

# M2CAT

Extracting reproducible simulation studies  
from model repositories using the  
CombineArchive Toolkit

## MARTIN SCHARM

Department of Systems Biology & Bioinformatics, University of Rostock

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



SBI Research Seminar  
SS 2015

## de.NBI Infrastructure

We will provide data management and support for systems biology projects, with a focus on provenance and reproducibility of experimental and modeling results. de.NBI-2380 is part of a large German Network for Bioinformatics Infrastructure. **WE ARE HIRING!**



## CombineArchive Toolkit

Sharing *in silico* experiments is essential for the advance of research in computational biology. The COMBINE archive is a digital container format to assess the management of numerous files and to enable the exchange of reproducible modeling results. We developed the CombineArchive Toolkit, consisting of a library, a web interface and a desktop application. It support scientists in creating, exploring, modifying, and sharing COMBINE archives.



## BiVeS

Armed with our method for difference detection, BiVeS is able to detect and communicate the differences in computational models. The differences are reported in several machine- and human-readable formats, closely related to be integrated in other tools.

## Version Control for Computational Models

With thousands of models available, a framework to track the differences between models and their versions is essential to compare and combine models. Focusing on SBML and CellML, we developed an algorithm to accurately detect and describe differences between versions of a model with respect to (i) the model's encoding, (ii) the structure of biological networks, and (iii) mathematical expressions.

## BudHat

BudHat showcases how BiVeS improves the understanding of a model's changes. BudHat calls BiVeS for the comparison two versions of a computational model and displays the obtained results in the web browser.



## SBGN-ED



SBGN is a markup language to describe models and exchange information about biological systems graphically. We will further develop methods and tools for SBGN-compliant visualization of model-related information. **WE ARE HIRING!**



**HARMONY**, April 19-23, 2015  
HARMONY is a hackathon-type meeting of the COMBINE Community, with a focus on development of the standards, interoperability and infrastructure. Instead of general discussions or oral presentations, the time is devoted to hands-on hacking and interaction between people focused on practical development of software and standards. The HARMONY 2015 is located at the Leuven Willebrord and it is hosted by the groups of Fab Schröder and Dagmar Waltemath.



**VW Summer School**, March 9-13, 2015  
During the 2015 Whole Cell summer school we aim to develop a standard-compliant, open version of the whole-cell model. Eleven tutors and 48 students will hack and code, model and simulate, layout and annotate the whole-cell model using openly available software and COMBINE standards. This event is funded by the Volkswagen Stiftung.

## Events

**ERASYS APP**  
**Workshop on Reproducible and Citable Data and Models**, September 14-16, 2015  
Computational biologists and experimentalists will learn about standards, citable data, about how to make scientific results sustainable, available through open repositories, and about how to find and reuse other people's works in a mixture of lectures and hands-on sessions. The workshop is funded by the ERASYS-APP program.



Martin Peters

Martin Scharm



Dagmar Waltemath



Srijana Kayastha



Ron Harrel



Vivek Garg

## Models as graphs

The increasing diversity of model-related data that is necessary to perform a simulation study leads to new challenges in model storage. We developed a concept for graph-based storage of models and model-related data. Graphs reflect the model's structure much better, enable linking of model-related data on the storage layer, and allow for an efficient search.

## Ontology of Differences

Changes in model versions are manifold and appear on different layers. We develop an ontology of differences occurring in model versions. It will support researchers in analyzing differences, discovering typical changes, summarizing major changes and providing statistics.

## 2MT

2MT is our web-based platform to demonstrate the capabilities of SEMS-related tools. It exemplifies how our model management solutions can be used in existing tools.

## Masymos

Containing SBML- and CellML models, linked semantic annotations (e.g., from ontologies), simulation descriptions, graphical representations and other available types of model-related data, our graph database Masymos can now be queried for complete simulation experiments.



## More

- Our retrieval engine for models applies Information Retrieval techniques to retrieve relevant models from MASYMOS. The proposed searching and retrieval techniques focus on the processing of model meta-information.

