**Annotation protocol (SemGen)**

**1. Provide free-text definitions for all codewords**

*This step simplifies subsequent annotation steps by removing the need to look up definitions in external sources such as journal articles. It provides a minimal, human-readable, semantic definition of the model.*

* Collect from source publication(s), in-line comments, communication with authors, etc.
* Spell out acronyms
* Do not require model users to go elsewhere to understand the biophysical meaning of a codeword or sub-model

**2. Confirm auto-added physical property annotations, add manually as needed**

*SemGen will attempt to accurately identify the OPB:Physical property that a codeword represents based on the codeword’s physical units and its mathematical relationship to other codewords; however, these assignments are not guaranteed to be correct.*

* Always use an OPB term
* In ambiguous cases, SemGen will not include a property annotation; it will be the user’s responsibility to specify it. In the future, SemGen will offer suggestions on which OPB terms might be appropriate.

**3. Annotate the codewords that represent properties of physical entities**

*Use the following recommended reference ontologies or create a custom term if an ontology term with the precise meaning is not available:*

* Foundational Model of Anatomy (FMA) for macromolecular structures on up
* Mouse Adult Gross Anatomy (MA) ontology for rodent-specific anatomy
* Cell Type Ontology (CL) for cell types not provided by the FMA
* GeneOntology:cellular component for subcellular structures not in the FMA
* Protein Ontology (PR) for proteins (UniProt can also be used to annotate proteins, but we recommend using terms from PR that are not taxon-specific)
* Chemical Entities of Biological Interest (ChEBI) for atoms and small molecules (e.g. metabolites)
* Ontology for Biomedical Investigations (OBI) for laboratory materials

**4. Annotate the codewords that represent properties of physical processes**

***Always create a custom term for physical processes.***

* Provide a precise, human-readable definition for the new term.
* Define the term logically by specifying the dynamical sources, sinks and mediating entities that participate in the process (SemGen provides an interface for this that allows the user to select from the physical entities entered in step 3).

***Rationale:*** *Currently there is no reference ontology for multi-scale biological processes that associates the processes with their physical entity participants. This information is crucial for automating model composition because it indicates how the processes affect the thermodynamic states of the physical entities in the system. This information is also critical for intelligently re-formulating conservation and flow equations during the model merging process. The long-term vision is to eventually create a reference ontology of logically-defined biological processes by “harvesting” the custom process terms in SemSim models and aggregating them into a single knowledge base.*

**5. Annotate the codewords that represent properties of physical forces**

***Conventionally, physical forces are not given unique names, so physical forces are always anonymous in SemSim models. Examples of forces include electrical potentials, chemical potentials, fluid pressures, and temperatures.***

* Define the force logically by specifying the dynamical sources and sinks that generate the force (SemGen provides an interface for this that allows the user to select from the physical entities entered in step 3).
* If a sink is unspecified, the sink is considered to be whatever "ground" is applicable, given the OPB force property term used in the composite annotation. For example, the ground for a codeword annotated as an *OPB:Electrical potential* would be the electrical ground; the ground for a codeword annotated as an *OPB:Temperature* would be the ambient temperature.

***Rationale:*** *There is no comprehensive set of knowledge resource terms that define specific physical forces, and so forces are defined logically within SemSim models by virtue of the thermodynamic sources and sinks that generate them.*

**6. Annotate the codewords that are properties of constitutive dependencies using terms subclassed under OPB:Constitutive property.**

*These include curve-shaping constants, reaction rate parameters, resistances, and other properties that are defined by the relationship between two or more disjoint physical properties.*

* Apply an OPB property to the codeword's composite annotation and leave the rest of the composite empty.
* In future versions, SemGen will be able to automatically determine the constitutive dependency associated with the physical property based on the OPB annotation.

**7. Delineate and annotate sub-models**

*A SemSim sub-model is defined as a model section comprised of one or more codewords and the computations that solve them. Any codewords required for these computations, but which are not explicitly included in the sub-model, go “along for the ride” as input parameters when extracting a submodel.*

* Provide a free-text definition for the sub-model
* Identify the codewords in the sub-model and any sub-models that it subsumes

**8. Provide model-level annotations**

The following are some basic recommendations for applying metadata to the model as a whole (such as curatorial information). The question of how this metadata should be structured and formatted across modeling formats remains open.

* Describe the taxon in which the simulated processes occur using an NCBI organismal taxonomy reference term
* Provide the PubMed ID and title of the model’s reference publication
* Provide a list of authors associated with the model
* Provide the contact name and email for those who annotated the model
* Provide a concise description of the model (the reference publication abstract could be used).

**9. Rules of thumb**

* Be as precise as possible with your annotations
* Use a composite annotation wherever possible
* If the codeword’s meaning cannot be captured with the composite annotation structure, use a singular annotation term that provides a precise definition. If a singular annotation term is not available, send a term request to the curators of the appropriate knowledge resource.
* Assert all custom entities and processes as a subclass of some reference ontology concept so that software tools can assess their semantic context.

With SemGen we aim to guide and constrain the annotation process to ensure a high level of quality and consistency in semantic annotation. If you have suggestions about how to improve quality and/or consistency, please contact Maxwell Neal: mneal { at } uw [ dot ] edu.