Presentation by Mike Hucka at

Third NeuroML Development Workshop,

London, UK, March 2011

http://www.neuroml.org/workshop2011.php

MIRIAM Resources

On behalf of Camille Laibe and the team @ EBI

Michael Hucka, Ph.D.
California Institute of Technology
Pasadena, California, USA

SBML = Systems Biology Markup Language

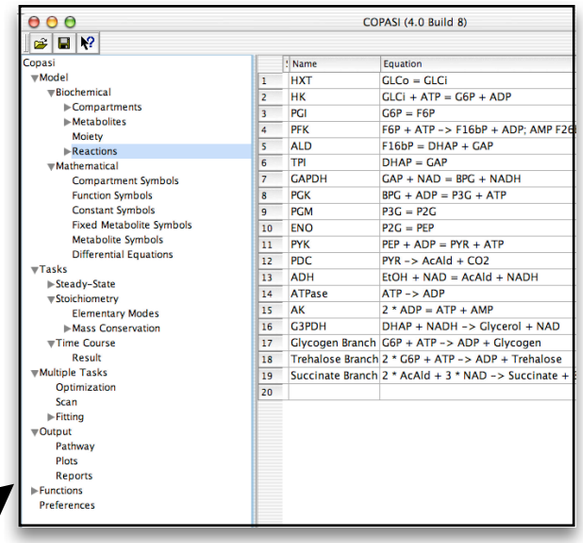
Format for representing quantitative models

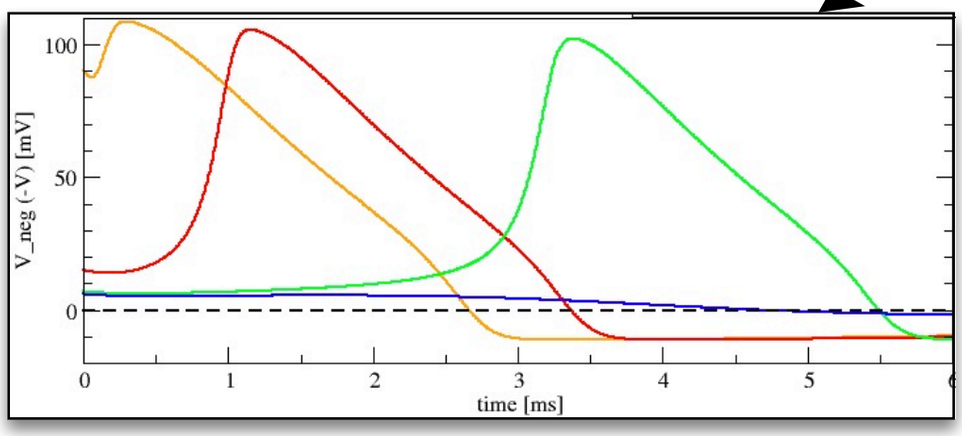
- Defines object model + rules for its use

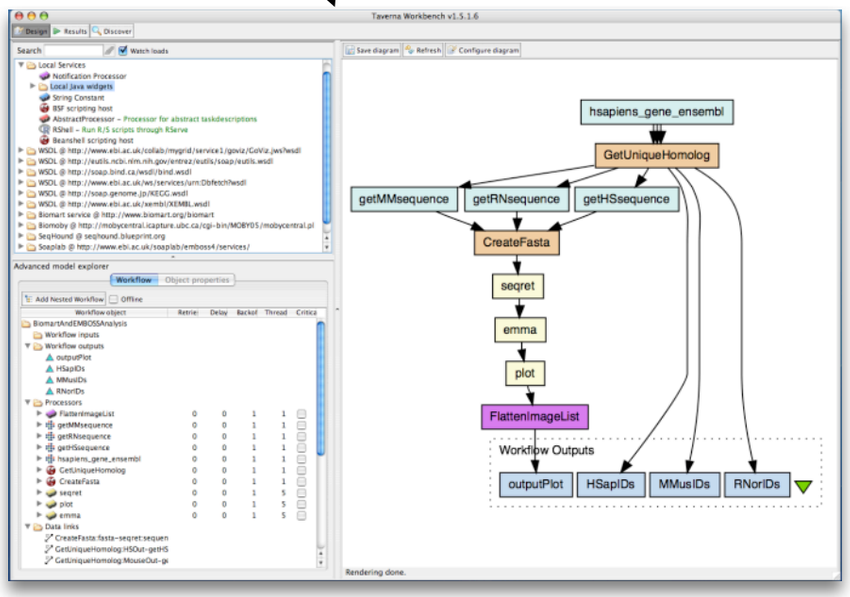
Neutral with respect to modeling framework

