A status update on COMBINE standardization activities, with a focus on SBML

Michael Hucka, Ph.D.

Department of Computing + Mathematical Sciences
California Institute of Technology
Pasadena, CA, USA

Email: mhucka@caltech.edu *Twitter: @mhucka*

NIH IMAG Model and Data Sharing Working Group webinar, 24 Jan. 2013

Outline

Introduction and motivation

COMBINE

SBML

SED-ML

SBGN

Conclusion

Outline

Introduction and motivation

COMBINE

SBML

SED-ML

SBGN

Conclusion

The many roles of computation in biological research

Instrument/device control, data management, data processing, database applications, statistical analysis, pattern matching, image processing, text mining, chemical structure prediction, genomic sequence analysis, proteomics, other *omics, molecular modeling, molecular dynamics, kinetic simulation, simulated evolution, phylogenetics, ... (to name only a subset)!

Focus here: **modeling and simulation**

Different tools ⇒ different interfaces & languages

Communication is better with standard exchange formats

Developing exchange standards is not easy

Diverse set of knowledge needed

- Scientific needs
- Technical implementation skills
- Practical experience

This has lead to different efforts for different facets of overall problem

Outline

Introduction and motivation

COMBINE

SBML

SED-ML

SBGN

Conclusion

Motivations for the creation of COMBINE

Realizations about the state of affairs in late-2000’s

- Many efforts overlapped, but lacked coordination
- Individual meetings meant more travel for many people
- Limited and fragile funding didn’t support solid, coherent base

COMBINE = Computational Modeling in Biology Network

Main objectives:

- Coordinate meetings
- Coordinate standards development
- Develop standard operating procedures and common tools
- Provide a recognized voice

COMBINE Standards

BioPAX SBML SED-ML SBGN CellML

Associated Standardization Efforts

MIRIAM SBO KISAO BioModels Qualifiers

Related Standardization Efforts

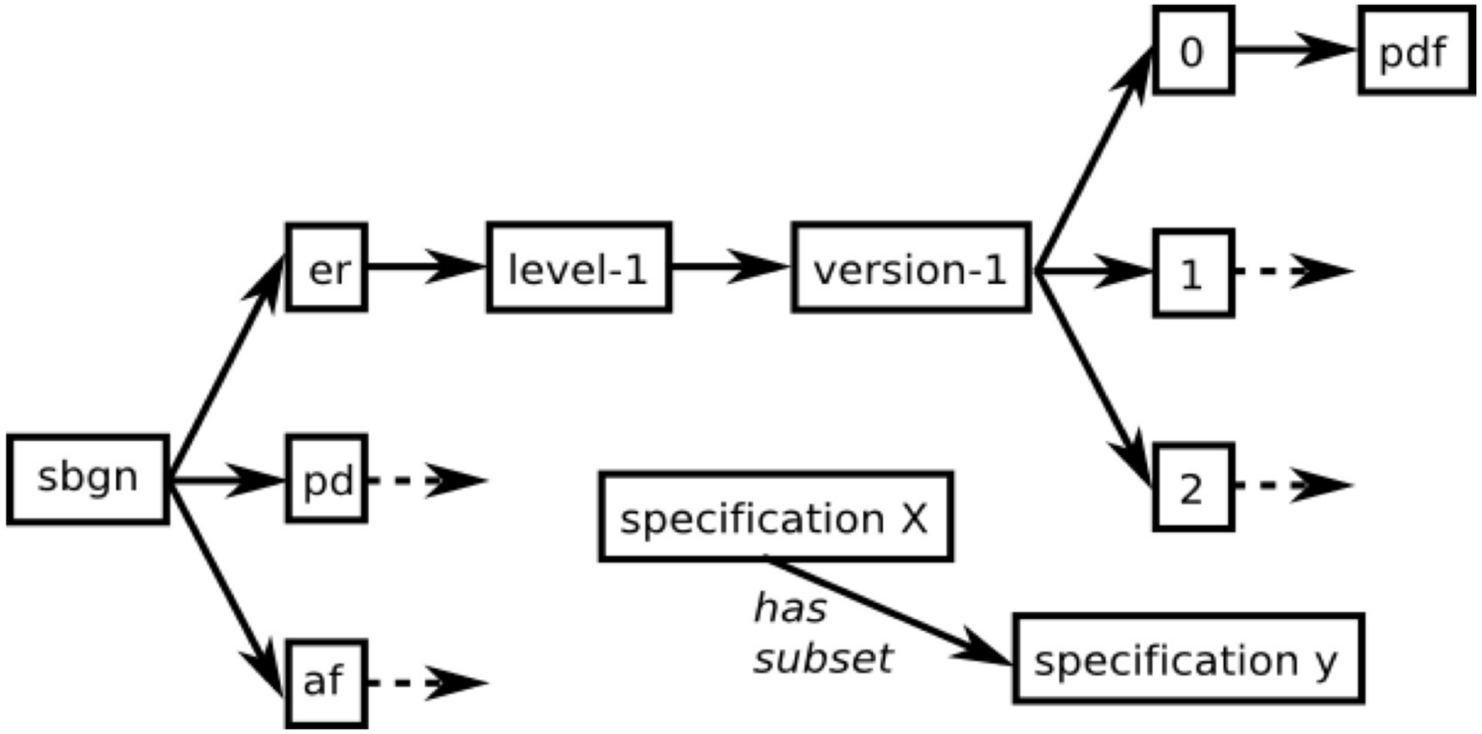
NeuroML NineML SBOL **FieldML** **NuML** **PSI-MI**

TEDDY **GPML**

Standardization efforts represented in COMBINE today

Example common infrastructure provided by COMBINE

Common URI scheme for specification documents



- E.g.: http://identifiers.org/combine.specifications/sbgn.er.level-1.version-1
  - Resolved and redirected to a page that lists where spec. is found
  - Actual documents can be stored anywhere

Some examples of goals for a common voice

Community standards are not standards branded by (e.g.) ISO

- Efforts are usually too small to undertake the process required
- However, *lack* of a standards body label is viewed negatively

COMBINE can act as a standardization body for the community

Act as a common point of contact for:

- Software developers
- Publishers
- Industry
- Funding bodies
- Legal entities

Outline

Introduction and motivation

COMBINE

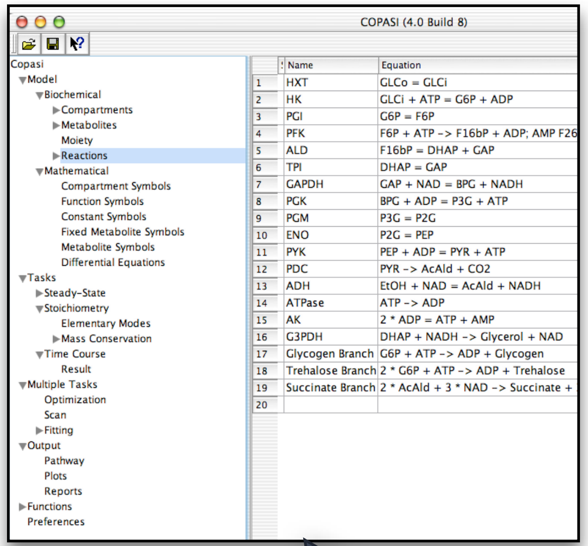
SBML

SED-ML

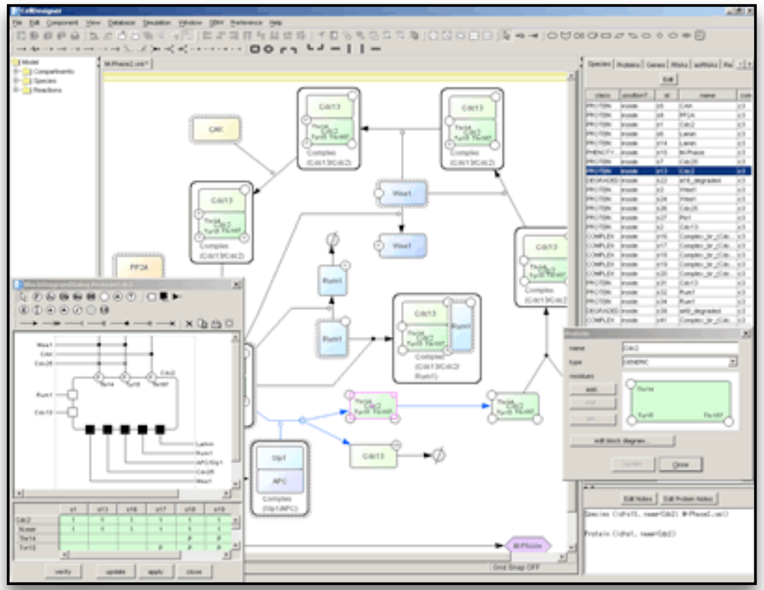
SBGN

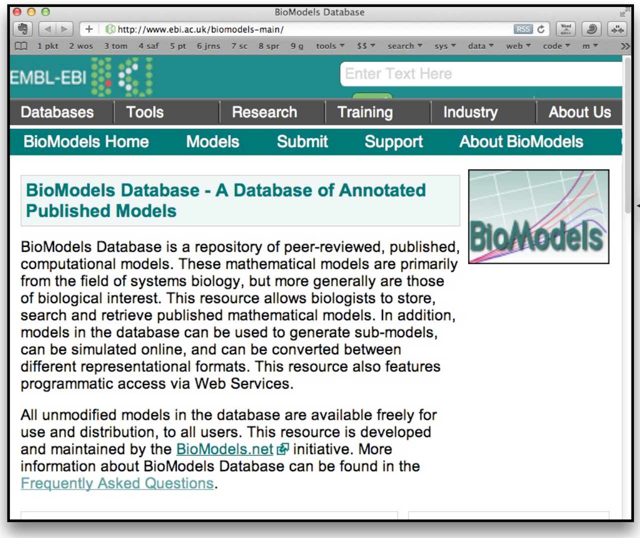
Conclusion

SBML: a lingua franca for software



	Name	Equation
1	HXT	GLCo = GLCi
2	HK	GLCi + ATP = G6P + ADP
3	PGI	G6P = F6P
4	PFK	F6P + ATP -> F16bP + ADP; AMP F26
5	ALD	F16bP = DHAP + GAP
6	TPI	DHAP = GAP
7	GAPDH	GAP + NAD = BPG + NADH
8	PGK	BPG + ADP = P3G + ATP
9	PGM	P3G = P2G
10	ENO	P2G = PEP
11	PYK	PEP + ADP = PYR + ATP
12	PDC	PYR -> AcAld + CO2
13	ADH	EtOH + NAD = AcAld + NADH
14	ATPase	ATP -> ADP
15	AK	2 * ADP = ATP + AMP
16	G3PDH	DHAP + NADH -> Glycerol + NAD
17	Glycogen Branch	G6P + ATP -> ADP + Glycogen
18	Trehalose Branch	2 * G6P + ATP -> ADP + Trehalose
19	Succinate Branch	2 * AcAld + 3 * NAD -> Succinate +
20		

