



SED-ML as BioUML workflow

Ivan Yevshin, Ilya Kiselev,
Fedor Kolpakov

Institute of Systems Biology, Ltd.,
Design Technological Institute of Digital Techniques, SBRAS
Novosibirsk, Russia

COMBINE 2015, October 12-16, Salt Lake City, USA

What is workflow

Workflow, as a concept, was defined in the business domain in 1996 by the Workflow Management Coalition as:

The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules.

Scientific workflow - “useful paradigm to describe, manage, and share complex scientific analyses”. They provide a high-level declarative way of specifying what a particular in silico experiment modelled by a workflow is set to achieve, not how it will be executed (Taverna web page).

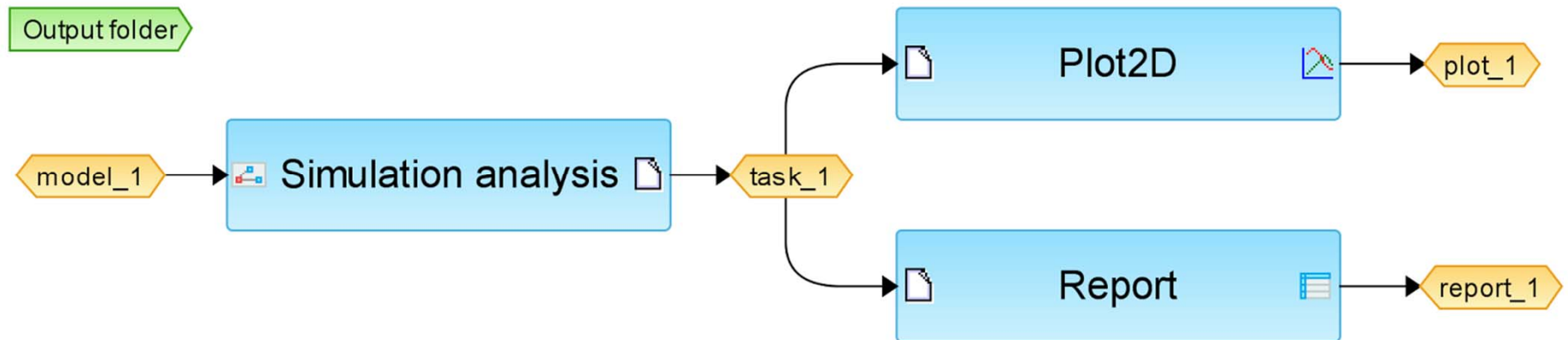
A workflow is a series of tools and dataset actions that run in sequence as a batch operation. (Galaxy web page)

SED-ML - is an XML-based format for encoding simulation setups, to ensure exchangeability and reproducibility of simulation experiments.

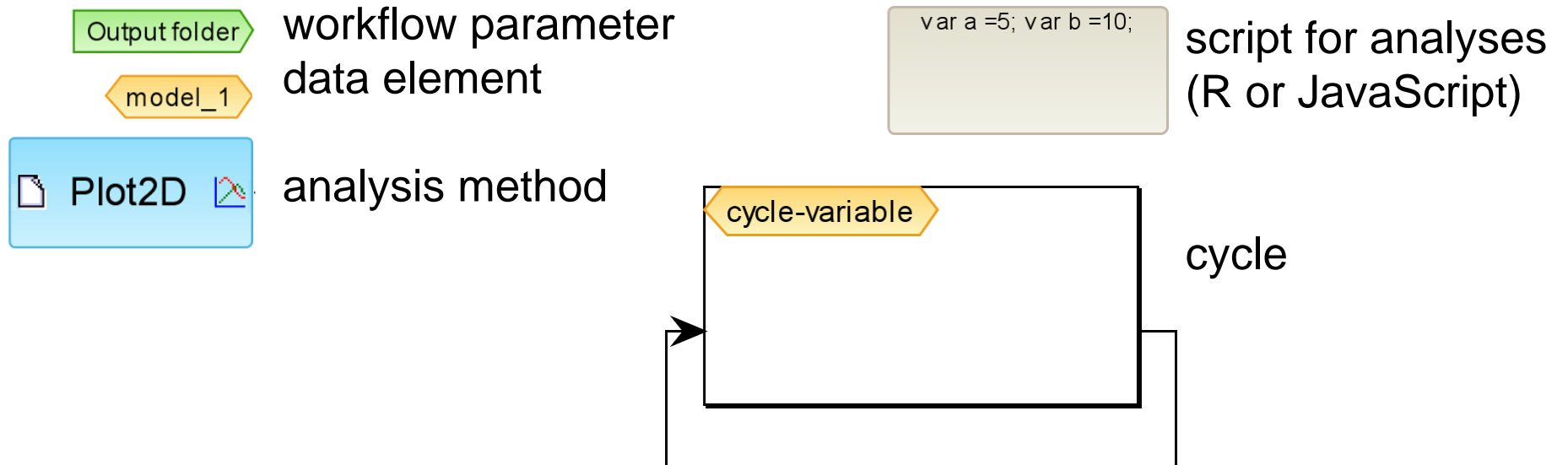
SED-ML is particular case of format for scientific workflow.

