

# COMBINE 2020: Opening Session

Chris J. Myers (chair), Dagmar Waltemath (vice-chair)

Gary Bader, Padraig Gleeson, Martin Golebiewski,

Sarah Keating, Matthias König, David P. Nickerson, and Falk Schreiber

COMBINE 2020  
October 5, 2020



# COMBINE Coordination Board



**Chris Myers**  
University of  
Colorado, USA  
SBOL



**Dagmar Waltemath**  
University Medicine  
Greifswald, Germany  
Annotations



**Gary Bader**  
University of  
Toronto, Canada  
BioPAX



**Padraig Gleeson**  
University College  
London, UK  
NeuroML



**Martin Golebiewski**  
HITS gGmbH,  
Germany



**Tom Gorochowski**  
University of  
Bristol, UK  
SBOL Visual



**Sarah Keating**  
University College  
London, UK  
SBML



**Matthias König**  
Humboldt-University  
Berlin, Germany  
SED-ML



**David Nickerson**  
University of  
Auckland, NZ  
CellML



**Falk Schreiber**  
University of  
Konstanz, Germany  
SBGN

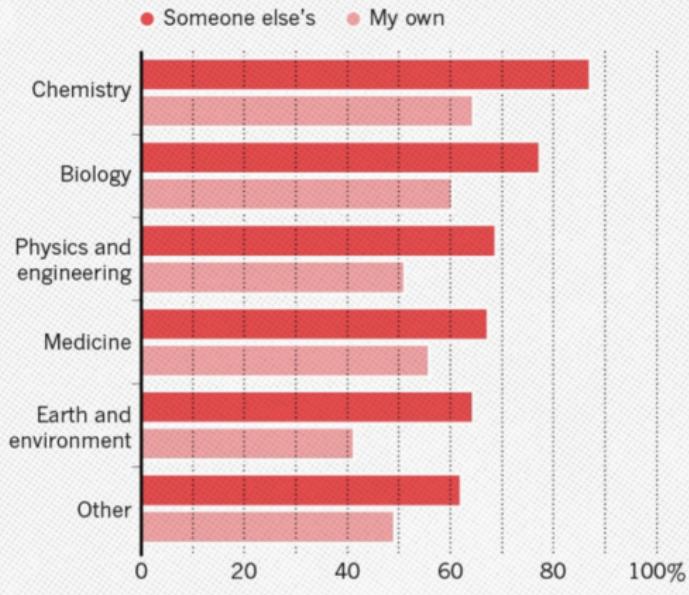
# COMBINE 2020

- Over 250 registrants from all over the world.
- 7 invited talks, 36 contributed talks, 10 lightning talks, 14 tutorials, and 22 breakouts.
- A 5-day 24 hours per day schedule:
  - Monday - live talks with 1 hour social breaks every 3 hours.
  - Tuesday - replays of talks except community session at 18:00 UTC.
  - Wednesday, Thursday, Friday - breakouts and tutorials.
- Please join our Slack channel ([combine-workspace.slack.com](https://combine-workspace.slack.com)).
- Note the common area Zoom channel for connecting with people.

# Reproducibility Crisis

## HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



(V. Simonyan, Center for Biologics Evaluation and Research FDA, USA)

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.*

– David Donoho, 1998

## Standards to the Rescue



(source <https://www.eaglegenomics.com/do-data-standards-really-matter/>)

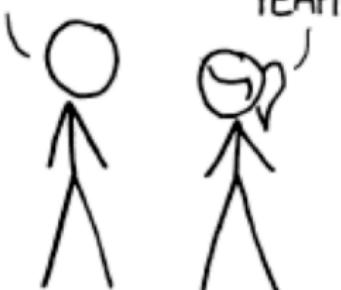
# Word of Warning

## HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:  
THERE ARE  
14 COMPETING  
STANDARDS.

14?! RIDICULOUS!  
WE NEED TO DEVELOP  
ONE UNIVERSAL STANDARD  
THAT COVERS EVERYONE'S  
USE CASES.



SOON:

SITUATION:  
THERE ARE  
15 COMPETING  
STANDARDS.

(source [xkcd.com](http://xkcd.com))

# Coordination of Standard Development in Systems/Synthetic Biology

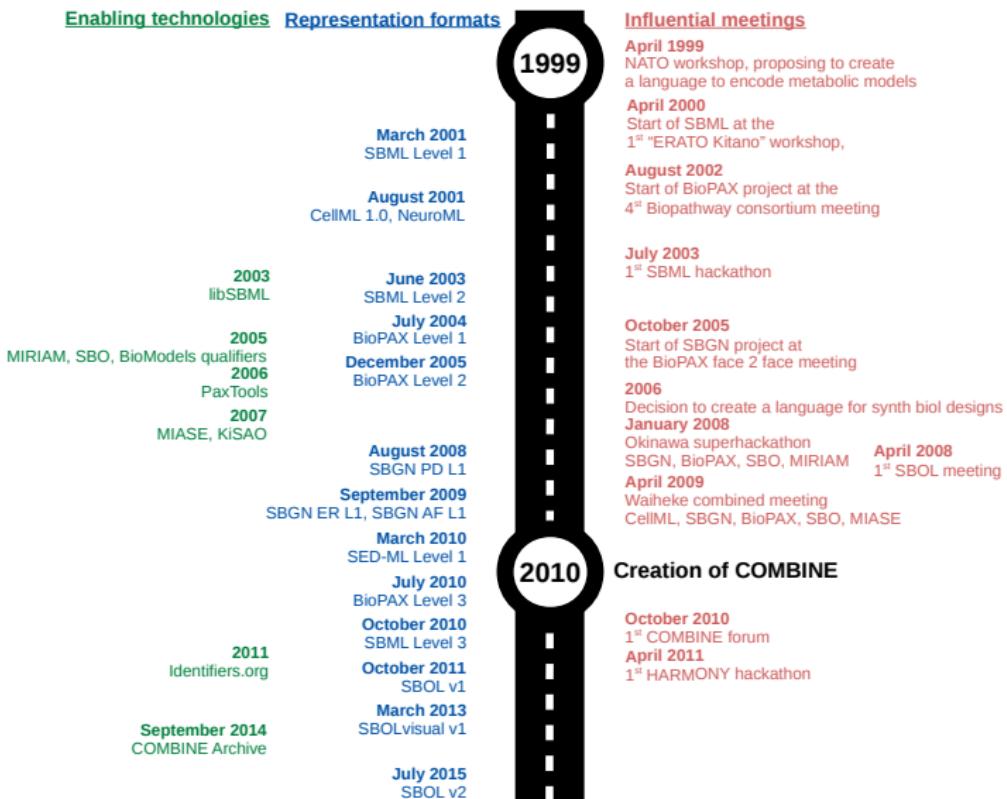


<http://co.mbine.org>

- Tasks and Actions:
  - Computational Modeling in Biology Network
  - Organizes meetings of standards: the COMBINE Forum & HARMONY
  - Training in application of standards (COMBINE tutorials)
  - Coordinate standards development
  - Develop common procedures and tools
  - Provide a recognized voice
  - Annual special issue in the Journal of Integrative Bioinformatics

