

Simulation Experiment Description Markup Language (SED-ML) : Level 1 Version 1 (Draft)

August 13, 2010

Editors

Dagmar Waltemath
Nicolas Le Novère
Frank T. Bergmann

Rostock University, Germany
European Bioinformatics Institute, UK
University of Washington, Seattle, USA

The latest release of the Level 1 Version 1 specification is available at
<http://biomodels.net/sed-ml#sedmlResources>

To discuss any aspect of the current SED-ML specification as well as
language details, please send your messages to the mailing list
sed-ml-discuss@lists.sourceforge.net.

To get subscribed to the mailing list, please write to the same address
sed-ml-discuss@lists.sourceforge.net.

To contact the authors of the SED-ML specification, please write to
sed-ml-editors@lists.sourceforge.net



1	Introduction	4
1.1	Motivation: A sample experiment	4
1.2	Conventions used in this document	6
2	Concepts used in SED-ML	11
2.1	The MathML Subset used in SED-ML	11
2.2	URI Scheme in SED-ML	12
2.3	KiSAO	13
2.4	SED-ML resources	14
3	General attributes and classes	15
3.1	The <code>xmlns</code> attribute	15
3.2	The <code>id</code> attribute	15
3.3	The <code>name</code> attribute	15
3.4	The SEDBase Class	15
3.5	The SED-ML Class	18
3.6	The reference Relations	20
3.7	The Variable Class	22
3.8	The Parameter Class	25
3.9	The ListOf containers	26
4	SED-ML Components	31
4.1	Model	31
4.2	Change	33
4.3	Simulation	37
4.4	Task	40
4.5	DataGenerator	41
4.6	Output	42
4.7	Curve	45
4.8	Surface	47
4.9	DataSet	49
A	SED-ML UML Overview	51
B	Overview of SED-ML	52
B.1	Conventions	52
C	XML Schema	58

D Examples	62
D.1 Le Loup Model (CellML)	62

1 Introduction

As Systems Biology transforms into one of the main fields in life sciences, the number of available computational models is growing at an ever increasing pace. At the same time, their size and complexity are also increasing. The need to build on existing studies by reusing models therefore becomes more imperative. It is now generally accepted that one needs to be able to exchange the biochemical and mathematical structure of models. The efforts to standardise the representation of computational models in various areas of biology, such as the *Systems Biology Markup Language* (SBML, [Hucka et al. \[2003\]](#)), *CellML* [Lloyd et al. \[2004\]](#) or *NeuroML* [Goddard et al. \[2001\]](#), result in an increase of the exchange and re-use of models. However, the description of the structure of models is not sufficient to enable the reproduction of simulation results. One also needs to describe the procedures the models are subjected to, as described by the *Minimum Information About a Simulation Experiment (MIASE)* [\[Waltemath et al.\]](#).

This document presents Level 1 Version 1 of the *Simulation Experiment Description Markup Language* (SED-ML), a format that allows for the encoding of simulation experiments. SED-ML files are encoded in the *eXtensible Markup Language* (XML) [\[Bray et al., 2006\]](#). The SED-ML language is defined by an XML Schema [\[Fallside et al., 2001\]](#).

1.1 Motivation: A sample experiment

To demonstrate how a simulation experiment can be described simply and effectively, we make use of a rather simple, though famous, model that may yet display rich and variable behaviors. The simulation example is taken from [\[Waltemath et al.\]](#).

The *repressilator* is a synthetic oscillating network of transcription regulators in *Escherichia coli* [\[Elowitz and Leibler, 2000\]](#). The network is composed of the three repressor genes Lactose Operon Repressor (lacI), Tetracycline Repressor (tetR) and Repressor CI (cI), which code for proteins binding to the promoter of the other, blocking their transcription. The three inhibitions together in tandem, form a cyclic negative-feedback loop. To describe the interactions of the molecular species involved in the network, the authors built a simple mathematical model of coupled first-order differential equations. All six molecular species included in the network (three mRNAs, three repressor proteins) participated in creation (transcription/translation) and degradation processes. The model was used to determine the influence of the various parameters on the dynamic behavior of the system. In particular, parameter values were sought which induce stable oscillations in the concentrations of the system components. Oscillations in the levels of the three repressor proteins are obtained by numerical integration.

1.1.1 A simple time-course simulation

The first experiment we intend to run on the model is the simulation that will lead to the oscillation shown in Figure 1c of the reference publication [\[Elowitz and Leibler, 2000\]](#). The according simulation experiment can be described as:

1. Import the model identified by the Unified Resource Identifier [\[Berners-Lee et al., 2005\]](#) `urn:miriam:biomodels.db:BIOMD0000000012`.
2. Select a deterministic method.
3. Run a uniform time course simulation for 1000 min with an output interval of 1 min.
4. Plot the amount of lacI, tetR and cI against time in a 2D Plot.

Following those steps and performing the simulation in the simulation tool COPASI [\[Hoops et al., 2006\]](#) let to the result shown in Figure 1 on the following page.

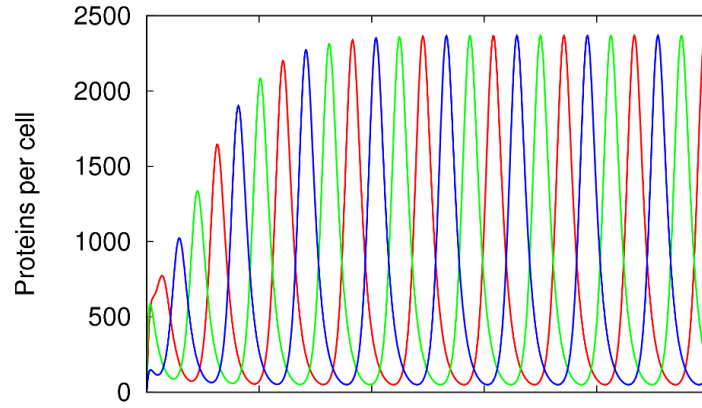


Figure 1: Time-course simulation of the repressilator model, imported from BioModels Database and simulated in COPASI. The number of repressor proteins *lacI*, *tetR* and *cI* is shown. (taken from [Waltemath et al.])

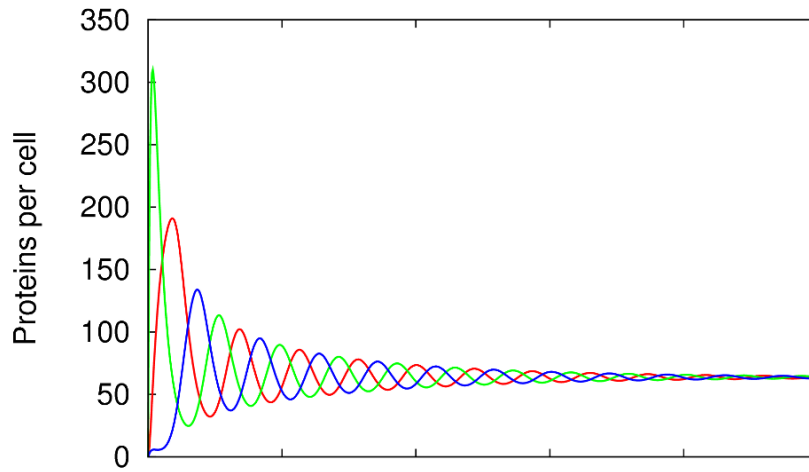


Figure 2: Time-course simulation of the repressilator model, imported from BioModels Database and simulated in COPASI after modification of the initial values of the protein copies per promoter and the leakiness in protein copies per promoter. The number of repressor proteins *lacI*, *tetR* and *cI* is shown. (taken from [Waltemath et al.])

1.1.2 Applying pre-processing

The fine-tuning of the model can be shown by adjusting parameters before simulation. When changing the initial values of the parameters *protein copies per promoter* and *leakiness in protein copies per promoter* the system's behavior switches from sustained oscillation to asymptotic steady-state. The adjustments leading to that behavior may be described as:

1. Import the model as above.
2. Change the value of the parameter `tps_repr` from “0.0005” to “1.3e-05”.
3. Change the value of the parameter `tps_active` from “0.5 “ to “ 0.013“.
4. Select a deterministic method.
5. Run a uniform time course for the duration of 1000 min with an output interval of 1 min.
6. Plot the amount of *lacI*, *tetR* and *cI* against time in a 2D Plot.

Figure 2 shows the result of the simulation.

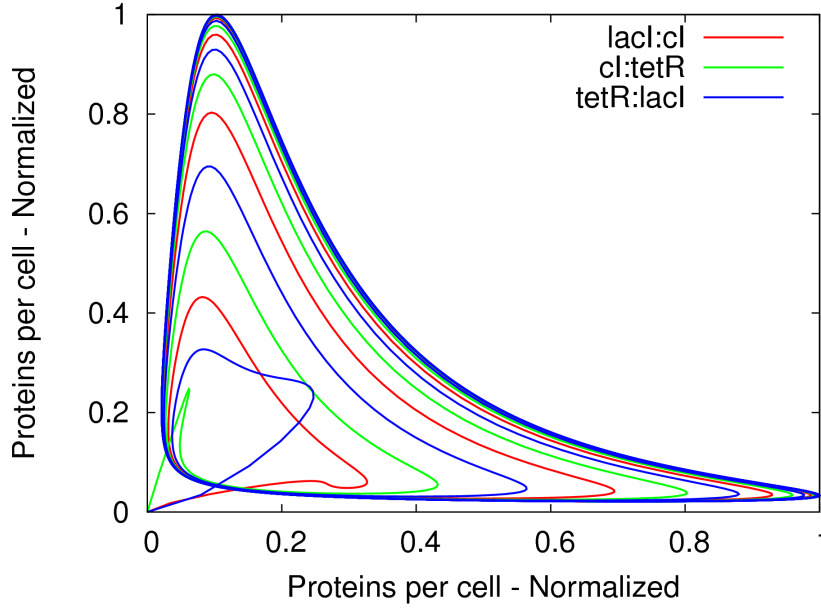


Figure 3: Time-course simulation of the repressilator model, imported from BioModels Database and simulated in COPASI, showing the normalized temporal evolution of repressor proteins *lacI*, *tetR* and *cI* in phase-plane. (taken from [Waltemath et al.])

1.1.3 Applying post-processing

However, the raw numerical output of the simulation steps may be subjected to data post-processing before plotting or reporting. In order to describe the production of a normalized plot of the time-course in the first example (section 1.1.1), depicting the influence of one variable on another (in phase-planes), one could define the following further steps:

(Please note that the description steps 1 - 4 remain as above)

5. Collect $PX(t)$ (*lacI*), $PY(t)$ (*tetR*) and $PZ(t)$ (*cI*).
6. Compute the highest value for each of the repressor proteins, $\max(PX(t))$, $\max(PY(t))$, $\max(PZ(t))$.
7. Normalize the data for each of the repressor proteins by dividing each time point by the maximum value, i.e. $PX(t)/\max(PX(t))$, $PY(t)/\max(PY(t))$, and $PZ(t)/\max(PZ(t))$.
8. Plot the normalized *lacI* protein in function of the normalized *cI*, the normalized *cI* in function of the normalized *tetR* protein, and the normalized *tetR* protein against the normalized *lacI* protein in a 2D plot.

Figure 3 illustrates the result of the simulation after post-processing of the output data.

1.2 Conventions used in this document

SED-ML is specified as an XML Schema [W3C, 2004]. We also provide a UML Class diagram representation of that XML Schema (refer to appendix A). UML class diagrams are a subset of the *Unified Markup Language* notation (UML, [OMG, 2009]). Sample experiment descriptions are given as XML snippets that comply with the XML Schema.

1.2.1 UML Classes

A SED-ML UML class (Figure 4) consists of a class name (**ClassName**) and a number of attributes (**attribute**) each of a specific data type (**type**). The SED-ML UML specification does not make use of UML operations.

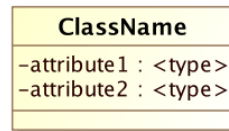


Figure 4: SED-ML UML Class with class names and attributes

SED-ML class names always begin with upper case letters, if they are composed of different words, the camel case style is used, as in e.g. **DataGenerator**.

1.2.2 UML Relationships

1.2.2.1 UML Relation Types

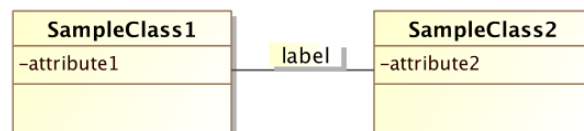


Figure 5: UML Class connectors

Links between classes specify the connection of objects with each other (Figure 5). The different relation types used in the SED-ML specification include aggregation, composite aggregation, and generalisation. The label on the line is called symbol (**label**) and describes the relation of the objects of both classes.

The [association](#) (Figure 6 on the next page) indicates the existence of a connection between the objects of the participating classes. Often associations are directed to show how the label should be read (in which direction). Associations can be uni-directional (one arrowhead), or bidirectional (zero or two arrowheads).

The [aggregation](#) (Figure 7 on the following page, top) indicates that the objects of the participating classes are connected in a way that one class (**Whole**) consists of several parts (**Part**). In an aggregation, the parts may be independent of the whole. For example, a car (**Whole**) has several parts called wheel (**Part**); however, the wheels can exist independently of the car while the car requires the wheels in order to function.

The [composite aggregation](#) (Figure 7 on the next page, bottom) indicates that the objects of the participating classes are connected in a way that one class (**Whole**) consists of several parts (**Part**). In contrast to the aggregation, the subelements (**Part**) are dependent on the parent class (**Whole**). An example is that a university (**Whole**) consists of a number of departments (**Part**) which have a so-called “lifetime responsibility” with the university, e.g. if the university vanishes, so will with it the departments [Bell \[2003\]](#).

The [generalisation](#) (Figure 8 on page 9) allows to extend classes (**BaseClass**) by additional properties. The derived class (**DerivedClass**) inherits all properties of the base class and defines additional ones.



Figure 6: UML Association

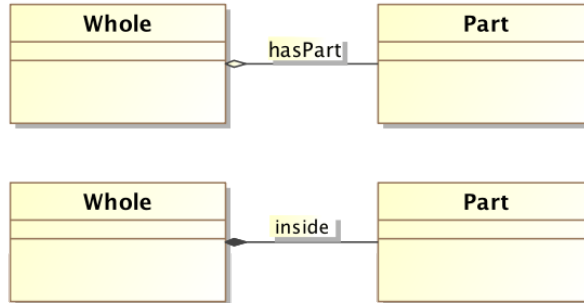


Figure 7: UML Aggregation

1.2.2.2 UML multiplicity

UML multiplicity defines the number of objects in one class that can be related to one object in the other class (also known as [cardinality](#)). Possible types of multiplicity include values (1), ranges (1..4), intervals (1,3,9), or combinations of ranges and intervals. The standard notation for “many” is the asterix (*).

Multiplicity can be defined for both sides of a relationship between classes. The default relationship is “many to many”. The example in [Figure 9 on the next page](#) expresses that a class is given by a professor, and a professor might give one to many classes.

1.2.3 XML Schema language elements

The main building blocks of an XML Schema specification are

- simple and complex types
- element specifications
- attribute specifications

XML Schema [definitions](#) create new types, [declarations](#) define new elements and attributes. The definition of new (simple and complex) types can be based on a number of already existing, predefined types (string, boolean, float). Simple types are restrictions or extensions of predefined types. Complex types describe how attributes can be assigned to elements and how elements can contain further elements. The SED-ML XML Schema only makes use of *complex type definitions*. An example for a complex type definition is given in [listing 1](#):

```

1 <xs:element name="computeChange">
2   <xs:complexType>
3     <xs:sequence>
4       <xs:element ref="listOfVariables" />
5       <xs:element ref="listOfParameters" />
6       <xs:element ref="math:math" />
7     </xs:sequence>
8     <xs:attribute name="target" type="xs:token" />
9   </xs:complexType>
10 </xs:element>
  
```

Listing 1: Complex Type definition of the SED-ML *computeChange* element

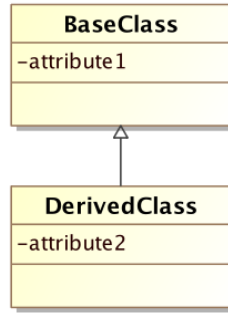


Figure 8: UML Generalisation

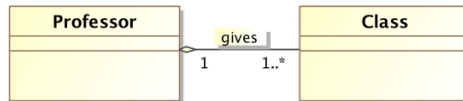


Figure 9: UML Multiplicity in an Aggregation

It shows the declaration of an element called **computeChange** that is used in SED-ML to change mathematical expressions. The element is defined using an *unnamed* complex type which is build of further elements called **listOfVariables**, **listOfParameters**, and **math**. Additionally, the element **computeChange** has an attribute **target** declared. Please note that the definition of the elements inside the complex type are only referred to and will be found elsewhere in the schema. The nesting of elements in the schema can be expressed using the **xs:sequence** (a sequence of elements), **xs:choice** (an alternative of elements to choose from), or **xs:all** (a set of elements that can occur in any order) concepts. The SED-ML XML Schema only uses the *sequence* of elements.

1.2.3.1 Multiplicities

The standard multiplicity for each defined **element** is 1. Explicit multiplicity is to be defined using the **minOccurs** and **maxOccurs** attributes inside the complex type definition, as shown in listing 2.

```

1 <xs:element name=" dataGenerator ">
2   <xs:complexType >
3     <xs:sequence >
4       <xs:element ref=" listOfVariables " minOccurs ="0" />
5       <xs:element ref=" listOfParameters " minOccurs ="0" />
6       <xs:element ref=" math:math " />
7     </xs:sequence >
8     [...]
9   </xs:complexType >
10 </xs:element >

```

Listing 2: Multiplicity for complex types in XML Schema

The **dataGenerator** type is build of a sequence of three elements: The **listOfVariables** element is not necessary for the definition of a valid **dataGenerator** XML structure (it may occur 0 times or once). The same is true for the **listOfParameters** element (it may as well occur 0 times or once). The **math** element, however, uses the implicit standard multiplicity – it must occur exactly 1 time in the **dataGenerator** specification.

1.2.3.2 Type extensions

XML Schema offers mechanics to restrict and extend previously defined complex types. Extensions add element or attribute declarations to existing types, while restrictions restrict the types by adding further characteristics and requirements (facets) to a type. An example for a

type extension is given in listing 3.

```
1 <xs:element name="sedML">
2   <xs:complexType>
3     <xs:complexContent>
4       <xs:extension base="SEDBase">
5         <xs:sequence>
6           <xs:element ref="listOfSimulations" />
7           <xs:element ref="listOfModels" />
8           <xs:element ref="listOfTasks" />
9           <xs:element ref="listOfDataGenerators" />
10          <xs:element ref="listOfOutputs" />
11        </xs:sequence>
12        <xs:attribute name="version" type="xs:decimal" use="required" fixed="0.1" />
13      </xs:extension>
14    </xs:complexContent>
15  </xs:complexType>
16 </xs:element>
```

Listing 3: *Definition of the sedML type through extension of SEDBase in SED-ML*

The **sedML** element is an extension of the previously defined **SEDBase** type. It extends **SEDBase** by a sequence of five additional elements (**listOfSimulations**, **listOfModels**, **listOfTasks**, **listOfDataGenerators**, and **listOfOutputs**) and a new attribute **versions**.

