ZIH Center for Information Services & High Performance Computing

TECHNISCHE UNIVERSITÄT DRESDEN

Heidelberg Institute for Theoretical Studies

HITS

<MulticellML/>

FAIR data exchange in the life sciences by standardization of heterogenous data and multicellular models

Martin Golebiewski

HITS gGmbH (Heidelberg, Germany)

COMBINE 2018, Boston, MA (USA), October 8th-12th, 2018

NORMSYS

Supported by: Federal Ministry for Economic Affairs and Energy

e:Med SYSTEMS MEDICINE

LiSyM Liver Systems Medicine

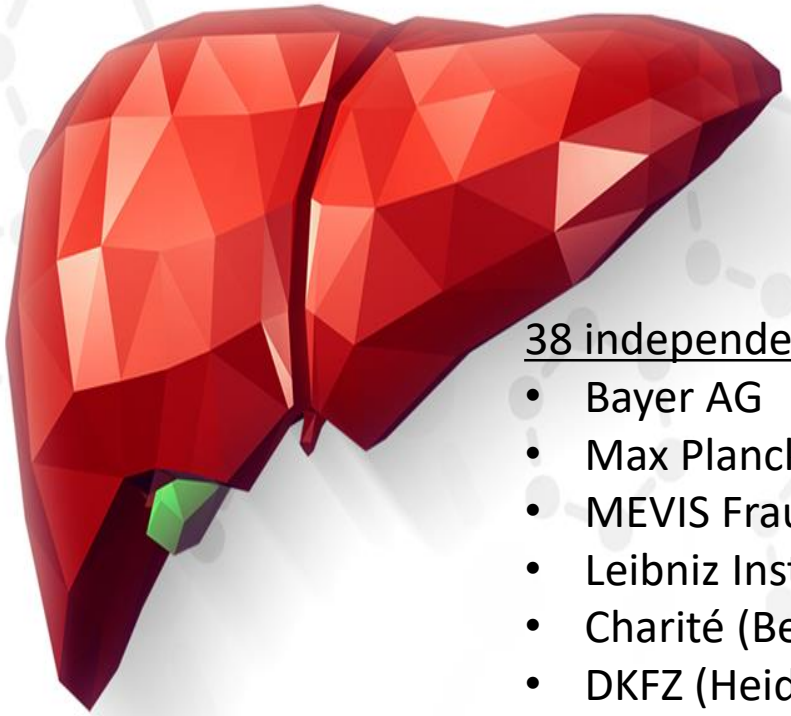
SPONSORED BY THE Federal Ministry of Education and Research

LiSyM
Liver Systems Medicine

Federal Ministry of Education and Research

LiSyM

Research Network on Systems Medicine for Liver Disease



<u>Multiple disciplines</u>

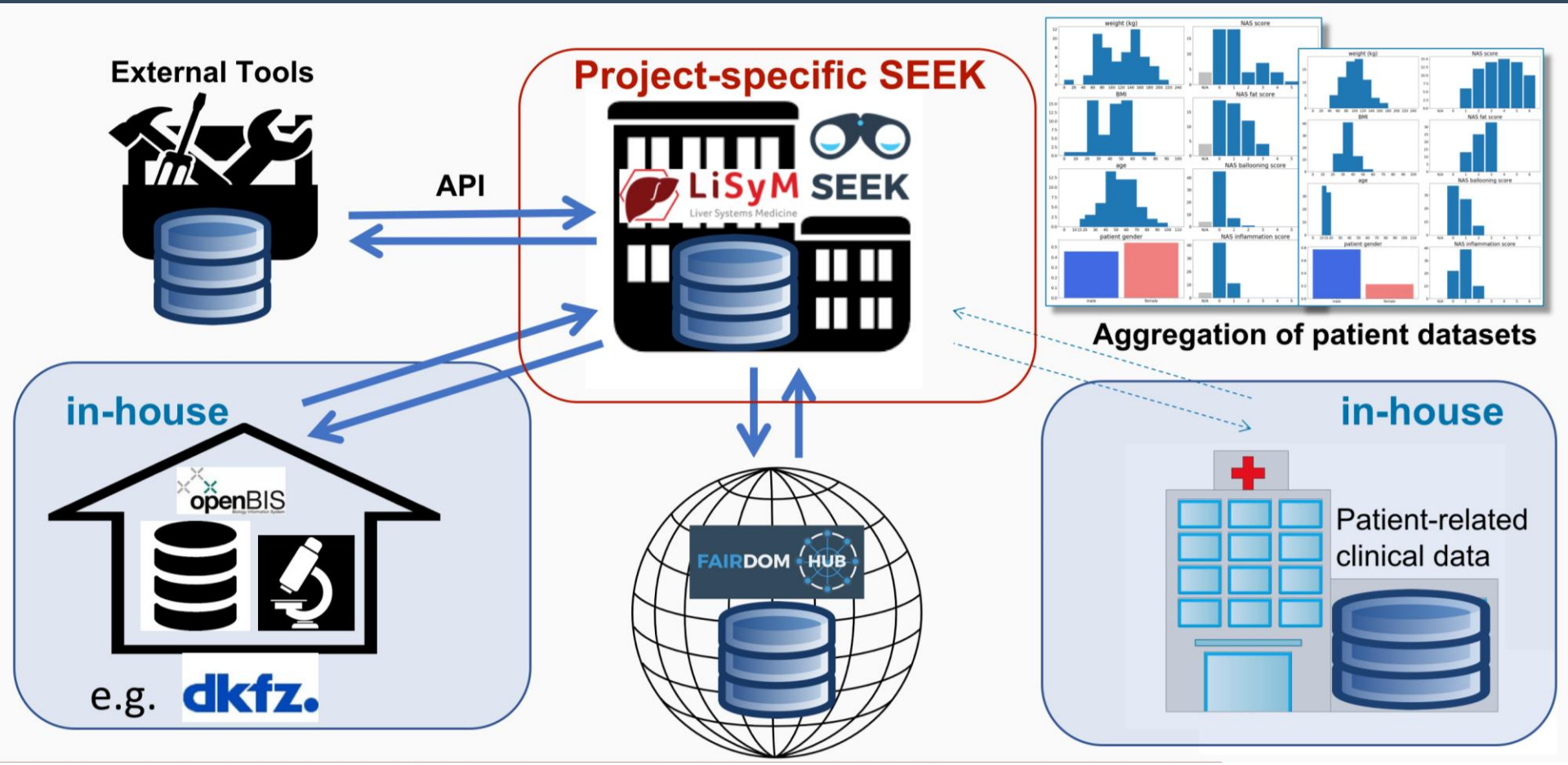
- Medicine
- Biology, Biochemistry
- Pharmacology
- Physics
- Bioinformatics
- Data management
- Industry

<u>38 independent research groups</u>

- Bayer AG
- Max Planck (Dresden and Berlin)
- MEVIS Fraunhofer (Bremen)
- Leibniz Institute IfaDo (Dortmund)
- Charité (Berlin)
- DKFZ (Heidelberg)
- <u>Hospitals:</u> Dresden, Kiel, Aachen, Homburg, Berlin, Heidelberg, Munich
- + 18 Universities