# COMBINE History

## Enabling technologies    Representation formats



# COMBINE Overview

Core Standards

Standards for Knowledge Representation

**BioPAX**



Standards for Visual Representation



Standards for Models and their Analyses



Associated Standards

Used by core standards

Projects



Infrastructure

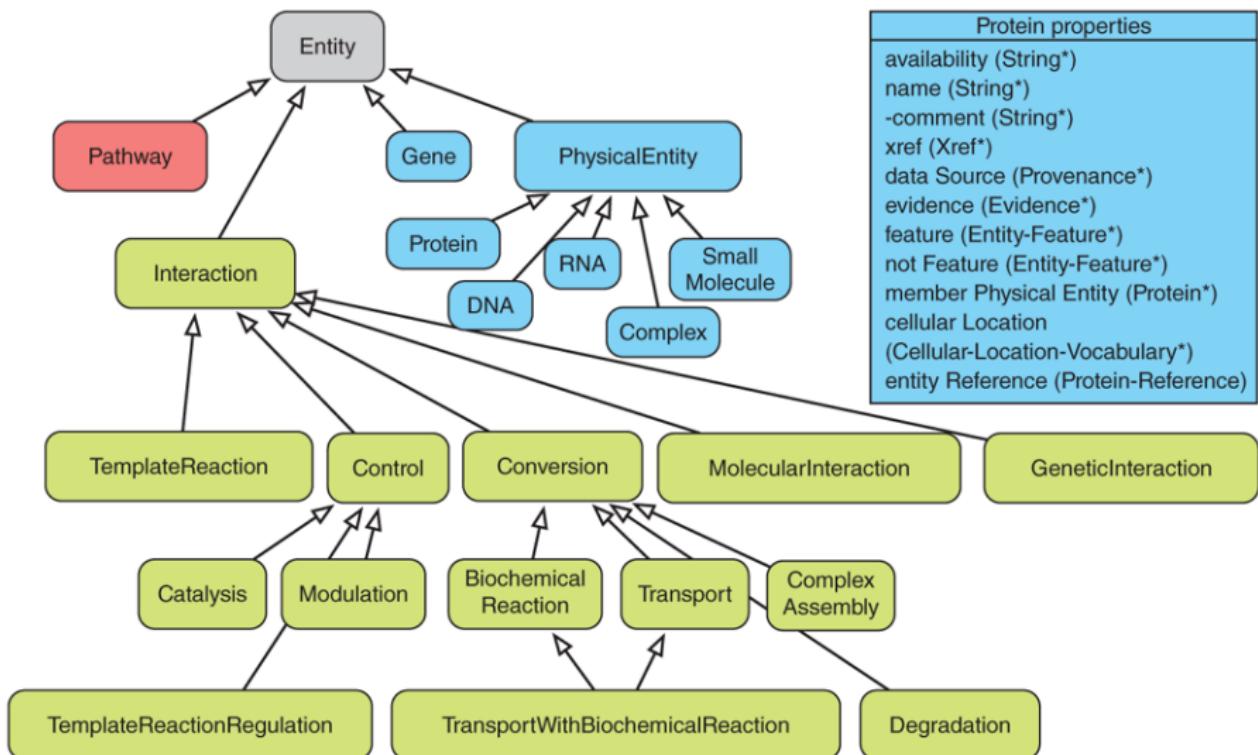
*BioModels.net* qualifiers



Controlled Vocabularies

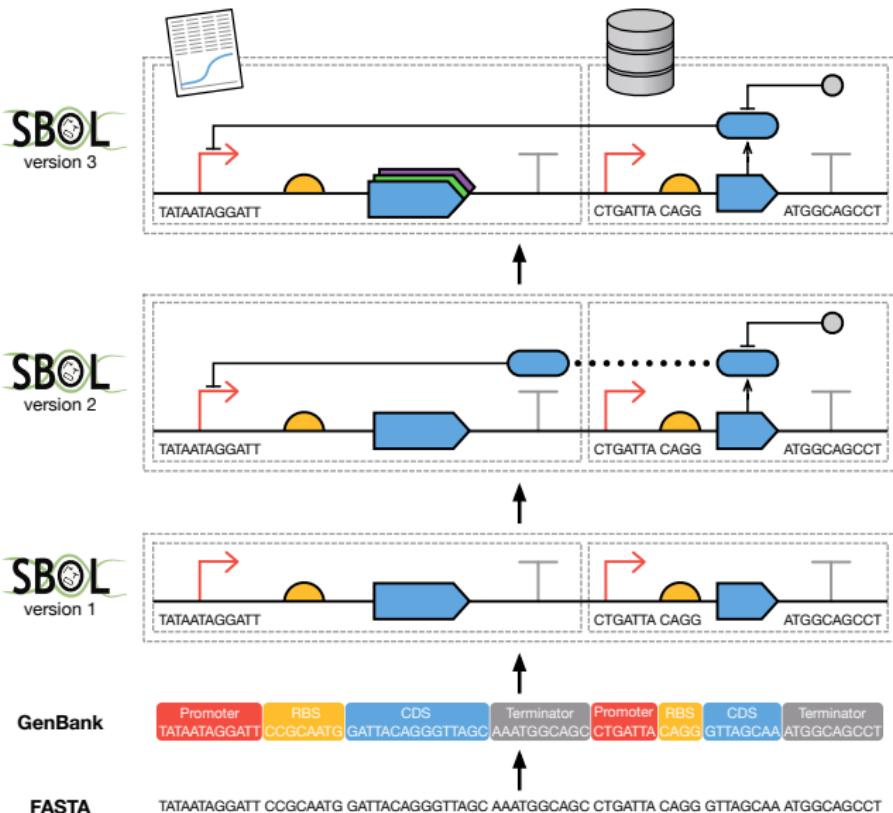


# BioPax: Biological Pathways



Demir et al., Nature Biotechnology (2010)

