Harmonizing semantic annotations for computational models in biology

Maxwell Lewis Neal ⓘD, Matthias König ⓘD, David Nickerson ⓘD,
Göksel Mısırlı ⓘD, Reza Kalbasi, Andreas Dräger ⓘD, Koray Atalag,
Vijayalakshmi Chelliah, Michael T. Cooling ⓘD, Daniel L. Cook,
Sharon Crook ⓘD, Miguel de Alba, Samuel H. Friedman ⓘD, Alan Garny,
John H. Gennari, Padraig Gleeson ⓘD, Martin Golebiewski, Michael Hucka ⓘD,
Nick Juty, Chris Myers, Brett G. Olivier ⓘD, Herbert M. Sauro, Martin Scharm,
Jacky L. Snoep, Vasundra Touré, Anil Wipat, Olaf Wolkenhauer ⓘD and
Dagmar Waltemath ⓘD

Corresponding author: Maxwell Lewis Neal, Seattle Children's Research Institute, Center for Global Infectious Disease Research, 307 Westlake Ave. N Suite 500, Seattle, WA 98109, USA. Tel: 206-884-3198; Fax: 206-884-3104; E-mail: Max.Neal@seattlechildrens.org

Abstract

Life science researchers use computational models to articulate and test hypotheses about the behavior of biological systems. Semantic annotation is a critical component for enhancing the interoperability and reusability of such models as well as for the integration of the data needed for model parameterization and validation. Encoded as machine-readable links to knowledge resource terms, semantic annotations describe the computational or biological meaning of what models and data represent. These annotations help researchers find and repurpose models, accelerate model composition and enable knowledge integration across model repositories and experimental data stores. However, realizing the potential benefits of semantic annotation requires the development of model annotation standards that adhere to a community-based annotation protocol. Without such standards, tool developers must account for a variety of annotation formats and approaches, a situation that can become prohibitively cumbersome and which can defeat the purpose of linking model elements to controlled knowledge resource terms. Currently, no consensus protocol for semantic annotation exists among the larger biological modeling community. Here, we report on the landscape of current annotation practices among the COmputational Modeling in BIology NEtwork community and provide a set of recommendations for building a consensus approach to semantic annotation.

Key words: semantic annotation; computational modeling; knowledge representation; modeling standards; data integration

Introduction

Biological researchers use computational models to articulate hypotheses about the organization and dynamics of biological systems. Modeling has become an essential component of life science research; it provides a basis for investigating system perturbations, predicting experimental outcomes and identifying a system's most critical processes. Given the increasingly tight coupling between computational biology and biomedical

research, the research community benefits when models can be easily repurposed across research groups. However, modelers use a variety of modeling languages and simulation platforms, not all of which interoperate coherently. Researchers aiming to repurpose a published model for their investigations must often resort to hand-coding the model anew, which is a time-consuming and error-prone process. Although much progress has been made in developing standardized model exchange formats such as the Systems Biology Markup Language (SBML) [1] and CellML [2], fundamental roadblocks associated with model reuse remain [3–6]. Here, we identify several critical bottlenecks that impede the reuse of models and model components as well as their integration. We argue that standardizing and expanding the semantic annotations on models remove these bottlenecks by helping researchers more quickly locate models, automate model compositions, translate between modeling formats and integrate the biological knowledge encoded in models.

Our perspective on semantic annotation is shaped by our participation in the COmputational Modeling in BIology NEtwork (COMBINE), a community of researchers developing standards for modeling in computational biology [7]. Collectively, we represent teams developing SBML, CellML, the Simulation Experiment Description Markup Language (SED-ML) [8], the Systems Biology Graphical Notation (SBGN) [9], the Synthetic Biology Open Language (SBOL) [10], NeuroML [11, 12], rule-based modeling languages [13, 14], MultiCellDS [15], the COMBINE archive [16], the Semantic Simulation (SemSim) architecture [17], FAIR-DOMHub [18], the SABIO-RK database [19], as well as BioModels [20] and the Physiome Model Repository [21, 22]. Here, we present a set of agreed-upon recommendations for harmonizing semantic annotations for computational models across these development efforts. Adopting these recommendations within our communities, together with the implementation of software libraries and tools that adhere to the recommendations, will vastly extend the possibilities for interpreting, comparing and evaluating computational models and thus enhance model reuse and repurposing.

Semantic annotations and their utility

In the context of computational modeling in biology, we define a semantic annotation as a computer-accessible metadata item that captures, entirely or in part, the meaning of a model, model component or data element. For example, an annotation on a model variable might indicate that it represents the *concentration of cytosolic glucose in a pancreatic beta cell* or it might describe a purely computational feature such as the *simulation time step*. To further clarify, we make a distinction between semantic annotations and other types of metadata that do not indicate meaning, such as curatorial annotations that indicate model authorship, provenance annotations that indicate the origins of model components (e.g. nanopublications) or tool-dependent annotations that indicate layout information for model visualization. Broadly, semantic annotations are a critical feature of the vision of the Semantic Web, wherein documents are linked to metadata describing the document's contents, thus facilitating search and retrieval as well as data interoperability [23]. Semantic annotations are used in many fields besides biological modeling, including the geosciences [24], music retrieval [25] and business process modeling [26].

As defined above, semantic annotations describe the meaning of a model's contents in a computer-accessible manner. These annotations are needed because, like much of the bio-

logical community, modelers and tool developers do not use a standardized set of names to indicate the meaning of model elements. One modeler may use a variable named 'X' to represent *cytosolic glucose concentration in a pancreatic beta cell*, and another modeler may use 'X' to indicate *blood flow rate through the aortic valve*. In the absence of community-wide naming conventions, machine-accessible metadata is used to capture the meaning of model elements. This metadata links model elements to terms from knowledge resources such as controlled vocabularies and ontologies, allowing different software tools to recognize when two models represent the same or similar biological features, which in turn helps researchers to align, reuse and merge models.

Annotations have contributed to the successful reuse and exploration of biological models and data in tasks such as comparison [27, 28], interpretation [29], retrieval [30–32], integration [33–37], simulation [38], translation between formats [29, 37, 39–42] and visualization [37, 43–45]. Semantic annotations are also a key component for model-driven design of synthetic biological systems where they are used in model composition tasks when constructing optimum biological systems built from models [41, 45–48].

