Model repositories and standard formats for model reusability

Dagmar Waltemath

Medical Informatics Laboratory, University Medicine of Greifswald

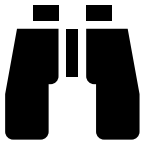
New Trends in Bioinformatics | September 1, 2020
T05: Computational modelling of cellular processes: regulatory vs metabolic systems

DIE DEUTSCHEN UNIVERSITÄTSKLINIKA®
Wir sind Spitzenmedizin

Dagmar Waltemath
Medical Informatics Laboratory
University Medicine Greifswald
dagmarwaltemath
0000-0002-5886-5563

Slides published on slideshare under CC BY SA

How this tutorial is going to work



Links on slides (get them via slideshare)
Ask me now or later

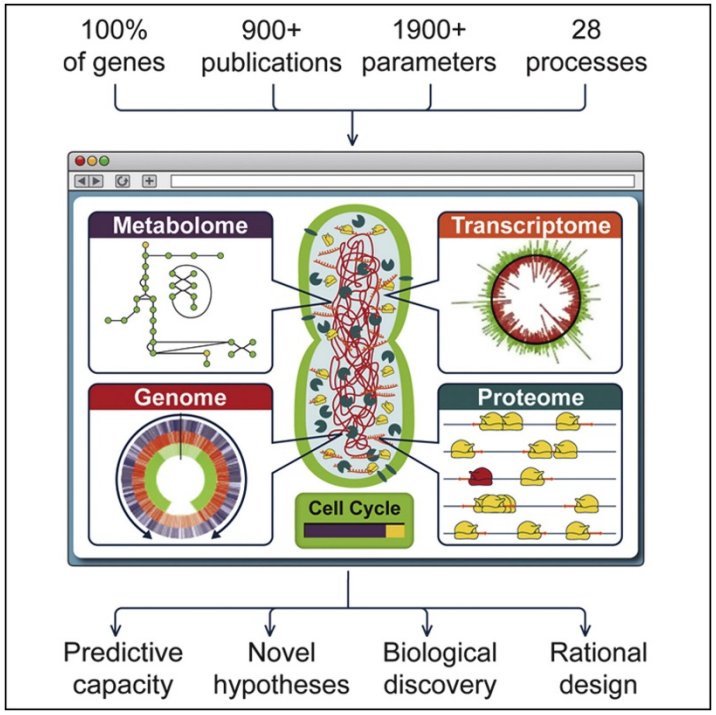
Ping me later or search online ☺

Anytime!

Anytime!

Disclaimer: Comic-style graphics in this presentation were done either by Anna Zhukova or by Martin Peters. Thank you very much!

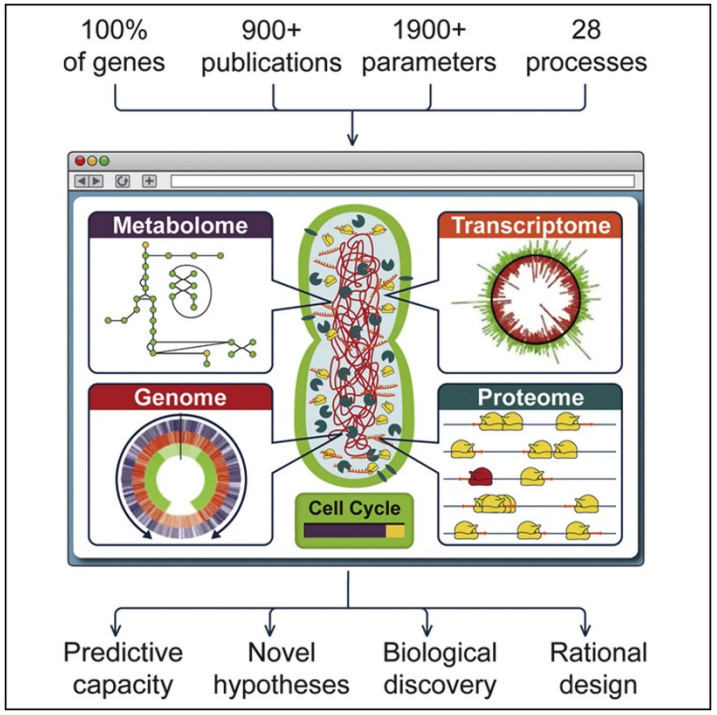
Example: Whole Cell Modeling



Cell 2012 150389-401 DOI: (10.1016/j.cell.2012.05.044)
Copyright © 2012 Elsevier Inc. Terms and Conditions

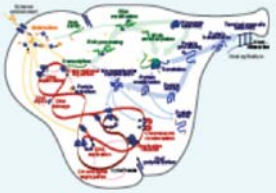
- First in silico Whole Cell Model
- Genome (525 genes), transcriptome, proteome and metabolome incorporated
- Describes whole life cycle of a single cell on molecular level, and
- predicts a wide range of cellular behaviors, and
- accounts for the specific function of every annotated gene product
- Based on 900+ publications
- Consists of 116 MATLAB files
- Incorporates over 1.900 experimentally observed parameters

All data and processing are available online, for download or one-stop exploration.



Cell 2012 150389-401 DOI: (10.1016/j.cell.2012.05.044)
Copyright © 2012 Elsevier Inc. Terms and Conditions

Mycoplasma genitalium



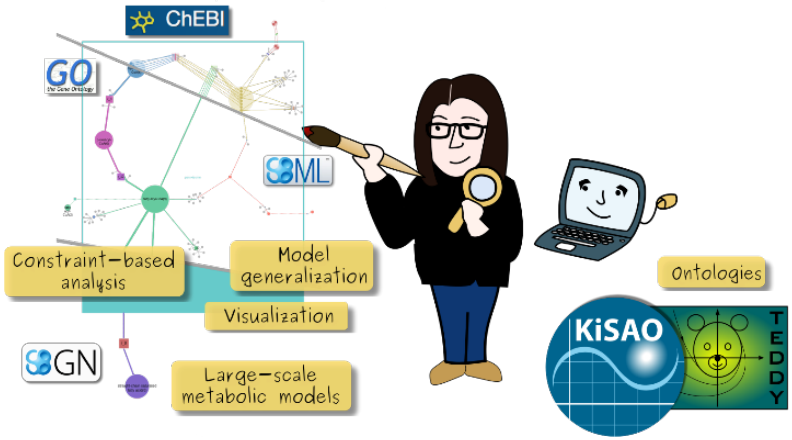
The *M. genitalium* WC model was the first model that represented each characterized gene function of an organism. The model is composed of 28 submodels of 28 cellular processes. In total, the model represents the functions of 401 genes, 722 compounds, 1,857 reactions, and 1,836 parameters. In addition to demonstrating the feasibility of WC models, the model has been used to gain new insights into cell cycle regulation, learn unknown parameters, and suggest new uses of existing antibiotics for *Mycoplasmas*.

