Presentation by Mike Hucka at
Third NeuroML Development Workshop,
London, UK, March 2011

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
<table>
<thead>
<tr>
<th>Representation format</th>
<th>Model</th>
<th>Procedures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>S8ML™</td>
<td><img src="S8ML.png" alt="" /></td>
<td><img src="SedML.png" alt="" /></td>
<td>SBRML</td>
</tr>
<tr>
<td>Minimal info requirements</td>
<td><img src="Miriam.png" alt="" /></td>
<td><img src="Miase.png" alt="" /></td>
<td>?</td>
</tr>
<tr>
<td>Semantics—Mathematical</td>
<td><img src="Sbo.png" alt="" /></td>
<td><img src="Kisa.png" alt="" /></td>
<td><img src="Teddy.png" alt="" /></td>
</tr>
<tr>
<td>Other annotations</td>
<td><img src="Miriam.png" alt="" /></td>
<td><img src="Miriam.png" alt="" /></td>
<td><img src="Miriam.png" alt="" /></td>
</tr>
<tr>
<td>Model</td>
<td>Procedures</td>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td><strong>Representation format</strong></td>
<td></td>
<td>SBRML</td>
<td></td>
</tr>
<tr>
<td>Minimal info requirements</td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Semantics—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annotations</td>
<td>Annotations</td>
<td>Annotations</td>
<td></td>
</tr>
</tbody>
</table>
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
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

<table>
<thead>
<tr>
<th>Reactions (20)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>binding MAPKK on Tyr site of MAPK</strong></td>
</tr>
<tr>
<td>$[\text{MAPK}] + [\text{MAPKK}] \leftrightarrow [\text{MAPK-MAPKK}_Y]$</td>
</tr>
<tr>
<td><strong>Math:</strong></td>
</tr>
<tr>
<td>$\text{cell} \times (k_1 \times M \times \text{MAPKK} - (k_1 \times M \times \text{MAPKK}_Y))$</td>
</tr>
<tr>
<td><strong>Annotations:</strong></td>
</tr>
<tr>
<td>set #1</td>
</tr>
<tr>
<td>bqbiol:isVersionOf</td>
</tr>
<tr>
<td>Gene Ontology mitogen-activated protein kinase kinase binding</td>
</tr>
<tr>
<td>Gene Ontology mitogen-activated protein kinase binding</td>
</tr>
<tr>
<td>Reactome REACT_1780</td>
</tr>
<tr>
<td>Reactome REACT_495</td>
</tr>
<tr>
<td><strong>tyr phosphorylation of MAPK</strong></td>
</tr>
<tr>
<td>$[\text{MAPK-MAPKK}_Y] \rightarrow [\text{MAPK-PY}] + [\text{MAPKK}]$</td>
</tr>
<tr>
<td><strong>binding of MAPKK on MAPK-PY</strong></td>
</tr>
<tr>
<td>$[\text{MAPK-PY}] + [\text{MAPKK}] \leftrightarrow [\text{MAPK-PY-MAPKK}]$</td>
</tr>
<tr>
<td><strong>thr phosphorylation of MAPK</strong></td>
</tr>
<tr>
<td>$[\text{MAPK-PY-MAPKK}] \rightarrow [\text{MAPK-PP}] + [\text{MAPKK}]$</td>
</tr>
<tr>
<td><strong>binding of MAPKK on Thr site of MAPK</strong></td>
</tr>
<tr>
<td>$[\text{MAPK}] + [\text{MAPKK}] \leftrightarrow [\text{MAPK-MAPKK}_T]$</td>
</tr>
</tbody>
</table>
E.g.: linking model entities to entities in external db’s

```plaintext
<table>
<thead>
<tr>
<th>Reaction</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>binding MAPKK on Tyr site of MAPK</td>
<td>[MAPK] + [MAPKK] ⇄ [MAPK_MAPKK_Y];</td>
</tr>
<tr>
<td>Math:</td>
<td>cell × (k1 × M × MAPKK - (k1 × M_MAPKK_Y))</td>
</tr>
<tr>
<td>Annotations:</td>
<td>set #1</td>
</tr>
<tr>
<td></td>
<td>bqbiol:isVersionOf Gene Ontology mitogen-activated protein kinase kinase binding</td>
</tr>
<tr>
<td></td>
<td>bqbiol:isHomologTo Reactome REACT_1780</td>
</tr>
<tr>
<td></td>
<td>Reactome REACT_495</td>
</tr>
<tr>
<td>tyr phosphorylation of MAPK</td>
<td>[MAPK_MAPKK_Y] → [MAPK-PY] + [MAPKK];</td>
</tr>
<tr>
<td>binding of MAPKK on MAPK-PY</td>
<td>[MAPK-PY] + [MAPKK] ⇄ [MAPK-PY_MAPKK];</td>
</tr>
<tr>
<td>thr phosphorylation of MAPK</td>
<td>[MAPK-PY_MAPKK] → [MAPK-PP] + [MAPKK];</td>
</tr>
<tr>
<td>binding of MAPKK on Thr site of MAPK</td>
<td>[MAPK] + [MAPKK] ⇄ [MAPK_MAPKK_T];</td>
</tr>
</tbody>
</table>
```
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

```xml
<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>
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>
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
“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
- Identifier of an entity within the resource
- Chosen by the creator of the entry in MIRIAM Resources
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
3. Resolution services for software via standard protocols (SOAP & REST)
Resource dictionary & resource resolution

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

Community-maintained

<table>
<thead>
<tr>
<th>Name</th>
<th>URI</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DMET</td>
<td>urn:miriam:3dmet</td>
<td>3DMET is a database collecting three-dimensional metabolites</td>
</tr>
<tr>
<td>Aclame</td>
<td>urn:miriam:aclame</td>
<td>ACLAME is a database dedicated to the collection of genetic elements (MGEs) from various sources, including genomes, plasmids and pathogens</td>
