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

March 25, 2011

Editors
Dagmar Waltemath
Frank T. Bergmann
Richard Adams
Nicolas Le Novère

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

The latest release of the Level 1 Version 1 specification is available at
http://sed-ml.org/

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

1.1 Motivation: A sample experiment

1.1.1 A simple time-course simulation

1.1.2 Applying pre-processing

1.1.3 Applying post-processing

1.2 Overview of SED-ML

1.2.1 Conventions

1.2.2 Models

1.2.3 Simulation setup

1.2.4 Task

1.2.5 Output

1.2.6 Data Generator

2 SED-ML technical specification

2.1 Conventions used in this document

2.1.1 UML Classes

2.1.2 UML Relationships

2.1.3 XML Schema language elements

2.1.4 Type extensions

2.2 Concepts used in SED-ML

2.2.1 MathML subset

2.2.2 URI Scheme

2.2.3 XPath usage

2.2.4 KiSAO

2.2.5 SED-ML resources

2.3 General attributes and classes

2.3.1 id

2.3.2 name

2.3.3 SEDBase

2.3.4 SED-ML top level element

2.3.5 Reference relations

2.3.6 Variable

2.3.7 Parameter

2.3.8 ListOf\textsuperscript{\textdegree} containers

2.4 SED-ML Components

2.4.1 Model

2.4.2 Change

2.4.3 Simulation

2.4.4 Task

2.4.5 DataGenerator

2.4.6 Output
2.4.7 Output components ........................................ 53

3 Acknowledgements ............................................. 59

A SED-ML UML Overview ....................................... 60

B XML Schema .................................................... 61

C Examples .......................................................... 67
   C.1 Le Loup Model (SBML) .................................... 68
   C.2 Le Loup Model (CellML) .................................. 70
   C.3 The IkappaB-NF-kappaB signaling module (SBML) .... 73

D SED-ML archive ............................................... 75
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], resulted in such 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., 2011].

This document presents Level 1 Version 1 of the Simulation Experiment Description Markup Language (SED-ML), a computer-readable format for encoding simulation experiments. SED-ML files are encoded in the eXtensible Markup Language (XML) [Bray et al., 2006]. The SED-ML format 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. [2011].

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 (URI) [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] led to the result shown in Figure 1.1.

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 $\text{tps}\_\text{repr}$ from “0.0005” to “1.3e-05”.

3. Change the value of the parameter $\text{tps}\_\text{active}$ 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 cl against time in a 2D Plot.

Figure 1.2 shows the result of the simulation.

1.1.3 Applying post-processing

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 given in section 1.1.2 above.)

5. Collect lacI(t), tetR(t) and cl(t).

6. Compute the highest value for each of the repressor proteins, max(lacI(t)), max(tetR(t)), max(cl(t)).

7. Normalize the data for each of the repressor proteins by dividing each time point by the maximum value, i.e. lacI(t)/max(lacI(t)), tetR(t)/max(tetR(t)) , and cl(t)/max(cl(t)).

8. Plot the normalized lacI protein as a function of the normalized cl, the normalized cl as a function of the normalized tetR protein, and the normalized tetR protein against the normalized lacI protein in a 2D plot.

Figure 1.3 on the following page illustrates the result of the simulation after post-processing of the output data.
Figure 1.3: 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. [2011])
1.2 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.

1.2.1 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 (Figure 1.4). Each collapsed subprocess has a corresponding expanded subprocess definition.

Figure 1.4: 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 split. Graphically, a gateway is represented as a diamond.

Figure 1.5: BPMN gateway types: Exclusive (left), parallel (right)

While there exist a number of different gateway types [White et al., 2004, p. 93], the SED-ML specification only uses the parallel and the exclusive gates (Figure 1.5).

Exclusive gateways – also denoted as decisions – allow the sequence flow to take two or more alternative paths (Figure 1.5, 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 gateway 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 1.5, 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 1.6).

![Figure 1.6: BPML connectors (left) and events (right).](image)

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 events 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, p. 30] show the full list of valid connections). SED-ML uses only a subset of available connectors, namely sequence flow, default flow, and unidirectional associations (Figure 1.6). *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 1.7. 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 (Section 1.2.2). In a third step, the simulation experiment *setups* are defined and stored into the same file (Section 1.2.3). To assign a setup to a number of models used in the experiment, these connections have to be defined and recorded (Section 1.2.4), called *task* in SED-ML. 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 (Section 1.2.5). In the end, the whole experiment is stored in the final SED-ML file. All collapsed processes are described in the following sections. Examples in XML are provided in the more technical description.

1.2.2 Models

To define a simulation experiment, first of all 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 database (e.g. Biomodels Database [Li et al., 2010] or CellML Repository [Beard et al., 2009]). All necessary changes to correctly simulate the model are defined, e.g., assigning new parameter values or updating the mathematics of the model (Figure 1.8). The procedure is repeated until all models participating in the experiment have been described. Each such model gets an internal SED-ML ID and an optional name.

1.2.3 Simulation setup

Secondly, the simulation setups (zero or many) used throughout the simulation experiment are described (Figure 1.9). Those may stem from various different types of simulation, e.g., steady 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 (Section 2.2.4). The setup definition is repeated until all different simulations have been described.

1.2.4 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 1.10). To do so, each

![Figure 1.10: The process of defining simulation task(s) in SED-ML](image)

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.

1.2.5 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 1.11), and also which data is part of the output. Therefore, SED-ML first defines a set of data generators (Figure 1.12), which

![Figure 1.11: The process of defining output(s) in SED-ML](image)
are then used to specify a particular result, i.e. output (Section 1.2.6).

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

1.2.6 Data Generator

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 must then be specified. In a SED-ML file, any number of data generators can be created for later re-use in the output definition.

Figure 1.12: The process of defining data generator(s) in SED-ML
2. SED-ML technical specification

This document represents the technical specification of SED-ML. We also provide an XML Schema [W3C, 2004] and a UML Class diagram representation of that XML Schema (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.

It should however be noted that some of the concepts of SED-ML cannot be captured using XML Schema alone. In these cases it is the specification that is considered the normative document.
2.1 Conventions used in this document

2.1.1 UML Classes

A SED-ML UML class (Figure 2.1) 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 2.1: SED-ML UML Class with class names and attributes](image)

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.

2.1.2 UML Relationships

2.1.2.1 UML Relation Types

![Figure 2.2: UML Class connectors](image)

Links between classes specify the connection of objects with each other (Figure 2.2). 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 2.3) 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).

![Figure 2.3: UML Association](image)

The aggregation (Figure 2.4 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 2.4 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, the departments will vanish with it [Bell, 2003].

The generalisation (Figure 2.5) allows to extend classes (BaseClass) by additional properties. The derived class (DerivedClass) inherits all properties of the base class and defines additional ones. In the given example, an instance of DerivedClass has two attributes attribute1 and attribute2.

2.1.2.2 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 2.6 expresses that a class is given by a professor, and a professor might give one to many classes.

2.1.3 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 current SED-ML XML Schema only makes use of complex type definitions. An example for a complex type definition is given in listing 2.1: 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 built 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 current SED-ML XML Schema only uses the sequence of elements.

2.1.3.1 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 2.2.

In this example, the dataGenerator type is built 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.
2.1.4 Type extensions

XML Schema offers mechanisms 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 2.3. 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 version.

```
<xs:element name="sedML">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="listOfSimulations" minOccurs="0" />  
          <xs:element ref="listOfModels" minOccurs="0" /> 
          <xs:element ref="listOfTasks" minOccurs="0" /> 
          <xs:element ref="listOfDataGenerators" minOccurs="0" /> 
          <xs:element ref="listOfOutputs" minOccurs="0" />
        </xs:sequence>
        <xs:attribute name="level" type="xs:decimal" use="required" fixed="1" />
        <xs:attribute name="version" type="xs:decimal" use="required" fixed="1" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

**Listing 2.3:** Definition of the sedML type through extension of SEDBase in SED-ML
2.2 Concepts used in SED-ML

2.2.1 MathML subset

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 a representation of mathematical expressions in other formats, such as SBML and CellML, two of the languages supported by SED-ML.

2.2.1.1 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`, `arccsc`, `arcsec`, `arccot`, `arcsinh`, `arccosh`, `arctanh`, `arcsech`, `arccsch`, `arccoth`
- **constants**: `true`, `false`, `notanumber`, `pi`, `infinity`, `exponentiale`
- **MathML annotations**: `semantics`, `annotation`, `annotation-xml`

2.2.1.2 MathML Symbols

All the operations listed above only operate on singular values. However, as one of SED-ML’s aims 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.

2.2.1.3 MathML functions

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

- **min**: Where the minimum of a variable represents the smallest value the simulation experiment yielded (Listing 2.4).

