Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A BRIEF HISTORY OF COMBINE

Chris J. Myers

University of Utah
50 S. Central Campus Dr., Rm. 2110
Salt Lake City, UT 84112, USA

Gary Bader

University of Toronto
160 College St., Rm. 602
Toronto, ON M5S 3E1, CANADA

Padraig Gleeson

University College London
London - WC1E 6BT
UNITED KINGDOM

Martin Golebiewski

HITS gGmbH
Schloss-Wolfsbrunnenweg 35
69118 Heidelberg, GERMANY

Michael Hucka

California Institute of Technology
1200 E. California Blvd.
Pasadena, CA 91125, USA

Nicolas Le Novère

The Babraham Institute
Cambridge CB22 3LF
UNITED KINGDOM

David P. Nickerson

University of Auckland
70 Symonds Street
Auckland, NEW ZEALAND

Falk Schreiber

University of Konstanz
Universitätsstr. 10
78464 Konstanz, GERMANY

Dagmar Waltemath

University of Rostock
Albert-Einstein-Str. 22
18059 Rostock, GERMANY

ABSTRACT

Standards for data exchange are critical to the development of any field. They enable researchers and practitioners to transport information reliably, to apply a variety of tools to their problems, and to reproduce scientific results. Over the past two decades, a range of standards have been developed to facilitate the exchange and reuse of information in the domain of representation and modeling of biological systems. These standards are complementary, so the interactions between their developers increased over time. By the end of the last decade, the community of researchers decided that more interoperability is required between the standards, and that common development is needed to make better use of effort, time, and money devoted to this activity. The COmputational MOdeling in Biology NEtwork (COMBINE) was created to enable the sharing of resources, tools, and other infrastructure. This paper provides a brief history of this endeavor and the challenges that remain.

1 INTRODUCTION

Standards for data exchange are critical to the development of any field. They support the reliable exchange of information, interoperability of various tools, and reproducibility of scientific results. The COmputational MOdeling in Biology NEtwork (COMBINE) was created to coordinate standardization efforts for modeling in biology (Hucka et al. 2015), initially systems biology but eventually also physiology, neuroscience, and synthetic biology. COMBINE began with four core standards, and it has since expanded to eight, and coordinates with a number of other related standardization efforts. In 2010, the first COMBINE Forum was held in Edinburgh as a satellite event to the Eleventh International Conference on Systems Biology (ICSB). Two years later, the COMBINE Forum began to be held as its own independent event, as it continues to do so today. The COMBINE efforts actually date back to more than a decade earlier, and some of the key milestones are highlighted in Figure 1.

