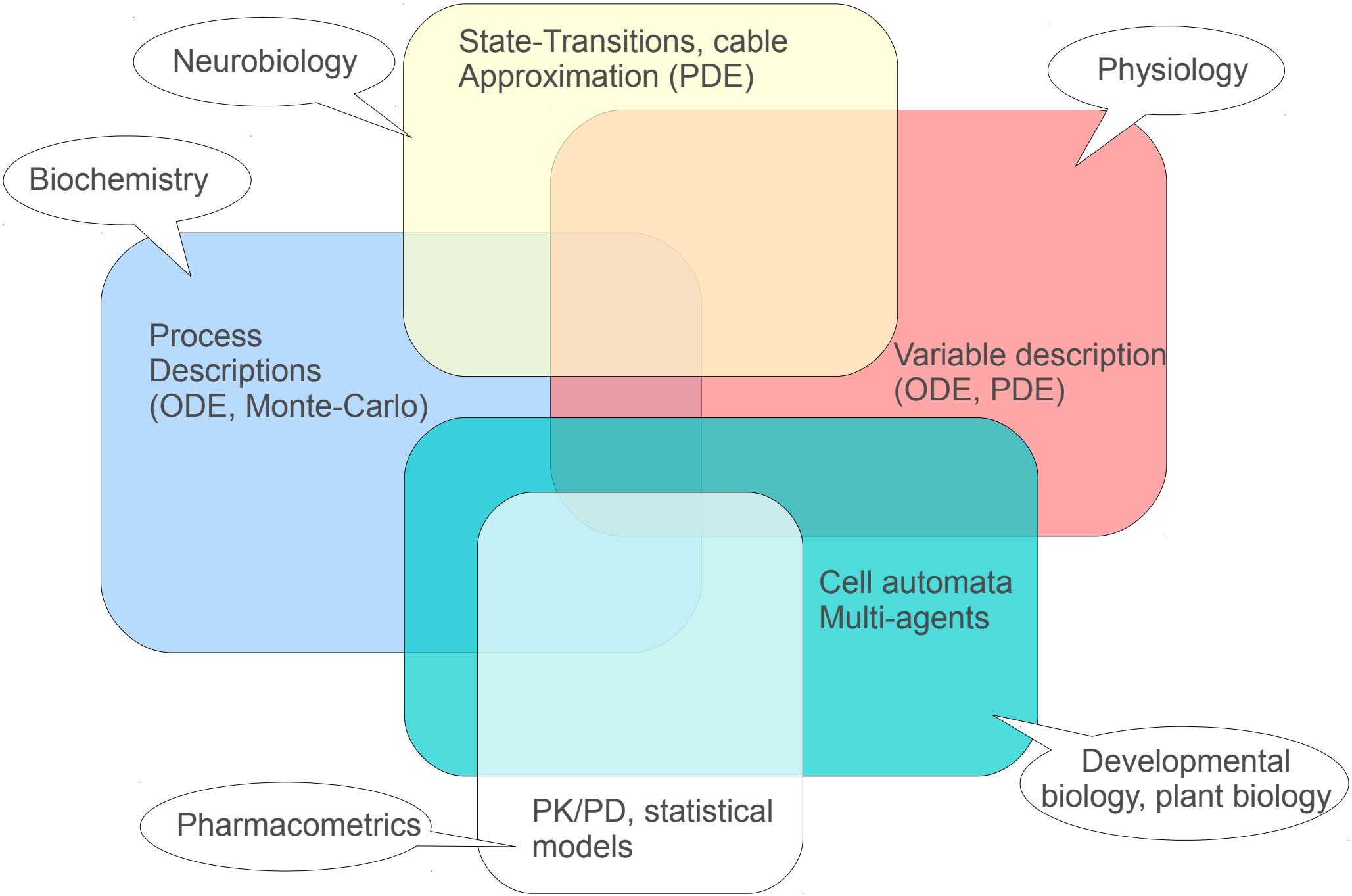
the combine

computational modeling in biology network

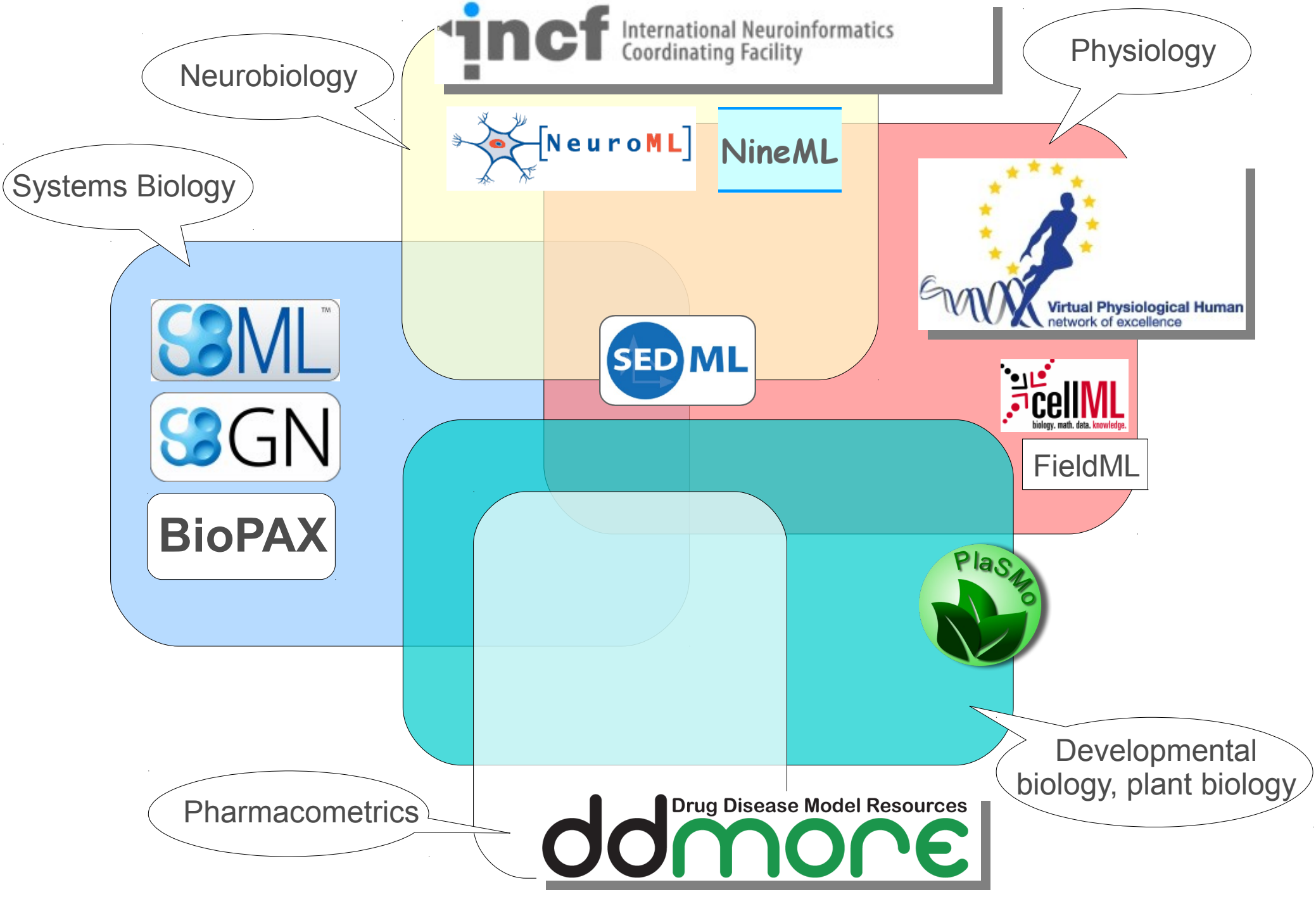
Standards for describing the whole life-cycle of modelling in the life sciences