```
1 <apply>
2  <csymbol encoding="text" definitionURL="http://sed-ml.org/#min">
3    min
4  </csymbol>
5  <ci> variableId </ci>
6 </apply>
```

Listing 2.4: Example for the use of the MathML min function.

- **max**: Where the maximum of a variables represents the largest value the simulation experiment yielded (listing 2.5).
- **sum**: All values of the variable returned by the simulation experiment are summed (listing 2.6).
- **product**: All values of the variable returned by the simulation experiment are multiplied (listing 2.7).
These represent the only exceptions. At this point SED-ML Level 1 Version 1 does not define a complete algebra of vector values. For more information see the description of the DataGenerator class.

2.2.2 URI Scheme

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 particular information that can unambiguously be identified.

2.2.2.1 Model references

The preferred way for referencing a model from a SED-ML file is adopted from the MIRIAM URI Scheme. MIRIAM enables identification of a data resource (in this case a model 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 Resources¹. The ID entry maps to a particular model in the model repository. That model is never deleted. A sample BioModels Database ID is BIOMD0000000048. Together with the data resource name it becomes unambiguously identifiable by the URN urn:miriam:biomodels.db:BIOMD0000000048 (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 so². For the above example of the urn:miriam:biomodels.db:BIOMD0000000048 model, the resolved URL may look like:

- http://biomodels.caltech.edu/BIOMD0000000048
- http://www.ebi.ac.uk/biomodels-main/BIOMD0000000048

¹http://www.ebi.ac.uk/miriam/
²http://www.ebi.ac.uk/miriam/
depending on the physical location of the resource chosen to resolve the URN.

An alternative means to obtain a model may be to provide a single resource containing necessary models and a SED-ML file. Although a specification of such a resource is beyond the scope of this document, one proposal – SED-ML archive format – is described in Appendix D. Further information on the source attribute referencing the model location is provided in Section 2.4.1.2.

### 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 being XML, if no standardised representation format has been used to encode the model. On the other hand, one can be as 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://sed-ml.org/](http://sed-ml.org/). Further information on the language attribute is provided in Section 2.4.1.1.

### 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://sed-ml.org/](http://sed-ml.org/). From that source, 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 semantic annotations, 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/](http://www.biomodels.net/qualifiers/).

### 2.2.3 XPath usage

XPath is a language for finding information in an XML document [Clarke and DeRose, 1999]. Within Level 1 Version 1, XPath version 1 expressions are used to identify nodes and attributes within an XML representation of a biological model in the following ways:

1. Within a Variable definition, where XPath identifies the model variable required for manipulation in SED-ML.
2. Within a Change definition, where XPath is used to identify the target XML to which a change should be applied.

For proper application, XPath expressions should contain prefixes that allow their resolution to the correct XML namespace within an XML document. For example, the XPath expression referring to a species \( X \) in an SBML model:

```
/sbml:sbml/sbml:model/sbml:listOfSpecies/sbml:species[@id='X']
```

is preferable to

```
/sbml/model/listOfSpecies/species[@id='X']
```

which will only be interpretable by standard XML software tools if the SBML file declares no namespaces.
2.2.4 KiSAO

An important aspect of a simulation experiment is the simulation algorithm used to solve the system. But the sole reference of a simulation algorithm through its name in form of a string is error prone and ambiguous. 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 a controlled vocabulary to refer to a particular simulation algorithm. One attempt to provide such a vocabulary is the *Kinetic Simulation Algorithm Ontology* (KiSAO, [Courtot et al., 2011]). 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. KiSAO is available from BioPortal at [http://purl.bioontology.org/ontology/KiSAO](http://purl.bioontology.org/ontology/KiSAO). The project homepage is at [http://www.biomodels.net/kisao/](http://www.biomodels.net/kisao/).

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.

2.2.5 SED-ML resources

2.3 General attributes and classes

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

2.3.1 id

Most objects in SED-ML carry an id attribute. The id attribute, if it exists for an object, is always required and identifies SED-ML constituents unambiguously. The data type for id is SId which is a datatype derived from the basic XML type string, but with restrictions about the characters permitted and the sequences in which those characters may appear. The definition is shown in Figure 2.7.

\[
\begin{align*}
\text{letter} & ::= 'a'..'z','A'..'Z' \\
\text{digit} & ::= '0'..'9' \\
\text{idChar} & ::= \text{letter} | \text{digit} | '~' \\
\text{SId} & ::= ( \text{letter} | '~' ) \text{idChar}^* \\
\end{align*}
\]

Figure 2.7: The definition of the type SId

For a detailed description see also the SBML specification on the “Type SId” [Hucka et al., 2010, p. 11]. All ids have a global scope, i.e. the id must be unambiguous throughout a whole SED-ML document.

An example for a defined id is given in Listing 2.8. The defined model carries the id m00001. If the model is referenced elsewhere in the SED-ML document, it is referred to by that id.

Listing 2.8: SED-ML identifier definition, e.g. for a model

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

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 2.9 extends the model definition in listing 2.8 by a model name.

Listing 2.9: SED-ML name definition, e.g. for a model

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

2.3.3 SEDBase

SEDBase is the base class of SED-ML Level 1 Version 1. All other classes are derived from it. As such it provides means to attach additional information on all other classes (Figure 2.8 on the next page). That information can be specified by human readable Notes or custom Annotations.

Table 2.1 on the following page shows all attributes and sub-elements for the SEDBase element as defined by the SED-ML Level 1 Version 1 XML Schema.
2.3.3.1 metaid

The main purpose of the metaid attribute is to attach semantic annotations in form of the Annotation class to SED-ML elements. The type of metaid is XML ID and as such the metaid attribute is globally unique throughout the whole SED-ML document.

An example showing how to link a semantic annotation to a SED-ML object via the metaid is given in the Annotation class description.

2.3.3.2 Notes

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 used in top level XHTML elements contained within the notes element. For further options of how to set the namespace and detailed examples, please refer to [Hucka et al., 2010, p. 14].

Table 2.2 shows all attributes and sub-elements for the Notes element as defined by the SED-ML Level 1 Version 1 XML Schema. Notes does not have any further sub-elements defined in SED-ML, nor attributes associated with it.

Listing 2.10 on the following page shows the use of the notes element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. 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.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>xmlns:string</td>
<td>“<a href="http://www.w3.org/1999/xhtml%E2%80%9D">http://www.w3.org/1999/xhtml”</a></td>
</tr>
</tbody>
</table>

| sub-elements  | well-formed content permitted in XHTML |

Table 2.2: Attributes and nested elements for Notes. xy* denotes optional elements and attributes.
The enclosed simulation description shows the oscillating behaviour of the Repressilator model using deterministic and stochastic simulators.

Listing 2.10: The notes element

2.3.3.3 Annotation

An annotation is considered a computer-processible piece of information. Annotations may contain any valid XML content. For further guidelines on how to use annotations, we would like to encourage the reading of the corresponding section in the SBML specification [Hucka et al., 2010, pp. 14-16]. The style of annotations in SED-ML is briefly described in Section 2.2.2.4 on page 20.

Table 2.3 shows all attributes and sub-elements for the Annotation element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>none in the SED-ML namespace</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Attributes and nested elements for Annotation. *xy* denotes optional elements and attributes.

Listing 2.11 shows the use of the annotation element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. 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 suggested by the MIRIAM reference standard. 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.
2.3.4 SED-ML top level element

Each SED-ML Level 1 Version 1 document has a main class called SED-ML which defines the document’s structure and content (Figure 2.9). It consists of several parts; the parts are all connected to the SED-ML class through aggregation: the Model class (for model specification, see Section 2.4.1), the Simulation class (for simulation setup specification, see Section 2.4.3), the Task class (for the linkage of models and simulation setups, see Section 2.4.4), the DataGenerator class (for the definition of post-processing, see Section 2.4.5), and the Output class (for the output specification, see Section 2.4.6). All of them are shown in Figure 2.9 and will be explained in more detail in the relevant sections of this document.

![Figure 2.9: The sub-classes of SED-ML](image)

Table 2.4 on the next page shows all attributes and sub-elements for the SED-ML element as defined by the SED-ML Level 1 Version 1 XML Schema.

A SED-ML document needs to have the SED-ML namespace defined through the mandatory xmlns attribute. In addition, the SED-ML level and version attributes are mandatory.

The basic XML structure of a SED-ML file is shown in listing 2.12 on the following page. 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).

2.3.4.1 xmlns

The xmlns attribute declares the namespace for the SED-ML document. The pre-defined namespace for SED-ML documents is http://sed-ml.org/.
### Table 2.4: Attributes and nested elements for SED-ML

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaID</td>
<td>page 23</td>
</tr>
<tr>
<td>xmlns</td>
<td>page 25</td>
</tr>
<tr>
<td>level</td>
<td>page 26</td>
</tr>
<tr>
<td>version</td>
<td>page 26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation</td>
<td>page 24</td>
</tr>
<tr>
<td>listOfModels</td>
<td>page 33</td>
</tr>
<tr>
<td>listOfSimulations</td>
<td>page 34</td>
</tr>
<tr>
<td>listOfTasks</td>
<td>page 34</td>
</tr>
<tr>
<td>listOfDataGenerators</td>
<td>page 35</td>
</tr>
<tr>
<td>listOfOutputs</td>
<td>page 35</td>
</tr>
</tbody>
</table>

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 2.2.1 on page 18.

SED-ML notes use the XHTML namespace `http://www.w3.org/1999/xhtml`. The Notes class is described in Section 2.3.3.2 on page 23.

Additional external namespaces might be used in annotations.

#### 2.3.4.2 level

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 2.12.

#### 2.3.4.3 version

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 2.12.

#### 2.3.5 Reference relations

The reference concept is used to refer to a particular element inside the SED-ML document. It may occur in five different ways in the SED-ML document:

1. as an association between two Models (modelReference),
2. as an association between a Variable and a Model (modelReference),
3. as an association between a Variable and a Task (taskReference),
4. as an association between a Task and the associated Model (modelReference) or
5. as an association between a Task and the Simulation (simulationReference).
6. as an association between an Output and a DataGenerator (dataReference),

The definition of a Task object demands a reference to a particular Model object (modelReference, see Section 2.3.5.1 on page 27); furthermore, the Task object must be associated with a particular Simulation object (simulationReference, see Section 2.3.5.3 on page 28).

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 2.3.5.1 on page 27).

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 2.3.5.2 on page 28).

Four different types of data references exist in SED-ML Level 1 Version 1. They are used depending on the type of output for the simulation. A 2d plot has an xDataReference and a yDataReference assigned. A 3D plot has in addition a zDataReference assigned. To define a report, each data column has a dataReference assigned.

2.3.5.1 modelReference

The modelReference either represents a relation between two Model objects, a Variable object and a Model object, or a relation between a Task object and a Model object.

The source attribute of a Model is allowed to reference either a URI or an SID to a second Model. Constructs where a model A refers to a model B and B to A are invalid.

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 the id of a model that is defined in the document. Listing 2.13 shows the use of the modelReference element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. In the example, a change is applied on model m0001. In