Universität  
Rostock



Traditio et Innovatio

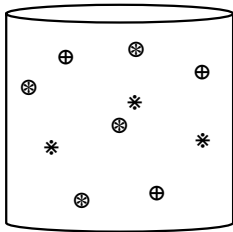


SYSTEMS BIOLOGY  
BIOINFORMATICS  
ROSTOCK

Improving the Management of Simulation Studies in Computational Biology

Martin Scharm, Vivek Garg, Srijana Kayastha, Martin Peters, Dagmar Waltemath





⊙ Models

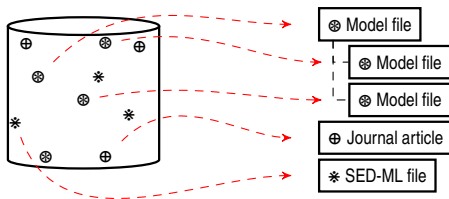
⊕ Documentation

\* Simulation descriptions

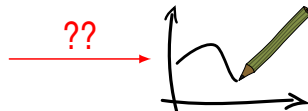
Database of  
models and  
related data

Data necessary to  
reproduce a  
simulation study

Reproduction is a  
**CHALLENGE!**



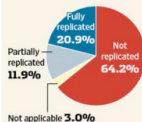
Extracting the data is  
already a challenge!



Understanding and  
using it is almost  
impossible.

### No Cure

When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds failed.



Source: Nature Reviews Drug Discovery

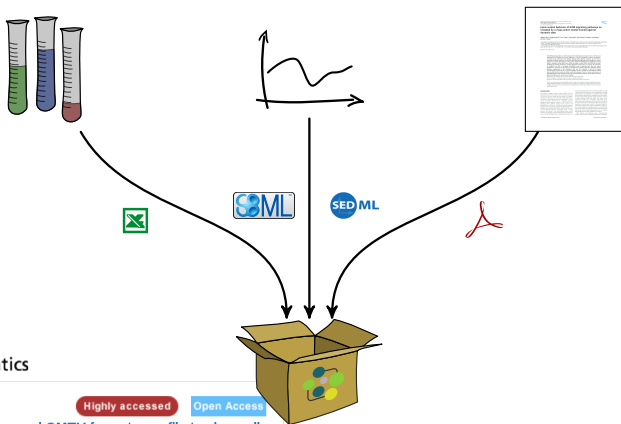
**Table 1: Reproducibility of research findings**

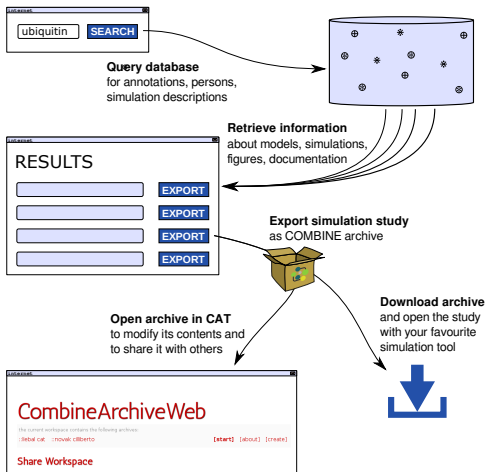
Preclinical research generates many secondary publications, even when results cannot be reproduced.

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.

[http://www.nature.com/nature/journal/v483/n7391/fig\\_tab/483531a\\_T1.html](http://www.nature.com/nature/journal/v483/n7391/fig_tab/483531a_T1.html)





Scharm et al. 2015: Extracting reproducible simulation studies from model repositories using the CombineArchive Toolkit, BTW 2015, Hamburg



## Search

for simulation studies in Masymos

## Retrieve

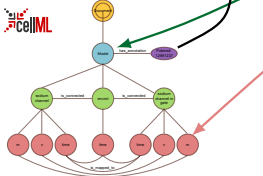
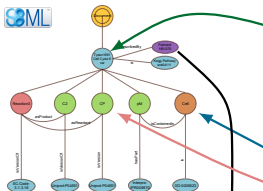
relevant results

## Export

the studies as COMBINE archives using the  
CombineArchive Toolkit



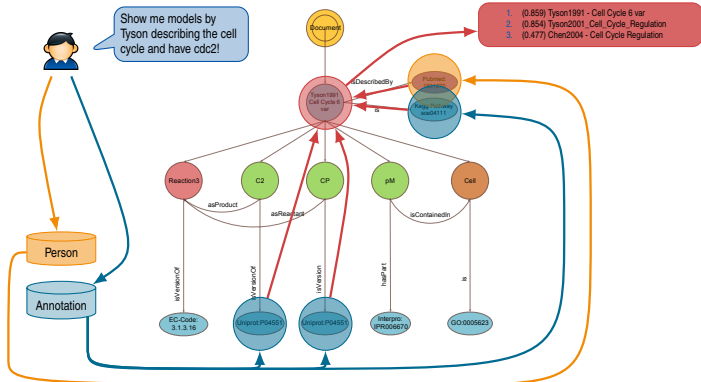
## Models



## Model related data



Henkel et al. 2014, Combining computational models, semantic annotations and simulation experiments in a graph database, Database



