Supplementary Information

This supplementary information belongs to "Food Safety Knowledge Markup Language (FSK-ML)" Software Developer Guide Version 3.1 (https://foodrisklabs.bfr.bund.de/fsk-ml-food-safety-knowledge-markup-language/).

1. FSK-SED-ML

FSK-SED-ML describes the simulation settings of an FSK-ML compliant model. FSK-SED-ML extends SED-ML, which is based on SBML, by introducing new classes and additional metadata. The main structure of the SED-ML format is maintained (`<listOfModels><listOfSimulations><listOfTasks><listOfDataGenerators><listOfOutputs>`, for further details see Bergmann, Cooper, Le Novere, Nickerson, and Waltemath (2015)). Note that this document uses the conventions defined in the OMEX specification document (see Bergmann et al. (2015) for details; also see Food Safety Knowledge Markup Language (FSK-ML) Software Developer Guide Version 3.0 (https://foodrisklabs.bfr.bund.de/fsk-ml-food-safety-knowledge-markup-language/) for a summary).

Language References

FSK-SED-ML makes use of an extra file types through the Internet media types, which is previously known as MIME type (Freed & Borenstein, 1996). In Table S1, the URIs for common scripting language is presented.

Table S1 Language references

Language	MIME type
R	https://iana.org/assignments/mediatypes/text/x-r
Python	https://iana.org/assignments/mediatypes/text/x-python
Matlab	https://iana.org/assignments/mediatypes/text/x-matlab

Class: SourceScript

SourceScript is a class that allows to reference external, language-specific scripts (see Figure S1 for the definition). The class contains the following attributes: (1) the path to the scripts (`src` in Figure S1) and (2) the corresponding languages (`language` in Figure S1).

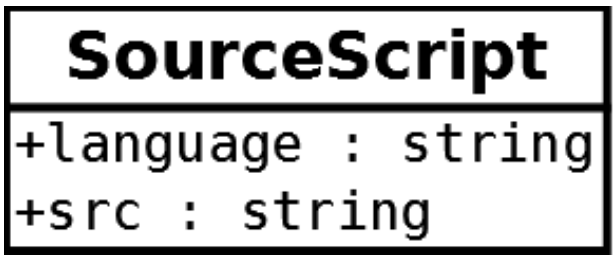

Figure S1 Definition of the SourceScript class

Attribute: `language`

The required `language`-attribute is of the data type `string`. This attribute is used to specify the scripting language of the code embedded within the **SourceScript** class or referenced through the `src`-attribute (also defined in the **SourceScript** class**)**. Common languages and their MIME types are listed in Table S1.

Attribute: `src`

For the optional source (`src`)-attribute, `anyURI` is used to indicate the path of a local script file. The path is resolved via the local file system or as a relative link. Given this attribute is provided, any free text in **SourceScript** is ignored. When the `src`-attribute is specified in a simulation script, the **Simulation** class should not include script code. If the `src`-attribute is empty, the script needs to be provided within **SourceScript** as free text.

Class: Simulation

The **Simulation** class defines the settings under which models is executed. In SED-ML the **Simulation** class acts as a container for defining the simulation experiments. Script-based models use the SED-ML simulation class **SteadyState** whose algorithm has an empty KISAO id (“”) because no algorithm is required within FSK-SED-ML. SED-ML simulation requires an algorithm id from the KISAO ontology, so called KISAO id, this id is included into FSK-SED-ML be compliant with SED-ML.

Simulations are further annotated by **FskSimulationType** and **SourceScript**.

Class: FskSimulationType

FskSimulationType is a class that allows providing metadata about the kind of simulation that will be performed (see

Figure S2 Definition of the FskSimulationType class.

for definition).

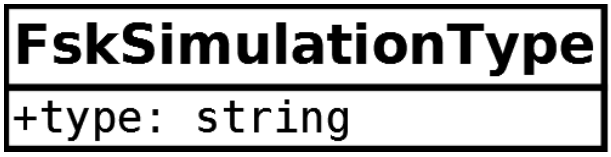

Figure S2 Definition of the FskSimulationType class.

Attribute: type

The required `type`-attribute is of the data type `string`. This attribute is used to specify the kind of simulation that will be performed. Allowed values for this attribute are `deterministic`, `statistic`, and `probabilistic` (see Table S2 for definition and Figure S3 for an example).

Table S2 Definition for the `type`-attribute values.