Availability: Download
Author: Covert Lab, Stanford
More info: Docs | Source | License | Tests
Reference: Karr JR et al. *Cell* 2012.

https://www.wholecell.org/

Key message

If your work is available, documented and open…



…then others can host it,
share it, cite it, and advertise it!

Biology simulation studies can comprise of many data items.

Original publication

COMBINE Archive

Article

Open Access

Dynamical modeling of syncytial mitotic cycles in *Drosophila* embryos

Laurence Calzone, Denis Thieffry, John J Tyson, Bela Novak

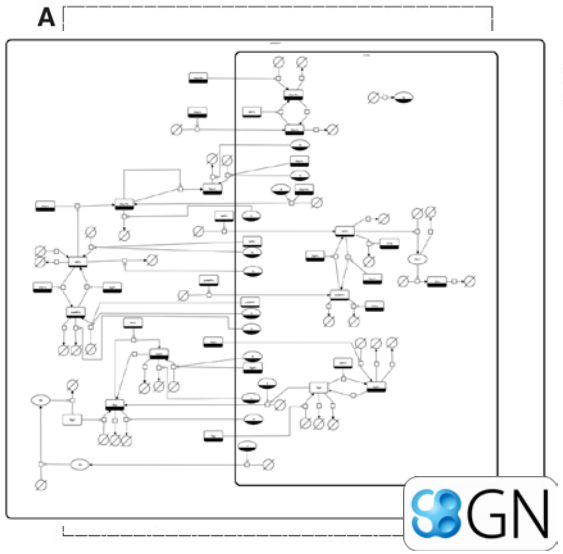
Author Affiliations

DOI 10.1038/msb4100171 | Published online 31.07.2007
Molecular Systems Biology (2007) 3, 131

Doi / PLOS

Versions, multi-file models and multi-site development add to the complexity!

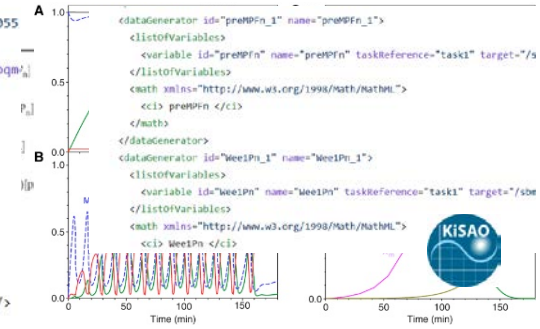
Model IDs




```
</listOfCompartments>
<listOfSpecies>
  <species id="MPFc" initialConcentration="1" name="MPFc" metaid="metaid_0000055
    <annotation>
      <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqm
        <rdf:Description rdf:about="#metaid_0000055">
          <bqbiol:hasPart>
            <rdf:Bag>
              <rdf:li rdf:resource="http://identifiers.org/uniprot/P23572"/>
              <rdf:li rdf:resource="http://identifiers.org/uniprot/P20439"/>
            </rdf:Bag>
          </bqbiol:hasPart>
          <bqbiol:isVersionOf>
            <rdf:Bag>
              <rdf:li rdf:resource="http://identifiers.org/obo.go/GO:0031387"/>
            </rdf:Bag>
          </bqbiol:isVersionOf>
```

SBO ChEBI GO the Gene Ontology

SBML CellML SBOL



SED ML

Visualisation

Model encoding

Simulation encoding

Source & full example: Scharm & Waltemath (2016) Fully featured archives

The Systems Biology Markup Language is a widely-understood model exchange format for many modeling types.