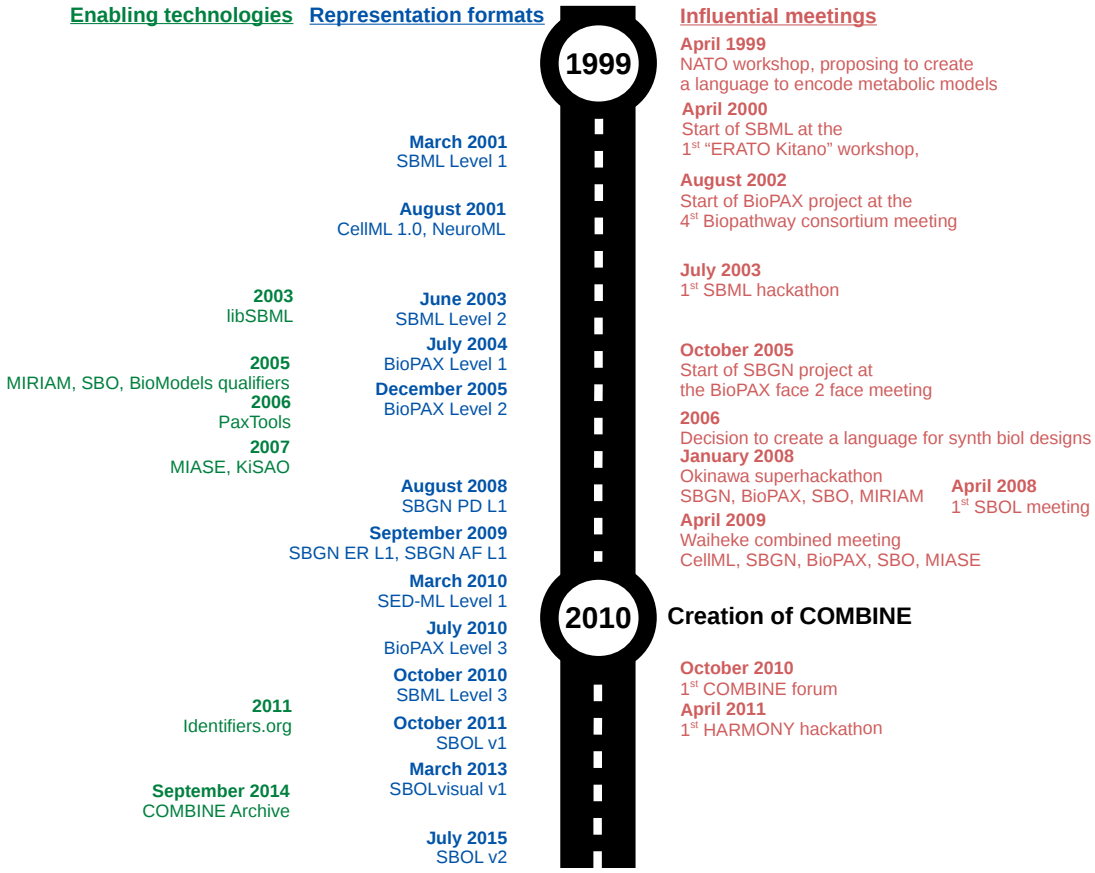

Figure 1: A brief timeline of the development of standards that are currently part of COMBINE. Initial and/or particularly important meetings are shown in red. Main releases of core standards are in blue; for clarity, subsequent revisions and extensions are omitted. The introduction of core supporting resources are in green. This list is not exhaustive, yet shows the community's sustained activity over nearly two decades.

COMBINE brings standard communities together naturally around activities that are mutually beneficial. These activities include making specification documents available from a common location, providing a central point of contact for inquiries about standards relevant to particular domains, and organizing regular face-to-face meetings. In addition to the COMBINE Forum, COMBINE organizes a "hackathon" style event called HARMONY (Hackathon on Resources for Modeling in Biology), as well as tutorials at ICSB

and related meetings. Finally, the COMBINE community interacts closely with scientists to represent state-of-the-art models, e.g. to provide a global map of human metabolism in the *Systems Biology Markup Language* (SBML) (Thiele et al. 2013), or to represent whole cell models in standard formats (Waltemath et al. 2016).

As shown in Figure 2, the standardization efforts in COMBINE are organized around eight core standards. These eight standards can be categorized into three application areas. The *Biological Pathways Exchange* (BioPAX) standard (Demir et al. 2010) and the *Synthetic Biology Open Language* (SBOL) (Galdzicki et al. 2014) are utilized for the exchange of biological knowledge. The *Systems Biology Graphical Notation* (SBGN) (Le Novère et al. 2009) and the *Synthetic Biology Open Language Visual* (SBOLv) (Quinn et al. 2015) specify rules for the exchange of visual representations. SBML (Hucka et al. 2003), *CellML* (Cuellar et al. 2003), *NeuroML* (Gleeson et al. 2010), and the *Simulation Experiment Description Markup Language* (SED-ML) (Waltemath et al. 2011b) encode computational models and their analyses. Finally, there are five associated standardization efforts that are leveraged by multiple core standards. These include the *COMBINE Archive*, which is used for bundling together separate files (Bergmann et al. 2014), *Identifiers.org* and *biomodels.net qualifiers*, which provide annotation infrastructure (Juty et al. 2012), and the *Systems Biology Ontology* (SBO) and the *Kinetic Simulation Algorithm Ontology* (KiSAO), which provide controlled vocabularies (Courtot et al. 2011). Besides these, there are numerous other related but independent standardization efforts (listed at co.mbine.org) that interact with (and may at some future time become part of) COMBINE.

Figure 2: Overview of the COMBINE standards; figure adapted from (Schreiber et al. 2016).

This paper elaborates on COMBINE standards for exchanging biological knowledge (Section 2), visual representations (Section 3), and models and analyses (Section 4), as well as associated standardization efforts that are leveraged by the core standards (Section 5). Finally, this paper discusses open challenges for the field going forward such that COMBINE standards can continue to contribute to reproducible research in computational systems and synthetic biology (Section 6).

2 STANDARDS FOR EXCHANGING BIOLOGICAL KNOWLEDGE

High-throughput experimental methods are generating a tremendous amount of biological data. Standards developed for the purpose of exchanging information learned from these experimental observations are said to encode biological knowledge. Such standards should encode this information in a way that is agnostic about how it is to be further analyzed. In other words, they should not be defined by a precise executable semantics. This type of information may be both structural in nature, such as the sequence of a DNA molecule, or functional in nature, such as an interaction between two proteins. There are two COMBINE core standards described in this section that meet this definition: BioPAX and SBOL.

2.1 Biological Pathways Exchange (BioPAX)

Over 15 workshops from 2002 to 2009 led to the development of three major levels of BioPAX. BioPAX Level 1 (released 2004) supports metabolic pathways, BioPAX Level 2 (2005) added support for molecular interactions, following the PSI-MI model (Hermjakob et al. 2004), and some aspects of signal transduction, such as protein post-translational modifications. BioPAX Level 3 (Demir et al. 2010) added support for gene regulation networks and genetic interactions and improved signal transduction support.

The Biological Pathway Exchange (BioPAX) format can be used to represent biological pathways at the molecular and cellular level. BioPAX facilitates the exchange and collection of pathway data among diverse sources. For instance, a pathway database made available in BioPAX format can be analyzed by a pathway analysis tool that reads BioPAX format, without the need to create new conversion software. Major pathway databases and pathway visualization and analysis tools increasingly support BioPAX as part of a growing community. BioPAX can represent metabolic and signaling pathways, molecular and genetic interactions and gene regulation networks. Metabolic pathways generally contain a set of biochemical reactions, focusing on the chemical modifications made to the small molecule substrates of enzymes. Signaling pathways propagate information across the cell, often via a series of protein covalent modifications, such as protein phosphorylation. Molecular interaction networks capture physical interactions amongst cellular molecules, such as protein-protein or protein-DNA. Genetic interactions connect genes whose combined genetic perturbation leads to a phenotypic result different than expected from the combination of single effects. For instance, in epistasis the phenotype of one gene masks that of another, indicating a possible gene ordering relationship. Gene regulation networks contain relationships between transcription or translation factors and the genes they regulate. More information can be found at biopax.org.

