CENTER FOR REPRODUCIBLE BIOMEDICAL MODELING

ABI Wednesday Forum

February 27, 2019

https://doi.org/10.17608/k6.auckland.7770065

UNIVERSITY of WASHINGTON

UCONN HEALTH

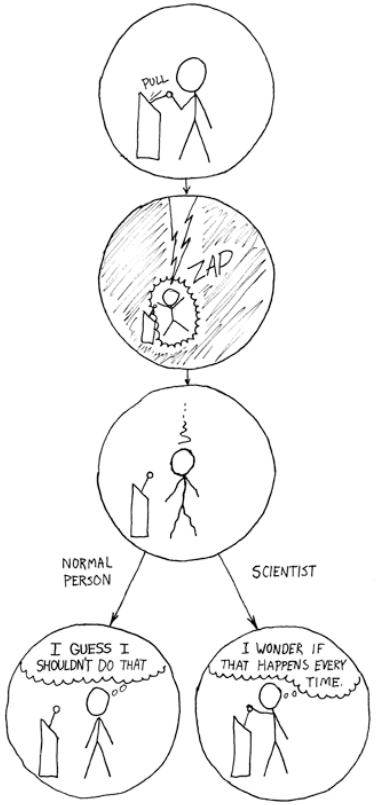
Mount Sinai

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Four Words that are Causing Problems

1. **Replicability**
2. **Repeatability**
3. **Reproducibility**
4. **Reusability**

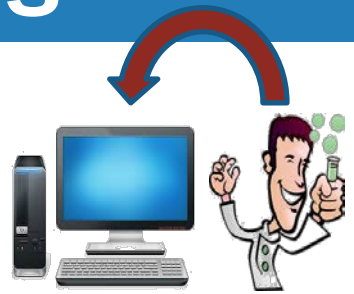
There is a battle going on to decide the meaning of the first three words. Even the US National Academy of Sciences has decided to write a report about it – to be published soon.



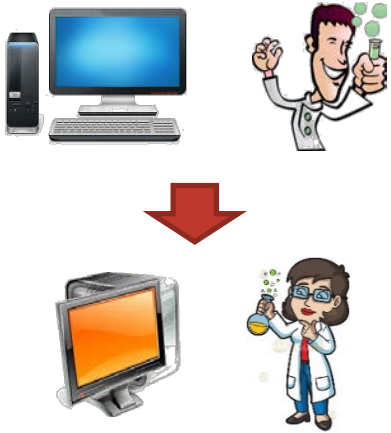
10.17608/k6.auckland.7770065

Two Extreme Scenarios

1. An experiment is carried out and is done again by the same author, using the same equipment, same methods, basically the same everything.



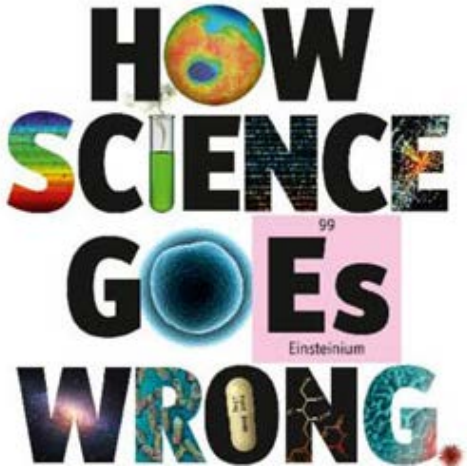
2. The experiment is carried out by a third-party using different equipment, different methods, etc. Basically, everything is different.



In between these two extremes are variants, For example, a third-party could use the same methods but implement them independently of the original author by reading the description given in the original paper.

10.17608/k6.auckland.7770065

What do we mean by reproducibility?



The results of a scientific experiment are **reproducible** if an **independent** investigator accessing published work can replicate them.

The results of a scientific experiment are **repeatable** if the **same** investigator with the same equipment etc. can repeat the results of the experiment.

Some consensus about **Replicability**: Different scientists, same experimental setup; it does not bring much to the table especially for computational experiments.

After some wrangling, Wikipedia is now consistent with these definitions.

These definitions also follow NIST, Six Sigma, ACM and FASEB.

A SIMPLE idea underpins science: “trust, but verify”. Results should always be subject to challenge from experiment. That simple but powerful idea has generated a vast body of knowledge.

10.17608/k6.auckland.7770065

Reproducibility of *in silico* experiments

Should be EASY(IER)!

The results of a scientific experiment are **reproducible** if an **independent** investigator accessing published work can replicate them.

- **Computational repeatability:** a result can be replicated with the same data and software.
- **Algorithmic reproducibility:** a result can be replicated with the same data and different software implementing the same algorithm.
- **Scientific reproducibility:** a result can be replicated with the same data and a different algorithm.
- **Empirical reproducibility:** a result can be replicated with independent data and algorithms.