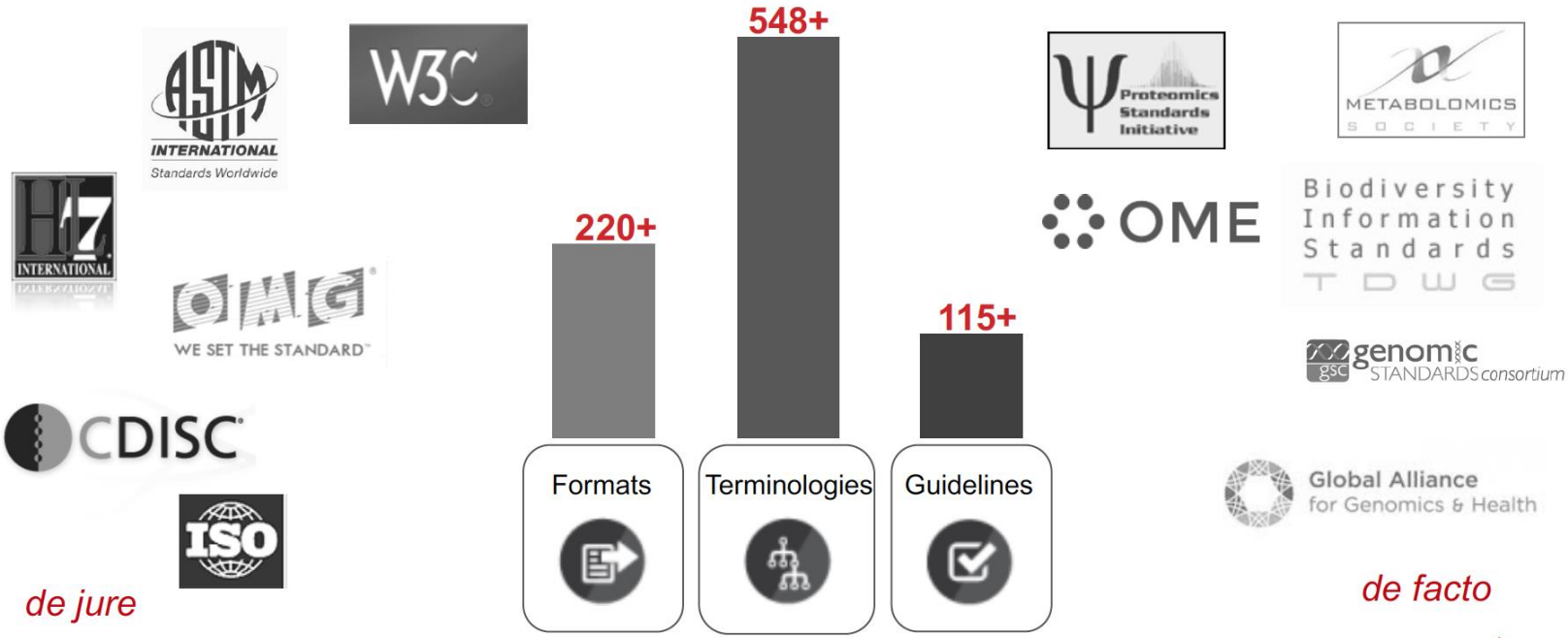
Supported by
The German Federal Ministry of Education and Research 2016-2020

Project-Specific Data Management Support: German Liver Systems Medicine Network (LiSyM)



HITS
MANCHESTER 1824 The University of Manchester
de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE
ERASysAPP ERA-Net for Applied Systems Biology
LiSyM Liver Systems Medicine
SPONSORED BY THE Federal Ministry of Education and Research

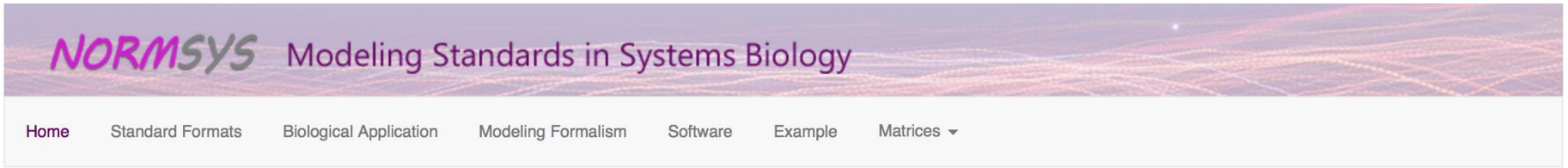
Community (de-facto) Standards in Life Sciences



FAIRsharing.org
standards, databases, policies
https://fairsharing.org

Source: Susanna-Assunta Sansone (University of Oxford, UK)

NORMSYS
Supported by:
Federal Ministry for Economic Affairs and Energy
HITS



The NormSys Registry for Modeling Standards

This registry aims at surveying standard formats for computational modeling in biology. It not only lists the standards, but also compares their major features, their possible fields of biological application and use cases (including model examples), as well as their relationships, commonalities and differences. This registry provides a common entry point for modelers and software developers who plan to apply the standards for their respective case of application, and serves them with detailed information and links to the standards, their specifications and APIs.

The information provided in this system does not claim to be complete or all-encompassing, nor can we guarantee any absence of defectiveness. However, we collect and assemble the information to the best of our knowledge and belief to to assist in selecting the appropriate standard format for your specific requirements. Please note that the system is work in progress and is constantly revised and extended. Any feedback and suggestions for corrections or improvements, as well as for new fields of applications to be included (with example models) are highly welcome.

http://normsys.h-its.org/

Format Classes

Systems Biology Markup Language (SBML)	CellML	Systems Biology Graphical Notation (SBGN)	Simulation Experiment Description Markup Language (SED-ML)
Pharmacometrics Markup Language (pharmML)	NeuroML	FieldML	Synthetic Biology Open Language (SBOL)

Formats / Details

NORMSYS
Supported by:
Federal Ministry for Economic Affairs and Energy
HITS

Show results for:

Select

Format

or filter by

Format class

Systems Biology Markup Language (S

Biological application

Modeling formalism

Software

Api language

Supported biological scale

- molecular
- cellular
- tissue
- organ
- organism
- ecosystem

Spatial representation

- Compartment
- Dimensions

Standard Formats

Displaying: 1 | Found: 1 | Total: 16

SBML L3V1 Core

Systems Biology Markup Language Level 3 Version 1 Core

Synopsis

Representation of biological processes as a set of processes, that are converting pools of entities into other pools entities.

Description

Systems Biology Markup Language (SBML), a free, open, XML-based format for representing biochemical reaction networks. SBML is a software-independent language for describing models common to research in many areas of computational biology, including cell signaling pathways, metabolic pathways, gene regulation, and others.*

*(M. Hucka et al. Bioinformatics (2003) 19 (4): 524-531)

Biological Applications | Class | Details | License | Links | Transformations | APIs | Validator

Publication Date

10/2010

Authors

Hucka, Michael
Sahle, Sven
T. Bergmann, Frank
C. Schaff, James
Hoops, Stefan
P. Smith, Lucian
M. Keating, Sarah
J. Wilkinson, Darren

Organizations

- combine

Biological Scales

Scale	molecular	cellular	tissue	organ	organism	ecosystem
Support	intrinsic	potential	potential	unknown	unknown	potential

NORMSYS
Supported by:
Federal Ministry for Economic Affairs and Energy
HITS

Show results for:

Select

Format

or filter by

Format class

Systems Biology Markup Language (S

Biological application

Modeling formalism

Software

Api language

Supported biological scale

- [ ] molecular

Standard Formats

Displaying: 1 | Found: 1 | Total: 16

SBML L3V1 Core

Systems Biology Markup Language Level 3 Version 1 Core

Synopsis

Representation of biological processes as a set of processes, that are converting pools of entities into other pools entities.

Description

Systems Biology Markup Language (SBML), a free, open, XML-based format for representing biochemical reaction networks. SBML is a software-independent language for describing models common to research in many areas of computational biology, including cell signaling pathways, metabolic pathways, gene regulation, and others.*

*(M. Hucka et al. Bioinformatics (2003) 19 (4): 524-531)

Biological Applications | Class | Details | License | Links | Transformations | APIs | Validator

Webpage

- SBML

Specification

- SBML
- The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core

Publication

- The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.

Model repository

- BioModels Database

Software Repository

- SBML Software Guide

NORMSYS

Supported by:

Federal Ministry for Economic Affairs and Energy

HITS

Show results for:

Select

Format

or filter by

Format class

Systems Biology Markup Language (S

Biological application

Modeling formalism

Software

Api language

Supported biological scale

- molecular
- cellular
- tissue
- organ
- organism
- ecosystem

Spatial representation

- Compartment

Standard Formats

Displaying: 1 | Found: 1 | Total: 16

SBML L3V1 Core

Systems Biology Markup Language Level 3 Version 1 Core

Synopsis

Representation of biological processes as a set of processes, that are converting pools of entities into other pools entities.