2.2 Synthetic Biology Open Language (SBOL)

Beginning in 2008, SBOL was developed to address the reproducibility problem in synthetic biology (Peccoud et al. 2011) by enabling the exchange of genetic design information. Led largely by Michal Galdzicki, the first version of SBOL was released in October 2011 (Galdzicki et al. 2014). Almost immediately, discussions began for the development of SBOL 2, which would ultimately extend SBOL to express designs composed of components other than DNA, as well as the interactions between these components. While many individuals contributed, Nicholas Roehner led the effort (Roehner et al. 2014). In 2014, SBOL began joining the COMBINE meetings, and in 2015, SBOL became an official COMBINE core standard.

Before SBOL, genetic designs could be exchanged using formats such as FASTA or GenBank, which are more tailored to reverse engineering projects, such as DNA sequencing, rather than forward engineering projects, such as genetic design. The FASTA format is simply a DNA sequence formated as a text string. GenBank, on the other hand, allows the annotation of specific locations on the DNA sequence to identify, for example, *promoters*, *coding sequences*, etc. GenBank, while an improvement, requires the full sequence presented in a flat manner with a limited vocabulary of sequence feature terms. SBOL 1 addresses each of these issues by allowing *DNA components* to be expressed hierarchically, enabling their order to be expressed qualitatively eliminating the need for full sequences, and leverages the *sequence ontology* (SO), which defines more than 2800 sequence feature terms (Eilbeck et al. 2005), to indicate the role of each component (Galdzicki et al. 2014). SBOL 2 generalizes the types of components that can be represented, and it allows for the specification of basic functional information (Bartley et al. 2015, Roehner et al. 2016). In particular, SBOL 2 components can be of type DNA, RNA, protein, etc. SBOL 2 also introduces the concept of a *module* that groups components that perform a desired function, as well as the ability to express *interactions* between these components. The type of these interactions and the roles of the *participants* in these interactions are denoted using terms from the SBO (Courtot et al. 2011), an associated COMBINE standard. Finally, SBOL 2 introduces a means to reference models expressed using a different data standard, such as SBML or CellML, described below. More information can be found at sbolstandard.org.

3 STANDARDS FOR THE EXCHANGE OF VISUAL REPRESENTATIONS

A visualization of a model or pathway can help humans to understand it better. A standard notation for visual representations is unique in that it does not need to include a description of a means to serialize it into a document form. Instead, a visualization standard defines the symbols, connectors, labels, and other components that should be used in a compliant visual representation. There are two COMBINE core standards described in this section that meet this definition: SBGN and SBOLv.

3.1 Systems Biology Graphical Notation (SBGN)

Initiated by Hiroaki Kitano, supported by funding from the Japanese New Energy and Industrial Technology Development Organization, discussions began in 2005 on standardizing the visual representation of biological networks. These discussions led to the development of SBGN. The specification of the SBGN *Process Description* (SBGN-PD) language (Moodie et al. 2015) was released first in 2008, followed by specifications of the SBGN *Entity Relationship* (SBGN-ER) language (Sorokin et al. 2015) and the SBGN *Activity Flow* (SBGN-AF) language (Mi et al. 2015) in 2009.

SBGN is a community-wide attempt to standardize the symbols used in the visual depiction of biochemical networks and cellular processes (Le Novère et al. 2009). SBGN has a well-developed set of symbols and associated semantics. To cover different aspects of biological systems, SBGN defines three different, but complementary, languages (see Fig. 3). The most widely used of these is SBGN-PD. This language describes molecular species and their interactions under consideration of temporal aspects in an unambiguous graphical form. The language categorizes molecule types by appropriately shaped symbols. For example, simple molecules are depicted using a circle, while proteins are depicted using a rectangle with rounded corners. SBGN-PD also includes a clear edge classification, with consistent representation of stimulation, inhibition, etc. The notation is quite extensive and includes provision for representing complexes, protein modification, catalytic reactions, and binding/unbinding of molecular complexes. In addition to the visual notation, there also exists a markup language format, SBGN-ML, that enables software tools to load and save diagrams in SBGN format (van Iersel et al. 2012). A growing number of tools support SBGN, see www.sbgn.org.