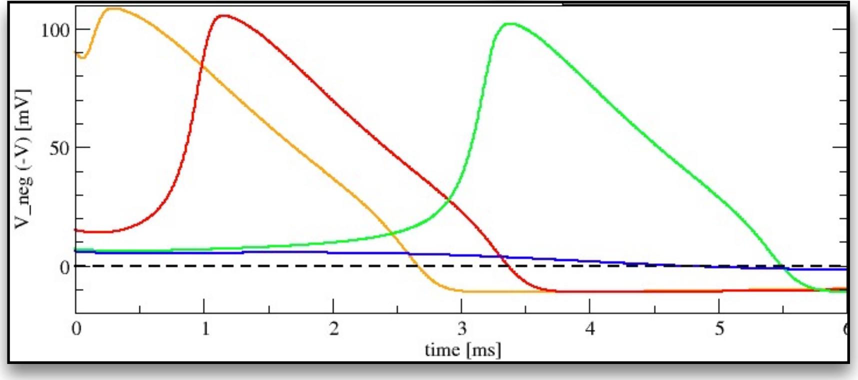


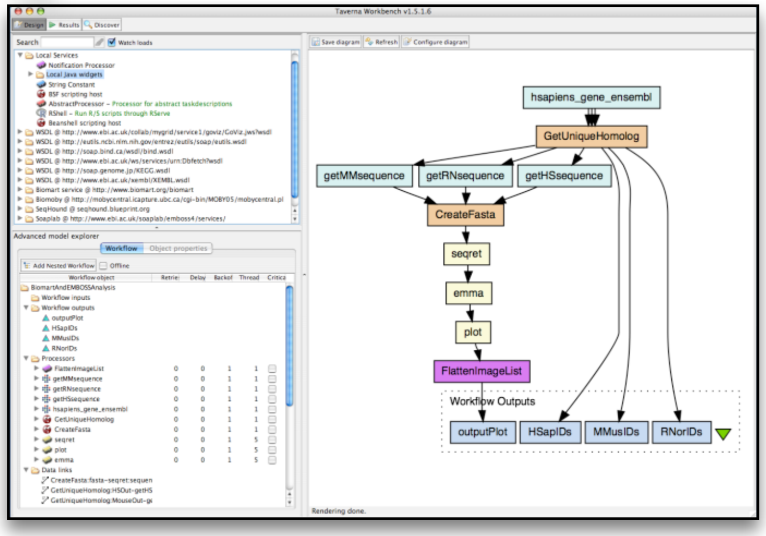


BioModels Database - A Database of Annotated Published Models

BioModels Database is a repository of peer-reviewed, published, computational models. These mathematical models are primarily from the field of systems biology, but more generally are those of biological interest. This resource allows biologists to store, search and retrieve published mathematical models. In addition, models in the database can be used to generate sub-models, can be simulated online, and can be converted between different representational formats. This resource also features programmatic access via Web Services.

All unmodified models in the database are available freely for use and distribution, to all users. This resource is developed and maintained by the BioModels.net initiative. More information about BioModels Database can be found in the Frequently Asked Questions.





SBML = Systems Biology Markup Language

Format for representing computational models of biological processes

- Data structures + usage principles + serialization to XML
- (Mostly) Declarative, not procedural—not a scripting language

Neutral with respect to modeling framework

- E.g., ODE, stochastic systems, etc.

Development started in 2000, with first specification distributed in 2001

The **process** is central

- Called a “reaction” in SBML
- Participants are pools of entities (**species**)

Models can further include:

- Other constants & variables
- Compartments
- Explicit math
- Discontinuous events
- Unit definitions
- Annotations

Basic SBML concepts are fairly simple

Traditional SBML models have been spatially homogeneous

- Metabolic network models
- Signaling pathway models
- Conductance-based models
- Neural models
- Pharmacokinetic/dynamics models
- Infectious diseases

Find examples in
BioModels Database
http://biomodels.net/biomodels

Development of SBML Level 3 *packages* is extending this scope

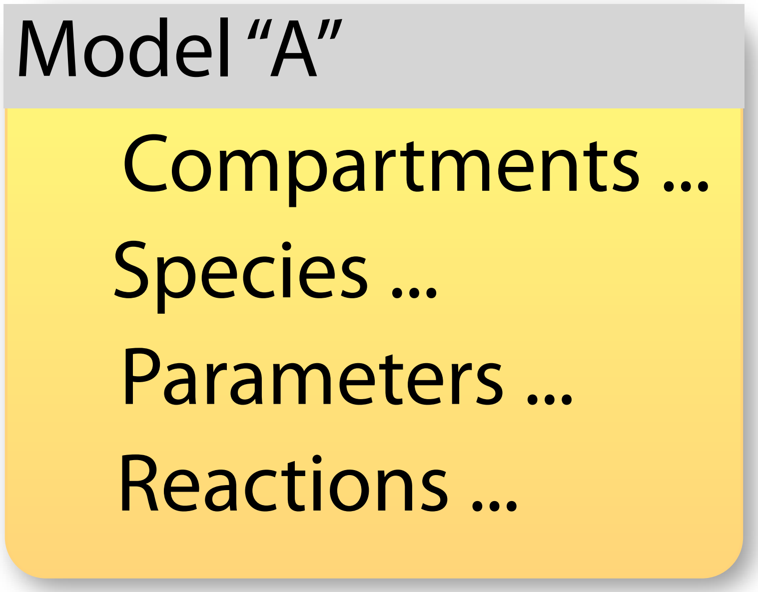
- E.g.: Spatially inhomogeneous models, also qualitative/logical

Scope of SBML encompasses many types of models

Level 3 package	What it enables	Status
Hierarchical model composition	Models containing submodels	✔
Flux balance constraints	Flux balance analysis models	✔
Qualitative models	Petri net models, Boolean models	RC
Spatial	Nonhomogeneous spatial models	draft
Multicomponent species	Entities w/ structure; also rule-based models	draft
Graph layout	Diagrams of models	draft
Graph rendering	Diagrams of models	draft
Distributions	Numerical values as statistical distributions	in dev
Groups	Arbitrary grouping of components	in dev
Annotations	Richer annotation syntax	
Dynamic structures	Creation & destruction of components	
Arrays & sets	Arrays or sets of entities	

Example: SBML Level 3 Hierarchical Model Composition

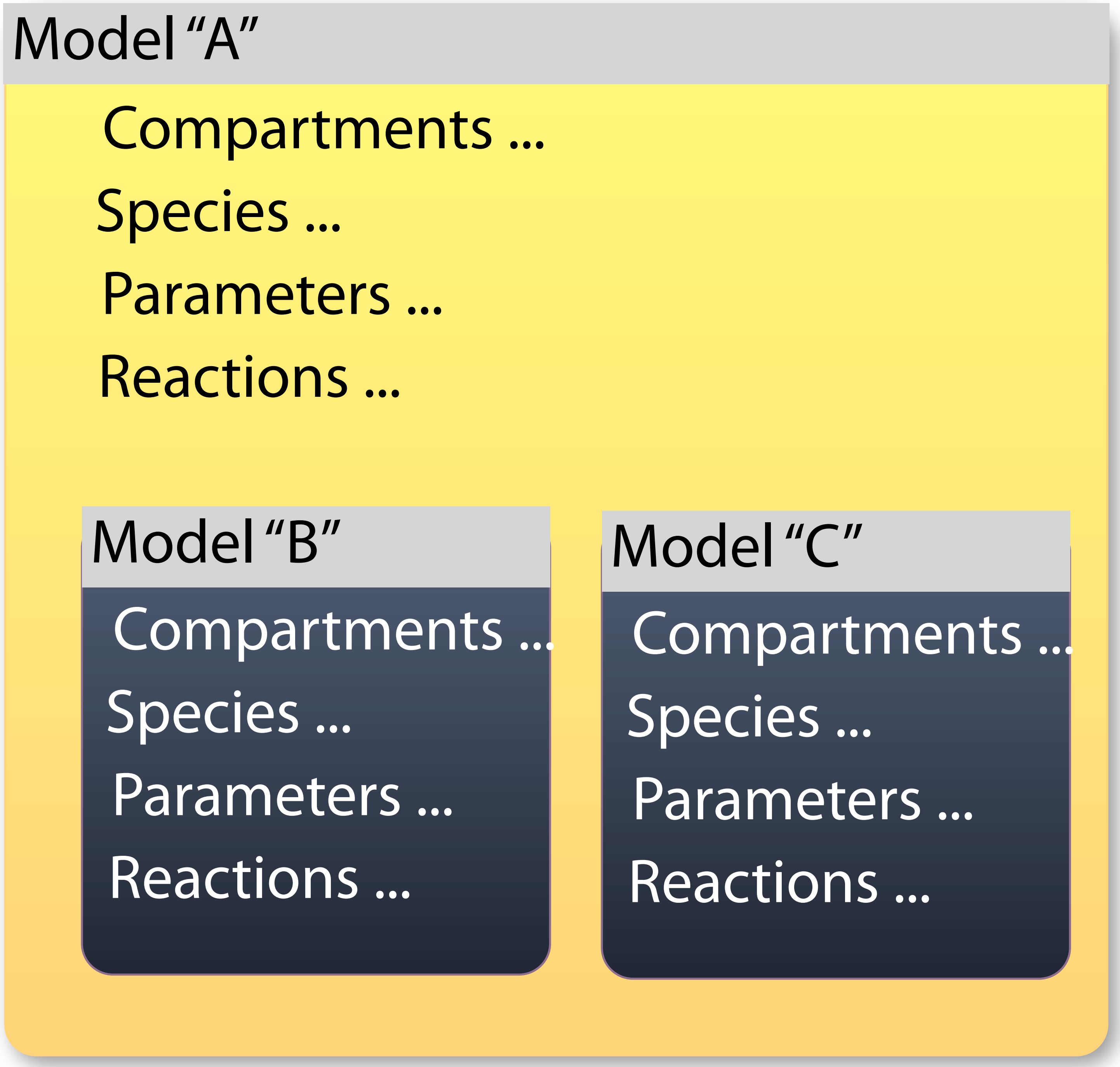
Core SBML



Model "A"

Compartments ...
Species ...
Parameters ...
Reactions ...

With hierarchical model composition



Model "A"

Compartments ...
Species ...
Parameters ...
Reactions ...