- ODE vs. stochastic vs. ...

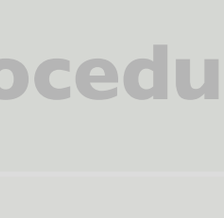
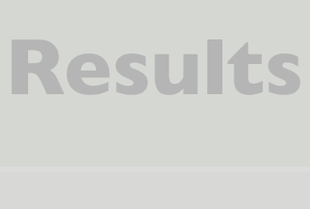
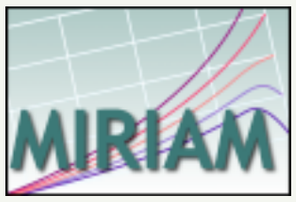
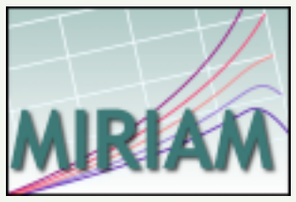
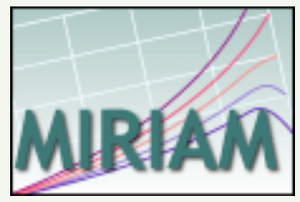
A lingua franca for software







	Model	Procedures	Results
Representation format	SBML	SED ML	SBRML
Minimal info requirements	MIRIAM	MIASE	?
Semantics— *Mathematical*	SBO	KiSAO	TEDDY
Other	MIRIAM annotations	MIRIAM annotations	MIRIAM annotations

	Model	Procedures	Results
Representation format			SBRML
Minimal info requirements			?
Semantics— *Mathematical*			
Other	 annotations	 annotations	 annotations

What is MIRIAM?

MIRIAM = "Minimum Information Requested In the Annotation of Models"

Addresses 2 general areas

- Requirements for reference correspondence
- Scheme for encoding annotations
  - Annotations for attributing model creators & sources
  - Annotations for referring to external data resources

MIRIAM is **not** specific to SBML

What is MIRIAM?

MIRIAM = "Minimum Information Requested In the Annotation of Models"

Addresses 2 general areas

- Requirements for reference correspondence
- **Scheme for encoding annotations**
  - Annotations for attributing model creators & sources
  - Annotations for referring to external data resources

MIRIAM is **not** specific to SBML

E.g.: linking model entities to entities in external db's

BioModels Home | Browse models | Submit | Sign in | Support | About BioModels

BIOMD0000000030 - Markevich2004_MAPK_AllRandomElementary

SBML formats | Other formats | Actions | Submit Model Comment/Bug

Model | Overview | **Math** | Physical entities | Parameters

Reactions (20)

binding MAPKK on Tyr site of MAPK [MAPK] + [MAPKK] ↔ [MAPK_MAPKK_Y];

Math: $cell \times (k1 \times M \times MAPKK - (k\_1 \times M\_MAPKK\_Y))$ (Detail:)

Annotations:

set #1	bqbiol:isVersionOf	**Gene Ontology** mitogen-activated protein kinase kinase binc **Gene Ontology** mitogen-activated protein kinase binding
	bqbiol:isHomologTo	**Reactome** REACT_1780 **Reactome** REACT_495

tyr phosphorylation of MAPK [MAPK_MAPKK_Y] → [MAPK-PY] + [MAPKK];

binding of MAPKK on MAPK-PY [MAPK-PY] + [MAPKK] ↔ [MAPK-PY_MAPKK];

thr phosphorylation of MAPK [MAPK-PY_MAPKK] → [MAPK-PP] + [MAPKK];

binding of MAPKK on Thr site of MAPK [MAPK] + [MAPKK] ↔ [MAPK_MAPKK_T];