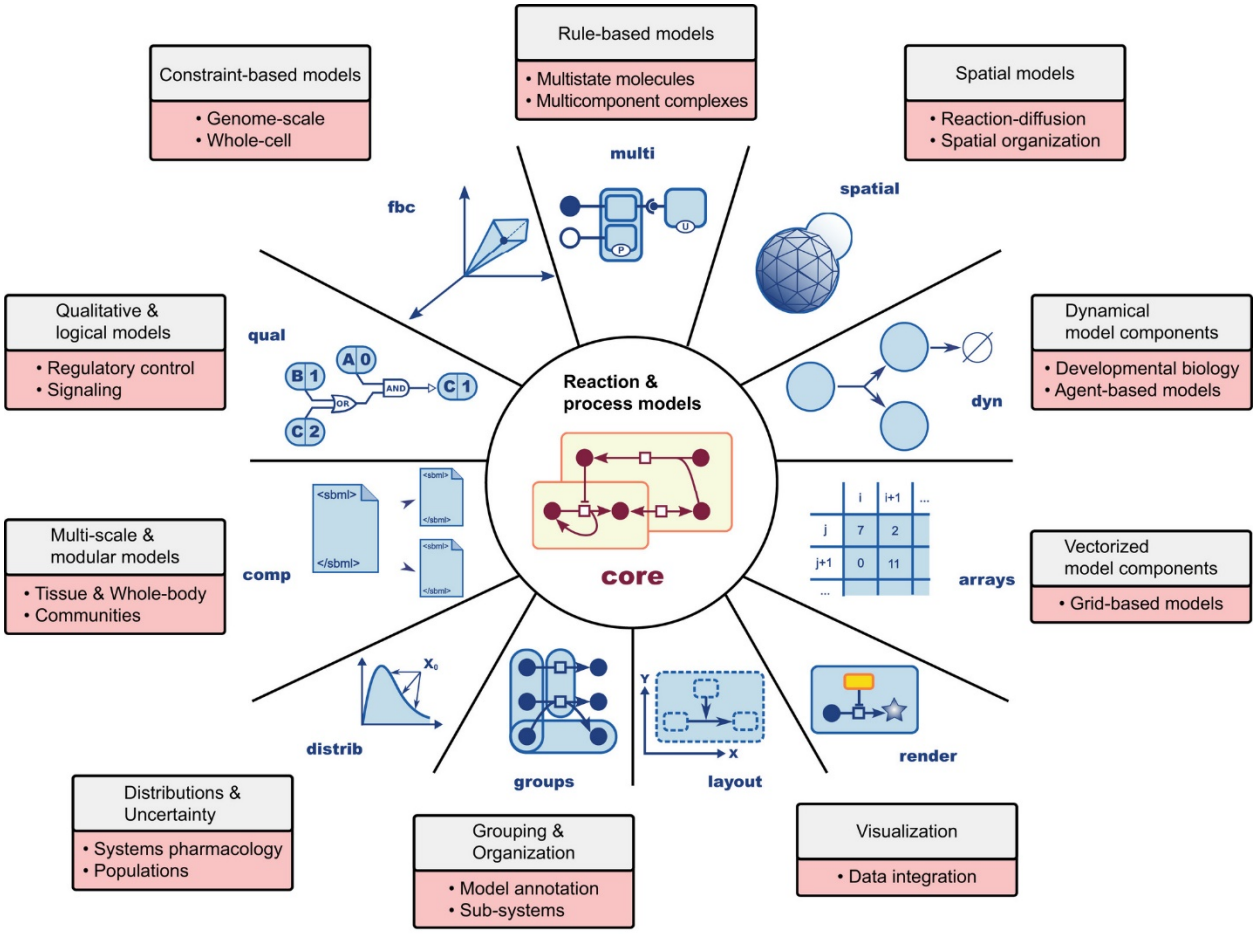


Fig: Keating (2020) SBML Level 3 Molecular Systems Biology

What does your data look like?

Photo by Jan Antonin Kolar on Unsplash

Follow the FAIR guiding principles when aiming to provide reusable simulation models.

- Identifiable data items
- Persistent
- Searchable

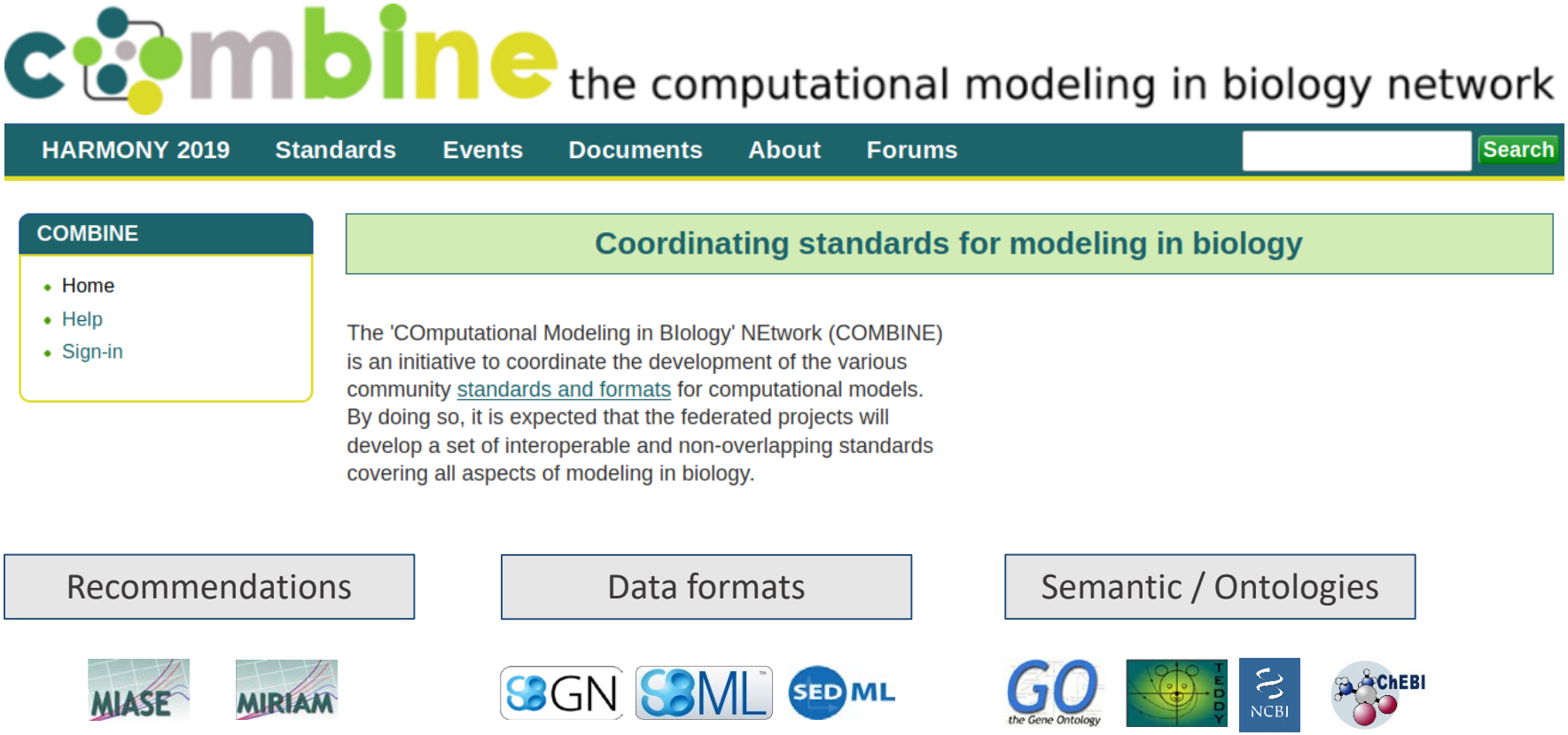
- Identifiers following standard protocols
- Authentification
- Access to meta data, even if data not accessible