2 Concepts used in SED-ML

2.1 The MathML Subset used in SED-ML

The SED-ML specification allows for the encoding of pre-processing applied to the computational model, as well as for the encoding of post processing applied to the raw simulation data before output. The corresponding mathematical expressions are encoded using MathML 2.0 [Carlisle et al., 2001]. MathML is an international standard for encoding mathematical expressions using XML. It is also used as representation of mathematical expressions in other formats, such as SBML and CellML, two of the languages supported by SED-ML.

2.1.1 MathML operations

In order to make the SED-ML format easier to adopt, at the beginning we restrict the MathML subset to the following operations:

- *token*: `cn`, `ci`, `csymbol`, `sep`
- *general*: `apply`, `piecewise`, `piece`, `otherwise`, `lambda`
- *relational operators*: `eq`, `neq`, `gt`, `lt`, `geq`, `leq`
- *arithmetic operators*: `plus`, `minus`, `times`, `divide`, `power`, `root`, `abs`, `exp`, `ln`, `log`, `floor`, `ceiling`, `factorial`
- *logical operators*: `and`, `or`, `xor`, `not`
- *qualifiers*: `degree`, `bvar`, `logbase`
- *trigonometric operators*: `sin`, `cos`, `tan`, `sec`, `csc`, `cot`, `sinh`, `cosh`, `tanh`, `sech`, `csch`, `coth`, `arcsin`, `arccos`, `arctan`, `arcsec`, `arccsc`, `arccot`, `arcsinh`, `arccosh`, `arctanh`, `arcsech`, `arccsch`, `arccoth`
- *constants*: `true`, `false`, `notanumber`, `pi`, `infinity`, `exponentiale`
- *MathML annotations*: `semantics`, `annotation`, `annotation-xml`

2.1.2 MathML Symbols

All the operations listed above only operate on singular values. However, as one of SED-ML's aim is to provide post processing on the results of simulation experiments, we need to enhance this basic set of operations by some aggregate functions. Therefore a defined set of MathML symbols that represent vector values are supported by SED-ML Level 1 Version 1. To simplify things for SED-ML L1V1 the only symbols to be used are the identifiers of variables defined in the `listOfVariables` of `DataGenerators`. These variables represent the data collected from the simulation experiment with the associated task.

2.1.3 MathML functions

The following aggregate functions are available for use in Level 1 Version 1.

- *min*: Where the minimum of a variable represents the smallest value the simulation experiment yielded. Example:
`<min> <ci> variableId </ci></min>`
- *max*: Where the maximum of a variables represents the largest value the simulation experiment yielded. Example:
`<max> <ci> variableId </ci></max>`
- *sum*: All values of the variable returned by the simulation experiment are added up. Example:

```
<sum> <ci> variableId </ci></sum>
```

- *product*: All values of the variable returned by the simulation experiment are multiplied.
Example:

```
<product> <ci> variableId </ci></product>
```

These represent the only exceptions. At this point SED-ML does not define a complete algebra of vector values. For more information see the description of the [DataGenerator](#) class.

2.2 URI Scheme in SED-ML

URIs are needed at different points in SED-ML Level 1 Version 1: Firstly, they are the preferred mechanism to refer to model encodings. Secondly, they are used to specify the language of the referenced model. Thirdly, they enable addressing implicit model variables. Finally, annotations of SED-ML elements should be provided with a standardised annotation scheme.

The use of a standardised URI Scheme ensures long-time availability of a particular information that can unambiguously be identified.

2.2.1 Model references

The preferred way for referencing a model from a SED-ML file is adopted from the [MIRIAM URI Scheme](#). MIRIAM allows to identify a data resource by a predefined URN. A data entry inside that resource is identified by an ID. That way each single model in a particular model repository can be unambiguously referenced. To become part of MIRIAM resources, a model repository must ensure permanent and consistent model references, that is stable IDs.

One model repository that is part of MIRIAM resources is the [BioModels Database](#) [Li et al., 2010]. Its data resource name in MIRIAM is `urn:miriam:biomodels.db`. To refer to a particular model, a standardised identifier scheme is defined in [MIRIAM Resources](#). The ID entry maps to a particular model in the model repository. That model is never deleted. A sample BioModels Database ID is `BIOMD00000000048`. Together with the data resource name it becomes unambiguously referable by the URN `urn:miriam:biomodels.db:BIOMD00000000048` (in this case referring to the 1999 Kholodenko model on EGFR signaling).

SED-ML recommends to follow the above scheme for model references, if possible. SED-ML does not specify how to resolve the URNs. However, MIRIAM Resources offers web services to do so ¹. For the above example of the `urn:miriam:biomodels.db:BIOMD00000000048` model, the resolved URL may look like:

- `http://biomodels.caltech.edu/BIOMD00000000048`
- `http://www.ebi.ac.uk/biomodels-main/BIOMD00000000048`

depending on the physical location of the resource chosen to resolve the URN.

Further information on the `source` attribute referencing the model location is provided in section [4.1.2](#).

2.2.2 Language references

To specify the language a model is encoded in, a set of pre-defined SED-ML URNs can be used. The structure of SED-ML language URNs is `urn:sedml:language:name.version`. SED-ML allows to specify a model representation format very generally as “XML”, if no standardised representation format has been used to encode the model. On the other hand, one can be as

¹<http://www.ebi.ac.uk/miriam>

specific as defining a model being in a particular version of a language, as “SBML Level 2, Version 2, Revision 1”.

The list of URNs is available from <http://www.biomodels.net/sed-ml/#sedmlLanguage>. Further information on the `language` attribute is provided in section 4.1.1.

2.2.3 Implicit variables

Some variables used in an experiment are not explicitly defined in the model, but may be implicitly contained in it. For example, to plot a variable’s behaviour over time, that variable is defined in an SBML model, while time is not explicitly defined.

To overcome this issue and allow SED-ML to refer to such variables in a common way, the notion of *implicit variables* is used. Those variables are called **symbols** in SED-ML. They are defined following the idea of MIRIAM URNs and using the SED-ML URN scheme. The structure of the URNs is `urn:sedml:symbol:implicit variable`. To refer from a SED-ML file to the definition of *time*, for example, the URN is `urn:sedml:symbol:time`.

The list of predefined symbols is available from the SED-ML site on <http://biomodels.net/sed-ml>. From that source, also a mapping of SED-ML symbols on possibly existing concepts in the single languages supported by SED-ML is provided

2.2.4 Annotations

When annotating SED-ML elements with the `annotation` concept, the [MIRIAM URI Scheme](#) should be used. In addition to providing the data type (e.g. PubMed) and the particular data entry inside that data type (e.g. `10415827`), the relation of the annotation to the annotated element should be described using the standardised [biomodels.net qualifier](#). The list of qualifiers, as well as further information about their usage, is available from <http://www.biomodels.net/qualifiers/>.

2.3 KiSAO

An important aspect of a simulation experiment is the simulation algorithm used to solve the system.

The sole reference of a simulation algorithm through its name in form of a string is error prone and unambiguous. Firstly, typing mistakes or language differences may make the identification of the intended algorithm difficult. Secondly, many algorithms exist with more than one name, having synonyms or various abbreviations that are commonly used.

These problems can be solved by using controlled vocabulary to refer to a particular simulation algorithm. One attempt to provide such a vocabulary is the *Kinetic Simulation Algorithm Ontology* (KiSAO, <http://www.ebi.ac.uk/compneur-srv/kisao/>). KiSAO is a community-driven approach of classifying and structuring simulation approaches by model characteristics and numerical characteristics. Model characteristics include, for instance, the type of variables used for the simulation (such as discrete or continuous variables) and the spatial resolution (spatial or non-spatial descriptions). Numerical characteristics specify whether the system’s behavior can be described as deterministic or stochastic, and whether the algorithms use fixed or adaptive time steps. Related algorithms are grouped together, producing classes of algorithms [Courtot et al., to be submitted].

Although work is still at an early stage, the use of KiSAO is recommended when referring to a simulation algorithm from a SED-ML description. However, the use of KiSAO for the moment is limited. One may look up the algorithm that was used in the simulation experiment (through resolving the KiSAO ID) and then try and use one algorithm that is as similar to the original one

as possible. KiSAO will become more supportive for SED-ML as soon as the ontology contains a wider range of relationships between different algorithms, as well as extended descriptions of the algorithm characteristics.

2.4 SED-ML resources

SED-ML is part of the biomodels.net initiative <http://www.biomodels.net>. Information on SED-ML can be found on www.biomodels.net/sed-ml.

The SED-ML XML Schema, the UML schema and related implementations, libraries, validators and so on can be found on the SED-ML sourceforge project page <http://sed-ml.svn.sourceforge.net/>.

3 General attributes and classes

In this section we introduce attributes and concepts used repeatedly throughout the SED-ML specification.

3.1 The `xmlns` attribute

The `xmlns` attribute contains namespace declarations for elements of external languages used in SED-ML.

First of all, `xmlns` declares the namespace for the SED-ML document. The pre-defined standard namespace is <http://www.biomodels.net/sed-ml>.

In addition, SED-ML makes use of the `MathML` namespace <http://www.w3.org/1998/Math/MathML> to enable the encoding of mathematical expressions in MathML 2.0. SED-ML uses a subset of MathML as described in section 2.1 on page 11.

SED-ML `notes` use the `xmlns` of XHTML <http://www.w3.org/1999/xhtml>. The `Notes` class is described in section 3.4.2 on page 16.

3.2 The `id` attribute

Most objects in SED-ML carry an `id` attribute. The `id` attribute, if existent for an object, is always required and identifies SED-ML constituents unambiguously. It is used to refer to a constituent from other constituents. The `id` data type is `String`. All `ids` have a global scope, i. e. the `id` must be unambiguous throughout a whole SED-ML document. As such it identifies the constituent it is related to. An example for a defined `id` is given in listing 4.

```
1 <model id="m00001" language="urn:sedml:language:sbml" source="urn:miriam:biomodels.db:BIOMD0000000012">
2   [MODEL DEFINITION]
3 </model>
```

Listing 4: SED-ML identifier definition, e. g. for a model

The defined model carries the ID `m00001`. If the model is used somewhere else in the SED-ML document, it is referred to by that ID.

3.3 The `name` attribute

Besides an `id`, a SED-ML constituent may carry an optional `name`. However, names do not have identifying character; several SED-ML constituents may carry the same name. The purpose of the `name` attribute is to keep a human-readable name of the constituent, e. g. for display to the user. In the XML Schema representation, names are of the data type `String`.

Listing 5 extends the model definition in listing 4 by a model name.

```
1 <model id="m00001" name="Circadian oscillator" language="urn:sedml:language:sbml" source="urn:miriam:biomodels.db:BIOMD0000000012">
2   [MODEL DEFINITION]
3 </model>
```

Listing 5: SED-ML name definition, e. g. for a model

3.4 The `SEDBase` Class

`SEDBase` is the base class of SED-ML Level 1 Version 1. All other classes are derived from it. It provides means to attach additional information on all other classes (Figure 10 on the following page). That information can be specified in form of human readable `Notes` or custom `Annotation` classes.

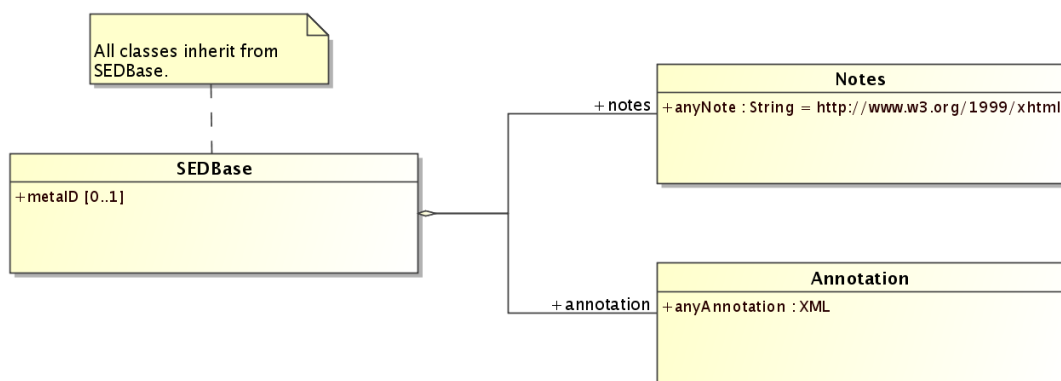


Figure 10: The *SEDBase* class

Table 3.4 shows all attributes and sub-elements for the [SEDBase](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaID ^o	page 16
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 1: Attributes and nested elements for [SEDBase](#). elements^o denotes optional elements.

3.4.1 metaid Attribute

The main purpose of the `metaid` attribute is to attach instances of the [Annotation](#) class to SED-ML elements. The `metaID` attribute is globally unique throughout the whole SED-ML document.

`metaID` is of type XML ID.

3.4.2 The Notes Class

A [note](#) is considered a human-readable description of the element it is assigned to. It serves to display information to the user. Instances of the [Notes](#) class may contain any valid XHTML [[Pemberton et al., 2002](#)], ranging from short comments to whole HTML pages for display in a Web browser. The namespace URL for XHTML content inside the [Notes](#) class is <http://www.w3.org/1999/xhtml>. It may either be declared in the [sedML XML element](#), or directly for use in each `notes` element of the XML file. For further options of how to set the namespace and detailed examples, please refer to ([[Hucka et al., 2010](#)], p. 14).

Table 3.4.2 on the following page shows all attributes and sub-elements for the [notes](#) element as defined by the SED-ML Level 1 Version 1 XML Schema. [Notes](#) has a mandatory attribute

attribute	description
xmlns	page 15
sub-elements	description
<i>none</i>	

Table 2: Attributes and nested elements for [Notes](#). elements^o denotes optional elements.

[xmlns](#) to declare the XHTML namespace. It does not have any further sub-elements nor attributes associated to it.

Listing 6 shows the use of the **notes** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <sedML [...]>
2 <notes http://www.w3.org/1999/xhtml>
3   The enclosed simulation description shows the oscillating behaviour of the Repressilator
   model using deterministic and stochastic simulators.
4 </notes>
5 </sedML>

```

Listing 6: The *notes* element

In this example, the namespace declaration is inside the **notes** element and the note is related to the **sedML** root element of the SED-ML file. A note may, however, occur inside *any* SED-ML XML element, except **note** itself and [annotation](#).

3.4.3 The Annotation Class

An [annotation](#) is considered a computer-processible piece of information. Annotations may contain any valid XML content. No additional namespace declaration is needed for use of the annotation element. However, for the type of XML inside annotation, the according namespaces should be declared, if necessary. For further guidelines on how to use annotations, we would like to encourage the reading of the according section in the SBML specification ([[Hucka et al., 2010](#)], pp. 14-16). The style of annotations in SED-ML is briefly described in section 2.2.4 on page 13.

Table 3.4.3 shows all attributes and sub-elements for the [Annotation](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
<i>none</i>	
sub-elements	description
<i>none</i>	

Table 3: Attributes and nested elements for [Annotation](#). elements^o denotes optional elements.

Listing 7 shows the use of the **annotation** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <sedML>
2   [...]
3   <model id="model1" metaID="#001" language="urn:sedml:language:cellml"
4     source="http://models.cellml.org/workspace/leloup_gonze_goldbeter_1999/@rawfile/
       d6613d7e1051b3eff2bb1d3d419a445bb8c754ad/leloup_gonze_goldbeter_1999_a.cellml" >
5   <annotation>
6     <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
7       xmlns:bqmodel="http://biomodels.net/model-qualifiers/">
8       <rdf:Description rdf:about="#001">

```

```

9      <bqmodel:isDescribedBy>
10      <rdf:Bag>
11      <rdf:li rdf:resource="urn:miriam:pubmed:10415827"/>
12      </rdf:Bag>
13      </bqmodel:isDescribedBy>
14      </rdf:Description>
15      </rdf:RDF>
16      </annotation>
17      </model>
18      [...]
19 </sedML>

```

Listing 7: *The annotation element*

In that example, a SED-ML `model` element is annotated with a reference to the original publication. The `model` contains an `annotation` that uses the biomodels.net `model-qualifier isDescribedBy` to link to the external resource `urn:miriam:pubmed:10415827`. In natural language the annotation content could be interpreted as “The model *is described by* the published article available from *pubmed* under ID *10643740*”. The example annotation follows the proposed [URI Scheme](https://pubmed.gov) suggested by MIRIAM. The MIRIAM URN can be resolved to the PubMed (<http://pubmed.gov>) publication with ID 10415827, namely the article “Alternating oscillations and chaos in a model of two coupled biochemical oscillators driving successive phases of the cell cycle.” published by Romond et al. in 1999.

3.5 The SED-ML Class

Each SED-ML Level 1 Version 1 document has a main class called SED-ML which defines the document’s structure and content (Figure 11). A SED-ML document needs to have the SED-

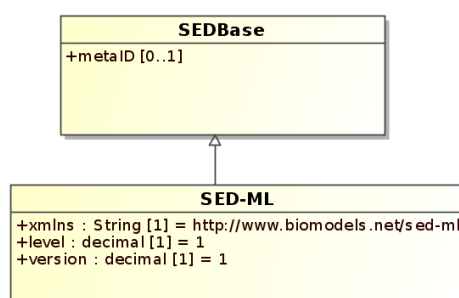


Figure 11: *The SED-ML class*

ML namespace defined through the mandatory `xmlns` attribute. In addition, the SED-ML `level` and `version` attributes are mandatory.

The SED-ML document is build of several parts which are all connected to the SED-ML class through aggregation: the `Model` class (for model specification, see section 4.1), the `Simulation` class (for simulation setup specification, see section 4.3), the `Task` class (for the linkage of models and simulation setups, see section 4.4), the `DataGenerator` class (for the definition of post-processing, see section 4.5), and the `Output` class (for the output specification, see section 4.6). All of them are shown in Figure 12 on the next page and will be explained in more detail in the according sections of this document.

Table 3.5 on the following page shows all attributes and sub-elements for the SED-ML element as defined by the SED-ML Level 1 Version 1 XML Schema.

The basic XML structure of a SED-ML file is shown in listing 8.