Nicolas Le Novère EMBL-EBI

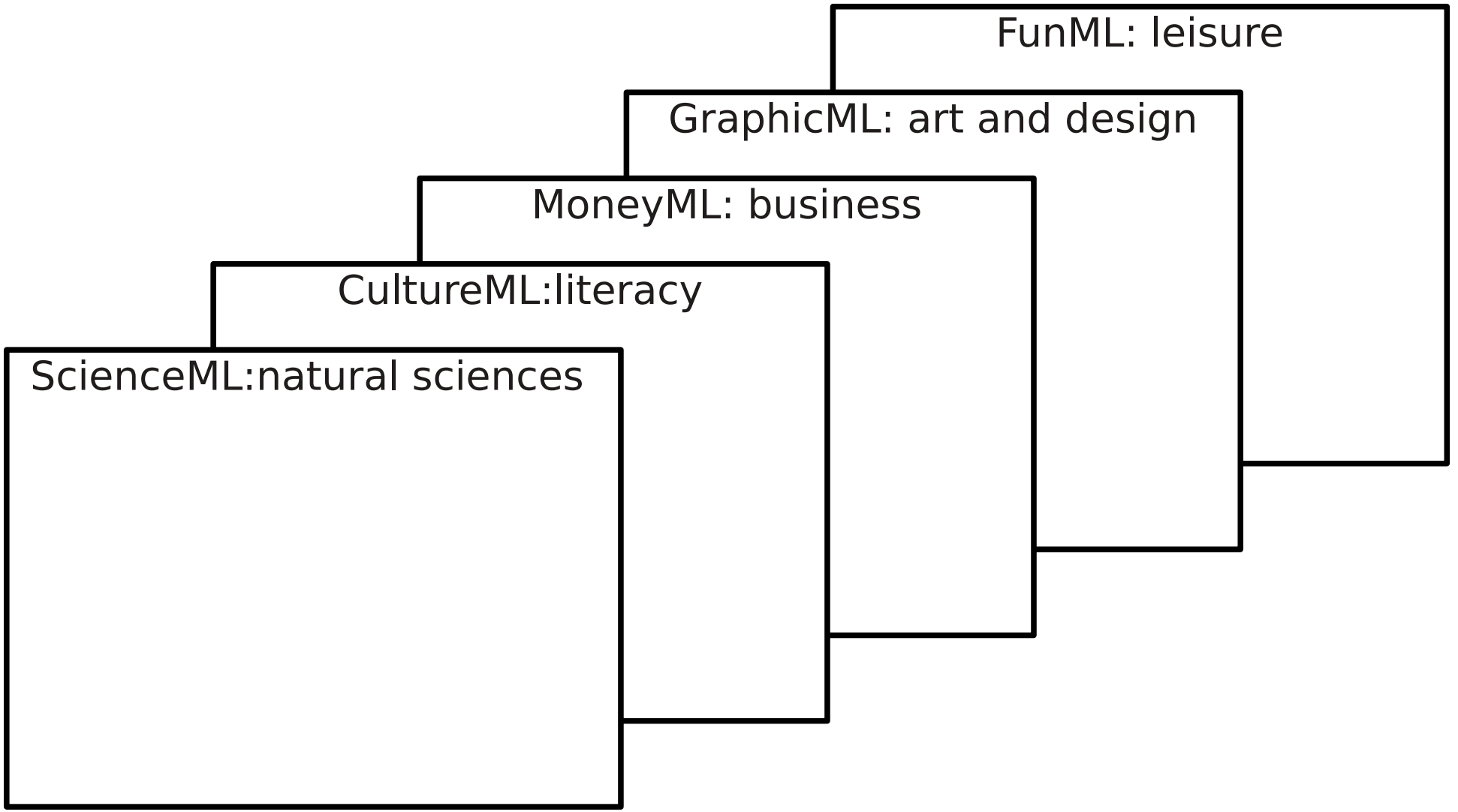
Many complementary modelling approaches



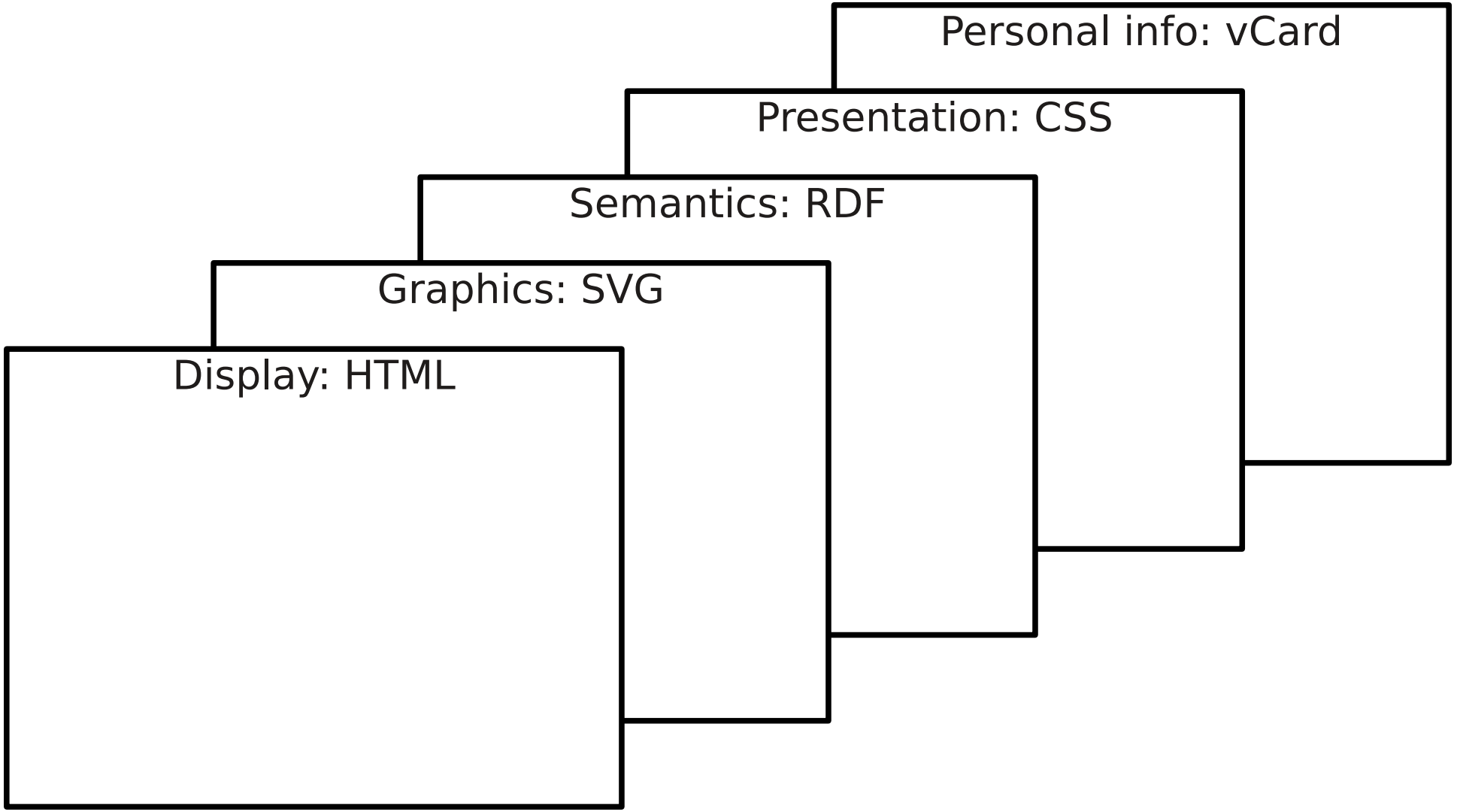
Parallel and redundant efforts



What if the world-wide web was built like this?



The correct way to do it



Threat to the development of standards for M&S

- Current efforts are entirely dependent on key people (SBML: Mike Hucka, CellML: Peter Hunter/Poul Nielsen, NeuroML: Padraig Gleeson, SBGN: NLN). Their disengagement probably means serious stalling.
- Current funding structure is fragile. Many different grants, sometimes only supporting meetings (SBGN), none of them are infrastructure rolling funding, often tied to individuals.
- Current efforts are not immune against intellectual property claims that would destroy the community (e.g. Caltech and SBML)
- Existing standards are developed with very different approaches, quality checks, and are based on completely different assumptions (e.g. NeuroML assumes some implicit knowledge while SBML requires explicit math).
- Specifications, APIs, test-suites need industry-grade support, incompatible with standard academic usages and possibilities.

http://co.mbine.org

the computational modeling in biology network

HARMONY 2012 | Standards | Documents | Tools | Events | About | Search

COMBINE

- Standards
- Events
- Documents
- Tools
- About
- Help
- Sign-in
- Home

lenov

- My account
- Create content
  - Page
  - Poll
- Recent posts
- Administer
- Help
- Log out

Home

View | Edit | Revisions | Access control

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 the 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.