10.17608/k6.auckland.7770065

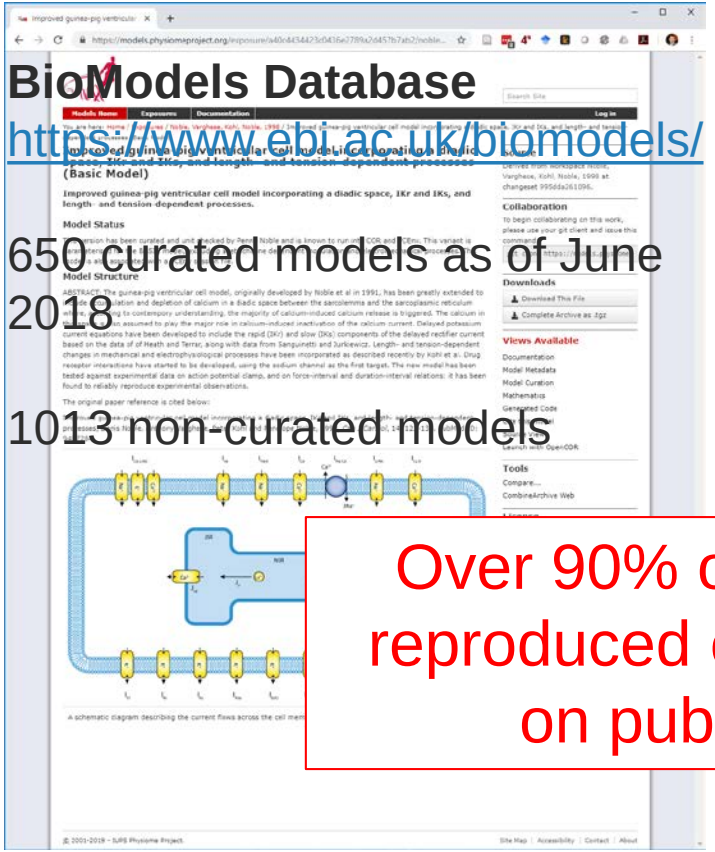
But is it?

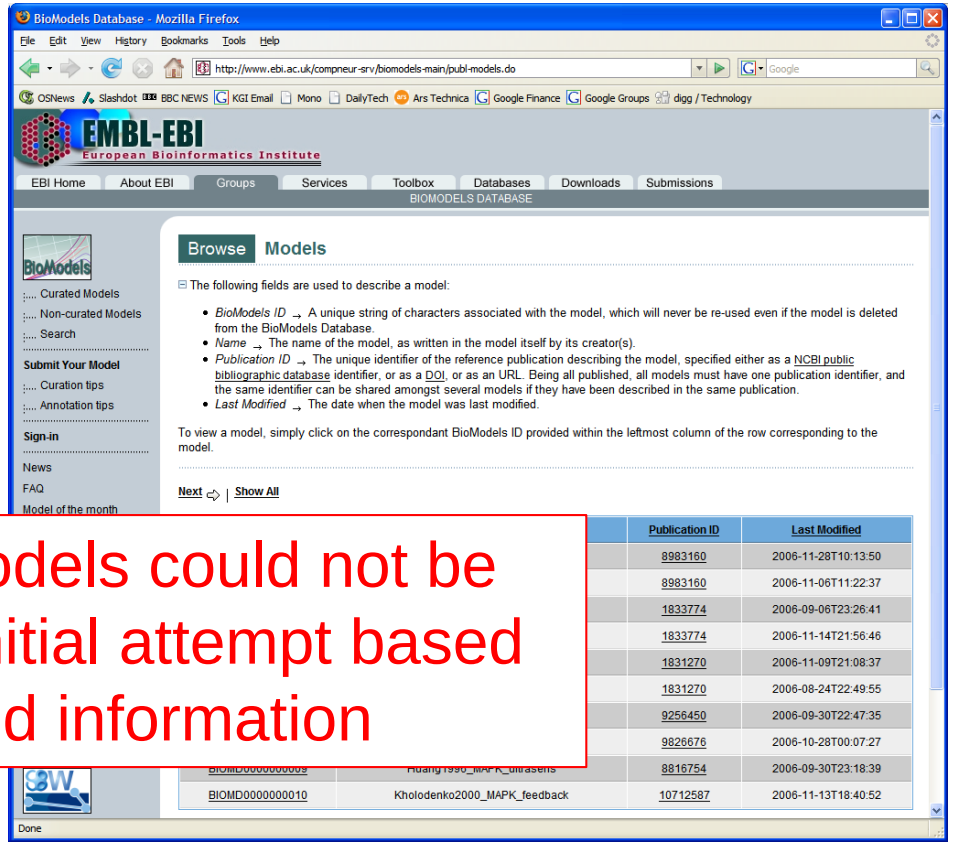
BioModels Database

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

650 curated models as of June 2018

1013 non-curated models





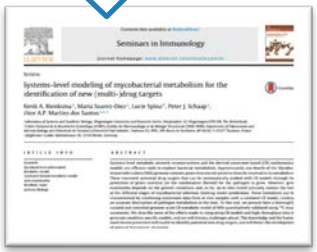
Over 90% of models could not be reproduced on initial attempt based on published information

10.17608/k6.auckland.7770065

Many Different Problems:

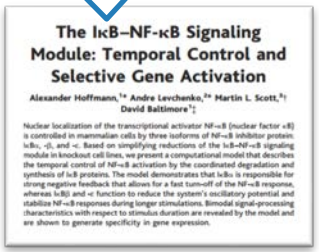
Incomplete parameters

Incomplete model definition

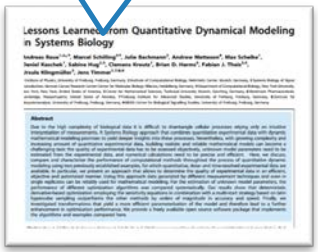


No parameter values

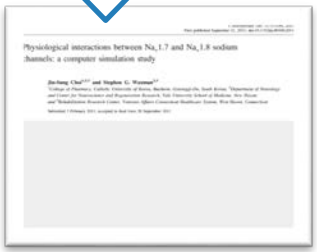
Irreproducible



Analysis procedure not described

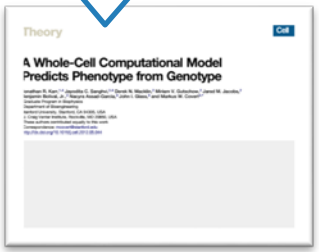


Parameter incorrectly annotated



Irreproducible

No language for describing large models



10.17608/k6.auckland.7770065

Executing Code != Computational Experiment

Why not use an executable language such as Matlab, Python, Java etc to exchange and reproduce models?

Recall that **reproducibility** requires the experiment be recreated independently. An executable language is really only good for **repeatability**.