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <sedML xmlns:math="http://www.w3.org/1998/Math/MathML"

```

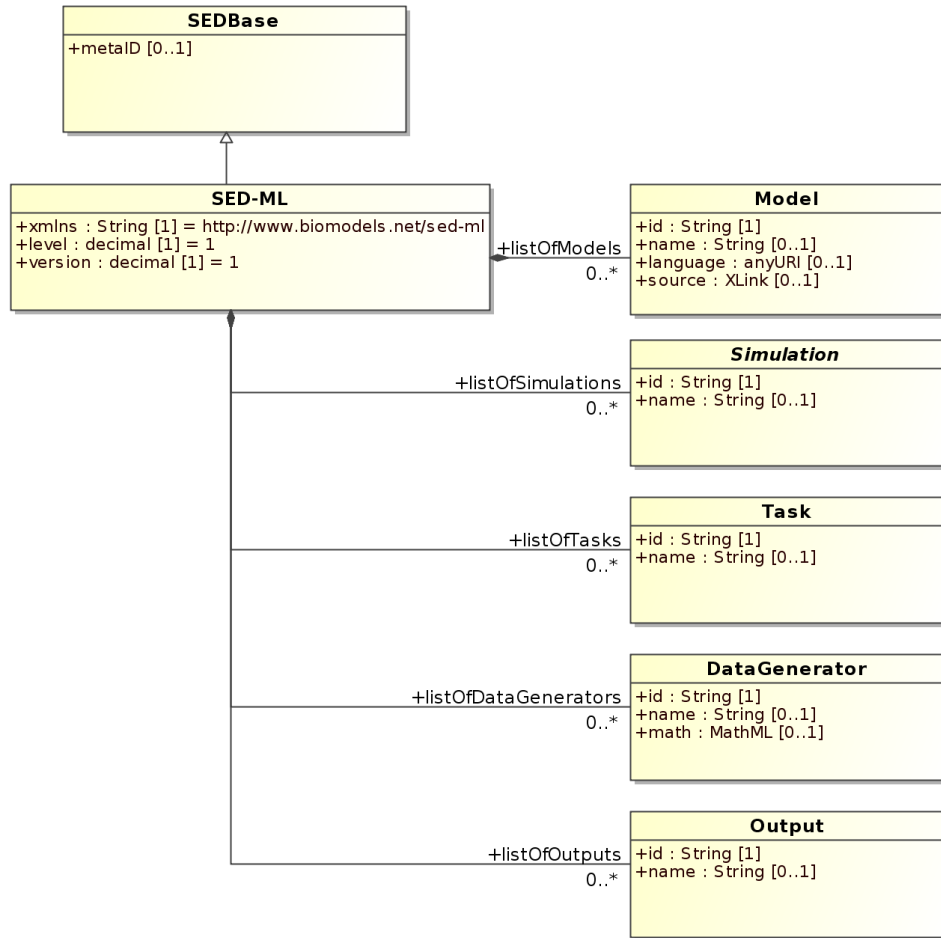


Figure 12: The sub-classes of *SED-ML*

attribute	description
metaID ^o	page 16
xmlns	page 15
level	page 20
version	page 20
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
model ^o	page 31
simulation ^o	page 37
task ^o	page 40
dataGenerator ^o	page 41
output ^o	page 42

Table 4: Attributes and nested elements for *SED-ML*. elements^o denotes optional elements.

```

3      xmlns="http://www.biomodels.net/sed-ml" level="1" version="1">
4    <listOfModels />
5    [MODEL REFERENCES AND APPLIED CHANGES]
6    <listOfSimulations />
7    [SIMULATION SETUPS]
8    <listOfTasks />
9    [MODELS LINKED TO SIMULATIONS]
10   <listOfDataGenerators />
11   [DEFINITION OF POST-PROCESSING]

```

```

12 <listOfOutputs />
13 [DEFINITION OF OUTPUT]
14 </sedML>

```

Listing 8: *The SED-ML root element*

The root element of each SED-ML XML file is the **sedML** element, encoding [version](#) and [level](#) of the file, and setting the necessary namespaces. Nested inside the **sedML** element are the five lists serving as containers for the encoded data ([listOfModels](#) for all models, [listOfSimulations](#) for all simulations, [listOfTasks](#) for all tasks, [listOfDataGenerators](#) for all post-processing definitions, and [listOfOutputs](#) for all output definitions).

3.5.1 The level attribute

The current SED-ML [level](#) is *level 1*. Major revisions containing substantial changes will lead to the definition of forthcoming levels.

The level attribute is **required** and its value is a **fixed** decimal. For SED-ML Level 1 Version 1 the value is set to 1, as shown in the example in listing 8.

3.5.2 The version attribute

The current SED-ML [version](#) is *version 1*. Minor revisions containing corrections and refinements of SED-ML elements will lead to the definition of forthcoming versions.

The version attribute is **required** and its value is a **fixed** decimal. For SED-ML Level 1 Version 1 the value is set to 1, as shown in the example in listing 8.

3.6 The reference Relations

The [reference](#) concept is used to refer to a particular element inside the SED-ML document. It may occur in four different ways in the SED-ML document:

1. as an association between a [Variable](#) and a [Model](#) ([modelReference](#))
2. as an association between a [Variable](#) and a [Task](#) ([taskReference](#))
3. as an association between a [Task](#) and the associated [Model](#) ([modelReference](#)) or
4. as an association between a [Task](#) and the [Simulation](#) ([simulationReference](#))

Depending on the use of the [reference](#) relation in connection with a [Variable](#) object, it may take different roles:

- a. The [reference](#) association might occur between a Variable object and a Model object, if the variable is to define a [Change](#). In that case the **variable** element contains a [modelReference](#) to refer to the particular model that contains the variable used to define the change (see section 3.6.1 on page 21).
- b. If the [reference](#) is used as an association between a Variable object and a Task object inside the [dataGenerator](#) class, then the **variable** element contains a [taskReference](#) to unambiguously refer to an observable in a given task (see section 3.6.2 on page 21).

The definition of a [Task](#) object demands a reference to a particular Model object ([modelReference](#), see 3.6.1 on page 21); furthermore, the Task object must be associated with a particular Simulation object ([simulationReference](#), see 3.6.3 on page 22).

3.6.1 modelReference

The `modelReference` represents a relation between a `Variable` object and a `Model` object, or a relation between a `Task` object and a `Model` object.

If pre-processing needs to be applied to a model before simulation, then the model update can be specified by creating a `Change` object. In the particular case that a change must be calculated with a mathematical function, variables need to be defined. To refer to an existing entity in a defined `Model`, the `modelReference` is used.

The `modelReference` attribute of the `variable` element contains a `modelID`. Listing 9 shows the use of the `modelReference` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```
1 <model id="m0001" [...]>
2   <listOfChanges>
3     <computeChange>
4       <listOfVariables>
5         <variable id="v1" modelReference="cellML" target="/cellml:model/cellml:component[
6           @cmeta:id='MP']/cellml:variable[@name='vsP']/@initial_value" />
7       </listOfVariables>
8       <listOfParameters [...] />
9       <math>
10        [CALCULATION OF CHANGE]
11      </math>
12    </computeChange>
13  </listOfChanges>
14  [...]
```

Listing 9: SED-ML `modelReference` attribute inside a variable definition of a `computeChange` element

In the example, a change is applied on model `m0001`. In the `computeChange` a list of variables is defined. One of those variable is `v1` which is defined in another model, namely `cellML`. To identify the variable in model `cellML` the XPath expression given in the `target` attribute.

The `modelReference` is as well used to define that a `Model` object is used in a particular `Task`. Listing 10 shows how this can be done for a sample SED-ML document.

```
1 <listOfTasks>
2   <task id="t1" name="Baseline" modelReference="model1" simulationReference="simulation1" />
3   <task id="t2" name="Modified" modelReference="model2" simulationReference="simulation1" />
4 </listOfTasks>
```

Listing 10: SED-ML `modelReference` definition inside a `task` element

The example defines two different tasks, the first one applies the simulation settings of `simulation1` on `model1`, the second one applies the same simulation settings on `model2`.

3.6.2 taskReference

`DataGenerator` objects are created to apply post-processing to the simulation results before simulation output.

For certain types of post-processing `Variable` objects need to be created. Those link to a defined `Task` from which the model that contains the variable of interest can be inferred. A `taskReference` association is used to realise that link from a `Variable` object inside a `DataGenerator` to a `Task` object. Listing 11 gives an example.

```
1 <listOfDataGenerators>
2   <dataGenerator id="tim3" name="tim mRNA (difference v1-v2+20)">
3     <listOfVariables>
4       <variable id="v1" taskReference="t1" [...] />
5     </listOfVariables>
6     <math />
7   </dataGenerator>
```

```
8 </listOfDataGenerators>
```

Listing 11: *SED-ML taskReference definition inside a dataGenerator element*

The example shows the definition of a variable **v1** in a **dataGenerator** element. The variable appears in the model that is used in task **t1**. The task definition of **t1** might look as follows:

```
1 <listOfTasks>
2   <task id="t1" name="task definition" modelReference="model1"
3     simulationReference="simulation1" />
4 </listOfTasks>
```

Task **t1** references the model **model1**. Therefore we can conclude that the variable **v1** defined in listing 11 targets an element of the model with ID **model1**. The targeting process itself will be explained in section 3.7.1 on page 23.

3.6.3 simulationReference

The **simulationReference** is used to refer to a particular **Simulation** in a **Task**. Listing 10 on page 21 shows how the reference to a defined simulation for a sample SED-ML document. In the example, both tasks **t1** and **t2** use the simulation settings defined in **simulation1** to run the experiment.

3.7 The Variable Class

SED-ML considers **variables** references to already existing entities, either existing in one of the defined **models** or externally defined **symbols** (Figure 13). If the variable is defined through a

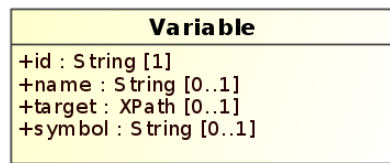


Figure 13: *The Variable class*

reference to a model constituent, such as an SBML species, then the reference is specified using the **target** attribute. If the variable is defined through a reference to an external entity, then the **symbol** attribute is used. It holds a SED-ML **URI**.

Table 3.7 on the following page shows all attributes and sub-elements for the **Variable** element as defined by the SED-ML Level 1 Version 1 XML Schema. A **variable** is always placed

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
target	page 23
symbol	page 24
taskReference	page 21
modelReference	page 21
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 5: Attributes and nested elements for *Variable*. elements^o denotes optional elements.

inside a *listOfVariables*. *Symbol* and *target* must not be used together in a single instance of *Variable*.

A *variable* element must contain a *taskReference* if it occurs inside a *listOfVariables* inside a *dataGenerator* element. A *variable* element must contain a *modelReference* if it occurs inside a *listOfVariables* inside a *computeChange* element. Listing 12 shows the use of the *variable* element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <sedML>
2   <listOfModels>
3     <model [...]>
4       <listOfChanges>
5         <computeChange target="TARGET ELEMENT OR ATTRIBUTE">
6           <listOfVariables>
7             <variable id="v1" name="maximum velocity"
8               target="XPath TO A MODEL ELEMENT OR ATTRIBUTE IN ANY SPECIFIED MODEL" />
9             [FURTHER VARIABLE DEFINITIONS]
10          </listOfVariables>
11        </computeChange>
12      </listOfChanges>
13    </model>
14  </listOfModels>
15  <listOfDataGenerators>
16    <dataGenerator [...]>
17      <listOfVariables>
18        <variable id="v2" name="time" symbol="urn:sedml:symbol:time" />
19        [FURTHER VARIABLE DEFINITIONS]
20      </listOfVariables>
21    </dataGenerator>
22  </listOfDataGenerators>
23  [...]
24 </sedML>

```

Listing 12: SED-ML variable definitions inside the *computeChange* element and inside the *dataGenerator* element

Listing 12 defines a variable *v1* (l. 7) to compute a change on a model constituent (referenced by the *target* attribute in line 5). The value of *v1* corresponds with the value of the targeted model constituent references by the *target* attribute in line 8. The second variable, *v2* (l. 21), is used inside a *dataGenerator*. As the variable is *time*, the *symbol* attribute is used to refer to the SED-ML URI for time (l. 21).

3.7.1 The target attribute

An instance of *Variable* refers to a model constituent inside a particular *model* through an *XPath* expression stored in the required *target* attribute. *XPath* allows to unambiguously identify an

element or attribute in an XML file.

Listing 13 shows the use of the **target** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```
1 <model id="m0001" language="urn:sedml:language:sbml"
2   source="urn:miriam:biomodels.db:BIOMD0000000012">
3   <listOfChanges>
4     <computeChange [...]>
5       <listOfVariables>
6         <variable id="v1" name="Tet protein" target="/sbml/listOfSpecies/species[@id="PY"]" />
7       </listOfVariables>
8     </computeChange>
9   </listOfChanges>
10 </model>
```

Listing 13: SED-ML target definition

Please note that the identifier and names inside the SED-ML document do not have to comply with the identifiers and names that the model and its constituents carry in the model definition. In the above example 12, the variable with ID **v1** is defined. It is described as the **TetR protein**. The reference points to a species in the referenced SBML model. The particular species can be identified through its ID in the SBML model, namely **PY**. However, SED-ML does not forbid to use identical identifiers and names as in the referenced models neither. The following is the same valid SED-ML example for the specification of a variable as the above in listing 12, but with different naming:

```
1 <model id="m0001" language="urn:sedml:language:sbml"
2   source="urn:miriam:biomodels.db:BIOMD0000000012">
3   <listOfChanges>
4     <computeChange [...] />
5     <listOfVariables>
6       <variable id="PY" name="TetR protein" target="/sbml/listOfSpecies/species[@id="PY"]" />
7     </listOfVariables>
8   </computeChange>
9 </listOfChanges>
10 </model>
```

Listing 14: SED-ML variable definition using the original model identifier and name in SED-ML

The XPath expression used in the **target** attribute unambiguously leads to the particular place in the XML SBML model – the species is to be found in the *sbml* element, and there inside the *listOfSpecies*:

```
1 <sbml [...]>
2   <listOfSpecies>
3     <species metaid="PY" id="PY" name="TetR protein" [...]>
4       [...]
5     </species>
6   </listOfSpecies>
7   [...]
8 </sbml>
```

Listing 15: Species definition in the referenced model (extracted from [urn:miriam:biomodels.db:BIOMD0000000012](http://miriam.org/BIOMD0000000012))

3.7.2 The symbol attribute

Symbols are predefined, implicit variables that can be called in a SED-ML file by referring to the defined URNs representing that variable's concept. The notion of implicit variables is explained in section 2.2.3 on page 13.

Listing 16 shows the use of the **symbol** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```
1 <model id="m0001" language="urn:sedml:language:sbml"
2   source="urn:miriam:biomodels.db:BIOMD0000000012">
```



```

3 <listOfChanges>
4   <computeChange [...]>
5     <listOfVariables>
6       <variable id="t1" name="time" symbol="urn:sedml:symbol:time" />
7     </listOfVariables>
8   </computeChange>
9 </listOfChanges>
10 </model>

```

Listing 16: *SED-ML symbol definition*

3.8 The Parameter Class

The SED-ML [Parameter](#) class creates instances with a constant value (Figure 14). SED-ML

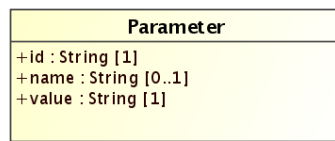


Figure 14: *The Parameter class*

uses parameters in two ways: Firstly, parameters may be defined in the [ComputeChange](#) class for describing the mathematical computation of a change of a model's observable. Secondly, parameters may be part of a [DataGenerator](#) specification. In both cases the parameter definitions are local to the particular class defining them.

Table 3.8 shows all attributes and sub-elements for the [parameter](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaID ^o	page 16
id	page 15
name ^o	page 15
value	page 26
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 6: *Attributes and nested elements for [parameter](#). elements^o denotes optional elements.*

A parameter can unambiguously be identified through it's given [id](#). It may additionally carry an optional [name](#). Each parameter has one associated [value](#).

Listing 17 shows the use of the [parameter](#) element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfParameters>
2   <parameter id="p1" name="KM" value="40" />
3 </listOfParameters>

```

Listing 17: *The definition of a parameter in SED-ML*

The listing shows the definition of a parameter [p1](#) with the [value](#)="40" assigned.

3.8.1 The value attribute

Each [parameter](#) defines a particular thing with a exactly one fixed [value](#). The `value` attribute of XML data type `String` is required for each `parameter` element.

3.9 The ListOf containers

SED-ML `listOf*` elements serve as containers for a collection of objects of the same type. For example, the `listOfModels` contains all [Model](#) objects of a SED-ML document. Lists do not carry any further semantics nor do they add additional attributes to the language. They might, however, be annotated with [Notes](#) and [Annotations](#) as they are derived from [SBase](#). All `listOf*` elements are optional in a SED-ML document.

3.9.1 listOfVariables: The variable definition container

SED-ML uses the [variable](#) concept to refer to existing entities inside a model. The container for all variables is `listOfVariable` (Figure 15). It includes all variables that need to be defined to either describe a change in the model by means of mathematical equations ([ComputeChange](#)) or to set up a [dataGeneratorClass](#).

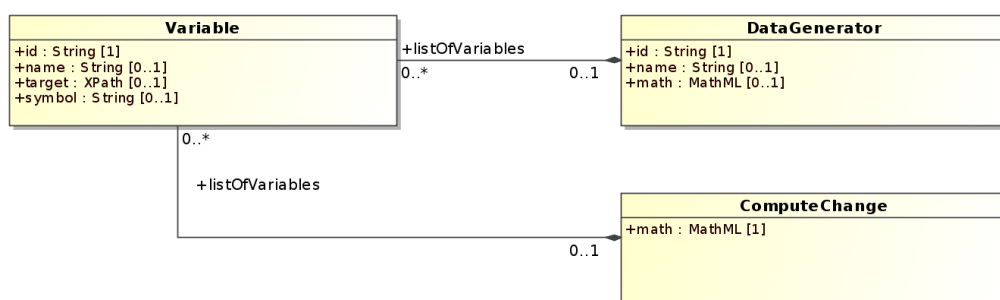


Figure 15: The SED-ML `listOfVariables` container

Listing 18 shows the use of the `listOfVariables` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```
1 <listOfVariables>
2   <variable id="v1" name="maximum velocity" target="/cellml:model/cellml:component[@cmeta:id
   = 'MP']/cellml:variable[@name='vsP']/@initial_value" />
3   <variable id="v2" symbol="urn:sedml:symbol:time" />
4 </listOfVariables>
```

Listing 18: SED-ML `listOfVariables` element

The `listOfVariables` is optional and may contain zero to many models.

3.9.2 listOfParameters: The parameter definition container

All parameters needed throughout the simulation experiment, either to apply a [Change](#) on a model prior to simulation or to set up a [DataGenerator](#), are defined inside a `listOfParameters` (Figure 16 on the following page).

Listing 19 shows the use of the `listOfParameters` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The element is optional and may contain zero to many parameters.