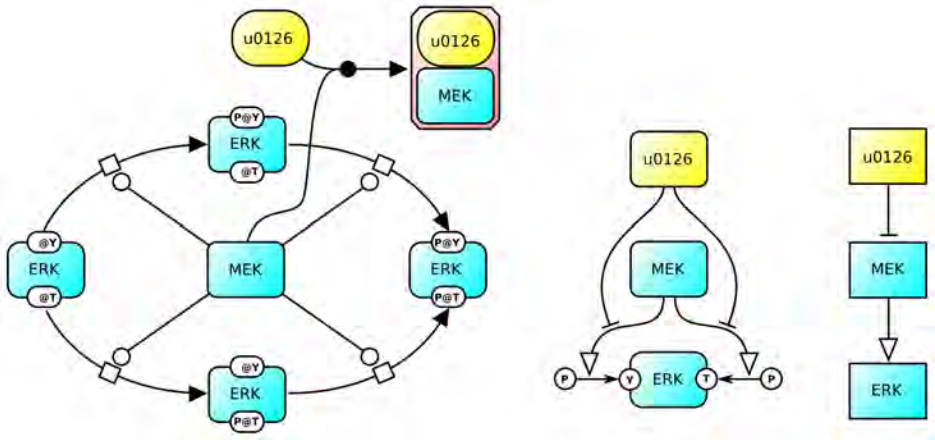

Figure 3: The three SBGN sublanguages providing different views onto the underlying biological system (courtesy of (Le Novère et al. 2009)). This example shows protein phosphorylation catalyzed by an enzyme and modulated by an inhibitor. (*Left to right*.) The SBGN-PD map shows temporal dependencies of biological interactions in detail, the SBGN-ER map displays relationships in which a given entity participates in a network, and the SBGN-AF map shows flow of information between biological entities in a network in an abstract way.

3.2 Synthetic Biology Open Language Visual (SBOLv)

SBOLv was developed from the beginning hand-in-hand with the SBOL knowledge representation format, described earlier. The effort involved many people, but Jackie Quinn carried the first version over the finish line in 2013 (Quinn et al. 2015). SBOLv joined COMBINE alongside SBOL as a core standard in 2015. The first version of SBOLv provides a set of 21 glyphs that represent common DNA components used in genetic designs (Quinn et al. 2015). An example SBOLv diagram generated by the VisBOL software (McLaughlin et al. 2016) is shown in Figure 4. While SBOLv does not have a serialization, conversion is possible from SBOL using a mapping of SO terms to SBOLv glyphs. SBOLv version 2 is actively being developed to visually represent SBOL 2 objects such as non-DNA components and interactions. SBGN already supports these types of objects, so the discussions between these communities within COMBINE are instrumental to ensure that their representations are consistent with each other. More information is available at sbolstandard.org.

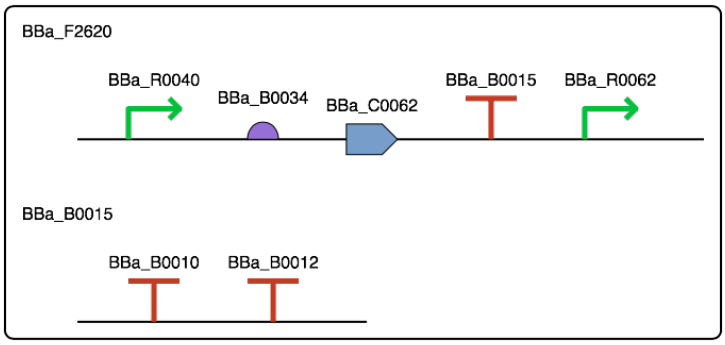

Figure 4: An example genetic design expressed using SBOLv. The bent arrows are *promoters*, the half circle is a *ribosome binding site*, the pointed box is a *coding sequence*, the tee-shape objects are *terminators*. This image was generated by the VisBol software (McLaughlin et al. 2016).

4 STANDARDS FOR THE EXCHANGE OF COMPUTATIONAL MODELS AND ANALYSES

A mathematical model enables a researcher to perform *in silico* experiments (i.e., on the computer rather than in a laboratory). A standard that includes a precise executable semantics that can be leveraged to produce simulations or other types of analyses is said to encode a computational model. There are three COMBINE core standards that meet this definition: SBML, CellML, and NeuroML. Additionally, in order to reproduce an analysis, it is also necessary to encode and exchange information about the analysis to be performed on the model. The last COMBINE core standard, SED-ML, fulfills that role.

4.1 Systems Biology Markup Language (SBML)

SBML grew out of a software interoperability project begun in 1999 at the California Institute of Technology. The community of partners quickly discovered that a means of model exchange would be required to achieve interoperability between the software tools being developed—and thus SBML was created. SBML Level 1 Version 1 was formally released in 2001, and during the subsequent years, SBML underwent a number of significant changes and improvements. The current version is SBML Level 3 Version 2. SBML is a representation format suitable for exchanging and storing formal models of biological processes (Hucka et al. 2003). While it was originally developed by the systems biology community, it has seen application more broadly in synthetic biology and other fields. The most recent generation of SBML is *Level 3*; it provides a modular format consisting of a standalone core and optional *packages* that provide additional constructs. Some SBML Level 3 packages add support for general features such as hierarchical composition (i.e.,

models constructed out of sub-models), and others provide direct support for employing different modeling paradigms such as flux balance analysis, qualitative models, rule-based models, spatial processes, and more. Software libraries such as libSBML (Bornstein, Keating, Jouraku, and Hucka 2008) and JSBML (Dräger et al. 2011) allow developers to implement support for SBML more easily in their software tools, and to date, SBML support has been implemented in over 280 software systems that allow researchers to create, annotate, simulate, visualize and store models. Hundreds of journals accept SBML as a format for supplementary materials, and researchers can also share models by depositing them in the BioModels Database (Chelliah, Laibe, and Novère 2013). More information is available at sbml.org.