Value for the `type`-attribute	Definition
`deterministic`	Deterministic simulations operate on predefined values for all model input parameters.
`statistic`	Statistic simulations create descriptive analysis of observational data. These simulations are meant to describe and analyse data.
`probabilistic`	Probabilistic simulations use probabilistic methods, like Monte Carlo simulations, to create model input parameters.

```
<listOfSimulations>
  <steadyState id="Simulation1">
    <algorithm kisaoID="" />
    <annotation>
      <fskSimulationType type="deterministic" />
      <sourcescript id="simulation1"
        language="https://iana.org/assignments/mediatypes/text/x-r">
        source("param.r")
        Npos <- 20
        source("model.r")
        result1 <- result

        Npos <- 30
```

```
        source("model.r")
        result2 <- result

        Npos <- 40
        source("model.r")
        result3 <- result
      </sourcescript>
    </annotation>
  </steadyState>
</listOfSimulations>
```

Figure S3 Example for the FskSimulationType class.

Simulations for SBML-based models are defined in this class. Supported SBML-based simulations are: (1) the "Time courses" simulations (only primary and primary-secondary models are supported) and (2) "Parameter scans". Of the predefined SED-ML classes the **SteadyState** is used and expanded by annotations.

Class: Model

The **Model** class references the model file that is used for simulations (see Figure S4 for definition and Figure S5 for an example). The `language`-attribute is a mandatory attribute. See Section 0 and 0 for definitions and details.

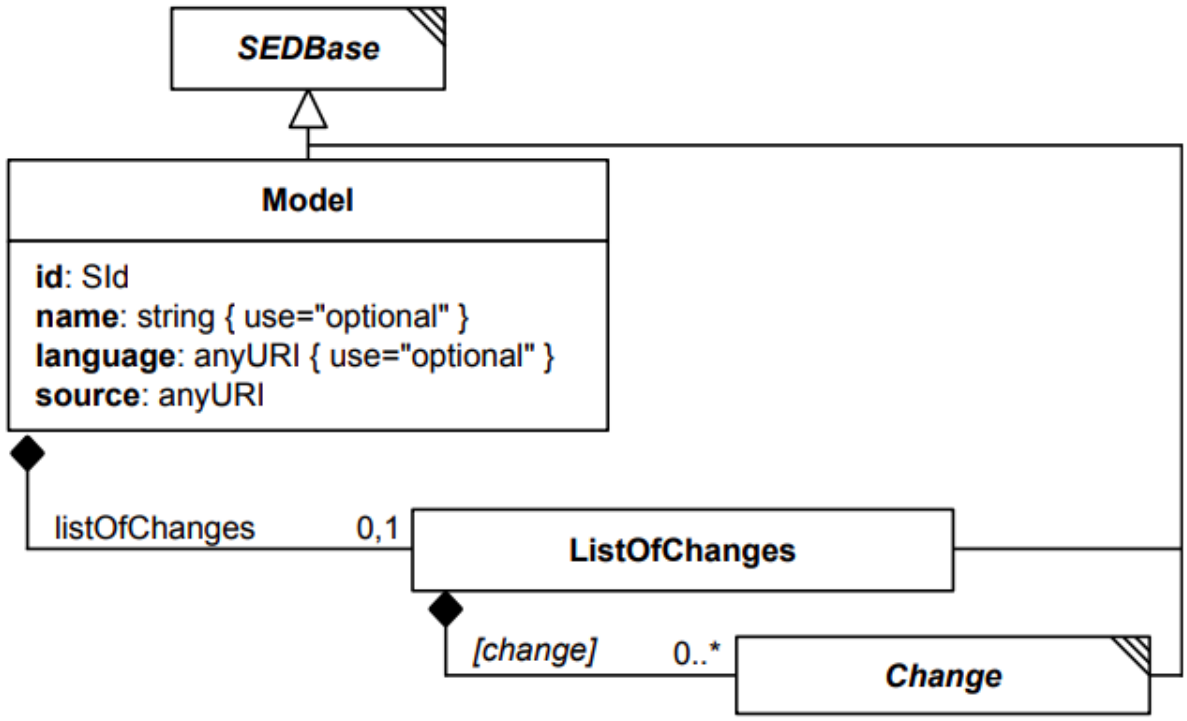

Figure S4 Definition of the Model class (the figure is taken from Bergmann et al. (2015)).

```
<listOfModels>
  <model id="model1" name="initialize_parents_flocks"
    language="https://iana.org/assignments/mediatypes/text/x-r"
    source="./model.r">
  </model>
</listOfModels>
```
Figure S5 Example for the Model class.