```
1 <listOfParameters>
2   <parameter id="p1" value="1" />
3   <parameter id="p2" name="Kadp_2" value="0.23" />
```

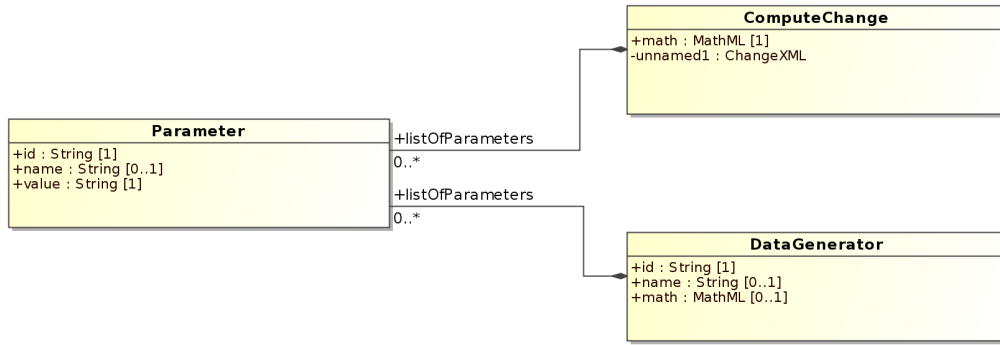


Figure 16: The SED-ML *listOfParameters* container

```
4 </listOfParameters>
```

Listing 19: SED-ML *listOfParameters* element

3.9.3 *listOfModels*: The model description container

In order to specify a simulation experiment, the participating models have to be defined. SED-ML uses the [listOfModels](#) container for all necessary models (Figure 17).

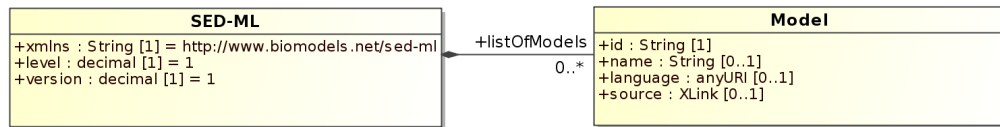


Figure 17: The SED-ML *listOfModels* container

Listing 20 shows the use of the `listOfModels` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The `listOfModels` is optional and may contain zero to many models.

```
1 <listOfModels>
2   <model id="m0001" language="urn:sedml:language:sbml"
3     source="urn:miriam:biomodels.db:BIOMD0000000012" />
4   <model id="m0002" language="urn:sedml:language:cellml"
5     source="http://www.cellml.org/models/leloup_gonze_goldbeter_1999_version02" />
6 </listOfModels>
```

Listing 20: SED-ML *listOfModels* element

3.9.4 *listOfChanges*: The change definition container

The [listOfChanges](#) contains the defined changes to be applied to a particular [model](#) (Figure 18). It always occurs as an optional subelement of the `model` element.

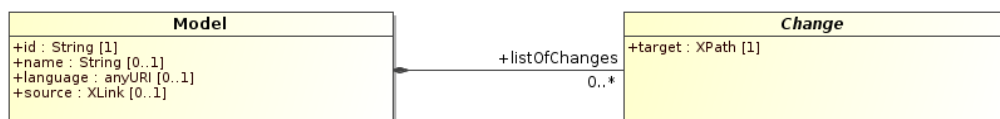


Figure 18: The SED-ML *listOfChanges* container

Listing 21 shows the use of the `listOfChanges` element in a SED-ML file as defined by the

SED-ML Level 1 Version 1 XML Schema. The `listOfChanges` is nested inside the `model` element.

```

1 <model id="m0001" [...]>
2   <listOfChanges>
3     [CHANGE DEFINITION]
4   </listOfChanges>
5 </model>

```

Listing 21: The SED-ML `listOfChanges` element, defining a change on a model

In the example, a change is defined on the model `m0001`. To encode the change, a `listOfChanges` is created that contains all changes, each stored in a single `change` element, as explained in the definition of the `Change` class.

3.9.5 `listOfSimulations`: The simulation description container

The `listOfSimulation` is the container for `simulation` descriptions (Figure 19).

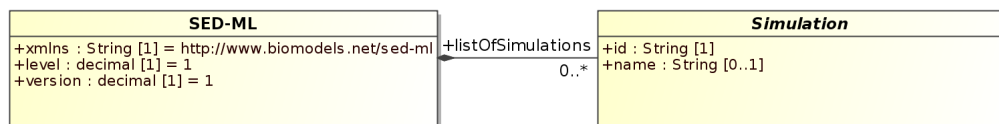


Figure 19: The `listOfSimulations` container

Listing 22 shows the use of the `listOfSimulation` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfSimulations>
2   <simulation id="s1" [...]>
3     [UNIFORM TIMECOURSE DEFINITION]
4   </simulation>
5   <simulation id="s2" [...]>
6     [UNIFORM TIMECOURSE DEFINITION]
7   </simulation>
8 </listOfSimulations>

```

Listing 22: The SED-ML `listOfSimulations` element, containing two simulation setups

Listing 22 shows the definition of two simulation setups, each of them encoded in a single `simulation` element. For all SED-ML Level 1 Version 1 documents, the encoded simulation definitions are instances of the `Uniform Timecourse` class.

3.9.6 `listOfTasks`: The task specification container

The `listOfTasks` container contains the defined tasks for the simulation experiment (Figure 20).

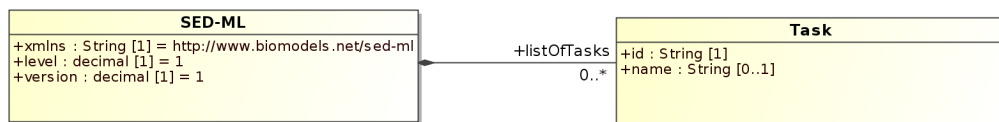


Figure 20: The SED-ML `listOfTasks` container

Listing ?? shows the use of the `listOfTasks` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfTasks>
2   <task id="t1" name="simulating v1" modelReference="m1" simulationReference="s1">
3     [FURTHER TASK DEFINITIONS]
4 </listOfTasks>

```

Listing 23: The SED-ML `listOfTasks` element, defining one task

The example shows the definition of one task `t1` in the `listOfTasks` container.

3.9.7 `listOfDataGenerators`: The post-processing container

In SED-ML, all variables/parameters/values that shall be used in the `Output` class need to be defined as a `dataGenerator` beforehand. The container for those data generators is the `listOfDataGenerators` (Figure 21).



Figure 21: The SED-ML `listOfDataGenerators` container

Listing 36 shows the use of the `listOfDataGenerators` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfDataGenerators>
2   <dataGenerator id="d1" name="time">
3     [DATA GENERATOR DEFINITION FOLLOWING]
4   </dataGenerator>
5   <dataGenerator id="LaCI" name="LaCI repressor">
6     [DATA GENERATOR DEFINITION FOLLOWING]
7   </dataGenerator>
8 </listOfDataGenerators>

```

Listing 24: The `listOfDataGenerators` element, defining two data generators `time` and `LaCI repressor`

3.9.8 `listOfOutputs`: The output specification container

The `listOfOutputs` container holds the output specifications for a simulation experiment.

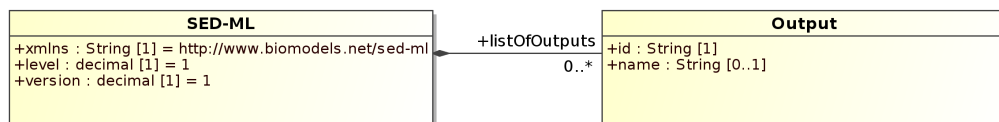


Figure 22: The SED-ML `listOfOutputs` container

The output can be defined as either a `report`, a `plot2D` or as a `3D plot`.

Listing 25 shows the use of the `listOfOutputs` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfOutputs>
2   <report id="report1">
3     [REPORT DEFINITION FOLLOWING]
4   </report>
5   <plot2D id="plot1">
6     [2D PLOT DEFINITION FOLLOWING]

```

```
7 </plot2D>
8 </listOfOutputs>
```

Listing 25: *The `listOfOutput` element*

The example shows the definition of two different outputs, one being a data table (report), the other one a 2D plot.

4 SED-ML Components

In this section we describe the major components of SED-ML. In the following we use the UML notation presented in section 1.2.1. In addition, we provide a detailed BNMP diagram with explanation of the SED-ML workflow in Appendix B and an XML Schema in C.

4.1 Model

The [Model](#) defines the models to be used in the simulation experiment (Figure 23).

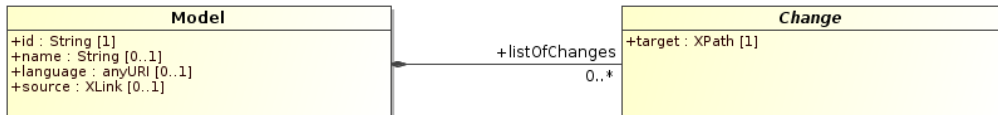


Figure 23: The SED-ML Model class

Each instance of the Model class has an unambiguous and mandatory [id](#). An additional, optional [name](#) may be given to the model.

The [language](#) may be specified, defining the format the model is encoded in, if such a format exists. Example formats are SBML, CellML, or myFancyLanguageML.

The [Model](#) class refers to the particular model of interest through the [source](#) attribute. The restrictions on the model reference are

- The model must be encoded in an XML format.
- To refer to the model encoding language, a reference to a valid definition of that XML format must be given ([language](#) attribute).
- To refer to a particular model in an external resource, an unambiguous reference must be given ([source](#) attribute).

Table 4.1 shows all attributes and sub-elements for the [model](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
language ^o	page 32
source	page 33
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
change ^o	page 33

Table 7: Attributes and nested elements for [model](#). elements^o denotes optional elements.

A model might need to undergo pre-processings before simulation. Those pre-processings are specified in the SED-ML [Change](#) class.

Listing 26 shows the use of the `model` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfModels>
2   <model id="m0001" language="urn:sedml:language:sbml"
3     source="urn:miriam:biomodels.db:BIOMD0000000012">
4     <listOfChanges>
5       <change>
6         [MODEL PRE-PROCESSING]
7       </change>
8     </listOfChanges>
9   </model>
10  <model id="m0002" language="urn:sedml:language:sbml" source="m0001">
11    [MODEL PRE-PROCESSING]
12  </model>
13  <model id="m0003" language="urn:sedml:language:cellml" source="http://www.cellml.org/
14    models/leloup_gonze_goldbeter_1999_version02">
15    [MODEL PRE-PROCESSING]
16  </model>
17 </listOfModels>

```

Listing 26: *SED-ML model element*

The above `listOfModels` contains three models: The first model `m0001` is the Repressilator model taken from BioModels Database. The original model is available from [urn:miriam:biomodels.db:BIOMD0000000012](http://miriam.biomodels.db:BIOMD0000000012). For the SED-ML simulation, the model might undergo pre-processings, described in the `change` element (ll. 5-7). Based on the description of the first model `m0001`, the second model is build. It refers to the model `m0001` in the `source` attribute, that is the modified version of the Repressilator model. `m0002` might then have even further changes applied to it on top of the changes defined in the pre-processing of `m0001`. The third model in the code example above (ll. 13-15) is a different model in CellML representation. `m0003` is the model available from the given URL in the `source` attribute. Again, it might have additional pre-processing applied to it before used in the simulation.

4.1.1 The language attribute

The evaluation of a SED-ML file will decide whether or not it can be used for a particular simulation environment. One crucial criterium is the particular model representation language used to encode the model. A simulation software usually only supports a small subset of the representation formats available to model biological systems computationally.

To help a software decide whether or not it supports a SED-ML description file, the information on the model encoding for each referenced model can be provided through the `language` attribute. As the description of a language name and version through an unrestricted `String` is error-prone. A prerequisite for a language to be fully supported by SED-ML is that a formalised language definition, e.g. an XML Schema, is provided online. SED-ML also defines a set of standard URIs to refer to particular language definitions (see again section 2.2.2 on page 12). The list of URNs for languages so far associated with SED-ML is available from the SED-ML web site on <http://biomodels.net/sed-ml>. To specify language and version, following the idea of MIRIAM URNs, the SED-ML URN scheme `urn:sedml:language:language name` is used. A model's language being "SBML Level 2 Version 2" can be referred to, for example, through the URN `urn:sedml:language:sbml.level-2.version-2`.

The `language` attribute is optional in the XML representation of a SED-ML file. If it is not explicitly defined in the SED-ML file, the default value for the `language` attribute is `urn:sedml:language:xml`, referring to any XML based model representation.

However, the use of the `language` attribute is strongly encouraged. Not only does it help a user decide whether or not he is able to run the simulation, that is to parse the model referenced in the SED-ML file. The language attribute is also needed to decide how to handle a particular implicit variable in the `Variable` class. The interpretation of implicit variables depends on the

language of the representation format. The concept of implicit variables has been introduced in section 2.2.3 on page 13.

4.1.2 The source attribute

To make the model available for the execution of a SED-ML file, the model **source** must be specified through an XLink. The XLink should preferably point to a public, consistent URI that contains the model description file and follows the proposed [URI Scheme](#). References to curated, open model bases are recommended, such as the BioModels Database. However, any resource registered with MIRIAM resources² can easily be referenced. Even without a MIRIAM URN, SED-ML can be used (see again section 2.2.1 on page 12).

An example for the definition of a model, and using the [URI scheme](#) is given in listing 27.

```
1 <model id="m1" name="repressilator" language="urn:sedml:language:sbml"
2   source="urn:miriam:biomodels.db:BIOMD0000000012">
3   <listOfChanges>
4     [DEFINE MODEL PRE-PROCESSING HERE]
5   </listOfChanges>
6 </model>
```

Listing 27: The SED-ML source element, using the URI scheme

The example defines one model `m1`. `urn:miriam:biomodels.db:BIOMD0000000012` defines the source of the model code. The MIRIAM URN can be resolved into the SBML model stored in BioModels Database under ID `BIOMD0000000012`.

An example for the definition of a model and using a URL is given in listing 28.

```
1 <model id="m1" name="repressilator" language="urn:sedml:language:cellml"
2   source="http://models.cellml.org/exposure/bba4e39f2c7ba8af51fd045463e7bdd3/aguda_b_1999.
3     cellml">
4   <listOfChanges />
5 </model>
```

Listing 28: The SED-ML source element, using a URL

In the example one model is defined. The **language** of the model is **CellML**. As the CellML model repository currently does not provide a MIRIAM URI for model reference, the URL pointing to the model code is used to refer to the model. The URL is given in the **source** element.

4.2 The Change Class

SED-ML not only allows to use the sole model for simulation, but on the contrary enables the description of [changes](#) to be made on the model before simulation (Figure 24 on the following page). Changes can be of three different types:

1. Changes on attributes of the model's XML representation ([ChangeAttribute](#))
2. Changes on any XML snippet of the model's XML representation ([ChangeXML](#))
3. Changes based on mathematical calculations ([ComputeChange](#))

The [Change](#) class is abstract and serves as the container for the different types of changes. Therefore, a SED-ML document will only contain the derived classes, i.e. [ChangeAttribute](#), [ChangeXML](#), or [ComputeChange](#).

Table 4.2 on the next page shows all attributes and sub-elements for the [change](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

²<http://www.ebi.ac.uk/miriam/main/>

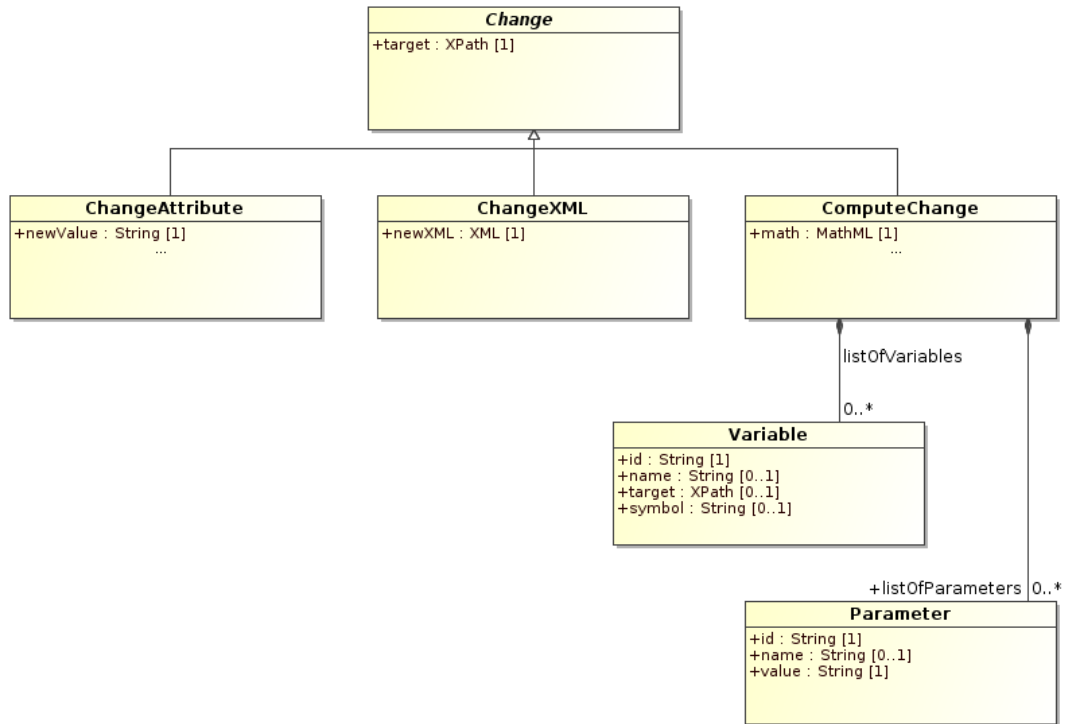


Figure 24: The SED-ML Change class

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
target	page 23
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
changeXML ^o	page 34
changeAttribute ^o	page 35
computeChange ^o	page 36

Table 8: Attributes and nested elements for *change*. elements^o denotes optional elements.

Each Change has a [target](#) attribute that holds a valid XPath expression pointing to the XML element or XML attribute that is to undergo the defined changes.

4.2.1 The ChangeXML Class

The [ChangeXML](#) class defines changes of any XML element in the model that can be addressed by a valid XPath expression. The XPath is specified in the required [target](#) attribute (see again section 3.7.1 on page [page 23](#)).

Table 4.2.1 on the following page shows all attributes and sub-elements for the [changeXml](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

4.2.2 The newXML attribute

The new piece of XML code that is to substitute the XML element addressed by the XPath is provided in the required **newXML** attribute.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
target	page 23
newXML	page 34
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 9: Attributes and nested elements for *changeXML*. elements^o denotes optional elements. , *newXML* is an attribute, probably should be made an element!?! (see problems in [listing 29](#))

An example for a change that adds an additional parameter to a model is given in [listing 29](#).

```

1 <model [...]>
2   <listOfChanges>
3     <changeXML target="/sbml:sbml/sbml:model/sbml:listOfParameters
4       /sbml:parameter[@id='V_mT']"
5       newXML="<parameter metaid="metaid_0000010" id="V_mT" value="0.7">
6         <parameter metaid="metaid_0000050" id="V_mT_2" value="4.6">" />
7     </listOfChanges>
8 </model>

```

Listing 29: The *changeXML* element with its *newXML* attribute

The code of the model is changed in the way that its parameter with ID *V_mT* is substituted by another two parameters *V_mT* and *V_mT_2*.