Description

Systems Biology Markup Language (SBML), a free, open, XML-based format for representing biochemical reaction networks. SBML is a software-independent language for describing models common to research in many areas of computational biology, including cell signaling pathways, metabolic pathways, gene regulation, and others.*

*(M. Hucka et al. Bioinformatics (2003) 19 (4): 524-531)

Biological Applications | Class | Details | License | Links | Transformations | APIs | Validator

Input \ Output	SBML L3V1 Core	CellML 1.1	SBGN PD L1 V1.3	NeuroML 2 beta 3	SBOL v2.0	SBGN-ML 0.2
SBML L3V1 Core	–	libAntimony	CellDesigner Arcadia	jNeuroML	–	SBML2SBGNML
CellML 1.1	libAntimony	–	–	–	–	–
SBGN PD L1 V1.3	–	–	–	–	–	–
NeuroML 2 beta 3	jNeuroML	–	–	–	–	–
SBOL v2.0	iBioSim	–	–	–	–	–
SBGN-ML 0.2	–	–	–	–	–	–

NORMSYS

Supported by:

Federal Ministry for Economic Affairs and Energy

HITS

NORMSYS Modeling Standards in Systems Biology

Home | Standard Formats | Biological Application | Modeling Formalism | Software | Example | Matrices

Biological Applications
Format Transformation

Biological Application \ Format	SBML L3V1 Core	CellML 1.1	SBGN ER L1 V1.2	SBGN PD L1 V1.3	SBGN AF L1 V1.0	MorphML v1.8.1	NeuroML 2 beta 3	PharmML v0.6	SBOL v2.0	SBOL Visual v1.0.0	ChannelML v1.8.1	Biophysics v1.8.1	NetworkML v1.8.1
Multi-organism Process	✔	✔	–	–	–	–	–	–	–	–	–	–	–
Cell Cycle	✔	✔	✔	–	–	–	–	–	–	–	–	–	–
Signaling	✔	✔	✔	✔	✔	–	–	–	–	–	–	–	–
Single Cell Morphology	–	–	–	–	–	✔	✔	–	–	–	–	–	–
Pharmacokinetic	✔	✔	–	–	–	–	–	✔	–	–	–	–	–
Pharmacodynamics	✔	✔	–	–	–	–	–	✔	–	–	–	–	–
Izhikevich-based Neuron Models	✔	–	–	–	–	–	✔	–	–	–	–	–	–
Synthetic Gene Regulatory Network	✔	–	✔	✔	✔	–	–	–	✔	✔	–	–	–
Metabolic Process	✔	✔	–	✔	–	–	–	✔	–	–	–	–	–
Immune Response	✔	✔	–	–	✔	–	–	–	–	–	–	–	–
Circadian Rhythm	✔	✔	✔	–	–	–	✔	–	–	–	–	–	–
Regulation of Gene Expression	✔	✔	✔	✔	✔	–	–	–	✔	✔	–	–	–
Electrophysiology	✔	✔	–	–	–	–	✔	–	–	–	✔	✔	–

NORMSYS Modeling Standards in Systems Biology

Home | Standard Formats | Biological Application | Modeling Formalism | Software | Example | Matrices | Validate Model (beta)

Given the FullXSD schema, the given model has 1 errors! ×

Given the FullXSD schema, the given model has 1 errors! ×

Validation

You can validate your xml model file against a selected format.
This validation is using libxml2 (http://xmlsoft.org/) to process the xml file using the Format's xsd schema (and dependencies).
After you select the Format your model is for, and upload a model file (*.xml), the file will be validated on this server. Upon completion you should see on top, if it was successful.
The two tabs below show the model file (with annotated errors if there are such). Alternatively, all errors can be inspected as a table.

Format

Systems Biology Markup Language Level 2 Version 5

Model file as xml

Datei auswählen Keine Date...usgewählt

Validate Model File

Model File | XML Errors 1 | External resource annotation 2 | RDF Graphs 37 | Alternative Validators

XML Errors; jump to: 7

External Resource Problmes; jump to:

```
1                   <?xml version='1.0' encoding='UTF-8' standalone='no'?>
2
3  <!-- This model was downloaded from BioModels Database -->
4  <!-- Mon Oct 10 15:40:40 BST 2016                      -->
5  <!-- http://www.ebi.ac.uk/biomodels/                   -->
```

NormSys Registry for Modelling Standards

- **Information** resource for community standards
- **Comparison** of their main characteristics and features
- **Classification** by fields of application (with examples)
- **Transformation** options between the standards
- **Bundled links** to corresponding web resources: specifications, websites, publications, repositories, APIs...
- **Faceted browsing** and search by different criteria
- **Validation** of models for several standards
- COMBINE standards and related efforts (for a start)

HITS

Universität Potsdam

LifeGlimmer GmbH

http://normsys.h-its.org/

Supported by:
Federal Ministry for Economic Affairs and Energy

Lessons learned from 10 years of declarative multicellular modeling

- MorpheusML
  - xml-language of open-source framework Morpheus [3,4], Specs in [5] by Jörn Starruß and Walter de Back
  - similar to combination of SBML L3 packages dyn(CBO)+spatial+comp
- Model repository [1] and 30+ scientific applications [2]

➢ Separation of bio/physics from spatial model-formalisms

➢ Entity-changing operations (birth/death of submodels)

➢ Lexical and spatial scoping (using cell-cell neighbor relations)

➢ Symbol dependency network to automatically schedule order of updates and generate sequential scheme

[1] https://imc.zih.tu-dresden.de/wiki/morpheus/doku.php?id=examples:examples
[2] https://academic.oup.com/bioinformatics/CrossRef-CitedBy/234757
[3] https://morpheus.gitlab.io
[4] Starruß et al., J. Stat. Phys. 128, 269 (2007). DOI: 10.1007/s10955-007-9298-9
[5] Starruß et al., Bioinformatics 30, 1331 (2014). DOI: http://dx.doi.org/10.1093/bioinformatics/btt772

HITS

Structure: separating bio/physics from spatial model formalisms

```
<?xml version='1.0' encoding='UTF-8'?>
<MultiCellML version="0.2" xmlns="…">
   <Model name="…" >
      <Description> (optional) </Description>
      <Space symbol="space"> (required)   ...physical, boundary conditions...   </Space>
      <Time symbol="time"> (optional)  </Time>
      <ListOfSpatialCompartmentTypes> (required)
         <GlobalType> (required) …properties, dynamics, nesting of comp.  </GlobalType>
         <SolidType> (optional, one or multiple) </SolidType>
         <LiquidType> (optional, one or multiple) </LiquidType>
         <CellType> (optional, one or multiple) …incl. SBML-defined dynamics… </CellType>
      </ListOfSpatialCompartmentTypes>
      <ListOfInteractions> (opt.) inter-entity processes and neighborhoods </ListOfInteractions>
   </Model>
   <Simulation>
      …parametrization of various spatial model formalisms (vertex, CPM, CBM…)
   </Simulation>
</MultiCellML>
```

HITS

Nesting of spatial compartments

```
<ListOfSpatialCompartmentTypes>
        <GlobalType>
                ...dynamical system definition from referenced SBML...
                <SpatialSubCompartmentTypes space-tiling="true">
                        <CompartmentType ref="ct1" />
                        <CompartmentType ref="ct2" />
                        <CompartmentType ref="medium" />
                </SpatialSubCompartmentTypes>
                <BoundaryValues default="medium">
                        <BoundaryValue axis= "x" value="solid1" />
                </BoundaryValues>
        </GlobalType>
```