4.2 CellML

The CellML language was originally developed at the University of Auckland to share models of cardiac cell dynamics across a number of sites around the world. CellML was designed to leverage existing XML-based languages to describe the mathematics (MathML), metadata (RDF), and links between resources (XLink). The CellML working group collaborated with a number of researchers at Physiome Sciences Inc. to publish the first version of CellML 1.0 in 2001. CellML became a COMBINE core standard in 2015.

CellML is an XML-based language that represents biological and mathematical models (Cuellar et al. 2003). CellML uses MathML for representing equations; and makes use of the Resource Description Framework (RDF) for the description of metadata – such as biological annotations (Cooling and Hunter 2015). It allows the sharing of the models and components within the models and enables reuse of the components. By doing so, existing models can be reused in new and novel modeling studies (Cooling, Nickerson, Nielsen, and Hunter 2016). With the increasing number of CellML models, it is essential to store them in a repository so that the results produced from these published models can be conveniently curated and annotated (Yu et al. 2011, Miller et al. 2011). More information is available at cellml.org.

4.3 NeuroML

Work on NeuroML began in 2001 (Goddard et al. 2001). Initially, NeuroML was designed for a specific software platform, Neosim (Howell et al. 2003). It evolved significantly, and the eventual first official version of NeuroML was made up of a set of related languages (MorphML, ChannelML, NetworkML) for defining the components of neuronal cell and network models (Gleeson et al. 2010). NeuroML is the newest COMBINE core standard, having joined in 2016.

NeuroML is a language in computational neuroscience for expressing models of neurons, ion channels, synapses, populations of cells, and network connections in a standardized form. In contrast to standards like SBML and CellML, it is intentionally domain specific; its scope has been defined by the physiological entities most often included in biophysically detailed neuronal modeling, and supported in neuronal simulators like NEURON (Carnevale and Hines 2006) and GENESIS (Bower and Beeman 1998). NeuroML version 2 (Cannon et al. 2014) is the latest version of the language. A significant addition for version 2 of the language was that it is now defined in terms of the *Low Entropy Model Specification* (LEMS) language. LEMS is a machine readable format which can be used to define the structure/parameters of components in the language (e.g., an Integrate and Fire cell will have parameters for reset voltage, time course of decay, etc.) as well as the dynamics of the components (how the state variable (voltage) of the cell changes in terms of the parameters). Only the structural elements (parameters, hierarchy of components) are required in the NeuroML file, but the LEMS dynamical definitions can be used behind the scenes when code is generated for execution on one of the target simulators. There are actively developed libraries for handling NeuroML 2/LEMS models in Java (jNeuroML) and Python (pyNeuroML & libNeuroML (Vella et al. 2014)). These allow reading and writing, validating files, and converting NeuroML into the target simulators. There is also some support for interaction with other COMBINE formats (e.g. import of SBML into LEMS, export of models into SBML or SED-ML) though many models in NeuroML are outside of the scope of the target supporting simulators for SBML, etc. More information is available at neuroml.org.

4.4 Simulation Experiments Description Markup Language (SED-ML)

In 2007, the idea of developing a standard format for simulation experiment encoding was proposed by Dagmar Waltemath and Nicolas Le Novère. The original version called MIASE-ML was presented at a "Super-hackathon" in Okinawa Japan in 2008, and the first version of SED-ML was officially released in 2011. Its current version is SED-ML Level 1 Version 2.

One key issue in computational biology today is the difficulty for a third-party to reproduce a set of published computational experiments. Curation efforts by the BioModels and CellML model repositories have shown that, for biomolecular models, it is often impossible to reproduce results published based solely on the encoding of the model structure and the written description of the experiment in a publication. This fact is usually due to missing or incorrect information in the text of the paper. Around 2010, when more models became available from open repositories and these models got larger and more complex, discussions started about encoding simulation experiments. In order to encourage reproducibility, the modeling community designed a set of guidelines, the Minimal Information About a Simulation Experiment (MIASE) (Waltemath et al. 2011a). One of the requirements listed by MIASE was to distribute together with the mathematical models, a precise description of the simulations and analyses to perform. SED-ML is an effort to encode this information in a structured format (Waltemath et al. 2011b). SED-ML allows one to describe a simulation experiment performed on a set of computational models expressed in languages such as SBML, CellML, or NeuroML. A SED-ML file points to the various models to simulate, how to change them before simulation, the simulation algorithms to use, the tasks to perform, and, finally, how to process and present the simulation results. More information is available at sed-ml.org.

5 ASSOCIATED STANDARDIZATION EFFORTS

One of the hallmarks of the COMBINE effort has been the development of additional associated standards that can be leveraged by multiple core standards. There are numerous common activities required by all the core standards including the archiving of documents and encoding of metadata. Therefore, the COMBINE community has through our unified efforts created standards in this space that meet the needs of the entire COMBINE community of standards.

One such associated standard is the COMBINE Archive (Bergmann et al. 2014), which was developed to exchange a set of documents that are necessary to fully describe a research project. For instance, to reproduce numerical results, one may need to have access to the model descriptions, in SBML, CellML or NeuroML, as well as the simulation procedures encoded in SED-ML. A model itself can be modular and described in several different files. One may also want to distribute biological pathways in a form amenable to automated reasoning, with BioPAX or SBOL, and also their visual representation in SBGN or SBOLv. The COMBINE Archive provides a means of encapsulating all necessary information using a variety of formats, together with associated metadata, in a single file.

