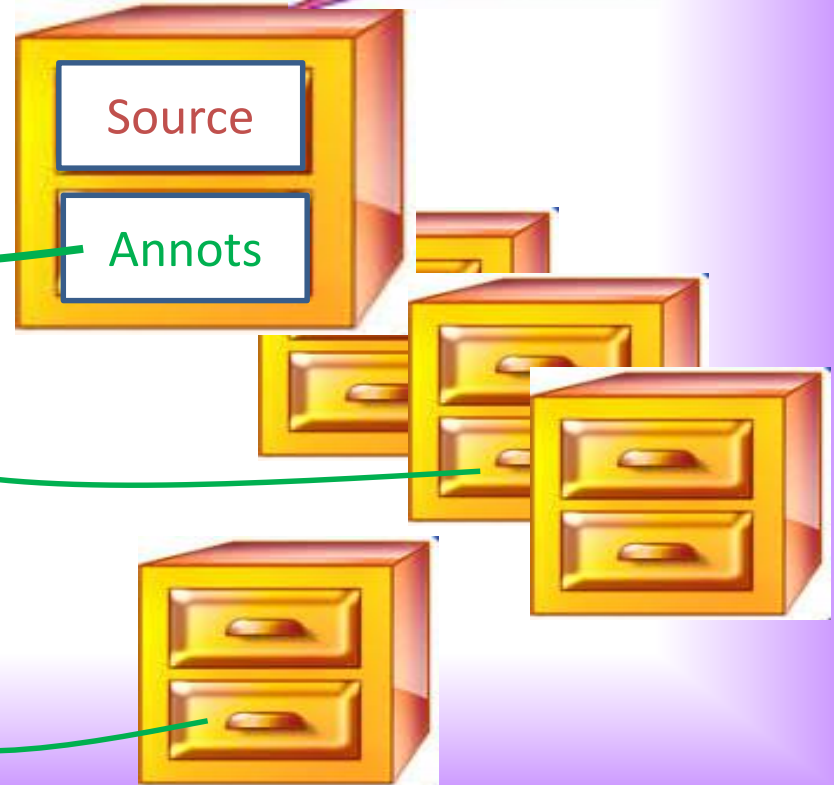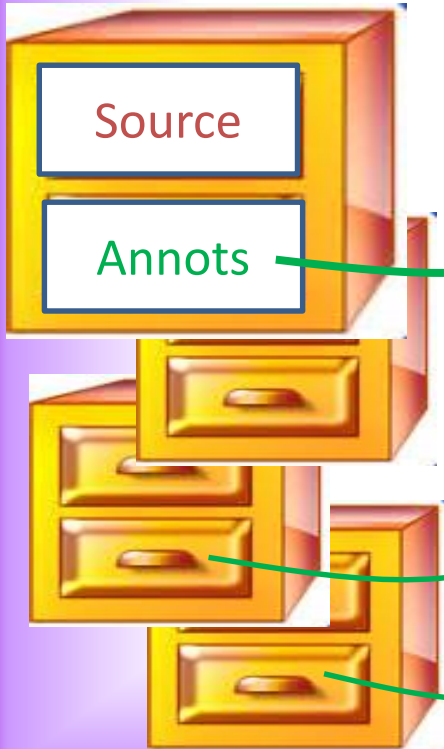
# Details: OMEX

- OMEX Combine archive:
  - Headers: manifest.xml, metatdata.rdf
  - Text files—e.g. Model source, Annotation file

# Details: OMEX

- OMEX Combine archive:
  - Headers: manifest.xml, metatdata.rdf
  - Text files—e.g. Model source, Annotation file

```
<?xml version="1.0" encoding="UTF-8"?>
<omexManifest xmlns="http://identifiers.org/combine.specifications/omex-manifest">
  <content location="." format="http://identifiers.org/combine.specifications/omex" />
  <content location="./manifest.xml" format="http://identifiers.org/combine.specifications/omex-manifest" />
  <content location="./metadata.rdf" format="http://identifiers.org/combine.specifications/omex-metadata" />
  <content location="./README.md" format="http://purl.org/NET/mediatypes/text/markdown" />
  <content location="./BIOMD0000000176.xml"
            format="http://identifiers.org/combine.specifications/sbml.level-2.version-1" />
  <content location="./BIOMD0000000176_example_semantics.rdf"
            format="http://identifiers.org/combine.specifications/omex-metadata" />
</omexManifest>
```

# Details: Annotation example

- D-glucopyranose & glucose transport;
 An RDF annotation:

@prefix bq: <http://biomodels.net/biology-qualifiers/> .
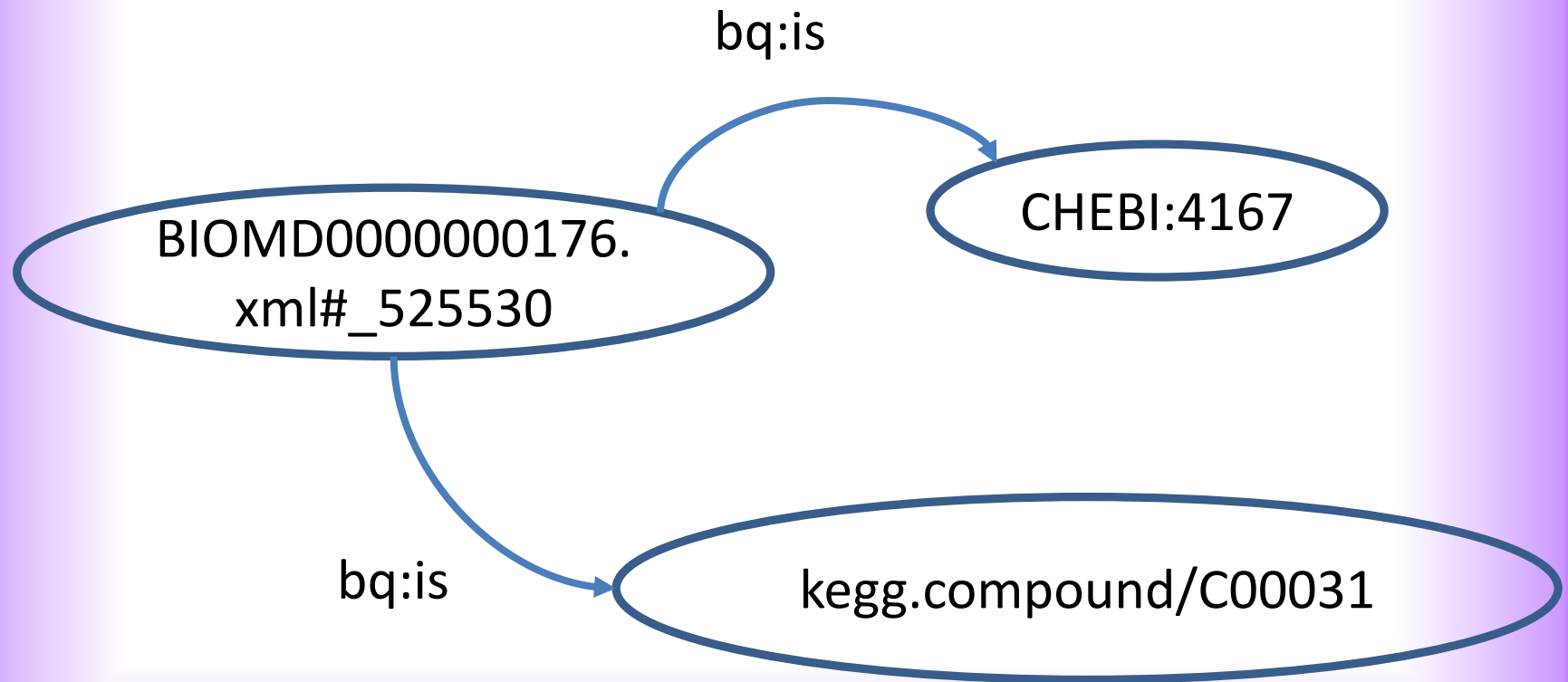
<file:///BIOMD0000000176.xml#_525530>  bq:is
        <http://identifiers.org/chebi/CHEBI:4167>,
        <http://identifiers.org/kegg.compound/C00031> .
<file:///BIOMD0000000176.xml#_525635>  bq:isVersionOf
        <http://identifiers.org/go/GO:0015758> .

# Details: Annotation example

- D-glucopyranose

# Details: Annotation example

- D-glucopyranose & glucose transport:

  @prefix bq: <http://biomodels.net/biology-qualifiers/> .

  <file:///BIOMD0000000176.xml#_525530>  bq:is
         <http://identifiers.org/chebi/CHEBI:4167>,
         <http://identifiers.org/kegg.compound/C00031> .
  <file:///BIOMD0000000176.xml#_525635>  bq:isVersionOf
         <http://identifiers.org/go/GO:0015758> .