BioUML workflows

Example:



Legend:



Workflow management system

A **workflow management system** provides an infrastructure for the set-up, performance and monitoring of a defined sequence of tasks, arranged as a workflow.

A **scientific workflow system** is a specialized form of a workflow management system designed specifically to **compose and execute a series of computational or data manipulation steps**, or workflow, in a scientific application.

Examples:

- **Galaxy** – mainly used for NGS data analysis,
- **Apache Taverna** - widely used in bioinformatics,
- **Kepler**,
- **KNIME**.

BioUML platform also provides powerful workflow management system used for analysis of wide range of biological data.

BioUML IDE as Workflow Management System

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

Repository Tree

- Databases
- Data
- Analyses
 - Admin
 - BSA
 - Binding regions
 - ChIPMunk
 - Creasemarks
 - DAE models
 - DAS
 - Data
 - GTRD
 - GeneModels
 - Import
 - Molecular networks
 - Optimization
 - RiboSeq
 - RiboseqArticleResult
 - Simulation

sa : Simple workflow

```
graph TD; OF[Output folder] --> M1[model_1]; M1 --> SA[Simulation analysis]; SA --> T1[task_1]; T1 --> R[Report]; R --> P2D[Plot2D]; R --> R1[report_1]; P2D --> P1[plot_1];
```

YAML

```
1 name: Simple workf
2 description: '
3 parameters:
4   Output folder: {
5 expressions:
6   model_1: {expres
7   plot_1: {type: I
8   report_1: {type:
9   task_1: {express
10 tasks:
11   Plot2D:
12     in: $task_1$
```

Element Info

Default

ID: [Simulation analysis](#)