Applying semantic annotations to address challenges in model reuse and integration

Several barriers impede the reuse of models and model components. Researchers cannot readily search across multiple repositories to find models of interest, and when a researcher does retrieve a relevant model, they must determine whether it (or part of it) can be repurposed for use in their modeling work. Similarity measures and pattern-matching algorithms can help identify relevant modeling components [28, 49], but existing proposals for cross-repository search and retrieval [32] are mostly theoretical and not yet applicable in practice. Applying semantic annotations in a standardized way across model repositories would provide a common ground for such cross-repository searches, allowing models encoded in diverse formats to be retrieved based on their shared semantic descriptions.

The amount of time required to select a model for reuse increases when the researcher must choose from a large set of potentially usable models. With only limited means to compare models automatically, researchers must manually assess the content, scope and underlying assumptions of each candidate model as well as the biological questions each was designed to answer. While semantic annotations cannot yet capture the purpose for which a model was built or modified, they can make the biological content of a model explicit and therefore help researchers decide whether or not to repurpose it. Semantic annotations can also be used to quantify the similarity between model elements [27, 28, 50–54], helping to ensure that the user is presented with the most relevant models following a search. Other types of annotations, such as those that capture provenance information [55–58], can also help researchers determine which models best meet their research needs and make comparisons between different model versions [59].

Another set of barriers is associated with model-to-model integration. Composing new models from existing models remains a mostly manual and error-prone process, and previous work has demonstrated how it can be accelerated using semantic annotations [31, 34, 36, 60]. By examining semantic annotations, software tools can recognize where models overlap in their biological content and then provide recommendations about how the models could be coupled.

An additional impediment to model reuse is the lack of integration between model repositories and online collections of experimental data. If such integration existed, modelers could readily find data appropriate for validating or constraining the models they repurpose, and experimentalists could find models for use in data analysis pipelines. If datasets and models were semantically annotated using a standard protocol, software tools could discover relevant connections between them and accelerate research. Efforts such as the SourceData project [61], which aims to make biological data discoverable through the use of metadata annotations, can help achieve this vision. Complementary efforts that link models with associated datasets also contribute to this goal. For example, the Simulation Database [62] component of the JWS Online Model Repository [63] or tellurium-web [64] provides standardized descriptions of reproducible simulation experiments (encoded in SED-ML), manually linked to model code and, if available, original experimental datasets. Using such standardized descriptions, researchers can readily reproduce and evaluate curated simulations against experimental datasets. The opportunities for integrating models and data could substantially increase if model and data repositories such as these used a standard semantic annotation protocol.

Examples of current semantic annotation practices

Participants across several modeling initiatives recognize the importance of semantic annotations; however, annotation practices vary across these efforts and are in need of standardization. To illustrate, we profile the semantic annotation practices of three initiatives within COMBINE in this section. These profiles are not meant to provide a comprehensive summary of all initiatives that focus on annotating biological models, but rather to offer examples of initiatives that, when considered together, demonstrate the need for harmonized annotation approaches.

BioModels

BioModels is a repository of publicly available models maintained at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI). The most recent BioModels release (#31) includes 640 curated models. The BioModels team has created guidelines for annotating the semantics of these models (https://www.ebi.ac.uk/biomodels-main/annotationtips). These guidelines indicate which components of curated models should be annotated and provide guidance on choosing appropriate reference terms. They have been created following the Minimum Information Requested In the Annotation of Models (MIRIAM) [65] guidelines.

Annotations within curated SBML models also follow the instructions given in SBML specification documents [66]. They are primarily encoded as Resource Description Framework (RDF) statements (https://www.w3.org/RDF/), although the SBML format also contains explicit constructs for associating Systems Biology Ontology [38] terms with model elements outside of the RDF content. The BioModels team has developed an in-house annotation tool for capturing the meaning of the biological aspects of SBML models using terms from a defined list of established knowledge resources such as UniProt [67], Chemical Entities of Biological Interest (ChEBI) [68] and the Gene Ontology [69]. Given that a model must be semantically annotated for

inclusion in the curated branch of BioModels, the repository contains a broad set of semantic annotations. For details on the number and granularity of these annotations, we refer the reader to Henkel *et al.* [32] and Alm *et al.* [27].

Additionally, the EMBL-EBI maintains a service at https://identifiers.org for resolving the Uniform Resource Identifiers (URIs) used in annotations [70]. Formatted according to the identifiers.org guidelines, URIs from various biomedical knowledge resources can be resolved online. Together with the COMBINE-maintained BioModels.net 'qualifiers' (https://co.mbine.org/standards/qualifiers), also known as 'predicates' or 'relations', they allow the construction of complete semantic annotations [65] linking elements of COMBINE formats (e.g. SBML, SED-ML or COMBINE archive metadata) to knowledge resource terms in order to define an element's biological meaning. Figure 1 shows an RDF-based semantic annotation on an example SBML model from BioModels.

The Physiome Model Repository

The Auckland Bioengineering Institute at the University of Auckland manages the Physiome Model Repository, which currently contains over 800 CellML models as well as models and simulation protocols encoded in various other formats. Annotation of CellML models is currently limited, but a collection of metadata specifications exists that provide recommendations and best practices for annotating models [72, 73]. Although the CellML metadata specification states that semantic annotations should be serialized externally, current tools used in the CellML community such as OpenCOR embed RDF/XML annotations in the CellML documents themselves, using identifiers.org URI formatting and BioModels.net qualifiers.

The CellML format focuses on representing the mathematical aspects of models, and it does not include biological constructs as in SBML models. Consequently, a model's variables must be linked to knowledge resource terms to capture the precise meaning of what a CellML model simulates. This presents challenges, as these variables represent concepts that can be very fine-grained (e.g. *concentration of cytosolic glucose in a pancreatic beta cell*), and publicly available knowledge resources do not provide adequate coverage for such concepts. Thus, the CellML group has been collaborating with the creators of the SemSim architecture, described next, to develop an annotation approach that uses *composite* annotations [17] to describe such fine-grained concepts.