Henkel et al. 2010: Ranked retrieval of Computational Biology models, BMC Bioinformatics

## M2CAT — From Masymos To CAT

Search for models and resources in the database:

### Search Results for Tyson: 15

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

hit for Tyson

**tyson\_1991.cellml**

Following resources available:

- **tyson\_1991.cellml** (model file)
- **tysonCellML\_simulation.xml** (simulation description)

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

**sveiczcer\_csikasznagy\_gyorffy\_tyson\_novak\_2000.cellml**

Following resources available:

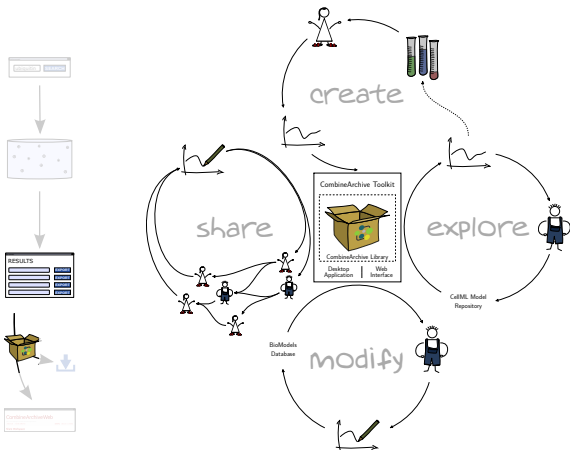
- **sveiczcer\_csikasznagy\_gyorffy\_tyson\_novak\_2000.cellml** (model file)

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

**tyson\_hong\_thron\_novak\_1999.cellml**

Model file  
Simulation description  
+ Additional information





Martin Scharm



Dagmar Waltemath



Florian Wendland



Tom Theile



Martin Peters



Markus Wolfien

Scharm et al. 2014: The CombineArchiveWeb application – A web based tool to handle files associated with modelling results, SWAT4LS, Berlin

[ceur-ws.org/Vol-1320/paper\\_19.pdf](http://ceur-ws.org/Vol-1320/paper_19.pdf)


# CombineArchiveWeb

the current workspace contains the following archives:

**tyson\_1991.cellml**

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

## Archive Content



various files from  
different resources

- tyson\_1991.ai
- tyson\_1991.cellml**
- tyson\_1991.png
- tyson\_1991.session.xml
- tyson\_1991.svg
- tyson\_1991.xul
- tysonCellML\_simulation.xml

FILES
META

**tyson\_1991.cellml**

file name: tyson\_1991.cellml

file path: /tyson\_1991.cellml

format: <http://identifiers.org/combine.specifications/cellml>

size: 33.0 KB

master: no

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

**OMEX entry**

**created:** 7/8/2006, 9:22:43 PM

**modified:** [7/7/2010, 11:55:35 PM] [7/7/2010, 1:12:16 PM] [4/6/2010, 2:52:46 AM] [12/14/2009, 11:16:53 PM] [6/17/2009, 6:15:17 AM] [5/28/2007, 3:32:24 PM] [7/8/2006, 9:22:43 PM]

**description:**

**creators:**

Hanne  
Hanne@hanne-nielsens-macbook.local

Catherine Lloyd  
c.lloyd@audland.ac.nz

pnr2.import  
nobody@models.cellml.org

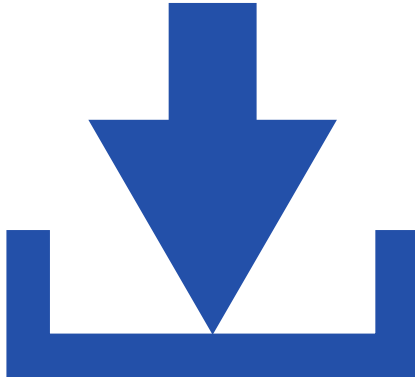
[\[Edit\]](#) [\[Delete\]](#)

as much meta  
as available

April 13, 2015

M2CAT | Martin Scharm

13





- M2CAT implements a workflow to extract reproducible simulation studies from model repositories
- It searches in Masymos <https://sems.uni-rostock.de/projects/masymos/>
- And creates and displays COMBINE archives using the CombineArchive Toolkit <https://sems.uni-rostock.de/projects/combinearchive/>
- all is available from our website: <http://sems.uni-rostock.de>

Thank you for your attention!

## SEMS group

Dagmar Waltemath

**Martin Peters**

Vivek Garg

Srijana Kayastha

Olaf Wolkenhauer



@SemsProject

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