Support for algebraic SBML-based models is also provided. Supported classes of SBML are: (1) **Species** for target values, (2) **Parameter** for dependent and constant parameters, and (3) **AssignmentRule** for implementation of equations for target values and dependent parameters. Boundaries for the parameters can be declared by usage of the **Constraint** class. When an already existing SBML-file is used, the SED-ML **Change** class can be used to make modifications. Note that the settings to run a simulation are defined in the class **Task**.

Class: Task

The **Task** class defines each simulation scenario through the combination of the classes **Model** with **Simulation**. It defines the order in which simulation scenario are executed (see Figure S6 for an example). If no additional **SourceScript** is defined, the referenced **Simulation**, which uses the default parameter, will be executed.

Each simulation scenario, defined in a **Task,** has to refer to simulation settings defined in a **Simulation**. Nevertheless, it is possible that simulation settings defined as a script can already contain the script command that calls a specific model. In general, a **Task** specifies which simulation setting is combined with which model.

```
<listOfTasks>
    <task id="Task1" name="FlocksSimu" modelReference="model1"
         simulationReference="Simulation1">
    </task>
    <task id="Task2" name="Model_Default" modelReference="model1"
         simulationReference="Simulation_Default">
      <annotation>
        <sourcescript id="task"
          language="https://iana.org/assignments/mediatypes/text/x-r">
            source("param.r")
            source("model.r")
       </sourcescript>
      </annotation>
    </task>
</listOfTasks>
```
Figure S6 Example of the Task class.

Class: DataGenerator

The **DataGenerator** class defines which values from which simulation scenario is considered for the output. A potential reason to process simulation results is to bring values in an appropriate form for later output. It is possible to reference to scripts.

Class: Output

The **Output** class defines how the values specified in the **Task** or **DataGenerator** are plotted (see Figure S7 for an example).

```
<listOfOutputs>
  <plot2D id=”plot1“>
    <annotation>
      <sourcescript id="script1"
        language="https://iana.org/assignments/mediatypes/text/x-python"
        src="visualization-script.py" />
    </annotation>
  </plot2D>
  <plot2D id=”plot2”>
    <annotation>
      <sourcescript id="script2"
        language="https://iana.org/assignments/mediatypes/text/x-r" >
        hist(result, breaks=50, main=”PREVALENCE OF PARENTS FLOCKS”, xlab=”Prevalence”,
              col=”32”)
      </sourcescript>
    </annotation>
  </plot2D>
</listOfOutputs>
```

Figure S7 Example of the Output class with referenced and embedded script examples.

2. Example for Code of a Full FSK-SED-ML-file

Figure S8 shows an example for a FSK-SED-ML-file. The example is based on the work of Plaza Rodriguez, Correia Carreira, and Kasbohrer (2018).

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Written by Sascha Bulik,
     Guido Correia Carreia
     Miguel de-Alba-Aparicio,
     Carolina Plaza-Rodríguez,
     Matthias Filter
 -->
<sedFSKML xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://sed-ml.org/sed-ml-L1-V2.xsd"
    xmlns="http://sed-ml.org/sed-ml/level1/version2"
    level="1" version="3">

    <listOfDataDescriptions>
      <dataDescription id="dataDescription1" name="model script"
        source="./model.r" format="https://iana.org/assignments/mediatypes/text/x-r">
```

```
        <listOfDataSources>
          <dataSource id="prevalenceSource" indexSet="Prevalence" />
        </listOfDataSources>
      </dataDescription>
    </listOfDataDescriptions>

    <listOfDataGenerators>
      <dataGenerator id="dgPrevalence">
        <listOfVariables>
          <variable id="Prevalence" modelReference="model1" target="#prevalenceSource" />
        </listOfVariables>
        <math xmlns="___">
          <math:ci>Prevalence</math:ci>
      </dataGenerator>
    </listOfDataGenerators>

    <listOfModels>
        <model id="model1" name="initialize_parents_flocks">
          <annotation>
            <sourcescript id="model"
              language="https://iana.org/assignments/mediatypes/text/x-r">
                result <- 100 * rbeta(n.iter, shape1=Npos+1,
                    shape2=Ntotal-Npos+1)
            </sourcescript>
          </annotation>
        </model>
    </listOfModels>

    <listOfSimulations>
      <steadyState id="Simulation1">
        <algorithm kisaoID="" />
        <annotation>
          <fskSimulationType type="deterministic" />
          <sourceScript id="simulation1"
            language="https://iana.org/assignments/mediatypes/text/x-r">
            n.iter <- 200
            Npos <- 30
            Ntotal <- 100
          </sourcescript>
        </annotation>
      </steadyState>
    </listOfSimulations>
    <listOfTasks>
        <task id="Task1" name="FlocksSimu" modelReference="model1"
            simulationReference="Simulation1">
        </task>
    </listOfTasks>