The SemSim architecture and SemGen

The SemSim architecture is a logical framework for capturing the biophysical meaning of what is represented in a biological model. Central to this architecture are composite annotations: logical statements that link multiple knowledge resource terms to precisely define a model element. The primary motivation behind the composite annotation approach is that biological models often simulate concepts that are not represented among the set of publicly available biological knowledge resources; therefore, annotators often cannot define a model element via a reference to a single controlled vocabulary term. With composite annotations, annotators can instead build a definition from multiple, more fundamental terms that are available in knowledge resources. For example, a model variable might simulate the *concentration of cytosolic glucose in a pancreatic beta cell*, but this

concept is not explicitly represented in any single publicly available knowledge resource. Applying the SemSim architecture, an annotator can create a composite annotation that captures the variable's biological meaning by linking the terms *Chemical concentration* from the Ontology of Physics for Biology (OPB) [74, 75], *glucose* from ChEBI, *Cytoplasm* from the Foundational Model of Anatomy (FMA) [76] and *Type B cell of pancreatic islet* from the FMA. Depending on annotator preferences, synonymous terms from alternative knowledge resources could be used to create this composite annotation. For example, Uberon [77] could be used as an alternative to the FMA.

The SemSim development group has created SemGen (https://github.com/SemBioProcess/SemGen), a software tool for applying SemSim-compliant composite annotations to computational models. SemGen also provides capabilities for semantics-based model composition wherein a model's annotations are leveraged to automate merging and extraction tasks [33, 34]. Using SemGen, composite annotations applied to a computational model can be stored within SemSim models, which are encoded in the RDF-based Web Ontology Language (OWL, https://www.w3.org/OWL/), or within SBML and CellML models as standard RDF. RDF is sufficiently expressive for serializing composite annotations, and SemSim models are encoded in OWL because the SemSim development group is exploring the use of automated inference algorithms for reasoning over the biological knowledge articulated in models [78]. The SemSim development group maintains an informal protocol for annotating a model with SemGen in accordance with the SemSim framework (https://github.com/SemBioProcess/SemGen/wiki/Annotation-protocol). These guidelines indicate which model components should be semantically annotated and at what level of detail. They also list recommended knowledge resources to use for different components of a composite annotation. For example, the OPB is recommended for physical property terms (*chemical concentration*, *fluid volume*, etc.), and ChEBI for small molecules.

As evidenced by these three initiatives, there is currently no consensus approach to semantic annotation being applied to the different projects under COMBINE. Every COMBINE initiative either uses a protocol specific to that initiative or has no established protocol for semantic annotation. A consensus approach is essential not only for integration across COMBINE standards but also for integration with external biological data representation efforts. For example, the *open*EHR (http://www.openehr.org/) and HL7 Fast Healthcare Interoperability Resources (https://www.hl7.org/fhir/) initiatives have established protocols for the semantic annotation of clinical data (called terminology binding). As discussed above, a consensus protocol for annotating models and data would greatly benefit researchers who apply models to understand clinical data. Working toward this goal, members of the *open*EHR and COMBINE communities have recently begun collaborating to coordinate their semantic annotation approaches. Thus, a primary challenge for the COMBINE community is to develop an approach to semantic annotation that will help ensure annotations among COMBINE standards are widely interoperable and consistent.

Recommendations

Through discussions and face-to-face meetings at COMBINE meetings, we have developed the following seven recommendations for harmonizing semantic annotations. They are summarized in Box 1.

Box 1. Summary of Recommendations.

i. *Use RDF, identifiers.org URIs and BioModels.net qualifiers to encode semantic annotations*
This recommendation indicates which technical standards to use for storage of semantic annotations.

ii. *Store annotations in a separate file*
This practice will help normalize the format in which annotations are stored across community initiatives. Storing annotations apart from the model also allows different research groups to annotate a model using different knowledge resources.

iii. *Establish a dedicated group for developing a software library that supports semantic annotation standards*
The proposed group will create and disseminate semantic annotation standards, develop standards-compliant software, promote consistency in annotation practices and ensure that advancements in annotation standards are coordinated across the modeling community.

iv. *Document which knowledge resources should be used for annotation and why*
Model annotators have different requirements for disambiguating model content, depending on their research goals. We encourage research groups to articulate their preferred knowledge resources on a group-by-group basis and provide publicly available documentation describing how specific knowledge resources should be used in annotations. This will help the broader community determine how to map annotations between research efforts.

v. *Establish a repository of reusable annotations*
This will help reduce the time required for annotation and promote inter-annotator consistency. Annotations in the repository could also be used for annotating experimental data.

vi. *Ensure high-quality semantic annotations through training and quality control processes*
Annotators should be trained so their annotations are specific, complete and as consistent as possible across model repositories. We encourage annotation initiatives to develop quality control protocols to ensure annotations adhere to community standards.

vii. *Establish and maintain collaborations with knowledge resource developers*
Concepts represented in biological models are not always present in existing knowledge resources. Establishing and maintaining collaborative relationships with knowledge resource developers will help ensure that new terms needed for annotation become available in a timely manner.

(i) Use RDF, identifiers.org URIs and BioModels.net qualifiers to encode semantic annotations

This recommendation addresses the issue of how to standardize the storage of semantic annotations within documents, an essential technical pre-requisite for harmonizing their representation across model annotation efforts.

```
<compartment id="MT_IMS" name="mitochondrial intermembrane space" metaid="metaid_MT_IMS" size="1">
 <annotation>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
   <rdf:Description rdf:about="#metaid_MT_IMS">
    <bqbiol:is>
     <rdf:Bag>
      <rdf:li rdf:resource="http://identifiers.org/go/GO:0005758"/>
     </rdf:Bag>
    </bqbiol:is>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</compartment>
```

Figure 1. Example RDF-based annotation from SBML model BIOMD0000000239 [71] in BioModels. The annotation block (indicated by a curly brace) defines the biological meaning of a physical compartment in the model. The RDF block within the *<annotation>* element links the compartment's metadata identifier 'metaid_MT_IMS' to the Gene Ontology term GO:0005758, which represents the mitochondrial intermembrane space. The use of the *bqbiol:is* predicate in this link indicates that the compartment is defined as the mitochondrial intermembrane space.

We recommend using RDF, a World Wide Web Consortium-recommended standard for representing information on the Web, for encoding annotations. RDF has emerged as the *de facto* standard for encoding semantic annotations among the COMBINE community; all COMBINE standards currently use it. While more expressive knowledge representation formats exist, our experience indicates that RDF is sufficiently expressive for articulating the kinds of semantic annotations required to catalyze the significant advances in model discovery, reuse and integration that we envision. RDF statements are built using subject-predicate-object triples that, as shown in Figure 1, can be used to assert relationships between model components and terms from online knowledge resources. A primer on RDF is available (https://www.w3.org/TR/rdf11-concepts/), and links to examples of RDF-encoded model annotations can be found in the Example Semantic Annotations section below.