- Formal, accesssible representation of data
- Qualified references

- Licensing
- Provenance
- Standards compliance

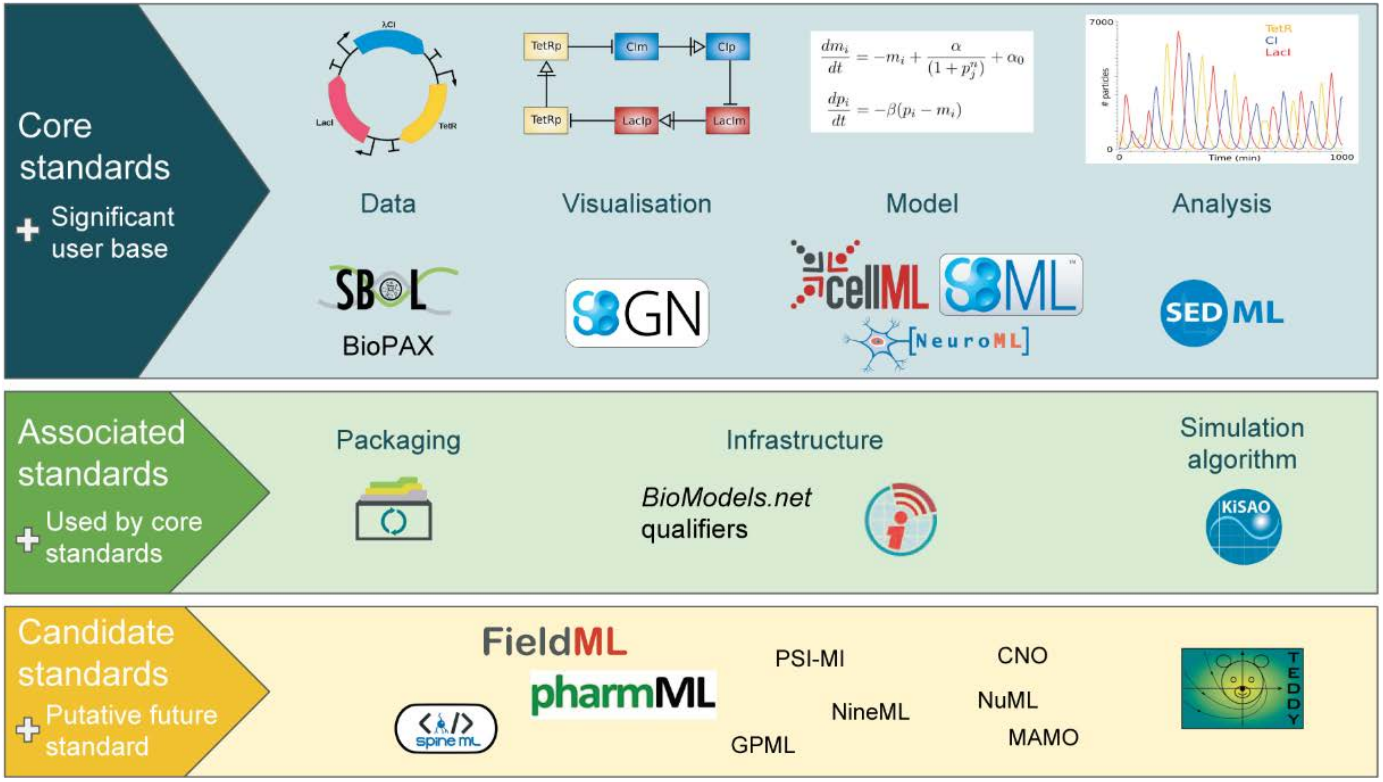
https://fairsharing.org/ | https://github.com/FAIRMetrics | Wilkinson et al (2016) FAIR guiding principles

Use community standards when publishing and exchanging model code.



https://co.mbine.org/ | Hucka et al (2015) COMBINE initiative | Waltemath et al (2020) COMBINE after 10 yrs

Check out our online resources to learn how to make your model code interoperable – and how you can get involved in development.



- Editorial Boards
- Software tool support
- Annual meetings
- Mailing lists
- Specifications

http://co.mbine.org/standards | https://co.mbine.org/tools | Fig.: Schreiber et al (2019) Latest COMBINE Specifications

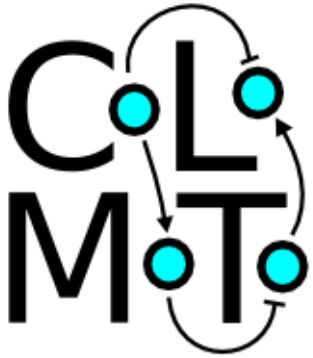
You can use standards-compliant models with many open libraries & tools.