The remaining associated standards relate to the encoding of metadata. These metadata annotations are essential in order for the information encoded within the COMBINE core standards to be useful, findable, and comparable. A fully annotated and curated record in a COMBINE core standard is a valuable resource for the exploration of biological systems, both in research and education. These annotations, for example, can be utilized to enable conversion between COMBINE standards (Büchel et al. 2012, Roehner et al. 2015, Nguyen et al. 2016, Rodriguez et al. 2016). A key element of COMBINE annotations are bio-ontologies. COMBINE maintains two such ontologies. The SBO is utilized to define terms commonly utilized in biological models, such as physical entities, processes, mathematical expressions, and parameters (Courtot et al. 2011). KISAO defines terms used in analysis of models, such as types of simulation algorithms and their parameters (Courtot et al. 2011). COMBINE also maintains a set of persistent identifiers (identifiers.org), which can be used for these and numerous other ontologies (Juty et al. 2012). Finally, the BioModels.net qualifiers link annotations in RDF format to elements inside the standard document, such as an XML file for a model (Juty et al. 2012).

6 ONGOING PROJECTS AND FUTURE OPPORTUNITIES

COMBINE has been highly successful in developing synergistic relationships between many efforts by fostering greater interaction between sub-communities and promoting greater interoperability between standards. Looking toward the future, we see many opportunities to develop even closer coupling between standards, as well as develop new resources.

- *SBML/CellML/SBOL/SED-ML*: A motivating principle of SBOL development has been to avoid reinventing formats when possible. For example, the SBOL community concluded that quantitative modeling is outside the scope of SBOL, since existing standards such as SBML already handle this well. Preliminary efforts have already shown that bridging SBOL and SBML is possible (Roehner and Myers 2013, Roehner et al. 2015, Nguyen et al. 2016). COMBINE can further coordinate these efforts to link SBOL with models represented in SBML or CellML, as well as with descriptions of simulation experiments described using SED-ML.
- *BioPAX/SBOL*: A major element of SBOL 2 is the inclusion of qualitative information about the interactions of DNA, RNA, protein, and other components in a genetic design. BioPAX also encodes similar types of relationships within the pathways they encode. COMBINE can explore conversions between these standards to enable data sharing within these communities.
- *SBGN/SBOL Visual*: Currently, SBOLv is restricted to DNA components, though simple interactions like activation and repression are handled in an ad hoc fashion. Generalized components and interactions within SBOL 2 necessitate an expansion of SBOLv to include these features. SBGN already includes a variety of biochemical and regulatory interactions. COMBINE can help promote greater interaction between SBGN and SBOLv, and if possible, produce a single standard representation format that meets the needs of both system and synthetic biologists.
- *Metadata and Annotations*: metadata and annotations connect data in a standard representation with additional information about that data. COMBINE should develop shared metadata guidelines and specifications, to improve interoperability between standards, as well as to increase consistency between those standards and other sources of biological and clerical information. A shared metadata and annotation framework would also enable the development of common software tools and thus remove the burden from each community to implement its own custom support for annotation.
- *Repositories*: the BioModels database provides a public repository of systems biology models, represented primarily in SBML (Chelliah, Laibe, and Novère 2013). The NormSys registry for modeling standards (normsys.h-its.org) surveys most of the COMBINE standards and compares their major features, possible fields of biological application and use cases. For synthetic biology data, there are also repositories under development, such as the iGEM Registry, JBEI-ICE (Ham et al. 2012), and SynBioHub (formally known as the SBOL Stack, (Madsen et al. 2016)). COMBINE can work to make these repositories interoperable. It should be possible to have genetic parts described in one repository and connect with a model in another. Ultimately, it would be desirable to create a single interface enabling a user to obtain all the necessary data from one information portal.
- *Journals*: The use of standards for DNA sequence data became commonplace when journals began to require them for publication. While there are journals that encourage the use of standards for modeling, there are currently no journals that require it. In order to encourage journals to require these standard data representations, the impact on the authors must be minimized. COMBINE, therefore, should continue to work towards a user friendly portal to their repositories that enables authors to easily deposit their models and designs. These interfaces should allow authors to provide their information in an intuitive way while storing their information using an appropriate standard.

While reproducibility remains a challenge in computational modeling in biology, the COMBINE community and the standards being developed within this community have the potential to make this a fully reproducible scientific endeavor.

ACKNOWLEDGMENTS

CM is supported by the National Science Foundation under Grant No. CCF-1218095 and DBI-1356041. MH is supported by NIGMS award R01070923. GDB is supported by NHGRI award U41HG006623. NLN is supported by BBSRC grant BB/P013384/1. DPN is supported by the Aotearoa Foundation and NIGMS award P50-GM094503. MG is supported by the German Federal Ministry of Education and Research (BMBF) through the Liver Systems Medicine (LiSyM) project and was also supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) through the NormSys project (FKZ 01FS14019). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or our other funding agencies.

REFERENCES

Bartley, B. et al. 2015. "Synthetic Biology Open Language (SBOL) Version 2.0.0". *Journal of Integrative Bioinformatics* 12 (2): 272.

Bergmann, F. T. et al. 2014. "COMBINE Archive and OMEX Format: One File to Share All Information to Reproduce a Modeling Project". *BMC Bioinformatics* 15 (1): 369.