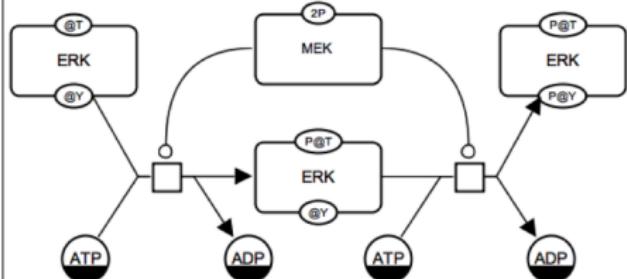
# Synthetic Biology Open Language (SBOL)



# Systems Biology Graphical Notation (SBGN)

## Process Description

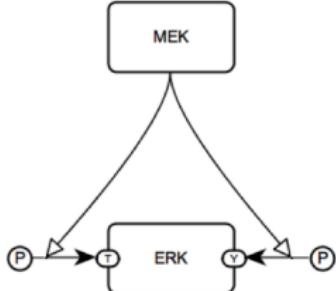
### maps



- ▶ Unambiguous
- ▶ Mechanistic
- ▶ Sequential
- ▶ Combinatorial explosion

## Entity Relationships

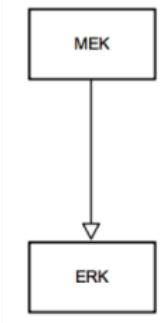
### maps



- ▶ Unambiguous
- ▶ Mechanistic
- ▶ Non-Sequential

## Activity Flow

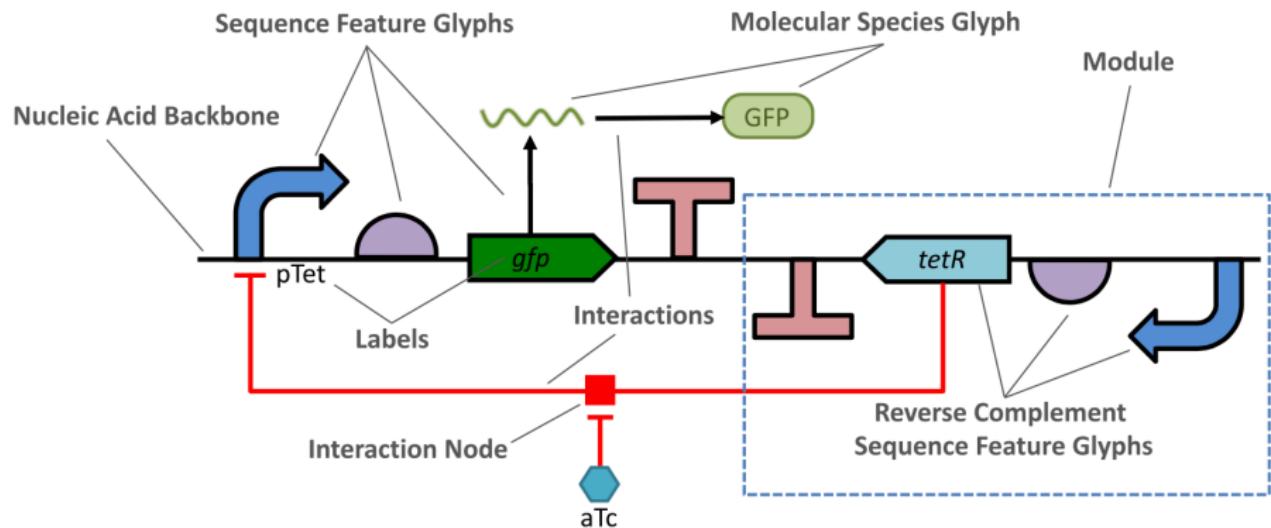
### maps



- ▶ Ambiguous
- ▶ Conceptual
- ▶ Sequential

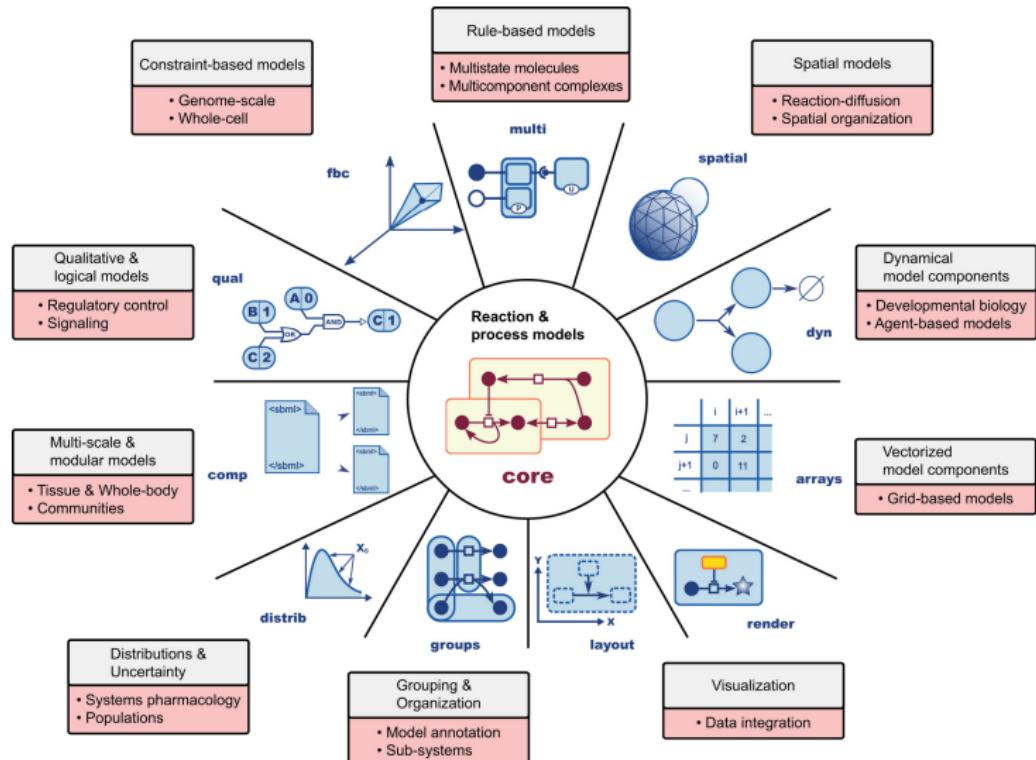
Le Novère et al., Nature Biotechnology (2009)