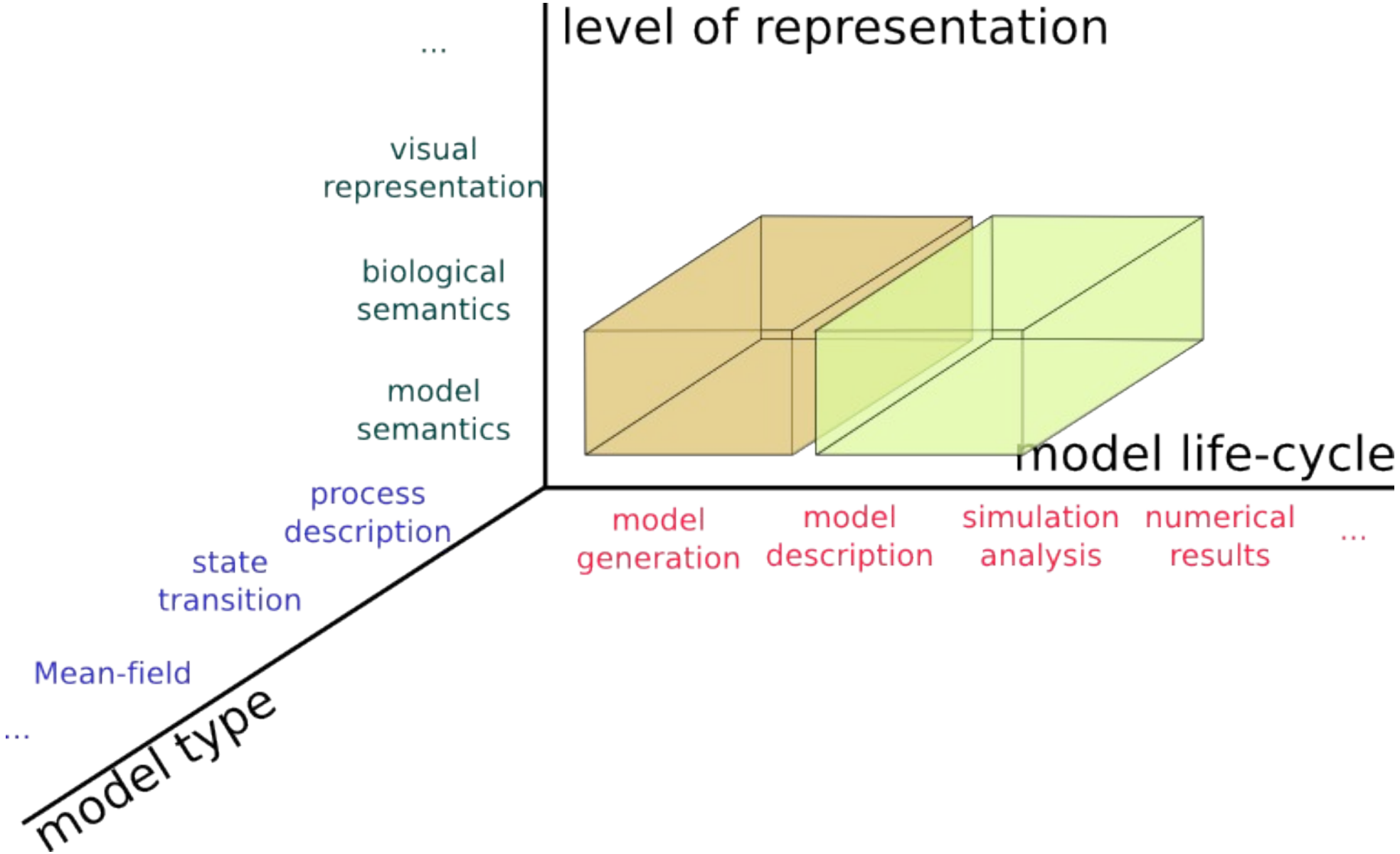
One of the initial activities of COMBINE is to coordinate the organization of scientific and technical events common to several standards.

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.

Vision: non-overlapping interoperable formats



COMBINE activities

- *Coordination* of standard development (no interference with the development itself)
  - Core set of COMBINE standards: stable, well-documented, (well-)supported

Criteria for inclusion of a standard in COMBINE

View | Edit | Revisions | Access control

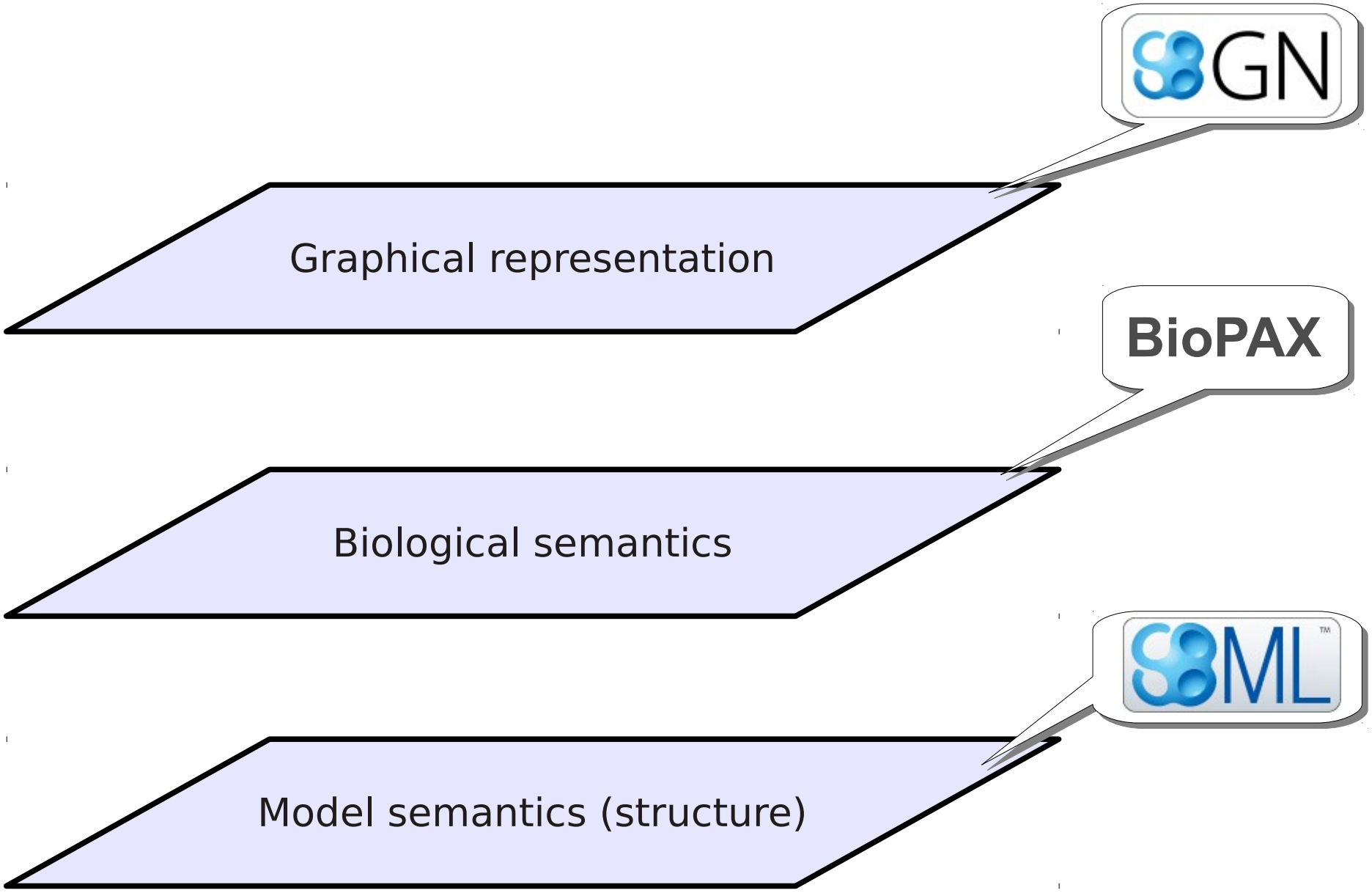
The aim of COMBINE is to help foster the development of a set of open, interoperable and non-overlapping standards in systems biology. In order to be included in the set of core COMBINE standards, an effort must, at a minimum, fulfill the following criteria:

1. The standard must cover aspects of modeling in biology that are significantly different from those already catered for by the existing set of COMBINE standards. This includes, but is not limited to, covering different modeling approaches, different biological scales or biological entities, different types of representations.
2. The standards must be described in technical specification documents as well as, when appropriate, formal specification languages (for instance, but not solely, XML Schemas, UML diagrams, and so on). The technical specifications must be precise enough to allow certifiable software support.
3. The specifications and other materials describing the specifications (including, for instance, XML Schemas and UML diagrams) must be publicly available free of charge to everyone and be unencumbered by licensing restrictions. If a specification material is covered by a license, it should allow use and redistribution by anyone. And example of a suitable license would be the creative commons with attribution.
4. The development of the standard must be open. The entire COMBINE community must be able to participate without exclusion. Proper ways of communication, must ensure that the community can express needs, criticisms and suggestions about all aspects of the standards. Examples include open development mailing lists with open archives, open source code repository, open website (no password protection).
5. The standard must be developed and used by more than a single team or organisation.
6. The development process must be led by editorial boards comprised of democratically elected members, possibly assisted by expert committees nominated by the editorial boards.
7. There must exist a mature software support, including standard API implementations, that facilitates the use of the standard. If possible, validation tools such as test suite, validators, etc., should exist as well.
8. The development of the standard must be stable; for instance, it must be supported by established teams and/or reliable funding sources. The standard must be actively developed.

The matrix of standard for models in systems biology

	Model descriptions	Simulations and analysis	results
Minimal requirements	MIRIAM	MIASE	
Data-models	SBML	SED ML	NuML
Terminologies	SBO	KISAO	TEDDY

Disentangling the level of discourse



BioPAX

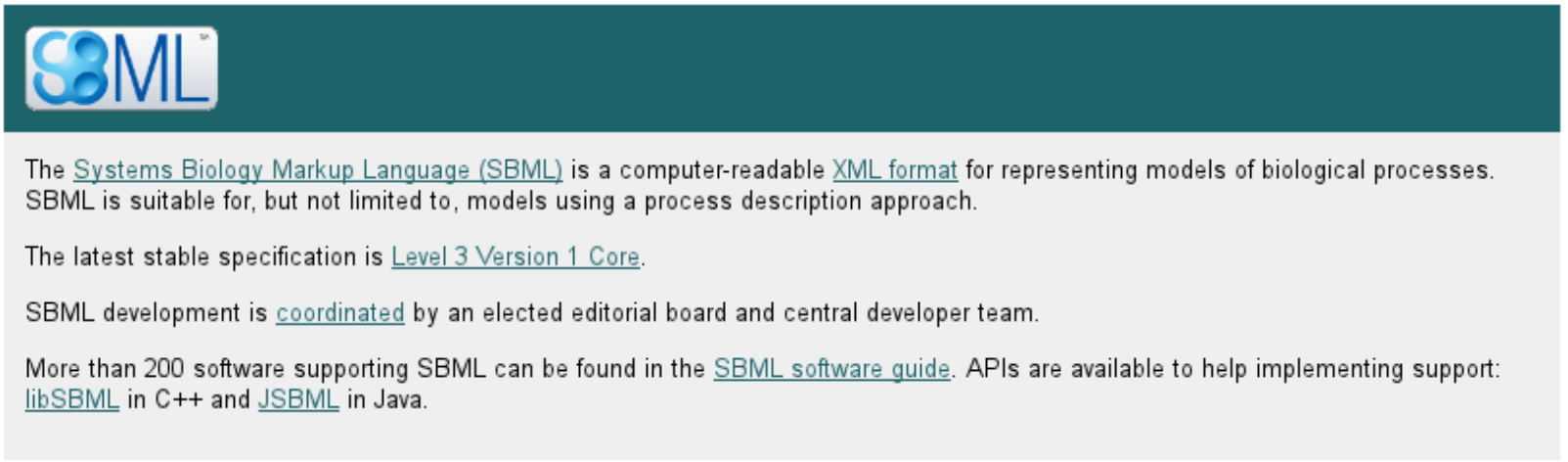
BioPAX is a standard language that aims to enable integration, exchange, visualization and analysis of biological pathway data. It is expressed in OWL.

The last specification of is BioPAX Level 3.

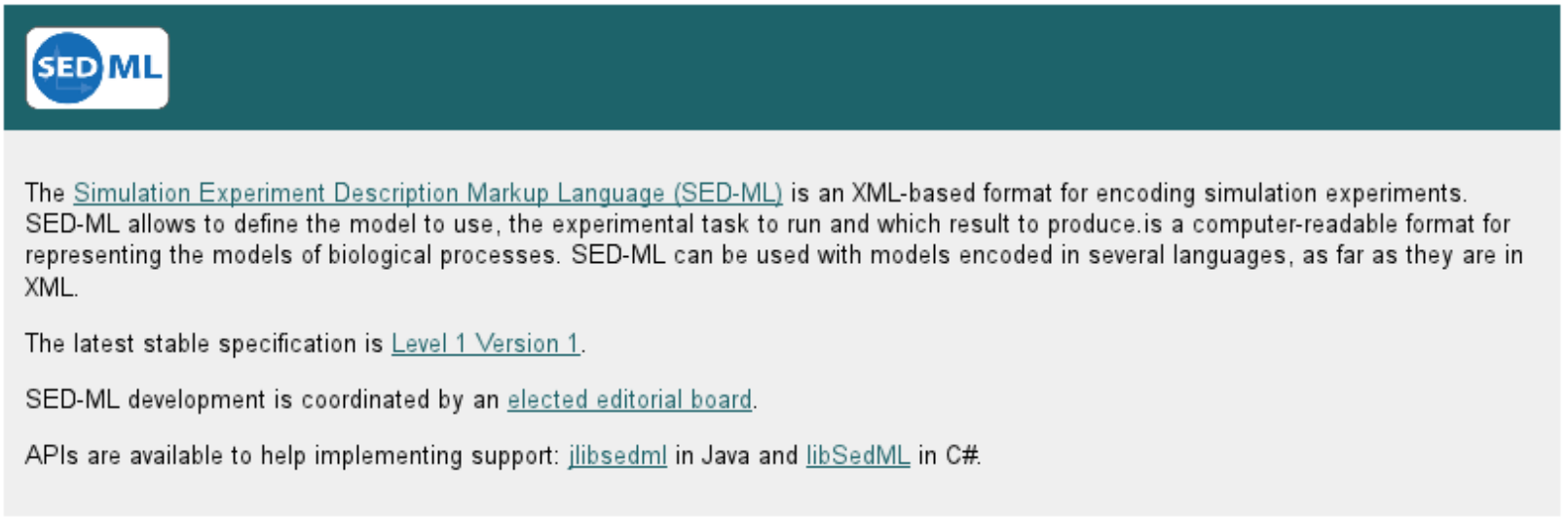
BioPAX development is coordinated by an elected editorial board and a Scientific Advisory Board.