4.2.3 The ChangeAttribute Class

The *ChangeAttribute* class allows to define updates on the XML attribute values of the corresponding model. The *ChangeXML* class covers the possibilities provided by the *ChangeAttribute* class. That is, everything that can be expressed by a *ChangeAttribute* construct can also be expressed by a *ChangeXML*. However, both concepts exist to allow for being very specific in defining changes. It is recommended to use the *ChangeAttribute* for any changes of an XML attribute, and to use the more general *ChangeXML* for all other cases.

Table 4.2.3 shows all attributes and sub-elements for the *changeAttribute* element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
target	page 23
newValue	page 36
derived class	description link
notes ^o	page 16
annotation ^o	page 17

Table 10: Attributes and nested elements for *ChangeAttribute*. elements^o denotes optional elements.

One required attribute of the *changeAttribute* element is the *target* of change, i. e. the location

of the addressed XML attribute.

4.2.4 The `newValue` attribute

The second required attribute in the `changeAttribute` element is `newValue`, which assigns a new value to the targeted XML attribute.

An example for an SBML model is the update of the initial concentration of a certain parameter, as shown in listing 30.

```

1 <model id="model1" name="Circadian Chaos" language="urn:sedml:language:sbml" source="
  urn:miriam:biomodels.db:BIOMD0000000021">
2 <listOfChanges>
3   <changeAttribute target="/sbml:sbml/sbml:model/sbml:listOfParameters/sbml:parameter[@id='
    V_mT']/@value" newValue="0.28"/>
4   <changeAttribute target="/sbml:sbml/sbml:model/sbml:listOfParameters/sbml:parameter[@id='
    V_dT']/@value" newValue="4.8"/>
5 </listOfChanges>
6 </model>

```

Listing 30: The `changeAttribute` element and its `newValue` attribute

4.2.5 The `ComputeChange` Class

The `ComputeChange` class allows to make changes on any element of the XML file addressable by an XPath expression. The changes are described by mathematical expressions using a [subset of MathML](#).

Table 4.2.5 shows all attributes and sub-elements for the `computeChange` element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
target	page 23
math	page 36
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
variable ^o	page 22
parameter ^o	page 25

Table 11: Attributes and nested elements for `computeChange`. elements^o denotes optional elements. , **math** currently is an attribute, probably should be turned into an element (see [listing 31](#))

The `target` attribute contains the XPath addressing the piece of XML that is to be changed. It is possible to introduce additional parameters for the mathematics. Therefore, the parameters first need to be defined in the `listOfParameters`. They are then referenced through their ID. To use model variables for the definition of a mathematical expression, those variables need to be defined in the `listOfVariables` first, and can then be addressed by their ID.

4.2.5.1 The `math` attribute

The `math` element encodes mathematical functions. If used as an attribute of the `ComputeChange` class, it computes the change of the element or attribute addressed by the `target` attribute.

Level 1 Version 1 supports the subset of MathML 2.0 shown in section 2.1.

Listing 31 shows the use of the `computeChange` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <model [...]>
2   <computeChange target="/sbml/model/listOfParameters/parameter[@id='w']">
3     <listOfVariables>
4       <variable id="camkii" name="active calcium/calmoduline kinase II"
5         target="/sbml/model[@id='calcium']/listOfSpecies/species[@id='KII']" />
6       <variable id="w" name="synaptic weight"
7         target="/sbml/model[@id='synapse']/listOfParameters/parameter[@id='w']" /
8     </listOfVariables>
9     <listOfParameters>
10      <parameter id="w0" name="synaptic weight change" value="1">
11      <parameter id="n" name="utrasensitivity to calcium" value="2">
12      <parameter id="K" name="sensitivity to calcium" value="1e-6">
13    </listOfParameters>
14    <math>
15      <apply>
16        <plus />
17        <ci>w</ci>
18        <apply>
19          <times />
20          <ci>w0</ci>
21          <apply>
22            <divide />
23            <apply>
24              <power />
25              <ci>camkii</ci>
26              <ci>n</ci>
27            </apply>
28            <apply>
29              <plus />
30              <apply>
31                <power />
32                <ci>K</ci>
33                <ci>n</ci>
34              </apply>
35              <apply>
36                <power />
37                <ci> camkii </ci>
38                <ci>n</ci>
39              </apply>
40            </apply>
41          </apply>
42        </apply>
43      </apply>
44    </math>
45  </computeChange>
46 </listOfChanges>
47 </model>

```

Listing 31: *The computeChange element*

The example in listing 31 computes a change on the model parameter w . To do so, it defines two variables. `camkii` corresponds to the value of the species `KII` defined in the model with ID `calcium`. `w` corresponds to the value of the parameter `w` defined in the model with ID `synapse`. In addition, three further parameter are defined: `w0=1`, `n=2`, and `K=1e-6`. The mathematical expression in the mathML then computes the change of w using the encoded equation:

$$w + w_0 \times \frac{camkii^n}{camkii^n + K^n}$$

4.3 The Simulation Class

A simulation is the execution of some defined algorithm(s). Simulations are described differently depending on the type of simulation experiment to be performed (Figure 25 on the next page). `Simulation` is an abstract class and serves as the container for the different types of simulation experiments. SED-ML Level 1 Version 1 offers the predefined simulation class

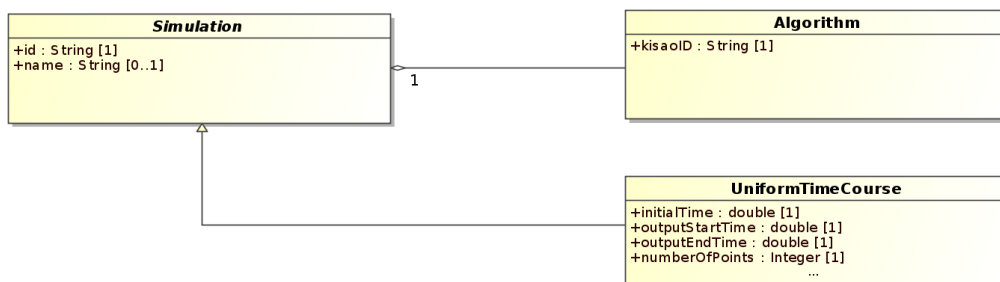


Figure 25: The SED-ML Simulation class

UniformTimeCourse. Further simulation classes are planned for future versions of SED-ML, including simulation classes for bifurcation analysis and parameter scans. Simulation algorithms used for the execution of a simulation setup are defined in the **Algorithm** class.

Table 4.3 shows all attributes and sub-elements for the **simulation** element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
algorithm	page 38

Table 12: Attributes and nested elements for **simulation**. elements^o denotes optional elements.

Listing 32 shows the use of the **simulation** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfSimulations>
2   <uniformTimeCourse id="s1" name="time course simulation of variable v1 over 100 minutes"
3     [...]>
4     <algorithm>
5       [ALGORITHM DEFINITION FOLLOWING]
6     </algorithm>
7   </uniformTimeCourse>
8   <uniformTimeCourse id="s2" name="time course definition for concentration of p" [...]>
9     [...]
10  </uniformTimeCourse>
11 </listOfSimulations>

```

Listing 32: The SED-ML **listOfSimulations** element, defining two different simulations

Two timcourses with uniform range are defined.

4.3.1 The Algorithm Class

SED-ML makes use of the **KiSA ontology** to refer to a term in the controlled vocabulary identifying the particular simulation algorithm to be used in the simulation.

Each instance of the **Simulation** class must contain one reference to a simulation algorithm. And each instance of the **Algorithm** class must contain a **KiSAO** reference to a simulation algorithm. The reference should define the simulation algorithm to be used in the simulation as precisely

as possible.

Table 4.3.1 shows all attributes and sub-elements for the [Algorithm](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
kisaoID	page 13
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 13: Attributes and nested elements for [algorithm](#). elements^o denotes optional elements.

The example given in code snippet 32, completed by algorithm definitions looks as in listing 33.

```

1 <listOfSimulations>
2   <uniformTimeCourse id="s1" name="time course simulation over 100 minutes" [...]>
3     <algorithm kisaoID="KiSAO:0000030" />
4   </uniformTimeCourse>
5   <uniformTimeCourse id="s2" name="time course definition for concentration of p" [...]>
6     <algorithm kisaoID="KiSAO:0000021" />
7   </uniformTimeCourse>
8 </listOfSimulations>

```

Listing 33: The SED-ML [algorithm](#) element, defining two different algorithms in the two defined simulations

For both simulations, one algorithm is defined. In the first simulation **s1** a deterministic simulation algorithm is used (Euler forward method), in the second simulation **s2** a stochastic one is used (Stochsim nearest neighbor).

4.3.2 The UniformTimeCourse Class

SED-ML Level 1 Version 1 so far only supports uniform time courses.

Table 4.3.2 shows all attributes and sub-elements for the [uniformTimeCourse](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
initialTime	page 40
outputStartTime	page 40
outputEndTime	page 40
numberOfPoints	page 40
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
algorithm	page 38

Table 14: Attributes and nested elements for [uniformTimeCourse](#). elements^o denotes optional elements.

Listing 34 shows the use of the `uniformTimeCourse` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfSimulations>
2   <uniformTimeCourse id="s1" name="time course simulation of variable v1 over 100 minutes"
3     initialTime="0" outputStartTime="0" outputEndTime="2500" numberOfPoints="1000">
4     <algorithm kisaoID="KISAO:0000030" />
5   </uniformTimeCourse>
6 </listOfSimulations>

```

Listing 34: The SED-ML `uniformTimeCourse` element, defining a uniform time course simulation over 2500 time units with 1000 simulation points, using the CVODE solver.

4.3.3 The `initialTime` attribute

tbw

4.3.4 The `outputStartTime` attribute

tbw

4.3.5 The `outputEndTime` attribute

tbw

4.3.6 The `numberOfPoints` attribute

tbw

4.4 The Task Class

A task in SED-ML links a `model` to a certain `simulation` description via the two according IDs (Figure 26), using the `modelReference` and the `simulationReference`. In SED-ML Level 1

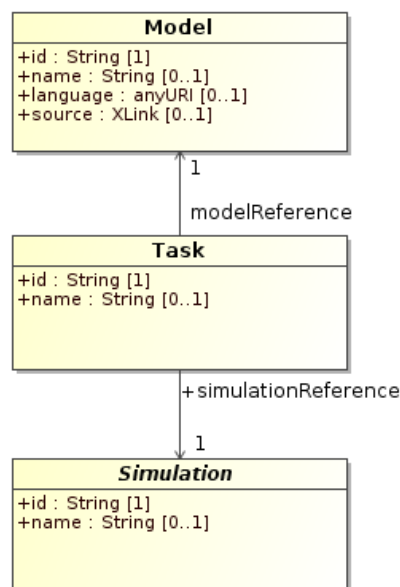


Figure 26: The SED-ML Task class

Version 1 it is only possible to link one simulation description to one model at a time. However, one can define as many tasks as needed within one experiment description, i.e. one SED-ML

file. Please note, that the tasks may be executed in any order, as XML does not have an ordering concept.

Table 4.4 shows all attributes and sub-elements for the [task](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
modelReference	page 21
simulationReference	page 22
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 15: Attributes and nested elements for [task](#). elements^o denotes optional elements.

Listing 35 shows the use of the `task` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfTasks>
2   <task id="t1" name="task definition" modelReference="model1"
3     simulationReference="simulation 1" />
4   <task id="t2" name="another task definition" modelReference="model2"
5     simulationReference="simulation 1" />
6 </listOfTasks>

```

Listing 35: The `task` element

In the example, a simulation setting *simulation1* is applied first to *model1* and then is applied to *model2*.

4.5 The DataGenerator Class

The [DataGenerator](#) class prepares the raw simulation results for later output (Figure 27). It stores the post-processings necessary to be applied to the simulation data. Those post-processing

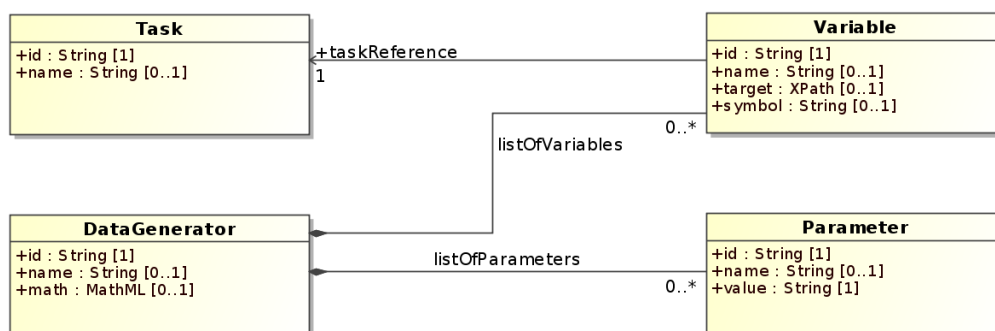


Figure 27: The SED-ML *DataGenerator* class

steps can be simple normalisation of data, but also mathematical calculations.

Table 4.5 on the following page shows all attributes and sub-elements for the [dataGenerator](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
math	page 36
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
variable ^o	page 22
parameter ^o	page 25

Table 16: *Attributes and nested elements for [dataGenerator](#). elements^o denotes optional elements.*

Each [dataGenerator](#) is identifiable within the experiment by its unambiguous [id](#). It can be further characterised by an optional [name](#). The [math](#) element contains a mathML expression for the calculation of the data generator. Mathematical functions available for the specification of [DataGenerator](#) variables are given in section 2.1 on [page 11](#). Within the mathematical expression, variables defined in the [listOfVariables](#) and parameters defined in the [listOfParameters](#) can be used.

Listing 36 shows the use of the `listOfDataGenerators` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```

1 <listOfDataGenerators>
2   <dataGenerator id="d1" name="time">
3     <listOfVariables>
4       <variable id="time" taskReference="task1" symbol="urn:sedml:symbol:time" />
5     </listOfVariables>
6     <listOfParameters />
7     <math xmlns="http://www.w3.org/1998/Math/MathML">
8       <ci> time </ci>
9     </math>
10  </dataGenerator>
11  <dataGenerator id="d2" name="LaCI repressor">
12    <listOfVariables>
13      <variable id="v1" taskReference="task1"
14        target="/sbml:sbml/sbml:model/sbml:listOfSpecies/
15          sbml:species[@id='PX']" />
16    </listOfVariables>
17    <math:math>
18      <math:ci>v1</math:ci>
19    </math:math>
20  </dataGenerator>
21 </listOfDataGenerators>

```

Listing 36: *The `listOfDataGenerators` element, defining two data generators time and LaCI repressor*

The `listOfDataGenerator` contains two `dataGenerator` elements. The first one, `d1`, refers to the task definition `t1` (which itself refers to a particular model), and from the corresponding model it reuses the symbol `time`. The second one, `d2`, references a particular species defined in the same model (and referred to via the `taskReference="t1"`). The model species with ID `PX` is reused for the data generator `d2` without further post-processing.

4.6 The Output Class

The [Output](#) class describes how the results of a simulation should be presented to the user ([Figure 28 on the next page](#)). It does not contain the data itself, but the type of output and the [data generators](#) used with the output type.

[Table 4.6 on the following page](#) shows all attributes and sub-elements for the [output](#) element as

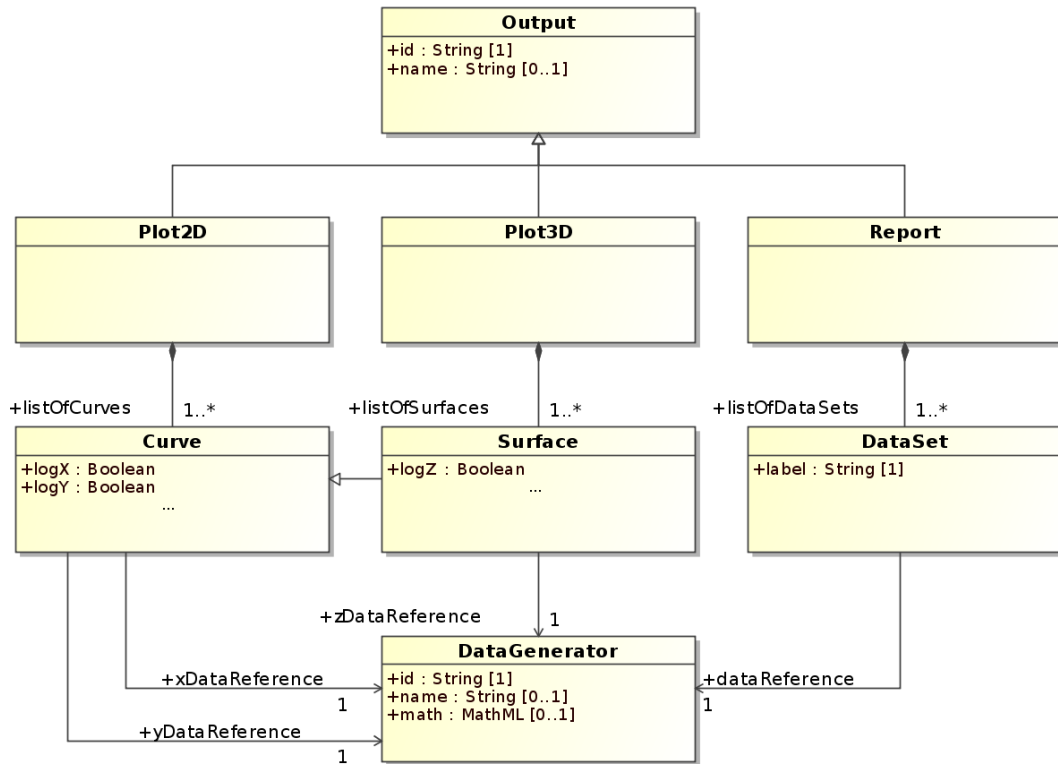


Figure 28: The SED-ML Output class

defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
plot2D ^o	page 43
plot3D ^o	page 44
report ^o	page 44

Table 17: Attributes and nested elements for *output*. elements^o denotes optional elements.

The types of output pre-defined in SED-ML are plots and [reports](#), also referrible as data tables. The output can be defined as a [2D plot](#) or alternatively as a [3D plot](#).

4.6.1 The Plot2D Class

Table [4.6.1 on the next page](#) shows all attributes and sub-elements for the [Plot2D](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

A 2D plot needs a data generator to be assigned to each of its two axes using the [xDataReference](#) to refer to a [dataGenerator](#) for the x-axis and using the [yDataReference](#) to refer to [dataGenerator](#) for the y-axis.



Figure 29: The SED-ML Plot2D class

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
curve	page 45

Table 18: Attributes and nested elements for [Plot2D](#). elements^o denotes optional elements.

An example for the definition of a 2D-plot with 1 defined curve is given in listing 37.