Title: Simulation analysis

Size: 2

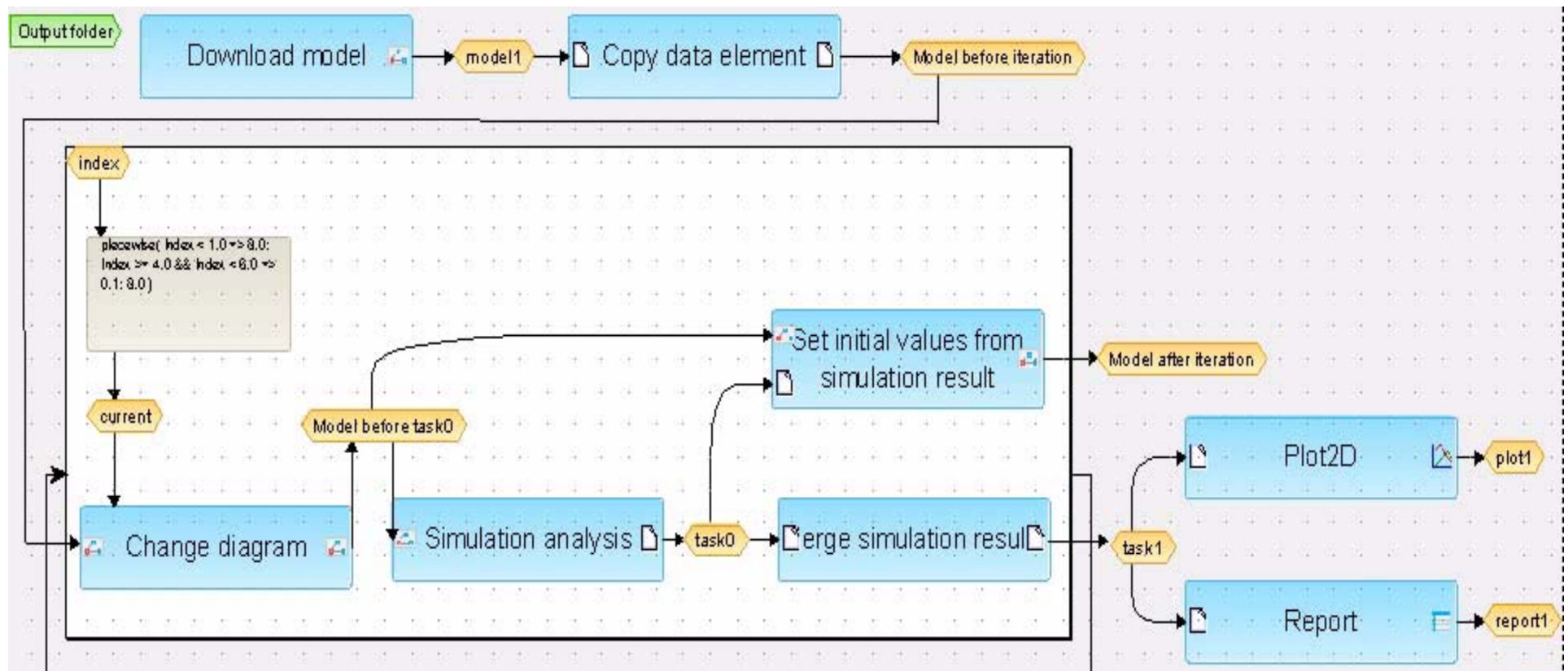
Complete name:
data/Collaboration/sa/Data/Diag
workflow/Simulation analysis

Property	Value
Simulation analysis	
Model	data/Collaboration/sa/Data/Diagrams/00002-sbml-l3v1
Simulation engine name	ODE Simulation Engine
Simulation engine	
Initial time	0
Completion time	5
Time increment	0.098
Simulator name	JVode
Simulator options	

