



# Format Overflow?

## Handling of Modeling Projects in Systems Biology

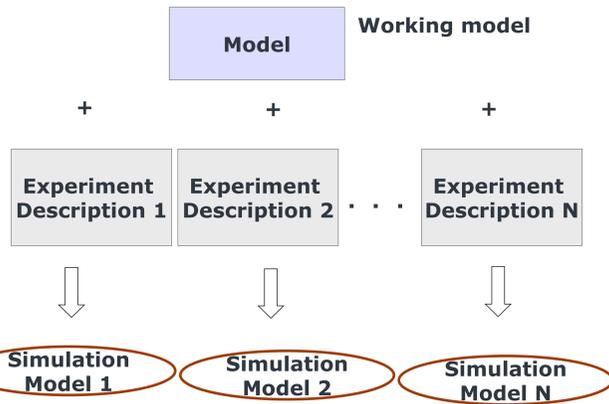
Arne T. Bittig & Henning Schmidt  
University of Rostock, Systems Biology and Bioinformatics Group

### Starting Point: SBML

- Milestone in the evolution of systems biology



- 1 SBML file: 1 single model
- Actual Systems Biology:
  - Different experiments on „same“ system (knockout, overexpression, change of concentration 1, knockout & c.o.c.,...)
- How to code the experiments?
  - 1 experiment = 1 model? (large number of models to keep track of, higher risk of modeling errors, hard to handle modeling projects, model overflow)
  - Flags and switches for all experiments (huge model, also quite impractical)
  - Separate experiment descriptions



- Wet-lab experimental settings analogue
- Automatic merging with model
- Once model finished: documentation of models validity

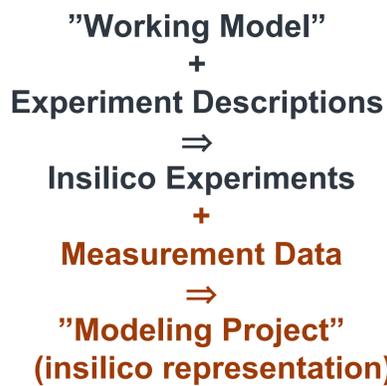
```
***** MODEL NAME
Simple model
***** MODEL STATES
d/dt (A) = -R
d/dt (B) = R
A(0) = 1
B(0) = 0
***** MODEL PARAMETERS
k1 = 0.5
***** MODEL REACTIONS
R = k1*A

***** EXPERIMENT NAME
Simple Experiment for simple model
***** EXPERIMENT INITIAL PARAMETER AND STATE SETTINGS
k1 = 2
***** EXPERIMENT PARAMETER CHANGES
***** EXPERIMENT STATE CHANGES
time=10, A=1

***** MODEL NAME
Simple model
***** MODEL STATES
d/dt (A) = -R
d/dt (B) = R
A(0) = 1
B(0) = 0
***** MODEL PARAMETERS
k1 = 2
***** MODEL REACTIONS
R = k1*A
***** MODEL EVENTS
StateChange_1 = ge(time,10),A,1
```

Successfully used in the sbtoolbox(2) for 2 years now

### 1 Level up: Modeling Project



- Combination model + data relevant for many (most?) simulation tools
  - Current state:
    - Messy and overly complicated combination of model + experiments (if present)
    - Proprietary modeling project representations (PottersWheel, SBtoolbox2)
  - Interoperable project format useful
- Possible starting point: MIASE [1] (Minimum Information About a Simulation Experiment)
  - Guidelines for descriptions of simulation runs (perturbed & unperturbed models)
  - Focus: reproduction of exact simulation procedures and resulting outputs (plots)
  - SedOM/SedML: Simulation experiment descript. object model/markup language
    - UML/xml-based representations of MIASE
- SedML as modeling project format
  - SedML independent of model description (only references to model in SBML, CellML, or other format)
  - Experiment descriptions already there (sort of: model perturbations as XPath expr. referring to SBML [2] tree of model)
  - Measurement data still missing