```

1 <plot2D>
2 <listOfCurves>
3 <curve>
4 [CURVE DEFINITION FOLLOWING]
5 </curve>
6 </listOfCurves>
7 </plot2D>

```

Listing 37: The *plot2D* element with the nested *listOfCurves* element

4.6.2 The Plot3D Class

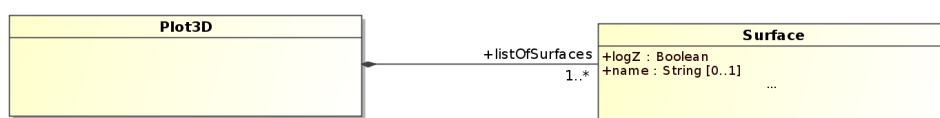


Figure 30: The SED-ML Plot3D class

Table 4.6.2 on the following page shows all attributes and sub-elements for the [Plot3D](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

A 3D plot is defined similarly to a 2D plot, but having 3 axes. To every axis a data generator is assigned using the [xDataReference](#), [yDataReference](#), and the [zDataReference](#). An example for a 3D plot with one defined surface is given in listing 38.

```

1 <plot3D>
2 <listOfSurfaces>
3 <surface>
4 [SURFACE DEFINITION FOLLOWING]
5 </surface>
6 </listOfSurfaces>
7 </plot3D>

```

Listing 38: The *plot3D* element with the nested *listOfSurfaces* element

4.6.3 The Report Class

Table 4.6.3 on page 46 shows all attributes and sub-elements for the [Report](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
surface	page 47

Table 19: Attributes and nested elements for *Plot3D*. elements^o denotes optional elements.

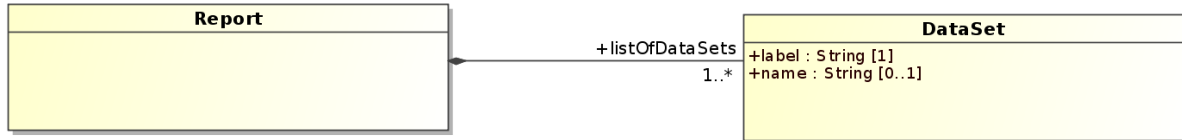


Figure 31: The SED-ML Report class

The [Report](#) class defines an output type that returns the simulation result in actual *numbers*. The particular columns of the report table are defined by creating a [DataSet](#) for each column.

Listing 39 provides an example for the definition of such a report with one column.

```

1 <report>
2   <listOfDataSets>
3     <dataSet>
4       [DATA REFERENCE FOLLOWING]
5     </dataSet>
6   </listOfDataSets>
7 </report>

```

Listing 39: The report element with the nested *listOfDataSets* element

The simulation result itself, i.e. concrete result numbers, are not stored in SED-ML, but the directive how to *calculate* them from the output of the simulator is provided through the [dataGenerator](#).

4.7 The Curve Class

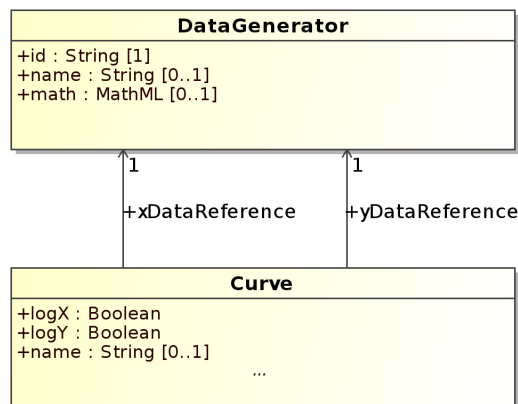


Figure 32: The SED-ML Curve class

Table 4.7 on the following page shows all attributes and sub-elements for the [Curve](#) element as

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
sub-elements	description
notes ^o	page 16
annotation ^o	page 17
dataSet	page 49

Table 20: Attributes and nested elements for [Report](#). elements^o denotes optional elements.

defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
logX	page 46
xDataReference	page 47
logY	page 46
yDataReference	page 47
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 21: Attributes and nested elements for [Curve](#). elements^o denotes optional elements.

4.7.1 The logX attribute

[logX](#) is an optional attribute of the [Curve](#) class and defines whether or not the data output on the x-axis is logarithmic or not. The data type of [logX](#) is boolean, the standard value is “false”. To make the output on the x-axis of a plot logarithmic, [logX](#) must be set to “true”, as shown in the sample listing 40:

```

1 <listOfCurves logX="true">
2   <curve id="c1" [...]>
3 </listOfCurves>

```

Listing 40: The SED-ML logX attribute, defining a logarithmic output on the x-axis of the according output

[logX](#) is also used in the definition of a [Surface](#) output.

4.7.2 The logY attribute

[logY](#) is an optional attribute of the [Curve](#) class and defines whether or not the data output on the y-axis is logarithmic or not. The data type of [logY](#) is boolean, the standard value is “false”. To make the output on the y-axis of a plot logarithmic, [logY](#) must be set to “true”, as shown in the sample listing 41:

```

1 <listOfCurves logY="true">
2   <curve id="c1" [...]>
3 </listOfCurves>

```

Listing 41: The SED-ML logY attribute, defining a logarithmic output on the y-axis of the according output

`logY` is also used in the definition of a `Surface` output.

4.7.3 The `xDataReference` attribute

The `xDataReference` is a mandatory attribute of the `Curve` object. It's content refers to a `dataGenerator` ID which denotes the `DataGenerator` object that is used to generate the output on the x-axis of a `Curve` in a `2D Plot`. `xDataReference` is also used in the definition of the x-axis of a `Surface` in a `3D Plot`. The `xDataReference` data type is string. However, the number of valid values for the `xDataReference` is restricted to the IDs of already defined `DataGenerator` objects.

An example for the definition of a curve is given in listing 42.

4.7.4 The `yDataReference` attribute

The `yDataReference` is a mandatory attribute of the `Curve` object. It's content refers to a `dataGenerator` ID which denotes the `DataGenerator` object that is used to generate the output on the y-axis of a `Curve` in a `2D Plot`. `yDataReference` is also used in the definition of the y-axis of a `Surface` in a `3D Plot`. The `yDataReference` data type is string. However, the number of valid values for the `yDataReference` is restricted to the IDs of already defined `DataGenerator` objects.

An example for the definition of a curve is given in the XML snippet in listing 42.

```
1 <listOfCurves>
2   <curve id="c1" name="v1 / time" xDataReference="dg1" yDataReference="dg2" logX="true" />
3 </listOfCurves>
```

Listing 42: The SED-ML curve element, defining the output curve showing the result of simulation for the referenced dataGenerators

Here, only one curve is created, results shown on the x-axis are generated by the data generator `dg1`, results shown on the y-axis are generated by the data generator `dg2`. Both `dg1` and `dg2` need to be already defined in the `listOfDataGenerators`. The x-axis is plotted logarithmically.

4.8 The Surface Class

A surface is a three-dimensional Figure representing a simulation result. Figure 33 shows the definition of the `Surface` class used to encode the information needed for the creation of such Figures in SED-ML.

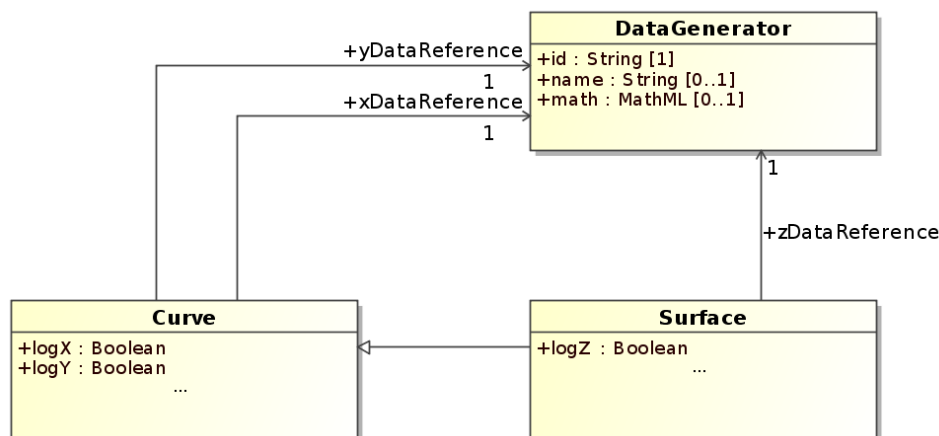


Figure 33: The SED-ML Surface class

Table 4.8 shows all attributes and sub-elements for the [Surface](#) element as defined by the SED-ML Level 1 Version 1 XML Schema. To define a surface, the three different axes have to

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
logX	page 46
xDataReference	page 47
logY	page 46
yDataReference	page 47
logZ	page 48
zDataReference	page 48
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 22: Attributes and nested elements for [Surface](#). elements^o denotes optional elements.

be defined, that is which data to plot on which axis and in which way. The aforementioned [xDataReference](#) and [yDataReference](#) attributes define the according dataGenerators for both the x- and y-axis of a surface. In addition, the [zDataReference](#) attribute defines the output for the z-axis. All axes might be logarithmic or not. This can be specified through the [logX](#), [logY](#), and the [logZ](#) attributes in the according dataReference elements.

4.8.1 The logZ attribute

[logZ](#) is an optional attribute of the [Surface](#) class and defines whether or not the data output on the z-axis is logarithmic or not. The data type of [logZ](#) is boolean, the standard value is “false”. To make the output on the z-axis of a surface plot logarithmic, [logZ](#) must be set to “true”, as shown in the sample listing 43:

```

1 <listOfSurfaces logZ="true">
2   <surface id="s1" [...>
3 </listOfSurfaces>
```

Listing 43: The SED-ML [logZ](#) attribute, defining a logarithmic output on the z-axis of the according output

4.8.2 The zDataReference attribute

The [zDataReference](#) is a mandatory attribute of the [Surface](#) object. It’s content refers to a dataGenerator ID which denotes the [DataGenerator](#) object that is used to generate the output on the z-axis of a [3D Plot](#). The [zDataReference](#) data type is string. However, the number of valid values for the [zDataReference](#) is restricted to the IDs of already defined [DataGenerator](#) objects.

An example for the definition of a surface is given in the XML snippet in listing 44.

```

1 <listOfSurfaces>
2   <surface id="s1" name="surface" algorithm="KiSA0:ID" xDataReference="dg1"
3     yDataReference="dg2" zDataReference="dg3">
4 </listOfSurfaces>
```

Listing 44: The SED-ML curve element, defining the output curve showing the result of the referenced task

Here, only one curve is created, results shown on the x-axis are generated by the data generator

dg1, results shown on the y-axis are generated by the data generator dg2, results shown on the z-axis are generated by the data generator dg3. All dg1, dg2 and dg3 need to be already defined in the [listOfDataGenerators](#).

4.9 The DataSet Class

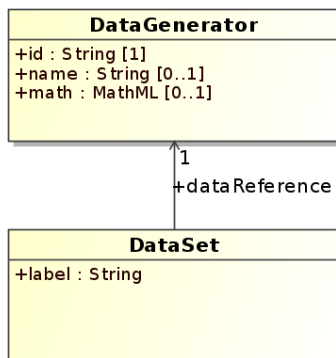


Figure 34: The SED-ML DataSet class

Table 4.9 shows all attributes and sub-elements for the [DataSet](#) element as defined by the SED-ML Level 1 Version 1 XML Schema.

attribute	description
metaid ^o	page 16
id	page 15
name ^o	page 15
dataReference	page 49
label	page 49
sub-elements	description
notes ^o	page 16
annotation ^o	page 17

Table 23: Attributes and nested elements for [DataSet](#). elements^o denotes optional elements.

4.9.1 The dataReference attribute

tbw

An example for the definition of a data set in form of a table is given in the XML snippet in [listing 45](#).