E.g.: linking model entities to entities in external db's

BioModels Home | Browse models | Submit | Sign in | Support | About BioModels

BIOMD0000000030 - Markevich2004_MAPK_AllRandomElementary

SBML formats | Other formats | Actions | Submit Model Comment/Bug

Model | Overview | **Math** | Physical entities | Parameters

Reactions (20)

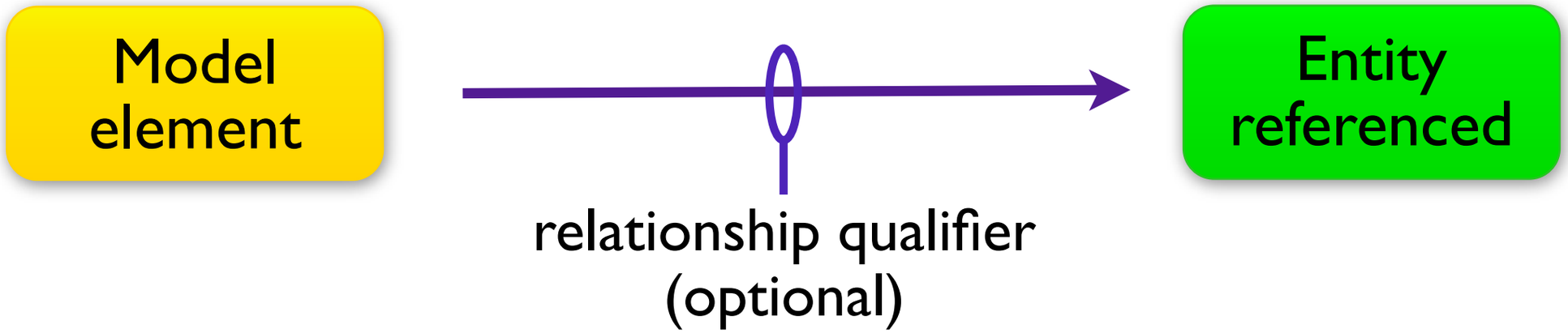
binding MAPKK on Tyr site of MAPK	[MAPK] + [MAPKK] ↔ [MAPK_MAPKK_Y];
Math:	$cell \times (k1 \times M \times MAPKK - (k\_1 \times M\_MAPKK\_Y))$ *(Detail:)*
Annotations:	set #1 — bqbiol:isVersionOf: **Gene Ontology** mitogen-activated protein kinase kinase bind; **Gene Ontology** mitogen-activated protein kinase binding bqbiol:isHomologTo: **Reactome** REACT_1780; **Reactome** REACT_495
tyr phosphorylation of MAPK	[MAPK_MAPKK_Y] → [MAPK-PY] + [MAPKK];
binding of MAPKK on MAPK-PY	[MAPK-PY] + [MAPKK] ↔ [MAPK-PY_MAPKK];
thr phosphorylation of MAPK	[MAPK-PY_MAPKK] → [MAPK-PP] + [MAPKK];
binding of MAPKK on Thr site of MAPK	[MAPK] + [MAPKK] ↔ [MAPK_MAPKK_T];

Why worry about standard ways of writing annotations?

Structured, machine-readable annotations increase your model's utility

- Allow more precise **identification** of model components
  - Understand model structure
  - Compare models
  - Integrate models
  - Search models
- Adds a **semantic layer**—integrates knowledge into the model
  - Understand the underlying biology
  - Reuse models
  - Convert models to other forms

Neat, yes? OK, *how* can you write such annotations?