    <listOfOutputs>
      <plot2D id="plot1">
        <annotation>
          <sourceScript id="visualization1"
            language="https://iana.org/assignments/mediatypes/text/x-r">
              hist(result, breaks=50,
                main="",
                xlab="Prevalence", col="32")
          </sourceScript>
        </annotation>
      </plot2D>
    </listOfOutputs>
</sedFSKML>
```

Figure S8 Example for a full FSK-SED-ML-document.

3. Controlled Vocabularies

Table S3 lists all controlled vocabularies from https://docs.google.com/spreadsheets/d/1C6N4-YWX9OMmNStd2rYlSUaVys-aiJGLj00cD44aVc8/edit#gid=1479548673 (effective 7th November 2019).

Table S3 Lists all controlled vocabularies from https://docs.google.com/spreadsheets/d/1C6N4-YWX9OMmNStd2rYlSUaVys-aiJGLj00cD44aVc8/edit#gid=1479548673 (effective 7th November 2019).

Source	RAKIP project
Rights	Creative Commons
Availability	Knowledge Junction (EFSA)/Zenodo
Format	RAKIP project
Publication Type	RIS format specifications
Publication Status	Bibliographic Ontology Specification
Language	SSD
Software	RAKIP project
Language written in	RAKIP project
Model Class	RAKIP project
Model Sub-Class	RAKIP project + SSD + ICRAd + FDA-iRisk® + PMM-Lab + OpenFSMR (class-specific controlled vocabularies)
Basic process	ICRA + FDA-iRisk® + PMM-Lab (class-specific controlled vocabularies)
Status	RAKIP project
Product-matrix name	SSD
Product-matrix unit	SSD and PMM-Lab
Method of production	SSD
Packaging	SSD
Product treatment	SSD
Country of origin	SSD
Area of origin	SSD
Fisheries area	SSD
Hazard type	SSD
Hazard name	SSD
Hazard unit	SSD and PMM-Lab
Hazard ind-sum	SSD
Population name	FOODON
Region	SSD

Country	SSD
Study Assay Technology Type	SSD
Accreditation procedure Ass.Tec	SSD
Sampling strategy	SSD
Type of sampling program	SSD
Sampling method	SSD
Lot size unit	SSD and PMM-Lab
Sampling point	SSD
Method tool to collect data	RAKIP project
Type of records	RAKIP project
Food descriptors	SSD
Laboratory accreditation	SSD
Laboratory country	SSD
Parameter classification	RAKIP project
Parameter unit	SSD and PMM-Lab
Parameter unit category	PMM-Lab
Parameter data type	RAKIP project
Parameter source	RAKIP project
Parameter subject	RAKIP project
Parameter distribution	probONTO
Model equation class-distr	FDA-iRISK®
Fitting procedure	RAKIP project
Type of exposure	RAKIP project

4. Metadata Schema of the HDF5-file

The metadata schema that is attached to the data sets and groups of the HDF5-file is resented in Table S4. This schema is based on DataCite Metadata Working Group (2019).

Table S4 Metadata schema of the HDF5-objects.

Property	Obligation	Notes
Identifier	Mandatory	Unique ID for referencing, could be DOI reference
Number of columns and rows	Mandatory	
Column and row names	Mandatory	

5. Funding

This work was supported by the AGINFRA PLUS project funded by the European Commission's Horizon 2020 research and innovation programme under grant agreement No 731001.

6. References

Bergmann, F. T., Cooper, J., Le Novere, N., Nickerson, D., & Waltemath, D. (2015). Simulation experiment description markup language (SED-ML) Level 1 Version 2. *Journal of integrative bioinformatics, 12*(2), 262. doi:10.2390/biecoll-jib-2015-262

DataCite Metadata Working Group. (2019). DataCite metadata schema documentation for the publication and citation of research data. *DataCite e.V., Version 4.3*. doi:https://doi.org/10.14454/7xq3-zf69

Freed, N., & Borenstein, N. (1996). Multipurpose internet mail extensions (MIME) part two: Media types. *rfc 2046, November*.

Plaza Rodriguez, C., Correia Carreira, G., & Kasbohrer, A. (2018). A probabilistic transmission model for the spread of extended-spectrum-beta-lactamase and AmpC-beta-lactamase-producing Escherichia coli in the broiler production chain. *Risk Anal, 38*(12), 2659-2682. doi:10.1111/risa.13145