Modeling & Simulation Software

tellurium

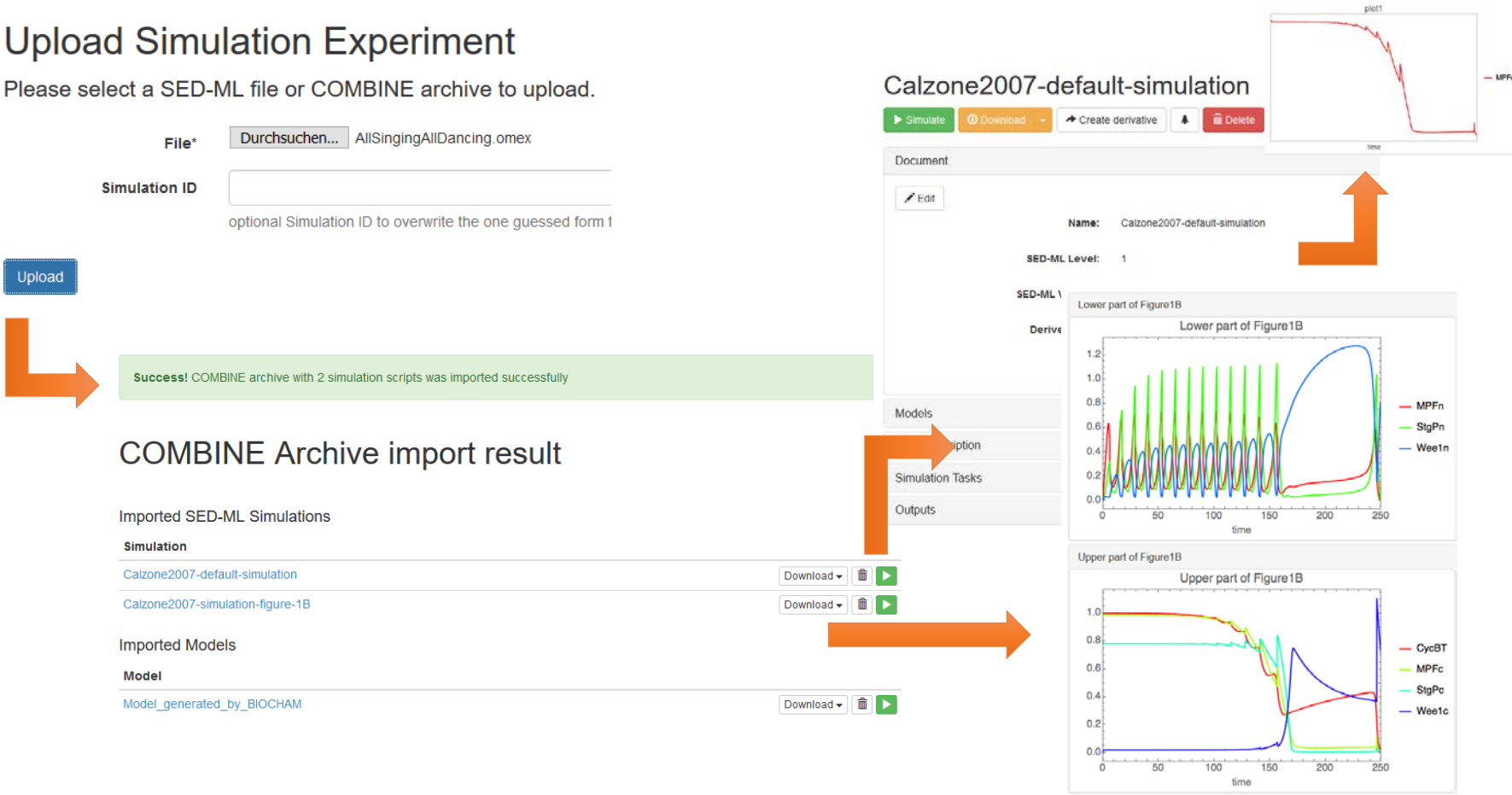
Repositories & Management tools

...



Hucka (2016) SBML software portfolio | Colomoto Consortium

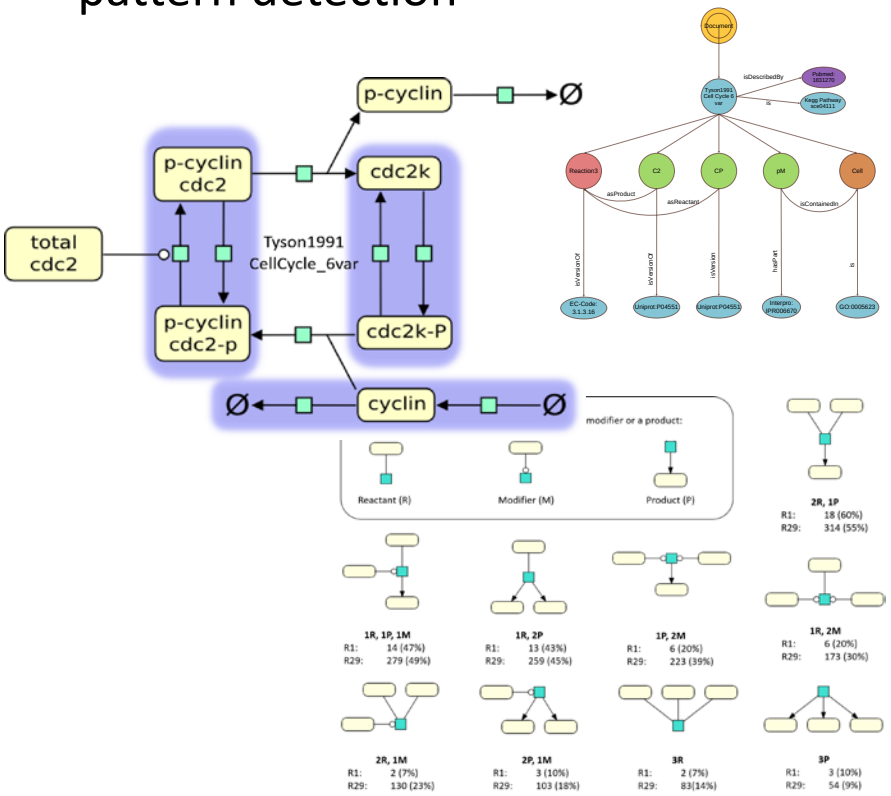
Automated tool chains for data and model reuse make your life easier.



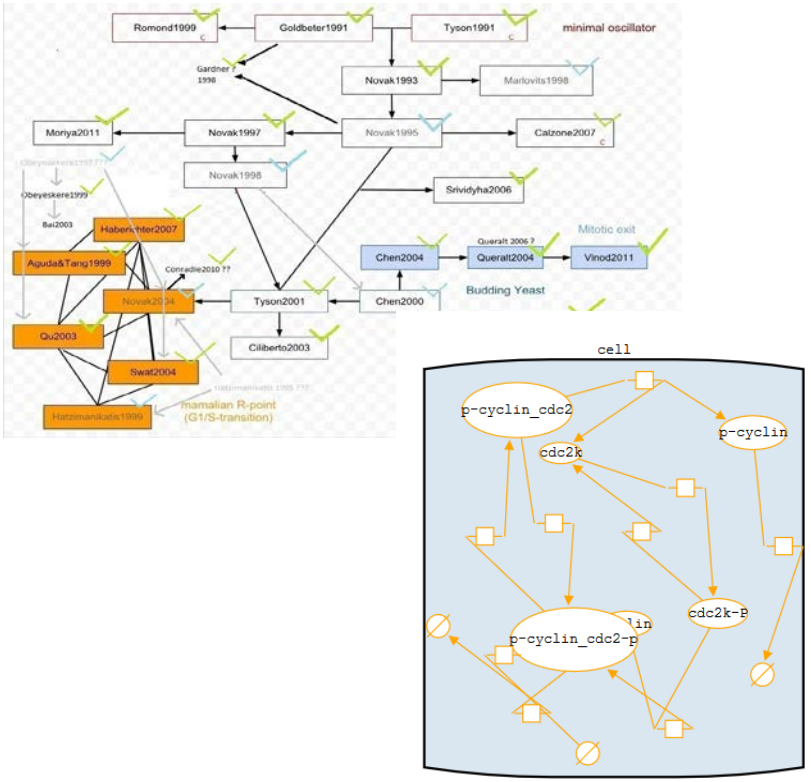
Scharm & Waltemath (2015) BTW [M2CAT] | Loading models into [JWS Online simulator]

The community develops new methods and algorithms for model management.