MIRIAM says: express it as a tuple

{ Data type identifier (Required) , Data item identifier (Required) , Annotation qualifier (Optional) }

Format:

URI chosen from agreed-upon list

Syntax & value space depends on data type

Controlled vocabulary term

SBML defines a syntax for annotations

```
<species metaid="metaid_0000009" id="species_3" compartment="c_1">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
             xmlns:bqbiol="http://biomodels.net/biology-qualifiers/" >
      <rdf:Description rdf:about="#metaid_0000009">
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A15996"/>
            <rdf:li rdf:resource="urn:miriam:kegg.compound:C00044"/>
          </rdf:Bag>
        </bqbiol:is>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

SBML defines a syntax for annotations

```
<species metaid="metaid_0000009" id="species_3" compartment="c_1">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
             xmlns:bqbiol="http://biomodels.net/biology-qualifiers/" >
      <rdf:Description rdf:about="#metaid_0000009
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A15996"/>
            <rdf:li rdf:resource="urn:miriam:kegg.compound:C00044"/>
          </rdf:Bag>
        </bqbiol:is>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

Data references

SBML defines a syntax for annotations

```
<species metaid="metaid_0000009" id="species_3" compartment="c_1">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
             xmlns:bqbiol="http://biomodels.net/biology-qualifiers/" >
      <rdf:Description rd                                    09">
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A15996"/>
            <rdf:li rdf:resource="urn:miriam:kegg.compound:C00044"/>
          </rdf:Bag>
        </bqbiol:is>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

Relationship qualifier

Effective, interoperable annotations need agreement

Intuitive but fundamentally **bad** approaches:

- Plain text
- Unregulated XML
- URLs

Qualities to seek in a **good** approach:

- Identifiers are **unique** and **unambiguous**
- Identifiers are resolvable to a unique **resource** and **entity** within it
- Identifiers are permanent and **perennial**
- Scheme conforms to or builds on existing standards
- Freely usable

MIRIAM Resources uses URIs as unique identifiers

"Term #1.1.1.1 (alcohol dehydrogenase) in the Enzyme Commission's *Enzyme Nomenclature* database"

⇒ urn:miriam:ec-code:1.1.1.1

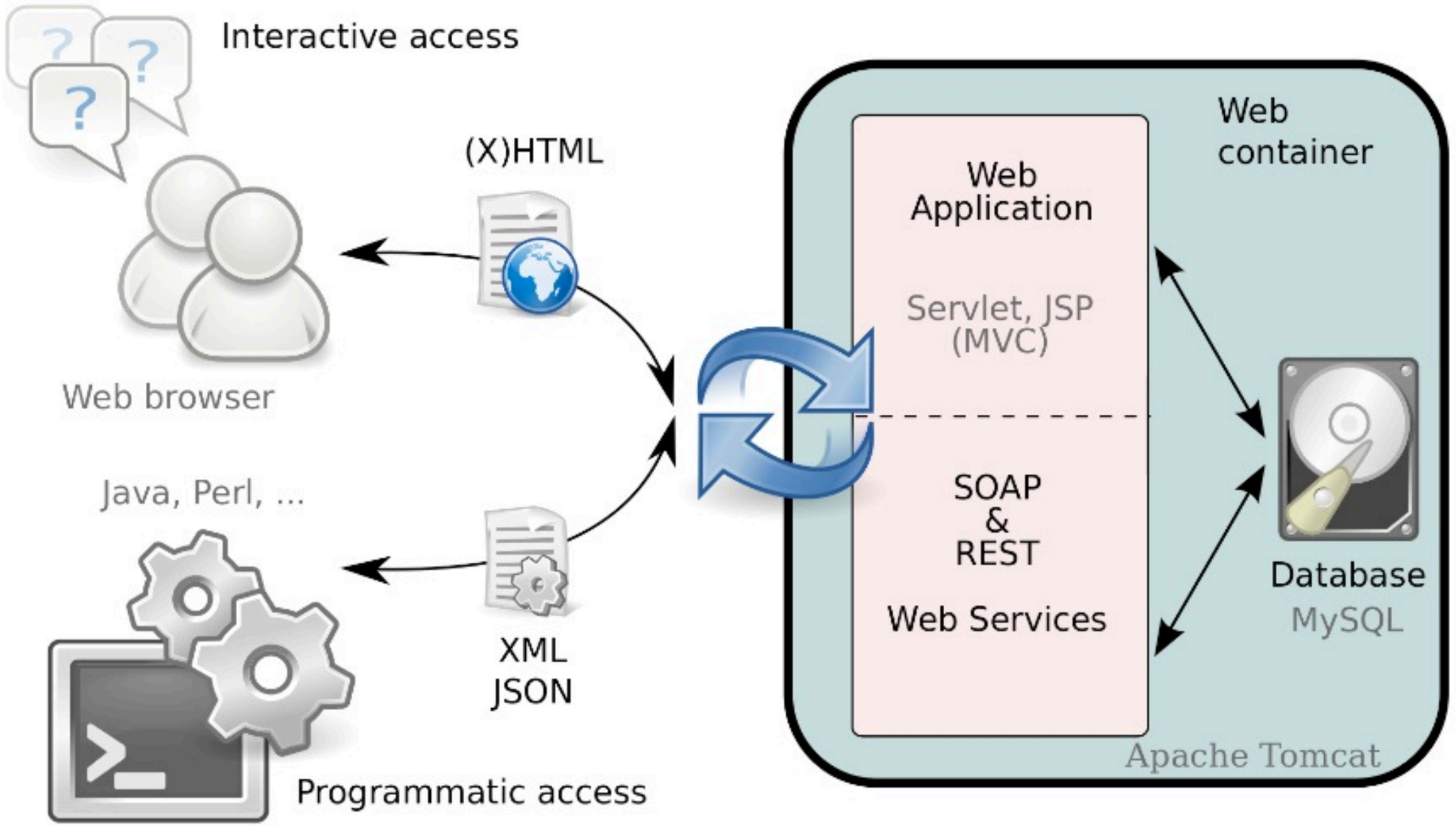
URI scheme established by the MIRIAM project

Chosen by the creator of the entry in MIRIAM Resources

Identifier of an entity within the resource

Main objectives of MIRIAM Resources

1. A central catalog of agreed-upon standard URIs for data types
2. A means for the user community to add and update entries
2. Resolution services for software via standard protocols (SOAP & REST)