BioPAX is supported by many pathway database or processing tools. An API is available to help implementing support: Paxtools

```
<rdf:RDF>
  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource=
       "http://www.biopax.org/release/biopax-level3.owl#"/>
  </owl:Ontology>
  <bp:Stoichiometry rdf:about="r3a_KK_STOICHIOMETRY">
    <bp:stoichiometricCoefficient rdf:datatype="xsd:float">
      1.0
    </bp:stoichiometricCoefficient>
    <bp:physicalEntity rdf:resource="r3a_KK"/>
  </bp:Stoichiometry>
  <bp:PhysicalEntity rdf:about="r3a_KK">
    <bp:cellularLocation rdf:resource="cytosol"/>
    <bp:memberPhysicalEntity rdf:resource="KK"/>
  </bp:PhysicalEntity>
```




```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2/version4"
      Level="2" version="4">
  <model name="simpleModel">
    <listOfCompartments>
      <compartment id="comp1" size="1"/>
    </listOfCompartments>
    <listOfSpecies>
      <species id="A" compartment="A" initialAmount="1"/>
      <species id="A" compartment="B" initialAmount="0"/>
    </listOfSpecies>
    <listOfReactions>
      <reaction id="AtoB">
        <listOfReactants>
```



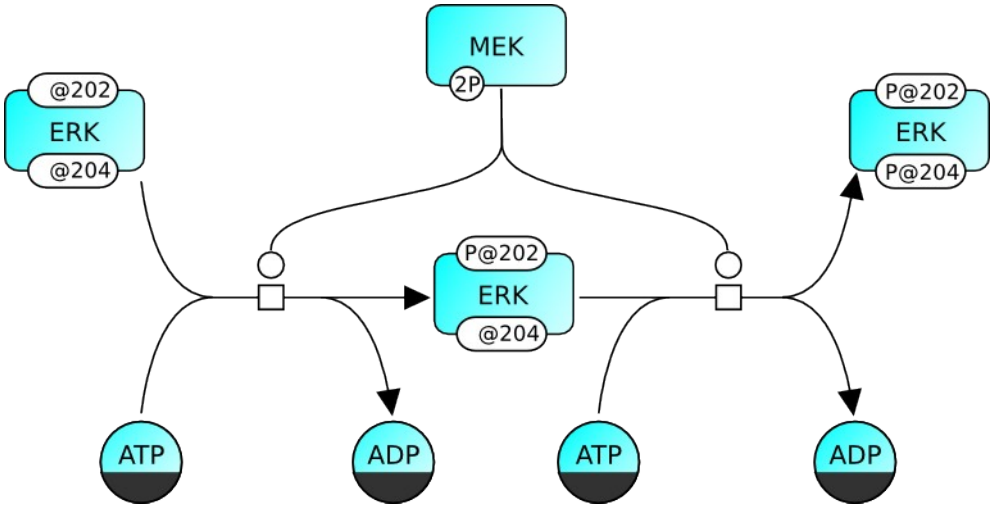

```
<?xml version="1.0" encoding="utf-8"?>
  <sedML xmlns:math="http://www.w3.org/1998/Math/MathML"
         xmlns="http://sed-ml.org/" level="1" version="1">
    <listOfSimulations>
      <uniformTimeCourse id="simulation1" initialTime="0"
 outputStartTime="0" outputEndTime="1000" numberOfPoints="1000">
        <algorithm kisaoID="KISAO:0000088" />
      </uniformTimeCourse>
    </listOfSimulations>
    <listOfModels>
      <model id="repressilator"
       language="urn:sedml:language:sbml.level-2.version-3"
       source="urn:miriam:biomodels.db:BIOMD0000000012" >
      </model>
```

The Systems Biology Graphical Notation (SBGN), is a set standard graphical languages to describe biological knowledge. It is currently made up of three languages describing Process Descriptions, Entity Relationships and Activity Flows.

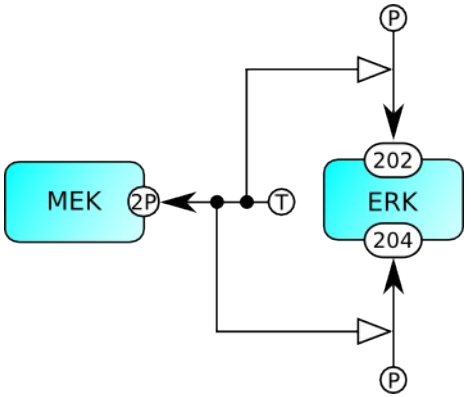
The last specifications are SBGN PD Level 1 Version 1.3, SBGN ER Level 1 Version 1.2 and SBGN AF Level 1 Version 1.

SBGN development is coordinated by an elected editorial board and a Scientific Committee.

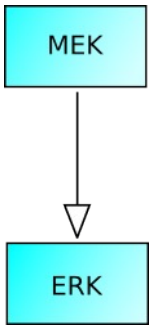
Several data resources and software claim support for SBGN. An API is available to help implementing support: libSBGN



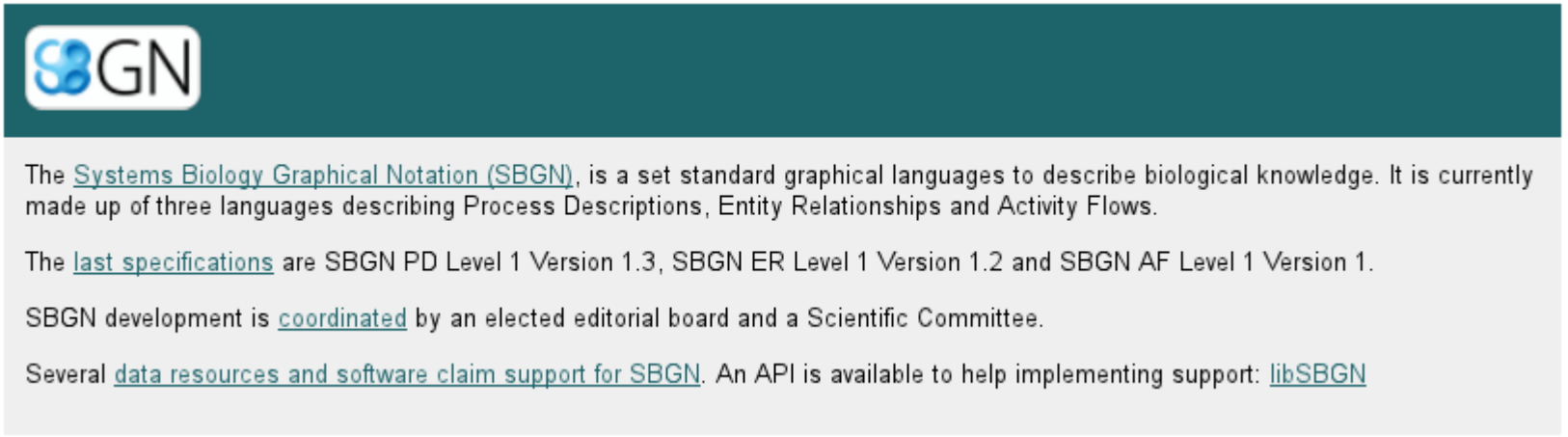
Process Descriptions



Entity Relationships



Activity flows




```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<sbgn xmlns="http://sbgn.org/libsbgn/pd/0.1">
    <map>
        <glyph id="glyph9" class="macromolecule">
            <label text="hexokinase"/>
            <bbox y="40.0" x="170.0" h="60.0" w="120.0"/>
        </glyph>
        <glyph id="glyph0" class="simple chemical">
            <label text="glucose"/>
            <bbox y="140.0" x="90.0" h="60.0" w="60.0"/>
        </glyph>
        <glyph id="glyph8" class="process">
            <bbox y="160.0" x="220.0" h="20.0" w="20.0"/>
            <port id="glyph8.1" y="170.0" x="210.0"/>
```

COMBINE activities

- *Coordination* of standard development (no interference with the development itself)
  - Core set of COMBINE standards: stable, well-documented, (well-)supported
  - Associated standardisation efforts: meant to be used across the core set

MIRIAM URIs

MIRIAM Unique Resource Identifiers allow one to uniquely and unambiguously identify an entity in a stable and perennial manner. MIRIAM Registry is a set of services and resources that provide support for generating, interpreting and resolving MIRIAM URIs. Through the new Identifiers.org technology, MIRIAM URIs can now be dereferenced in a flexible and robust way.

