

# MIRIAM

## Minimum Information Requested In the Annotation of (biochemical) Models

Camille Laibe

Okinawa, Jan 2008



## MIRIAM Standard:

proposed guidelines for curation and annotation of  
quantitative models

cf. Nicolas Le Novère *et al.* Minimum Information Requested in the  
Annotation of biochemical Models (MIRIAM). *Nature Biotechnology*,  
2005





- The model must be encoded in a public, standardized, machine-readable format (SBML, CellML, GENESIS ...)
- The model must comply with the standard in which it is encoded
- The model must be clearly related to a single reference description. If a model is composed from different parts, there should still be a description of the derived/combined model





- The encoded model structure must reflect the biological processes listed in the reference description
- The model must be instantiated in a simulation: All quantitative attributes have to be defined, including initial conditions
- When instantiated, the model must be able to reproduce all results given in the reference description within an epsilon (algorithms, round-up errors)





- The model has to be named
- A citation of the reference description must be joined (complete citation, unique identifier, unambiguous URL). The citation should permit to identify the authors of the model
- The name and contact of model creators must be joined
- The date and time of creation and last modification should be specified. An history is useful but not required
- The model should be linked to a precise statement about the terms of distribution. MIRIAM does not require “freedom of use” or “no cost”





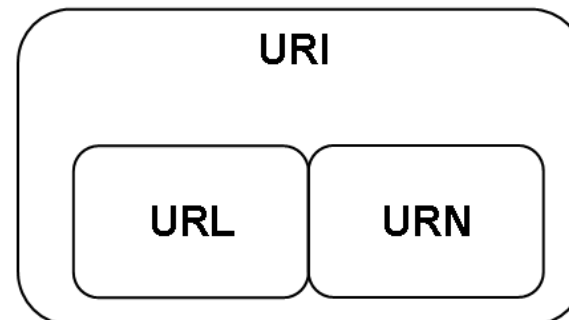
- The annotation must permit to unambiguously relate a piece of knowledge to a model constituent
- The referenced information should be described using a triplet {data-type, identifier, qualifier}
  - The data type should be written as a *Unique Resource Identifier*
  - The identifier is analysed within the framework of the data-type
  - Data-type and Identifier can be combined in a single URI:
    - <http://www.myDataType.org/#myIdentifier>
    - <urn:lsid:myDataType.org:myIdentifier>
  - Qualifiers (optional) should refine the link between the model constituent and the piece of knowledge: "has a", "is version of", "is homolog to" etc.





one URI (Uniform Resource Identifier) two "syntaxes" (W3C standards)

- URN (Uniform Resource Name)
  - urn:ietf:rfc:2141
- URL (Uniform Resource Locator)
  - <http://www.wikipedia.org/>
  - **Not** (in this context) a valid address used in a web browser!





```
[...]
<species metaid="metaid_0000006"
  id="L_EGFR"
  compartment="compartment"
  initialConcentration="0">
  <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_0000006">
        <bqbiol:hasPart>
          <rdf:Bag>
            <rdf:li rdf:resource="http://www.uniprot.org/#P07522" />
            <rdf:li rdf:resource="http://www.uniprot.org/#Q9QX70_RAT" />
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
[...]
```



## Computational Models:

- composed of several components
- perennial identification of elements
- multiple online resources to access these elements





Different resources (associated to a physical location) linked to one data type and identified with a unique and perennial URI.

URI:

- <http://www.uniprot.org/#Q9QX70>

(Epidermal growth factor receptor)

URLs (physical addresses):