Resource dictionary & resource resolution

http://www.ebi.ac.uk/miriam

Community-maintained

EBI › Groups › Computational Neurobiology › Research › MIRIAM Resources

MIRIAM Resources

Browse the data types

Brief overview of the different data types stored in *MIRIAM Database*.

Next page ⇨

Name	URI	Definition
3DMET	urn:miriam:3dmet	3DMET is a database collecting three-dim metabolites
Aclame	urn:miriam:aclame	ACLAME is a database dedicated to the colle genetic elements (MGEs) from various sour genomes, plasmids and
Anatomical Therapeutic Chemical	urn:miriam:atc	The Anatomical Therapeutic Chemical (ATC) c substances into different groups according to act and their therapeutic, pharmacological ar classified in groups at five different levels; Dr groups (1st level), with pharmacological/thera 3rd and 4th levels are chemical/pharmacolog 5th level is the chemical substance. The Anato classification system and the Defined Daily D

Resource dictionary & resource resolution

http://www.ebi.ac.uk/miriam

Community-maintained

EBI › Groups › Computational Neurobiology › Research › MIRIAM Resources

MIRIAM Resources

Browse the data types

Brief overview of the different

Next page

Name	UR
3DMET	urn:miriam
Aclame	urn:miriam
Anatomical Therapeutic Chemical	urn:miria

Data type: *3DMET*

General | Tags | Example Usage | Web Services

General information about the data type

Name		
Identifier		MIR:00000066
Name		3DMET
URIs		
MIRIAM URN		urn:miriam:3dmet
Deprecated		*No deprecated URI*
Information		
Definition		3DMET is a database collecting three-dimensional structures of natural metabolites.
Identifier Pattern		^B\d{5}$
Physical Locations		
Resource MIR:00100095	**Access URL**	http://www.3dmet.dna.affrc.go.jp/html/**$id**.html [Example: B00162]
	Website	http://www.3dmet.dna.affrc.go.jp/
	Description	3DMET database
	Institution	Bioinformatics Research Unit, National Institute of Agrobiological Sicences, Japan
References		
URL(s)		http://www.jsbi.org/journal/GIW04/GIW04P058.pdf
Miscellaneous		
Date of creation		2009-01-23 10:00:42 GMT
Date of last modification		2009-12-28 18:23:44 GMT

Go back to the list of data types

Suggest modifications to this data type

classification system and the Defined Daily D

MIRIAM identifiers now in use by many other projects

Data resources

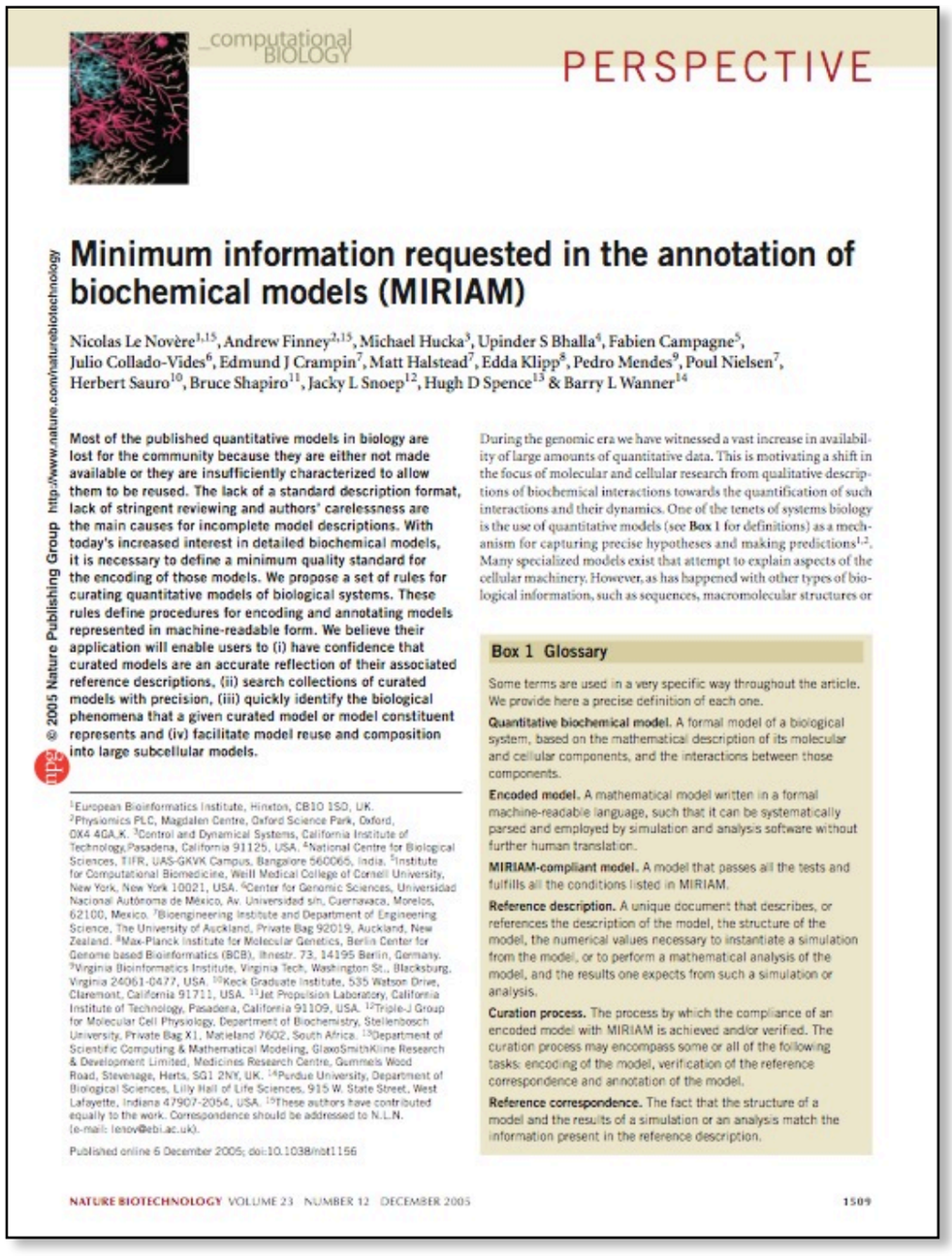
- BioModels Database (kinetic models)
- PSI Consortium (protein interaction)
- Reactome (pathways)
- Pathway Commons (pathways)
- SABIO-RK (reaction kinetic
- Yeast consensus model data
- Human consensus model da
- E-MeP (structural genomics

relationship to PURL
relationship to LSID

Application software

- ARCADIA
- BioUML
- COPASI
- Cpath
- libAnnotationSBML
- libSBML
- PathTest
- Saint
- SBML2BioPAX
- SBML2LaTeX
- SBMLeditor
- semanticSBML
- Snazer
- SBW
- The Virtual Cell

MIRIAM



_computational BIOLOGY