```xml
<model id="m0001" [...]>
  <listOfChanges>
    <computeChange>
      <listOfVariables>
        <variable id="v1" modelReference="cellML" target="/cellml:model/cellml:component[@cmeta:id='MP ']/cellml:variable[@name='vsP ']/@initial_value" />
        [...] />
      </listOfVariables>
      <listOfParameters [...] />
      <math>
        [CALCULATION OF CHANGE]
      </math>
    </computeChange>
    [...] />
  </listOfChanges>
</model>
```

Listing 2.13: SED-ML modelReference attribute inside a variable definition of a computeChange element

the computeChange element a list of variables is defined. One of those variable is v1 which is defined in
another model (cellML). The XPath expression given in the target attribute identifies the variable in
the model which carries the ID cellML.

The modelReference is also used to indicate that a Model object is used in a particular Task. Listing
2.14 shows how this can be done for a sample SED-ML document. The example defines two different

```
<listOfTasks>
  <task id="t1" name="Baseline" modelReference="modell" simulationReference="simulation1" />
  <task id="t2" name="Modified" modelReference="model2" simulationReference="simulation1" />
</listOfTasks>
```

Listing 2.14: SED-ML modelReference definition inside a task element

tasks; the first one applies the simulation settings of simulation1 on modell, the second one applies the
same simulation settings on model2.

2.3.5.2 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. These 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 2.15
gives an example. The example shows the definition of a variable v1 in a dataGenerator element. The

```
<listOfDataGenerators>
  <dataGenerator id="tim3" name="tim mRNA (difference v1-v2+20)">
    <listOfVariables>
      <variable id="v1" taskReference="t1" />  <!-- ... -->
    </listOfVariables>
    <math [...]/>
  </dataGenerator>
</listOfDataGenerators>
```

Listing 2.15: SED-ML taskReference definition inside a dataGenerator element

variable appears in the model that is used in task t1. The task definition of t1 might look as shown in
Listing 2.16. Task t1 references the model modell. Therefore we can conclude that the variable v1
defined in listing 2.15 targets an element of the model with ID modell. The targeting process itself will
be explained in section 2.3.6.1 on page 30.

2.3.5.3 simulationReference

The simulationReference is used to refer to a particular Simulation in a Task. Listing 2.14 shows 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.

2.3.5.4 dataReference

The dataReference is used to refer to a particular DataGenerator instance from an Output instance.
Listing 2.17 shows the reference to a defined data set for a sample SED-ML document. In the example,
the output type is a 2D plot, which defines one curve with id c1. A curve must refer to two different
data generators which describe how to procure the data that is to be plotted on the x-axis and y-axis
respectively.

2.3.6 Variable

Variables are references to already existing entities, either existing in one of the defined models or externally defined symbols (Figure 2.10). If the variable is defined through a 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. A variable is always placed inside a listOfVariables. Symbol and target must not be used together in a single instance of Variable.

Table 2.5 shows all attributes and sub-elements for the Variable element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name</td>
<td>page 22</td>
</tr>
<tr>
<td>target</td>
<td>page 30</td>
</tr>
<tr>
<td>symbol</td>
<td>page 31</td>
</tr>
<tr>
<td>taskReference</td>
<td>page 28</td>
</tr>
<tr>
<td>modelReference</td>
<td>page 27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation</td>
<td>page 24</td>
</tr>
</tbody>
</table>

Table 2.5: Attributes and nested elements for Variable. "xy" denotes optional elements and attributes.

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 2.18 on the following page shows the use of the variable element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.18 defines a variable v1 (line 7) to compute a change on a model constituent (referenced by the target attribute on computeChange 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 (line 21), is used inside a dataGenerator. As the variable is time as used in task1, the symbol attribute is used to refer to the SED-ML URI for time (line 21).
2.3.6.1 target

An instance of Variable refers to a model constituent inside a particular model through an XPath expression stored in the required target attribute. XPath unambiguously identifies an element or attribute in an XML file.

Listing 2.19 shows the use of the target element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. It should be noted 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 listing 2.19, 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 Listing 2.20 is another valid example for the specification of a variable, but uses the same naming in the variable definition as in the original model (as opposed to Listing 2.19):

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 (Listing 2.21 on the following page).
2.3.6.2 symbol

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 2.2.2.3 on page 20.

Listing 2.22 shows the use of the symbol element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The example encodes a computed change of model m001. To specify that change, a symbol is defined (i.e. the SED-ML symbol for time is assigned to the variable t1). How to compute the change itself is explained in Section 2.4.2.6.

2.3.7 Parameter

The SED-ML Parameter class creates instances with a constant value (Figure 2.11). SED-ML 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 2.6 on the next page shows all attributes and sub-elements for the parameter element as defined by the SED-ML Level 1 Version 1 XML Schema.

A parameter can unambiguously be identified through its given id. It may additionally carry an optional name. Each parameter has one associated value.

Listing 2.23 on the following page shows the use of the parameter element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The listing shows the definition of a parameter p1 with the value="40" assigned.

2.3.7.1 value

Each parameter has exactly one fixed value. The value attribute of XML data type Double is required for each parameter element.
Table 2.6: Attributes and nested elements for parameter. *xy* denotes optional elements and attributes.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaID</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name</td>
<td>page 22</td>
</tr>
<tr>
<td>value</td>
<td>page 31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation</td>
<td>page 24</td>
</tr>
</tbody>
</table>

Listing 2.23: The definition of a parameter in SED-ML

2.3.8 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.

2.3.8.1 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 listOfVariables (Figure 2.12). 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 dataGenerator.

Listing 2.24 on the following page shows the use of the listOfVariables element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The listOfVariables is optional and may contain zero to many variables.

![Figure 2.12: The SED-ML listOfVariables container](image)
2.3.8.2 listOfVariables: The parameter definition container

All parameters needed throughout the simulation experiment, either to compute a change on a model prior to simulation (ComputeChange) or to set up a DataGenerator, are defined inside a listOfVariables (Figure 2.13).

Listing 2.24: SED-ML listOfVariables element

```
<listOfVariables>
  <variable id="v1" name="maximum velocity" taskReference="task1"
            target="/cellml:model/cellml:component[@cmeta:id='MP']/
cellml:variable[@name='vsP']/@initial_value" />
  <variable id="v2" taskReference="task2" symbol="urn:sedml:symbol:time" />
</listOfVariables>
```

Figure 2.13: The SED-ML listOfVariables container

Listing 2.25 shows the use of the listOfVariables 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.

```
<listOfVariables>
  <parameter id="p1" value="1" />
  <parameter id="p2" name="Kadp_2" value="6.23" />
</listOfVariables>
```

Listing 2.25: SED-ML listOfVariables element

2.3.8.3 listOfParameters: The parameter description container

In order to specify a simulation experiment, the participating models have to be defined. SED-ML uses the listOfParameters container for all necessary models (Figure 2.14 on the next page).

Listing 2.26 on the following page 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 listOfParameters is optional and may contain zero to many models. However, if the Level 1 Version 1 document contains one or more Task elements, at least one Model element must be defined to which the Task element refers (Section 2.3.5.1 on page 27).