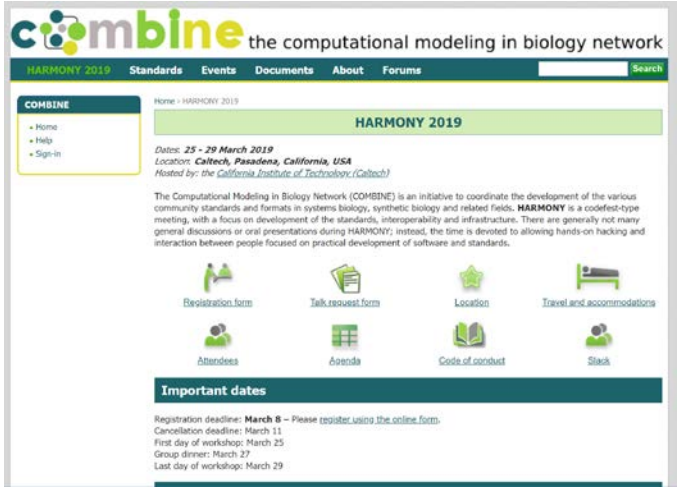
1. To reproduce a model in a different programming language it would need to be manually translated to another language. This can be difficult and error prone.
2. There is no means to share such models because other groups might use different programming languages, APIs, etc.
3. *Combining such models into larger models is extremely difficult.*
4. *It is difficult to annotate models that use an executable language.*

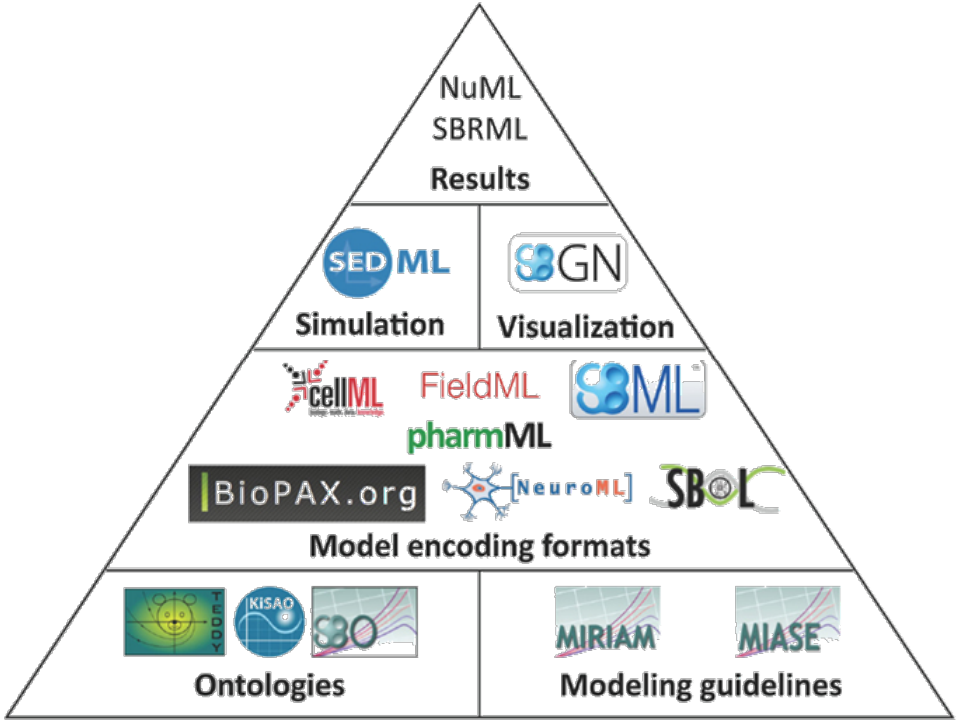
10.17608/k6.auckland.7770065

What's the solution?

There is no complete solution but many of the issues can be resolved by using community based modelling standards.

These standards fall under the umbrella of the COMBINE Standards (http://co.mbine.org/)





