



Extracting reproducible simulation studies from model repositories using the CombineArchive Toolkit

Martin Scharm and Dagmar Waltemat





Pedro Mendes @gepasi · 1h
to read a computational paper without access to the software is an act of faith -
you either believe it or not, no chance to check it

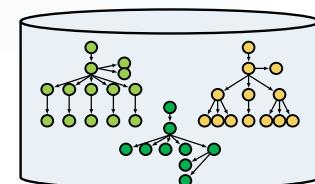
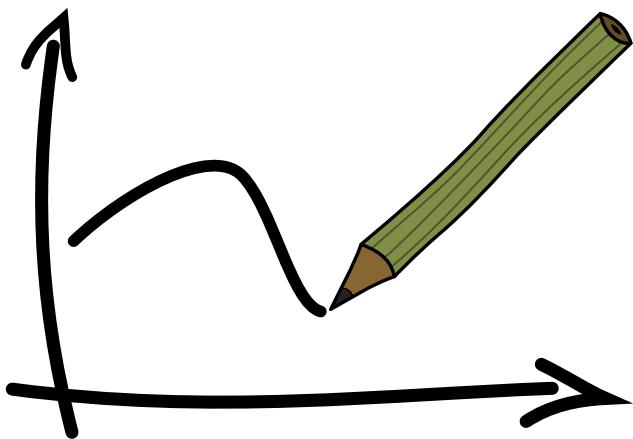
◀ ⬆⬇ 1 ★ ...



Reproducible simulation studies in computational biology

Pedro Mendes @gepasi · 1h
to read a computational paper without access to the software is an act of faith -
you either believe it or not, no chance to check it

◀ ↕ 1 ⋆ ⋮



MASYMOS



CAT



The need for better reproducibility of simulation studies

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

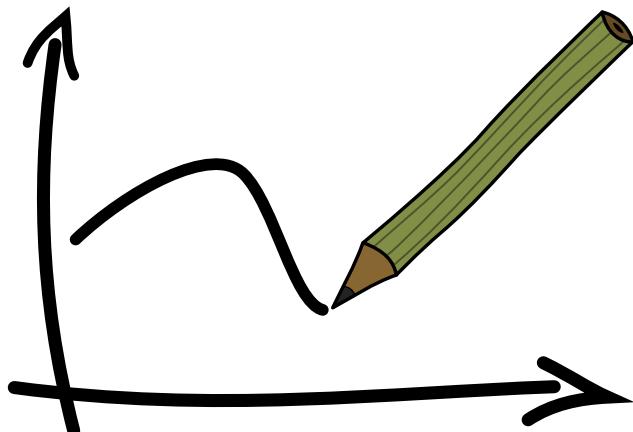
Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

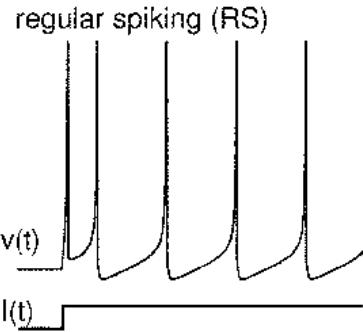
Fifty-three papers were deemed 'landmark' studies [...]. It was acknowledged from the outset that some of the data might not hold up, because papers were deliberately selected that described something completely new, such as fresh approaches to targeting cancers or alternative clinical uses for existing therapeutics. Nevertheless, scientific findings were **confirmed in only 6 (11%) cases**.

http://www.nature.com/nature/journal/v483/n7391/fig_tab/483531a_T1.html

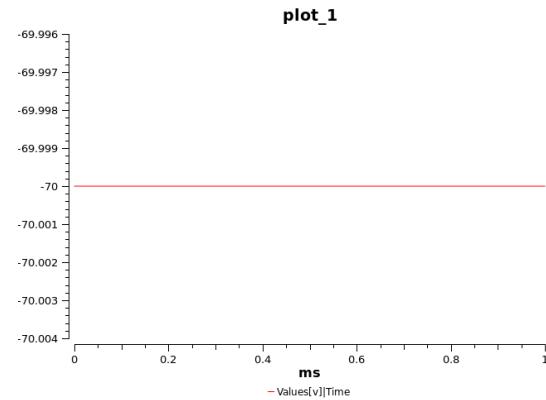


Ongoing initiatives

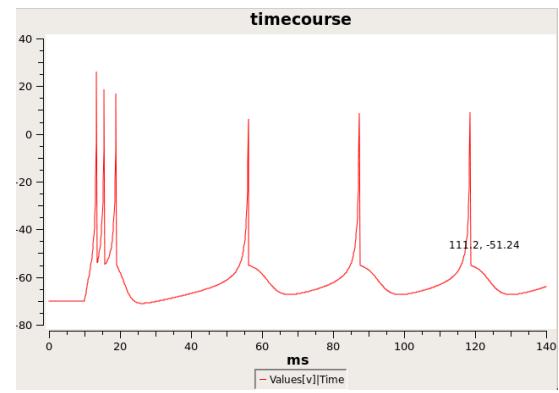
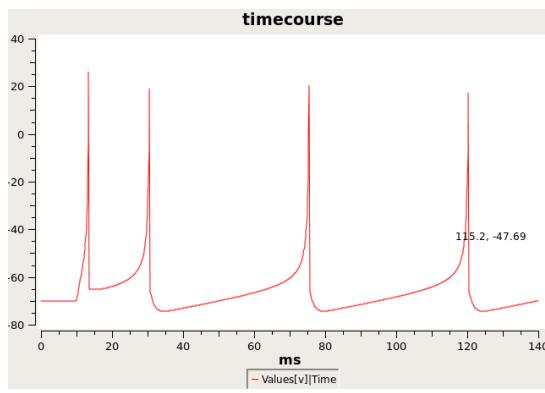
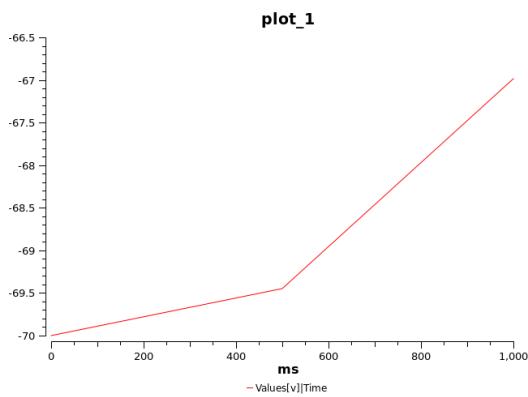
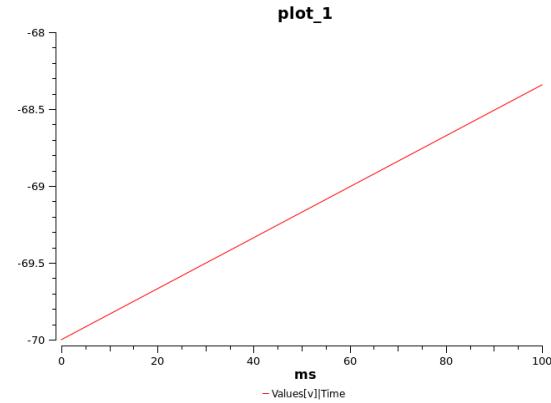