# SBOL Visual (SBOLv)



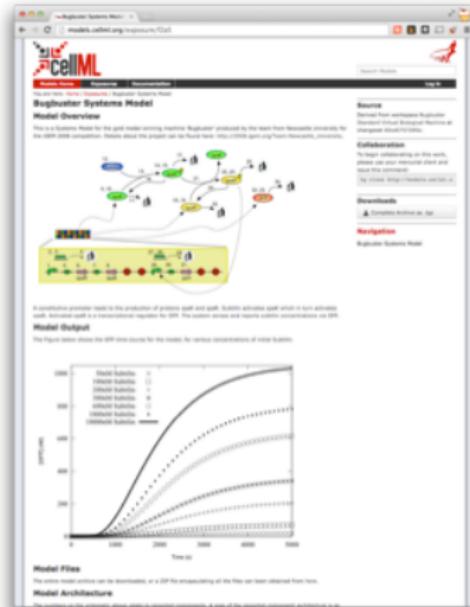
Beal et al., ACS Synthetic Biology (2019)

# Systems Biology Markup Language (SBML)



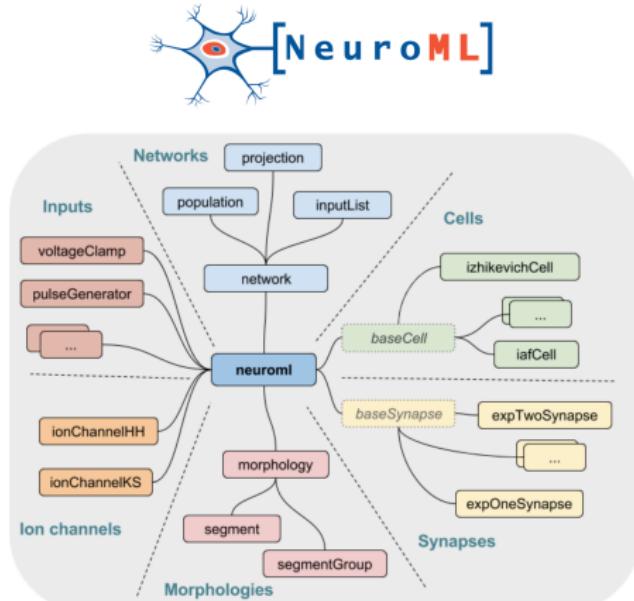
Keating, Waltemath, et al., Molecular Systems Biology (2020)

# CellML

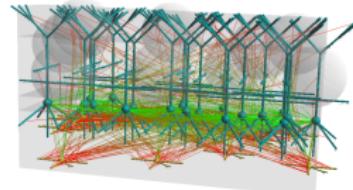
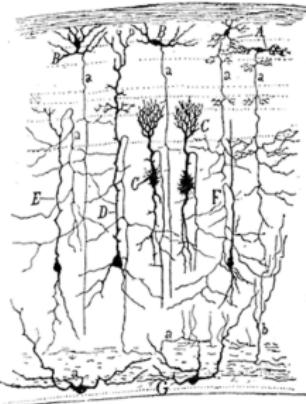


Cuellar et al., Simulation (2003)

# NeuroML: Computational Neuroscience Models

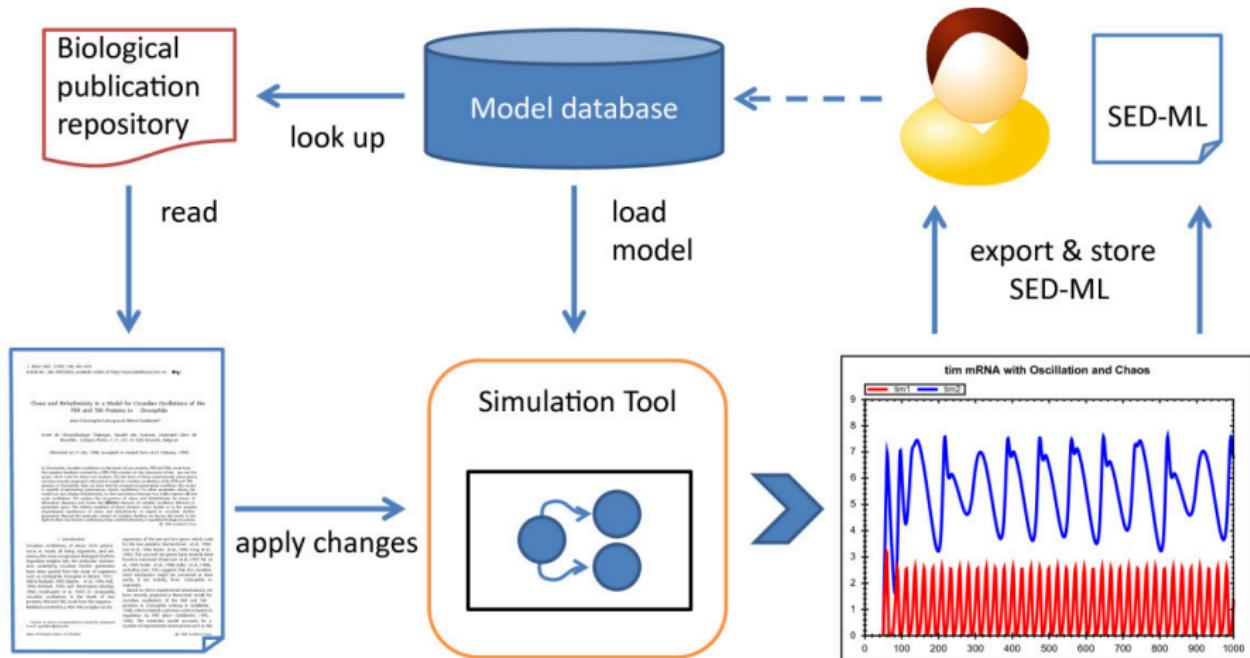


The NeuroML language includes all the elements required to construct models biological neuronal circuits



Cannon et al., Frontiers in Neuroinformatics (2014)

# Simulation Expt. Description Markup Lang. (SED-ML)



Waltemath et al., BMC Systems Biology (2011)

# COMBINE Archive

**COMBINE Archive format =**  
single file that supports exchange  
of all information necessary for any  
modeling and simulation  
experiment