We recommend using the identifiers.org URI format when referencing knowledge resource terms in RDF statements. Identifiers.org supports a vast set of biological knowledge resources used for semantic annotation, including databases as well as ontologies, and because identifiers.org-formatted URIs are web-resolvable. These URIs are widely adopted by the biological research community, particularly to facilitate scientific data interoperability and are beginning to show uptake among high-impact publishing groups [79, 80]. The identifiers.org services are capable of more complex URI resolution compared to alternative services. For example, identifiers.org is specifically built to address downtime and changing endpoints and directs users to an alternative site for a given data record as long as one is listed (one-to-many mappings) whereas persistent uniform resource locator services specify only one endpoint for URI resolution (one-to-one mappings). We also recommend using identifiers.org-formatted URIs because they use a uniform and simple nested structure that facilitates generation and parsing, and because identifiers.org reuses data providers' record identifiers.

We recommend the use of BioModels.net qualifiers in annotation statements that define model elements. These existing qualifiers provide a basic level of coverage needed for articulating semantic annotations in models, and they are specifically intended for use in statements that link *computational abstractions* of physical phenomena to knowledge resource terms representing the material manifestations of those phenomena. We would welcome efforts to unify BioModels.net qualifiers with alternative collections of qualifiers that do not have an explicit focus on linking elements in computer simulations with physical phenomena, such as the Relations Ontology [81].

(ii) Store annotations in a separate file

Current practice among the COMBINE community is to store semantic annotations within the same file that specifies a model's computational aspects. This practice is in line with the initial MIRIAM guidelines put forth for annotating models [65]; however, there are several reasons why we recommend storing semantic annotations in a separate file. First, we wish to normalize the format in which annotations are stored across the different COMBINE standards. Currently, the exact format used to store annotations within model files differs slightly from standard to standard. Normalizing the format will simplify the development of software that provides programmatic manipulation of semantic annotations and will allow for better separation between modeling and annotation tasks. This will also remove the burden of supporting semantic annotation from the software teams that are developing software libraries for specific COMBINE standards. We recommend establishing a separate development group that will be responsible for creating and maintaining a common data model and annotation library (Recommendation 3). If the community decides to change the way semantic annotations are stored or processed, this should not require each team of developers to adjust their software. Instead, there should be one development project focused on the programmatic manipulation of semantic annotations, and the development team for that annotation library should implement the community's decisions.

We also recommend storing annotation files separately because we recognize that different research groups may have different preferences for which knowledge resources to use for annotation. Externalizing annotations in a separate file allows a single model file to be referenced by multiple annotation files, allowing different research groups to describe the same modeling resource in different ways. This approach follows the vision of the COMBINE archive, wherein multiple types of modeling files are archived together (as in a zip file) to make simulation experiments readily reproducible and shareable among research groups [16]. When sharing models, we recommend that annotations be distributed along with the files they annotate, and COMBINE archives provide a standardized way to bundle such files together.

We recognize that storing annotations in a separate file requires keeping them synchronized. For example, if a variable identifier changes in the model file, that change should be reflected in the annotation file(s) as well. However, ensuring synchrony is an issue regardless of whether or not annotations are stored in a separate file, and we recommend that the community encourage the development of tools that help ensure coordination between a model's computational aspects and its semantic annotations.

An additional advantage of storing annotations in a separate file is that the RDF content can be serialized in various formats including XML or Turtle, whereas currently the serialization is dictated by the model format.

(iii) Establish a dedicated group for developing a software library that supports semantic annotation standards

Programmatically creating, applying and managing semantic annotations on models require substantial software development. The modeling community, especially model curators, would benefit from the creation of a single dedicated group that focuses on creating and disseminating semantic annotation standards and developing standards-compliant software. The organization and operation of the new working group would follow the model of other successful software development efforts within COMBINE, such as those that focus on developing the widely adopted SBML libraries libSBML [82] and JSBML [83]. Using this model, participants within the development group can come from different research labs, and funding to support the group's efforts can come from multiple sources. Centralized software development systems such as GitHub can be used to develop a semantic annotation library collaboratively among members of the group, and working group meetings at COMBINE conferences/hackathons provide venues for promoting group cohesion and for interfacing with other standards development efforts. Online forums can be used to resolve outstanding software issues, propose enhancements, discuss the working group's priorities, etc. Although this operational model has worked in the past, it is crucial that funding for the effort is as continuous as possible. It is also important that individuals contributing to the effort can be compensated through alternative funding mechanisms in case a specific funding source ends.

The semantic annotation group's role would be to promote consistency in annotation practices and help ensure that advancements in annotation standards are coordinated across the modeling community. It would also work to ensure the longevity of annotation tools beyond individual software projects within COMBINE. Given that several tools already provide model annotation capabilities, including OpenCOR [84], SemGen [17, 34], COPASI [85], JWS Online [63], CellDesigner [86] and the annotation tools used internally by the BiGG Models Database [87] and BioModels, the working group's focus would not be to create a one-size-fits-all semantic annotation toolset, but rather provide guidance and core software infrastructure to help ensure that users and developers of modeling software adhere to the community's annotation standards.

This centralized semantic annotation software package should harmonize with other annotation standards, such as the curatorial metadata standards of MIRIAM and should support compliance-checking. We would also encourage package developers to provide multilingual support so that the lexical content associated with models and annotations can be readily converted into a wide variety of languages. This would further enhance researchers' abilities to comprehend the meaning of model elements.

(iv) Document which knowledge resources should be used for annotation and why

The COMBINE community currently uses terms from a variety of biomedical knowledge resources to capture the meaning of model elements. Some of these knowledge resources overlap in content, and there is no broad consensus about which resources should be used for an annotation. Therefore, the same biophysical concept represented in two different models might be annotated against different knowledge resource terms. This undermines the community's ability to compare and compose models in an automated fashion, as well as convert between standard formats. Ideally, synonymous content within models should be annotated using the same set of reference terms and qualifiers. Otherwise, the developers of model comparison and composition tools must rely on mappings between knowledge resources to identify synonymous terms. This can be burdensome, given the increased computational cost and the challenges associated with ontology mapping and alignment [88]. It is in the interest of the community to agree upon a set of core knowledge resources that will be used for annotation and also define the scope of use for each resource.