```
<listOfParameters>
  <parameter id="p1" value="1" />
  <parameter id="p2" name="Kadp_2" value="6.23" />
</listOfParameters>
```

Listing 2.26: SED-ML listOfParameters element

2.3.8.4 listOfModels: The model description container

The listOfModels contains the defined changes to be applied to a particular model (Figure 2.15 on the next page). It always occurs as an optional subelement of the model element.

Listing 2.27 on the following page 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 nested inside the model element.
2.3.8.5 listOfSimulations: The simulation description container

The listOfSimulations element is the container for simulation descriptions (Figure 2.16 on the next page).

Listing 2.28 shows the use of the listOfSimulation element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. For all SED-ML Level 1 Version 1 documents, the encoded simulation definitions are instances of the Uniform Timecourse class. The listOfSimulations is optional and may contain zero to many simulations. However, if the Level 1 Version 1 document contains one or more Task elements, at least one Simulation element must be defined to which the Task element refers - see section 2.3.5.3 on page 28.

2.3.8.6 listOfTasks: The task specification container

The listOfTasks element contains the defined tasks for the simulation experiment (Figure 2.17 on the next page).

Listing 2.29 on the following page shows the use of the listOfTasks element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The listOfTasks is optional and may contain zero to many tasks. However, if the Level 1 Version 1 document contains a DataGenerator element with at
least one Variable element, at least one Task must be defined to which variable(s) in the DataGenerator element refers - see section 2.3.5.2 on page 28.

2.3.8.7 listOfDataGenerators: The post-processing container

In SED-ML, all variable- and parameter 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 2.18 on the next page).

Listing 2.30 shows the use of the listOfDataGenerators element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

The listOfDataGenerators is optional and in general may contain zero to many DataGenerators. However, if the Level 1 Version 1 document contains an Output element, at least one DataGenerator must be defined to which the Output element refers - see section 2.3.5.4 on page 28.

2.3.8.8 listOfOutputs: The output specification container

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

The output can be defined as either a report, a plot2D or as a plot3D.
Listing 2.31 shows the use of the `listOfOutputs` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The `listOfOutputs` is optional and may contain zero to many outputs.

```
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 2.31: The `listOfOutputs` element
2.4 SED-ML Components

In this section we describe the major components of SED-ML. We use the UML notation presented in section 2.1.1, and we show the use of SED-ML with XML examples. In addition, we provide a detailed BNMP diagram with explanation of the SED-ML workflow in Appendix 1.2 and an XML Schema in Appendix B.

2.4.1 Model

The Model class defines the models to be used in the simulation experiment (Figure 2.20).

![Figure 2.20: The SED-ML Model class](image)

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 or CellML.

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).

A model might need to undergo preprocessing before simulation. Those pre-processings are specified in the SED-ML Change class.

Table 2.7 shows all attributes and sub-elements for the model element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid&quot;</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name&quot;</td>
<td>page 22</td>
</tr>
<tr>
<td>language&quot;</td>
<td>page 38</td>
</tr>
<tr>
<td>source</td>
<td>page 38</td>
</tr>
<tr>
<td>sub-elements</td>
<td></td>
</tr>
<tr>
<td>notes&quot;</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation&quot;</td>
<td>page 24</td>
</tr>
<tr>
<td>change&quot;</td>
<td>page 39</td>
</tr>
</tbody>
</table>

Table 2.7: Attributes and nested elements for model. xy" denotes optional elements and attributes.

Listing 2.32 on the following page shows the use of the model element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

The above list of models 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:
BIOND0000000012. For the SED-ML simulation, the model might undergo preprocessing, described in the change element (lines 5-7). Based on the description of the first model m0001, the second model is built. 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 (lines 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.

### 2.4.1.1 language

The evaluation of a SED-ML document is required in order for software to decide whether or not it can be used for a particular simulation environment. One crucial criterion 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 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. The list of URNs for languages so far associated with SED-ML is available from the SED-ML web site on [http://sed-ml.org/](http://sed-ml.org/) (Section 2.2.2.2 on page 20). 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 for two reasons. Firstly, it helps 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. Secondly, the language attribute is also needed to decide how to handle the implicit variables in the Variable class, as the interpretation of implicit variables depends on the language of the representation format. The concept of implicit variables has been introduced in Section 2.2.2.3 on page 20.

### 2.4.1.2 source

To make a model available for the execution of a SED-ML file, the model source must be specified through either a URI or a reference to an SID of an existing Model.

The URI should preferably point to a public, consistent location that provides 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 resources\(^3\) can easily be referenced. Even without a MIRIAM URN, SED-ML can be used (Section 2.2.2.1 on page 19).

An example for the definition of a model, and using the URI scheme is given in Listing 2.33. The

```xml
<model id="m1" name="repressilator" language="urn:sedml:language:sbml"
  source="urn:miriam:biomodels.db:BIOMD0000000012">
  <listOfChanges>
    [MODEL PRE-PROCESSING]
  </listOfChanges>
</model>
```

Listing 2.33: The SED-ML source element, using the URI scheme

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 using the MIRIAM web service. The resulting URL is [http://www.ebi.ac.uk/biomodels-main/BIOMD0000000012](http://www.ebi.ac.uk/biomodels-main/BIOMD0000000012).

An example for the definition of a model and using a URL is given in Listing 2.34. In the example one

```xml
<model id="m1" name="repressilator" language="urn:sedml:language:cellml"
  source="http://models.cellml.org/exposure/bbae39f2c7baaaf51fd9454d6e73bbd3/apuda_b_1999.cellml">
  <listOfChanges />
</model>
```

Listing 2.34: The SED-ML source element, using a URL

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 attribute.

### 2.4.2 Change

SED-ML not only allows to use the sole model for simulation, but also enables the description of changes to be made on the model before simulation (Figure 2.21 on the next page). Changes can be of three distinct types:

1. Changes on attributes of the model’s XML representation (ChangeAttribute)
2. Changes on any XML snippet of the model’s XML representation (AddXML, ChangeXML, RemoveXML)
3. Changes based on mathematical calculations (ComputeChange)

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

Table 2.8 on the following page shows all attributes and sub-elements for the change element as defined by the SED-ML Level 1 Version 1 XML Schema.

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.

### 2.4.2.1 NewXML

The newXML element provides a piece of XML code (Figure 2.21 on the next page). NewXML must hold a valid piece of XML which after insertion into the original model must lead to a valid model file, according to the model language specification (as given by the language attribute).

\(^3\)http://www.ebi.ac.uk/miriam/main/
Table 2.8: Attributes and nested elements for change. \(xy^o\) denotes optional elements and attributes.

Table 2.9 shows all attributes and sub-elements for the newXML element as defined by the SED-ML Level 1 Version 1 XML Schema.

The newXML element is used at two different places inside SED-ML Level 1 Version 1:

1. If it is used as a sub-element of the addXML element, then the XML it contains is to be inserted as a child of the XML element addressed by the XPath.
2. If it is used as a sub-element of the changeXML element, then the XML it contains is to replace the XML element addressed by the XPath.

Examples are given in the relevant change class definitions.
2.4.2.2 AddXML

The AddXML class specifies a snippet of XML that is to be added as a child of the specified XPath target attribute (Figure 2.22). The new piece of XML code is provided by the NewXML class.

![Fig 2.22: The SED-ML AddXML class](image)

Table 2.10 shows all attributes and sub-elements for the addXML element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name</td>
<td>page 22</td>
</tr>
<tr>
<td>target</td>
<td>page 30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation</td>
<td>page 24</td>
</tr>
<tr>
<td>newXML</td>
<td>page 39</td>
</tr>
</tbody>
</table>

Table 2.10: Attributes and nested elements for addXML. * denotes optional elements and attributes.

An example for a change that adds an additional parameter to a model is given in listing 2.35. The code of the model is changed so that a parameter with ID VₘT is added to its list of parameters. The newXML element adds an additional XML element to the original model. The element’s name is parameter and it is added to the existing parent element listOfParameters that is addressed by the XPath expression in the target attribute.

2.4.2.3 ChangeXML

The ChangeXML class defines changes of any XML element in the model that can be addressed by a valid XPath expression (Figure 2.23 on the next page). The XPath is specified in the required target attribute (Section 2.3.6.1 on page page 30). The change of XML is specified in the NewXML class.

Table 2.11 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.

An example for a change that adds an additional parameter to a model is given in listing 2.36. The code of the model is changed in the way that its parameter with ID VₘT is substituted by two other parameters VₘT₁ and VₘT₂. The target attribute defines that the parameter with ID VₘT is to be changed. The newXML element then specifies the XML that is to be exchanged for that parameter.

2.4.2.4 RemoveXML

The RemoveXML class can be used to delete the XML element of the model that is addressed by the XPath expression (Figure 2.24 on page 43).
Listing 2.35: The addXML element with its newXML sub-element

```xml
<Listing 2.35: The addXML element with its newXML sub-element

```xml
<br>
```xml
<model language="urn:sedml:language:sbml" [..]>
  <addXML target="/sbml:sbml/sbml:model/sbml:listOfParameters">
    <newXML>
      <parameter metaid="metaid_0000010" id="V_mT" value="0.7" />
    </newXML>
  </addXML>
</model>
```

Figure 2.23: The ChangeXML class

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaId^</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name^</td>
<td>page 22</td>
</tr>
<tr>
<td>target</td>
<td>page 30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes^</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation^</td>
<td>page 24</td>
</tr>
<tr>
<td>newXML</td>
<td>page 39</td>
</tr>
</tbody>
</table>

Table 2.11: Attributes and nested elements for changeXML. \(^{xy}\) denotes optional elements and attributes.

Listing 2.36: The changeXML element

```xml
<Listing 2.36: The changeXML element

```xml
<br>
```xml
<model [..]>
  <listOfChanges>
    <changeXML target="/sbml:sbml/sbml:model/sbml:listOfParameters/sbml:parameter[@id='V_mT']">
      <newXML>
        <parameter metaid="metaid_0000010" id="V_mT_1" value="0.7" />
        <parameter metaid="metaid_0000050" id="V_mT_2" value="4.6" />
      </newXML>
    </changeXML>
  </listOfChanges>
</model>
```

The XPath is specified in the required target attribute.