# Details: Linking to source

- Search in source code for _525530:

```
<species compartment="cyto" id="GLCi
        initialConcentration="0.097652231064563"
        metaid="_525530">
<annotation>
        …
```
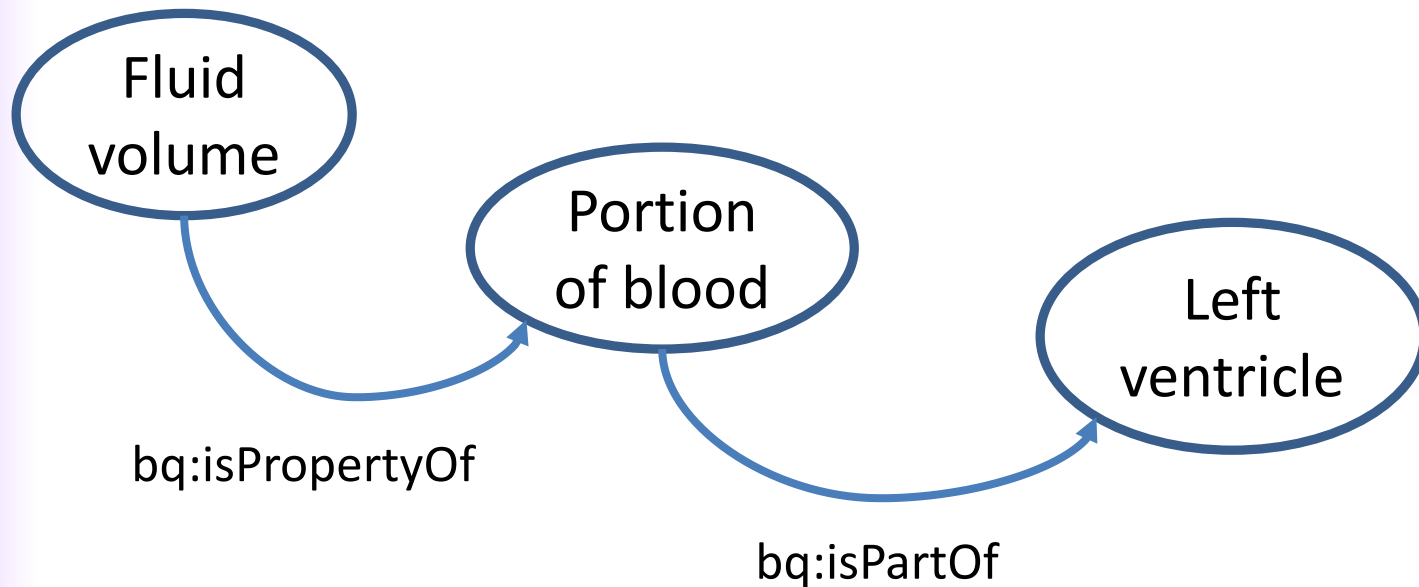
# Details: Example 2

- Left ventricular blood volume:

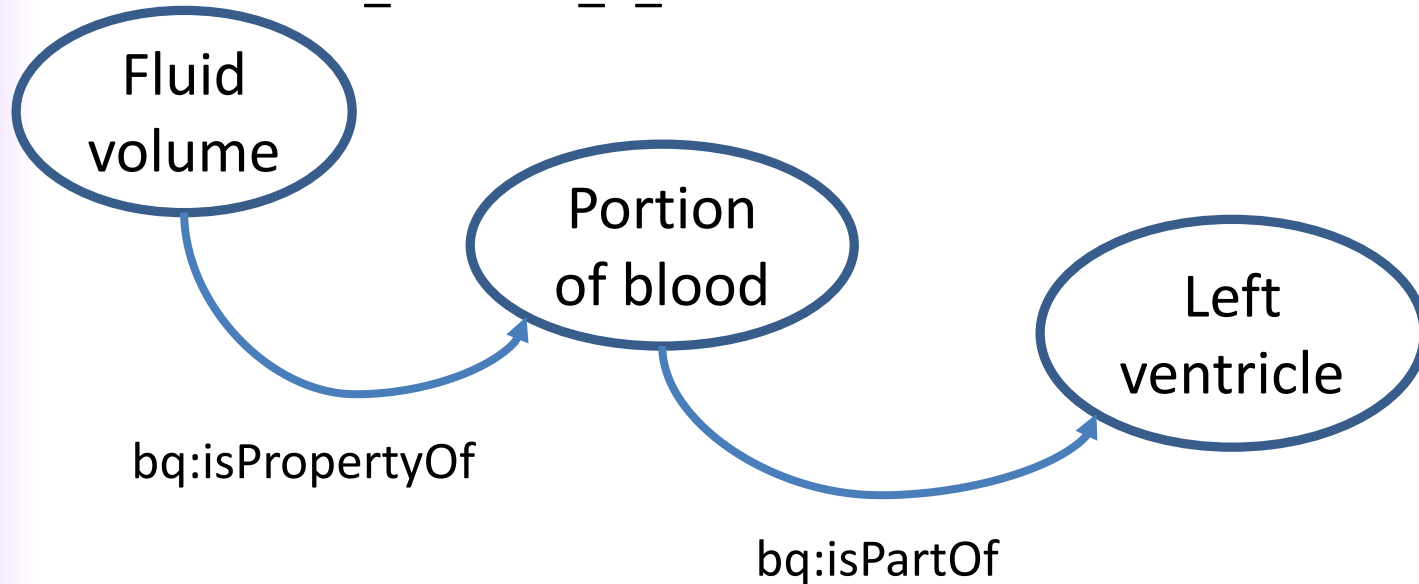  (a composite annotation)