```

1 <listOfDataSets>
2   <dataSet id="d1" name="v1 over time" dataReference="dg1" label="_1">
3 </listOfDataSets>

```

Listing 45: The SED-ML dataSet element, defining a data set containing the result of the referenced task

4.9.2 The label attribute

Each data set in a [Report](#) does have to carry an unambiguous [label](#). The label is then used to refer to a particular data set in a report for later use.

Acknowledgements

The SED-ML specification has been developed with the input of many people. Main contributors of the current specification include Richard Adams, Frank Bergmann, Stefan Hoops, Nicolas Le Novère, Ion Moraru, Sven Sahle, Henning Schmidt and Dagmar Waltemath.

Moreover, we would like to thank all the participants of the meetings where SED-ML has been discussed as well as the subscribers of the sed-ml-discuss mailing list.

A SED-ML UML Overview

Figure 35 shows the complete UML diagram of the SED-ML. It gives the full picture of all implemented classes (see the XML Schema definition in 58).

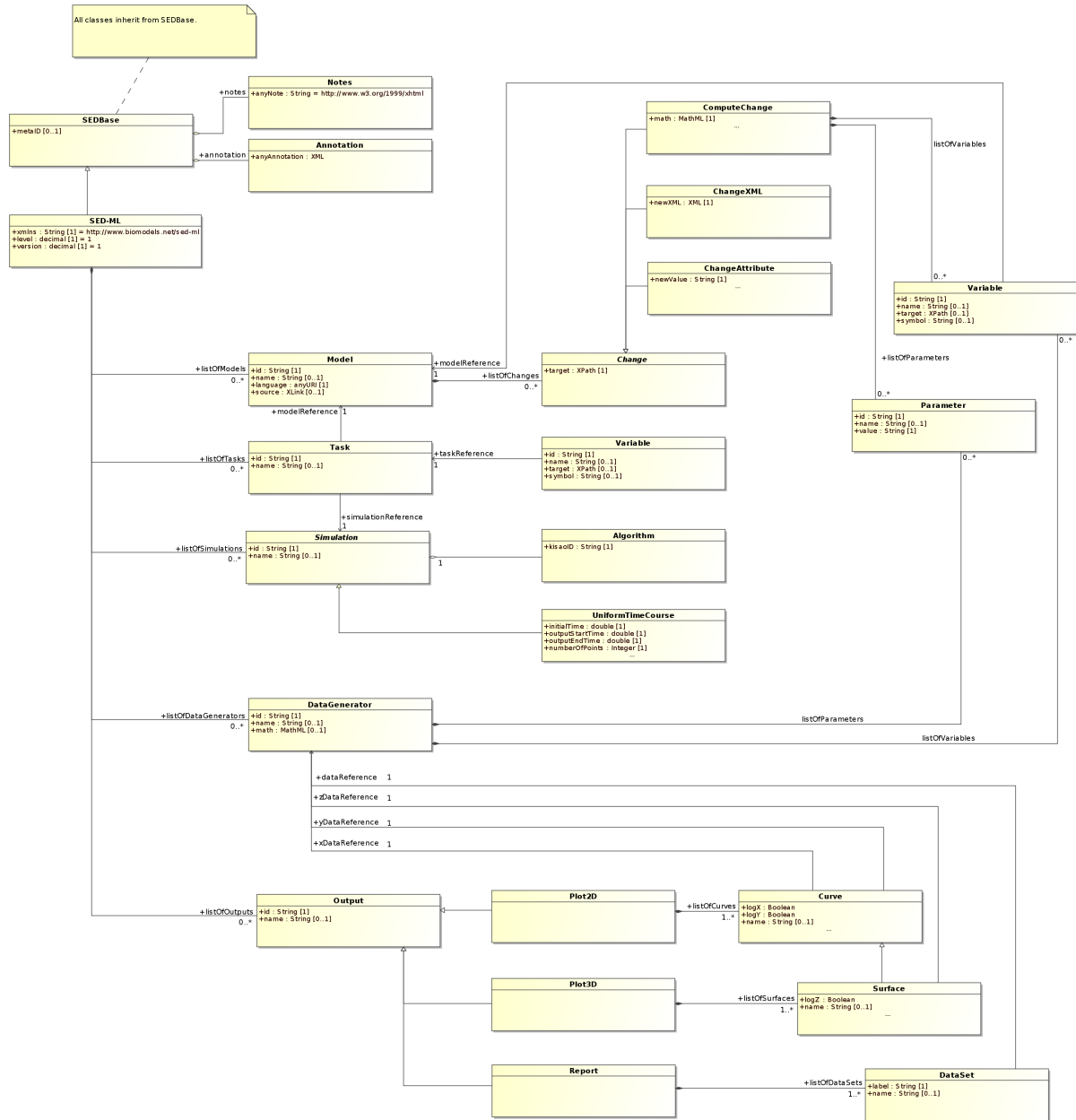


Figure 35: The SED-ML UML class diagram

B Overview of SED-ML

The *Simulation Experiment Description Markup Language* (SED-ML) is an XML-based format for the description of simulation experiments. It serves to store information about the simulation experiment performed on one or more models with a given set of outputs. Support for SED-ML compliant simulation descriptions will enable the exchange of simulation experiments across tools.

B.1 Conventions

The Business Process Modeling Notation Version 1.2 (BPMN) was initially intended to describe internal business procedures (processes) in a graphical way. However, we will use BPMN to graphically describe the steps and processes of setting up a simulation experiment description. The major parts of BPMN that are used to specify SED-ML are activities, gateways, events, data, and documentation.

An *activity* is “work that is performed on a [...] process”, for example “Specify the simulation settings”. Activities may be atomic or non-atomic. SED-ML in particular makes use of the *task* activities, i.e., specific work units that need to be performed. Non-atomic tasks might be collapsed or expanded in the graphical representation (see Figure 36). Each collapsed subprocess has a corresponding expanded subprocess definition.

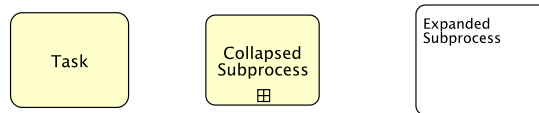


Figure 36: BPMN activities: task, collapsed process, expanded subprocess

Gateways serve as means to control the flow of sequence in the diagram. As the term already implies, a gateway needs some “mechanism that either allows or disallows passage through” [White et al., 2004]. The result of a gateway pass-through can be that processes are merged or splitted. Graphically, a gateway is represented as a diamond.

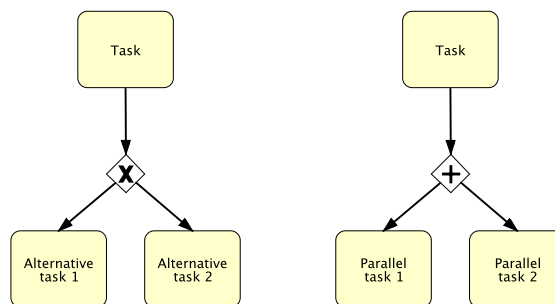


Figure 37: BPML gateway types: Exclusive (left), parallel (right)

While there exist a number of different gateway types (see [White et al., 2004], pp. 93), the SED-ML specification only uses the parallel and the exclusive gates (see Figure 37).

Exclusive gateways – also denoted as decisions – allow the sequence flow to take two or more

alternative paths (Figure 37 on the previous page, left hand side). However, *only one* of the paths may be chosen (not more). Sometimes two alternative branches need to be merged together again, in which case the exclusive gate must be used as well: The sequence flow continues as soon as *one* of the incoming processes send a signal. An exclusive gateways is marked by an **X** in the graphical notation.

Parallel gateways, “provide a mechanism to synchronize parallel flow and to create parallel flow” [White et al., 2004] (Figure 37 on the preceding page, right hand side). They are used to show parallel paths in the workflow; even if sometimes not required they might help in understanding the process. Synchronisation allows to start two processes in parallel at the same time in the sequence flow: The sequence flow will continue with *all* processes leaving the parallel gateway. Joining two processes with a parallel gateway is also possible: the process flow will only continue after a signal has arrived from *all* processes coming in the parallel gateway. A parallel gateway is marked by a + in the graphical notation.

Events mark everything happening during the execution of the sequence flow, usually they interrupt the business process, having some cause or impact on the execution. From the broad range of events that BPMN offers, SED-ML only uses a small subset, namely the start event and the end event (Figure 38).

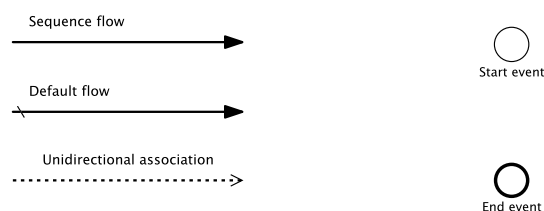


Figure 38: BPML connectors (left) and events (right).

All events are graphically drawn as small circles. A *start event* is drawn with a single thin line and mark the start of a process, it can not have any incoming sequence flow. Start event may be triggered by different mechanisms, for the case of SED-ML the untyped start event (no marker inside the circle) is used. The trigger to start the process is “Create new simulation experiment”. The *end event* is marked with a thick line. It indicates the end of a process. SED-ML specification makes use of the untyped end event (no marker inside the circle). The end event is used to show the end of sub-processes as well as processes. If the end of a sub-process is reached, the sequence flow returns to the according parent process.

Connectors are used to combine different BPMN objects with each other ([White et al., 2004] page 30 shows the full list of valid connections). SED-ML uses only a subset of available connectors, namely sequence flow, default flow, and unidirectional associations (Figure 38). *Sequence flow* defines the execution order of activities. *Default flow* marks the default branch to be chosen if other conditions leave various possibilities for further execution of the sequence flow. A *unidirectional association* is used to indicate that a data object is modified, i.e. read and written during the execution of an activity [Business Process Technology group, 2009].

The rough SED-ML workflow is shown in Figure 39. The process of defining a SED-ML simulation experiment starts by initialising the experiment and creating a new sed-ml file. Afterwards, the **models** needed for the simulation are specified and stored into the existing sed-ml file (see section B.1.1). In a third step, the simulation experiment **setups** are defined and stored into the same file (see section B.1.2). To assign a setup to a number of models used in the experiment, these connections have to be defined and recorded (see section B.1.3), called **task** in SED-ML.

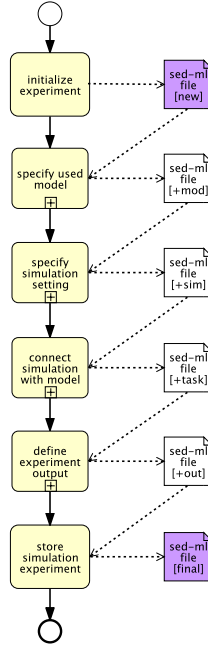


Figure 39: *The process of defining a simulation experiment in SED-ML (overview)*

After simulation, the **output** should be defined, based on the specified tasks and performed simulation experiment. The information is added to the existing SED-ML file (see section B.1.4). In the end, the whole experiment is stored in the final SED-ML file. All collapsed processes are described in the following. Examples in XML are provided in the more technical description

B.1.1 Models

To define a simulation experiment, first a new SED-ML file is created. The models to be used in the experiment (zero or many) are referenced, using a link to a model description in some open, curated model base (such as Biomedicals Database [Le Novère et al. \[2006\]](#), CellML Repository [Beard et al. \[2009\]](#), or alike). Changes that are necessary to simulate the model correctly are defined, e.g. assigning new parameter values or updating the mathematics of the model (Figure 40). The procedure is repeated until all models participating in the experiment

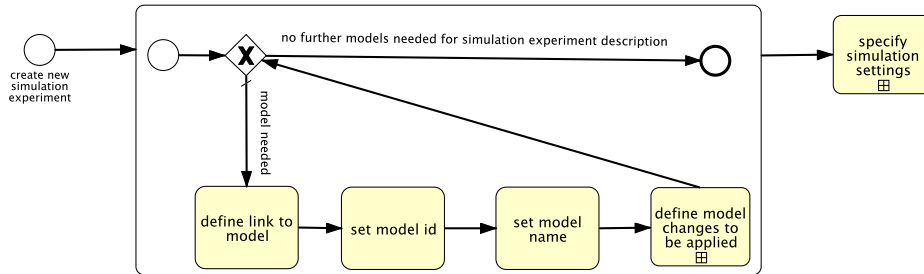


Figure 40: *The process of defining model(s) in SED-ML*

have been described. Each model used gets an internal SED-ML ID and an optional name.

B.1.2 Simulation setup

Secondly, the simulation setups (zero or many) used throughout the simulation experiment are described (Figure 41). Those may stem from various different types of simulation, e.g. steady

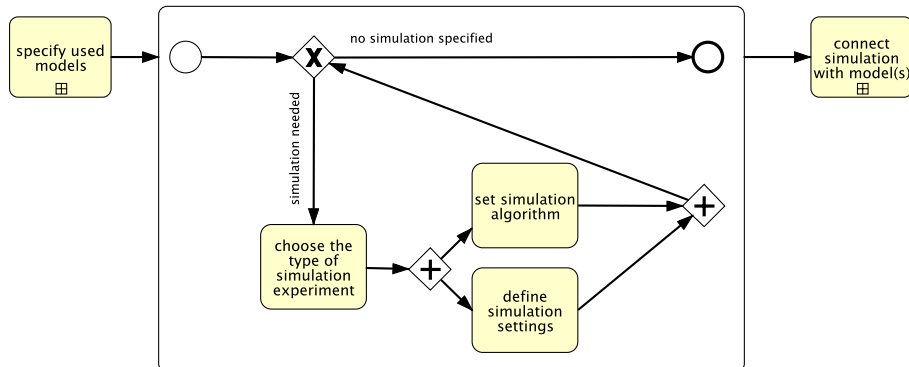


Figure 41: *The process of defining simulation(s) in SED-ML*

state analysis or bifurcation. Depending on the specific type of experiment, the information encoded for the simulation setup might differ. Thus, the definition of simulation settings is specific to the simulation experiment.

In a simple case the experiment consists of one simulation, but it can get far more complex. For example, one might define a nested sequence of simulations, in which case every simulation has to be defined separately. Each simulation setup gets its own internal ID and an optional name. For each of the setups, the simulation algorithm to be used for that simulation is defined through a reference to a well-defined algorithm name, e.g. an ontology or controlled vocabulary. One approach to define such a controlled vocabulary of simulation algorithms is the *Kinetic Simulation Algorithm Ontology* (KiSAO, Köhn and Le Novère [2008]). The setup definition is repeated until all different simulations have been described.

B.1.3 Task

SED-ML allows to apply one defined simulation setting to one defined model at a time. However, any number of **tasks** may be defined inside a simulation experiment description (Figure 42). To do so, each task refers to one of the formerly specified models and to one of the formerly specified simulation setups. Each task has its own ID and an optional name. The process of task definition is repeated until all tasks have been defined.

The current SED-ML does not allow to nest or order tasks. However, these features are evaluated for future versions of SED-ML.

B.1.4 Output

The SED-ML finally consists of output definitions that describe what kind of output the experiment uses to present the simulation result to the user, i.e. a plot or a data table (Figure 43), and also which data is part of the output. Therefore, SED-ML first defines a set of **data generators** (Figure 44), which are then used to specify a particular result, i.e. output (see section B.1.5).

The SED-ML specification comes with three pre-defined types of outputs: 2D- and 3D plots, and reports. All use the aforementioned data generators to specify the information to be plotted

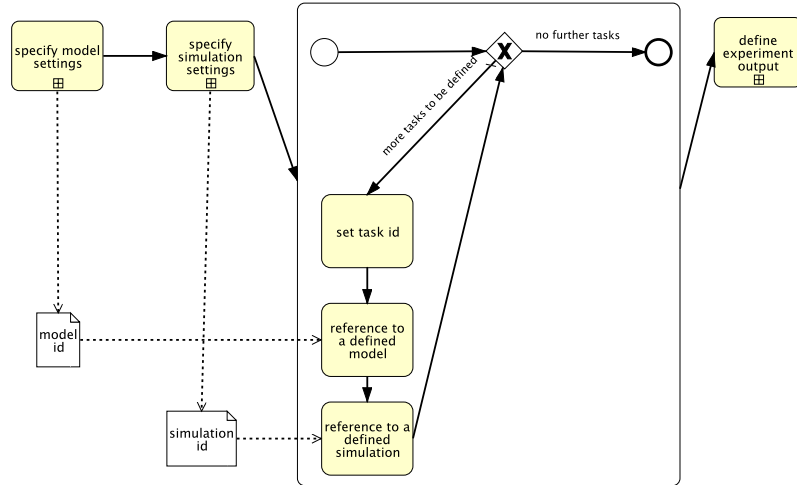


Figure 42: *The process of defining simulation task(s) in SED-ML*

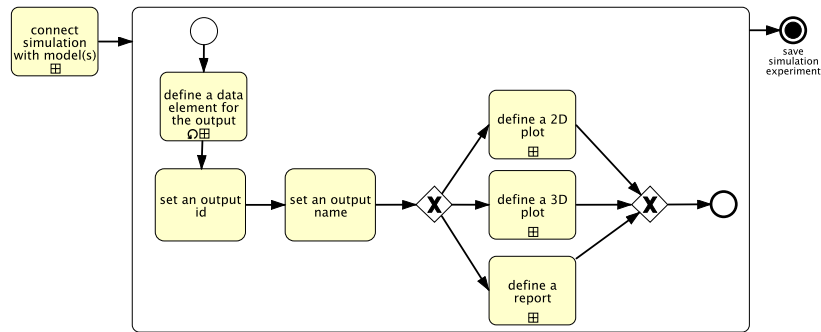


Figure 43: *The process of defining output(s) in SED-ML*

on the different axes, or in the table comlumns respectively.

B.1.5 Data Generator

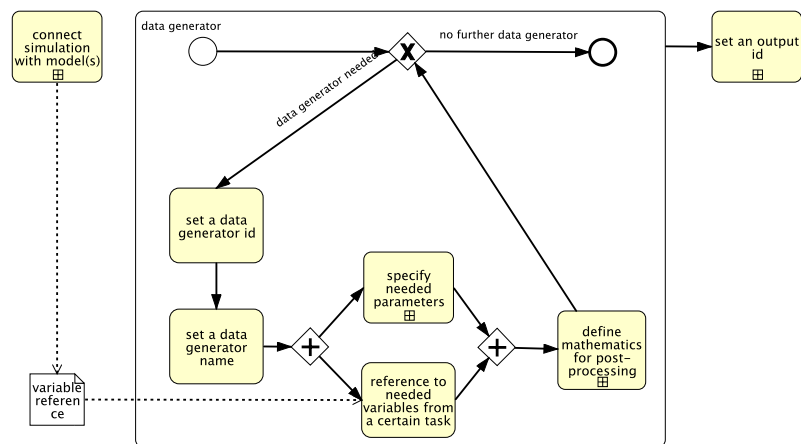


Figure 44: *The process of defining data generator(s) in SED-ML*

A data generator may use data elements, e. g. variables, or parameters, that either (1) have been taken directly from the model, or (2) have been generated in a post-processing step. If post-processing needs to be applied, variables and parameters from the various, previously defined models may be used, but also existing global parameters, such as *time*. If the variables are taken from existing models, a reference to the model and the particular variable needs to be given. If post-processing is necessary, a reference to an existing variable or parameter, including other data generators, has to be provided. Additional mathematical rules to be applied on the referred variable or parameter needs then to be specified. In a SED-ML file, any number of data generators can be created for later re-use in the output definition.

C XML Schema

Listing 46 shows the full SED-ML XML Schema. The code is commented inline.

```
1 <xs:schema targetNamespace="http://www.biomodels.net/sed-ml" xmlns="http://www.biomodels.net/sed-ml" xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:math="http://www.w3.org/1998/Math/MathML">
2 <!-- mathML import -->
3 <xs:import namespace="http://www.w3.org/1998/Math/MathML" schemaLocation="sbml-mathml.xsd" />
4
5 <!-- global element declarations -->
6 <!-- variable definition -->
7 <xs:element name="variable">
8   <xs:complexType>
9     <xs:attribute name="taskReference" type="xs:string" use="optional" />
10    <xs:attribute name="modelReference" type="xs:string" use="optional" />
11    <xs:attribute name="name" type="xs:string" use="optional" />
12    <xs:attribute name="target" type="xs:token" />
13    <xs:attribute name="id" type="xs:string" />
14  </xs:complexType>
15 </xs:element>
16 <!-- parameter definition -->
17 <xs:element name="parameter">
18   <xs:complexType>
19     <xs:attribute name="id" type="xs:string" />
20     <xs:attribute name="name" type="xs:string" use="optional" />
21     <xs:attribute name="value" type="xs:string" use="required" />
22   </xs:complexType>
23 </xs:element>
24 <!-- general simulation class definition -->
25 <xs:element name="anySimulation">
26   <xs:complexType>
27     <xs:sequence>
28       <xs:element name="simulationProperty">
29         <xs:complexType>
30           <xs:attribute name="propertyName" type="xs:string" use="required" />
31           <xs:attribute name="propertyValue" type="xs:string" use="required" />
32         </xs:complexType>
33       </xs:element>
34     </xs:sequence>
35   </xs:complexType>
36 </xs:element>
37 <!-- uniform time course simulation definition -->
38 <xs:element name="uniformTimeCourse">
39   <xs:complexType>
40     <xs:attribute name="id" type="xs:string" />
41     <xs:attribute name="outputStartTime" type="xs:double" use="required" />
42     <xs:attribute name="outputEndTime" type="xs:double" use="required" />
43     <xs:attribute name="numberOfPoints" type="xs:integer" use="required" />
44     <xs:attribute name="initialTime" type="xs:double" use="required" />
45     <xs:attribute name="algorithm" type="xs:string" />
46   </xs:complexType>
47 </xs:element>
48 <!-- task definition -->
49 <xs:element name="task">
50   <xs:complexType>
51     <xs:attribute name="simulationReference" type="xs:string" use="required" />
52     <xs:attribute name="name" type="xs:string" use="required" />
53     <xs:attribute name="modelReference" type="xs:string" use="required" />
54     <xs:attribute name="id" type="xs:string" />
55   </xs:complexType>
56 </xs:element>
57 <!-- notes element definition -->
58 <xs:element name="notes" type="xs:string" />
59 <!-- annotation element definition -->
60 <xs:element name="annotation" type="xs:string" />
61
62 <!-- complex type SEDBase definition -->
63 <xs:complexType name="SEDBase">
64   <xs:annotation>
65     <xs:documentation xml:lang="en">
66       The SEDBase type is the base type of all main types in SED-ML. It serves as a container
67       for the annotation of any part of the experiment description.
68     </xs:documentation>
69   </xs:annotation>
70 </xs:complexType>
```

```

68 </xs:annotation>
69 <xs:sequence>
70 <xs:element ref="notes" minOccurs="0" />
71 <xs:element ref="annotation" minOccurs="0" />
72 </xs:sequence>
73 </xs:complexType>
74
75 <!-- root element sedML definition -->
76 <xs:element name="sedML">
77 <xs:complexType>
78 <xs:complexContent>
79 <xs:extension base="SEDBase">
80 <xs:sequence>
81 <xs:element ref="listOfSimulations" />
82 <xs:element ref="listOfModels" />
83 <xs:element ref="listOfTasks" />
84 <xs:element ref="listOfDataGenerators" />
85 <xs:element ref="listOfOutputs" />
86 </xs:sequence>
87 <xs:attribute name="version" type="xs:decimal" use="required" fixed="0.1" />
88 </xs:extension>
89 </xs:complexContent>
90 </xs:complexType>
91 </xs:element>
92 <!-- 2D plot definition -->
93 <xs:element name="plot2D">
94 <xs:complexType>
95 <xs:sequence>
96 <xs:element ref="listOfCurves" />
97 </xs:sequence>
98 <xs:attribute name="name" type="xs:string" use="required" />
99 <xs:attribute name="id" type="xs:string" />
100 </xs:complexType>
101 </xs:element>
102 <!-- 3D plot definition -->
103 <xs:element name="plot3D">
104 <xs:complexType>
105 <xs:sequence>
106 <xs:element ref="listOfSurfaces" />
107 </xs:sequence>
108 <xs:attribute name="name" type="xs:string" use="required" />
109 <xs:attribute name="id" type="xs:string" />
110 </xs:complexType>
111 </xs:element>
112 <!-- report definition -->
113 <xs:element name="report">
114 <xs:complexType>
115 <xs:sequence>
116 <xs:element ref="listOfDataSets" />
117 </xs:sequence>
118 <xs:attribute name="name" type="xs:string" use="required" />
119 <xs:attribute name="id" type="xs:string" />
120 </xs:complexType>
121 </xs:element>
122 <!-- model element definition -->
123 <xs:element name="model">
124 <xs:complexType>
125 <xs:sequence>
126 <xs:element ref="listOfChanges" minOccurs="0" />
127 </xs:sequence>
128 <xs:attribute name="type" type="xs:string" use="required" />
129 <xs:attribute name="source" use="required" type="xs:string" />
130 <xs:attribute name="name" type="xs:string" use="required" />
131 <xs:attribute name="id" type="xs:string" />
132 </xs:complexType>
133 </xs:element>
134
135 <!-- <xs:element name="math" type="math:Math" /> -->
136 <!-- list definitions -->
137 <xs:element name="listOfVariables">
138 <xs:complexType>
139 <xs:sequence>
140 <xs:element ref="variable" />
141 </xs:sequence>
142 </xs:complexType>

```