### Measurement Data Format

- Standard not in sight
  - Mostly two-dimensional data (components x time points), xml not too suitable
- Needed, primarily:
  - Representation of time series data
  - Representation of steady-state data
  - Variance, error bounds
- For additional convenience: storage format for simulation data

### Implementation

- Proof-of-concept impl. in SBtoolbox2
  - Im/exports SedML project description (modified) and SBML models instead of proprietary (mainly text-based) formats
  - Ignores plot generation info (so far)
- Each model has an associated
  - list of changes (=exp. settings) w.r.t. original model (=SBML file) (SedML)
  - list of measurements = reference(s) to file(s) with experimental data obtained using the experimental settings (new)
- Measurement data: xls or csv files

	A	B	C	D	E	F	G	H
1 Name	Measurement Example 2							
2 Notes	Just some notes in a single line							
3 Componentnotes								
4 Components	time	A	A+	A-	Component B	Component C		
5 Values	0	0.0172	0.01892	0.01548	0.0116	0.0009		
6	49.0552	0.0171892	0.0189081	0.0154703	0.0115591	0.000865671		
7	99.9524	0.0170654	0.0187939	0.0153769	0.0114354	0.000835473		
8	128.814	0.0173518	0.019087	0.0156166	0.0118688	0.000825422		
9	151.362	0.0169693	0.0186663	0.0152724	0.0109388	0.00104578		
10	160.548	0.0181534	0.0199687	0.0163381	0.0125635	0.00104271		
11	173.206	0.0170857	0.0187943	0.0153772	0.0119458	0.000671465		
12	186.74	0.0146325	0.0160957	0.0131652	0.0089325	0.0010277		
13	189.85	0.0150251	0.0165276	0.0135226	0.0086028	0.00138207		
14	190.64	0.0151959	0.0167155	0.0136763	0.00811683	0.00149481		
15	199.735	0.0173497	0.0190847	0.0156147	0.00929493	0.00299474		
16	207.51	0.0185011	0.0204612	0.016741	0.00937849	0.00270628		
17	211.924	0.0190048	0.0209053	0.0171043	0.0092062	0.00270628		
18	217.317	0.0190811	0.0209892	0.017173	0.00912062	0.00270628		
19	226.539	0.0161404	0.0177545	0.0145264	0.0111411	0.000550005		
20	235.881	0.0135297	0.0148827	0.0121768	0.0071788	0.000783395		
21	241.166	0.0139325	0.0153258	0.0125393	0.00671213	0.00146087		
22	241.697	0.0140435	0.0154479	0.0126392	0.00672502	0.00150506		
23	242.971	0.0143499	0.0157849	0.0129149	0.00679063	0.00178496		
24	255.948	0.017554	0.0193094	0.0157985	0.00790143	0.00468672		

- Scope question
  - One project = one model + some exp.s
  - Or: one project = a few models + several experiments?

### Summary

- Bioinformaticians, database experts:
  - Focus on model storage & annotation, reproduction of simulation results, etc.
  - Much effort put in already:



- Modelers, systems biologists:
  - Universal measurement data format & experiment descriptions needed
  - Not much progress so far
- Inter-disciplinary dialogue possible
  - as indicated by the proof-of-concept implementation in SBtoolbox2
  - Objectives need to be reconciled
    - Who needs the plot reproduction features of SedML without included experimental data?
    - Which features of modeling projects in other software tools are still missing?

### References

- [1] D Köhn, N le Novere. SED-ML – An XML Format for the Implementation of the MIASE Guidelines, to appear in LNCS CMSB08
- [2] M Hucka et al. The system biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics, 2003, 19(4), 524-531 www.sbml.org

### Acknowledgements

EC FP6 – AMPKIN  
BMBF – Forsys Partner



### Contact

Henning Schmidt  
FORSYS Research Group hIMOsYS  
Systems Biology and Bioinformatics  
University of Rostock  
Albert Einstein Str. 21  
18059 Rostock, Germany  
henning.schmidt@uni-rostock.de