MIRIAM URIs are used by SBML and BioPAX controlled annotation schemes.

identifiers.org

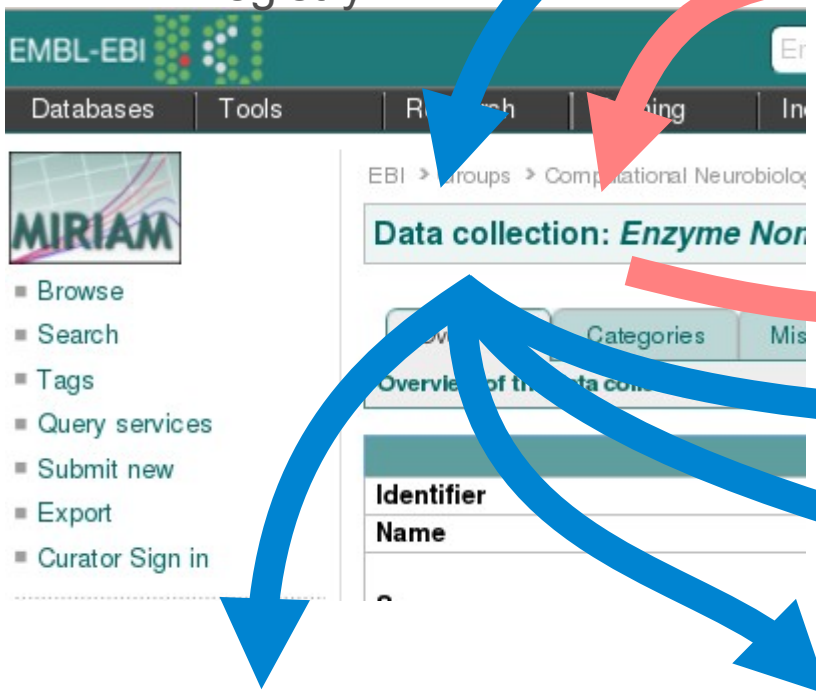
http://identifiers.org/obo.go/GO:0000186

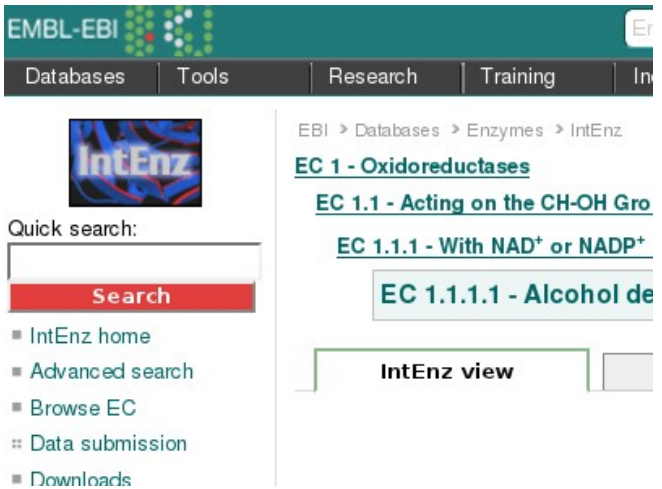
http://identifiers.org/pubmed/22140103

http://identifiers.org/ec-code/1.1.1.1

http://identifiers.org/ec-code/1.1.1.1
?resource=MIR:00100001

MIRIAM Registry



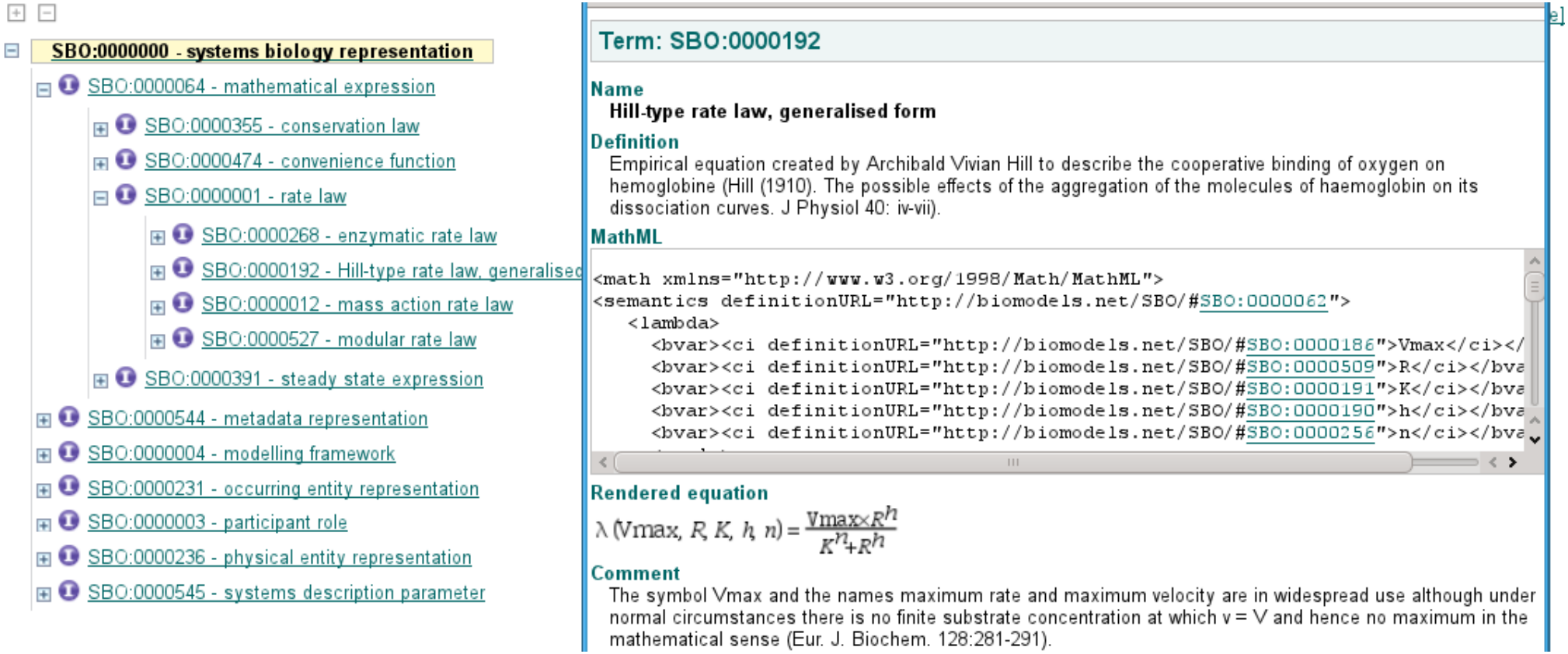


Systems Biology Ontology

The Systems Biology Ontology (SBO) is a set of controlled, relational vocabularies of terms commonly used in Systems Biology, and in particular in computational modeling.

Each element of an SBML file carries an optional attribute sboTerm which value must be a term from SBO.

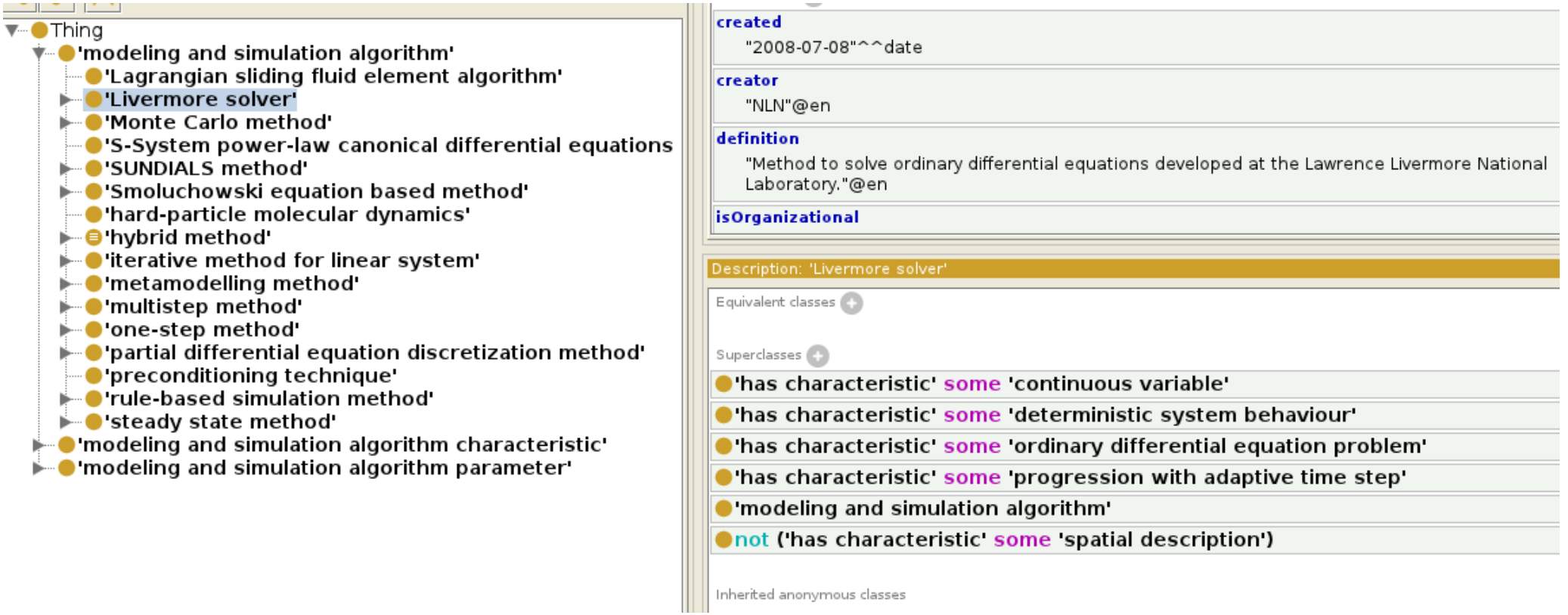
Each symbol of SBGN is associated with an SBO term.