Reported result (BM127)



1 ms (standard, COPASI)



The whole cell model (900+ publications, 1900+ parameters)

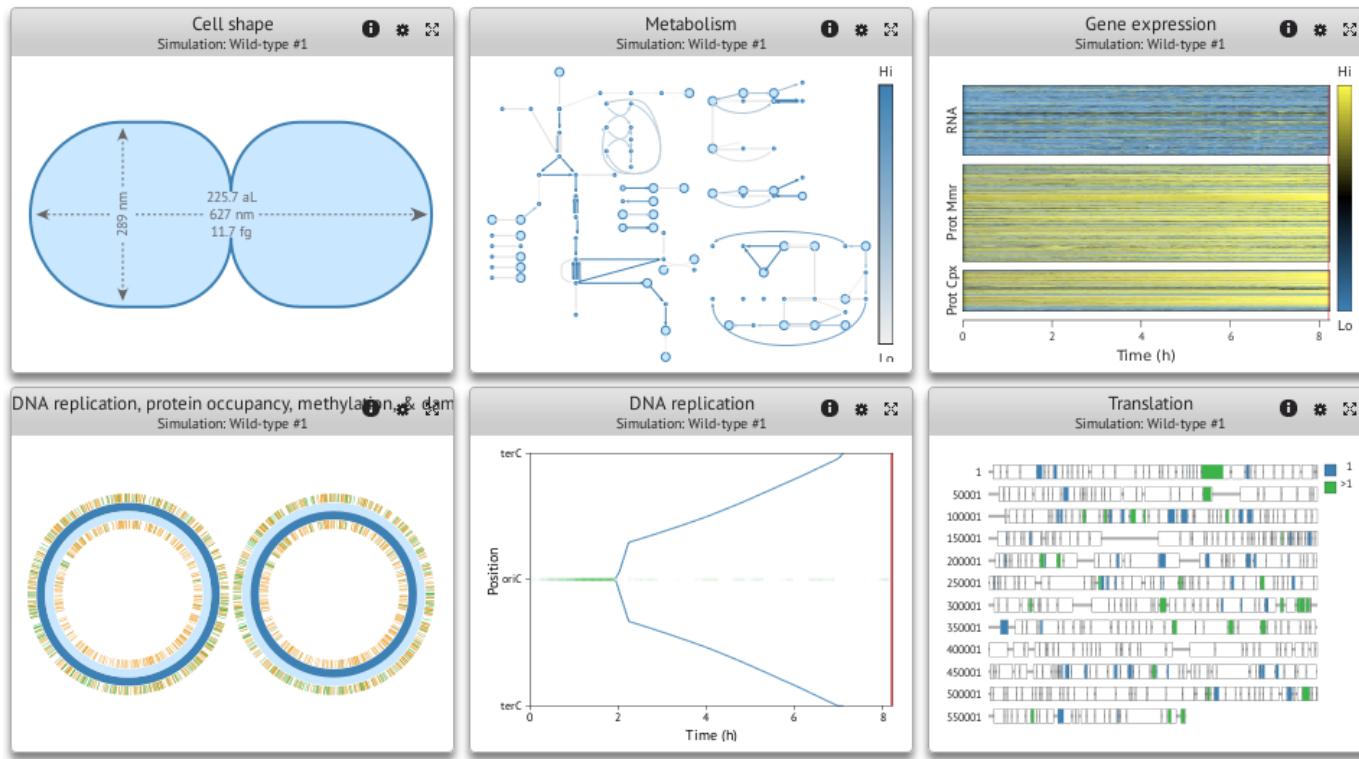


Fig.: <http://wholecellviz.stanford.edu/viz.php#cellCycle>



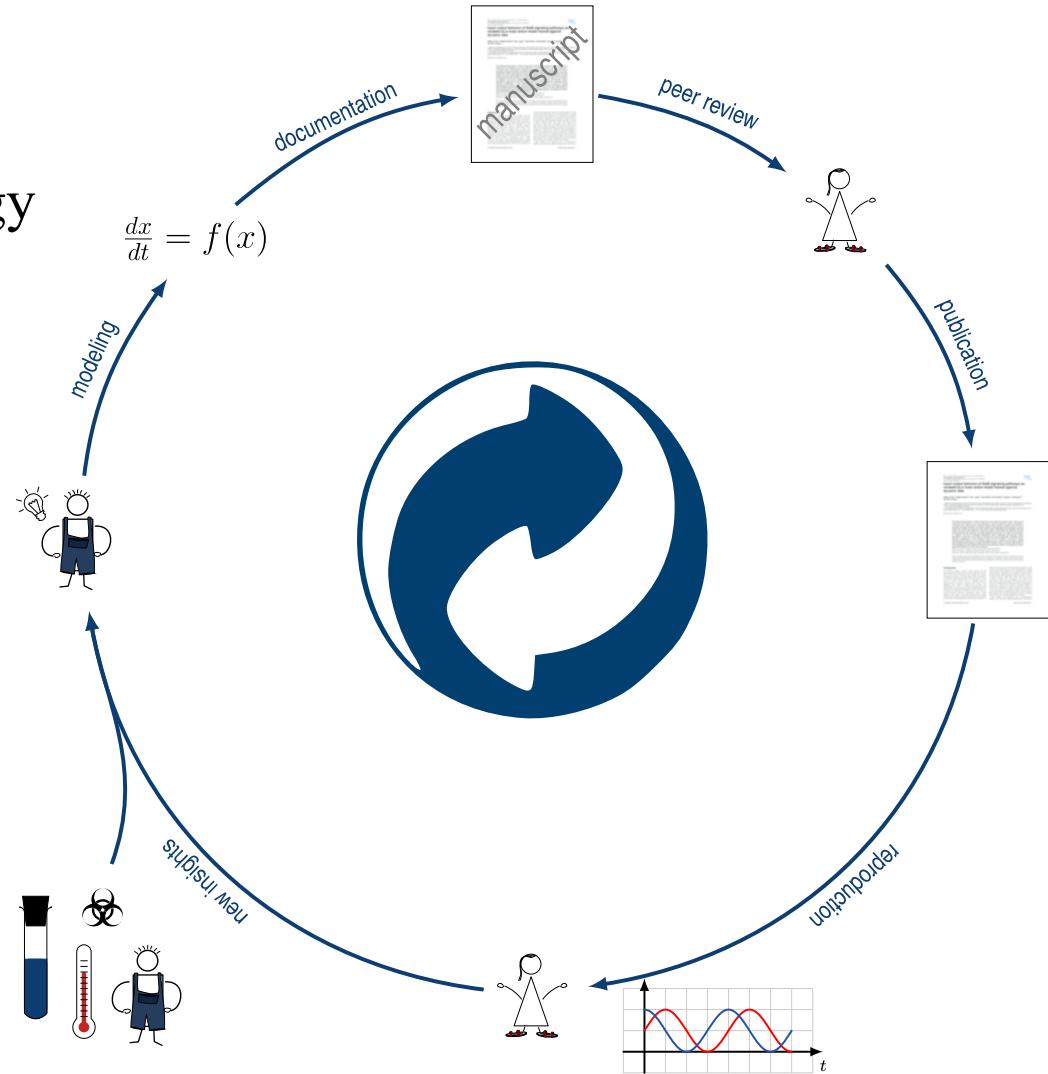
SEMS investigates...

simulation experiment management system

- Development of standard formats for data exchange
- Suitable concepts for storage, linking and retrieval of model-related data and meta-data