Similarity measures & pattern detection



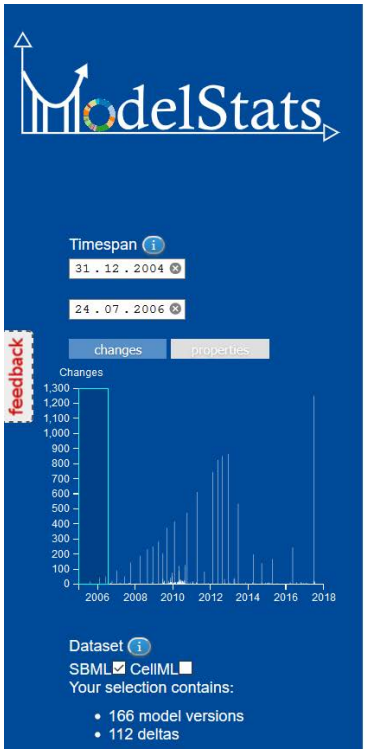
Phylogeny & version control

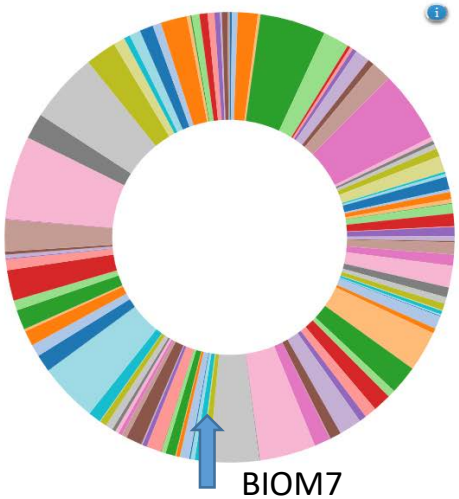


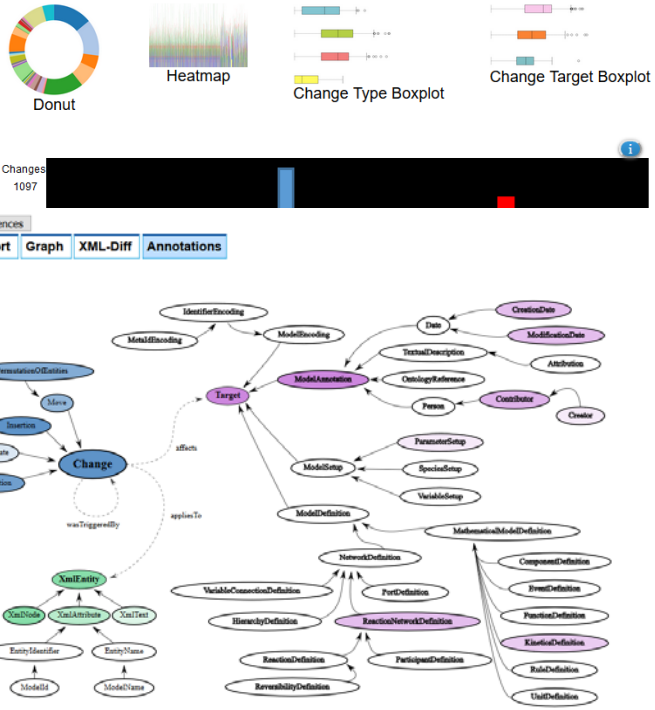
Henkel et al (2015) Graph-based storage | Lambusch et al (2018) Pattern detection | Scharm et al (2016) model version control

The community develops new methods and algorithms for model management.

Statistics on model sizes & changes





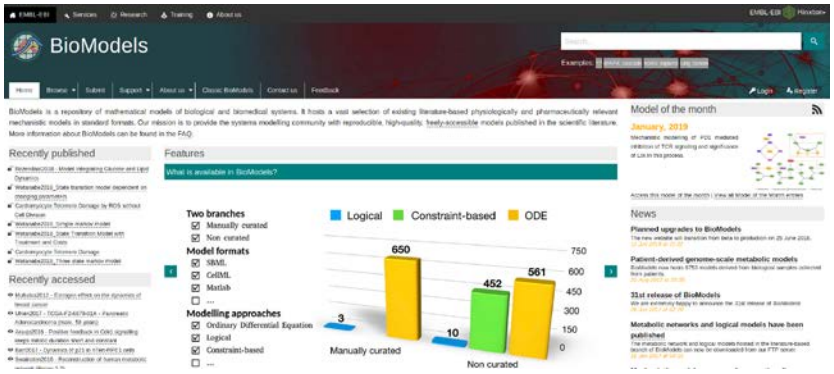


Try it out at: https://most.bio.informatik.uni-rostock.de

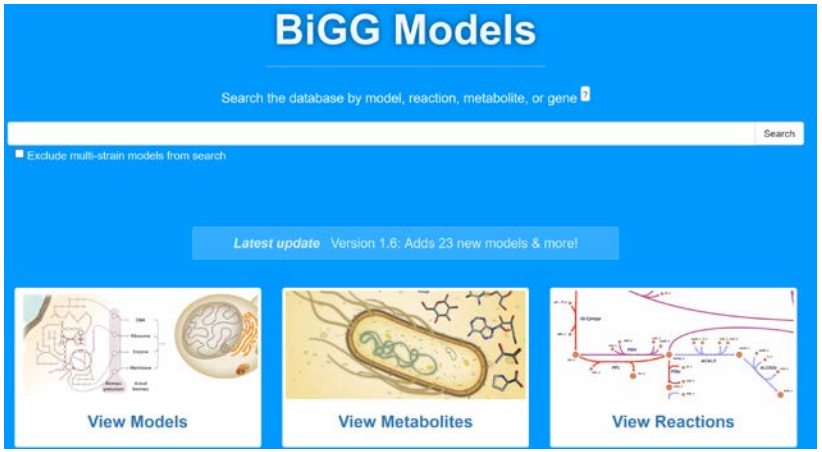
And where is your data?