Table 2.12 on the following page shows all attributes and sub-elements for the removeXML element as defined by the SED-ML Level 1 Version 1 XML Schema.

An example for the removal of an XML element from a model is given in Listing 2.37.

Listing 2.37: The removeXML element

```xml
<Listing 2.37: The removeXML element

```xml
<br>
```xml
<model [..]>
  <listOfChanges>
    <removeXML target="/sbml:sbml/sbml:model/sbml:listOfReactions/sbml:reaction[@id='J1']" />
  </listOfChanges>
</model>
```
The code of the model is changed by deleting the reaction with ID $V_{mT}$ from the model’s list of reactions.

2.4.2.5 ChangeAttribute

The ChangeAttribute class allows to define updates on the XML attribute values of the corresponding model (Figure 2.25).

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’s value, and to use the more general ChangeXML for all other cases.

ChangeAttribute requires to specify the target of change, i.e. the location of the addressed XML attribute, and also the new value of that attribute.

Table 2.13 shows all attributes and sub-elements for the changeAttribute element as defined by the SED-ML Level 1 Version 1 XML Schema.
2.4.2.5.1 **newValue**

The mandatory **newValue** attribute assigns a new value to the targeted XML attribute. The example in Listing 2.38 shows the update of the initial concentration of two parameters inside an SBML model.

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

*Listing 2.38: The changeAttribute element and its newValue attribute*

2.4.2.6 **ComputeChange**

The **ComputeChange** class permits to change, prior to the experiment, the value of any element or attribute of a model addressable by an XPath expression, based on a calculation (Figure 2.26). The changes are described by mathematical expressions using a subset of MathML (see section 2.2.1 on page 18). The computation can use the value of variables from any model defined in the simulation experiment. Those variables need to be defined, and can then be addressed by their ID. A variable used in a **ComputeChange** must carry a modelReference attribute (page 27) but no taskReference attribute (page 28). To carry out the calculation it may be necessary to introduce additional parameters, that are not defined in any of the model used by the experiment. This is done through the parameter class, thereafter referred to by their ID. Finally, the change itself is specified using an instance of the Math class.

Table 2.14 on the following page shows all attributes and sub-elements for the computeChange element as defined by the SED-ML Level 1 Version 1 XML Schema.

2.4.2.6.1 **Math**

The **Math** element encodes mathematical functions. If used as an element 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 2.2.1.

Listing 2.39 on the next page shows the use of the **computeChange** element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

The example in listing 2.39 computes a change of the variable **sensor** of the model **model2**. To do so, it uses the value of the variable **regulator** coming from model **model1**. In addition, the calculation used two additional parameters, the cooperativity **n**, and the sensitivity **K**. The mathematical expression in
Table 2.14: Attributes and nested elements for computeChange. \$xy^o\$ denotes optional elements and attributes.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid$o$</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name$o$</td>
<td>page 22</td>
</tr>
<tr>
<td>target</td>
<td>page 30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes$o$</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation$o$</td>
<td>page 24</td>
</tr>
<tr>
<td>listOfVariables$o$</td>
<td>page 32</td>
</tr>
<tr>
<td>listOfParameters$o$</td>
<td>page 33</td>
</tr>
<tr>
<td>math</td>
<td>page 44</td>
</tr>
</tbody>
</table>

Listing 2.39: The computeChange element

```
<model [...]>
  <computeChange target="/sbml:sbml/sbml:model/sbml:listOfParameters/sbml:parameter[@id='sensor']">
    <listOfVariables>
      <variable modelReference="model1" id="R" name="regulator"
                  target="/sbml:sbml/sbml:model/sbml:listOfSpecies/sbml:species[@id='regulator']"/>
      <variable modelReference="model2" id="S" name="sensor"
                  target="/sbml:sbml/sbml:model/sbml:listOfParameters/sbml:parameter[@id='sensor']"/>
    </listOfVariables>
    <listOfParameters>
      <parameter id="n" name="cooperativity" value="2">
        <parameter id="K" name="sensitivity" value="1e-6">
          <math>
            <apply>
              <times/>
              <ci>S</ci>
              <apply>
                <divide/>
                <apply>
                  <power/>
                  <ci>R</ci>
                  <ci>n</ci>
                </apply>
                <apply>
                  <plus/>
                  <apply>
                    <power/>
                    <ci>K</ci>
                    <ci>n</ci>
                  </apply>
                  <apply>
                    <power/>
                    <ci>R</ci>
                    <ci>n</ci>
                  </apply>
                </apply>
              </apply>
            </apply>
          </math>
        </parameter>
      </parameter>
    </listOfParameters>
  </computeChange>
</model>
```

the mathML then computes the new initial value of sensor using the equation:

\[ S = S \times \frac{R^n}{K^n + R^n} \]

2.4.3 Simulation

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 2.27 on the following 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 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 2.15 shows all attributes and sub-elements for the simulation element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid*</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name*</td>
<td>page 22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes*</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation*</td>
<td>page 24</td>
</tr>
<tr>
<td>algorithm</td>
<td>page 48</td>
</tr>
</tbody>
</table>

Table 2.15: Attributes and nested elements for simulation. \(xy^{*}\) denotes optional elements and attributes.

Listing 2.40 shows the use of the simulation element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. Two timecourses with uniform range are defined.

```
1 <listOfSimulations>
2  <uniformTimeCourse [...]>
3   [SIMULATION SPECIFICATION]
4  </uniformTimeCourse>
5  <uniformTimeCourse [...]>
6   [SIMULATION SPECIFICATION]
7  </uniformTimeCourse>
8 </listOfSimulations>
```

Listing 2.40: The SED-ML listOfSimulations element, defining two different simulations

2.4.3.1 UniformTimeCourse

SED-ML Level 1 Version 1 so far only supports the encoding of uniform time course experiments.

Table 2.16 on the next page shows all attributes and sub-elements for the uniformTimeCourse element as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.41 on the following page shows the use of the uniformTimeCourse element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

2.4.3.1.1 initialTime

The attribute initialTime of type \textit{double} represents the time from which to start the simulation. Usually this will be 0. Listing 2.41 shows an example.
2.4.3.1.2 outputStartTime

Sometimes a researcher is not interested in simulation results at the start of the simulation (i.e. the initial time). To accommodate this in SED-ML the `uniformTimeCourse` class uses the attribute `outputStartTime` of type `double`. To be valid the `outputStartTime` cannot be before `initialTime`. For an example, see Listing 2.41.

2.4.3.1.3 outputEndTime

The attribute `outputEndTime` of type `double` marks the end time of the simulation. See Listing 2.41 for an example.

2.4.3.1.4 numberOfPoints

When executed, the `uniformTimeCourse` simulation produces output on a regular grid starting with `outputStartTime` and ending with `outputEndTime`. The attribute `numberOfPoints` of type `integer` describes the number of points expected in the result. Software interpreting the `uniformTimeCourse` is expected to produce a first `outputPoint` at time `outputStartTime` with the initial values of the model to be simulated, and then `numberOfPoints` `outputPoints` with the results of the simulation. Thus a total of `numberOfPoints + 1` `outputPoints` will be produced.

Just because the output points lie on the regular grid described above, this does not mean that the simulation algorithm has to work with the same step size. Usually the step size the simulator chooses will be adaptive and much smaller than the required output step size. On the other hand a stochastic simulator might not have any new events occurring between two grid points. Nevertheless the simulator has to produce data on this regular grid. For an example, see Listing 2.41.
2.4.3.2 Algorithm

SED-ML makes use of the KiSAO ontology (Section 2.2.4 on page 21) 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 (Figure 2.29).

![Algorithm class](image)

Figure 2.29: The Algorithm class

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, and should be defined in the correct syntax, as defined by the regular expression KISAO:[0-9]{7}.

Table 2.17 shows all attributes and sub-elements for the Algorithm element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaId</td>
<td>page 23</td>
</tr>
<tr>
<td>kisaoID</td>
<td>page 21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation</td>
<td>page 24</td>
</tr>
</tbody>
</table>

Table 2.17: Attributes and nested elements for algorithm. *xy* denotes optional elements and attributes.

The example given in code snippet in Listing 2.40, completed by algorithm definitions results in the code given in Listing 2.42. For both simulations, one algorithm is defined. In the first simulation s1 a deterministic approach has been chosen (Euler forward method), in the second simulation s2 a stochastic approach is used (Stochsim nearest neighbor).

2.4.4 Task

A task in SED-ML links a model to a certain simulation description via their respective identifiers (Figure 2.30 on the next page), using the modelReference and the simulationReference. In SED-ML Level 1 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. Please note that the tasks may be executed in any order, as XML does not have an ordering concept.

Table 2.18 on the following page shows all attributes and sub-elements for the task element as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.43 on the next page shows the use of the task element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

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

2.4.5 DataGenerator

The DataGenerator class prepares the raw simulation results for later output (Figure 2.31 on page 50). It encodes the post-processing to be applied to the simulation data. The post-processing steps could be anything, from simple normalisations of data to mathematical calculations. Each instance of the DataGenerator class is identifiable within the experiment by its unambiguous id. It can be further char-
Listing 2.42: The SED-ML algorithm element for two different time course simulations, defining two different algorithms. KISAO:0000030 refers to the Euler forward method; KISAO:0000021 refers to the StochSim nearest neighbor algorithm.