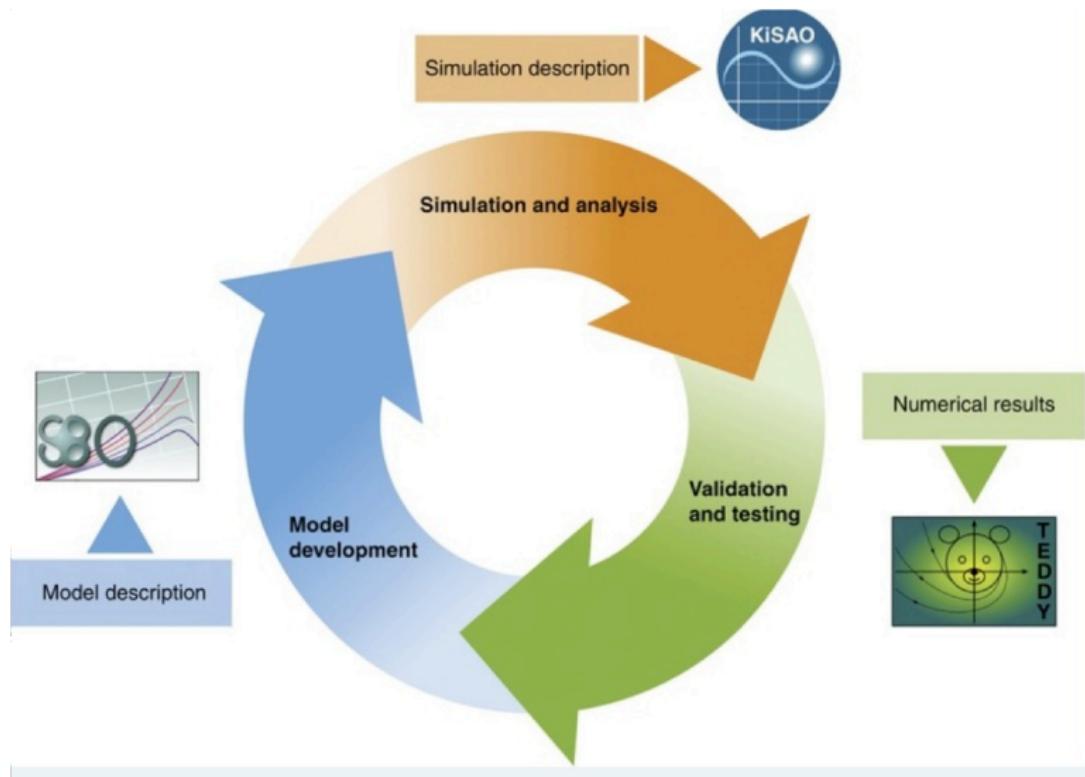
- Not SBML-specific at all
- Not programming-language specific
- Not domain specific



**OMEX** = file format for COMBINE Archive

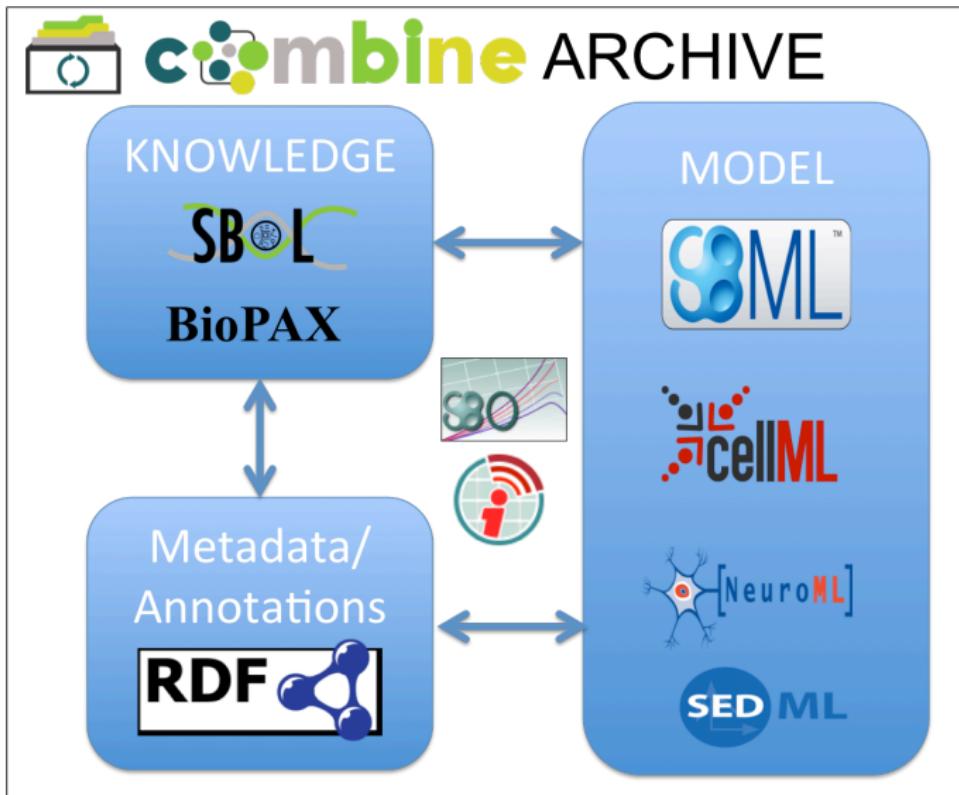
- ZIP file containing manifest file (in XML form) + other files
- Use of ZIP leverages many existing programming libraries