That being said, we recognize that different research groups have different requirements for disambiguating model content. For example, UniProt might be a more useful resource compared to the Protein Ontology (PRO) [89] for a group that needs to disambiguate proteins by amino acid sequence. For a group that does not require sequence information and wishes to use taxon-neutral protein entities, the PRO might be more useful. Furthermore, the value of different knowledge resources may change over time. Therefore, we do not propose a short list of recommended knowledge resources here. Instead, we encourage research groups to make recommendations on a group-by-group basis and provide publicly available documentation describing how specific knowledge resources should be used in annotations (see, for example, the SemSim development group's annotation protocol mentioned above). This will help those outside the research group determine how to map annotations between research efforts. Such mappings represent perhaps the most significant challenge associated with this approach, and we therefore stress the importance of creating and maintaining formal mappings between knowledge resources, such as those provided by BioPortal (https://bioportal.bioontology.org/) and the EMBL-EBI Ontology Xref Service (http://www.ebi.ac.uk/spot/oxo/).

We recommend that research groups use knowledge resources that are publicly and freely available, programmatically accessible via web services and mapped to other resources used within the community that overlap in content.

(v) Establish a repository of reusable annotations

The process of annotating a model is time-consuming and susceptible to inter-annotator variability. We, therefore, recommend the creation of an online repository of curated, reusable annotations and integration between this repository and software tools. The repository would allow annotators to quickly find and reuse valid, standards-compliant annotations previously applied to

curated models, reducing annotation time and promoting consistency among annotators. This feature is especially relevant for applying composite annotations; they take longer to create compared to singular annotations and require linking multiple knowledge resource terms, resulting in more opportunities for inter-annotator inconsistency. We envision this repository would mine RDF annotation statements from curated models (e.g. in BioModels or the PMR) and integrate with annotation software tools so that annotators can quickly link elements in their models/datasets to knowledge resource terms by reusing these statements.

Such a repository could also provide the basis for automating the annotation process; once an annotator begins annotating a model, software tools could search the repository for curated models that have the same annotations and make suggestions for unannotated model elements. This approach would also improve inter-annotator consistency; see Recommendation 4 above.

By utilizing the mappings that exist between knowledge resources (e.g. between ChEBI and the Kyoto Encyclopedia of Genes and Genomes [90]), a centralized repository of annotations could also identify and auto-generate synonymous annotations. This feature would help minimize redundant annotations within the repository and allow annotators to retrieve annotations that only use references to their preferred knowledge resources. Additionally, as discussed in Recommendation 3, we encourage the developers of an annotation repository to provide multilingual support.

An additional benefit of such a repository is that the annotations it contains could be used for annotating experimental data. Therefore, the repository could act as a hub for accessing biological models, data and associated metadata. Given this, we recommend that the repository's development team interfaces not only with members of the COMBINE community but with other external data sharing/integration efforts such as the Findable, Accessible, Interoperable, Reusable [91] initiative and Bioschemas [92]. This will help ensure that the repository's content is readily discoverable, accessible and integrable by a wide community of users and will assist in developing effective metadata governance approaches. We anticipate that outreach will be a significant challenge that repository developers must address.

(vi) Ensure high-quality semantic annotations through training and quality control processes

The success of new software tools for expediting cross-repository model retrieval and model composition depends on linking models to thorough, precise semantic descriptions. Therefore, we recommend that annotators, whether they are the modelers themselves or curation team members, receive training so that their annotations are specific, complete and as consistent as possible across repositories. Annotators must not only thoroughly understand the meaning of a model's computational elements, but also the scope, organization and limitations of the knowledge resources used for semantic annotation as well as the technical aspects of encoding annotations.

We also encourage all groups annotating models to develop quality control protocols that ensure annotations are encoded correctly; to identify overly generic, overly specific and missing annotations; and to check that the annotations accurately capture the meaning of the model elements to which they

are linked. We also recommend the enhancement of existing annotation software tools [36, 93, 94] so they prevent the application of inconsistent annotations and can auto-suggest annotations. Another interesting approach for promoting high-quality annotations is the use of Web 2.0 advancements and collaborative technologies to evaluate semantic annotations with user-friendly and interactive infrastructures (see, for example, Kalbasi *et al.* [24]).

Implementing these proposed training and quality control processes can be a challenge. Funds must be obtained to finance their implementation and maintenance. Furthermore, implementation will require specialized expertise to train annotators and develop quality control practices. Therefore, the biological modeling community would benefit from new funding initiatives that aim to establish training programs and quality control operations for annotation practices.

(vii) Establish and maintain collaborations with knowledge resource developers

Modelers do not always organize a system for study in the same way as knowledge resource developers, and sometimes the biological concepts represented in a model are not present in existing knowledge resources. Many biological concepts in models are too fine-grained to be included and maintained in knowledge resources aiming to provide broader classifications for biophysical phenomena. Thus, annotating a model may require sending term requests to resource developers. Therefore, the modeling community should establish and maintain connections with these developers so that terms needed for annotation can be added to resources or modified in a timely fashion. For example, the Kinetic Simulation Algorithm Ontology [38] term submission system on the COMBINE website (http://co.mbine.org/standards/kisao) connects to a SourceForge ticket system. An ongoing dialog between modelers and resource developers will also help modelers better understand the scope and content of knowledge resources, which is a critical component of Recommendation 6. These collaborations, although they will require resource investment by all parties involved, will likely be mutually beneficial: model annotators will be able to annotate faster and more thoroughly, and knowledge resource developers will be able to improve the coverage of their resources via feedback from annotation efforts.

Example semantic annotations

To provide concrete examples of semantic annotations that adhere to our recommendations, we have created two example COMBINE archives that contain annotated models. These archives are included as Supplementary Data, and their latest versions are available at the COMBINE GitHub repository (https://github.com/combine-org/Annotations). One of the archives contains an SBML model retrieved from BioModels (BIOMD0000000176) that simulates glycolysis in yeast [95]. The other contains a CellML model retrieved from the Physiome Model Repository that simulates hemodynamics in the human circulatory system [96]. These two models were chosen because, together, they provide examples of semantic annotations across physical scales and because the phenomena they simulate will be familiar to many biological researchers. In the future, we plan to create example semantic annotations for additional COMBINE standards (e.g. NeuroML and SBGN) and post them to the COMBINE GitHub repository as well. Developers within the COMBINE

community are currently enhancing existing tools to support reading and writing semantic annotations that adhere to our recommendations, and recent versions of SemGen and tellurium-web support reading the example archives provided here.