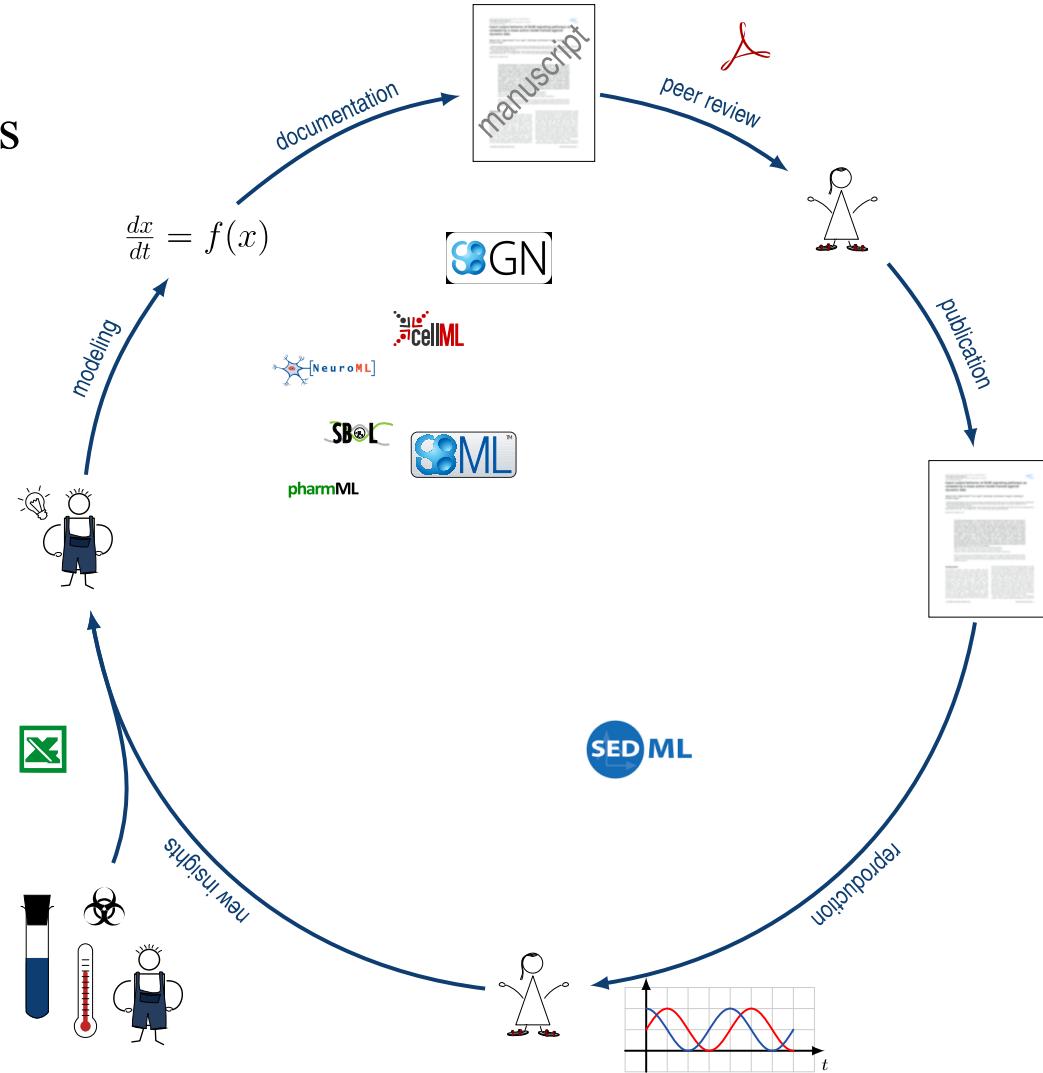


Model-related data in the systems biology workflow

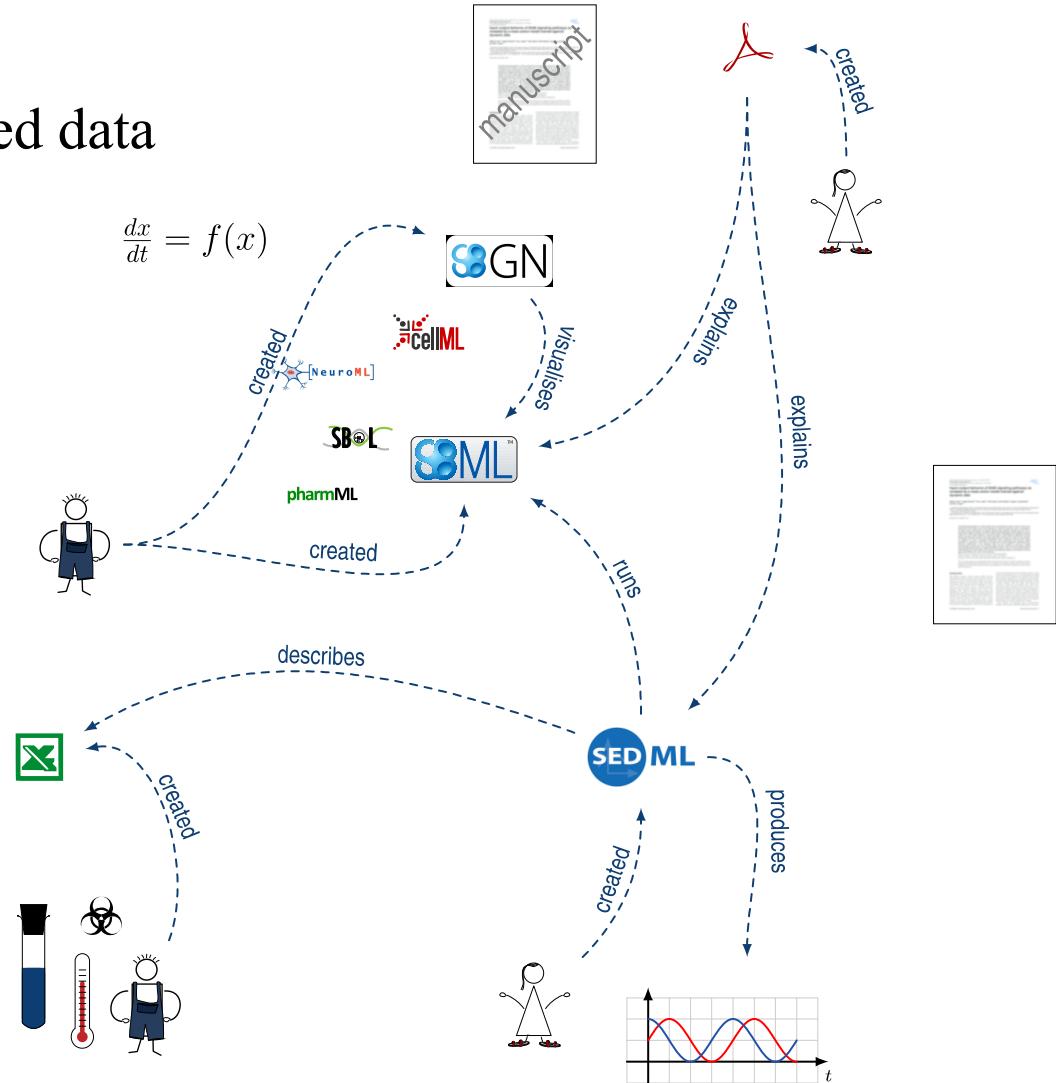




combine standards



Linking model-related data