Listing 2.43: The task element

Table 2.18: Attributes and nested elements for task. \(xy^o\) denotes optional elements and attributes.

Table 2.19 on the next page shows all attributes and sub-elements for the dataGenerator element as defined by the SED-ML Level 1 Version 1 XML Schema.
Figure 2.31: The SED-ML DataGenerator class

Table 2.19: Attributes and nested elements for dataGenerator. *xy* denotes optional elements and attributes.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metakP</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name*</td>
<td>page 22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>math</td>
<td>page 44</td>
</tr>
<tr>
<td>notes*</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation*</td>
<td>page 24</td>
</tr>
<tr>
<td>variable*</td>
<td>page 29</td>
</tr>
<tr>
<td>parameter*</td>
<td>page 31</td>
</tr>
</tbody>
</table>

Listing 2.44 shows the use of the dataGenerator element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The listOfDataGenerators contains two dataGenerator elements.

```xml
<listOfDataGenerators>
  <dataGenerator id="d1" name="time">
    <listOfVariables>
      <variable id="time" taskReference="task1" symbol="urn:sedml:symbol:time" />
    </listOfVariables>
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <ci> time </ci>
    </math>
  </dataGenerator>
  <dataGenerator id="d2" name="LaCI repressor">
    <listOfVariables>
      <variable id="v1" taskReference="task1" target="/sbml:sbml/sbml:model/sbml:listOfSpecies/sbml:species[@id='PX ']">
      </variable>
    </listOfVariables>
    <math:math>
      <math:ci>v1</math:ci>
    </math:math>
  </dataGenerator>
</listOfDataGenerators>
```

Listing 2.44: Definition of two dataGenerator elements, time and LaCI repressor

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 taskReferences="t1") . The model species with ID
PX is reused for the data generator d2 without further post-processing.

### 2.4.6 Output

The **Output** class describes how the results of a simulation should be presented to the user (Figure 2.32). It does not contain the data itself, but the type of output and the data generators used to produce a particular output.

The types of output pre-defined in SED-ML Level 1 Version 1 are plots and reports. The output can be defined as a 2D plot or alternatively as a 3D plot.

Note that even though the terms “2D plot” and “3D plot” are used, the exact type of plot is not specified. In other words, whether the 3D plot represents a surface plot, or three dimensional lines in space, cannot be distinguished by SED-ML alone. It is expected that applications use annotations for this purpose.

Table 2.20 shows all attributes and sub-elements for the output element as defined by the SED-ML Level 1 Version 1 XML Schema.

![Figure 2.32: The SED-ML Output class](image)

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid’</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name’</td>
<td>page 22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes’</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation’</td>
<td>page 24</td>
</tr>
<tr>
<td>plot2D’</td>
<td>page 52</td>
</tr>
<tr>
<td>plot3D’</td>
<td>page 52</td>
</tr>
<tr>
<td>report’</td>
<td>page 52</td>
</tr>
</tbody>
</table>

*xy’ denotes optional elements and attributes.*
2.4.6.1 Plot2D

A 2 dimensional plot (Figure 2.33) contains a number of curve definitions.

![Plot2D diagram]

**Figure 2.33:** The SED-ML Plot2D class

Table 2.21 shows all attributes and sub-elements for the plot2D element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes(^a)</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation(^a)</td>
<td>page 24</td>
</tr>
<tr>
<td>curve</td>
<td>page 53</td>
</tr>
</tbody>
</table>

Table 2.21: Attributes and nested elements for plot2D. \(^a\) denotes optional elements and attributes.

Listing 2.45 shows the use of the listofCurves element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The listing shows the definition of a 2 dimensional plot containing one curve element inside the listofCurves. The curve definition follows in Section 2.4.7.1 on page 53.

```xml
<plot2D>
  <listOfCurves>
    <curve>
      [CURVE DEFINITION]
    </curve>
    [FURTHER CURVE DEFINITIONS]
  </listOfCurves>
</plot2D>
```

Listing 2.45: The plot2D element with the nested listofCurves element

2.4.6.2 Plot3D

A 3 dimensional plot (Figure 2.34 on the following page) contains a number of surface definitions.

Table 2.22 on the next page shows all attributes and sub-elements for the plot3D element as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.46 on the following page shows the use of the plot3D element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. The example defines a surface for the 3 dimensional plot. The surface definition follows in Section 2.4.7.2 on page 55.

2.4.6.3 The Report class

The Report class defines a data table consisting of several single instances of the DataSet class (Figure 2.35 on page 54). Its output returns the simulation result in actual numbers. The particular columns of the report table are defined by creating an instance of the DataSet class for each column.

Table 2.23 on page 54 shows all attributes and sub-elements for the report element as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.47 on the next page shows the use of the listofDataSets element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.
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.

The encoding of simulation results is outside the scope of SED-ML, but other efforts exist, for example the Systems Biology Result Markup Language (SBRML, [Dada et al., 2010]).

### 2.4.7 Output components

#### 2.4.7.1 Curve

One or more instances of the Curve class define a 2D plot. A curve needs a data generator reference to refer to the data that will be plotted on the x-axis, using the xDataReference. A second data generator reference is needed to refer to the data that will be plotted on the y-axis, using the yDataReference.

Table 2.24 on page 55 shows all attributes and sub-elements for the curve element as defined by the SED-ML Level 1 Version 1 XML Schema.

Listing 2.48 shows the use of the curve element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema. 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 \texttt{listOfDataGenerators}. The x-axis is plotted logarithmically.

2.4.7.1.1 \texttt{logX}

\texttt{logX} is a required attribute of the \texttt{Curve} class and defines whether or not the data output on the x-axis is logarithmic. The data type of \texttt{logX} is \texttt{boolean}. To make the output on the x-axis of a plot logarithmic, \texttt{logX} must be set to “true”, as shown in the sample Listing 2.48.

\texttt{logX} is also used in the definition of a \texttt{Surface} output.

2.4.7.1.2 \texttt{logY}

\texttt{logY} is a required attribute of the \texttt{Curve} class and defines whether or not the data output on the y-axis is logarithmic. The data type of \texttt{logY} is \texttt{boolean}. To make the output on the y-axis of a plot logarithmic, \texttt{logY} must be set to “true”, as shown in the sample Listing 2.48.

\texttt{logY} is also used in the definition of a \texttt{Surface} output.

2.4.7.1.3 \texttt{xDataReference}

The \texttt{xDataReference} is a mandatory attribute of the \texttt{Curve} object. Its content refers to a dataGenerator ID which denotes the \texttt{DataGenerator} object that is used to generate the output on the x-axis of a \texttt{Curve} in a 2D Plot. The \texttt{xDataReference} data type is \texttt{string}. However, the valid values for the \texttt{xDataReference} are restricted to the IDs of already defined \texttt{DataGenerator} objects.

An example for the definition of a curve is given in Listing 2.48. \texttt{xDataReference} is also used in the definition of the x-axis of a \texttt{Surface} in a 3D Plot.
Table 2.24: Attributes and nested elements for curve. \( xy \) denotes optional elements and attributes.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaId(^\circ)</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name(^\circ)</td>
<td>page 22</td>
</tr>
<tr>
<td>logX</td>
<td>page 54</td>
</tr>
<tr>
<td>xDataReference</td>
<td>page 54</td>
</tr>
<tr>
<td>logY</td>
<td>page 54</td>
</tr>
<tr>
<td>yDataReference</td>
<td>page 55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes(^\circ)</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation(^\circ)</td>
<td>page 24</td>
</tr>
</tbody>
</table>

2.4.7.1.4 \( y \text{DataReference} \)

The \( y \text{DataReference} \) is a mandatory attribute of the Curve object. Its 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. The \( y \text{DataReference} \) data type is \textit{string}. However, the number of valid values for the \( y \text{DataReference} \) is restricted to the IDs of already defined DataGenerator objects.

An example for the definition of a curve is given in listing 2.48. \( y \text{DataReference} \) is also used in the definition of the y-axis of a Surface in a 3D Plot.

2.4.7.2 Surface

A surface is a three-dimensional figure representing a simulation result (Figure 2.37).

Creating an instance of the Surface class demands the definition of three different axes, that is which data to plot on which axis and in which way. The aforementioned xDataReference and yDataReference attributes define the according data generators 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.

Table 2.25 on the next page shows all attributes and sub-elements for the surface element as defined by the SED-ML Level 1 Version 1 XML Schema. Listing 2.49 on the following page shows the use of the surface element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

Here, only one surface is created, results shown on the x-axis are generated by the data generator \( \text{dg1} \), results shown on the y-axis are generated by the data generator \( \text{dg2} \), and results shown on the z-axis are generated by the data generator \( \text{dg3} \). All \( \text{dg1} \), \( \text{dg2} \) and \( \text{dg3} \) need to be already defined in the \text{listOfDataGenerators}. 
### Table 2.25: Attributes and nested elements for surface. xy° denotes optional elements and attributes.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name°</td>
<td>page 22</td>
</tr>
<tr>
<td>logX</td>
<td>page 54</td>
</tr>
<tr>
<td>xDataReference</td>
<td>page 54</td>
</tr>
<tr>
<td>logY</td>
<td>page 54</td>
</tr>
<tr>
<td>yDataReference</td>
<td>page 55</td>
</tr>
<tr>
<td>logZ</td>
<td>page 56</td>
</tr>
<tr>
<td>zDataReference</td>
<td>page 56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes°</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation°</td>
<td>page 24</td>
</tr>
</tbody>
</table>

logZ is a required attribute of the Surface class and defines whether or not the data output on the z-axis is logarithmic. The data type of logZ is boolean. 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 2.49.