# Details: Example 2

- Left ventricular blood volume:

(a composite annotation)

# Details: Example 2

- Left ventricular blood volume:

  (a composite annotation)

Smith2004.cellml#left_ventricle_V_lv

Fluid volume

Portion of blood

Left ventricle

bq:isPropertyOf

bq:isPartOf

# Details: Example 2

- Left ventricular blood volume:

@prefix bq: <http://biomodels.net/biology-qualifiers/> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix annot: <http://PKB.org/Smith2004_semantics.rdf#>

<file:///Smith2004.cellml#left_ventricle_V_lv>
        bq:isVersionOf <http://identifiers.org/opb/OPB_00154> ;
        bq:isPropertyOf annot:entity_0;
        dc:description "Left ventricular blood volume" .
annot:entity_0 bq:is <http://identifiers.org/fma/FMA:9670>;
        bq:isPartOf local:entity_1;
        dc:description "portion of blood" .
annot:entity_1  bq:is
        <http://identifiers.org/fma/FMA:9466>;
        dc:description "cavity of left ventricle".

# Details: Linking to source

- Search in source code for left_ventricle_V_lv:

```
<component name="left_ventricle">
    <variable cmeta:id="left_ventricle_P_lv"
        name="P_lv" public_interface="out" units="kPa"/>
    <variable cmeta:id="left_ventricle_V_lv"
        initial_value="94.6812" name="V_lv"
        public_interface="out" units="mL"/>
....
```

# …what next?

- Curatorial work

  - Create lots of examples—show the value!

  - Synergy with  CENTER FOR REPRODUCIBLE BIOMEDICAL MODELING

- Develop software for these annotations

  - The OMEX Metadata specification

  - Build a library for tool developers

- Use case development & tool integration

# Questions?

# Details: OMEX

- OMEX archive:
  - Headers: manifest.xml, metatdata.rdf
  - Text files—e.g. Model source, Annotation file

```
<?xml version="1.0" encoding="UTF-8"?>
<omexManifest xmlns="http://identifiers.org/combine.specifications/omex-manifest">
 <content location="." format="http://identifiers.org/combine.specifications/omex" />
 <content location="./manifest.xml" format="http://identifiers.org/combine.specifications/omex-manifest" />
 <content location="./metadata.rdf" format="http://identifiers.org/combine.specifications/omex-metadata" />
 <content location="./README.md" format="http://purl.org/NET/mediatypes/text/markdown" />
 <content location="./smith_chase_nokes_shaw_wake_2004.cellml"
          format="http://identifiers.org/combine.specifications/cellml" />
 <content location="./smith_chase_nokes_shaw_wake_2004_example_semantics.rdf"
          format="http://identifiers.org/combine.specifications/omex-metadata" />
</omexManifest>
```
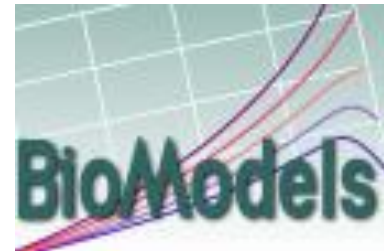
# To build the Physiome, we need….

# Model repositories

- Cell ML repository
www.cellml.org/

- SBML and BioModels
www.ebi.ac.uk/biomodels/

# Model search/integration

With semGen and annotated models, we can do a semantic search, merge models and succeed!

# Model search/integration

With semGen and annotated models, we can do a semantic search, merge models and succeed!

Or not….

# Reproducibility

- Reproducing results is not enough
(a good first step, but…)

- Reuse requires
  - Finding models
  - Understanding models and their assumptions
  - Modifying models to fit new needs/contexts

# The center for reproducibility

Promises:

- Curators & annotators and a pipeline
- Easy access to biosimulation environments
- New ontologies and standards for assumptions and provenance of models
- Data annotation
- Visualizations of models

# Questions?