```

143 </xs:element>
144 <xs:element name="listOfParameters">
145   <xs:complexType>
146     <xs:sequence>
147       <xs:element ref="parameter" />
148     </xs:sequence>
149   </xs:complexType>
150 </xs:element>
151 <xs:element name="listOfTasks">
152   <xs:complexType>
153     <xs:sequence>
154       <xs:element ref="task" maxOccurs="unbounded" />
155     </xs:sequence>
156   </xs:complexType>
157 </xs:element>
158 <xs:element name="listOfSimulations">
159   <xs:complexType>
160     <xs:sequence>
161       <xs:element ref="uniformTimeCourse" minOccurs="0" maxOccurs="unbounded"/>
162       <xs:element ref="anySimulation" minOccurs="0" maxOccurs="unbounded"/>
163     </xs:sequence>
164   </xs:complexType>
165 </xs:element>
166 <xs:element name="listOfOutputs">
167   <xs:complexType>
168     <xs:sequence minOccurs="0">
169       <xs:element ref="plot2D" minOccurs="0" />
170       <xs:element ref="plot3D" minOccurs="0" />
171       <xs:element ref="report" minOccurs="0" />
172     </xs:sequence>
173   </xs:complexType>
174 </xs:element>
175 <xs:element name="listOfModels">
176   <xs:complexType>
177     <xs:sequence>
178       <xs:element ref="model" maxOccurs="unbounded" />
179     </xs:sequence>
180   </xs:complexType>
181 </xs:element>
182 <xs:element name="listOfDataGenerators">
183   <xs:complexType>
184     <xs:sequence>
185       <xs:element ref="dataGenerator" maxOccurs="unbounded" />
186     </xs:sequence>
187   </xs:complexType>
188 </xs:element>
189 <xs:element name="listOfCurves">
190   <xs:complexType>
191     <xs:sequence>
192       <xs:element ref="curve" maxOccurs="unbounded" />
193     </xs:sequence>
194   </xs:complexType>
195 </xs:element>
196 <xs:element name="listOfSurfaces">
197   <xs:complexType>
198     <xs:sequence>
199       <xs:element ref="surface" maxOccurs="unbounded" />
200     </xs:sequence>
201   </xs:complexType>
202 </xs:element>
203 <xs:element name="listOfDataSets">
204   <xs:complexType>
205     <xs:sequence>
206       <xs:element ref="dataSet" maxOccurs="unbounded" />
207     </xs:sequence>
208   </xs:complexType>
209 </xs:element>
210 <xs:element name="listOfChanges">
211   <xs:complexType>
212     <xs:sequence>
213       <xs:element ref="changeAttribute" minOccurs="0" maxOccurs="unbounded" />
214       <xs:element ref="changeXML" minOccurs="0" maxOccurs="unbounded" />
215       <xs:element ref="changeMath" minOccurs="0" maxOccurs="unbounded" />
216     </xs:sequence>
217   </xs:complexType>

```

```

218 </xs:element>
219 <!-- data generator definition -->
220 <xs:element name="dataGenerator">
221   <xs:complexType>
222     <xs:sequence>
223       <xs:element ref="listOfVariables" minOccurs="0" />
224       <xs:element ref="listOfParameters" minOccurs="0" />
225       <xs:element ref="math:math" />
226     </xs:sequence>
227     <xs:attribute name="name" type="xs:string" use="required" />
228     <xs:attribute name="id" type="xs:string" />
229   </xs:complexType>
230 </xs:element>
231 <!-- curve definition (2D plot) -->
232 <xs:element name="curve">
233   <xs:complexType>
234     <xs:attribute name="yDataReference" type="xs:string" use="required" />
235     <xs:attribute name="xDataReference" type="xs:string" use="required" />
236     <xs:attribute name="logY" use="required" type="xs:boolean" />
237     <xs:attribute name="logX" use="required" type="xs:boolean" />
238   </xs:complexType>
239 </xs:element>
240 <!-- surface definition (3D plot) -->
241 <xs:element name="surface">
242   <xs:complexType>
243     <xs:attribute name="yDataReference" type="xs:string" use="required" />
244     <xs:attribute name="xDataReference" type="xs:string" use="required" />
245     <xs:attribute name="zDataReference" type="xs:string" use="required" />
246     <xs:attribute name="logY" use="required" type="xs:boolean" />
247     <xs:attribute name="logX" use="required" type="xs:boolean" />
248     <xs:attribute name="logZ" use="required" type="xs:boolean" />
249   </xs:complexType>
250 </xs:element>
251 <!-- data set definition (report) -->
252 <xs:element name="dataSet">
253   <xs:complexType>
254     <xs:attribute name="dataReference" type="xs:string" use="required" />
255   </xs:complexType>
256 </xs:element>
257 <!-- definition of attribute changes inside a model -->
258 <xs:element name="changeAttribute">
259   <xs:complexType>
260     <xs:attribute name="target" type="xs:token" />
261     <xs:attribute name="newValue" type="xs:string" use="required" />
262   </xs:complexType>
263 </xs:element>
264 <!-- definition of changes in the XML structure of the model -->
265 <xs:element name="changeXML">
266   <xs:complexType>
267     <xs:attribute name="target" type="xs:token" />
268     <xs:attribute name="newValue" type="xs:string" use="required" />
269   </xs:complexType>
270 </xs:element>
271 <!-- definition of changes in the mathematics of the model -->
272 <xs:element name="changeMath">
273   <xs:complexType>
274     <xs:sequence>
275       <xs:element ref="listOfVariables" />
276       <xs:element ref="listOfParameters" />
277       <xs:element ref="math:math" />
278     </xs:sequence>
279     <xs:attribute name="target" type="xs:token" />
280   </xs:complexType>
281 </xs:element>
282 </xs:schema>

```

Listing 46: SED-ML XML Schema definition

D Examples

D.1 Le Loup Model (CellML)

The following example provides a SED-ML description for the simulation of the model based on the publication by Leoup, Gonze and Goldbeter “Limit Cycle Models for Circadian Rhythms Based on Transcriptional Regulation in *Drosophila* and *Neurospora*” (PubMed ID: 10643740). The model source code is taken from the CellML Model Repository [Lloyd et al., 2008].

The original model used in the simulation experiment is referred to using a URL (http://models.cellml.org/workspace/leloup_gonze_goldbeter_1999/@rawfile/d6613d7e1051b3eff2bb1d3d419a445bb8c754ad/leloup_gonze_goldbeter_1999_b.cellml, ll. 15-16). In order to set up the model some pre-processing needs to be applied: Those are defined in the `listOfChanges` from ll. 17-25. All changes defined update particular parameter values in the model.

A second model is defined in l. 28 of the example, using `model1` as a source and applying even further changes to it, in this case updating two more model parameters.

One simulation setup is defined in the `listOfSimulations`. It is a `uniformTimeCourse` over 180 time units, using 1000 simulation points. The algorithm used is the CVODE solver, as denoted by the KiSAO ID `KiSAO:0000019`.

A number of `dataGenerators` are defined in ll. 42-92. Those are the prerequisite for defining the output of the simulation. The first `dataGenerator` named `tim1` in l. 45 maps on the `Mt` entity in the model that is used in `task1` which here is the model with ID `model1`. The second `dataGenerator` named `per-tim` in l. 57 maps on the `CN` entity in `model1`. Finally the third and fourth `dataGenerators` map on the `Mt` and `per-tim` entity respectively in the updated model with ID `model2`.

The `output` defined in the experiment consists of a 2D plot with two different curves (ll. 96-102). Both curves plot the `per-tim` concentration against the `tim` concentration. In the first curve the original parametrisation (as given in `model1`) is used, in the second curve the updated one is used (as given in `model2`).

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <sedML version="0.1" xmlns="http://www.biomodels.net/sed-ml" xmlns:math="http://www.w3.org
  /1998/Math/MathML">
3 <!-- textual information about the experiment (optional) -->
4 <notes>Comparing Limit Cycles and strange attractors for oscillation in Drosophila
5 </notes>
6 <!-- definition of simulation setup -->
7 <listOfSimulations>
8 <!-- definition of a uniform time course over 180 time units using the deterministic
  CVODE solver (KiSAO:0000019) -->
9 <uniformTimeCourse id="simulation1" algorithm="KiSAO:0000019" initialTime="0"
  outputStartTime="0" outputEndTime="180" numberOfPoints="1000" />
10 </listOfSimulations>
11 <!-- definition of models used during the experiment -->
12 <listOfModels>
13 <!-- reference to a cellML model -->
14 <model id="model1" name="Circadian Oscillations" type="CellML"
15 source="http://models.cellml.org/workspace/leloup_gonze_goldbeter_1999/@rawfile/
  d6613d7e1051b3eff2bb1d3d419a445bb8c754ad/leloup_gonze_goldbeter_1999_b.cellml" >
16 <!-- definition of changes to be applied to the original model (changing initial
  conditions) -->
17 <listOfChanges>
18 <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='MP']/cellml:variable
  [@name='vsP']/@initial_value" newValue="1"/>
19 <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='MP']/cellml:variable
  [@name='vmP']/@initial_value" newValue="0.7"/>
20 <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='P2']/cellml:variable
  [@name='vdP']/@initial_value" newValue="2"/>
21 <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='T2']/cellml:variable
  [@name='vdT']/@initial_value" newValue="2"/>
22 <changeAttribute target="/cellml:model/cellml:component[@name='parameters']/
  cellml:variable[@name='k1']/@initial_value" newValue="0.6"/>
23 <changeAttribute target="/cellml:model/cellml:component[@name='parameters']/
  cellml:variable[@name='K4P']/@initial_value" newValue="1"/>
24 <changeAttribute target="/cellml:model/cellml:component[@name='parameters']/
  cellml:variable[@name='K4T']/@initial_value" newValue="1"/>
25 </listOfChanges>
26 </model>
27 <!-- reference to the above model (model1) with additional changes of initial values of
  MY and T2 -->
28 <model id="model2" name="Circadian Chaos" type="CellML" source="model1">
29 <listOfChanges>
```

```

30     <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='MT']/cellml:variable
31       [name='vmT']/@initial_value" newValue="0.28"/>
32     <changeAttribute target="/cellml:model/cellml:component[@cmeta:id='T2']/cellml:variable
33       [name='vdT']/@initial_value" newValue="4.8"/>
34   </listOfChanges>
35 </model>
36 </listOfModels>
37 <!-- definition of tasks (combining simulation setup and model) -->
38 <listOfTasks>
39   <!-- limit cycle on model1 -->
40   <task id="task1" name="Limit Cycle" modelReference="model1" simulationReference="
41     simulation1"/>
42   <!-- strange attractors on the further perturbed model model2 -->
43   <task id="task2" name="Strange attractors" modelReference="model2" simulationReference="
44     simulation1"/>
45 </listOfTasks>
46 <!-- definition of the data generators needed to produce the output -->
47 <listOfDataGenerators>
48   <!-- definition of data generator for tim mRNA -->
49   <dataGenerator id="tim1" name="tim mRNA">
50     <listOfVariables>
51       <variable id="v1" taskReference="task1" target="/cellml:model/cellml:component[
52         @cmeta:id='MT']" />
53     </listOfVariables>
54     <math:math>
55       <math:apply>
56         <math:plus />
57         <math:ci>v1</math:ci>
58       </math:apply>
59     </math:math>
60   </dataGenerator>
61   <!-- definition of data generator for the nuclear PER-TIM complex -->
62   <dataGenerator id="per-tim" name="nuclear PER-TIM complex">
63     <listOfVariables>
64       <variable id="v1" taskReference="task1" target="/cellml:model/cellml:component[
65         @cmeta:id='CN']" />
66     </listOfVariables>
67     <math:math>
68       <math:apply>
69         <math:plus />
70         <math:ci>v1</math:ci>
71       </math:apply>
72     </math:math>
73   </dataGenerator>
74   <!-- definition of data generator for pertubated tim mRNA -->
75   <dataGenerator id="tim2" name="tim mRNA (changed parameters)">
76     <listOfVariables>
77       <variable id="v2" taskReference="task2" target="/cellml:model/cellml:component[
78         @cmeta:id='MT']" />
79     </listOfVariables>
80     <math:math>
81       <math:apply>
82         <math:plus />
83         <math:ci>v2</math:ci>
84       </math:apply>
85     </math:math>
86   </dataGenerator>
87   <!-- definition of data generator for perturbed nuclear PER-TIM complex -->
88   <dataGenerator id="per-tim2" name="nuclear PER-TIM complex">
89     <listOfVariables>
90       <variable id="v1" taskReference="task2" target="/cellml:model/cellml:component[
91         @cmeta:id='CN']" />
92     </listOfVariables>
93     <math:math>
94       <math:apply>
95         <math:plus />
96         <math:ci>v1</math:ci>
97       </math:apply>
98     </math:math>
99   </dataGenerator>
100 </listOfDataGenerators>
101 <!-- output definition -->
102 <listOfOutputs>
103   <!-- definition of a 2D plot to show the tim mRNA concentration with different initial
104     conditions -->

```

```

96 <plot2D id="plot1" name="tim mRNA with Oscillation and Chaos">
97 <!-- definition of two output curves, both plotting per-tim (original and perturbed)
    against the tim concentration (original and perturbed) -->
98 <listOfCurves>
99 <curve logX="false" logY="false" xDataReference="per-tim" yDataReference="tim1" />
100 <curve logX="false" logY="false" xDataReference="per-tim2" yDataReference="tim2" />
101 </listOfCurves>
102 </plot2D>
103 </listOfOutputs>
104 </sedML>

```

Listing 47: *LeLoup Model Simulation Description in SED-ML*

References

- D. A. Beard, R. Britten, M. T. Cooling, A. Garny, M. D. Halstead, P. J. Hunter, J. Lawson, C. M. Lloyd, J. Marsh, A. Miller, D. P. Nickerson, P. M. Nielsen, T. Nomura, S. Subramaniam, S. M. Wimalaratne, and T. Yu. Cellml metadata standards, associated tools and repositories. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 367 (1895):1845–1867, May 2009. ISSN 1364-503X. doi: 10.1098/rsta.2008.0310. URL <http://dx.doi.org/10.1098/rsta.2008.0310>.
- Donald Bell. Uml basics, part iii: The class diagram. IBM, the rational edge, 2003. http://download.boulder.ibm.com/ibmdl/pub/software/dw/rationaledge/nov03/t_modelinguml_db.pdf.
- T Berners-Lee, R Fielding, and L Masinter. Uniform resource identifier (URI): Generic syntax, 2005. URL <http://www.ietf.org/rfc/rfc3986.txt>.
- T Bray, J Paoli, CM Sperberg-McQueen, E Maler, F Yergeau, and J Cowan. Extensible markup language (XML) 1.1 (second edition), 2006. URL <http://www.w3.org/TR/xml11/>.
- Business Process Technology group. Bpmn – business process modeling notation. poster, 2009. URL <http://bpt.hpi.uni-potsdam.de/Public/BPMNCorner>.
- D. Carlisle, P. Ion, R. Miner, and N. Poppelier. Mathematical markup language (mathml) version 2.0. *W3C Recommendation*, 21, 2001.
- Mélanie Courtot, Dagmar Waltemath, Christian Knüpfer, et al. Controlled vocabularies and semantics in systems biology. to be submitted.
- MB Elowitz and S Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.
- D.C. Fallside, P. Walmsley, et al. XML schema part 0: Primer. *W3C recommendation*, 2, 2001.
- N Goddard, M Hucka, F Howell, H Cornelis, K Skankar, and D Beeman. Towards neuroml: Model description methods for collaborative modeling in neuroscience. *Phil. Trans. Royal Society series B*, 356:1209–1228, 2001.
- Stefan Hoops, Sven Sahle, Christine Lee, Jurgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. COPASI - a COMplex PATHway Simulator. *Bioinformatics (Oxford, England)*, 22(24):3067–3074, December 2006. ISSN 1460-2059. doi: 10.1093/bioinformatics/btl485. URL <http://dx.doi.org/10.1093/bioinformatics/btl485>.
- M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, March 2003. ISSN 1367-4803. doi: 10.1093/bioinformatics/btg015. URL <http://dx.doi.org/10.1093/bioinformatics/btg015>.
- Michael Hucka, Michael Hucka, Frank T. Bergmann, Stefan Hoops, Sarah Keating, Sven Sahle, and Darren J. Wilkinson. The systems biology markup language (sbml): Language specification for level 3 version 1 core (release 1 candidate). *Nature Precedings*, January 2010. ISSN 1756-0357. doi: 10.1038/npre.2010.4123.1. URL <http://dx.doi.org/10.1038/npre.2010.4123.1>.

- D. Köhn and N. Le Novère. The KInetic Simulation Algorithm Ontology (KiSAO)-A Proposal for the Classification of Simulation Algorithms in Systems Biology. Eingeladener Vortrag, 2008.
- N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, 34(Database issue), January 2006. ISSN 1362-4962. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16381960.
- Chen Li, Marco Donizelli, Nicolas Rodriguez, Harish Dharuri, Lukas Endler, Vijayalakshmi Chelliah, Lu Li, Enuo He, Arnaud Henry, Melanie Stefan, Jacky Snoep, Michael Hucka, Nicolas Le Novère, and Camille Laibe. Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4(1): 92+, June 2010. ISSN 1752-0509. doi: 10.1186/1752-0509-4-92. URL <http://dx.doi.org/10.1186/1752-0509-4-92>.
- Catherine M. Lloyd, Matt D. B. Halstead, and Poul F. Nielsen. Cellml: its future, present and past. *Prog Biophys Mol Biol*, 85:433–450, 2004.
- C.M. Lloyd, J.R. Lawson, P.J. Hunter, and P.F. Nielsen. The CellML model repository. *Bioinformatics*, 24(18):2122, 2008.
- OMG. *UML 2.2 Superstructure and Infrastructure*, February 2009. URL <http://www.omg.org/spec/UML/2.2/>.
- S. Pemberton et al. XHTML 1.0: The Extensible HyperText Markup Language—W3C Recommendation 26 January 2000. *World Wide Web Consortium (W3C)(August 2002)*, 2002.
- W3C. Xml schema part 1: Structures second edition. W3C Recommendation, October 2004. URL <http://www.w3.org/TR/xmlschema-1/>.
- D. Waltemath, R. Adams, D.A. Beard, F.T. Bergmann, U.S. Bhalla, R. Britten, V. Chelliah, M.T. Cooling, J. Cooper, E. Crampin, A. Garny, S. Hoops, M. Hucka, P. Hunter, E. Klipp, C. Laibe, A. Miller, i. Moraru, D. Nickerson, P. Nielsen, M. Nikolski, S. Sahle, H. Sauro, H. Schmidt, J.L. Snoep, D. Tolle, O. Wolkenhauer, and N. Le Novère. Minimum information about a simulation experiment (miase). *submitted to PLoS Comput Biol*.
- S.A. White et al. Business process modeling notation (BPMN) version 1.0. *Business Process Management Initiative, BPMI. org*, 2004.