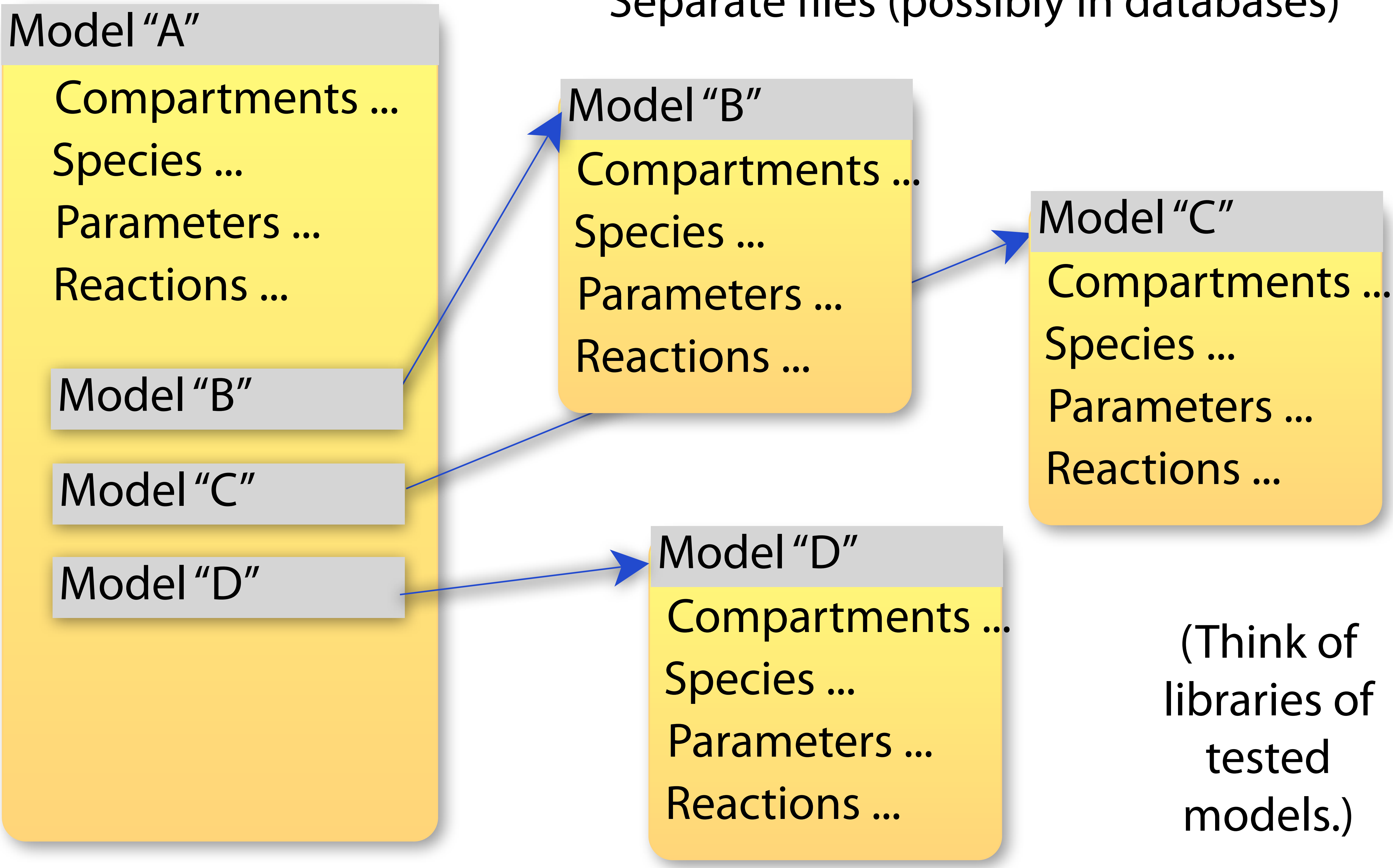
Model "B"

Compartments ...
Species ...
Parameters ...
Reactions ...

Model "C"

Compartments ...
Species ...
Parameters ...
Reactions ...

The ‘comp’ package supports multiple arrangements



Separate files (possibly in databases)

Model “A”
Compartments ...
Species ...
Parameters ...
Reactions ...
Model “B”
Model “C”
Model “D”

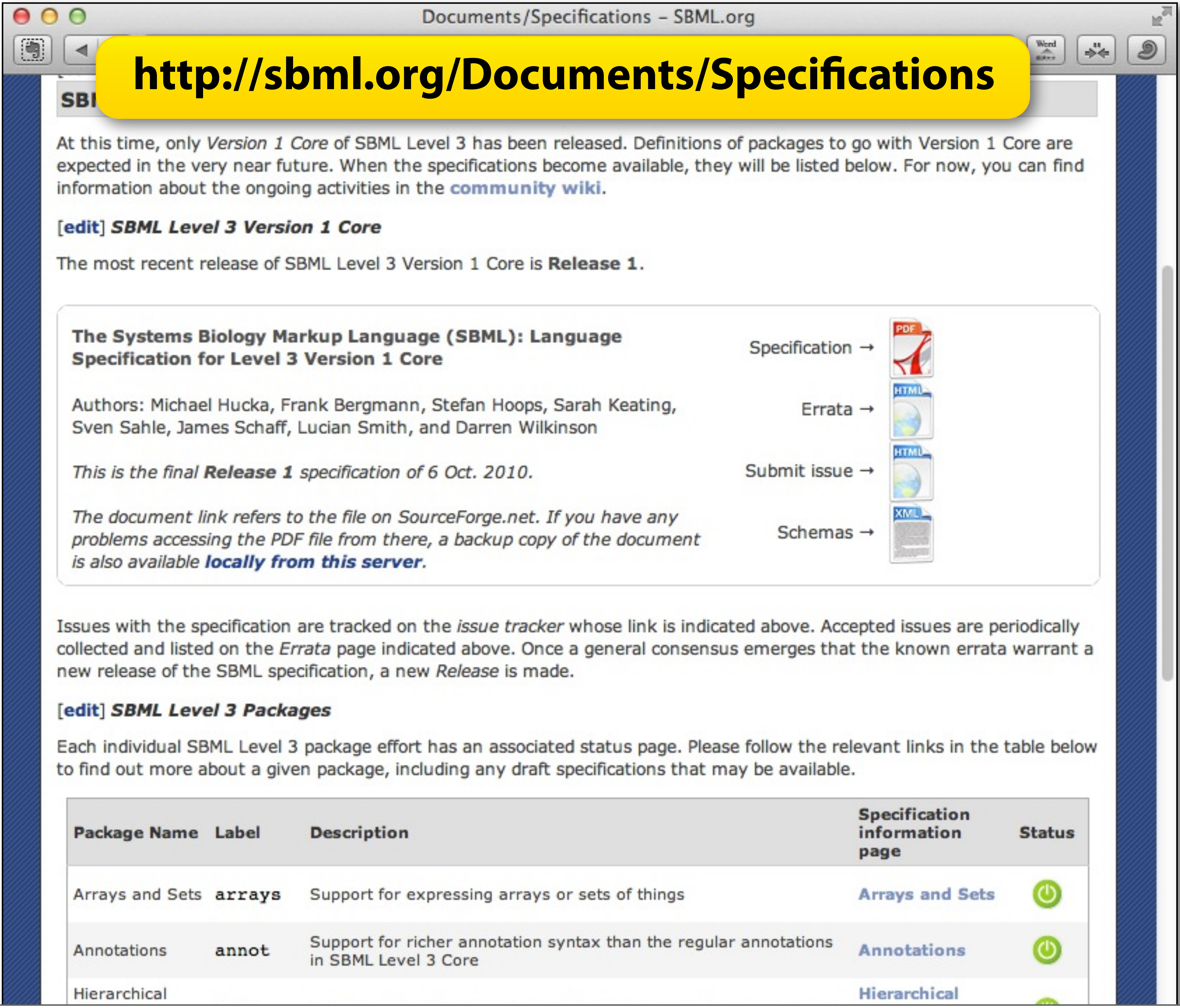
Model “B”
Compartments ...
Species ...
Parameters ...
Reactions ...

Model “C”
Compartments ...
Species ...
Parameters ...
Reactions ...

Model “D”
Compartments ...
Species ...
Parameters ...
Reactions ...

(Think of libraries of tested models.)

Where to find package information & documents



http://sbml.org/Documents/Specifications

SBI

At this time, only *Version 1 Core* of SBML Level 3 has been released. Definitions of packages to go with Version 1 Core are expected in the very near future. When the specifications become available, they will be listed below. For now, you can find information about the ongoing activities in the **community wiki**.

[edit] ***SBML Level 3 Version 1 Core***

The most recent release of SBML Level 3 Version 1 Core is **Release 1**.

The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core

Authors: Michael Hucka, Frank Bergmann, Stefan Hoops, Sarah Keating, Sven Sahle, James Schaff, Lucian Smith, and Darren Wilkinson

*This is the final **Release 1** specification of 6 Oct. 2010.*

*The document link refers to the file on SourceForge.net. If you have any problems accessing the PDF file from there, a backup copy of the document is also available **locally from this server**.*

Specification → PDF

Errata → HTML

Submit issue → HTML

Schemas → XML

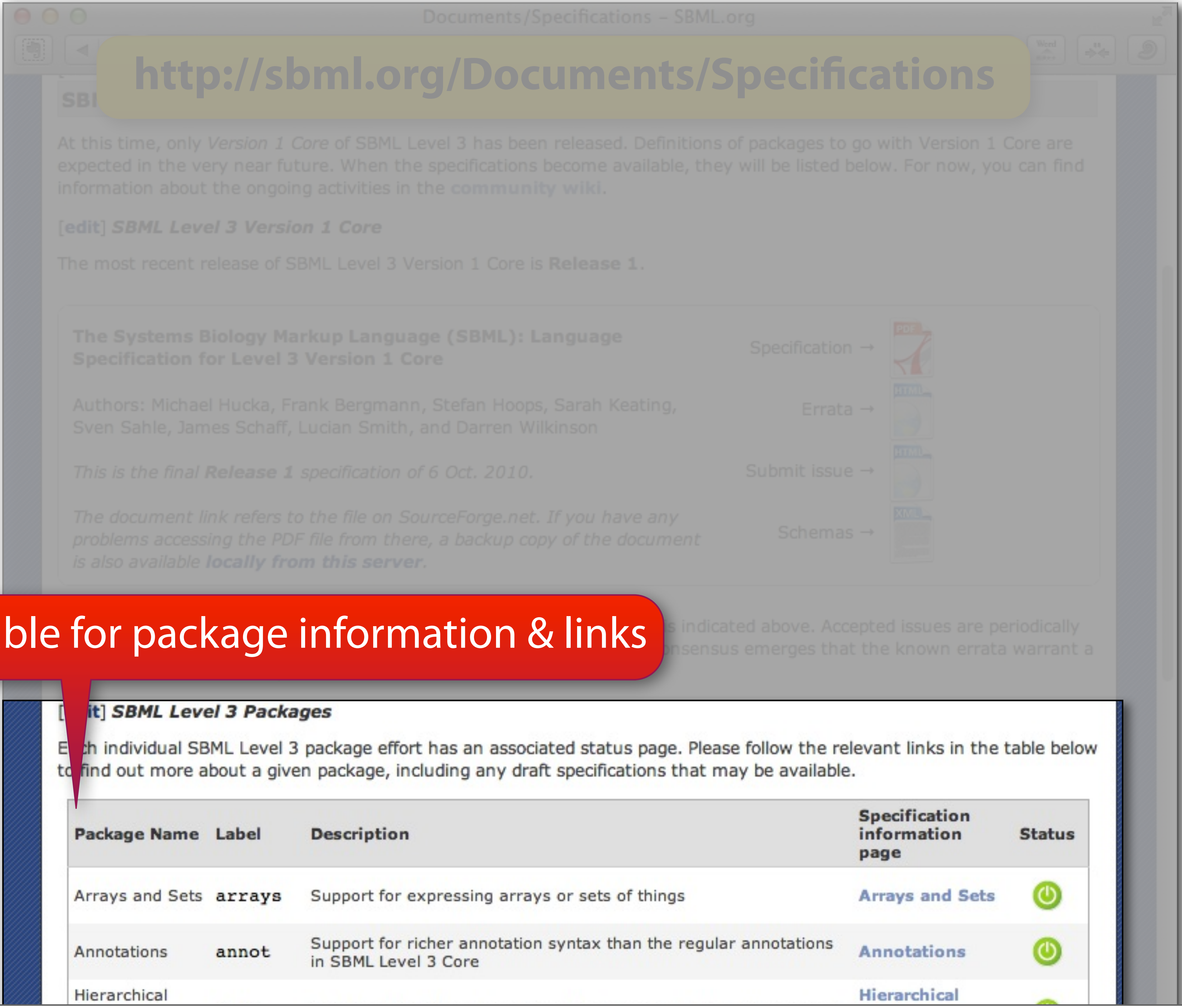
Issues with the specification are tracked on the *issue tracker* whose link is indicated above. Accepted issues are periodically collected and listed on the *Errata* page indicated above. Once a general consensus emerges that the known errata warrant a new release of the SBML specification, a new *Release* is made.