# Ontologies and Controlled Vocabulary

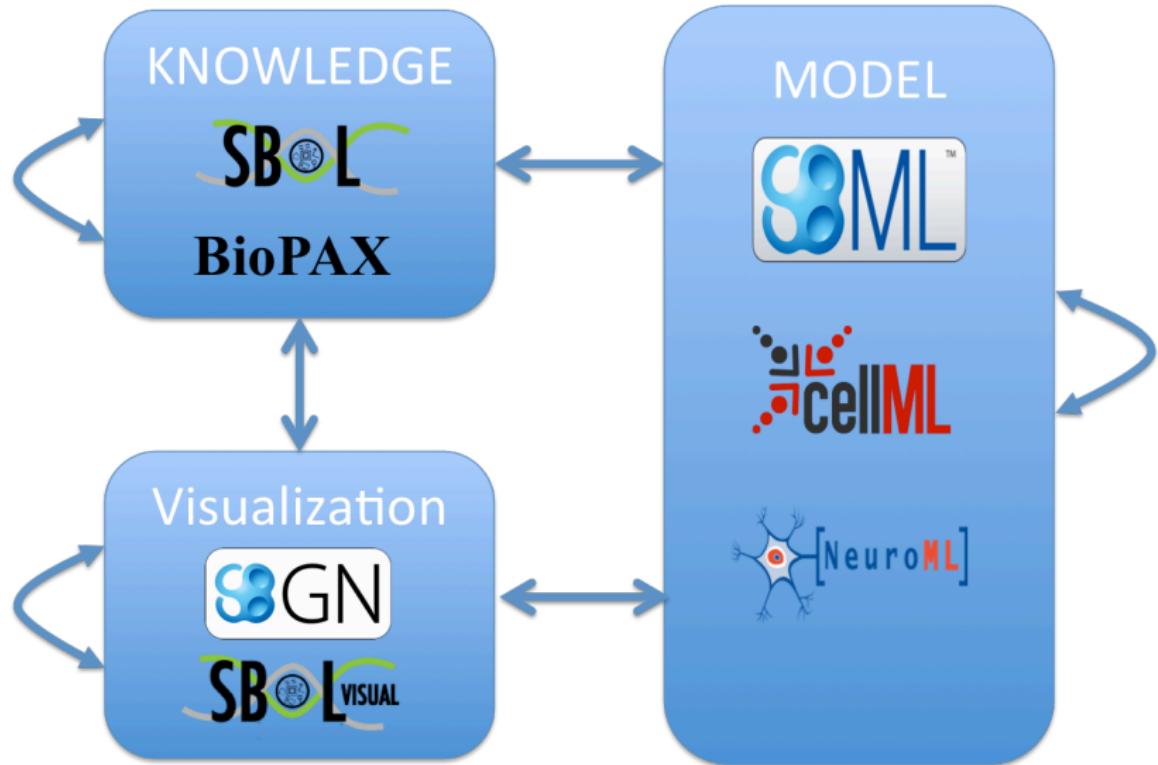


Courtot, et al., Mol Syst Biol. (2011)

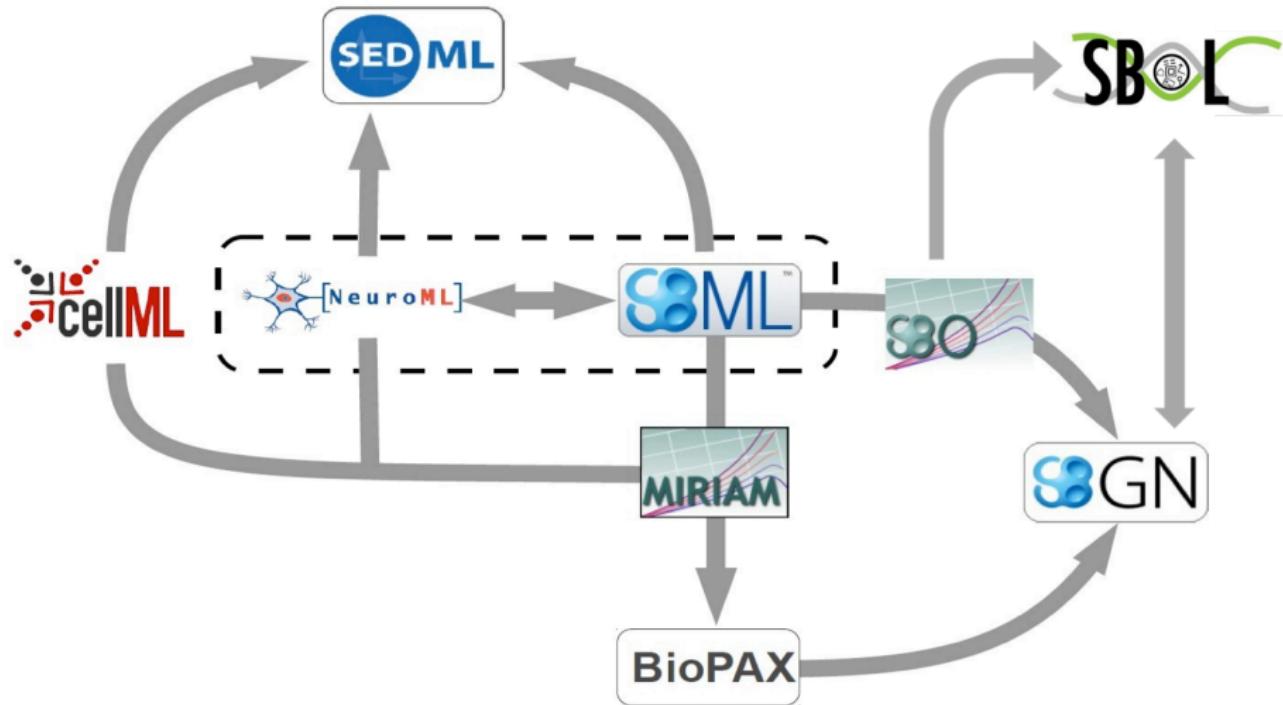
# Unified Metadata and Annotations



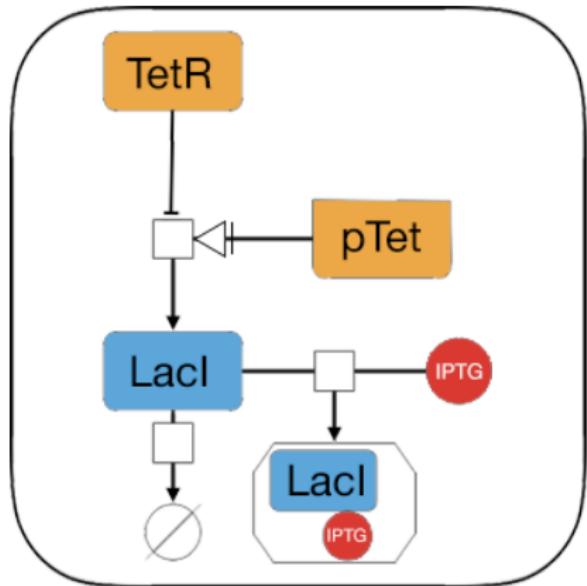
# Data Standard Conversions



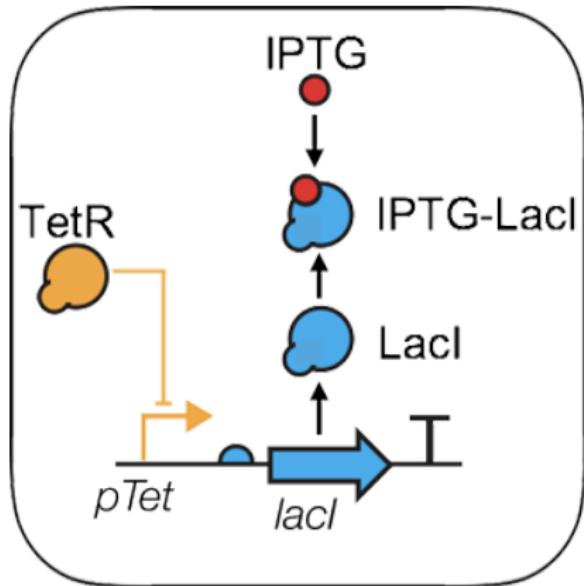
# Interfacing and Interoperability of Standards