- <http://www.ebi.uniprot.org/entry/Q9QX70>

- <http://us.expasy.org/uniprot/Q9QX70>

- <http://www.pir.uniprot.org/cgi-bin/upEntry?id=Q9QX70>





URI	Example of alternative physical locations
<b>Taxonomy</b>	
<a href="http://www.ncbi.nlm.nih.gov/Taxonomy/#9606">http://www.ncbi.nlm.nih.gov/Taxonomy/#9606</a>	<a href="http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9606">http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9606</a> (ref. 32)
<a href="urn:lsid:ncbi.nlm.nih.gov:Taxonomy:9606">urn:lsid:ncbi.nlm.nih.gov:Taxonomy:9606</a>	<a href="http://www.ebi.ac.uk/newt/display?search=9606">http://www.ebi.ac.uk/newt/display?search=9606</a> (ref. 33)
<b>Gene Ontology</b>	
<a href="http://www.geneontology.org/#GO:0045202">http://www.geneontology.org/#GO:0045202</a>	<a href="http://www.ebi.ac.uk/ego/DisplayGoTerm?id=GO:0045202">http://www.ebi.ac.uk/ego/DisplayGoTerm?id=GO:0045202</a>
<a href="urn:lsid:geneontology.org:GO:0045202">urn:lsid:geneontology.org:GO:0045202</a>	<a href="http://www.godatabase.org/cgi-bin/amigo/go.cgi?view=details&amp;query=GO:0045202">http://www.godatabase.org/cgi-bin/amigo/go.cgi?view=details&amp;query=GO:0045202</a>
<b>UniProt</b>	
<a href="http://www.uniprot.org/#P62158">http://www.uniprot.org/#P62158</a>	<a href="http://www.ebi.uniprot.org/entry/P62158">http://www.ebi.uniprot.org/entry/P62158</a> (ref. 28)
<a href="urn:lsid:uniprot.org:P62158">urn:lsid:uniprot.org:P62158</a>	<a href="http://us.expasy.org/uniprot/P62158">http://us.expasy.org/uniprot/P62158</a> (ref. 40)
	<a href="http://www.pir.uniprot.org/cgi-bin/upEntry?id=P62158">http://www.pir.uniprot.org/cgi-bin/upEntry?id=P62158</a> (ref. 41)
<b>EC code</b>	
<a href="http://www.ebi.ac.uk/intenz/EC 1.1.1.1">http://www.ebi.ac.uk/intenz/EC 1.1.1.1</a>	<a href="http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&amp;ec=1.1.1.1">http://www.ebi.ac.uk/intenz/query?cmd=SearchEC&amp;ec=1.1.1.1</a> (ref. 42)
<a href="urn:lsid:ebi.ac.uk:intenz:EC 1.1.1.1">urn:lsid:ebi.ac.uk:intenz:EC 1.1.1.1</a>	<a href="http://www.genome.jp/dbget-bin/www_bget?ec:1.1.1.1">http://www.genome.jp/dbget-bin/www_bget?ec:1.1.1.1</a> (ref. 35)
	<a href="http://www.chem.qmul.ac.uk/iubmb/enzyme/EC1/1/1/1.html">http://www.chem.qmul.ac.uk/iubmb/enzyme/EC1/1/1/1.html</a>
	<a href="http://us.expasy.org/cgi-bin/nicezyme.pl?1.1.1.1">http://us.expasy.org/cgi-bin/nicezyme.pl?1.1.1.1</a> (ref. 43)





- Curator's point of view
  - From the name of a data type and the identifier of an element (a model constituent, for example), be able to generate the URI to put in the annotation
- Software developer's point of view
  - Be able to convert a URI (data type and element identifier) into addresses (physical locations hyperlinks) towards precise pieces of knowledge





- the scientific community has to agree on a set of valid and standard URIs for all the useful existed data types
- creates an online catalogue of these data-types, their resources (physical locations) and other meaningful information
- maintains this resource updated





- stable and perennial identifier
- name
- synonym(s)
- official URI(s)
- deprecated URI(s)
- definition
- pattern of the elements (regular expression)
- resource(s) (cf. next slide)
- documentation(s)
- **example of annotations in various formats**



- stable and perennial identifier
- link towards a precise element of the resource (without the ID of the element)
- link towards the main page of the resource
- general information
- institution managing the resource
- location of the institution (country)