[edit] ***SBML Level 3 Packages***

Each individual SBML Level 3 package effort has an associated status page. Please follow the relevant links in the table below to find out more about a given package, including any draft specifications that may be available.

Package Name	Label	Description	Specification information page	Status
Arrays and Sets	`arrays`	Support for expressing arrays or sets of things	**Arrays and Sets**	●
Annotations	`annot`	Support for richer annotation syntax than the regular annotations in SBML Level 3 Core	**Annotations**	●
Hierarchical			**Hierarchical**	

Where to find package information & documents



http://sbml.org/Documents/Specifications

SBI

At this time, only *Version 1 Core* of SBML Level 3 has been released. Definitions of packages to go with Version 1 Core are expected in the very near future. When the specifications become available, they will be listed below. For now, you can find information about the ongoing activities in the **community wiki**.

[edit] ***SBML Level 3 Version 1 Core***

The most recent release of SBML Level 3 Version 1 Core is **Release 1**.

The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core

Authors: Michael Hucka, Frank Bergmann, Stefan Hoops, Sarah Keating, Sven Sahle, James Schaff, Lucian Smith, and Darren Wilkinson

*This is the final **Release 1** specification of 6 Oct. 2010.*

*The document link refers to the file on SourceForge.net. If you have any problems accessing the PDF file from there, a backup copy of the document is also available **locally from this server**.*

s indicated above. Accepted issues are periodically
nsensus emerges that the known errata warrant a

Table for package information & links

[it] ***SBML Level 3 Packages***

E ch individual SBML Level 3 package effort has an associated status page. Please follow the relevant links in the table below to find out more about a given package, including any draft specifications that may be available.

Package Name	Label	Description	Specification information page	Status
Arrays and Sets	`arrays`	Support for expressing arrays or sets of things	Arrays and Sets	⏻
Annotations	`annot`	Support for richer annotation syntax than the regular annotations in SBML Level 3 Core	Annotations	⏻
Hierarchical			Hierarchical	

Where to find software applications compatible with SBML

Welcome to the portal for the **Systems Biology Markup Language** (**SBML**), a free and open interchange format for computer models of biological processes. SBML is useful for models of metabolism, cell signaling, and more. It has been in development by an international community since the year 2000.

For the curious
What *is* SBML? Read our **introduction**, then perhaps browse the **mailing lists** to glimpse what's happening with SBML today.

For modelers
Looking for software that supports SBML? Our **software guide** lists over **220** systems. Are you instead looking for models? Visit **BioModels Database**, where you can find hundreds!

For software developers
Interested in supporting SBML in your software? Read our **basic introduction** and then the **SBML specifications** to understand how to use SBML. After that, you may want to look at **libSBML**.

No matter how you use SBML, we invite you to sign up for news updates either through our **RSS feed**, our **Twitter feed**, or one of the **mailing lists**, and get involved with **community efforts** to help keep improving SBML. You can also call attention to your project's support of SBML by displaying the **SBML logo**.

SBML would not have been possible without support from **multiple agencies and organizations**, as well as intellectual contributions from many motivated individuals, including the **major contributors** who are shaping SBML Level 3

SBML News

COMBINE 2011
(14 Jun.'11) Registration is now open for **COMBINE 2011**, being held in Heidelberg, DE, on Sep. 3–7.

SBML Test Suite updated
(6 Jun.'11) Changes include addition of SBML L3v1 Core test cases, SED-ML files, and updates to the online interface.

Older news ...

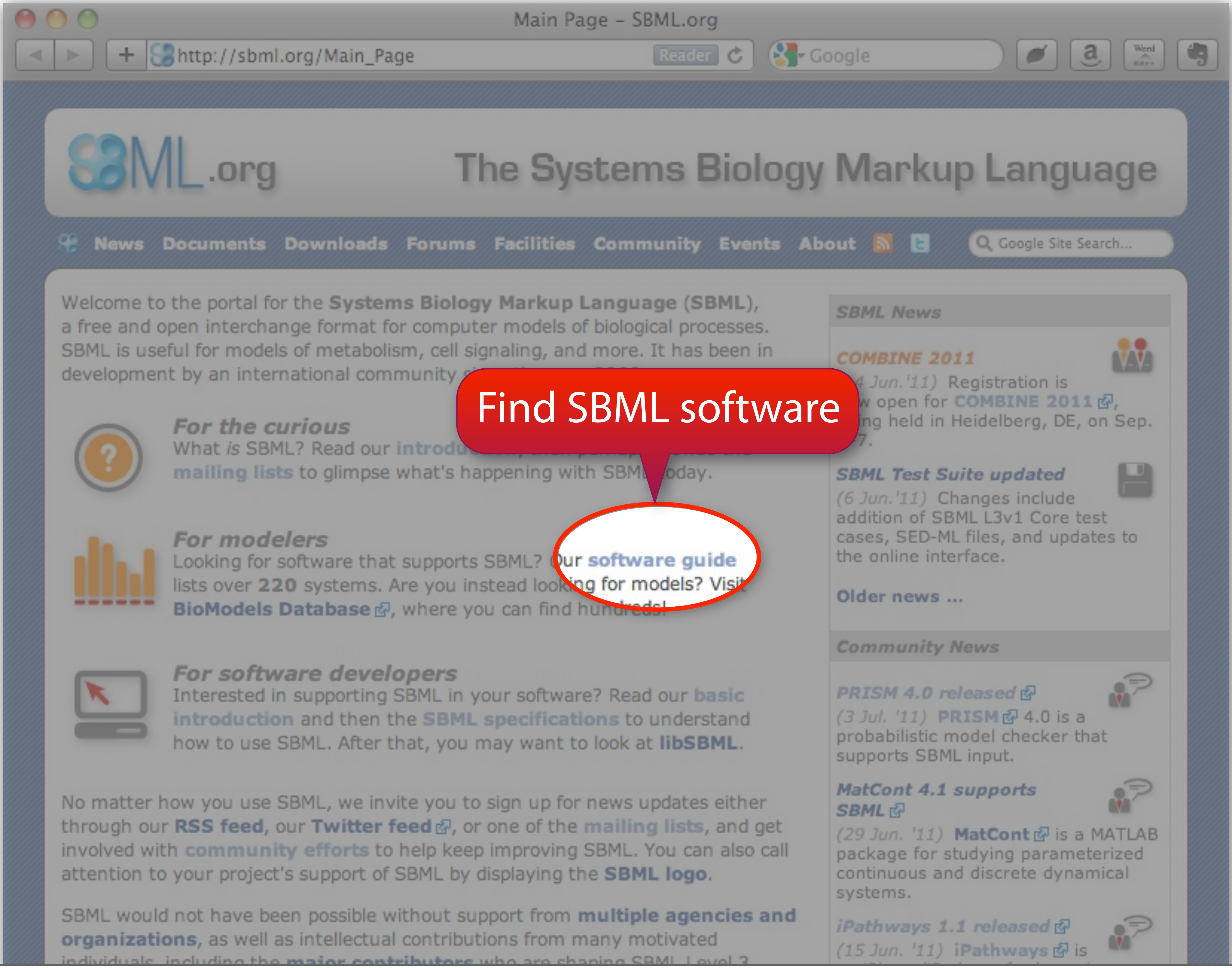
Community News

PRISM 4.0 released
(3 Jul. '11) **PRISM** 4.0 is a probabilistic model checker that supports SBML input.

MatCont 4.1 supports SBML
(29 Jun. '11) **MatCont** is a MATLAB package for studying parameterized continuous and discrete dynamical systems.

iPathways 1.1 released
(15 Jun. '11) **iPathways** is

Where to find software applications compatible with SBML



The Systems Biology Markup Language

Welcome to the portal for the **Systems Biology Markup Language (SBML)**, a free and open interchange format for computer models of biological processes. SBML is useful for models of metabolism, cell signaling, and more. It has been in development by an international community s

For the curious
What *is* SBML? Read our introdu
mailing lists to glimpse what's happening with SBM oday.

For modelers
Looking for software that supports SBML? Our **software guide** lists over **220** systems. Are you instead looking for models? Visit **BioModels Database**, where you can find hundreds!

For software developers
Interested in supporting SBML in your software? Read our **basic introduction** and then the **SBML specifications** to understand how to use SBML. After that, you may want to look at **libSBML**.

No matter how you use SBML, we invite you to sign up for news updates either through our **RSS feed**, our **Twitter feed**, or one of the **mailing lists**, and get involved with **community efforts** to help keep improving SBML. You can also call attention to your project's support of SBML by displaying the **SBML logo**.

SBML would not have been possible without support from **multiple agencies and organizations**, as well as intellectual contributions from many motivated individuals, including the **major contributors** who are shaping SBML Level 3

SBML News

COMBINE 2011
4 Jun.'11) Registration is w open for **COMBINE 2011**, ng held in Heidelberg, DE, on Sep. 7.

SBML Test Suite updated
(6 Jun.'11) Changes include addition of SBML L3v1 Core test cases, SED-ML files, and updates to the online interface.

Older news ...