PERSPECTIVE

Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère[1,15], Andrew Finney[2,15], Michael Hucka[3], Upinder S Bhalla[4], Fabien Campagne[5], Julio Collado-Vides[6], Edmund J Crampin[7], Matt Halstead[7], Edda Klipp[8], Pedro Mendes[9], Poul Nielsen[7], Herbert Sauro[10], Bruce Shapiro[11], Jacky L Snoep[12], Hugh D Spence[13] & Barry L Wanner[14]

Most of the published quantitative models in biology are lost for the community because they are either not made available or they are insufficiently characterized to allow them to be reused. The lack of a standard description format, lack of stringent reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's increased interest in detailed biochemical models, it is necessary to define a minimum quality standard for the encoding of those models. We propose a set of rules for curating quantitative models of biological systems. These rules define procedures for encoding and annotating models represented in machine-readable form. We believe their application will enable users to (i) have confidence that curated models are an accurate reflection of their associated reference descriptions, (ii) search collections of curated models with precision, (iii) quickly identify the biological phenomena that a given curated model or model constituent represents and (iv) facilitate model reuse and composition into large subcellular models.

[1]European Bioinformatics Institute, Hinxton, CB10 1SD, UK. [2]Physiomics PLC, Magdalen Centre, Oxford Science Park, Oxford, OX4 4GA,K. [3]Control and Dynamical Systems, California Institute of Technology,Pasadena, California 91125, USA. [4]National Centre for Biological Sciences, TIFR, UAS-GKVK Campus, Bangalore 560065, India. [5]Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, New York 10021, USA. [6]Center for Genomic Sciences, Universidad Nacional Autónoma de México, Av. Universidad s/n, Cuernavaca, Morelos, 62100, Mexico. [7]Bioengineering Institute and Department of Engineering Science, The University of Auckland, Private Bag 92019, Auckland, New Zealand. [8]Max-Planck Institute for Molecular Genetics, Berlin Center for Genome based Bioinformatics (BCB), Ihnestr. 73, 14195 Berlin, Germany. [9]Virginia Bioinformatics Institute, Virginia Tech, Washington St., Blacksburg, Virginia 24061-0477, USA. [10]Keck Graduate Institute, 535 Watson Drive, Claremont, California 91711, USA. [11]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [12]Triple-J Group for Molecular Cell Physiology, Department of Biochemistry, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. [13]Department of Scientific Computing & Mathematical Modeling, GlaxoSmithKline Research & Development Limited, Medicines Research Centre, Gunnels Wood Road, Stevenage, Herts, SG1 2NY, UK. [14]Purdue University, Department of Biological Sciences, Lilly Hall of Life Sciences, 915 W. State Street, West Lafayette, Indiana 47907-2054, USA. [15]These authors have contributed equally to the work. Correspondence should be addressed to N.L.N. (e-mail: lenov@ebi.ac.uk).