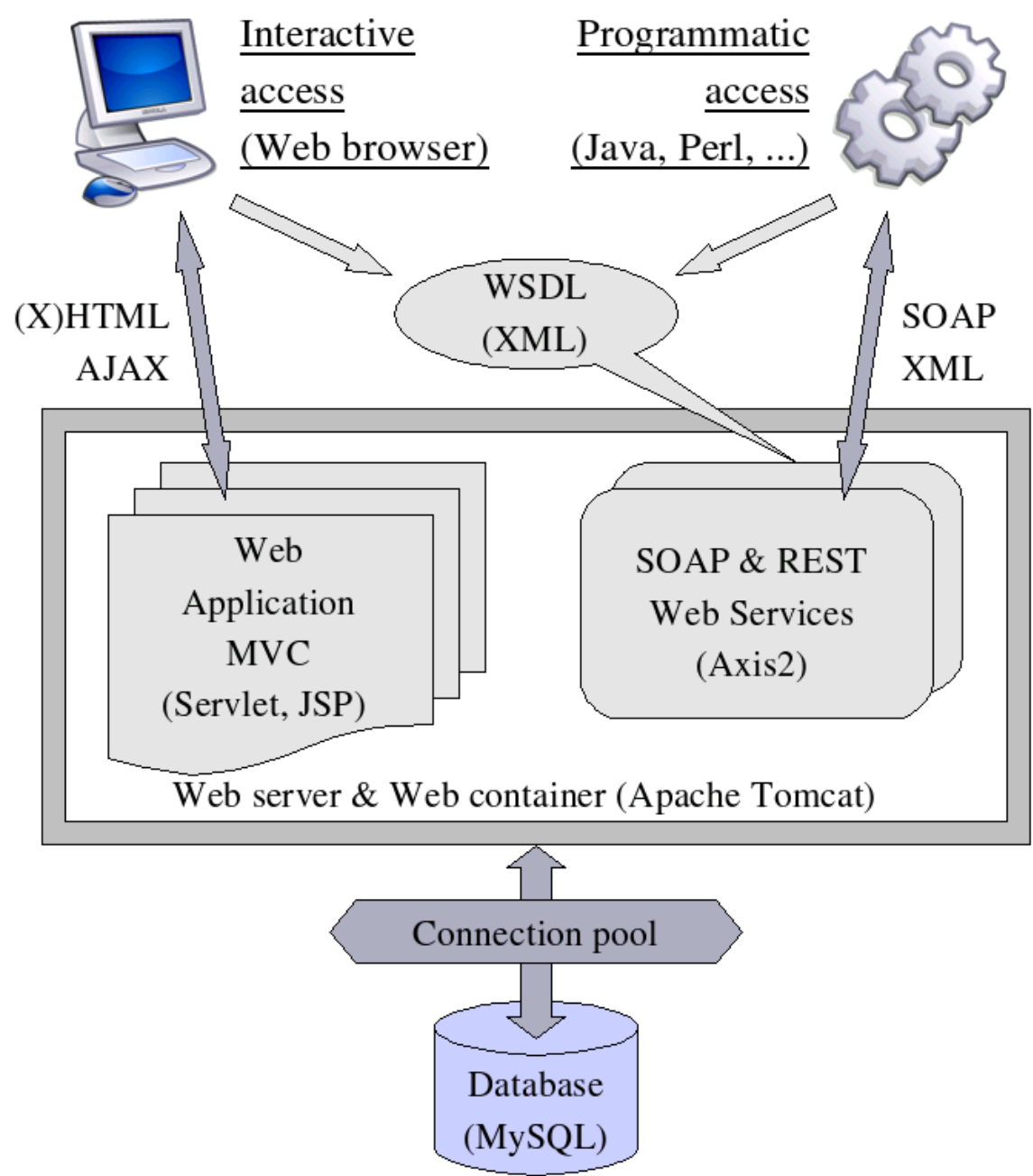
- <http://www.ebi.ac.uk/miriam/>
  - browsing
  - editing
  - export (XML)
  - **Web Services**
  - Java library
- <http://sourceforge.net/projects/miriam/>
  - bug tracker
  - source code





- Provide a query system available through a piece of software
- Based on standard protocols (interoperability)
- Queries:
  - retrieve all the information about a data-type
  - get the physical locations of the knowledge about a precise element based on its URI
  - get the URI from a data-type name and element identifier
  - check if a URI is deprecated
  - give the official URI corresponding to a deprecated one
  - ...





- BioModels Database
- E-MeP
- SABIO-RK
- LGICdb
- ...



Thank you for your attention