Kinetic Simulation Algorithm Ontology

The Kinetic Simulation Algorithm Ontology (KiSAO) describes existing algorithms and their inter-relationships through their characteristics and parameters.

KiSAO is used in SED-ML, which allows simulation software to automatically choose the best algorithm available to perform a simulation and unambiguously refer to it.



BioModels.net qualifiers

BioModels.net qualifiers are standardized relationships (predicates) that specify the relation between an object represented in a description language and the external resource used to annotate it. The relationship is rarely one-to-one, and the information content of an annotation is greatly increased if one knows what it represents, rather than only know it is "related to" the model component.

encodes, encodement

The biological entity represented by the model element encodes, directly or transitivity, the subject of the referenced resource (biological entity B). This relation may be used to express, for example, that a specific DNA sequence encodes a particular protein.

hasPart, part

The biological entity represented by the model element includes the subject of the referenced resource (biological entity B), either physically or logically. This relation might be used to link a complex to the description of its components.

hasProperty, property **[new]**

The subject of the referenced resource (biological entity B) is a property of the biological entity represented by the model element. This relation might be used when a biological entity exhibits a certain enzymatic activity or exerts a specific function.

hasVersion, version

The subject of the referenced resource (biological entity B) is a version or an instance of the biological entity represented by the model element. This relation may be used to represent an isoform of modified form of a biological entity.

is, identity

The biological entity represented by the model element has identity with the subject of the referenced resource (modeling object B).

COMBINE activities

- *Coordination* of standard development (no interference with the development itself)
  - Core set of COMBINE standards: stable, well-documented, (well-)supported
  - Associated standardization efforts: meant to be used across the core set
  - Related standardization efforts: Potential members of the core set, or format to interoperate with

The CellML language is an XML markup language to store and exchange computer-based mathematical models. CellML is being developed by the Auckland Bioengineering Institute at the University of Auckland and affiliated research groups.

FieldML

FieldML's (Field Modelling/Markup Language) goal is to be a declarative language for building hierarchical models represented by generalized mathematical fields. Its primary use will be to represent the dynamic geometry and solution fields from computational models of cells, tissues and organs.

NeuroML is an XML format that facilitates the exchange of neuronal models, including electrical behavior, morphology, and connectivity.

NineML

The Network Interchange for Neuroscience Modeling Language (NineML) - is a language developed by the International Neuroinformatics Coordinating Facility (INCF) and designed for the description of large networks of spiking neurons.

NuML

The Numerical Markup Language (NuML) (pronounce "neumeul" and not "new em el", that sounds like NewML) is a simple XML format to exchange multidimensional arrays of numbers to be used with model and simulation descriptions. NuML was initially developed as part of the Systems Biology Results Markup Language (SBRML).

PSI-MI

The Proteomics Standards Initiative Molecular Interaction XML Format is a a data exchange format for molecular interactions developed by the the HUPO Proteomics Standards Initiative

The Synthetic Biology Open Language is a language for the description and the exchange of synthetic biological parts, devices, and systems.

COMBINE activities

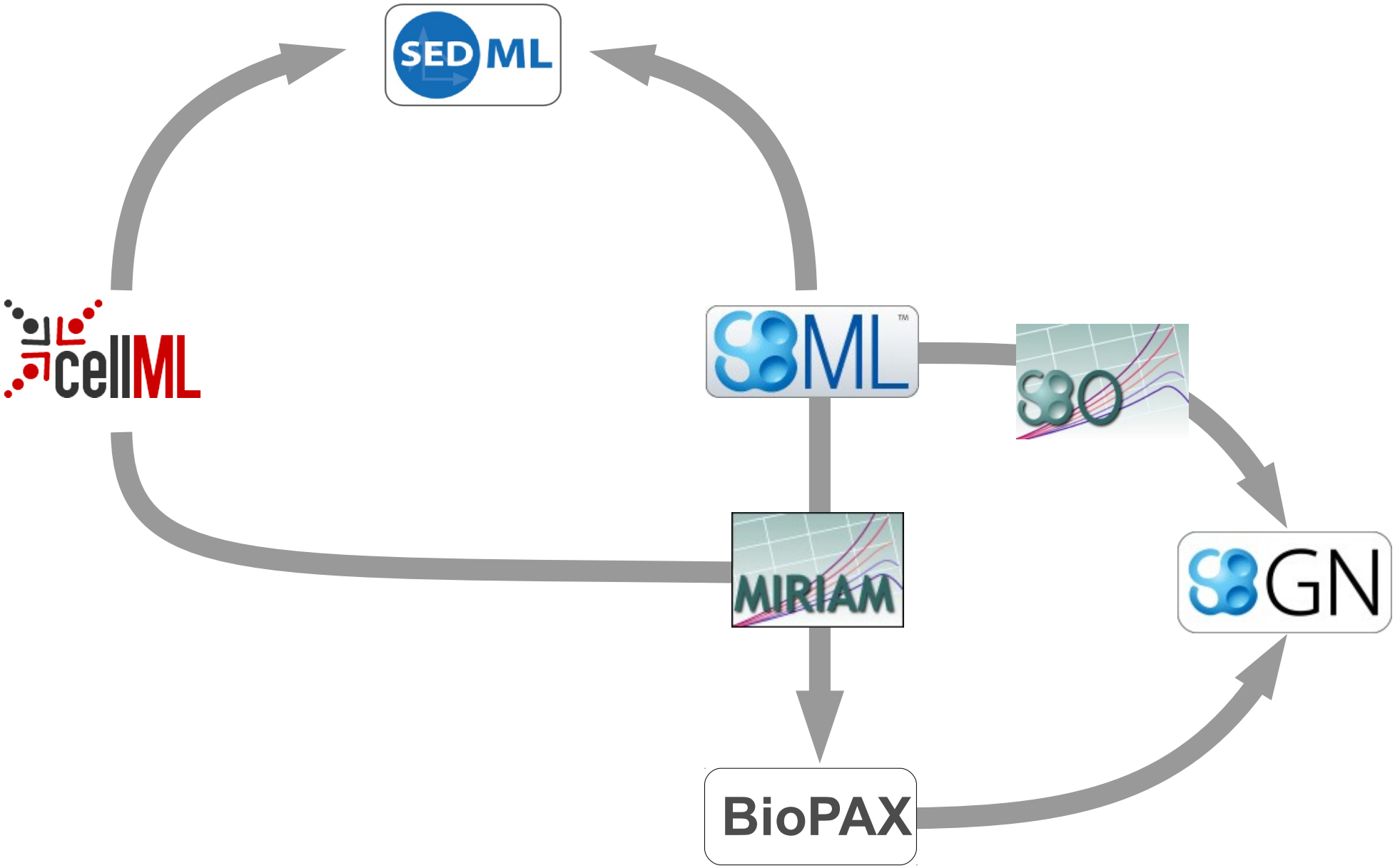
- *Coordination* of standard development (no interference with the development itself)
  - Core set of COMBINE standards: stable, well-documented, (well-)supported
  - Associated standardization efforts: meant to be used across the core set
  - Related standardization efforts: Potential members of the core set, or format to interoperate with
- Organisation of joint meetings, replacing the former standard specific ones
  - COMBINE forum: replace SBML and SBGN forums, BioPAX F2F
  - HARMONY hackathon: replace SBML and SBGN hackathons