Published online 6 December 2005; doi:10.1038/nbt1156

During the genomic era we have witnessed a vast increase in availability of large amounts of quantitative data. This is motivating a shift in the focus of molecular and cellular research from qualitative descriptions of biochemical interactions towards the quantification of such interactions and their dynamics. One of the tenets of systems biology is the use of quantitative models (see **Box 1** for definitions) as a mechanism for capturing precise hypotheses and making predictions[1,2]. Many specialized models exist that attempt to explain aspects of the cellular machinery. However, as has happened with other types of biological information, such as sequences, macromolecular structures or

Box 1 Glossary

Some terms are used in a very specific way throughout the article. We provide here a precise definition of each one.

Quantitative biochemical model. A formal model of a biological system, based on the mathematical description of its molecular and cellular components, and the interactions between those components.

Encoded model. A mathematical model written in a formal machine-readable language, such that it can be systematically parsed and employed by simulation and analysis software without further human translation.

MIRIAM-compliant model. A model that passes all the tests and fulfills all the conditions listed in MIRIAM.

Reference description. A unique document that describes, or references the description of the model, the structure of the model, the numerical values necessary to instantiate a simulation from the model, or to perform a mathematical analysis of the model, and the results one expects from such a simulation or analysis.

Curation process. The process by which the compliance of an encoded model with MIRIAM is achieved and/or verified. The curation process may encompass some or all of the following tasks: encoding of the model, verification of the reference correspondence and annotation of the model.

Reference correspondence. The fact that the structure of a model and the results of a simulation or an analysis match the information present in the reference description.

Le Novère et al., *Nature Biotech.*, 23(12), 2005.

MIRIAM Resoures

BMC Systems Biology

BioMed Central

Database

Open Access

MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology

Camille Laibe and Nicolas Le Novère*

Address: European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

Email: Camille Laibe - laibe@ebi.ac.uk; Nicolas Le Novère* - lenov@ebi.ac.uk

* Corresponding author

Published: 13 December 2007

BMC Systems Biology 2007, 1:58 doi:10.1186/1752-0509-1-58