10.17608/k6.auckland.7770065

Many pieces exist, but…

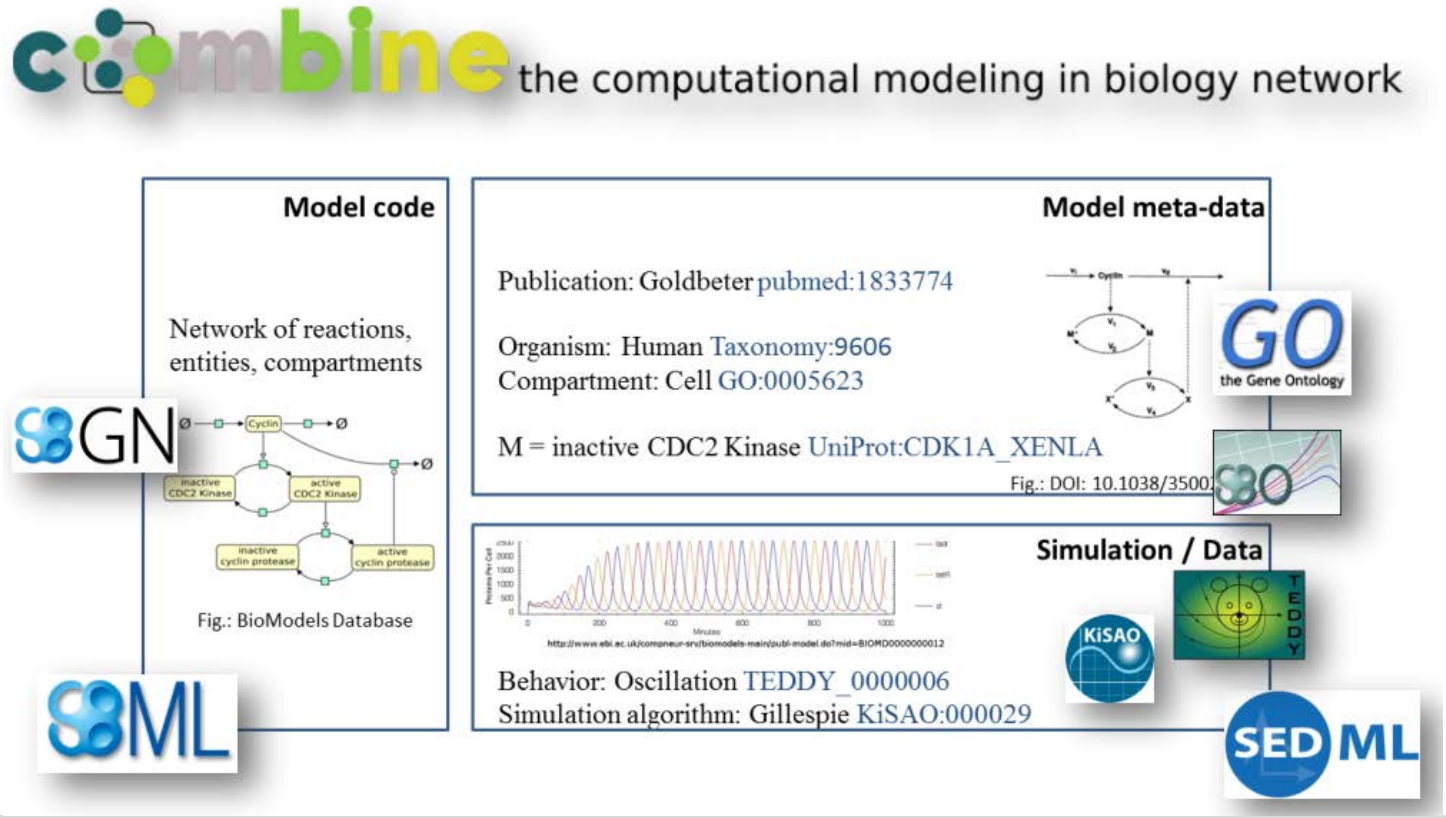
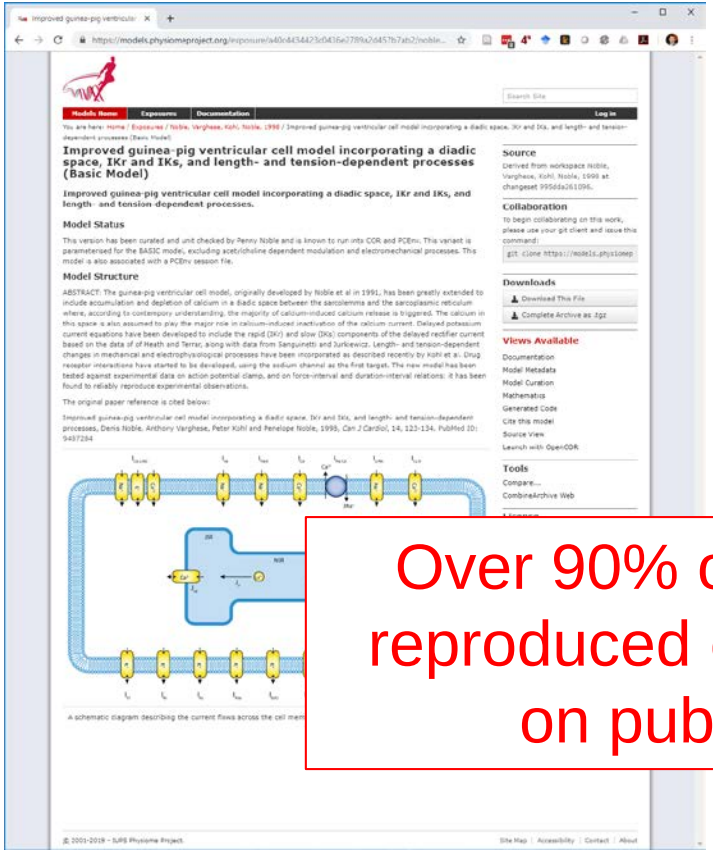
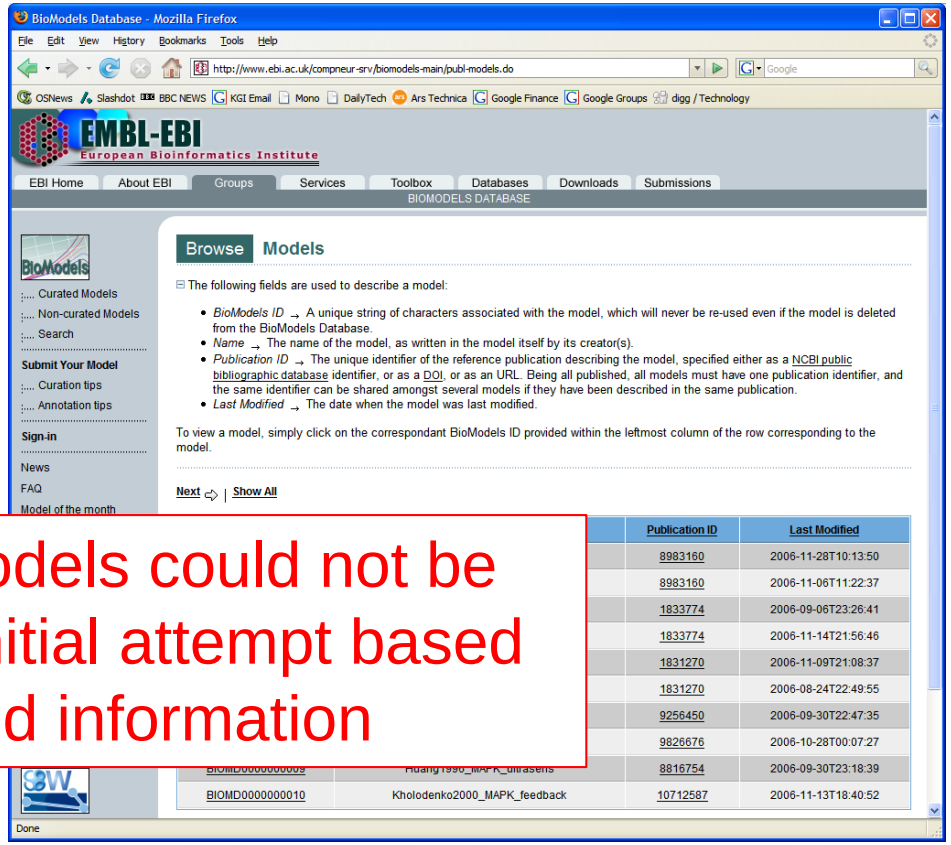