-> Nesting tree of arbitrary depth incl. all spatial compartment types

Cell type declaration

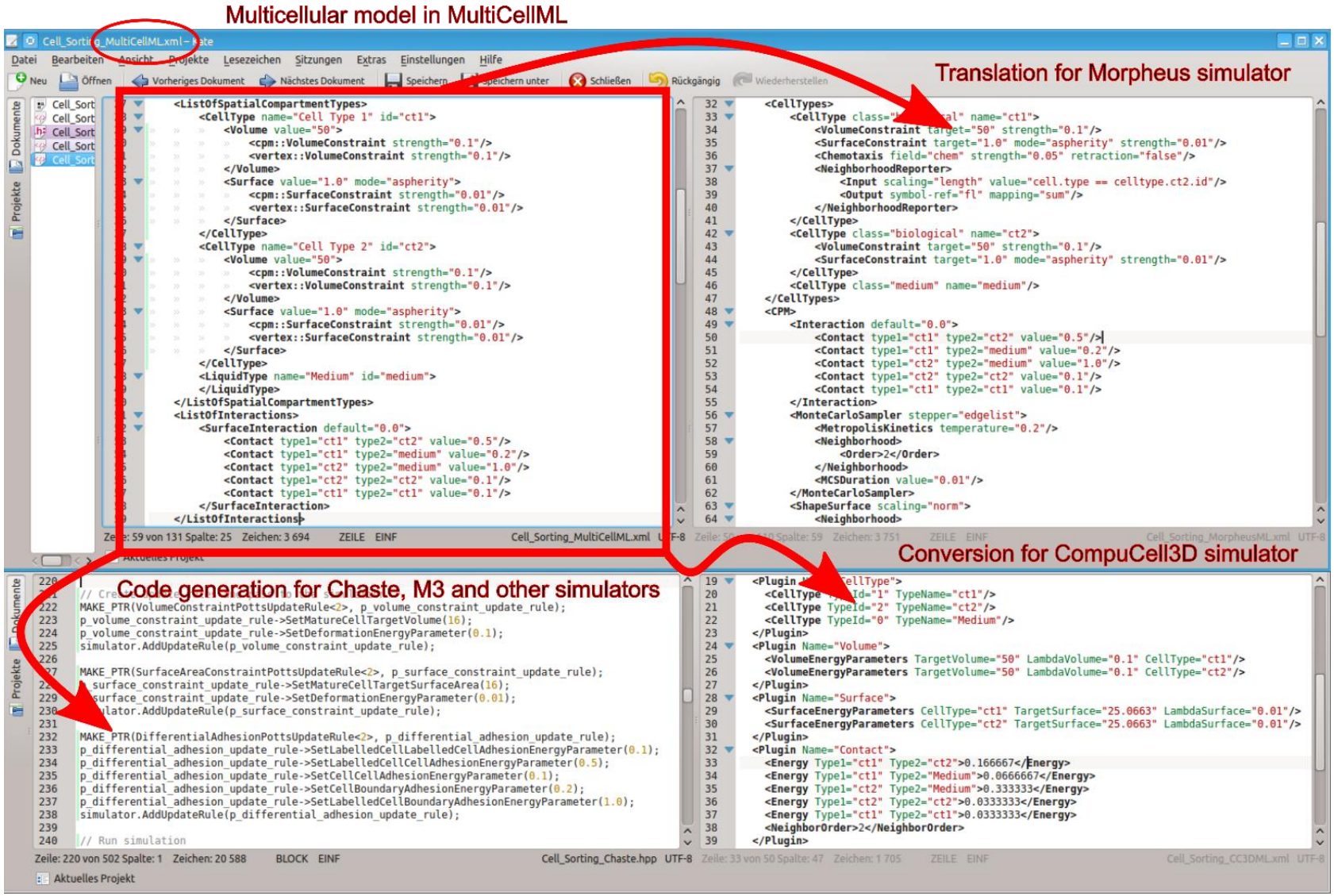
```
<CellType name="..." id="...">
        <Volume value="...">
                <cpm::VolumeConstraint strength="..."/>
                <vertex::VolumeConstraint strength="..."/>
        </Volume>
        <Surface value="..." mode="absolute/aspherity">
                <cpm::SurfaceConstraint strength="..."/>
                <vertex::SurfaceConstraint strength="..."/>
        </Surface>
        <SurfaceCompartment>
                ... optional, dynamical system definition for states along object shape...
        </SurfaceCompartment>

        ...optional, dynamical system definition from referenced SBML...

        <ListOfEntityProcesses>
                ...processes changing number of entities of given CellType, following CBO...
        </ListOfEntityProcesses>
</CellType>
```

(optional framework-specific parameters)

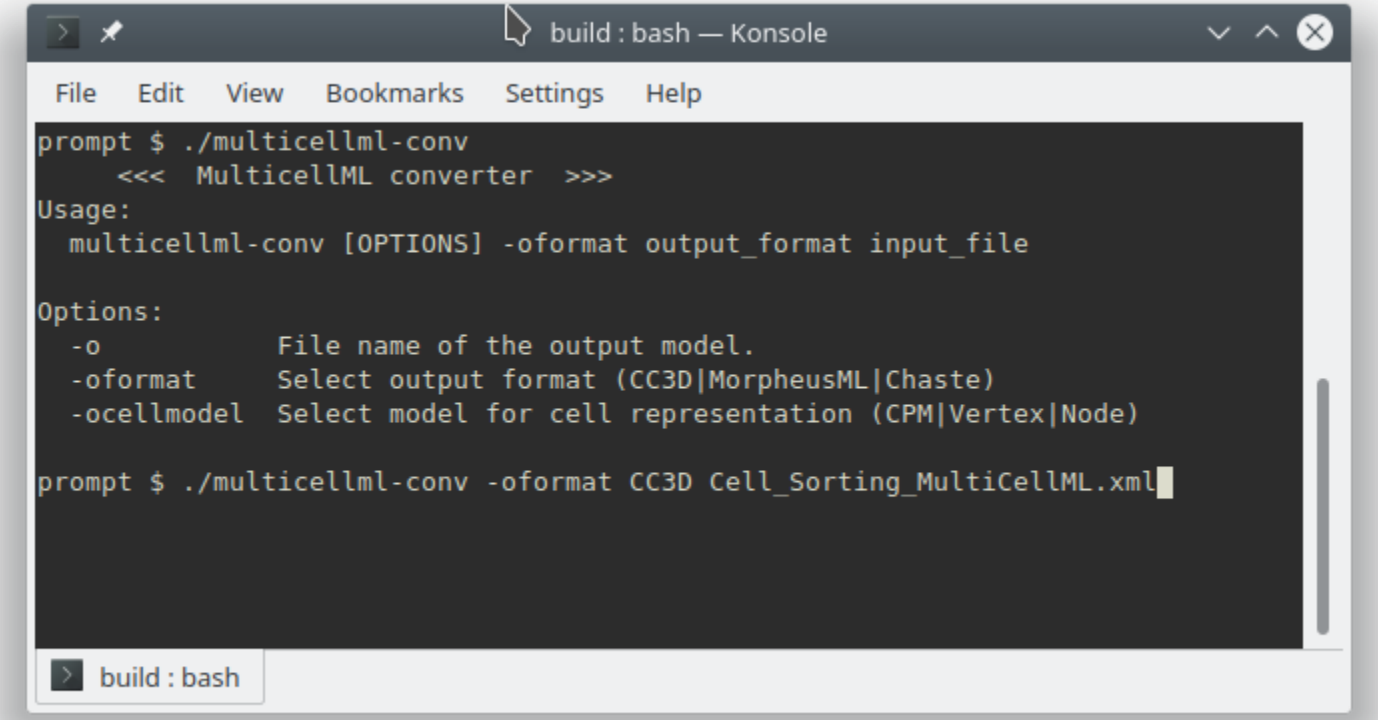
Parametrization of model formalisms

```
<Simulation>
    <StopTime value="1.0"/>
    <RandomSeed value="0"/>
    <DefaultTemporalSolver solver="…" time-step="1.0">
    <SpatialFramework>
        <Vertex>
            …
        </Vertex>
        <CPM>
            <Lattice class="square/hexagonal/cubic">
                <NodeLength value="1.0"/>
            </Lattice>
            <MonteCarloSampler stepper="edgelist">
                <MCSDuration value="1.0"/>
                <Neighborhood>
                    <Order>2</Order>
                </Neighborhood>
                <MetropolisKinetics temperature="10.0"/>
            </MonteCarloSampler>
        </CPM>
        …
    </SpatialFramework>
```

