The SBML Level 3 Annotation package: an initial proposal

Allyson Lister, Neil Swainston, Dagmar Waltemath
et al.

COMBINE 2010, Edinburgh, UK
7 October 2010

Overview

- Background

- Limitations of existing annotations

- Proposed solutions

- Semi-proposed solutions / open issues

- SBML does not contains biological semantics
  - What does species name="glc" mean?

- Annotations were introduced in Level 2
  - Also supported in Level 3 Core

- MIRIAM standards specify recommendations of how to use Annotations
  - Based on RDF
  - **NOT SBML specific** – applicable to CellML, BioPAX, any XML?

- Resource Descriptor Framework
  - W3C standard
  - Assign triples to describe resources

- SUBJECT has PREDICATE whose value is OBJECT
  - SUBJECT: model element
  - PREDICATE: property of the element
  - OBJECT: value of the property (e.g. MIRIAM URI)

- SPECIES (complex) has PART whose value is UNIPROT:P00001

Current annotation examples

```
<model metaid="_180340" id="GMO" name="Goldbeter1991_MinMitOscil">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#_180340">
    <dc:creator>
     <rdf:Bag>
      <rdf:li rdf:parseType="Resource">
       <vCard:N rdf:parseType="Resource">
        <vCard:Family>Shapiro</vCard:Family>
        <vCard:Given>Bruce</vCard:Given>
       </vCard:N>
      </rdf:li>
     </rdf:Bag>
    </dc:creator>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</annotation>
```

Current annotation examples

```xml
<species id="Ca_calmodulin" metaid="cacam">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#cacam">
    <bqbiol:hasPart>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
      <rdf:li rdf:resource="urn:miriam:kegg.compound:C00076"/>
     </rdf:Bag>
    </bqbiol:hasPart>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Limitations of current annotations

- NO Statements about attributes

- NO Statements about statements

- NO Relations between statements

- NO Negative statements

- SO… SBML Level 3 Annot package proposal

- Initial discussions (May Workshop)

- Level 3 Proposal with first draft version
http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Annotations

- Three categories
  - Solutions
  - Proposals
  - Open issues

Proposed solutions

- Statements about attributes

- Statements about statements

- Relations between statements

Statements about attributes

```xml
<species id="abc" metaid="meta_abc" compartment="cytoplasm"
   initialConcentration="53.4">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_abc">
    <bqbiol:isDescribedBy>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:pubmed:15387819"/>
     </rdf:Bag>
    </bqbiol:isDescribedBy>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Statements about attributes

- Use XPath to specify attribute that is the SUBJECT of the RDF triple

```
<species metaid="meta_abc" id="abc" compartment="cytoplasm"
   initialConcentration="53.4">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="xref://species[@id='abc']/@compartment">
    ..
```

- May have to restrict the allowed XPath statements

```
<rdf:Description rdf:about="xref://species[7]/@compartment">
```

Statements about statements

- Case study: who added an annotation?
  - Annotation of annotations

- Use RDF Reification
  - Existing standard of "chaining / linking" RDF statements
  - Assign an rdf:ID to a statement, and reference this in subsequent statements through the rdf:about attribute
  - Involves a very minor change to the current annotation standard – add to L3 Core?

Statements about statements

```
<species id="abc" metaid="meta_abc">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_abc">
    <bqbiol:isDescribedBy>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:pubmed:15387819"/>
     </rdf:Bag>
    </bqbiol:isDescribedBy>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Statements about statements

```
<species id="abc" metaid="meta_abc">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_abc">
    <bqbiol:isDescribedBy rdf:ID="statement1">
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:pubmed:15387819"/>
     </rdf:Bag>
    </bqbiol:isDescribedBy>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Statements about statements

```xml
<species id="abc" metaid="meta_abc">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_abc">
    <bqbiol:isDescribedBy rdf:ID="statement1">
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:pubmed:15387819"/>
     </rdf:Bag>
    </bqbiol:isDescribedBy>
   </rdf:Description>
   <rdf:Description rdf:about="#statement1">
    <dc:creator>...</dc:creator>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Statements about statements

- Other application: protein modifications
  - Cannot currently define "Protein X is modified by modification Y in position Z"

- With RDF Reification, we can link more complex statements by defining a blank node, A:
  - PROTEIN_X has MODIFICATION whose value is A
  - A has MODIFIER whose value is Y
  - A has POSITION whose value is Z

Relations between statements

- Alternative annotations...