### 2.4.7.2.2 zDataReference

The zDataReference is a mandatory attribute of the Surface object. Its 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 valid values for the zDataReference are restricted to the IDs of already defined DataGenerator objects.

An example using the zDataReference is given in Listing 2.49 on page 56.

### 2.4.7.3 DataSet

The DataSet class holds definitions of data to be used in the Report class (Figure 2.38). Data sets are labeled references to instances of the DataGenerator class.
Table 2.26 shows all attributes and sub-elements for the dataSet element as defined by the SED-ML Level 1 Version 1 XML Schema.

<table>
<thead>
<tr>
<th>attribute</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>metaid</td>
<td>page 23</td>
</tr>
<tr>
<td>id</td>
<td>page 22</td>
</tr>
<tr>
<td>name*</td>
<td>page 22</td>
</tr>
<tr>
<td>dataReference</td>
<td>page 57</td>
</tr>
<tr>
<td>label</td>
<td>page 57</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-elements</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>notes*</td>
<td>page 23</td>
</tr>
<tr>
<td>annotation*</td>
<td>page 24</td>
</tr>
</tbody>
</table>

Table 2.26: Attributes and nested elements for dataSet. xy* denotes optional elements and attributes.

2.4.7.3.1 label

Each data set in a Report does have to carry an unambiguous label. A label is a human readable descriptor of a data set for use in a report. For example, for a tabular data set of time series results, the label could be the column heading.

2.4.7.3.2 dataReference

The dataReference attribute contains the ID of a dataGenerator element and as such represents a link to it. The data produced by that particular data generator fills the according data set in the report.
List of Data Sets

Listing 2.50 shows the use of the `dataSet` element in a SED-ML file as defined by the SED-ML Level 1 Version 1 XML Schema.

```xml
<listOfDataSets>
  <dataSet id="d1" name="v1 over time" dataReference="dg1" label="_1">
  </listOfDataSets>

---

Listing 2.50: The SED-ML `dataSet` element, defining a data set containing the result of the referenced task
3. 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.

Thanks to David Nickerson for feedback and help with Example C2.

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 A.1 shows the complete UML diagram of the SED-ML. It gives the full picture of all implemented classes (see the XML Schema definition on page 61).

Figure A.1: The SED-ML UML class diagram
Listing B.1 shows the full SED-ML XML Schema. The code is commented inline.

```xml
<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">
    <xs:import namespace="http://www.w3.org/1998/Math/MathML"
        schemaLocation="sbml-mathml.xsd" />

    <xs:simpleType name="SId">
        <xs:annotation>
            <xs:documentation>
                The type SId is used throughout SED-ML as the type of the 'id' attributes on model elements.
            </xs:documentation>
        </xs:annotation>
        <xs:restriction base="xs:string">
            <xs:pattern value="(_|[a-z]|[A-Z])(_|[a-z]|[A-Z]|[0-9])*" />
        </xs:restriction>
    </xs:simpleType>

    <!-- attribute group for elements with ID/name att -->
    <xs:attributeGroup name="idGroup">
        <xs:attribute name="id" use="required" type="SId"></xs:attribute>
        <xs:attribute name="name" use="optional" type="xs:string"></xs:attribute>
    </xs:attributeGroup>

    <!-- SED Base class -->
    <xs:complexType name="SEDBase">
        <xs:annotation>
            <xs:documentation xml:lang="en">
                The SEDBase type is the base type of all main types in SED-ML. It serves as a container for the annotation of any part of the experiment description.
            </xs:documentation>
        </xs:annotation>
        <xs:sequence>
            <xs:element ref="notes" minOccurs="0" />
            <xs:element ref="annotation" minOccurs="0" />
        </xs:sequence>
        <xs:attribute name="metaid" type="xs:ID" use="optional" />
    </xs:complexType>

    <!-- notes and annotations -->
    <xs:element name="notes">
        <xs:complexType name="SedML">
            <xs:complexType>
                <xs:extension base="SEDBase">
                    <xs:sequence>
                        <xs:element ref="listOfSimulations" minOccurs="0" />
                        <xs:element ref="listOfModels" minOccurs="0" />
                        <xs:element ref="listOfTasks" minOccurs="0" />
                        <xs:element ref="listOfDataGenerators" minOccurs="0" />
                        <xs:element ref="listOfOutputs" minOccurs="0" />
                    </xs:sequence>
                    <xs:attribute name="level" type="xs:decimal" use="required" fixed="1" />
                    <xs:attribute name="version" type="xs:decimal" use="required" fixed="1" />
                </xs:extension>
            </xs:complexType>
        </xs:element>
    </xs:element>
</xs:complexType>
```

B. XML Schema
<xs:complexType>
  <xs:sequence>
    <xs:element name="annotation">
      <xs:complexType>
        <xs:sequence>
          <xs:any namespace="http://www.w3.org/1999/xhtml" processContents="skip" minOccurs="0" maxOccurs="unbounded" />
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
</xs:complexType>

<xs:element name="variable">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <!-- at least one of taskReference or modelReference must be set -->
        <xs:attribute name="taskReference" type="SId" use="optional" />
        <xs:attribute name="modelReference" type="SId" use="optional" />
        <xs:attribute name="target" type="xs:token" use="optional" />
        <xs:attribute name="symbol" type="xs:string" use="optional" />
        <xs:attributeGroup ref="idGroup" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>

<xs:element name="parameter">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:attributeGroup ref="idGroup" />
        <xs:attribute name="value" type="xs:double" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>

<xs:element name="algorithm">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:attribute name="kisaoID" type="KisaoType" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>

<xs:element name="uniformTimeCourse">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="algorithm" />
        </xs:sequence>
        <xs:attributeGroup ref="idGroup" />
        <xs:attribute name="outputStartTime" type="xs:double" use="required" />
        <xs:attribute name="outputEndTime" type="xs:double" use="required" />
        <xs:attribute name="numberOfPoints" type="xs:integer" use="required" />
        <xs:attribute name="initialTime" type="xs:double" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>

<xs:element name="task">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:attributeGroup ref="idGroup" />
        <xs:attribute name="simulationReference" type="SId" use="required" />
        <xs:attribute name="modelReference" type="SId" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:sequence>
  <xs:element ref="changeAttribute" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="changeXML" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="addXML" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="removeXML" minOccurs="0" maxOccurs="unbounded" />
  <xs:element ref="computeChange" minOccurs="0" maxOccurs="unbounded" />
</xs:sequence>
</xs:extension>
</xs:complexType>
</xs:element>
<xs:element name="newXML">
  <xs:complexType>
    <xs:sequence>
      <xs:any processContents="skip" minOccurs="1" maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="changeAttribute">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:attribute name="target" use="required" type="xs:token" />
        <xs:attribute name="newValue" type="xs:string" use="required" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="changeXML">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="newXML" />
        </xs:sequence>
        <xs:attribute name="target" use="required" type="xs:token" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="addXML">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="newXML" />
        </xs:sequence>
        <xs:attribute name="target" use="required" type="xs:token" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="removeXML">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:attribute name="target" use="required" type="xs:token" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="computeChange">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="listOfVariables" minOccurs="0" />
          <xs:element ref="listOfParameters" minOccurs="0" />
          <xs:element ref="math" />
        </xs:sequence>
        <xs:attribute name="target" use="required" type="xs:token" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
<xs:element name="dataGenerator">
  <xs:complexType>
    <xs:complexContent>
      <xs:extension base="SEDBase">
        <xs:sequence>
          <xs:element ref="listOfVariables" minOccurs="0" />
          <xs:element ref="listOfParameters" minOccurs="0" />
          <xs:element ref="math" />
        </xs:sequence>
        <xs:attribute name="target" use="required" type="xs:token" />
      </xs:extension>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
Listing B.1: The SED-ML XML Schema definition
C. Examples

This appendix present a few examples SED-ML uses. These examples are only illustrative and do not intend to demonstrate the full capabilities of SED-ML. Please read the specification for a for a more comprehensive view (Chapter 2).

The current examples make use of models encoded in SBML and CellML. It is important to remember that SED-ML is not restricted to those formats, but can be used with models encoded in many formats, so long as they are serialized in XML. A list of formats known to have been used, at least tentatively, with SED-ML is available on http://sed-ml.org/.
C.1 Le Loup Model (SBML)

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).

This model is referenced by its SED-ML ID model1 and refers to the model with the MIRIAM URN urn:miriam:biomodels.db:BIOMD0000000021. Software applications interpreting this example know how to dereference this URN and access the model in BioModels Database [Le Novère et al., 2006].

A second model is defined in l. 11 of the example, using model1 as a source and applying even further changes to it, in this case updating two model parameters.

One simulation setup is defined in the listOfSimulations. It is a uniformTimeCourse over 380 time units, providing 1000 output points. The algorithm used is the CVODE solver, as denoted by the KiSAO ID KISAO:0000019.

A number of dataGenerators are defined in ll. 23-62. Those are the prerequisite for defining the outputs of the simulation. The first dataGenerator named time collects the simulation time. tim1 in l. 31 maps on the Mt entity in the model that is used in task1 which here is the model with ID model1. The dataGenerator named per-tim1 in l. 39 maps on the Cn entity in model1. Finally the fourth and fifth 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 three 2D plots. The first plot has two different curves (ll. 65-70) and provides the time course of the simulation using the tim mRNA concentrations from both simulation experiments. The second plot shows the per-tim concentration against the tim concentration for the oscillating model. And the third plot shows the same plot for the chaotic model. The resulting three plots are shown in Figure C.1.

Figure C.1: The simulation result gained from the simulation description given in listing C.1