Figure from **Dagmar Waltemath**

10.17608/k6.auckland.7770065

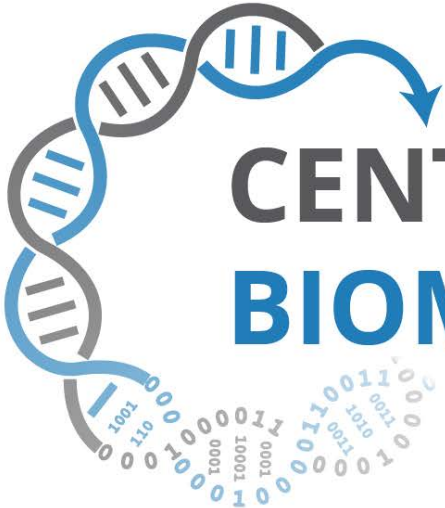




Over 90% of models could not be reproduced on initial attempt based on published information

10.17608/k6.auckland.7770065

And this is where our new Center comes in



CENTER FOR REPRODUCIBLE BIOMEDICAL MODELING

https://reproduciblebiomodels.org/

10.17608/k6.auckland.7770065

Overview of the center

Center Team

Herbert Sauro
U Washington
Director

Jonathan Karr
Mount Sinai
TR&D 1

John Gennari
U Washington
TR&D 2

Ion Moraru
UConn Health
TR&D 3

David Nickerson
ABI
Curation Service

Support by NIBIB and NIGMS:

10.17608/k6.auckland.7770065

External Advisory Board

- **Gary Bader**
  University of Toronto
- **Ahmet Erdemir**
  Cleveland Clinic
- **Juliana Friere**
  New York University
- **Bill Lytton**
  SUNY Downstate Medicine
- **Andrew McCulloch**
  UC San Diego
- **Pedro Mendes**
  UConn Health

10.17608/k6.auckland.7770065

Goals

Long-term

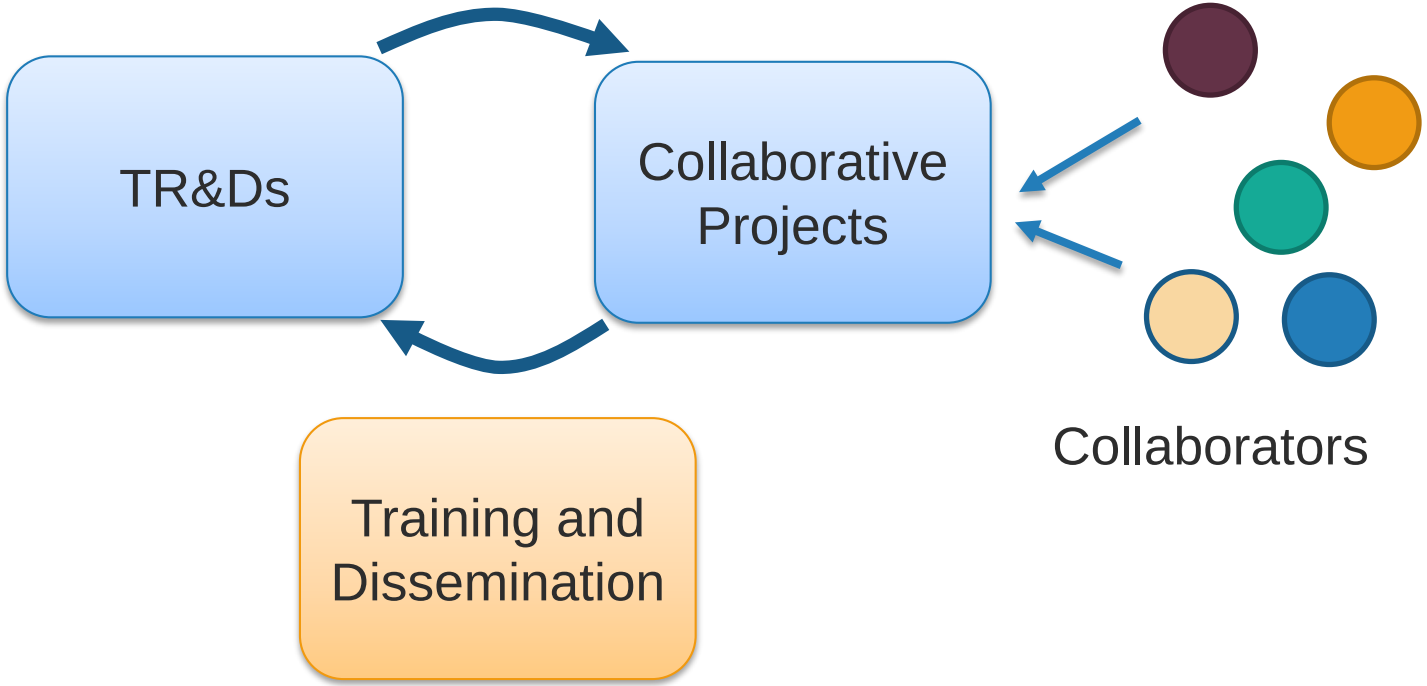
- Enable more comprehensive and more predictive models that advance precision medicine and synthetic biology