Community News

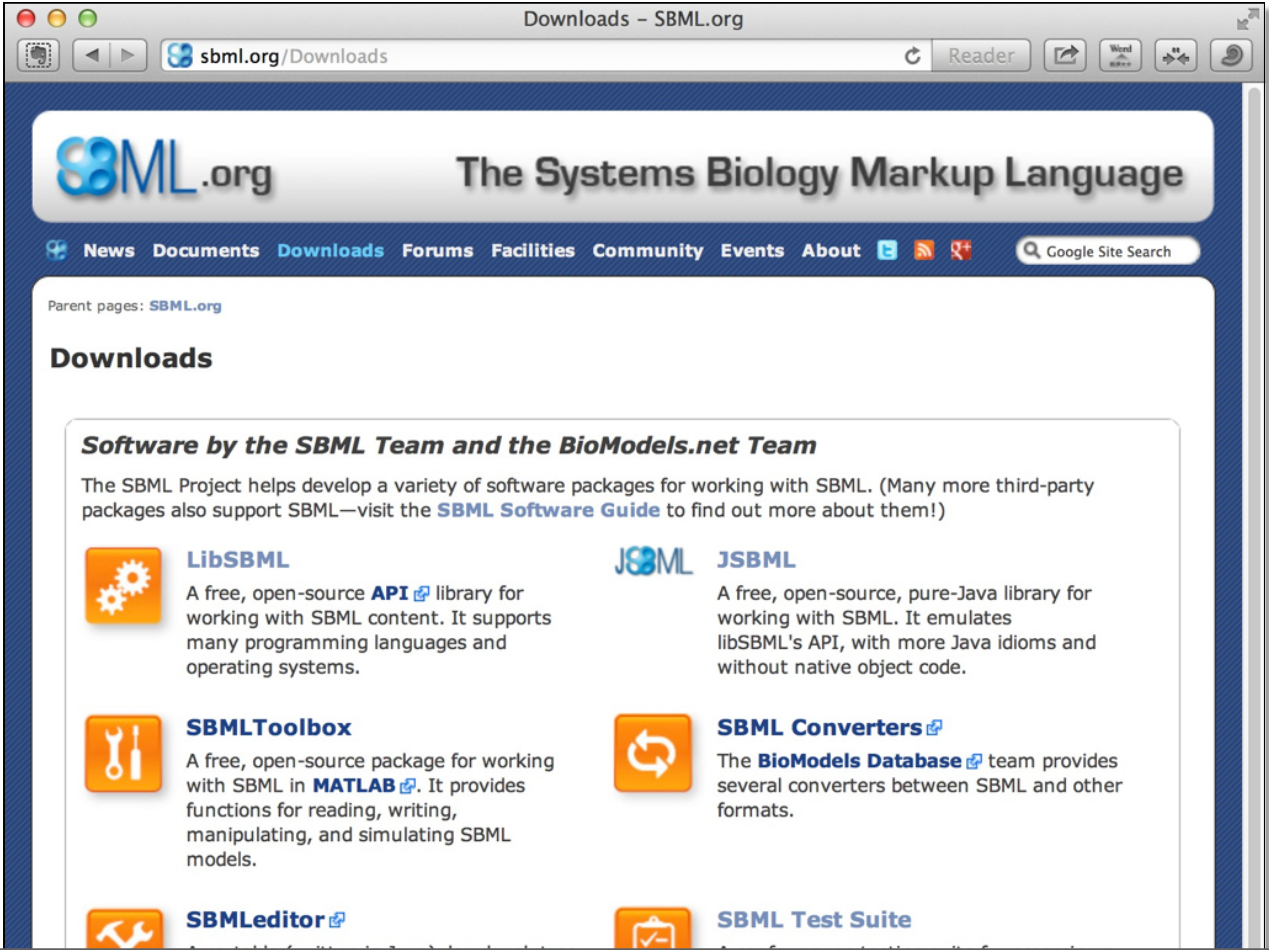
PRISM 4.0 released
(3 Jul. '11) **PRISM** 4.0 is a probabilistic model checker that supports SBML input.

MatCont 4.1 supports SBML
(29 Jun. '11) **MatCont** is a MATLAB package for studying parameterized continuous and discrete dynamical systems.

iPathways 1.1 released
(15 Jun. '11) **iPathways** is

Find SBML software

Where to find libraries for implementing SBML support



Parent pages: **SBML.org**

Downloads

Software by the SBML Team and the BioModels.net Team

The SBML Project helps develop a variety of software packages for working with SBML. (Many more third-party packages also support SBML—visit the **SBML Software Guide** to find out more about them!)

LibSBML

A free, open-source **API** library for working with SBML content. It supports many programming languages and operating systems.

JSBML

A free, open-source, pure-Java library for working with SBML. It emulates libSBML's API, with more Java idioms and without native object code.

SBMLToolbox

A free, open-source package for working with SBML in **MATLAB**. It provides functions for reading, writing, manipulating, and simulating SBML models.

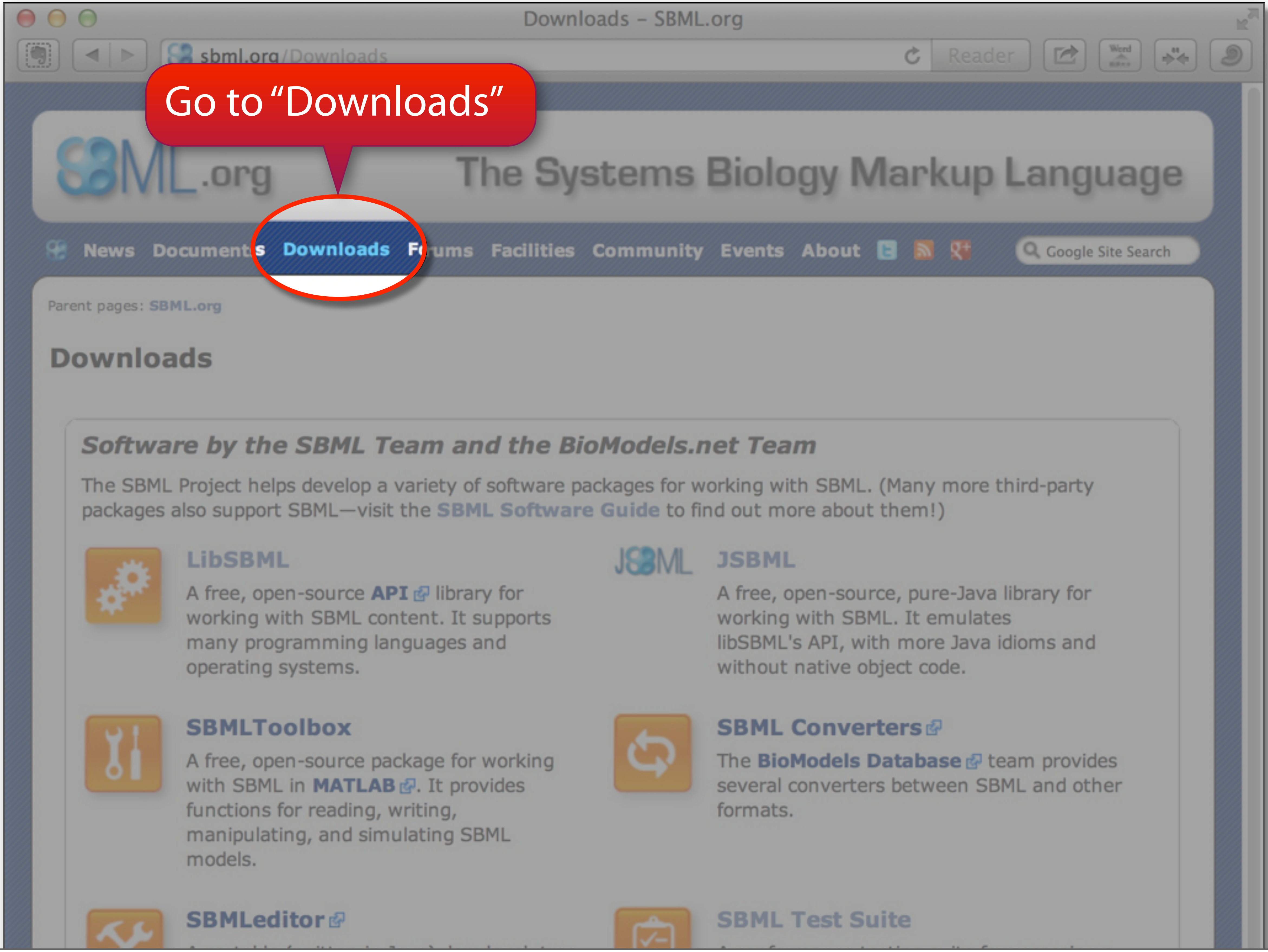
SBML Converters

The **BioModels Database** team provides several converters between SBML and other formats.

SBMLeditor

SBML Test Suite

Where to find libraries for implementing SBML support



Go to “Downloads”

The Systems Biology Markup Language

Parent pages: SBML.org

Downloads

Software by the SBML Team and the BioModels.net Team

The SBML Project helps develop a variety of software packages for working with SBML. (Many more third-party packages also support SBML—visit the **SBML Software Guide** to find out more about them!)

LibSBML

A free, open-source **API** library for working with SBML content. It supports many programming languages and operating systems.

JSBML

A free, open-source, pure-Java library for working with SBML. It emulates libSBML's API, with more Java idioms and without native object code.

SBMLToolbox

A free, open-source package for working with SBML in **MATLAB**. It provides functions for reading, writing, manipulating, and simulating SBML models.

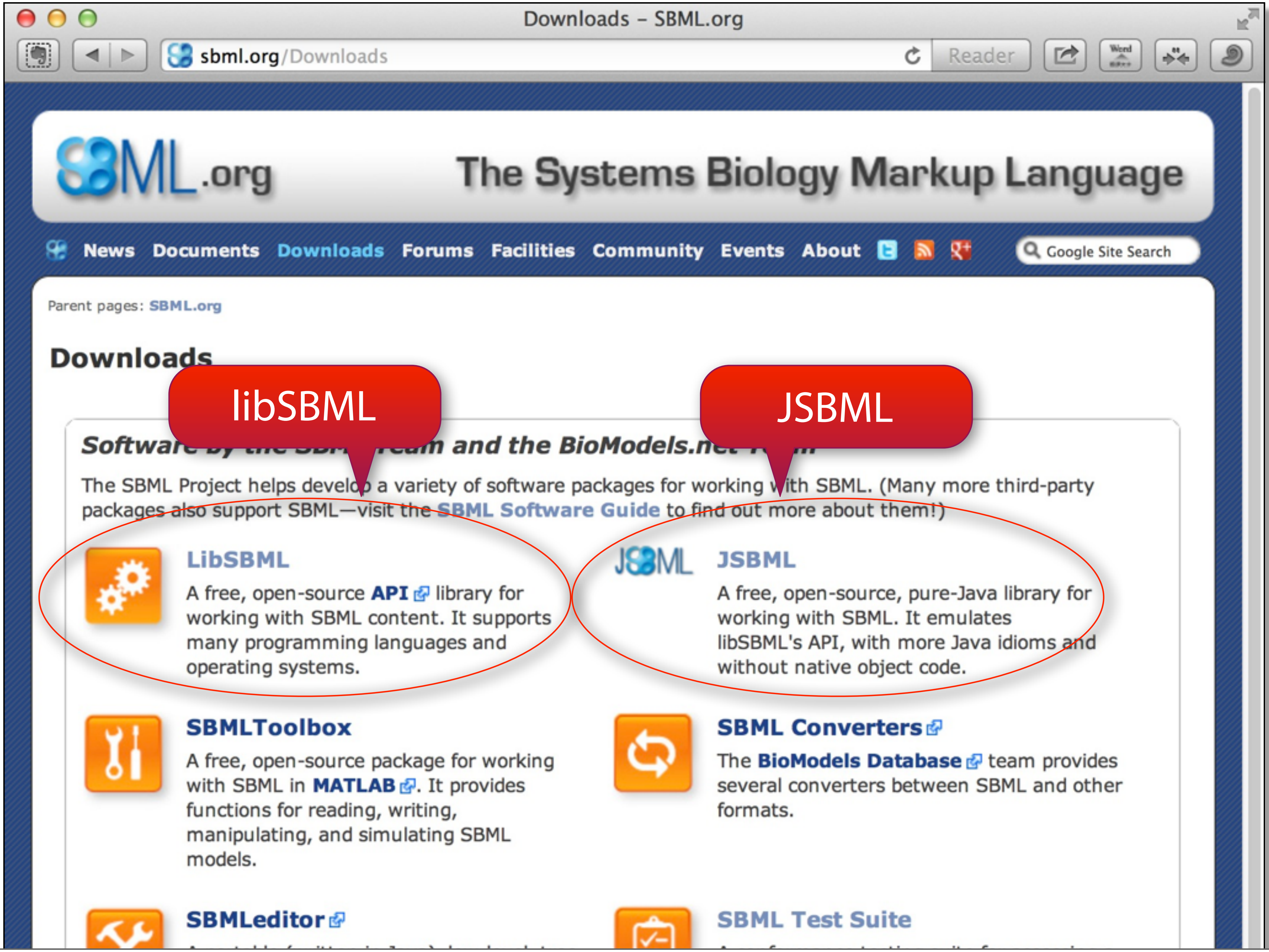
SBML Converters

The **BioModels Database** team provides several converters between SBML and other formats.