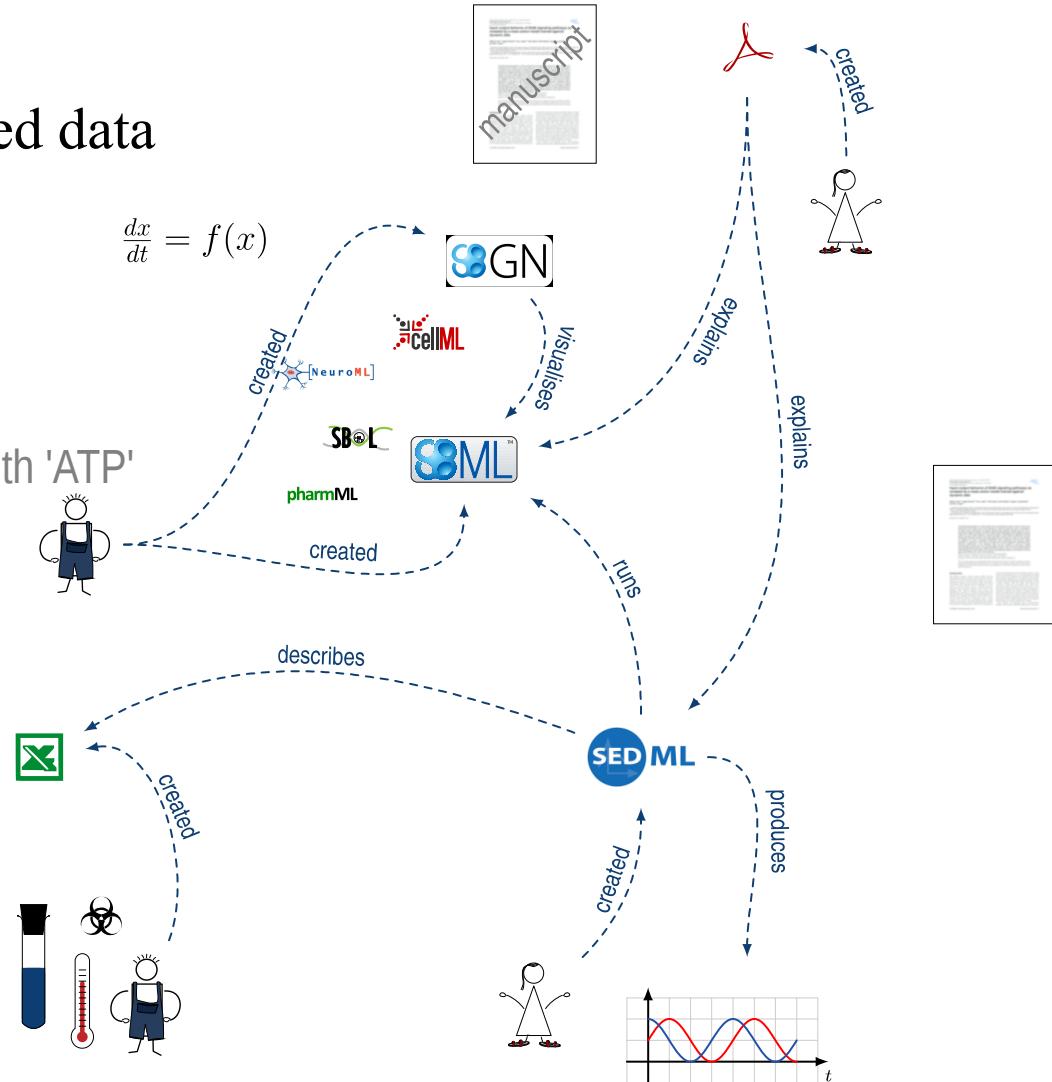
Linking model-related data

Which are the most frequently used GO annotations in my model set?
(Alm et al 2014)