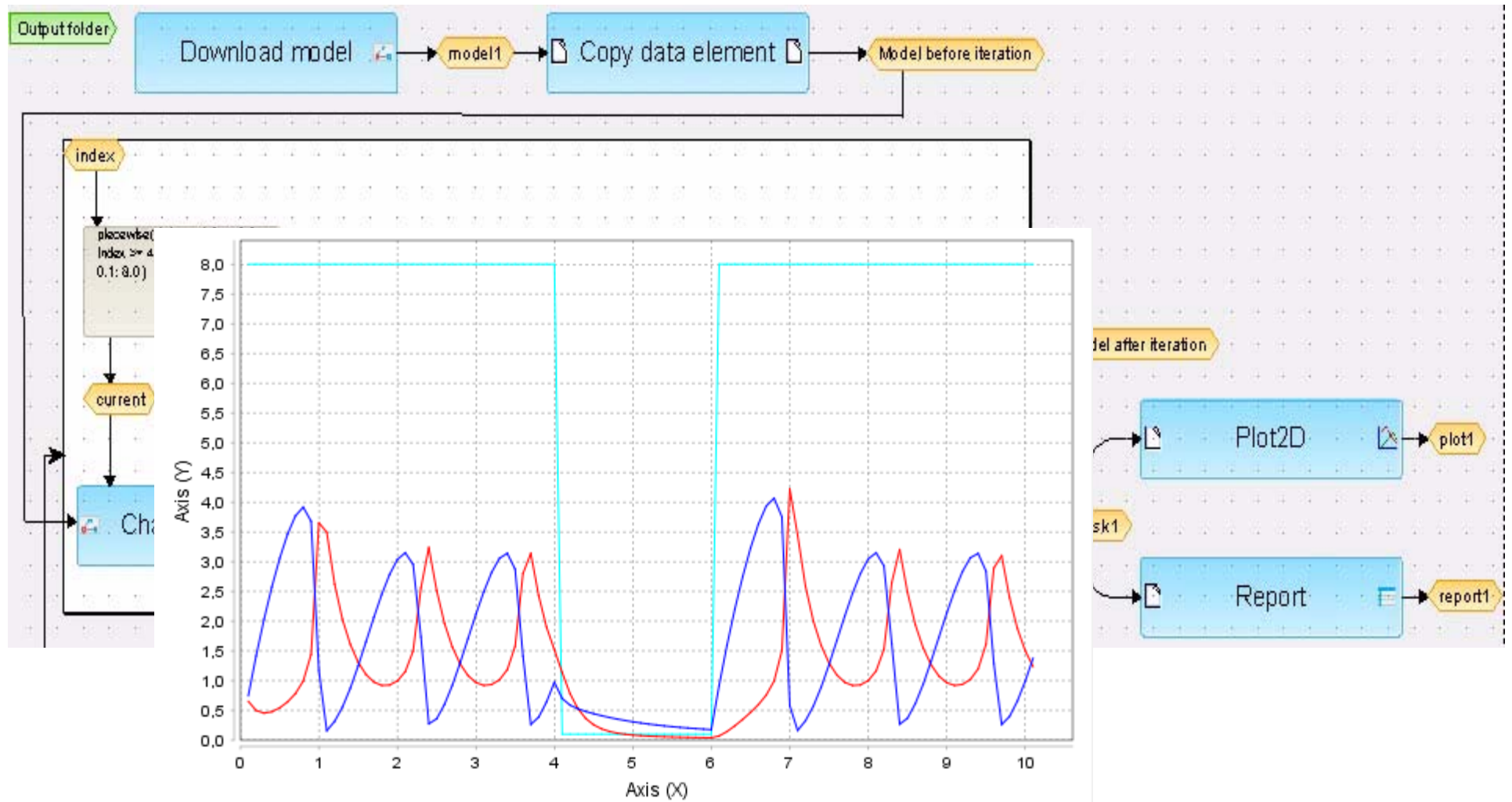
SED-ML Analyses

BioUML analysis	SED-ML element
Simulation	Task
Change diagram	listOfChanges
Report	Report, DataGenerator
Plot2d	Plot2D, DataGenerator
Merge simulation results	Repeated Task
Apply simulation results	
Download model	Model

SED-ML Repeated Task example



SED-ML Repeated Task example



YAML - YAML Ain't Markup Language

Clark Evans, 2001

YAML is a human friendly data serialization standard for all programming languages.

Data structure hierarchy is maintained by outline indentation.

(www.yaml.org)

```
receipt:      Oz-Ware Purchase Invoice
date:         2012-08-06
customer:
  given:      Dorothy
  family:     Gale

items:
  - part_no:   A4786
    descrip:   Water Bucket (Filled)
    price:     1.47
    quantity:  4

  - part_no:   E1628
    descrip:   High Heeled "Ruby"
    size:      8
    price:     100.27
    quantity:  1

bill-to:  &id001
street:  |
        123 Tornado Alley
        Suite 16
city:    East Centerville
state:   KS
```

YAML

29 lines

name: Simple workflow

description: ''

parameters:

Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:

model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-l3v1}

plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}

report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}

task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:

Plot2D:

in: \$task_1\$

out: \$plot_1\$

param:

Curves:

- {Title: c_S1_1, expressionX: time, expressionY: \$compartment.S1/\$compartment}

- {Title: c_S2_1, expressionX: time, expressionY: \$compartment.S2/\$compartment}

Report:

in: \$task_1\$

out: \$report_1\$

param:

Columns:

- {Name: Time, Expression: time}

- {Name: S1, Expression: \$compartment.S1/\$compartment}

- {Name: S2, Expression: \$compartment.S2/\$compartment}

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

SED-ML

66 lines

```
<?xml version="1.0" encoding="UTF-8"?>
<sedML xmlns="http://sed-ml.org/" xmlns:math="http://www.w3.org/1998/Math/MathML"
version="1" level="1">
  <listOfSimulations>
    <uniformTimeCourse id="simulation_1" initialTime="0.0" outputStartTime="0.0"
outputEndTime="5.0" numberOfPoints="51">
      <algorithm kisaID="KISAO:0000019" />
    </uniformTimeCourse>
  </listOfSimulations>
  <listOfModels>
    <model id="model_1" language="urn:sedml:language:sbml" source="00002-sbml-l3v1.xml" />
  </listOfModels>
  <listOfTasks>
    <task id="task_1" modelReference="model_1" simulationReference="simulation_1" />
  </listOfTasks>
  <listOfDataGenerators>
    <dataGenerator id="data_generator_1" name="data_generator_1">
      <listOfVariables>
        <variable id="time" name="time" taskReference="task_1" symbol="urn:sedml:symbol:time" />
      </listOfVariables>
      <math:math>
        <math:ci>time</math:ci>
      </math:math>
    </dataGenerator>
```

...