Short-term

- Make modeling more reproducible, comprehensible, reusable, composable, collaborative, and scalable
- Develop technological solutions to the barriers to modeling
- Integrate the technology into user-friendly solutions
- Push researchers to use these tools
- Partner with journals

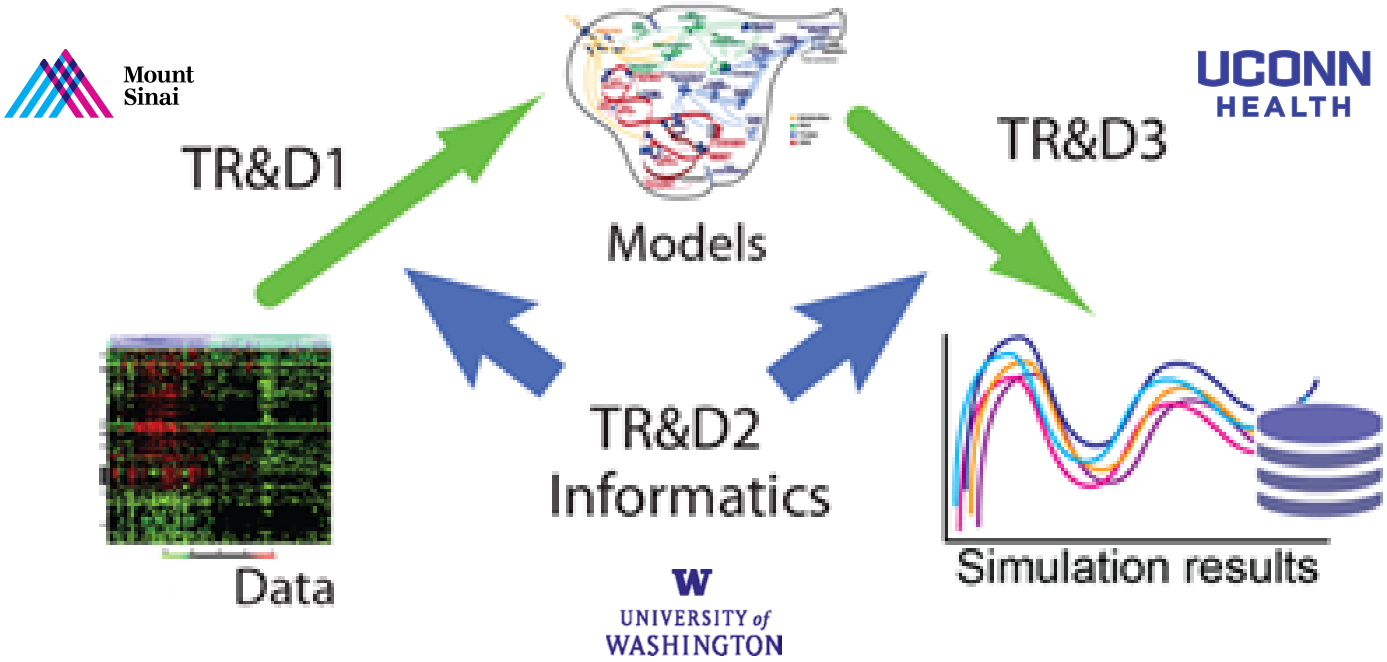
10.17608/k6.auckland.7770065

Center organization



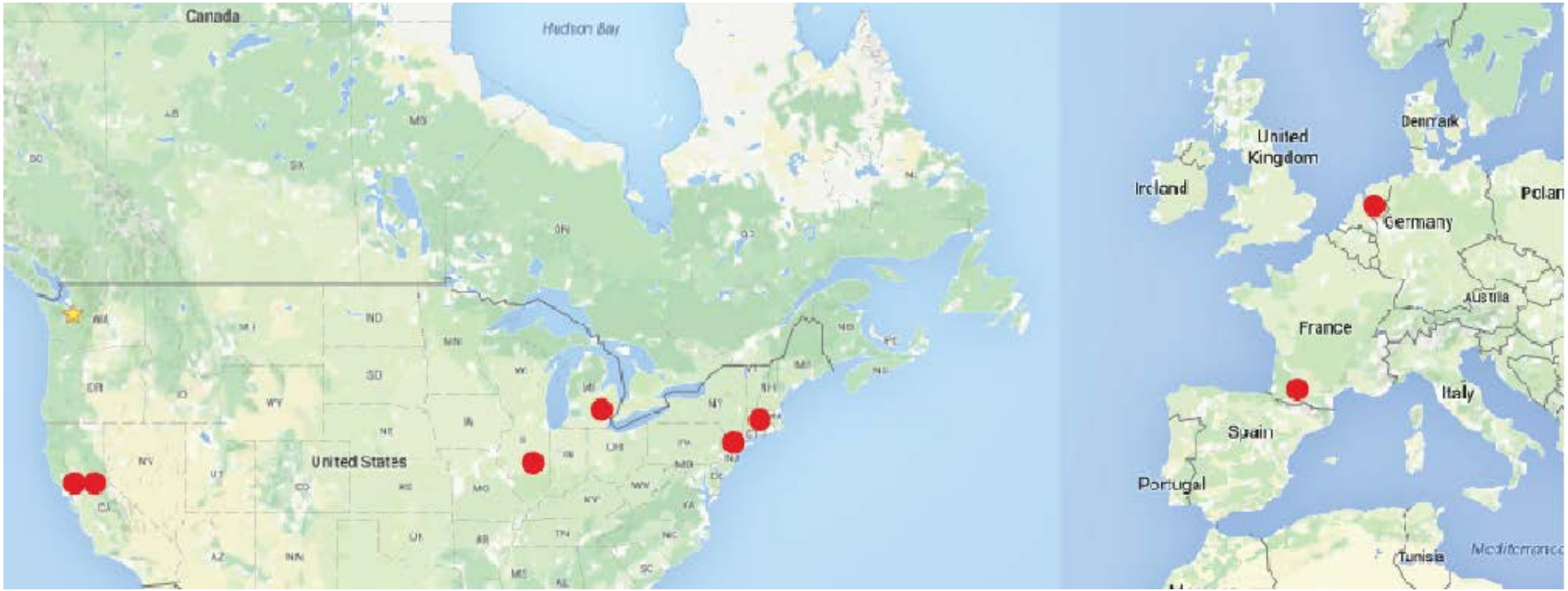
10.17608/k6.auckland.7770065