SBMLeditor

SBML Test Suite

Where to find libraries for implementing SBML support



Parent pages: SBML.org

Downloads

libSBML

JSBML

Software by the SBML Team and the BioModels.net Team

The SBML Project helps develop a variety of software packages for working with SBML. (Many more third-party packages also support SBML—visit the **SBML Software Guide** to find out more about them!)

LibSBML

A free, open-source **API** library for working with SBML content. It supports many programming languages and operating systems.

JSBML

A free, open-source, pure-Java library for working with SBML. It emulates libSBML's API, with more Java idioms and without native object code.

SBMLToolbox

A free, open-source package for working with SBML in **MATLAB**. It provides functions for reading, writing, manipulating, and simulating SBML models.

SBML Converters

The **BioModels Database** team provides several converters between SBML and other formats.

SBMLeditor

SBML Test Suite

Outline

Introduction and motivation

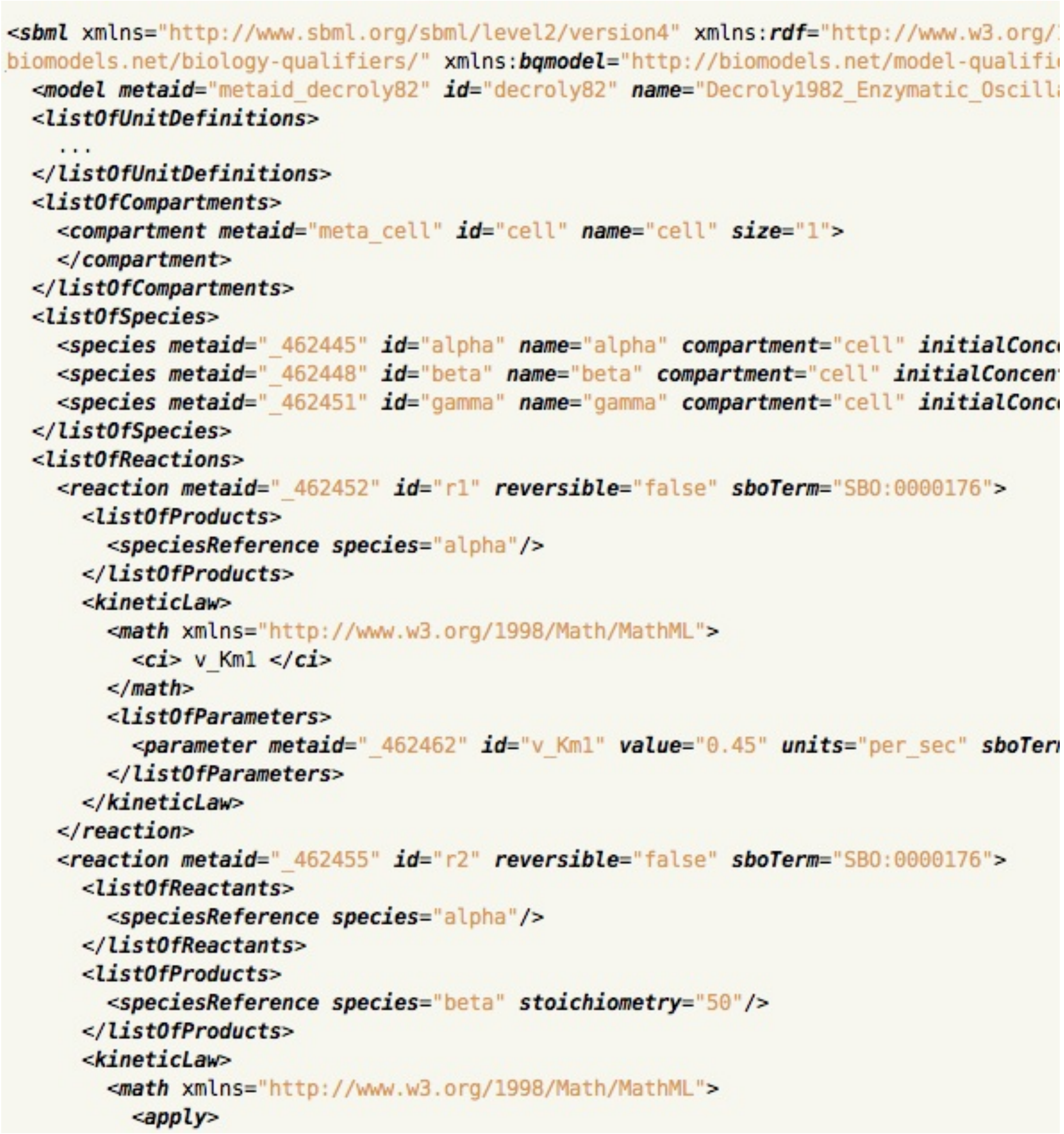
COMBINE

SBML

SED-ML

SBGN

Conclusion



```
<sbml xmlns="http://www.sbml.org/sbml/level2/version4" xmlns:rdf="http://www.w3.org/
biomodels.net/biology-qualifiers/" xmlns:bqmodel="http://biomodels.net/model-qualifi
  <model metaid="metaid_decroly82" id="decroly82" name="Decroly1982_Enzymatic_Oscill
  <listOfUnitDefinitions>
    ...
  </listOfUnitDefinitions>
  <listOfCompartments>
    <compartment metaid="meta_cell" id="cell" name="cell" size="1">
    </compartment>
  </listOfCompartments>
  <listOfSpecies>
    <species metaid="_462445" id="alpha" name="alpha" compartment="cell" initialConc
    <species metaid="_462448" id="beta" name="beta" compartment="cell" initialConcen
    <species metaid="_462451" id="gamma" name="gamma" compartment="cell" initialConc
  </listOfSpecies>
  <listOfReactions>
    <reaction metaid="_462452" id="r1" reversible="false" sboTerm="SBO:0000176">
      <listOfProducts>
        <speciesReference species="alpha"/>
      </listOfProducts>
      <kineticLaw>
        <math xmlns="http://www.w3.org/1998/Math/MathML">
          <ci> v_Km1 </ci>
        </math>
        <listOfParameters>
          <parameter metaid="_462462" id="v_Km1" value="0.45" units="per_sec" sboTer
        </listOfParameters>
      </kineticLaw>
    </reaction>
    <reaction metaid="_462455" id="r2" reversible="false" sboTerm="SBO:0000176">
      <listOfReactants>
        <speciesReference species="alpha"/>
      </listOfReactants>
      <listOfProducts>
        <speciesReference species="beta" stoichiometry="50"/>
      </listOfProducts>
      <kineticLaw>
        <math xmlns="http://www.w3.org/1998/Math/MathML">
          <apply>
```

BIOMD0000000319 in BioModels Database

?

Decroly & Goldbeter, PNAS, 1982

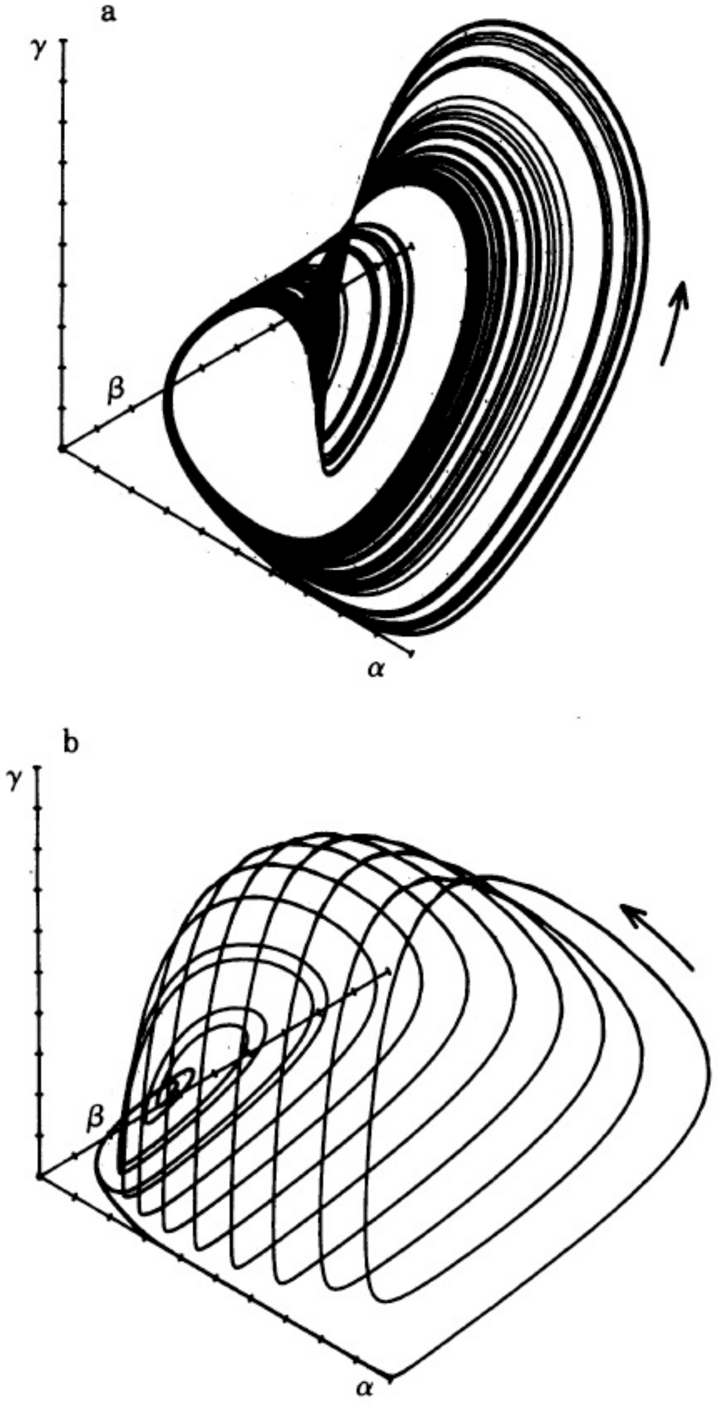


FIG. 4. Trajectories in the phase space (α, β, γ) associated with chaos (a) and with complex periodic behavior (b). The curves correspond to the substrate evolution depicted in Fig. 2 c and d, respectively, and have been obtained by integration of the kinetic equations from t = 0–5,000 sec. The ranges of variation of α, β, and γ in a are α = 28.44–50.6, β = 50.05–351.1, and γ = 0.05–2.28 and in b are α = 28.18–190.5, β = 0.14–604.0, and γ = 0.00014–8.8.