Bornstein, B. J., S. M. Keating, A. Jouraku, and M. Hucka. 2008. "LibSBML: an API Library for SBML". *Bioinformatics* 24 (6): 880–881.

Bower, J., and D. Beeman. 1998. "The Book of Genesis: Exploring Realistic Neural Models with the GEneral NEural SImulation System". *Telos, Springer, New York*.

Büchel, F. et al. 2012. "Qualitative Translation of Relations from BioPAX to SBML Qual". *Bioinformatics* 28 (20): 2648–2653.

Cannon, R. C. et al. 2014. "LEMS: a Language for Expressing Complex Biological Models in Concise and Hierarchical form and its Use in Underpinning NeuroML 2". *Frontiers in Neuroinformatics* 8:79.

Carnevale, N. T., and M. L. Hines. 2006. *The NEURON Book*. Cambridge University Press.

Chelliah, V., C. Laibe, and N. L. Novère. 2013. "BioModels Database: A Repository of Mathematical Models of Biological Processes.". *In silico Systems Biology*:189–199.

Cooling, M. T., and P. Hunter. 2015. "The CellML Metadata Framework 2.0 Specification". *Journal of Integrative Bioinformatics* 12 (2): 260.

Cooling, M. T., D. P. Nickerson, P. M. Nielsen, and P. J. Hunter. 2016. "Modular modelling with Physiome Standards: Modular Modelling with Physiome Standards". *The Journal of Physiology*.

Courtot, M. et al. 2011. "Controlled Vocabularies and Semantics in Systems Biology". *Molecular Systems Biology* 7 (1): 543.

Cuellar, A. et al. 2003. "An Overview of CellML 1.1, a Biological Model Description Language". *Simulation* 79 (12): 740–747.

Demir, E. et al. 2010. "The BioPAX Community Standard for Pathway Data Sharing". *Nature Biotechnology* 28 (9): 935–942.

Dräger, A. et al. 2011. "JSBML: a Flexible Java Library for Working with SBML". *Bioinformatics* 27 (15): 2167–2168.

Eilbeck, K. et al. 2005. "The Sequence Ontology: A Tool for the Unification of Genome Annotations". *Genome Biology* 6 (R44).

Galdzicki, M. et al. 2014. "SBOL: A Community Standard for Communicating Designs in Synthetic Biology". *Nature Biotechnology* 32 (6).

Gleeson, P. et al. 2010. "NeuroML: a Language for Describing Data Driven Models of Neurons and Networks with a High Degree of Biological Detail". *PLoS Computational Biology* 6 (6): e1000815.

Goddard, N. et al. 2001. "Towards NeuroML: Model Description Methods for Collaborative Modelling in Neuroscience". *Philosophical Transactions of the Royal Society B* 356 (1412): 1209–1228.

Ham, T. S. et al. 2012. "Design, Implementation and Practice of JBEI-ICE: an Open Source Biological Part Registry Platform and Tools". *Nucleic Acids Research* 40 (18): e141.

Hermjakob, H. et al. 2004. "The HUPO PSI's Molecular Interaction Formata Community Standard for the Representation of Protein Interaction Data". *Nature Biotechnology* 22 (2): 177–183.

Howell, F. et al. 2003. "Linking Computational Neuroscience Simulation Tools—a Pragmatic Approach to Component-based Development". *Neurocomputing* 52:289–294.

Hucka, M. et al. 2003. "The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models". *Bioinformatics* 19 (4): 524–531.

Hucka, M. et al. 2015. "Promoting Coordinated Development of Community-based Information Standards for Modeling in Biology: the COMBINE Initiative". *Frontiers in Bioengineering and Biotechnology* 3:19.

Juty, N. et al. 2012. "Identifiers.org and MIRIAM Registry: Community Resources to Provide Persistent Identification". *Nucleic Acids Research* 40 (D1): D580–D586.

Le Novère, N. et al. 2009. "The Systems Biology Graphical Notation". *Nature Biotechnology* 27:735–41.

Madsen, C. et al. 2016. "The SBOL Stack: A Platform for Storing, Publishing, and Sharing Synthetic Biology Designs". *ACS Synthetic Biology* 5 (6): 487–497.

McLaughlin, J. et al. 2016. "VisBOL: Web-based Tools for Synthetic Biology Design Visualization". *ACS Synthetic Biology* 5 (8): 874–876.

Mi, H. et al. 2015. "Systems Biology Graphical Notation: Activity Flow Language Level 1 Version 1.2". *Journal of Integrative Bioinformatics* 12 (2): 340–381.

Miller, A. K. et al. 2011, Jan. "Revision History Aware Repositories of Computational Models of Biological Systems". *BMC Bioinformatics* 12.

Moodie, S. et al. 2015. "Systems Biology Graphical Notation: Process Description Language Level 1 Version 1.3". *Journal of Integrative Bioinformatics* 12 (2): 213–280.

Nguyen, T. et al. 2016. "A Converter from the Systems Biology Markup Language to the Synthetic Biology Open Language". *ACS Synthetic Biology* 5 (6): 479–486.

Peccoud, J. et al. 2011. "Essential Information for Synthetic DNA Sequences". *Nature Biotechnology* 29 (22).

Quinn, J. et al. 2015. "SBOL Visual: a Graphical Language for Genetic Designs". *PLoS Biology* 13 (12).