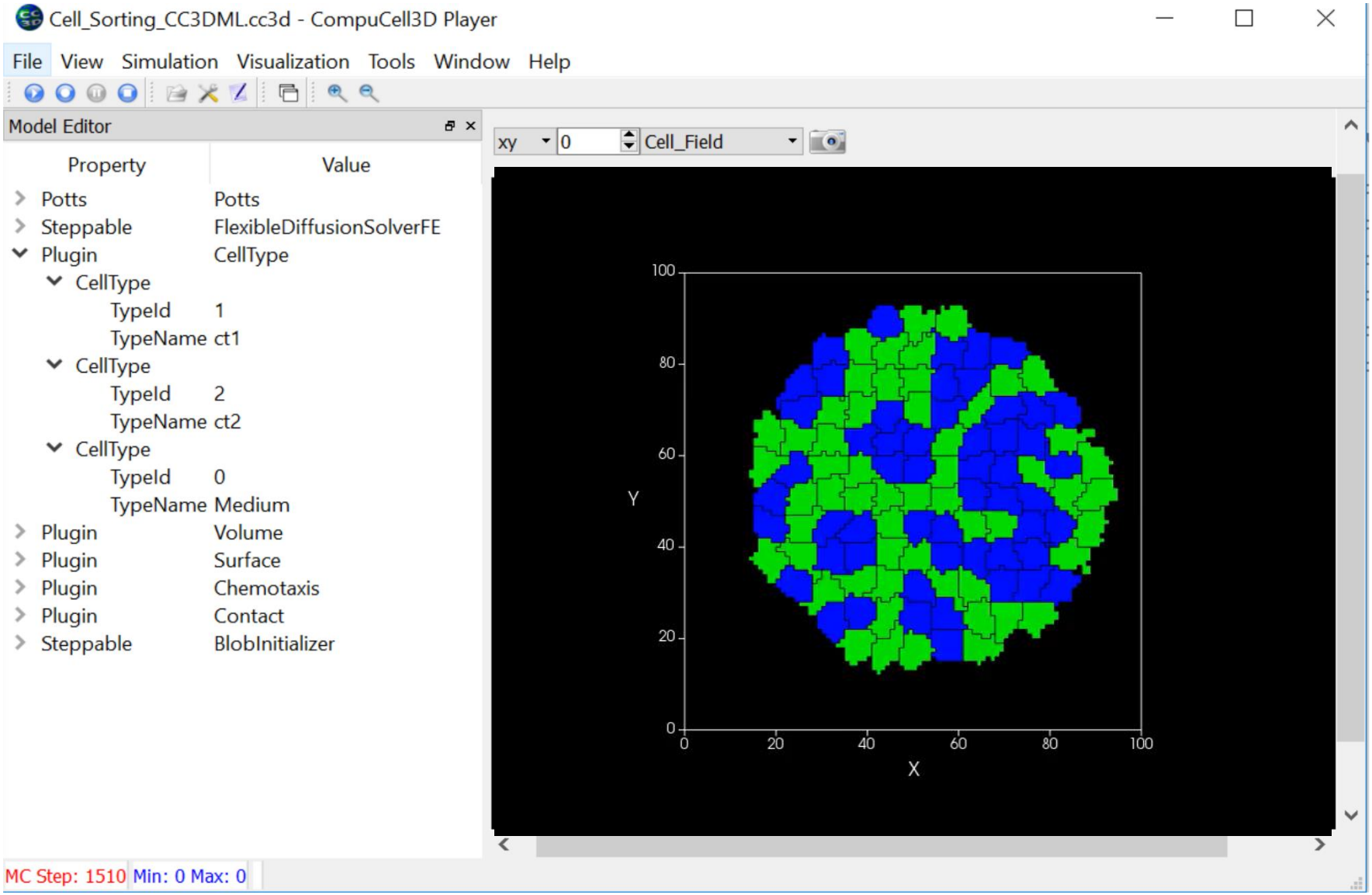
Model example in MultiCellML and derived simulator-specific codes



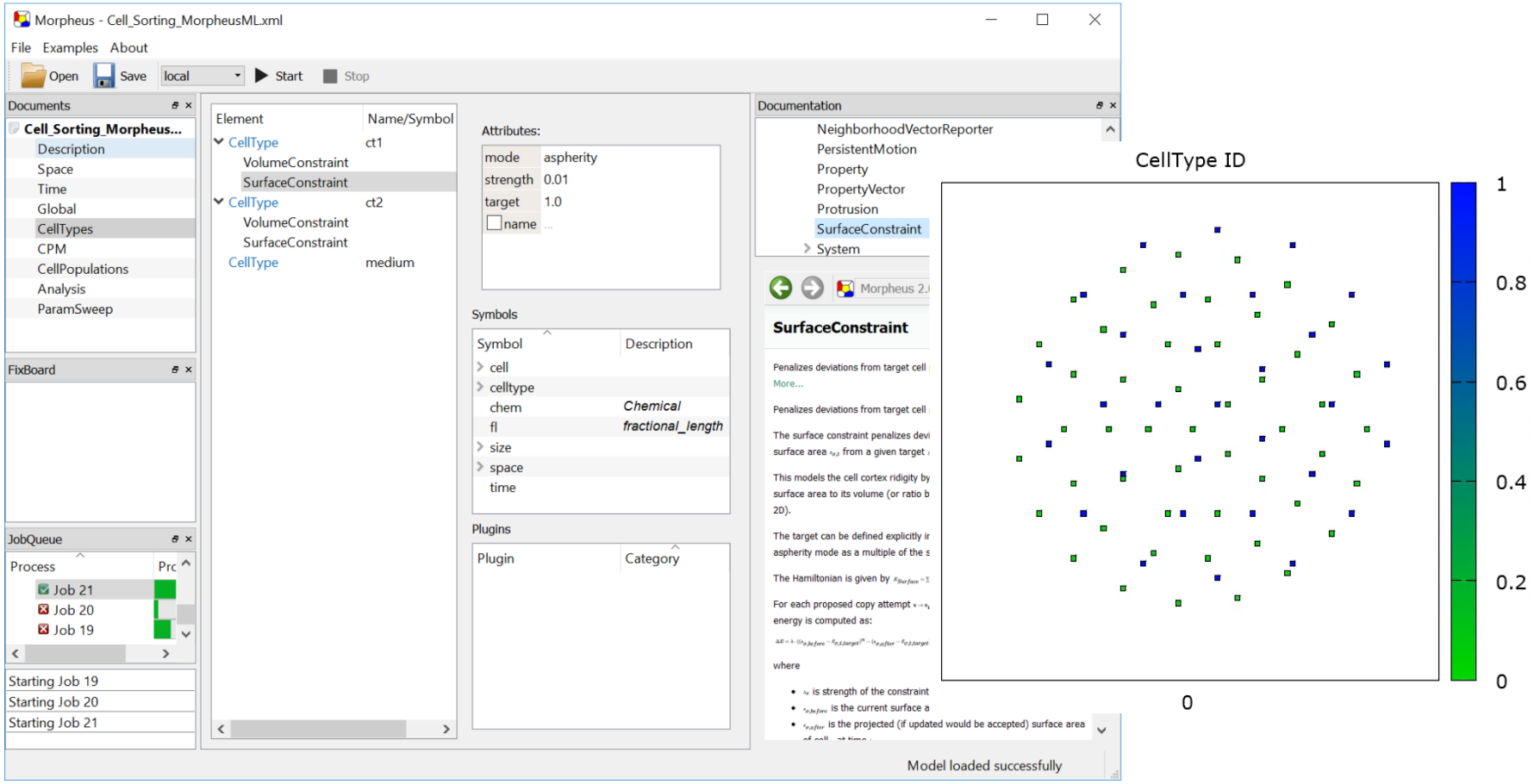
Converter using libMultiCellML as testbed for specification



Simulation in CompuCell3D



Simulation in Morpheus



MultiCellML Roadmap

- Oct. 2018 v0.3 – Jörn's talk and live demo at ICSB, Lyon

- Mid 2019 v0.7
  - benchmark models from [1] declarable
  - libMultiCellML -> Morpheus, CC3D, Chaste
  - exploration of SBML L3 ext. compatibility
  - draft specs (incl. CPM namespace package)
  - hackathon in Dresden/Heidelberg, Germany for MultiCellML support by simulators

- End 2019 v1.0
  - consolidated specs (incl. CPM, Vertex, CBM)
  - repository for multicellular models

- End 2020 v2.0
  - support for Units, MathML, annotations
  - language bindings to libMultiCellML

[1] Osborne et al., Comparing individual-based approaches to modelling the self-organization of multi-cellular tissues. PLoS Comput Biol 13(2): e1005387. DOI: 10.1371/journal.pcbi.1005387

ISO/IEC 23092 - MPEG-G (Genomic Information Representation)

MPEG-G at a Glance

ISO/IEC 23092

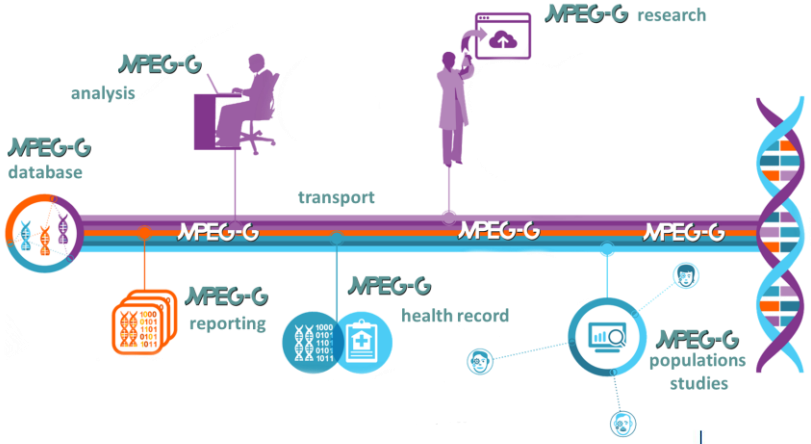
Developed by ISO/IEC JTC 1/SC 29/WG 11
a.k.a. Moving Picture Experts Group (MPEG)