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```

graph LR
    OF[Output folder] --> M1[model_1]
    M1 --> SA[Simulation analysis]
    SA --> T1[task_1]
    T1 --> R[Report]
    T1 --> P2D[Plot2D]
    R --> R1[report_1]
    P2D --> P1[plot_1]
  
```

parameters:
 Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:
 model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}
 plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}
 report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}
 task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:
 Plot2D:
 in: \$task_1\$
 out: \$plot_1\$
 param:
 Curves:
 - {Title: c_S1_1, expressionX: time, expressionY: \$compartment.S1/\$compartment}
 - {Title: c_S2_1, expressionX: time, expressionY: \$compartment.S2/\$compartment}

Report:
 in: \$task_1\$
 out: \$report_1\$
 param:
 Columns:
 - {Name: Time, Expression: time}
 - {Name: S1, Expression: \$compartment.S1/\$compartment}
 - {Name: S2, Expression: \$compartment.S2/\$compartment}

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

Applica... Clipboard Search ... Script Search ... Tasks SQL edi... Descrip... Refere... History Workflow YAML % Layout FBC table

View parts selection Property Inspector

3:12 INS

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```

graph LR
    OF[Output folder] --> M1[model_1]
    M1 --> SA[Simulation analysis]
    SA --> T1[task_1]
    T1 --> R[Report]
    T1 --> P2D[Plot2D]
    R --> R1[report_1]
    P2D --> P1[plot_1]
  
```

parameters:
 Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:
 model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}
 plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}
 report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}
 task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:

Plot2D:

```

in: $task_1$
out: $plot_1$
param:
  Curves:
    - {Title: c_S1_1, expressionX: time, expressionY: $compartment.S1/$compartment}
    - {Title: c_S2_1, expressionX: time, expressionY: $compartment.S2/$compartment}
  
```

Report:

```

in: $task_1$
out: $report_1$
param:
  Columns:
    - {Name: Time, Expression: time}
    - {Name: S1, Expression: $compartment.S1/$compartment}
    - {Name: S2, Expression: $compartment.S2/$compartment}
  
```

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

3:12 INS

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```

graph LR
    OF[Output folder] --> M1[model_1]
    M1 --> SA[Simulation analysis]
    SA --> T1[task_1]
    T1 --> R[Report]
    T1 --> P2D[Plot2D]
    R --> R1[report_1]
    P2D --> P1[plot_1]
  
```

parameters:

Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:

model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}

plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}

report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}

task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:

Plot2D:

in: \$task_1\$

out: \$plot_1\$

param:

Curves:

- {Title: c_S1_1, expressionX: time, expressionY: \$compartment.S1/\$compartment}
- {Title: c_S2_1, expressionX: time, expressionY: \$compartment.S2/\$compartment}

Report:

in: \$task_1\$

out: \$report_1\$

param:

Columns:

- {Name: Time, Expression: time}
- {Name: S1, Expression: \$compartment.S1/\$compartment}
- {Name: S2, Expression: \$compartment.S2/\$compartment}

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

3:12 INS

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```

graph LR
    OF[Output folder] --> M1[model_1]
    M1 --> SA[Simulation analysis]
    SA --> T1[task_1]
    T1 --> R[Report]
    T1 --> P2D[Plot2D]
    R --> R1[report_1]
    P2D --> P1[plot_1]
  
```

parameters:
 Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:
 model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}
 plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}
 report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}
 task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:
 Plot2D:
 in: \$task_1\$
 out: \$plot_1\$
 param:
 Curves:
 - {Title: c_S1_1, expressionX: time, expressionY: \$compartment.S1/\$compartment}
 - {Title: c_S2_1, expressionX: time, expressionY: \$compartment.S2/\$compartment}