```
<species id="glc" metaid="meta_glc" name="Glucose">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_glc">
    <bqbiol:is>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3417234"/>
      <rdf:li rdf:resource="urn:miriam:kegg.compound:C00234"/>
     </rdf:Bag>
    </bqbiol:is>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Relations between statements

- Complexes...

```
<species id="Ca_calmodulin" metaid="cacam">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#cacam">
    <bqbiol:hasPart>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
      <rdf:li rdf:resource="urn:miriam:kegg.compound:C00076"/>
     </rdf:Bag>
    </bqbiol:hasPart>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Relations between statements

- Complexes...

```xml
<species id="Ca_calmodulin" metaid="cacam">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#cacam">
    <bqbiol:hasPart>
     <rdf:Bag>
      <rdf:li rdf:resource="urn:miriam:uniprot:P62158"/>
      <rdf:li rdf:resource="urn:miriam:kegg.compound:C00076"/>
     </rdf:Bag>
    </bqbiol:hasPart>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Relation between statements

- RDF Containers and Collections describe groups

- RDF Containers describe an open set:
    - rdf:Bag   unordered set
    - rdf:Seq   ordered set
    - rdf:Alt   alternatives ("or")

- RDF Collections describe a closed set:
    - rdf:List

Relation between statements

- L2 and L3 Core annotations limit groups to rdf:Bag (unordered set)

- Proposal: expand to allow all Containers and Collections

  - rdf:Alt -> Alternative annotations (ChEBI, KEGG)
  - rdf:List -> Components of a complex

Open issues

- Negative statements

- Distinction between current and new L3 Annot package annotations

Negative statements

- If an annotation is absent, are we representing…
    - An unknown un/known?
    - A known unknown?

- Protein X is not annotated to show that it is modified by phosphorylation
- Protein X is annotated to show that it is not modified by phosphorylation

Negative statements

- Problem: RDF statements are positive

- Proposed solution 1
  - Use Boolean statements to specify NOT (and AND, OR, XOR)
  - Not supported in RDF (invalid)

Negative statements

- Proposed solution 2
  - Update predicates to define negation
  - PROTEIN X has NO_MODIFICATION whose value is PHOSPHORYLATION

Negative statements

- Proposed solution 3
  - Used closed set (rdf:List) and predefined resource rdf:nil

```
<species id="x" metaid="meta_x" name="Protein X">
 <annotation>
  <rdf:RDF>
   <rdf:Description rdf:about="#meta_x">
    <bqbiol:modification>
     <rdf:List>
      <rdf:li rdf:resource="rdf:nil"/>
     </rdf:List>
    </bqbiol:modification>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
```

Distinction between current and new L3 Annot package annotations

- Proposal 1: child of annotation

```
<annotation>
  [SBML EXISTING ANNOTATION]
  <annot:annotation xmlns:annot="http://www.sbml.org/sbml/extension/
   annot/version1">
   <rdf:RDF>
    [ANY VALID RDF AS DEFINED IN THE ANNOT PACKAGE SPEC]
   </rdf:RDF>
  </annot:annotation>
</annotation>
```

Distinction between existing and new L3 Annot package annotations

- Proposal 2: sibling of annotation

```
<annotation>
 [SBML EXISTING ANNOTATION]
</annotation>
<annot:annotation xmlns:annot="http://www.sbml.org/sbml/extension/annot/
    version1">
  <rdf:RDF>
  [ANY VALID RDF AS DEFINED IN THE ANNOT PACKAGE SPEC]
  </rdf:RDF>
</annot:annotation>
```

Additional issues

- **New predicates** required

- Previous example considered modifications
  - Currently no predicates regarding modifications supported by MIRIAM standard

- Presumably, other requests…

- Also predicates should be **properties** of the subject
  - **Nouns** not verbs

Conclusion

- Extended Annotation package proposal in development

- Only intended to proposal extensions that will actually be used

- Solutions required for some issues – please contribute if interested
  - http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Annotations

Thanks...

- Wolfram Liebermeister
- Falko Krause
- Marvin Schulz
- Christian Knuepfer
- Ron Henkel
- Stefan Hoops
- Frank Bergmann
- Nick Juty

- Nicolas Le Novère
- Michael Hucka
- Camille Laibe
- Morgan Taschuk
- Goksel Misirli
- Sarah Keating
- Catherine Lloyd

The SBML Level 3 Annotation package: an initial proposal

Allyson Lister, Neil Swainston, Dagmar Waltemath
et al.