TR&Ds



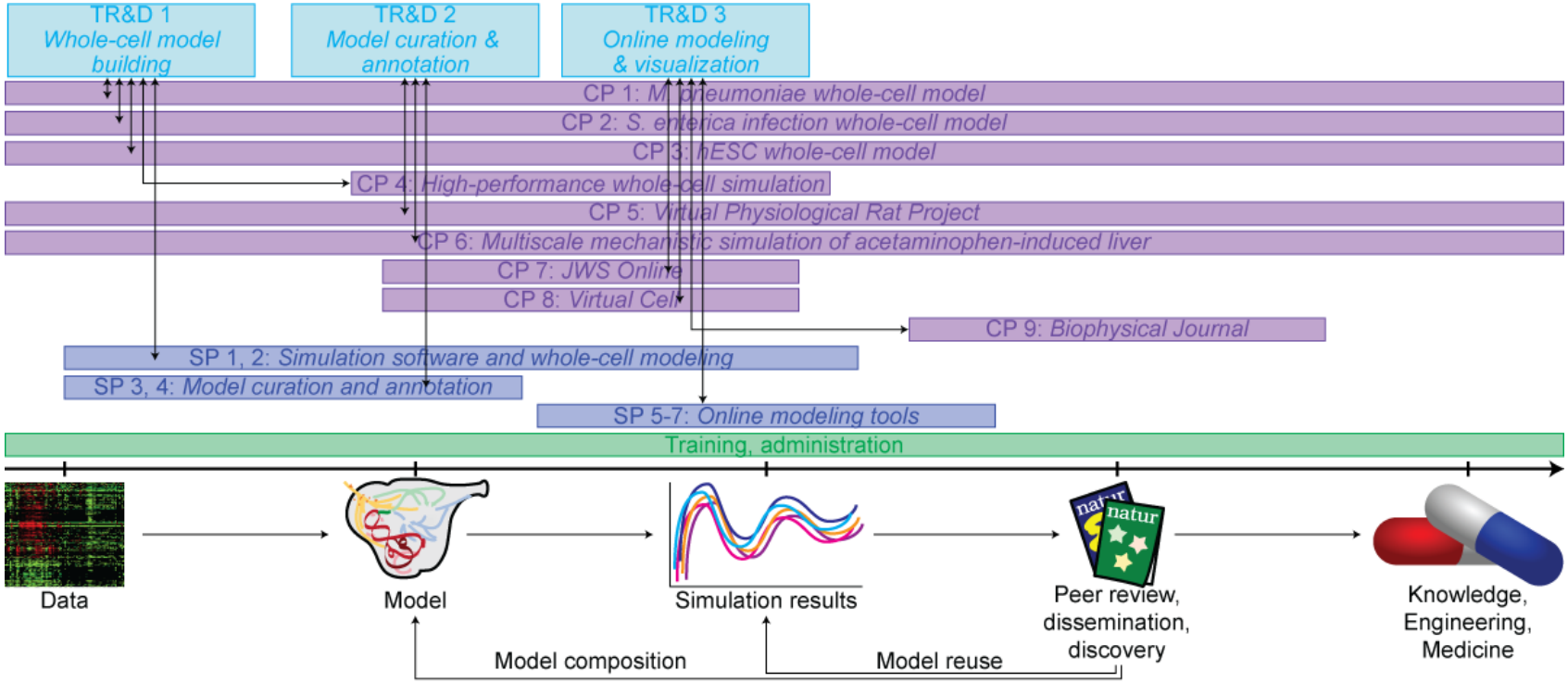
10.17608/k6.auckland.7770065

Driving collaborative projects



10.17608/k6.auckland.7770065

TR&Ds span every modeling phase



10.17608/k6.auckland.7770065

Training and dissemination

10.17608/k6.auckland.7770065

Center funding

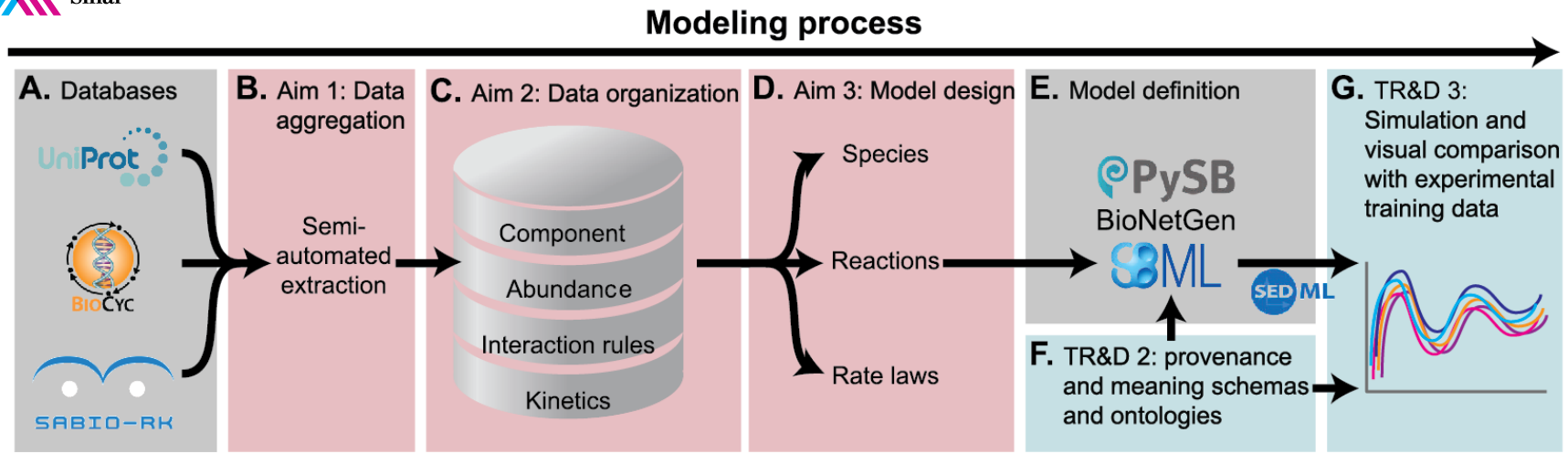
- $6.5 million for 5 years
- Each core has R01-scale funding
- Funds for workshops
- Funds for project management