We have serialized semantic annotations for both of our example models as separate files within COMBINE archives. The annotations within these standalone files are encoded as RDF subject-predicate-object statements and the URIs for the subjects in these statements serve as pointers to the specific model elements that are annotated. To construct these pointers, we use the relative model file name for the URI base and the 'metaid' attribute on the annotated model element for the URI fragment. For example, the URI referencing the cytosol compartment in the SBML model is ./BIOMD0000000176.xml#_525523. This approach allows us to uniquely reference model elements within a COMBINE archive, even if elements from two different models share the same metaid. These URI pointers establish the link between the model file and its annotations, and the original model code is left unedited. By asserting links this way, a model file does not have to be updated when it is annotated, and synchronization of pointers across multiple files is not required.

Our examples are based on initial progress toward a full technical specification of how external annotation files should be implemented for COMBINE standards. However, this specification is still under development, and several implementation details must be resolved before it can be completed, including how annotations should be implemented in BioPAX [97] and SBOL documents. Regardless, we hope that these examples will provide an initial reference point for standardization of a COMBINE-wide semantic annotation protocol. We invite those interested in participating in this standardization process to join the 'COMBINE-annot' online forum (https://groups.google.com/forum/#!forum/combine-annot), which provides a venue for addressing issues related to the annotation of models and data.

Discussion

As discussed above, semantic annotations can accelerate model reuse by enhancing search and retrieval of models. If a standard semantic annotation protocol is applied to publicly available models, the modeling community can develop more advanced tools that can search across model repositories, improve the relevance of search results and provide users with information that helps them decide whether an existing model is suitable for reuse in a modeling project. A common annotation protocol is also a critical component of model composition. Applying the protocol, model-merging tools such as SemGen can recognize the biological commonalities between models and then use that information to compare the models' biological content and guide their assembly. Without a harmonized approach to semantic annotation, modelers would be required to identify the biological overlap between models manually.

Given that data is an essential component of model-based research, we strongly encourage the computational biology community to develop protocols for data annotation. We hope that the annotation approach articulated here, and implemented in the example annotation files described above, provides a basis from which to build such protocols. It is in the interest of both modelers and experimentalists to adopt a shared approach for semantic annotation: by linking annotations in models and data sources, modelers could discover valuable datasets for tuning and validating their models, and experimentalists could discover simulation models that integrate knowledge about their field of study into systems-level perspectives.

We recognize that it takes time and effort to annotate models accurately and consistently. Therefore, we urge the community to develop incentives and tools to make it easier and faster to annotate models. Incentives might include giving credit to annotators in model metadata and on model repositories. New tools that automate compliance-checking of annotations (e.g. https://normsys.h-its.org/validate and https://github.com/opencobra/memote) could help flag errors, omissions and overgeneralizations.

Despite these challenges, we believe that semantic annotation is one of the keys for transforming the entire field of biological modeling into one in which models do not languish within single research labs, but are readily shared among the broader research community, easily incorporated into new modeling projects and contain reusable components. As the biotechnology community undertakes more investigations into systems-level biology, standardized semantic annotations will reduce the activation energy required to initiate, repurpose and extend model-based research projects. We hope that the recommendations articulated here help move the greater modeling community, including groups outside of COMBINE, toward the goal of broad model reuse and interoperability.

Key points

- Semantic annotations are a critical component for achieving broad interoperability and reusability of computational models in biology.
- Semantic annotation practices differ across groups within the larger biological modeling community; a consensus protocol for semantic annotation is needed.
- We present seven recommendations for harmonizing semantic annotation practices across the biological modeling community, along with example annotation files that adhere to the recommendations.

References

1. Hucka M, Finney A, Sauro HM, *et al*. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;**19**:524–31.

2. Cuellar AA, Lloyd CM, Nielsen PF, *et al*. An overview of CellML 1.1, a biological model description language. *Simulation* 2003;**79**:740–7.

3. Finkelstein A, Hetherington J, Margoninski O, *et al*. Computational challenges of systems biology. *IEEE Comput Soc* 2004;**37**:26–33.

4. Cooper J, Vik JO, Waltemath D. A call for virtual experiments: accelerating the scientific process. *Prog Biophys Mol Biol* 2015;**117**:99–106.

5. Neal ML, Cooling MT, Smith LP, *et al*. A reappraisal of how to build modular, reusable models of biological systems. *PLoS Comput Biol* 2014;**10**:e1003849.

6. Waltemath D, Wolkenhauer O. How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine. *IEEE Trans Biomed Eng* 2016;**63**:1999–2006.

7. Hucka M, Nickerson DP, Bader GD, *et al*. Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative. *Front Bioeng Biotechnol* 2015;**3**:19.

8. Waltemath D, Adams R, Bergmann FT, *et al*. Reproducible computational biology experiments with SED-ML - The Simulation Experiment Description Markup Language. *BMC Syst Biol* 2011;**5**:198.

9. Le Novère N, Hucka M, Mi H. The systems biology graphical notation. *Nat Comput Biol* 2009;**27**:735–41.

10. Galdzicki M, Clancy KP, Oberortner E, *et al*. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol* 2014;**32**:545–50.

11. Gleeson P, Crook S, Cannon RC, *et al*. NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Comput Biol* 2010;**6**:1–19.

12. Cannon RC, Gleeson P, Crook S, *et al*. LEMS: a language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Front Neuroinform* 2014;**8**:79.

13. Danos V, Feret J, Fontana W, *et al*. Scalable simulation of cellular signaling networks. In: *Program. Lang. Syst. APLAS 2007*. Lect. Notes Comput. Sci., Vol. 4807, 2007, 139–57. Springer-Verlag Berlin Heidelberg, Singapore.

14. Faeder JR, Blinov ML, Hlavacek WS. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol Biol* 2009;**500**:113–67.

15. Friedman SH, Anderson ARA, Bortz DM, *et al*. MultiCellDS: a standard and a community for sharing multicellular data. bioRxiv 2016

16. Bergmann FT, Adams R, Moodie S, *et al*. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. *BMC Bioinformatics* 2014; **15**:369.

17. Gennari JH, Neal ML, Galdzicki M, *et al*. Multiple ontologies in action: composite annotations for biosimulation models. *J Biomed Inform* 2011;**44**:146–54.

18. Wolstencroft K, Krebs O, Snoep JL, *et al*. FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. *Nucleic Acids Res* 2017;**45**:D404–7.

19. Wittig U, Kania R, Golebiewski M, *et al*. SABIO-RK–database for biochemical reaction kinetics. *Nucleic Acids Res* 2012;**40**:D790–6.