Part 1 DIS	Part 2 DIS	Part 3 CD	Part 4 CD	Part 5 WD
Transport and Storage of Genomic Information	Coding of Genomic Information	APIs	Reference Software	Conformance
The technology to transport and access data	The compressed representation	The standard interfaces with genomic data applications and formats	The standard support to the implementation of applications	The methodology to test compliance with the standard

Standardization Workplan

October 2017	Committee Draft
July 2018	Draft International Standard (DIS)
January 2019	Final Draft International Standard (FDIS)

International Standard (IS)



Heidelberg Institute for Theoretical Studies
HITS

ISO/AWI 20691

Requirements for data formatting and description in the life sciences for downstream data processing and integration workflows

ISO/TC 276 Biotechnology WG 5 (Data Processing and Integration) works on a draft for a new ISO standard in the life sciences:

Reference framework („hub“) standard for (non-ISO) community standards:

- Rules and application guidelines for community standards for the formatting and documentation of life science data and computer models
- Recommended workflows for the structured processing, storing and integration of data and corresponding computer models
- Catalogue of criteria and requirements for life science data formats and model formats as prerequisites for a framework of interoperable standards

Towards MultiCellML

Jörn Starruß, Walter de Back, Lutz Brsuch

Center for Information Services and High Performance Computing, TU Dresden, Germany

NORMSYS

Heidelberg Institute for Theoretical Studies

HITS

http://normsys.h-its.org/

Concept & project lead: Martin Golebiewski
Design: Jill Zander
Implementation: Nils Wötzel
Content: Martin Golebiewski, Alexander Nikolaew

Collaboration partners:
Susanne Hollmann & Bernd Müller-Röber (University of Potsdam)
Babette Regierer (SB ScienceManagement UG, Berlin)
Vitor Martins dos Santos (LifeGlimmer GmbH, Berlin)

ISO

DIN

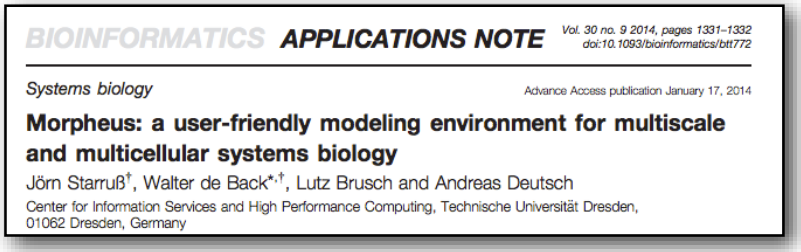
combine

Supported by:
Federal Ministry for Economic Affairs and Energy

Appendix
Overview of Morpheus

Publication

Starruß et al., *Bioinformatics*, 2014



BIOINFORMATICS **APPLICATIONS NOTE** *Vol. 30 no. 9 2014, pages 1331–1332*
doi:10.1093/bioinformatics/btt772

Systems biology Advance Access publication January 17, 2014

Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology

Jörn Starruß[†], Walter de Back[*,†], Lutz Brusch and Andreas Deutsch
Center for Information Services and High Performance Computing, Technische Universität Dresden, 01062 Dresden, Germany

Available for

Developers

Jörn Starruß (core)
Walter de Back (co)

Website

Google morpheus dresden

References

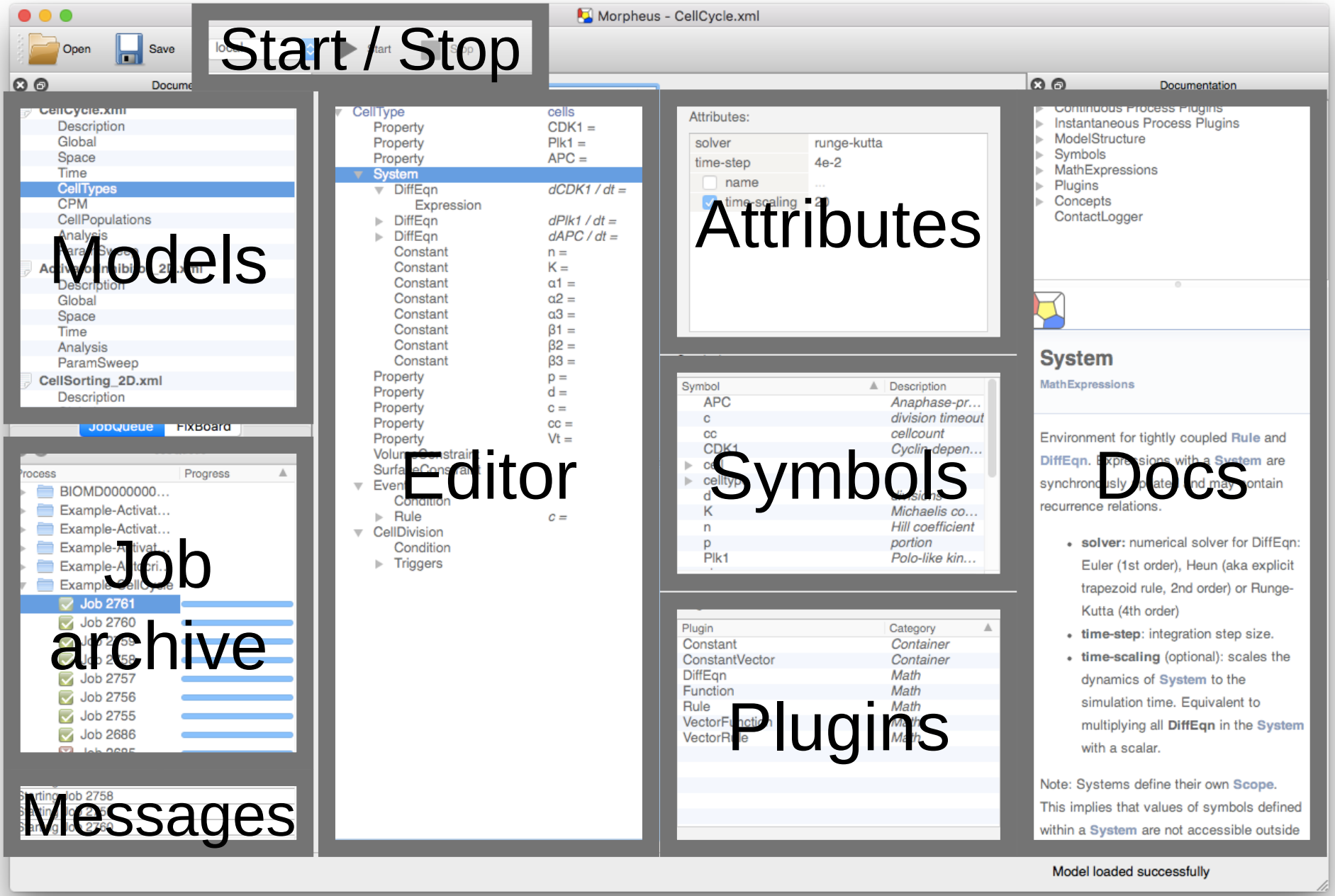
de Back et al., *J RS Interface*, 2012
de Back et al., *BMC Sys Biol*, 2013
Starruß et al., *Bioinformatics*, 2014
Herberg et al., *Cytometry A*, 2015