# Unified Visualizations (SBGN/SBOLv)

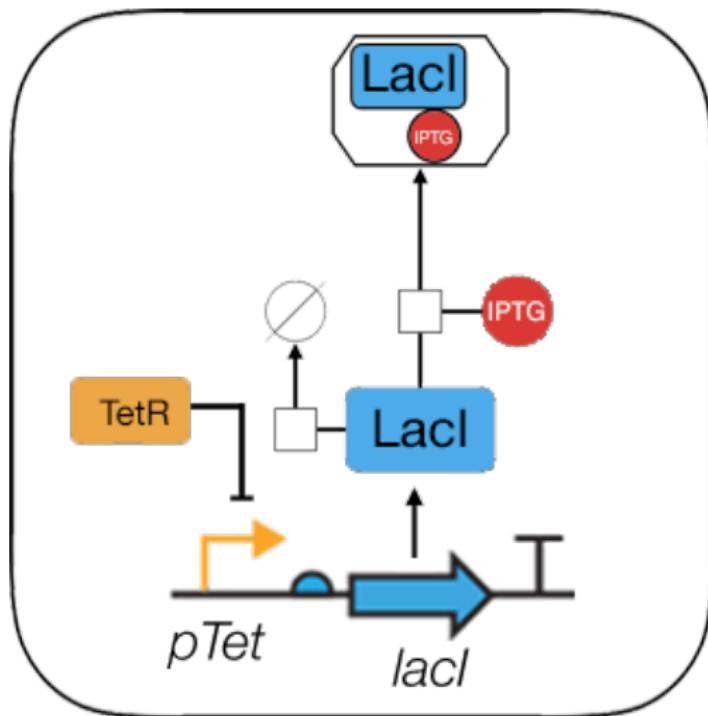


SBGN



SBOLv

# Unified Visualizations (SBGN/SBOLv)



Unified SBGN/SBOLv

## Repositories

 **Pathway Commons**  
 Access and discover data integrated from public pathway and interactions databases.  
 Pathway Commons, a web resource for biological pathway data.  
Data Tools FAQ Contact

Apps

<b>Search</b> Search the entire collection of pathways <input type="text" value="Names or gene IDs (e.g. 'glycolysis', 'TP53')"/> <span style="font-size: 2em;">Q</span>	<b>PCViz</b> Get details about genes and their interactions <input type="text" value="Gene IDs (e.g. MDM2 TP53)"/> <span style="font-size: 2em;">Q</span>
--	---

<http://www.pathwaycommons.org>

The screenshot shows the homepage of the Open-Source Brain website. It features a large green globe icon with the text "Modelling the brain, together". Below the header, there's a sub-headline "Open-Source Brain is a resource for sharing and collaboratively developing computational models of neural systems." To the left, a box titled "Learn more about the OSB" contains a red brain image and a link to "the OSB website". In the center, a box titled "Learn about the Douglas-Hallay model" shows a black brain image with a yellow circle and a link to "the model". To the right, a box titled "Simulate electrophysiologically detailed cell models" shows a dark brain image and a link to "the simulator". At the bottom, there's a call-to-action button "Or create an account to add your own models and run simulations!" followed by social media icons for GitHub, Bitbucket, and LinkedIn. Below this, a section titled "Try the new Neurophysiology Without Borders PAVIS Explorer" has a link to "run it here on HIVE". A note at the bottom states "The Open-Source Brain® page has been published by NeuroConstruct Dawson et al. 2010".

<http://opensourcebrain.org>

The screenshot shows the SynBioHub website. At the top, there's a navigation bar with icons for home, submit, about, and submissions. Below it is a large header with the SynBioHub logo and name. The main content area has three main sections: 1) A search section titled "Search for useful parts and designs" with a magnifying glass icon, a search input field, and a "Search" button. It also includes a "Browse Public Designs" link. 2) An "Upload your design for safekeeping" section with an "Upload Design" button. 3) A "Share designs for publication or collaboration" section with a "Manage Submissions" button.

<https://synbiohub.org>

**BioModels Database**

Model Search      Model List      Model Submission      Contact

The BioModels Database is a repository of mathematical models of biological processes. Models described from literature are manually curated and enriched with cross-references. All models are provided to the public domain. More information about BioModels Database can be found in the FAQ.

Model entries in the BioModels Database

- Curated (10,000 models)
- Published (1,000 models)
- Submitted (100,000 models)

Recent dataset genome-scale metabolic models

- [1] Human cell cycle model (1,000 metabolites)
- [2] Human cell cycle model (1,000 metabolites)

Models automatically generated from pathway resources (metabolic)

- [1] Human cell cycle model (1,000 metabolites)
- [2] Human cell cycle model (1,000 metabolites)

Models (111,000 models)

Non-metabolic (17,231 models)

Multi-gene (1,000 models)

Model search

Advanced search

Teaching

Alternative views

Disease Ontology (Diseases)

Bone Ontology (Bones)

Advanced search

Search

---

Model of the month

CellNet 2011 (first published genome-scale metabolic model of the yeast cell cycle) was selected as the model of the month for January 2012. The model describes the cell cycle progression of yeast using a genome-scale metabolic model.

Access to the model of the month

Notice

2012-01-01 Patient derived genome scale metabolic models now fully available. Model based phenotypic platemaps

2012-01-01 "Model of the month": CellNet 2011 (first published genome-scale metabolic model of the yeast cell cycle) was selected as the model of the month for January 2012. The model describes the cell cycle progression of yeast using a genome-scale metabolic model.