Rodriguez, N. et al. 2016. "The Systems Biology Format Converter". *BMC Bioinformatics* 17 (1): 154.

Roehner, N. et al. 2014. "Proposed Data Model for the Next Version of the Synthetic Biology Open Language". *ACS Synthetic Biology* 4 (1): 57–71.

Roehner, N. et al. 2015. "Generating Systems Biology Markup Language Models from the Synthetic Biology Open Language". *ACS Synthetic Biology* 4 (8): 873–879.

Roehner, N. et al. 2016. "Sharing Structure and Function in Biological Design with SBOL 2.0". *ACS Synthetic Biology* 5 (6): 498–506.

Roehner, N., and C. J. Myers. 2013. "A Methodology to Annotate Systems Biology Markup Language Models with the Synthetic Biology Open Language". *ACS Synthetic Biology* 3 (2): 57–66.

Schreiber, F. et al. 2016. "Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2016". *Journal of Integrative Bioinformatics* 13 (3): 289.

Sorokin, A. et al. 2015. "Systems Biology Graphical Notation: Entity Relationship Language Level 1 Version 2". *Journal of Integrative Bioinformatics* 12 (2): 281–339.

Thiele, I. et al. 2013. "A Community-driven Global Reconstruction of Human Metabolism". *Nature Biotechnology* 31 (5): 419–425.

van Iersel, M. P. et al. 2012. "Software Support for SBGN Maps: SBGN-ML and LibSBGN". *Bioinformatics* 28 (15): 2016–2021.

Vella, M. et al. 2014. "libNeuroML and PyLEMS: Using Python to Combine Procedural and Declarative Modeling Approaches in Computational Neuroscience". *Frontiers in Neuroinformatics* 8:38.

Waltemath, D. et al. 2011a. "Minimum Information About a Simulation Experiment (MIASE).". *PLoS Computational Biology* 7:e1001122.

Waltemath, D. et al. 2011b. "Reproducible Computational Biology Experiments with SED-ML-the Simulation Experiment Description Markup Language". *BMC Systems Biology* 5 (1): 198.

Waltemath, D. et al. 2016. "Toward Community Standards and Software for Whole-Cell Modeling". *IEEE Transactions on Biomedical Engineering* 63 (10): 2007–2014.
Yu, T. et al. 2011. "The Physiome Model Repository 2". *Bioinformatics* 27 (5): 743–744.

AUTHOR BIOGRAPHIES

CHRIS J. MYERS has served as an editor for the SBML standard and is on the steering committee for the SBOL standard. His research group develops numerous software tools that utilize and support SBML, SED-ML, and SBOL, including, iBioSim, SBOLDesigner, libSBOLj, and the SBOL Validator/Converter. He is the COMBINE coordinator for SBOL and SBOLv. His email address is myers@ece.utah.edu.

GARY D. BADER works on biological network analysis and pathway information resources. He has been involved in leading development of protein interaction and pathway databases and standards, including the BioPAX biological pathway exchange language, and Pathway Commons. He is the COMBINE coordinator for BioPAX. His email address is gary.bader@utoronto.ca.

PADRAIG GLEESON is a member of the NeuroML Editorial Board, and has been an active technical contributor to the standard for many years. He is also the main developer of the application neuroConstruct and is project manager of the Open Source Brain Initiative. He is the COMBINE coordinator for NeuroML. His email address is p.gleeson@ucl.ac.uk.

MARTIN GOLEBIEWSKI is involved in data management projects for systems biology and systems medicine, and in the reaction kinetics database SABIO-RK. He is part of the CHARME network, chairs the data processing and integration working group of the ISO committee for biotechnology (ISO/TC 276) and is the COMBINE coordinator for meta-data standards. His email address is martin.golebiewski@h-its.org.

MICHAEL HUCKA is co-author of numerous software libraries and systems such as libSBML, JSBML, and MOCCASIN, and the maintainer of SBML-related resources. such as the SBML.org portal. Dr. Hucka is the SBML Team leader, and was one of the original SBML Editors as well as a co-founder of COMBINE. He is the COMBINE coordinator for SBML. His email address is mhucka@caltech.edu.

NICOLAS LE NOVÈRE coordinated the development of key software tools to support computational systems biology research, such as BioModels and Identifiers.org, and was a key developer of standards, including SBML, SBGN and the MIRIAM guidelines. He was one of the co-founders of COMBINE. His email address is n.lenovere@gmail.com.

DAVID P. NICKERSON leads the Auckland Renal Physiome project, and he is involved in the development and application of the CellML and SED-ML standards and related software tools. He is also involved in the Physiome Model Repository. He is the COMBINE coordinator for CellML. His email address is d.nickerson@auckland.ac.nz.

FALK SCHREIBER works on immersive analytics of biological data, network science for biological systems, graphical standards for systems biology, and modeling of metabolism. His group develops standards-related software such as Vanted and SBGN-ED. He served twice as SBGN editor and is the COMBINE coordinator for SBGN. His email address is falk.schreiber@uni-konstanz.de.

DAGMAR WALTEMATH works on strategies for data and model management in computational biology. She has been actively involved in the development of SED-ML, SBML, and the COMBINE Archive format. She is the COMBINE coordinator for SED-ML. Her email address is dagmar.waltemath@uni-rostock.de.