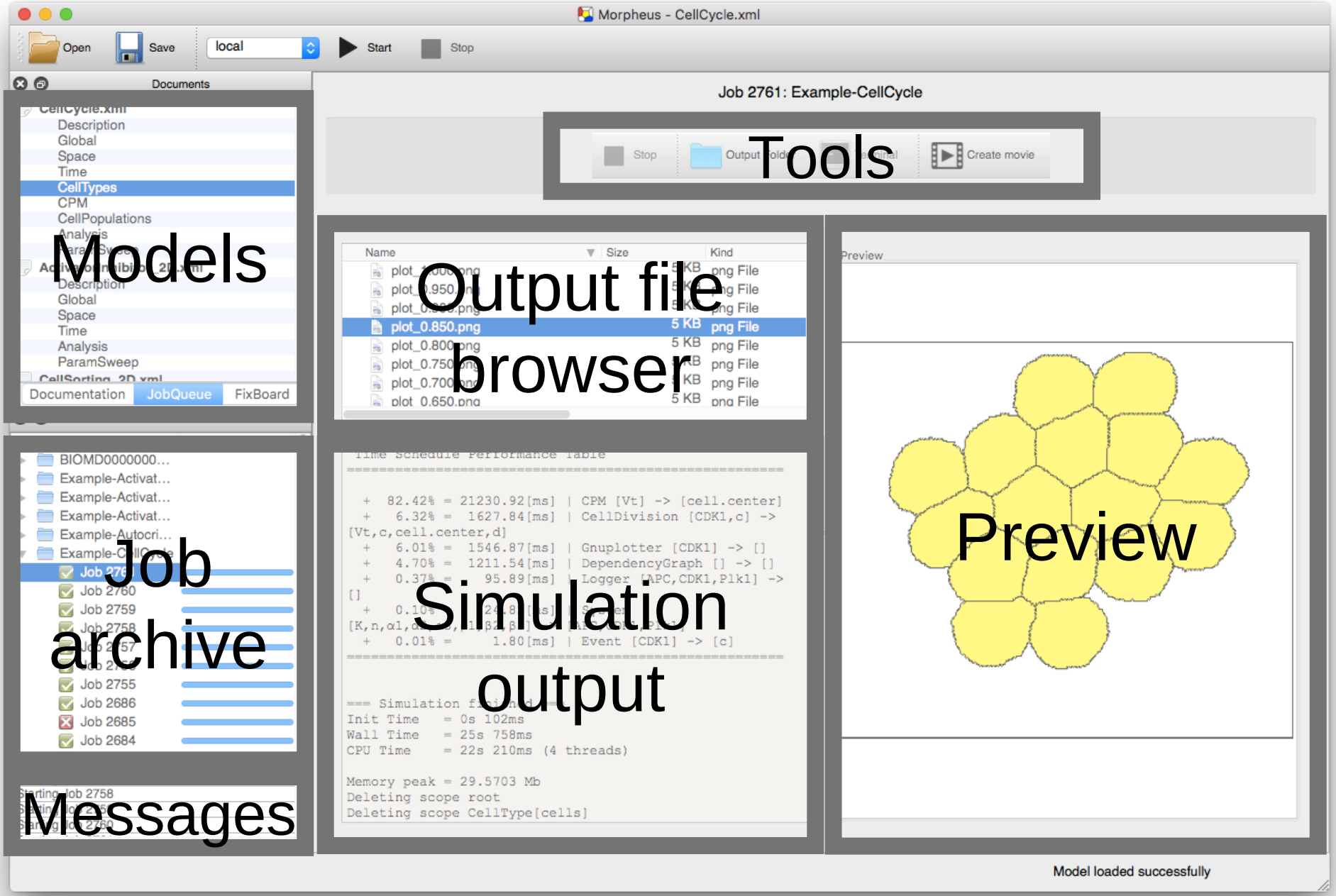
Morpheus 2.0

Open source GPL license
Plugin development
GitLab repository

Funding

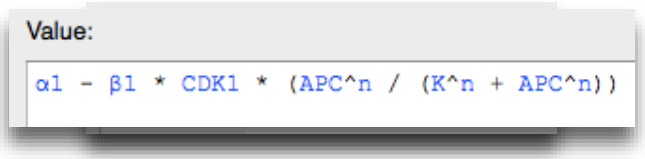
Bundesministerium für Bildung und Forschung
virtual liver network


Morpheus - CellCycle.xml
Open
Save
Start / Stop
Models
Editor
Attributes
Symbols
Docs
Job archive
Messages
Plugins
Documentation
Attributes:
solver
runge-kutta
time-step
4e-2
name
time-scaling
Symbol
Description
APC
Anaphase-pr…
c
division timeout
cc
cellcount
CDK1
Cyclin depen…
K
Michaelis co…
n
Hill coefficient
p
portion
Plk1
Polo-like kin…
Plugin
Category
Constant
Container
ConstantVector
Container
DiffEqn
Math
Function
Math
Rule
Math
Instantaneous Process Plugins
ModelStructure
Symbols
MathExpressions
Plugins
Concepts
ContactLogger
System
MathExpressions
Environment for tightly coupled Rule and DiffEqn. Expressions with a System are synchronously updated and may contain recurrence relations.
solver: numerical solver for DiffEqn: Euler (1st order), Heun (aka explicit trapezoid rule, 2nd order) or Runge-Kutta (4th order)
time-step: integration step size.
time-scaling (optional): scales the dynamics of System to the simulation time. Equivalent to multiplying all DiffEqn in the System with a scalar.
Note: Systems define their own Scope. This implies that values of symbols defined within a System are not accessible outside
Model loaded successfully


Morpheus - CellCycle.xml
Open
Save
local
Start
Stop
Documents
Description
Global
Space
Time
CellTypes
CPM
CellPopulations
Analysis
ParamSweep
Documentation
JobQueue
FixBoard
Job 2761: Example-CellCycle
Stop
Create movie
Tools
Models
Name
Size
Kind
plot_0.850.png
5 KB
png File
Output file browser
Preview
BIOMD0000000...
Example-Activat...
Example-Activat...
Example-Activat...
Example-Autocri...
Job 2760
Job 2759
Job 2755
Job 2686
Job 2685
Job 2684
Job archive
Simulation output
Messages
Starting Job 2758
Init Time = 0s 102ms
Wall Time = 25s 758ms
CPU Time = 22s 210ms (4 threads)
Memory peak = 29.5703 Mb
Deleting scope root
Deleting scope CellType[cells]
Model loaded successfully