</tr>
<tr>
<td>Anatomical Therapeutic Chemical</td>
<td>urn:miriam:atc</td>
<td>The Anatomical Therapeutic Chemical (ATC) classification system and the Defined Daily Dose system classify substances into different groups according to their use in therapy and treatment, with the fourth level classifying in groups at five different levels; Different substances are classified in groups (1st level), with pharmacological/therapeutic use, such as drugs for the treatment of specific diseases or conditions. The 3rd and 4th levels are chemical/pharmacological, and the 5th level is the chemical substance.</td>
</tr>
</tbody>
</table>
Resource dictionary & resource resolution

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

Data type: 3DMET

<table>
<thead>
<tr>
<th>Name</th>
<th>URIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DMET</td>
<td>urn:miriam:3dmet</td>
</tr>
</tbody>
</table>

Deprecated: No deprecated URI

Definition: 3DMET is a database collecting three-dimensional structures of natural metabolites.

Identifier Pattern: ^[Bc]{5}$

Physical Locations:

<table>
<thead>
<tr>
<th>Resource MIR:00100095</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access URL</td>
</tr>
<tr>
<td>3DMET database</td>
</tr>
</tbody>
</table>

Institution: Bioinformatics Research Unit, National Institute of Agrobiological Sciences, Japan

References:

URL(s):


Miscellaneous:

Date of creation: 2009-01-23 10:00:42 GMT
Date of last modification: 2009-12-28 18:23:44 GMT
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 kinetics)
- Yeast consensus model database
- Human consensus model database
- E-MeP (structural genomics)

Application software
- ARCADIA
- BioUML
- COPASI
- Cpath
- libAnnotationSBML
- libSBML
- PathTest
- Saint
- SBML2BioPAX
- SBML2LaTeX
- SBMLEditor
- semanticSBML
- Snazer
- SBW
- The Virtual Cell
Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère,1,2 Andrew Finney1,2,3,4,5 Michaelchu4,5,6 Uppin2 J Bhatia,4 FabienCampagne,6,7 JohnColbath-Hicks,1,2 EdwardJ Crampel,2,8 Maciek Hauser,9 Edika Kipypt,4 PedroMendes,2 PaulWaters,1,2 HerbertSauro,9 BrunoShapiro1,3,9 JackieTjosvold,8 HughDStevens2,9,10 SivaraLWanni11

Most of the published quantitative models in biology are used for the community because they are often not made available or they are insufficiently classified to enable them to be reused. The lack of a standard description format, lack of tight integration between model curators and the lacks in web tools for importing model curations, makes it necessary to define a minimum quality standard for the reusability of these models. We propose a set of rules for quantifying the minimum requirements of acceptable models. These rules define procedures for encoding and annotating models represented in concise markup forms. It defines their applications that will be useful to biologists to have confidence that annotated models are an accurate reflection of the associated models descriptions, the search of evidence of correctness of models, and to facilitate model reuse and composition into larger subcellular models.

During the development we have identified a set of concepts available at the interface of molecular and cellular research. This is involving a shift in the focus of molecular and cellular research from qualitative description of biological interactions towards a description of their interactions and their dynamics. One of the intended systems biology is to set of rules that models used for exploration and simulation model predict protagonists and make predictions based on this understanding and making predictions for systems biology.

Many specific models exist that attempt to explain properties of the cellular and molecular processes that have been described in the literature. This is not a comprehensive list of the biological processes, but a list of the types of biological information, such as qualitative or quantitative description of the interactions of cellular components or biological processes.

Box 2. Glossary

Gene terms are used in a very specific way throughout the text. This provides a precise definition of terms.

Quantitative biological model: A formal model of biological process, based on the mathematical description of its molecules and cellular components, and the interactions between these components.

Encoded model: A mathematical model written in a formal modeling language, such that it can be automatically and efficiently translated into any analysis software without further human translation.

MIRIAM compliant model: A model that allows the use of the tools and facilities at the level of the tools.

References:


The people behind MIRIAM Resources

Camille Laibe
Nicolas Le Novère
Nick Juty