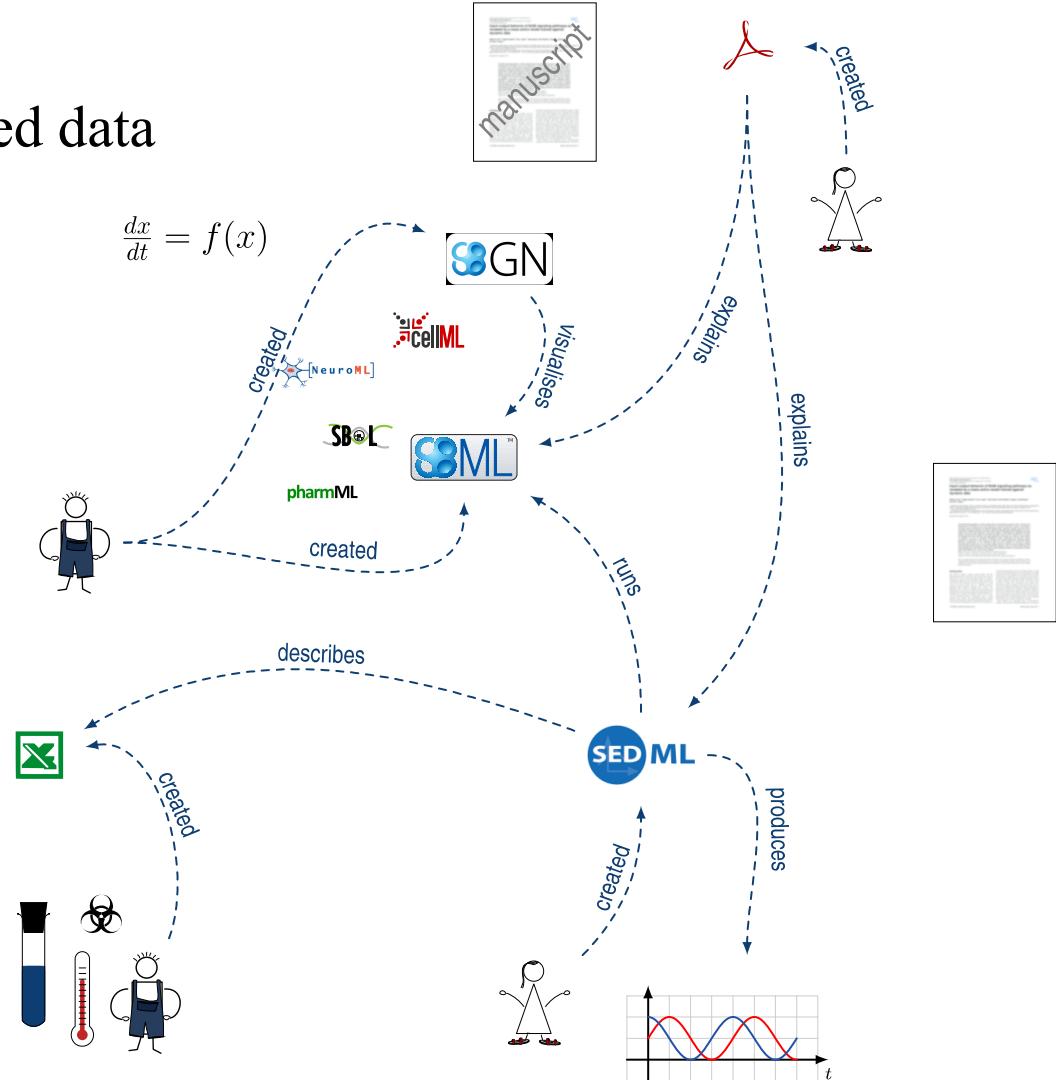
Which models contain reactions with 'ATP' as reactant and 'ADP' as product?
(Henkel et al 2015)

Give me all the files I need to run this simulation study.