Need to capture the processes applied to models

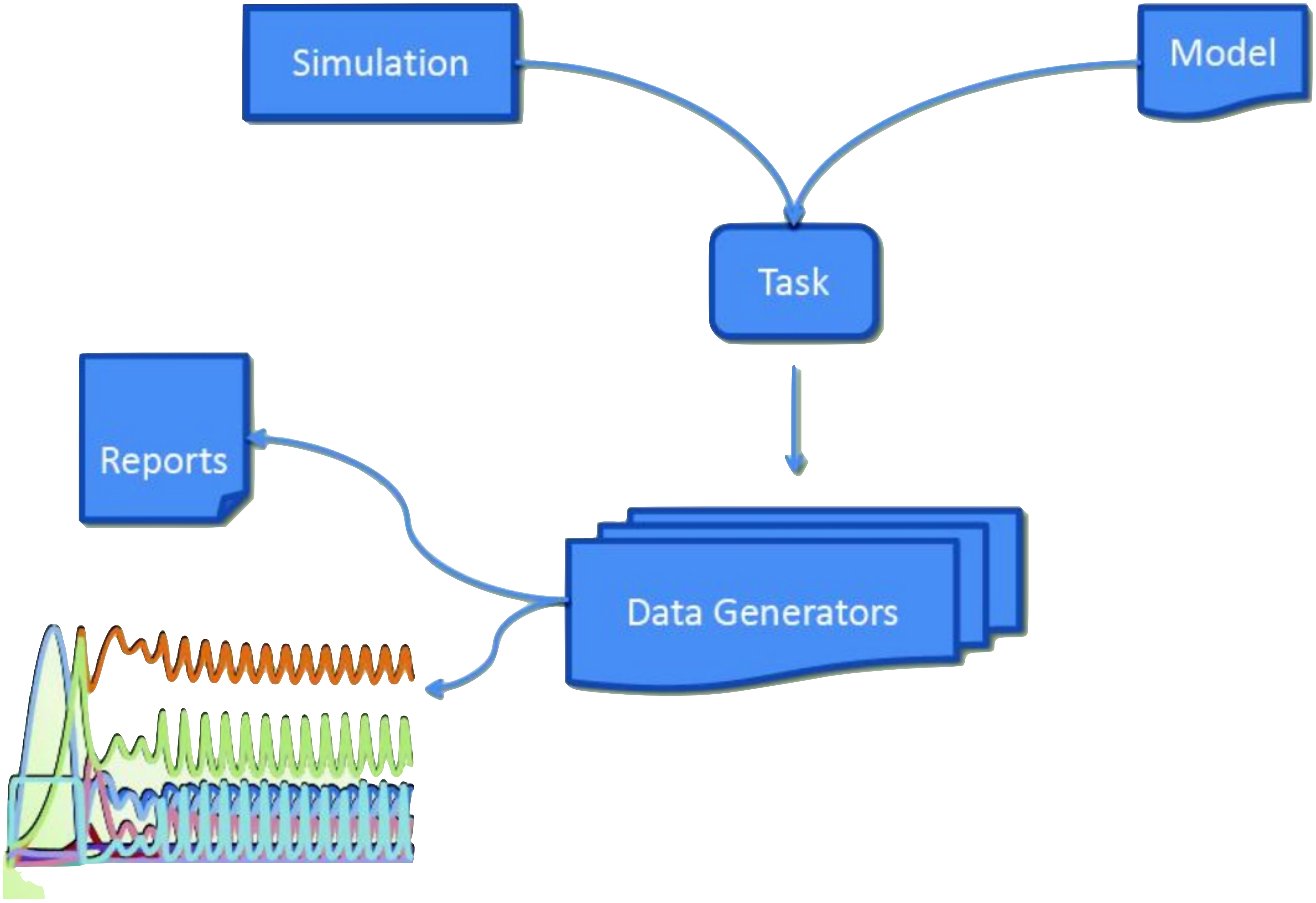
SED-ML = Simulation Experiment Description ML

Application-independent format to capture procedures, algorithms, parameter values

- Neutral format for encoding the steps to go from model to output

Can be used for

- Simulation experiments encoding parametrizations & perturbations
- Simulations using more than one model and/or method
- Data manipulations to produce plot(s)



Simulation

Model

Task

Reports

Data Generators

Waltemath et al., BMC Systems Biology, 2011

Basic components of SED-ML Level 1 Version 1

Software apps & libraries available for SED-ML Level 1 v.1

Some SED-ML-compatible software today:

- libSedML
- jlibsedml
- SBW Simulation Tool
- CellDesigner
- Web tools
- others

libSedML - Libraries and Tools supporting SED-ML

```
<?xml version="1.0" encoding="utf-8"?>
<sedML version="0.1" xmlns="http://www.biomodels.net/sed-ml"
  <notes>Changing a system from oscillation to chaos</notes>
  <listOfSimulations>
    <uniformTimeCourse id="simulation1"
    algorithm="KiSAO:0000071" initialTime="0" outputStartTime
  </listOfSimulations>
  <listOfModels>
    <model id="model1" name="Circadian Oscillations" type="SBM
    <model id="model2" name="Circadian Chaos" type="SBML" sou
      <listOfChanges>
        <changeAttribute target="/sbml:sbml/sbml:model/sbml:l
```

SED-ML - Simulation Experiment Description - Markup Language

The SED-ML format allows to store *in silico* simulation experiments independently from the

http://sedml.org

Outline

- Introduction and motivation
- COMBINE
- SBML
- SED-ML
- SBGN
- Conclusion

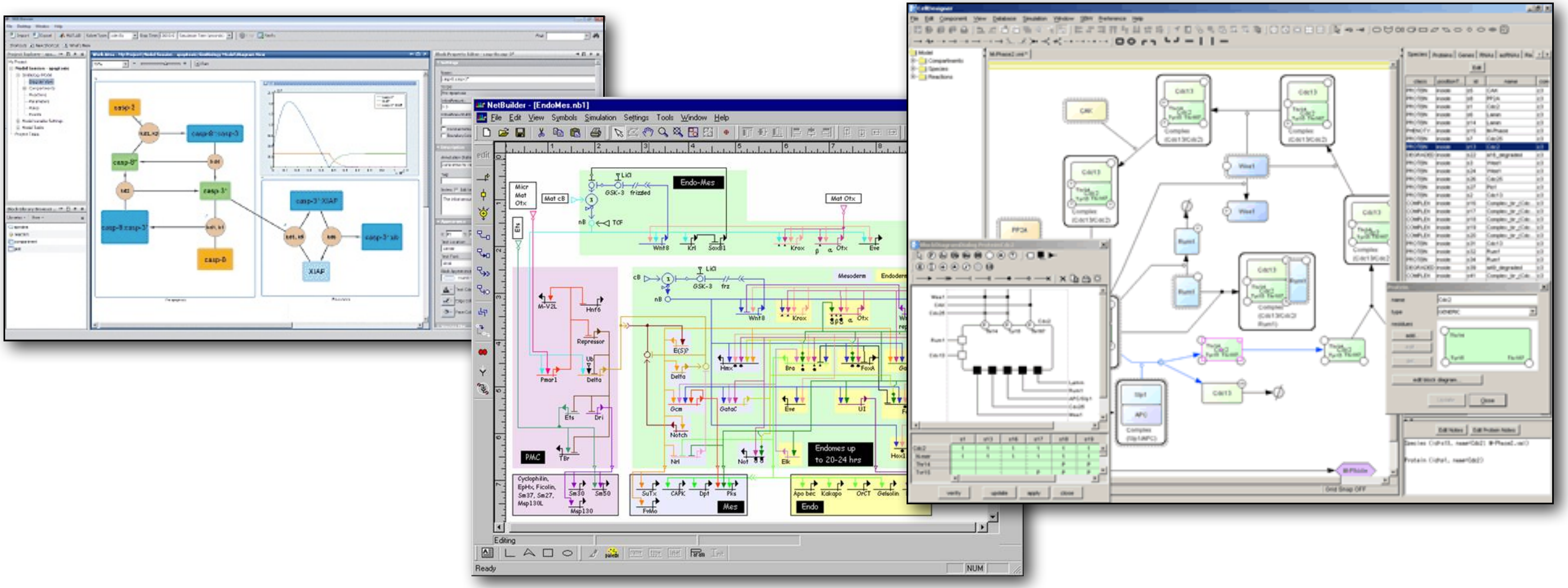
Graphical representation of models

Today: broad variation in graphical notation used in biological diagrams

- Between authors, between journals, even people in same group

However, standard notations would offer benefits:

- Consistency = easier to read diagrams with less ambiguity
- Software support: verification of correctness, translation to math



SBGN = Systems Biology Graphical Notation

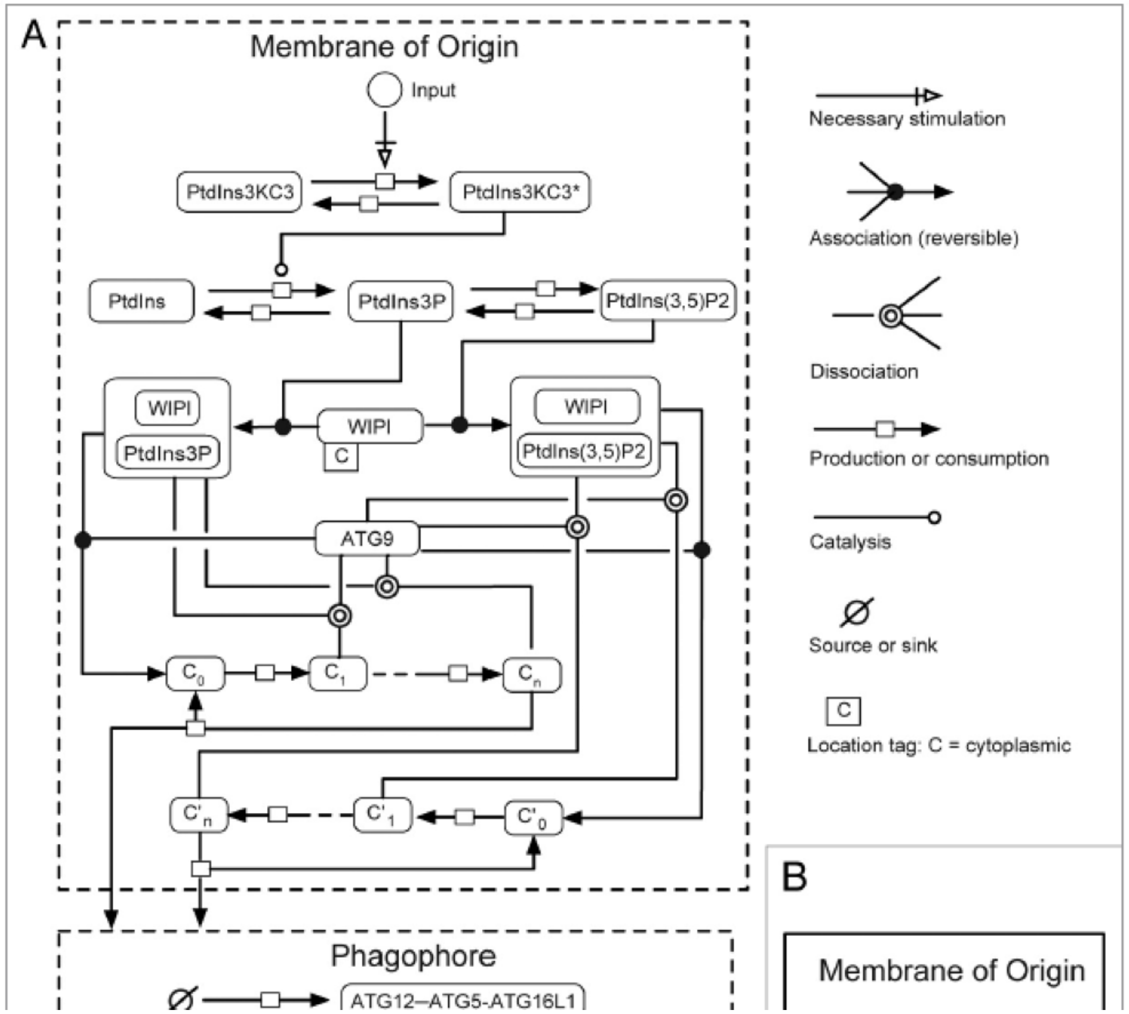
Goal: standardize the graphical notation in diagrams of biological processes