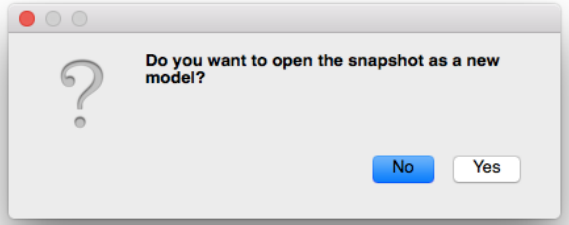
Graphical user interface

Features



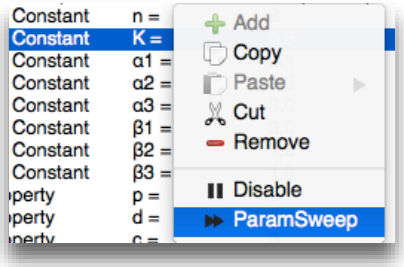
Mathematical expressions

In familiar in-fix notation



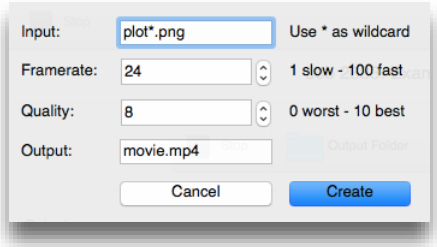
Re-use models

Stored in job archive



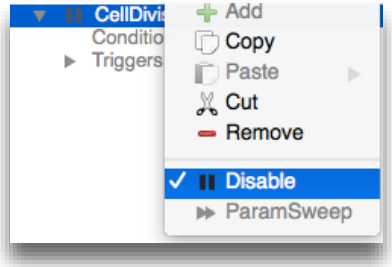
Parameter exploration

Batch simulation within GUI



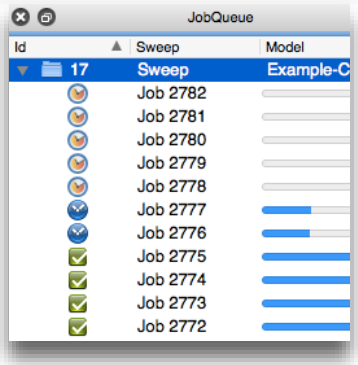
Create movies

From image sequences



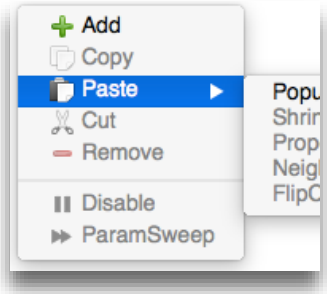
Temporarily disable elements

Testing and debugging



Copy-paste elements

Between different models

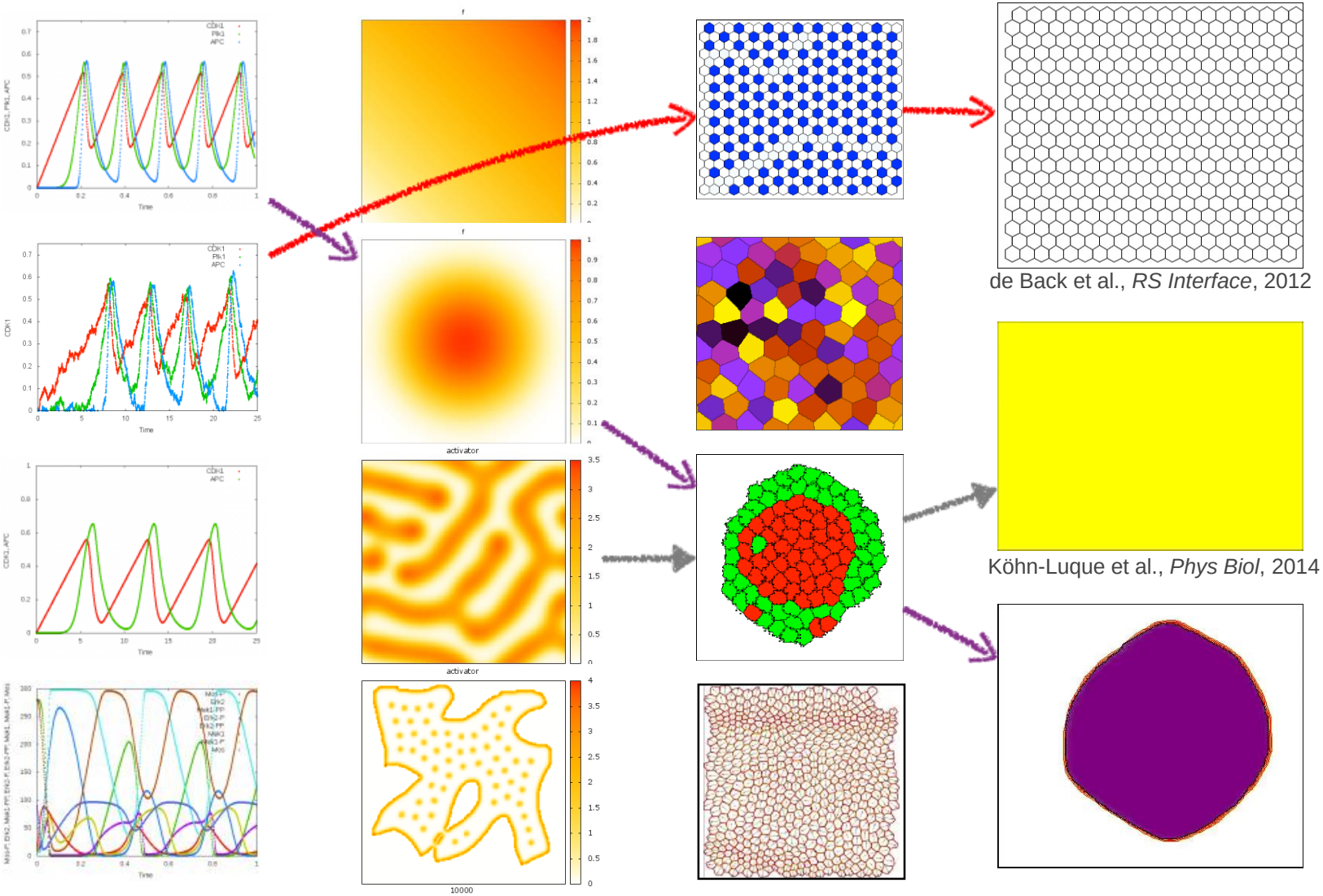


Job scheduling

Multithreading and parallel simulation

Multi-scale models

Coupling model formalisms



de Back et al., *RS Interface*, 2012

Köhn-Luque et al., *Phys Biol*, 2014

Graphical user interface

Data export and visualization

Gnuplotter

Visualize cell Properties and

Logger

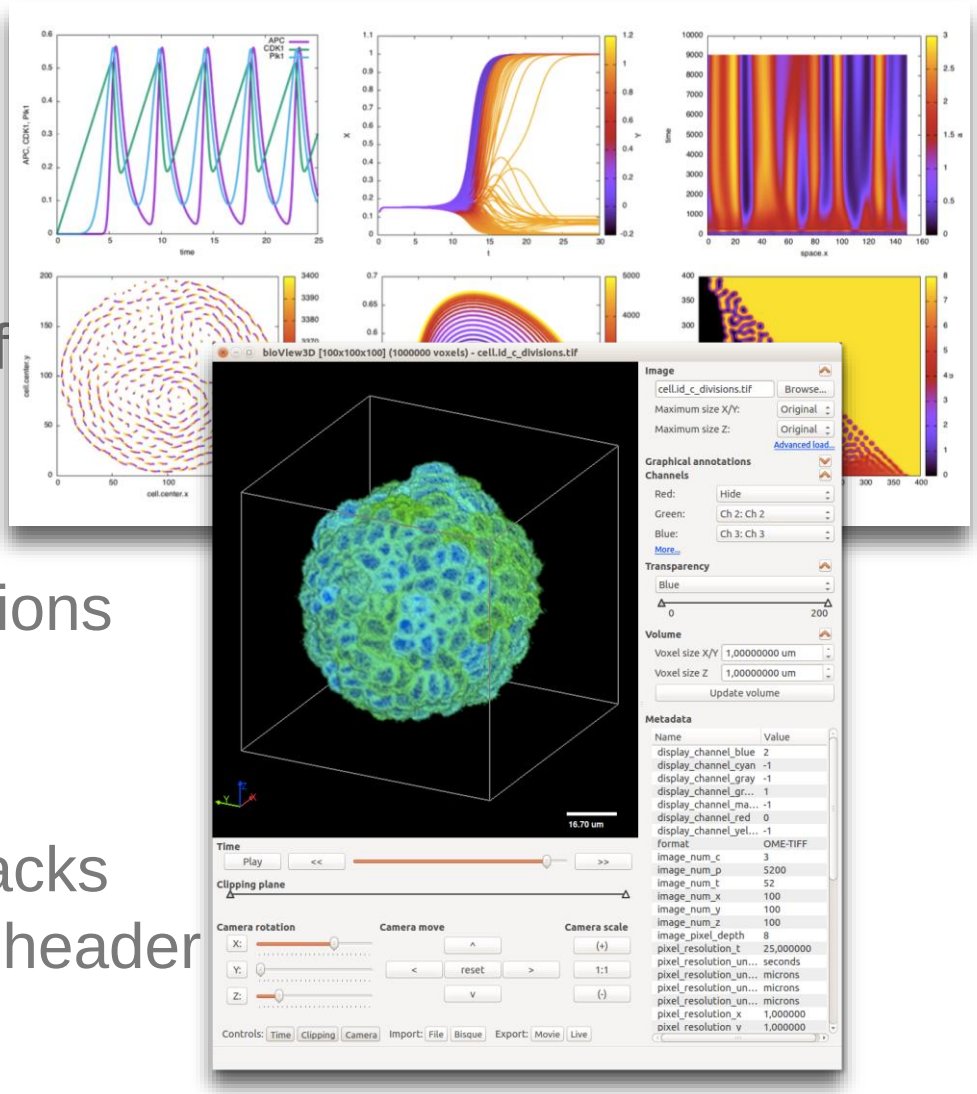
Data export in csv or matrix f
and versatile tool for plotting

HistogramLogger

Compute frequency distributions
and visualization

TIFFPlotter

Export multichannel TIFF stacks
with OME (open microsopy) header



Morpheus in education
Teaching mathematical and computational modeling

Lecture course

Introduction to mathematical biology

TU Dresden

Practical course

Pattern formation in tissues

DIGS-BB (biomedicine and bioengineering)

Summer schools

Math. modeling of tissue mechanics

Euro. Mathematical Society (EMS) and ESMTB

Multiscale Biology

UK Multi-scale biology network, Nottingham

OpenMultiMed Training School in Erlangen

Friedrich Alexander University Erlangen-Nürnberg

Workshops

Computational stem cell biology

German Stem Cell Network (GSCN), and at Helmholtz Centres in Munich, Braunschweig

Multi-scale modeling of biological systems

National University La Plata, Buenos Aires

Code Repository

gitlab.com/morpheus.lab

Git repository
distributed version control

BSD License
permissive open-source

Issue tracker
bugs and features

Documentation
plugin development

Community
distributed development