Find good candidates for features describing my set of models. (Alm et al 2015)



Linking model-related data



Give me all the files I need to run this simulation study.

MASYMOS

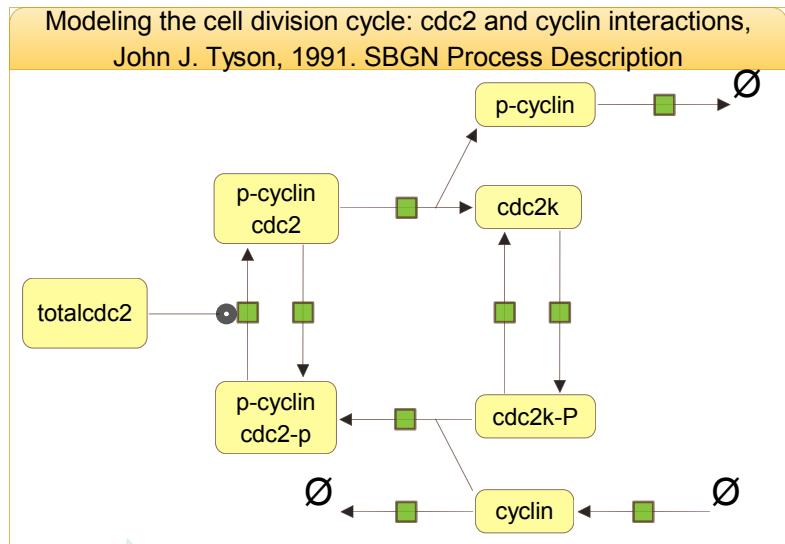
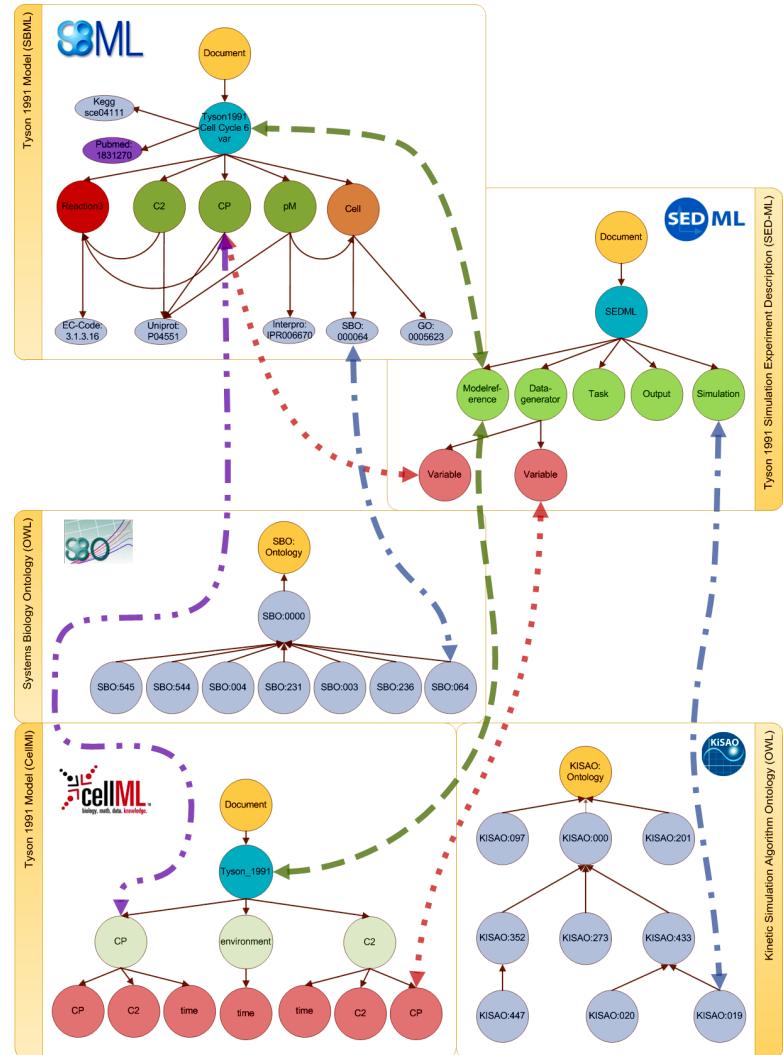
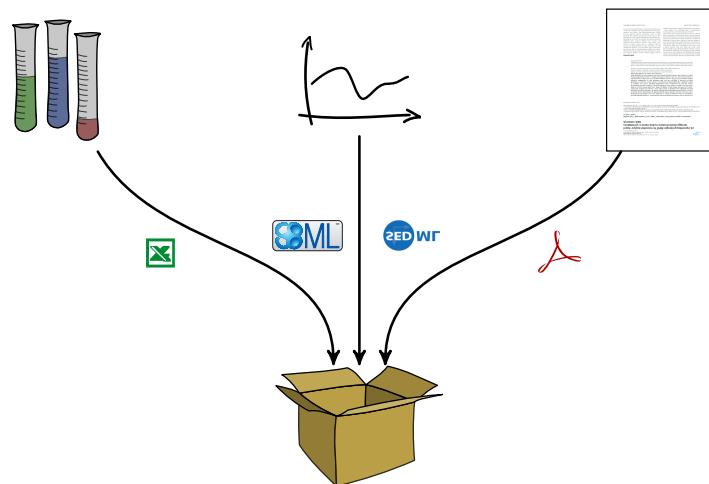