20. Glont M, Nguyen TVN, Graesslin M, *et al*. BioModels: expanding horizons to include more modelling approaches and formats. *Nucleic Acids Research* 2017;**46**:D1248–D1253.

21. Yu T, Lloyd CM, Nickerson DP, *et al*. The physiome model repository 2. *Bioinformatics* 2011;**27**:743–4.

22. Nickerson D, Atalag K, de Bono B, *et al*. The Human Physiome: how standards, software and innovative service infrastructures are providing the building blocks to make it achievable. *Interface Focus* 2016;**6**:20150103.

23. Berners-Lee T, Hendler J, Lassila O. The semantic web. *Sci Am* 2001;**284**:34–43.

24. Kalbasi R, Janowicz K, Reitsma F, *et al*. Collaborative ontology development for the geosciences. *Trans GIS* 2014;**18**: 834–51.

25. Turnbull D, Barrington L, Torres D, *et al*. Semantic annotation and retrieval of music and sound effects. *IEEE Trans Audio Speech Lang Process* 2008;**16**:467–76.

26. Born M, Dörr F, Weber I. User-friendly semantic annotation in business process modeling. *Web Inf. Syst. Eng. 2007 Work. WISE 2007*. Lect. Notes Comput. Sci., Vol. 4832, 2007, 260–71. Springer-Verlag Berlin Heidelberg, Nancy, France.

27. Alm R, Waltemath D, Wolfien M, *et al*. Annotation-based feature extraction from sets of SBML models. *J Biomed Semantics* 2015;**6**:20.

28. Henkel R, Hoehndorf R, Kacprowski T, *et al*. Notions of similarity for systems biology models. *Briefings in Bioinformatics* 2016;**19**:77–88 bbw090.

29. Büchel F, Wrzodek C, Mittag F, *et al*. Qualitative translation of relations from BioPAX to SBML qual. *Bioinformatics* 2012;**28**:2648–53.

30. Henkel R, Endler L, Peters A, *et al*. Ranked retrieval of computational biology models. *BMC Bioinformatics* 2010;**11**: 423.

31. Schulz M, Krause F, Le Novère N, *et al*. Retrieval, alignment, and clustering of computational models based on semantic annotations. *Mol Syst Biol* 2011;**7**:512.

32. Henkel R, Wolkenhauer O, Walthemath D. Combining computational models, semantic annotations and simulation experiments in a graph database. *Database* 2015;**2015**: bau130.

33. Beard DA, Neal ML, Tabesh-Saleki N, *et al*. Multi-scale modeling and data integration in the virtual physiological rat project. *Ann Biomed Eng* 2012;**40**:2365–78.

34. Neal ML, Carlson BE, Thompson CT, *et al*. Semantics-based composition of integrated cardiomyocyte models motivated by real-world use cases. *PLoS One* 2015;**10**:e0145621.

35. Chelliah V, Endler L, Juty N, *et al*. Data integration and semantic enrichment of systems biology models and simulations. *Data Integr. Life Sci. DILS 2009*. Lect. Notes Comput. Sci., Vol. 5647, 2009, 5–15. Springer-Verlag Berlin Heidelberg, Manchester, UK.

36. Krause F, Uhlendorf J, Lubitz T, *et al*. Annotation and merging of SBML models with semanticSBML. *Bioinformatics* 2010;**26**:421–2.

37. Touré V, Mazein A, Waltemath D, *et al*. STON: exploring biological pathways using the SBGN standard and graph databases. *BMC Bioinformatics* 2016;**17**:494.

38. Courtot M, Juty N, Knupfer C, *et al*. Controlled vocabularies and semantics in systems biology. *Mol Syst Biol* 2014;**7**:543.

39. Ruebenacker O, Moraru II, Schaff JC, *et al*. Integrating BioPAX pathway knowledge with SBML models. *IET Syst Biol* 2009;**3**:317–28.

40. Nguyen T, Roehner N, Zundel Z, *et al*. A converter from the systems biology markup language to the synthetic biology open language. *ACS Synth Biol* 2016;**5**:479–86.

41. Roehner N, Zhang Z, Nguyen T, *et al*. Generating systems biology markup language models from the synthetic biology open language. *ACS Synth Biol* 2015;**4**:873–9.

42. Rodriguez N, Pettit J-B, Dalle Pezze P, *et al*. The systems biology format converter. *BMC Bioinformatics* 2016;**17**:154.

43. Wimalaratne SM, Halstead MDB, Lloyd CM, *et al*. A method for visualizing CellML models. *Bioinformatics* 2009;**25**:3012–9.

44. Wrzodek C, Dräger A, Zell A. KEGGtranslator: visualizing and converting the KEGG PATHWAY database to various formats. *Bioinformatics* 2011;**27**:2314–5.

45. Misirli G, Cavaliere M, Waites W, *et al*. Annotation of rule-based models with formal semantics to enable creation, analysis, reuse and visualization. *Bioinformatics* 2015;**32**:908–17.

46. Misirli G, Hallinan J, Wipat A. Composable modular models for synthetic biology. *J Emerg Technol Comput Syst* 2014;**11**:22:1–19.

47. Misirli G, Hallinan JS, Yu T, *et al*. Model annotation for synthetic biology: automating model to nucleotide sequence conversion. *Bioinformatics* 2011;**27**:973–9.

48. Hallinan JS, Gilfellon O, Misirli G, *et al*. Tuning receiver characteristics in bacterial quorum communication: an evolutionary approach using standard virtual biological parts. In: *IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB*, 2014, IEEE, Honolulu, HI, USA.

49. Lambusch F, Waltemath D, Wolkenhauer O, *et al*. Identifying frequent patterns in biochemical reaction networks - a workflow. *PeerJ Prepr* 2017;**5**:e1479v3.

50. Pedersen T, Pakhomov SVS, Patwardhan S, *et al*. Measures of semantic similarity and relatedness in the biomedical domain. *J Biomed Inform* 2007;**40**:288–99.

51. Pesquita C, Faria D, Falcao AO, *et al*. Semantic similarity in biomedical ontologies. *PLoS Comput Biol* 2009;**5**:e1000443.

52. Batet M, Sánchez D, Valls A. An ontology-based measure to compute semantic similarity in biomedicine. *J Biomed Inform* 2011;**44**:118–25.

53. Lord PW, Stevens RD, Brass A, *et al*. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* 2003;**19**:1275–83.

54. Hoehndorf R, Schofield PN, Gkoutos GV. The role of ontologies in biological and biomedical research: a functional perspective. *Brief Bioinform* 2015;**16**:1069–80.