2012-01-01 Metabolic network analysis reveals metabolic pathways associated with cell cycle genes

2012-01-01 Model on neurogenesis reveals gene activity patterns associated with cell cycle genes

<http://biomodels.net>

**Source**  
Kumar et al., *Nature*, 2006; 440: 103-106. © 2006 Nature Publishing Group

**Collaboration**  
Highly collaborative on this work, Kumar et al. also worked on other projects involving genetic oscillations.

**Downloads**  
[Download PDF](#) | [Download XML](#) | [Download ZIP](#)

**Article Available**

**Abstract**  
Abstract: Networks of interacting proteins carry out many complex functions in living cells, but the design principles underlying these networks remain elusive. Here we present a computational approach to the problem—the design of synthetic gene networks that exhibit sustained oscillations.

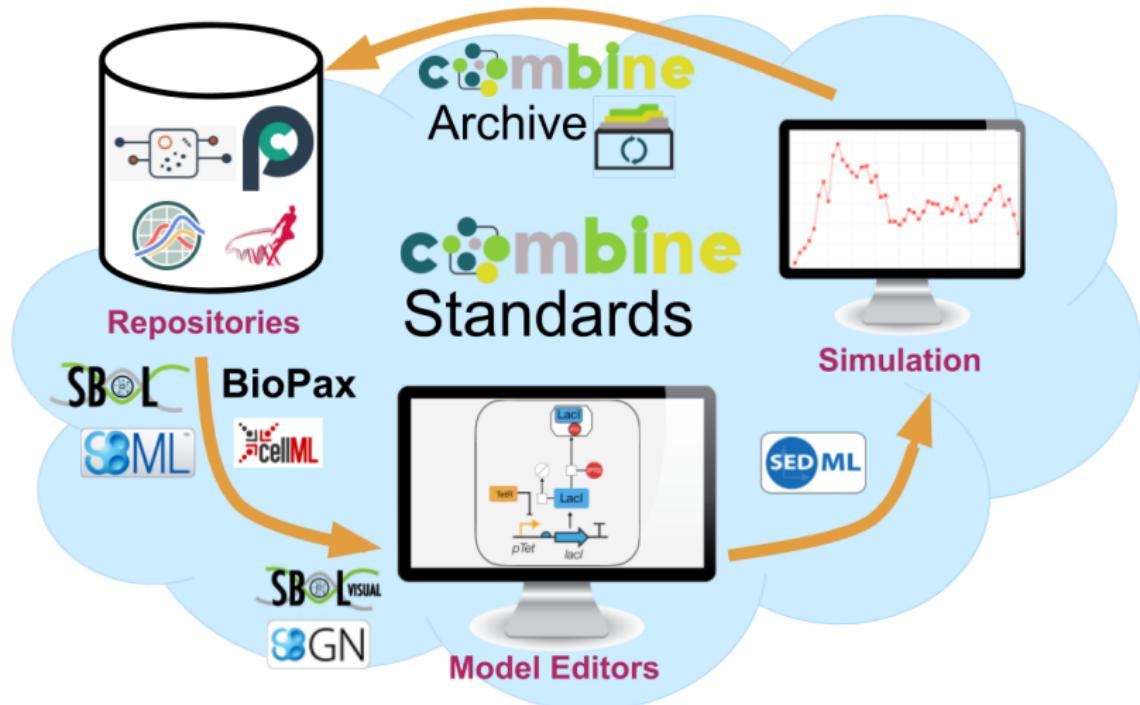
We used a transcriptional regulatory system that did not yet have any natural interest to build in oscillating behaviour. We chose to work with the yeast *Saccharomyces cerevisiae*. The resulting oscillations, with typical periods of hours, are slower than the circadian oscillations in plants and animals, but faster than those of most bacteria. The oscillations are driven by a negative feedback loop involving periodic overexpression of identical Realizations of the components. Such 'self-reinforcing' design may help to explain the prevalence of oscillations in biological systems.

The complete original paper reference is:

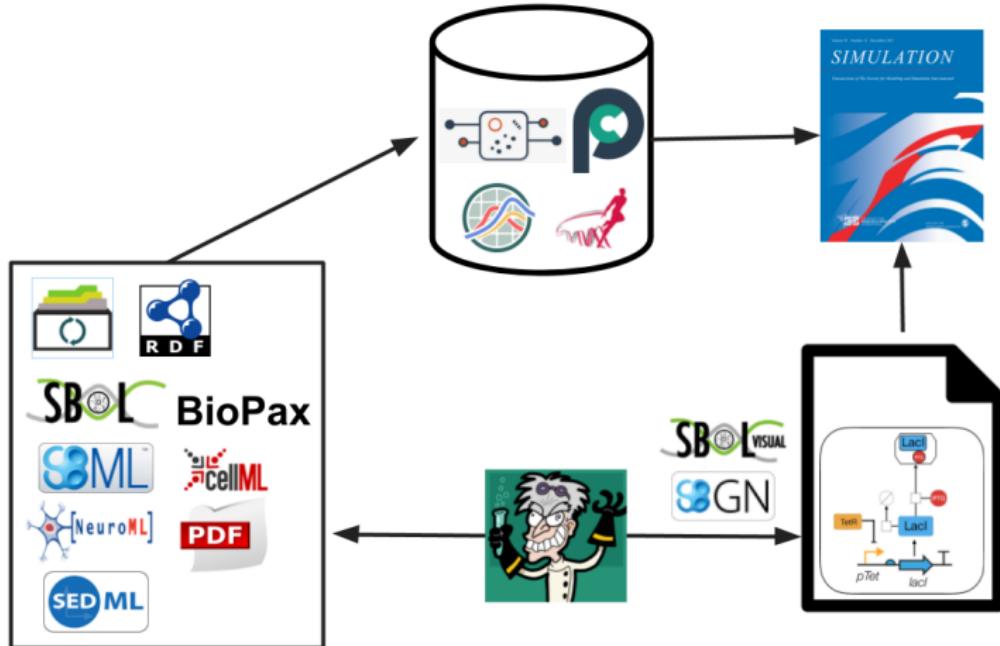
Kumar et al., *A synthetic network of transcriptional regulators, Mihir B. Shetty and Sanjivani Joshi, 2006, Nature Biotechnology, Volume 24, Number 2, pages 103-106, March 2006*.

<https://models.physiomeproject.org>

# Standard Enabled System/Synthetic Biology Workflows



# Journal Workflow for Reproducibility



# Conclusion

*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.*