Fig: Henkel et al (2015)





combine Archive (Toolkit)



Software

Highly accessed

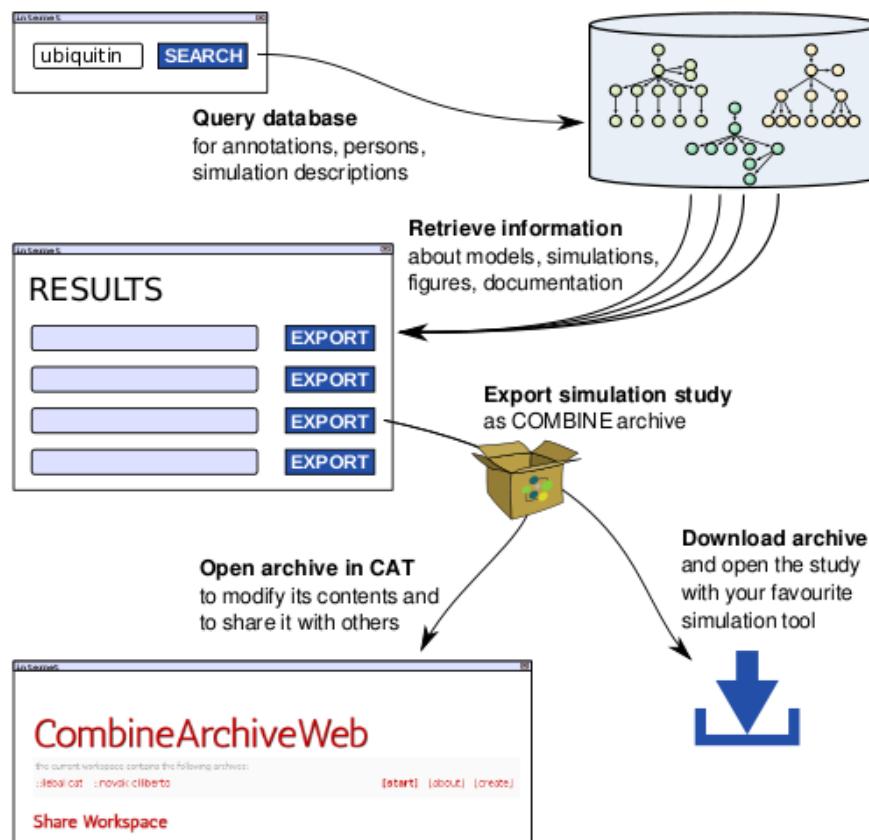
Open Access

**COMBINE archive and OMEX format: one file to share all information
to reproduce a modeling project**

Frank T Bergmann¹, Richard Adams², Stuart Moodie^{3,4}, Jonathan Cooper⁵, Mihai Glont³, Martin Golebiewski⁶, Michael Hucka⁷, Camille Laibe³, Andrew K Miller⁸, David P Nickerson⁸, Brett G Olivier⁹, Nicolas Rodriguez¹⁰, Herbert M Sauro¹¹, Martin Scharm¹², Stian Soiland-Reyes¹³, Dagmar Waltemath¹², Florent Yvon³ and Nicolas Le Novère^{10,3*}



Extracting simulation studies from model repositories





Extracting simulation studies from model repositories

M2CAT — From Masymos To CAT

[M2CAT](#) is a web based tool to export reproducible research results. It links the graph based database [MASYMO](#)s to the [CombineArchiveToolkit](#) (CAT).

Search for models and resources in the database:

Search Results for ubiquitin: 20

(search results are currently showing at most 20 hits.)

[BIOMD0000000293.xml](#)

Following resources available:

- [BIOMD0000000293.xml](#) (model file)
- [BIOMD0000000293.png](#) (reaction network)
- [BIOMD0000000293-curation.png](#) (curation result)

[download CombineArchive \(open in CAT\)](#)

meta data about you

First Name:	Dagmar
Last Name:	Waltemath
Mail:	dw@test.de
Organization:	uros

<http://m2cat.sems.uni-rostock.de>



Extracting simulation studies from model repositories

[BIOMD0000000056.xml](#)

Following resources available:

- BIOMD0000000056.xml (model file)
- BIOMD0000000056_fig2sedml.xml (simulation description)
- BIOMD0000000056.png (reaction network)
- BIOMD0000000056-curation.png (curation result)

[download CombineArchive \(open in CAT\)](#)



CombineArchiveWeb

the current workspace contains the following archives:

[::biomd0000000056.xml](#)

[\[start\]](#) [\[about\]](#) [\[create\]](#)

id: 07587466-b528-44a2-a9ce-ffd62c1726c1

name: BIOMD0000000056.xml

[\[Download\]](#) [\[Edit\]](#) [\[Delete\]](#)

UPLOAD FILES

Archive Content



BIOMD0000000056.xml

file name: BIOMD0000000056.xml

file path: /model/BIOMD0000000056.xml

format: <http://identifiers.org/combine.specifications/sbml.level-2.version-1>

size: 223 KB

master: no

[\[Add OMEX meta\]](#) [\[Add RDF/XML meta\]](#) [\[Download\]](#) [\[Edit\]](#) [\[Delete\]](#)

OMEX entry

FILES



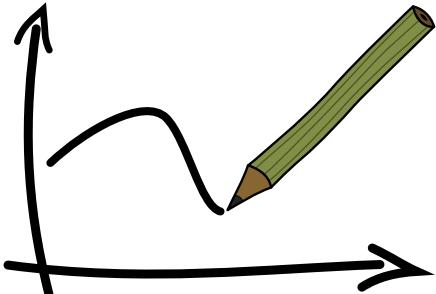
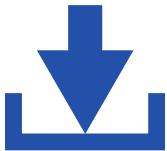
Extracting simulation studies from model repositories