55. Ciccarese P, Soiland-Reyes S, Belhajjame K, *et al*. PAV ontology: provenance, authoring and versioning. *J Biomed Semantics* 2013;**4**:37.

56. Moreau L, Groth P. Provenance: an introduction to PROV. *Synth Lect Semant Web Theory Technol* 2013;**3**:1–129.

57. Scharm M, Waltemath D, Mendes P, *et al*. COMODI: an ontology to characterise differences in versions of computational models in biology. *J Biomed Semantics* 2016;**7**:46.

58. Ruscheinski A, Uhrmacher AM. Provenance in modeling and simulation studies - bridging gaps. In: *Winter Simul. Conf.* 2017, IEEE, Las Vegas, NV, USA.

59. Scharm M, Wolkenhauer O, Waltemath D. An algorithm to detect and communicate the differences in computational models describing biological systems. *Bioinformatics* 2016;**32**:563–70.

60. Büchel F, Rodriguez N, Swainston N, *et al*. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol* 2013;**7**:116.

61. Liechti R, George N, Götz L, *et al*. SourceData: a semantic platform for curating and searching figures. *Nat Methods* 2017;**14**:1021.

62. Peters M, Eicher JJ, van Niekerk DD, *et al*. The JWS online simulation database. *Bioinformatics* 2017;**33**:1589–90.

63. Olivier BG, Snoep JL. Web-based kinetic modelling using JWS online. *Bioinformatics* 2004;**20**:2143–4.

64. König M, Grzegorzewski J. tellurium-web-v0.2.0: an online database and simulation tool for reproducible simulation experiments in computational biology. *Zenodo* 2018. https://doi.org/10.5281/zenodo.1136218.

65. Le NN, Finney A, Hucka M, *et al*. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 2005;**23**:1509–15.

66. Hucka M, Bergmann FT, Hoops S, *et al*. The Systems Biology Markup Language (SBML): language specification for level 3 version 1 Core. *J Integr Bioinform* 2015;**12**:266.

67. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res* 2015;**43**:D204–12.

68. Degtyarenko K, Matos P, Ennis M, *et al*. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2007;D344–50.

69. Ashburner M, Ball CA, Blake JA, *et al*. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–29.

70. Juty N, Le Novère N, Laibe C. Identifiers.org and MIRIAM registry: community resources to provide persistent identification. *Nucleic Acids Res* 2012;**40**:D580–6.

71. Jiang N, Cox RD, Hancock JM. A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic *β* cells. *Mamm Genome* 2007;**18**:508–20.

72. Cooling MT, Hunter P. The CellML metadata framework 2.0 specification. *J Integr Bioinform* 2015;**12**:86–103.

73. Beard DA, Britten R, Cooling MT, *et al*. CellML metadata standards, associated tools and repositories. *Philos Trans R Soc A Math Phys Eng Sci* 2009;**367**:1845–67.

74. Cook DL, Bookstein FL, Gennari JH. Physical properties of biological entities: an introduction to the ontology of physics for biology. *PLoS One* 2011;**6**:e28708.

75. Cook DL, Neal ML, Bookstein FL, *et al*. Ontology of physics for biology: representing physical dependencies as a basis for biological processes. *J Biomed Semantics* 2013;**4**:41.

76. Rosse C, Mejino JLV. A reference ontology for bioinformatics: the foundational model of anatomy. *J Biomed Inform* 2003;**36**:478–500.

77. Mungall CJ, Torniai C, Gkoutos GV, *et al*. Uberon, an integrative multi-species anatomy ontology. *Genome Biol* 2012;**13**:R5.

78. Neal ML, Gennari JH, Cook DL. Qualitative causal analyses of biosimulation models. *CEUR Workshop Proc* 2016;**1747**.

79. Wimalaratne SM, Juty N, Kunze J, *et al*. Uniform resolution of compact identifiers for biomedical data. *Sci Data* 2018;**5**:180029.

80. On the road to robust data citation. *Sci Data* 2018;**5**:180095.

81. Smith B, Ceusters W, Klagges B, *et al*. Relations in biomedical ontologies. *Genome Biol* 2005;**6**:R46.

82. Bornstein BJ, Keating SM, Jouraku A, *et al*. LibSBML: an API library for SBML. *Bioinformatics* 2008;**24**:880.

83. Rodriguez N, Thomas A, Watanabe L, *et al*. JSBML 1.0: providing a smorgasbord of options to encode systems biology models. *Bioinformatics* 2015;**31**:3383–6.

84. Garny A, Hunter PJ. OpenCOR: a modular and interoperable approach to computational biology. *Front Physiol* 2015;**6**:26.

85. Hoops S, Sahle S, Gauges R, *et al*. COPASI—a COmplex PAthway SImulator. *Bioinformatics* 2006;**22**:3067–74.

86. Matsuoka Y, Funahashi A, Ghosh S, *et al*. Modeling and simulation using CellDesigner. *Methods Mol Biol* 2014; **1164**:121–45.

87. King ZA, Lu J, Dräger A, *et al*. BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 2016;**44**:D515–22.

88. Shvaiko P, Euzenat J. Ontology matching: state of the art and future challenges. *IEEE Trans Knowl Data Eng* 2013;**25**: 158–76.

89. Natale DA, Arighi CN, Blake JA, *et al*. Protein Ontology (PRO): enhancing and scaling up the representation of protein entities. *Nucleic Acids Res* 2017;**45**:D339–46.

90. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.

91. Wilkinson MD, Dumontier M. Aalbersberg IjJ, *et al*. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 2016;**3**:160018.

92. Gray AJG, Goble CA, Jiménez R. Bioschemas: from potato salad to protein annotation. *Int Semant Web Conf* 2017.

93. Lister AL, Pocock M, Taschuk M, *et al*. Saint: a lightweight integration environment for model annotation. *Bioinformatics* 2009;**25**:3026–7.

94. Römer M, Eichner J, Dräger A, *et al*. ZBIT bioinformatics toolbox: a web-platform for systems biology and expression data analysis. *PLoS One* 2016;**11**:e0149263.

95. Conant GC, Wolfe KH. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol* 2007;**3**:129.

96. Smith BW, Chase JG, Nokes RI, *et al*. Minimal haemodynamic system model including ventricular interaction and valve dynamics. *Med Eng Phys* 2004;**26**:131–9.

97. Demir E, Cary MP, Paley S, *et al*. The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 2010;**28**:935–42.