3 sublanguages to describe different facets of a model

- **Process Diagram**: causal sequences of processes & their results
  - A node represents a given state of an entity
- **Entity Relationship**: interactions bet. entities regardless of sequence
  - A node represents an entity regardless of state
- **Activity Flow**: information flowing from one entity to another
  - Hybrid — shows flow of activity without state transitions

Languages reuse same symbols, but their interpretations are different

SBGN support today



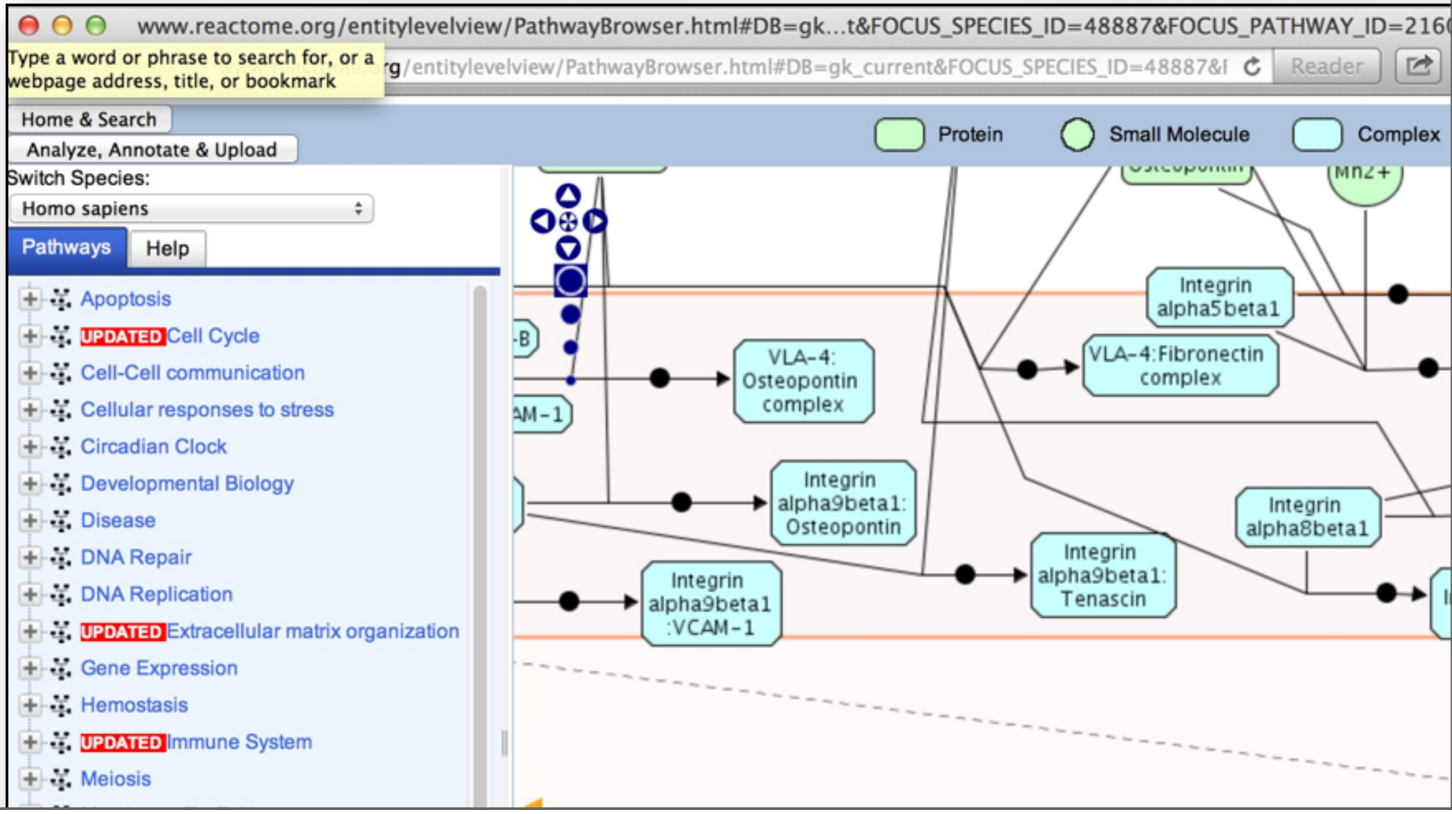
Martin et al., Autophagy, Jan. 2013

Being used in publications

Numerous software tools and databases

- API libraries are under development

See **http://sbgn.org** for more



Reactome Database — http://reactome.org

Outline

Introduction and motivation

COMBINE

SBML

SED-ML

SBGN

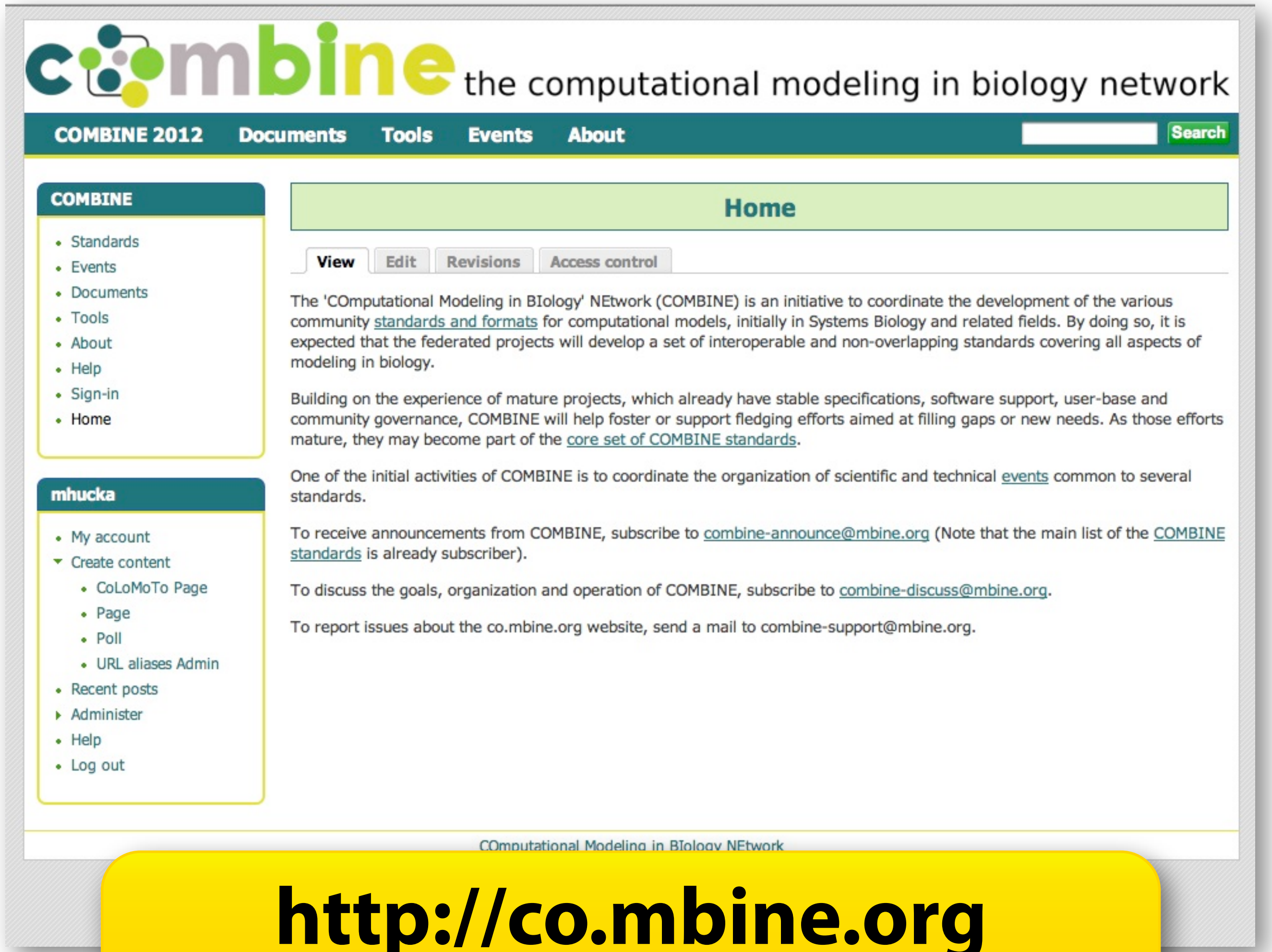
Conclusion

Attendees at SBML 10th Anniversary Symposium, Edinburgh, 2010

Such standards are the work of a great community

Get involved and make things better!

COMBINE (Computational Modeling in Biology Network)



Home

The 'COmputational Modeling in BIology' NEtwork (COMBINE) is an initiative to coordinate the development of the various community standards and formats for computational models, initially in Systems Biology and related fields. By doing so, it is expected that the federated projects will develop a set of interoperable and non-overlapping standards covering all aspects of modeling in biology.

Building on the experience of mature projects, which already have stable specifications, software support, user-base and community governance, COMBINE will help foster or support fledging efforts aimed at filling gaps or new needs. As those efforts mature, they may become part of the core set of COMBINE standards.

One of the initial activities of COMBINE is to coordinate the organization of scientific and technical events common to several standards.

To receive announcements from COMBINE, subscribe to combine-announce@mbine.org (Note that the main list of the COMBINE standards is already subscriber).

To discuss the goals, organization and operation of COMBINE, subscribe to combine-discuss@mbine.org.

To report issues about the co.mbine.org website, send a mail to combine-support@mbine.org.

COmputational Modeling in BIology NEtwork

http://co.mbine.org

Upcoming: HARMONY at U. Connecticut Health Center, May 20–23

- HARMONY = Hackathon on Resources for Modeling in Biology

COMBINE meeting planned for later this year

SBML was made possible thanks to funding from:

National Institute of General Medical Sciences (USA)
European Molecular Biology Laboratory (EMBL)
JST ERATO Kitano Symbiotic Systems Project (Japan) (to 2003)
JST ERATO-SORST Program (Japan)
ELIXIR (UK)
Beckman Institute, Caltech (USA)
Keio University (Japan)
International Joint Research Program of NEDO (Japan)
Japanese Ministry of Agriculture
Japanese Ministry of Educ., Culture, Sports, Science and Tech.
BBSRC (UK)
National Science Foundation (USA)
DARPA IPTO Bio-SPICE Bio-Computation Program (USA)
Air Force Office of Scientific Research (USA)
STRI, University of Hertfordshire (UK)
Molecular Sciences Institute (USA)

I'd like your feedback!

You can use this anonymous form:

http://tinyurl.com/mhuckafeedback