[BIOMD0000000056.xml](#)

Following resources available:

- BIOMD0000000056.xml (model file)
- BIOMD0000000056_fig2sedml.xml (simulation description)
- BIOMD0000000056.png (reaction network)
- BIOMD0000000056-curation.png (curation result)

[download CombineArchive \(open in CAT\)](#)



CombineArchiveWeb

the current workspace contains the following archives:

[::biomd0000000056.xml](#)

[\[start\]](#) [\[about\]](#) [\[create\]](#)

id: 07587466-b528-44a2-a9ce-ffd62c1726c1

name: BIOMD0000000056.xml

[\[Download\]](#) [\[Edit\]](#) [\[Delete\]](#)

UPLOAD FILES

Archive Content



BIOMD0000000056.xml

file name: BIOMD0000000056.xml

file path: /model/BIOMD0000000056.xml

format: <http://identifiers.org/combine.specifications/sbml.level-2.version-1>

size: 223 KB

master: no

[\[Add OMEX meta\]](#) [\[Add RDF/XML meta\]](#) [\[Download\]](#) [\[Edit\]](#) [\[Delete\]](#)

OMEX entry

Thank you for your attention



Martin Scharm
Vivek Garg
Srijana Kayastha
Fabienne Lambusch
Martin Peters
Markus Wolfien
Dagmar Waltemath

Rebekka Alm
Ron Henkel



All cool graphics
by Martin Peters.

**ERASYS-APP Workshop:
Reproducible and Citable
Data and Models (SEP 14-16)**

We're hiring! de.NBI:SYSBIO (BMBF infrastructure) | SBGN-ED+ (e:Bio



Bibliography

(Alm et al 2015) Alm, Rebekka, et al. "Annotation-Based Feature Extraction from Sets of SBML Models." Data Integration in the Life Sciences. Springer International Publishing, 2014.

(Alm et al 2014) Alm, Rebekka, et al. "Annotation-based feature extraction from sets of SBML models." Journal of Biomedical Semantics, accepted for publication, 2015.

(Henkel et al 2015) Henkel, Ron, et al. "Combining computational models, semantic annotations, and simulation experiments in a graph database." Oxford DATABASE, accepted for publication, 2015.

(Scharm et al 2014) Scharm, Martin, et al. The CombineArchiveWeb application—A web based tool to handle files associated with modelling results. No. e639v1. PeerJ PrePrints, 2014.



COMBINE Standards

Core standards



Associated Standardization Efforts



Associated standards

Related standards

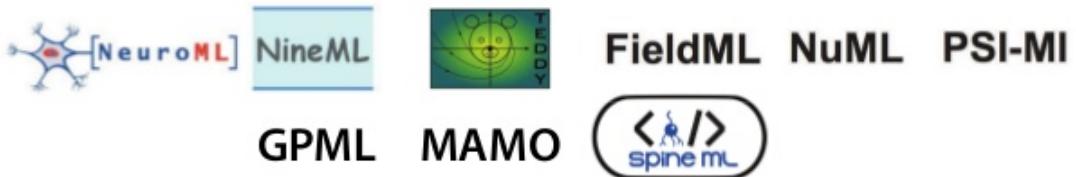
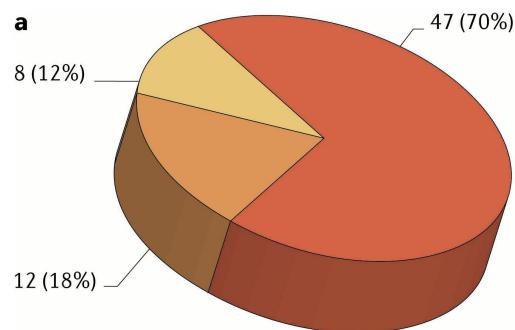
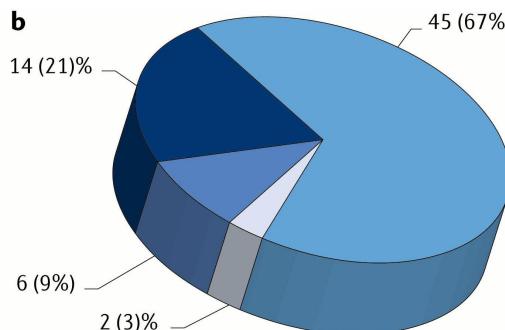


Fig. Adapted from M. Hucka, slideshare. More info
<http://journal.frontiersin.org/article/10.3389/fbioe.2015.00019/abstract>

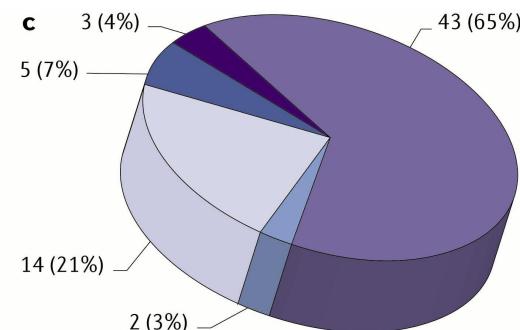
The need for better reproducibility of simulation studies



- Oncology
- Women's health
- Cardiovascular



- Model adapted to internal needs
- Literature data transferred to another indication
- Not applicable
- Model reproduced 1:1



- Inconsistencies
- Not applicable
- Literature data are in line with in-house data
- Main data set was reproducible
- Some results were reproducible

d

	Model reproduced 1:1	Model adapted to internal needs (cell line, assays)	Literature data transferred to another indication	Not applicable
In-house data in line with published results	1 (7%)	12 (86%)	0	1 (7%)
Inconsistencies that led to project termination	11 (26%)	26 (60%)	2 (5%)	4 (9%)

Nature Reviews | Drug Discovery

http://www.nature.com/nrd/journal/v10/n9/fig_tab/nrd3439-c1_F1.html