```xml
<?xml version="1.0" encoding="utf-8"?>
<!-- Written by libSedML v1.1.4992.22172 see http://libsedml.sf.net -->
ENDORSED
<sedML xmlns="http://www.biomodels.net/sed-ml">
  <listOfSimulations>
    <uniformTimeCourse id="simulation1" initialTime="0" outputStartTime="0" outputEndTime="380" numberOfPoints="1000">
      <algorithm kisaoID="KISAO:0000019"/>
    </uniformTimeCourse>
  </listOfSimulations>
</sedML>
```
Listing C.1: LeLoup Model Simulation Description in SED-ML
C.2 Le Loup Model (CellML)

The following example provides a SED-ML description for the simulation of the model based on the publication by Leloup, Gonze and Goldbeter “Limit Cycle Models for Circadian Rhythms Based on Transcriptional Regulation in Drosophila and Neurospora” (PubMed ID: 10643740). Whereas the previous example used SBML to encode the simulation experiment, here the model 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/7606a47e222bc3b3d9117baa08d2e7246d67eedd/leloup_gonze_goldbeter_1999_a.cellml, ll. 14).

A second model is defined in l. 15 of the example, using model1 as a source and applying even further changes to it, in this case updating two model parameters.

One simulation setup is defined in the listOfSimulations. It is a uniformTimeCourse over 380 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. 27-73. Those are the prerequisite for defining the output of the simulation. The dataGenerator named tim1 in l. 37 maps on the Mt entity in the model that is used in task1, which here is the model with ID model1. The dataGenerator named per-tim in l. 46 maps on the CN entity in model1. Finally the fourth and fifth 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 three 2D plots (ll. 74-91). They reproduce the same output as the previous example.

![Tim mRNA with Oscillation and Chaos](image1)

![Tim mRNA limit cycle (oscillations)](image2)

![Tim mRNA limit cycle (chaos)](image3)

Figure C.2: The simulation result gained from the simulation description given in listing C.2

```xml
<?xml version="1.0" encoding="utf-8"?>
<sedML xmlns="http://sed-ml.org/"
xmlns:math="http://www.w3.org/1998/Math/MathML" level="1" version="1">
<notes><p xmlns="http://www.w3.org/1999/xhtml">Comparing Limit Cycles and strange attractors for oscillation in Drosophila</p></notes>
<listOfSimulations>
  <uniformTimeCourse id="simulation1" algorithm="KiSAO:0000019"
initialTime="0" outputStartTime="0" outputEndTime="380"
numberOfPoints="1000" />
  <algorithm kisaoId="KiSAO:0000019"/>
</uniformTimeCourse>
</sedML>
```
<listOfModels>
  <model id="model2" name="Circadian Chaos" language="urn:sedml:language:cellml" source="model1">
    <listOfChanges>
      <changeAttribute target="/cellml:model/cellml:component[@name='MT']/cellml:variable[@name='vmT']/@initial_value" newValue="0.28"/>
      <changeAttribute target="/cellml:model/cellml:component[@name='T2']/cellml:variable[@name='vdT']/@initial_value" newValue="4.8"/>
    </listOfChanges>
  </model>
</listOfModels>

<listOfTasks>
  <task id="task1" name="Limit Cycle" modelReference="model1" simulationReference="simulation1"/>
  <task id="task2" name="Strange attractors" modelReference="model2" simulationReference="simulation1"/>
</listOfTasks>

<listOfDataGenerators>
  <dataGenerator id="time" name="time">
    <listOfVariables>
      <variable id="t" taskReference="task1" target="/cellml:model/cellml:component[@name='environment']/cellml:variable[@name='time']"/>
    </listOfVariables>
    <math:math>
      \text{t}
    </math:math>
  </dataGenerator>
  <dataGenerator id="tim1" name="tim mRNA">
    <listOfVariables>
      <variable id="v0" taskReference="task1" target="/cellml:model/cellml:component[@name='MT']/cellml:variable[@name='MT']"/>
    </listOfVariables>
    <math:math>
      \text{v0}
    </math:math>
  </dataGenerator>
  <dataGenerator id="per_tim" name="nuclear PER -TIM complex">
    <listOfVariables>
      <variable id="v1" taskReference="task1" target="/cellml:model/cellml:component[@name='CN']/cellml:variable[@name='CN']"/>
    </listOfVariables>
    <math:math>
      \text{v1}
    </math:math>
  </dataGenerator>
  <dataGenerator id="tim2" name="tim mRNA (changed parameters)"/>
    <listOfVariables>
      <variable id="v2" taskReference="task2" target="/cellml:model/cellml:component[@name='MT']/cellml:variable[@name='MT']"/>
    </listOfVariables>
    <math:math>
      \text{v2}
    </math:math>
  </dataGenerator>
  <dataGenerator id="per_tim2" name="nuclear PER -TIM complex">
    <listOfVariables>
      <variable id="v3" taskReference="task2" target="/cellml:model/cellml:component[@name='CN']/cellml:variable[@name='CN']"/>
    </listOfVariables>
    <math:math>
      \text{v3}
    </math:math>
  </dataGenerator>
</listOfDataGenerators>

<listOfOutputs>
  <plot2D id="plot1" name="tim mRNA with Oscillation and Chaos">
    <listOfCurves>
      <curve id="c1" logX="false" logY="false" xDataReference="time" yDataReference="tim1"/>
      <curve id="c2" logX="false" logY="false" xDataReference="time" yDataReference="tim2"/>
    </listOfCurves>
  </plot2D>
  <plot2D id="plot2" name="tim mRNA limit cycle (Oscillation)">
    <listOfCurves>
      <curve id="c3" logX="false" logY="false" xDataReference="per_tim" yDataReference="tim1"/>
    </listOfCurves>
  </plot2D>
  <plot2D id="plot3" name="tim mRNA limit cycle (chaos)">
    <listOfCurves>
      <curve id="c4" logX="false" logY="false" xDataReference="per_tim2" yDataReference="tim2"/>
    </listOfCurves>
  </plot2D>
</listOfOutputs>
Listing C.2: *LeLoup Model Simulation Description in SED-ML*
The following example provides a SED-ML description for the simulation of the IkappaB-NF-kappaB signaling module based on the publication by Hoffmann, Levchenko, Scott and Baltimore “The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation.” (PubMed ID: 12424381)

This model is referenced by its SED-ML ID model1 and refers to the model with the MIRIAM URN urn:miriam:biomodels.db:BIOMD0000000140. Software applications interpreting this example know how to dereference this URN and access the model in BioModels Database [Le Novère et al., 2006].

The simulation description specifies one simulation simulation1, which is a uniform timecourse simulation that simulates the model for 41 hours. task1 then applies this simulation to the model.

As output this simulation description collects four parameters: Total_NFkBn, Total_IkBbeta, Total_IkBeps and Total_IkBalpha. These variables are to be plotted against the simulation time and displayed in four separate plots, as shown in Figure C.3.

Figure C.3: The simulation result gained from the simulation description given in listing C.3

The SED-ML description of the simulation experiment is given in listing C.3.
Listing C.3: IkappaB-NF-kappaB signaling Model Simulation Description in SED-ML
D. SED-ML archive

A SED-ML archive is a self-contained repository of all the resources necessary to run a simulation and display its output. It is a convenient alternative if a model source URI cannot be resolved, or if an end-user is offline.

A SED-ML archive is a zipped folder containing one SED-ML file and any number of model files. By convention, the name of the archive will be the name of the SED-ML file contained in the archive, with the suffix “.sedx”. Each model file contained in the archive is referred to by a relative URI in the source attribute of the SED-ML document’s model element.

For example, the contents of an archive, when unzipped, may be as follows:

Name of archive: Mysedml.sedx
Contents:

- Mysedml.xml
- model1.xml
- model2.xml

Listing D.1 shows how a model would be referenced in the SED-ML file Mysedml.xml in the above example.

Future versions of SED-ML may expand the permitted contents of an archive to include experimental data files or other resources.

```
1 <listOfModels>
2 <model id="m0001" language="urn:sedml:language:sbml"
3    source="model1.xml">
4    <listOfChanges>
5      <change>
6      [MODEL PRE-PROCESSING]
7      </change>
8    </listOfChanges>
9  </model>
10 </listOfModels>
```

Listing D.1: Usage of relative URIs to reference a model in a SED-ML archive

While the SED-ML archive is not normative it is supported in currently available implementations of SED-ML such as LibSedML (http://libsedml.sf.net) and jlibsedml (http://sf.net/projects/jlibsedml).
Bibliography