Photo by Warren Wong on Unsplash

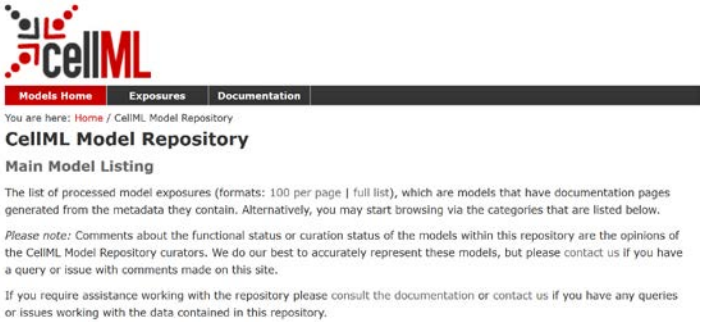
Different open model repositories give you access to curated models.



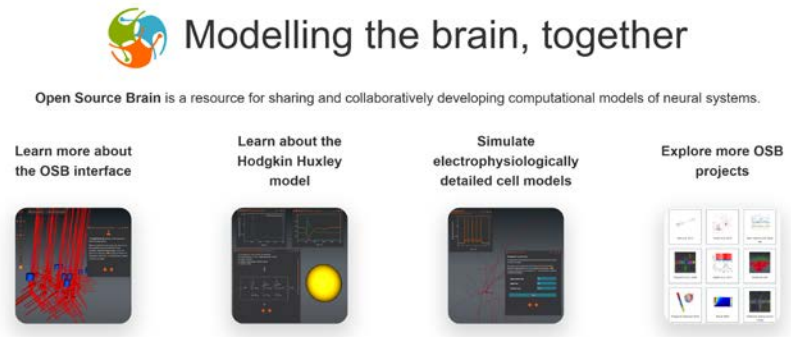
https://www.ebi.ac.uk/biomodels/



http://bigg.ucsd.edu/



https://models.cellml.org/cellml



http://www.opensourcebrain.org/

Different open model repositories give you access to curated models.

BioModels

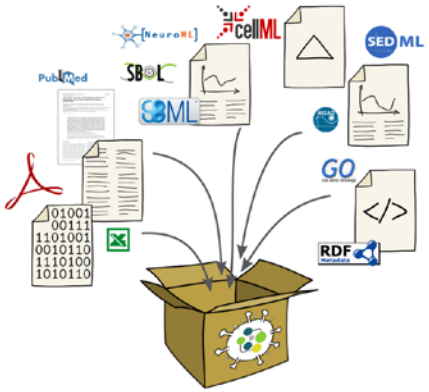
Individual Models | Search...

What is this box used for? | Examples: MAPK cascade, homo sapiens, lung cancer | Search tips

Home | Browse | Submit | Support | About us | Contact us | Feedback | Login | Register

Reproducible simulation studies targeting COVID-19

In a collaboration between the University of Greifswald, the Humboldt-University Berlin, code ahoi, and the BioModels database at EMBL-EBI, we aim to rapidly disseminate simulation studies of COVID-19 models to the research community, in interoperable formats and in high quality.



List of curated models

BIOMD0000000955	Giordano2020 - Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy
BIOMD0000000956	Bertozzi2020 - The challenges of modeling and forecasting the spread of COVID-19
BIOMD0000000957	Roda2020 - Why is it difficult to accurately predict the COVID-19 epidemic?
BIOMD0000000958	Ndairou2020 - Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan
BIOMD0000000960	Paiva2020 - A data-driven model to describe and forecast the dynamics of COVID-19 transmission
BIOMD0000000962	Zhao2020 - Modeling the epidemic dynamics and control of COVID-19 outbreak in China

EOSC secretariat.eu

https://www.ebi.ac.uk/biomodels/covid-19 (through COVID-19 fast track funding)

You can follow defined steps to make your models better reusable.

1. Decide for a repository
2. Check accepted formats and upload procedures
3. Prepare your model (use standards, annotate, document, provide an archive containing the files to reproduce all virtual studies described in the paper)
4. Upload your model and (Cross fingers for positive curation results)
5. Receive URI for further reference

Waltemath et al (2013) Reproducibility of Model-Based Results in Systems Biology (or ask me)

More help needed? Check out online guidelines!

10 tips for building useful SBGN maps

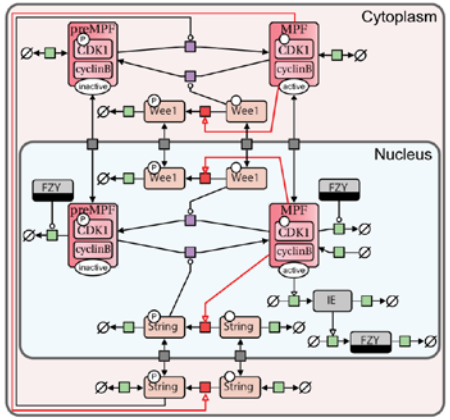
EDUCATION

Quick tips for creating effective and impactful biological pathways using the Systems Biology Graphical Notation

Vasundra Touré[1]*¤, Nicolas Le Novère[2], Dagmar Waltemath[1], Olaf Wolkenhauer[1,3]

1 Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany, 2 Babraham Institute, Babraham Research Campus, Cambridge, United Kingdom, 3 Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre, Stellenbosch, South Africa

¤ Current address: Department of Biology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
* vasundra.toure@ntnu.no



Building fully featured COMBINE archives

F1000Research F1000Research 2016, 5:2421 Last updated: 11 OCT 2016

DATA NOTE

A fully featured COMBINE archive of a simulation study on syncytial mitotic cycles in *Drosophila* embryos [version 1; referees: 1 approved, 2 approved with reservations]

Martin Scharm, Dagmar Waltemath
Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany

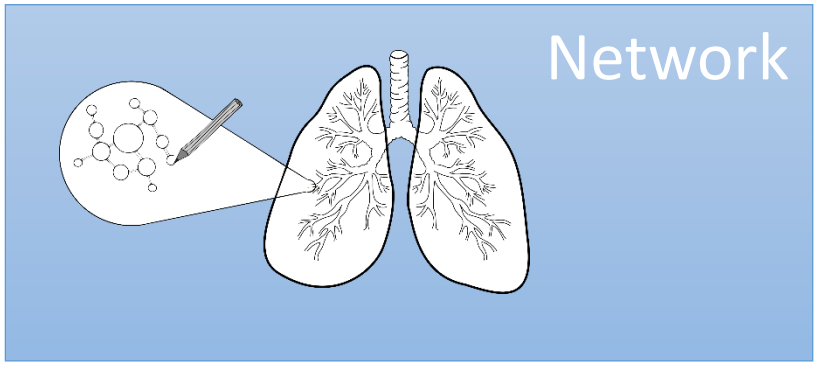
File	Format	Description
manifest.xml	Omex	Skeleton, automatically generated by WebCAT
metadata.rdf	Omex	Skeleton, automatically generated by WebCAT
README.md	Markdown	Human readable information for users stumbling upon the archive
model/		
BIOMD0000000144.xml	SBML L2V1	origin: www.ebi.ac.uk/biomodels-main/download?mid=BIOMD0000000144
calzone_2007.svg	SVG	origin: models.celiml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_2007.ai	Illustrator	origin: models.celiml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_2007.png	PNG	origin: models.celiml.org/workspace/calzone_thieffry_tyson_novak_2007
calzone_thieffry_tyson_novak_2007.celiml	CellML 1.0	origin: models.celiml.org/workspace/calzone_thieffry_tyson_novak_2007
sbgn/Calzone2007.gml	GML	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.graphml	GraphML	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.pdf	PDF	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.png	PNG	SBGN compliant figure generated using SBGN-ED
sbgn/Calzone2007.sbgn	SBGN-ML	SBGN-ML encoded figure generated using SBGN-ED
experiment/		
Calzone2007-default-simulation.xml	SED-ML L1V1	Simulation description generated using SED-ML Web Tools
Calzone2007-simulation-figure-1B.xml	SED-ML L1V1	Simulation description generated using SED-ML Web Tools based on Calzone2007-default-simulation.xml
documentation/		
Calzone2007.pdf	PDF	Scientific publication *"Dynamical modeling of syncytial mitotic cycles in Drosophila embryos"* obtained from msb.embopress.org/content/3/1/131
Calzone2007-supplementary-material.pdf	PDF	Supplementary information for the publication obtained from msb.embopress.org/content/3/1/131
result/		
Fig1B-bottom-COPASI.svg	SVG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in COPASI
Fig1B-top-COPASI.svg	SVG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in COPASI
Fig1B-bottom-webtools.png	PNG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in SED-ML Web Tools
Fig1B-top-webtools.png	PNG	Image generated by executing Calzone2007-simulation-figure-1B.xml on BIOMD0000000144.xml in SED-ML Web Tools

Even more help needed?

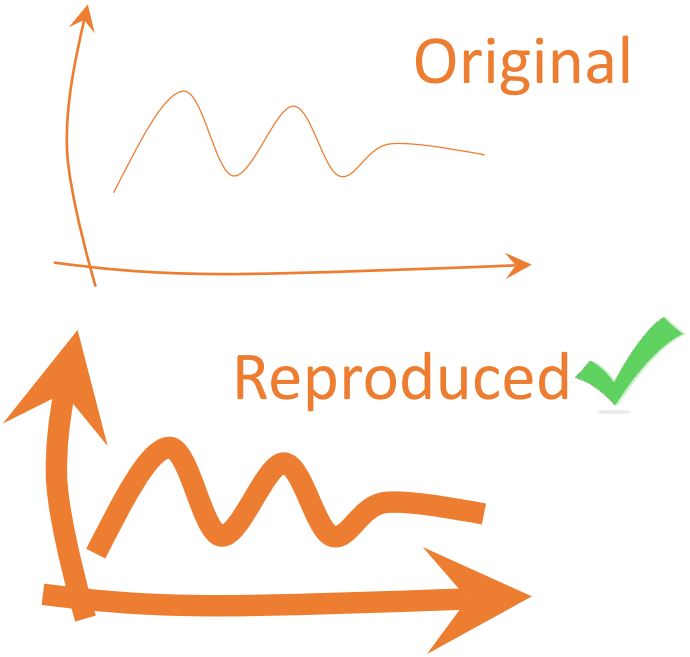
http://co.mbine.org/

@combine_coord

Summary

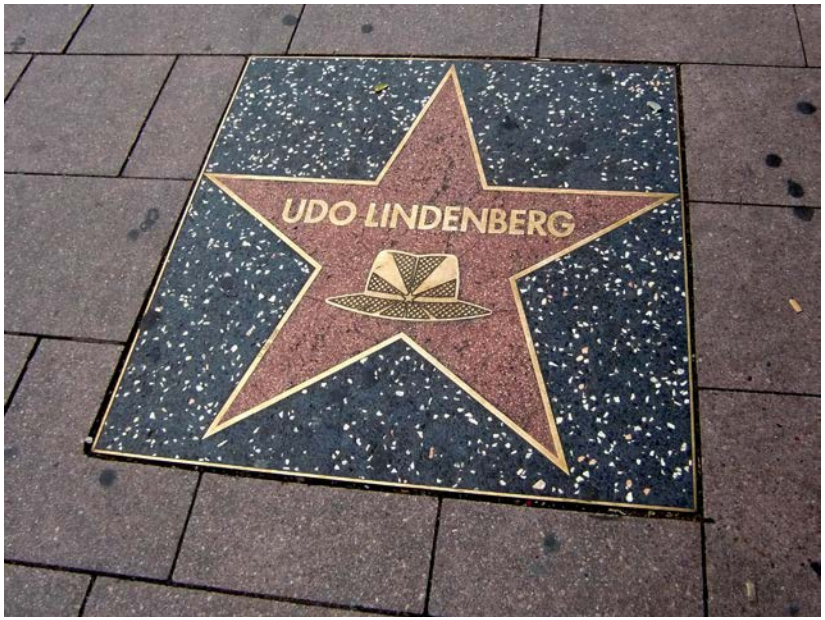


COMBINE Archive



(Animation courtesy T. Gebhardt, Human Disease 2016)

Be recognised! Be a good scientist!



- Standardise your model code.
- Make your simulation studies openly available.
- Reuse other people‘s work.
- Join the COMBINE community to contribute to the projects you saw on my slides (and many more).

COMBINE 2020 - ONLINE FORUM

Dates: ***5-9 October 2020***
Location: ***Online (virtual) meeting***

Virtual & free of charge

Call for contributions open until Sep 8

Fig: Waltemath et al. (2020) Meeting report , JIB