- COMBINE forum 2010
  - 6-9 October 2010, Edinburgh
  - 82 attendees
- COMBINE forum 2011
  - 3-7 September 2011, Heidelberg
  - 83 attendees
- COMBINE forum 2012
  - 14-19 August 2012, Toronto
  - http://co.mbine.org/events/COMBINE_2012
- HARMONY 2011
  - 18-22 April 2011, New-York city
  - 59 attendees
- HARMONY 2012
  - 21-25 May 2012, Maastricht
  - http://co.mbine.org/events/HARMONY_2012
  - **Registration open**

COMBINE activities

- *Coordination* of standard development (no interference with the development itself)
  - Core set of COMBINE standards: stable, well-documented, (well-)supported
  - Associated standardization efforts: meant to be used across the core set
  - Related standardization efforts: Potential members of the core set, or format to interoperate with
- Organisation of joint meetings, replacing the former standard specific ones
  - COMBINE forum: replace SBML and SBGN forums, BioPAX F2F
  - HARMONY hackathon: replace SBML and SBGN hackathons
- Develop Standard Operating Procedures and best practices (started)
- Act as the voice of computational modeling standards for funders, publishers and policy makers (wishful thinking)
- Act as a funding structure (just a dream)

Fledging standards interoperability



Coordination

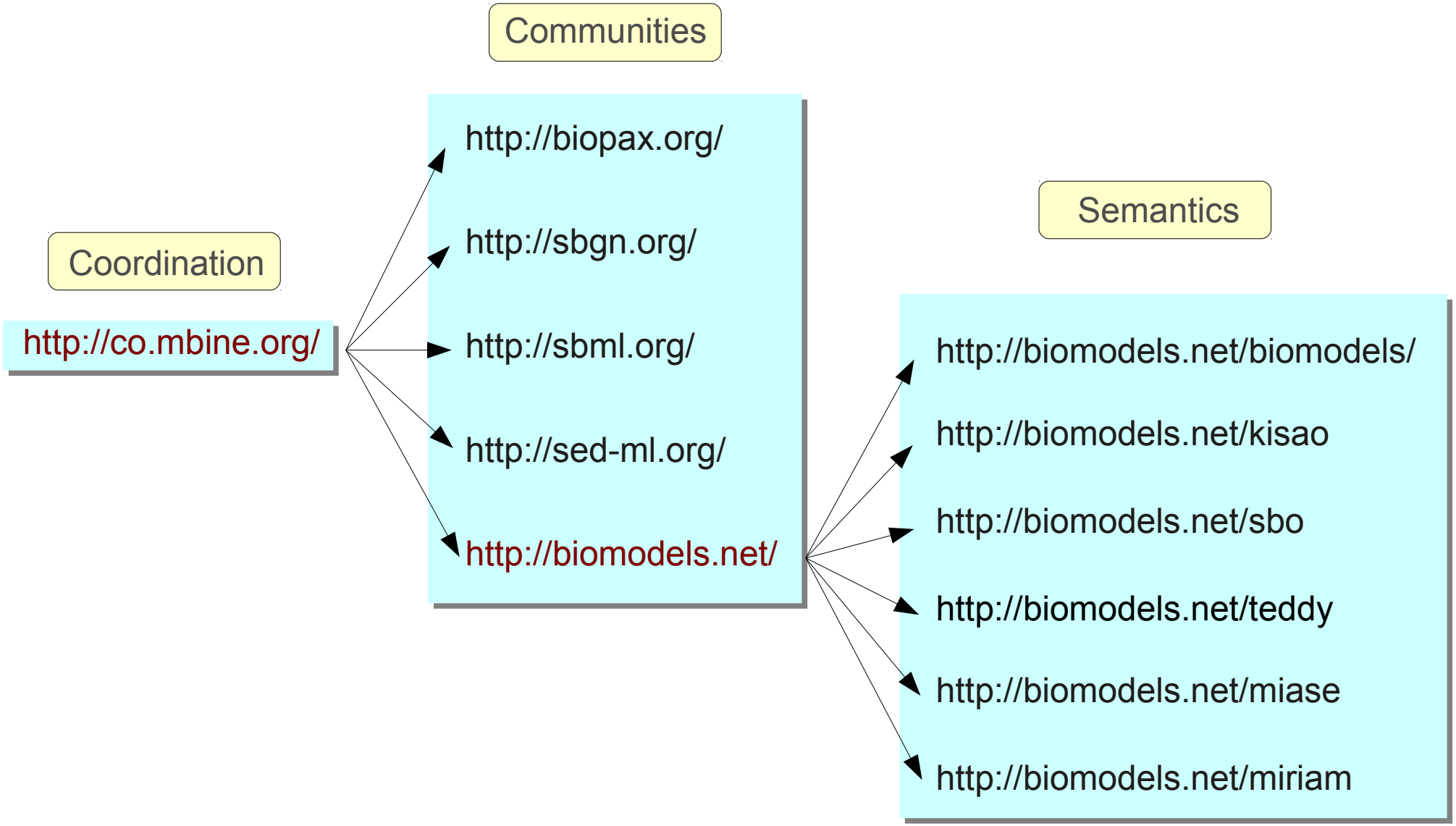
The various COMBINE activities are organized by a variety of individuals acknowledged on the relevant pages. Pending the development of a proper administrative structure, the global effort is led by three acting coordinators:

Gary D. Bader (Ph.D. Biochemistry) works on biological network analysis and pathway information resources as an Assistant Professor at The Donnelly Centre at the University of Toronto. He has been involved in leading development of protein interaction and pathway databases and standards, including the BioPAX biological pathways exchange language.

Michael Hucka (Ph.D. in Computer Science and Engineering), staff scientist at Caltech. He has chaired the SBML effort by community consensus since 2003. Today he works on all aspects of SBML and is involved with BioModels.net activities.

Nicolas Le Novère (Ph.D. in Molecular Pharmacology). Leads a research group at the European Bioinformatics Institute. His interests include neural signal transduction and computational modeling of biological processes. His group maintains BioModels Database and SBO. He is also involved in the development of SBML and SBGN.

Where to find more information?



What about NeuroML?

What about NeuroML?

- Currently a *Related Standardization Effort*

What about NeuroML?

- Currently a *Related Standardization Effort*
- Should join the *core set of COMBINE standards*

What about NeuroML?

- Currently a *Related Standardization Effort*
- Should join the *core set of COMBINE standards*
- Compliance with criteria
  - **1) cover different aspects of biology**
  - **2) technical specification**
  - 3) all specifications documents free and open
  - **4) development must be open to all**
  - **5) more than one team actively developing**
  - **6) democratically elected editorial board**
  - **7) mature software support including API implementation**
  - **8) development stable and financial support available**

Acknowledgements

SBML editors: *Frank Bergmann*, *Andrew Finney, Stefan Hoops*, *Michael Hucka*, Nicolas Le Novère, Sarah Keating, Chris Myers, *Sven Sahle*, *Herbert Sauro*, Jim Schaff, Lucian Smith, *Darren Wilkinson*

SBGN editors: Emek Demir, Nicolas Le Novère, *Huaiyu Mi*, *Stuart Moodie*, Falk Schreiber, Anatoly Sorokin, Alice Villéger

SED-ML editors: Richard Adams, Franck Bergmann, *Nicolas Le Novère,* Andrew Miller, David Nickerson, Dagmar Waltemath

BioPAX editors: Emek Demir, Peter d'Eustachio, Huaiyu Mi, Oliver Ruebenackacker, Andrea Splendiani

Metadata: Mélanie Courtot, Nick Juty, Camille Laibe, Dagmar Waltemath, Anna Zhukova

The whole community of Computational Systems Biology

The EBI group Computational Systems Neurobiology