Report:
 in: \$task_1\$
 out: \$report_1\$
 param:
 Columns:
 - {Name: Time, Expression: time}
 - {Name: S1, Expression: \$compartment.S1/\$compartment}
 - {Name: S2, Expression: \$compartment.S2/\$compartment}

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

3:12 INS

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```
graph LR; OF[Output folder] --> M1[model_1]; M1 --> SA[Simulation analysis]; SA --> T1[task_1]; T1 --> R[Report]; T1 --> P2D[Plot2D]; R --> R1[report_1]; P2D --> P1[plot_1];
```

parameters:
Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:
model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}
plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}
report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}
task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:

Plot2D:

- in: \$task_1\$
- out: \$plot_1\$
- param:
 - Curves:
 - {Title: c_S1_1, expressionX: time, expressionY: \$compartment.S1/\$compartment}
 - {Title: c_S2_1, expressionX: time, expressionY: \$compartment.S2/\$compartment}

Report:

- in: \$task_1\$
- out: \$report_1\$
- param:
 - Columns:
 - {Name: Time, Expression: time}
 - {Name: S1, Expression: \$compartment.S1/\$compartment}
 - {Name: S2, Expression: \$compartment.S2/\$compartment}

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

3:12 INS

BioUML 201411181905

Файл Правка Вид Переход Источник Databases Data Реорганизация кода Выполнить Отладка Группа Сервис Окно Справка Поиск (Ctrl+I)

sa : Simple workflow

```

graph LR
    OF[Output folder] --> M1[model_1]
    M1 --> SA[Simulation analysis]
    SA --> R[Report]
    SA --> P2D[Plot2D]
    R --> R1[report_1]
    P2D --> P1[plot_1]
  
```

parameters:
 Output folder: {role: Output, elementType: Folder, defaultValue: data/Collaboration/sa/Data/Diagrams/Simple workflow results}

expressions:
 model_1: {expression: data/Collaboration/sa/Data/Diagrams/00002-sbml-13v1}
 plot_1: {type: Data element (auto-open), expression: \$Output folder\$/plot_1}
 report_1: {type: Data element (auto-open), expression: \$Output folder\$/report_1}
 task_1: {expression: \$Output folder\$/Simulation results/task_1}

tasks:

Plot2D:

```

in: $task_1$
out: $plot_1$
param:
  Curves:
    - {Title: c_S1_1, expressionX: time, expressionY: $compartment.S1/$compartment}
    - {Title: c_S2_1, expressionX: time, expressionY: $compartment.S2/$compartment}
  
```

Report:

```

in: $task_1$
out: $report_1$
param:
  Columns:
    - {Name: Time, Expression: time}
    - {Name: S1, Expression: $compartment.S1/$compartment}
    - {Name: S2, Expression: $compartment.S2/$compartment}
  
```

Simulation analysis: {in: \$model_1\$, out: \$task_1\$}

Applica... Clipboard Search ... Script Search ... Tasks SQL edi... Descrip... Refere... History Workflow YAML % Layout FBC table

Repository Tree

Element Info

View parts selection

Property Inspector

3:12 INS

Summary

- ✓ Sed-ML l1 v2 import\export
- ✓ Visual + text sync editing
- ✓ Combination with other BioUML analysis and scripts (model automatic generation, parameter fitting, simulation result analysis)

Summary

- ✓ Sed-ML l1 v2 import\export
- ✓ Visual + text sync editing
- ✓ Combination with other BioUML analysis and scripts (model automatic generation, parameter fitting, simulation result analysis)
- ? Workflows seems to be more complicated than corresponding SED-ML
- ? Not every workflow can be exported to SED-ML.

Towards SED-ML 2

Motivation:

- standards become more and more complicated, e.g. no tool supports all SBML packages;
- complicated models, requiring multiple simulation types;
- we need to use different types of tasks (e.g. parameter fitting).

Towards SED-ML 2

Motivation:

- standards become more and more complicated, e.g. no tool supports all SBML packages;
- complicated models, requiring multiple simulation types;
- we need to use different types of tasks (e.g. parameter fitting).

Solution:

- possibility to add new tasks or analysis methods;
- possibility to explicitly specify software tool and its parameters for each task.
- possible frameworks to call different tools:
 - SBW – systems biology workbench;
 - command line and Galaxy core.

Thank you for your attention