Received: 25 June 2007

Accepted: 13 December 2007

This article is available from: http://www.biomedcentral.com/1752-0509/1/58

Abstract

Background: The *Minimal Information Requested In the Annotation of biochemical Models* (MIRIAM) is a set of guidelines for the annotation and curation processes of computational models, in order to facilitate their exchange and reuse. An important part of the standard consists in the controlled annotation of model components, based on Uniform Resource Identifiers. In order to enable interoperability of this annotation, the community has to agree on a set of standard URIs, corresponding to recognised data types. *MIRIAM Resources* are being developed to support the use of those URIs.

Results: *MIRIAM Resources* are a set of on-line services created to catalogue data types, their URIs and the corresponding physical URLs (or resources), whether data types are controlled vocabularies or primary data resources. *MIRIAM Resources* are composed of several components: *MIRIAM Database* stores the information, *MIRIAM Web Services* allows to programmatically access the database, *MIRIAM Library* provides an access to the Web Services and *MIRIAM Web Application* is a way to access the data (human browsing) and also to edit or add entries.

Conclusions: The project *MIRIAM Resources* allows an easy access to MIRIAM URIs and the associated information and is therefore crucial to foster a general use of MIRIAM annotations in computational models of biological processes.

Background

Computational Systems Biology relies on developing large quantitative models of biological processes. Because of their size and complexity, those models need to be exchanged and reused, rather than rewritten. Standard formats have been created by the community to encode Systems Biology models, such as SBML [1], CellML [2] or BioPAX [3]. However, the fact that a model is syntactically correct does not ensure its semantic accuracy. Moreover, because of thematic or personal preferences, the terminology used to name model components varies widely. The community had therefore to define a set of guidelines to improve the quality of models aimed to be exchanged. The *Minimal Information Requested In the Annotation of biochemical Models* (MIRIAM) [4] fulfils this need by providing a standard for the annotation and curation of biochemical models.

MIRIAM is a project of the international initiative BioModels.net [5], which aims are multiple: define agreed-upon standards for model curation, define agreed-upon vocabularies for annotating models with connections to biological data resources and provide a free access to published, peer-reviewed, annotated, computational models.

Laibe & Le Novère et al., *BMC Systems Biology*, 13 Dec. 2007.

The people behind MIRIAM Resources

Camille Laibe

Nicolas Le Novère

Nick Juty