10.17608/k6.auckland.7770065

TR&D 1: Scalable model construction

TR&D 1: Model Construction

Mount Sinai



TR&D 1 will develop tools for reproducibly building models. This will include (1) aggregating large and heterogeneous data needed to build models, (2) organizing this data for model construction, and (3) designing models from this data.

10.17608/k6.auckland.7770065

TR&D 1: goals

- Facilitate the construction of more comprehensive and more accurate models
  - CP 1: Mycoplasma pneumoniae
  - CP 3: Human embryonic stem cells

10.17608/k6.auckland.7770065

TR&D 1: goals

- Overcome the most immediate barriers
  - Lack of data for modeling
  - Inability to identify relevant data for modeling
  - Disconnect between data and models
  - Incomposability of separately developed models
  - Insufficient metadata for composition
  - Inability to model collaboratively

10.17608/k6.auckland.7770065

TR&D 1: philosophy

- Modeling should be collaborative and composable from the ground up
- Modeling tools should be modular, composable, and easy to use
- Technology development should be motivated by specific models

10.17608/k6.auckland.7770065

TR&D 1: aims

- Develop an integrated database of data for modeling
- Develop tools for identifying relevant data for a specific model
- Develop a framework for organizing the data needed for a model
- Develop a framework for programmatically constructing models from these datasets
- Deploy these tools as web-based tools and Python libraries

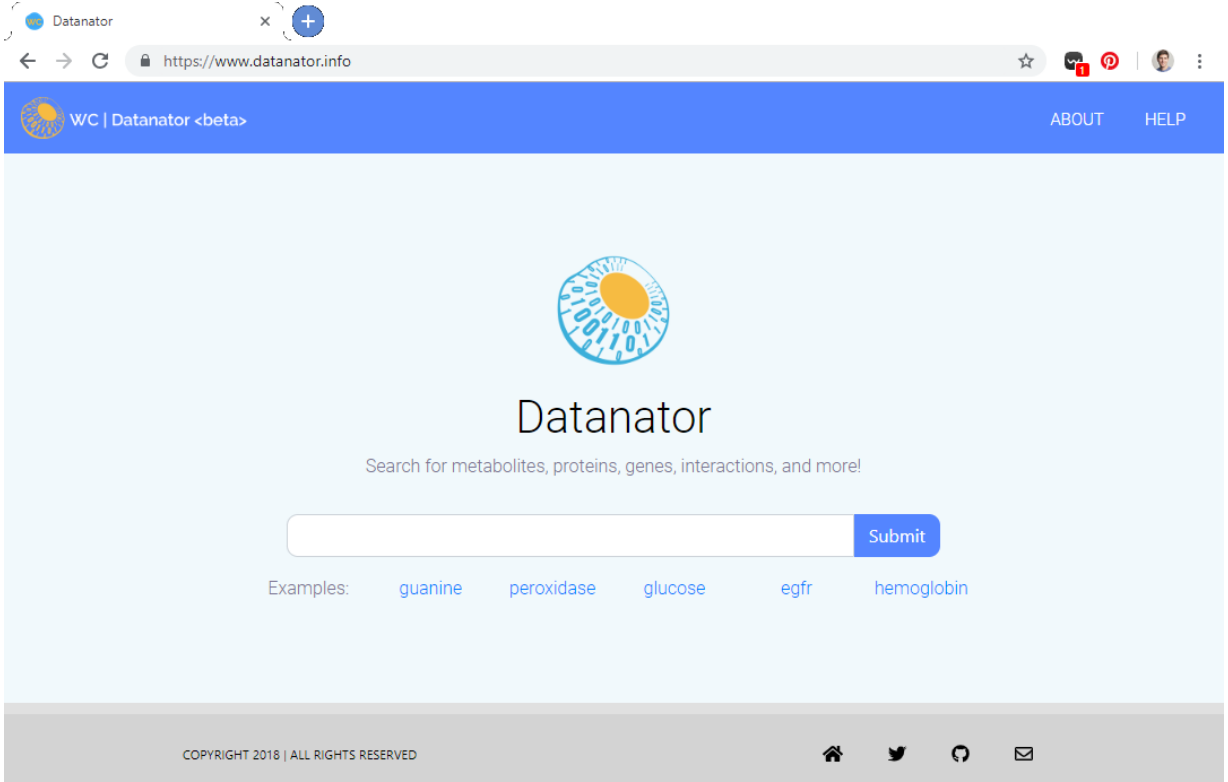
10.17608/k6.auckland.7770065

TR&D 1: progress

- Developed an integrated database of most essential data
- Developed tools to discover relevant data about a specific organism and condition
- Begun to develop web interface to browse and search data
- Developing tools for extracting data for a specific model
- Developing a data model to describe the data used for specific modeling projects
- Developing a framework for programmatically constructing models from these datasets

10.17608/k6.auckland.7770065

Datanator website

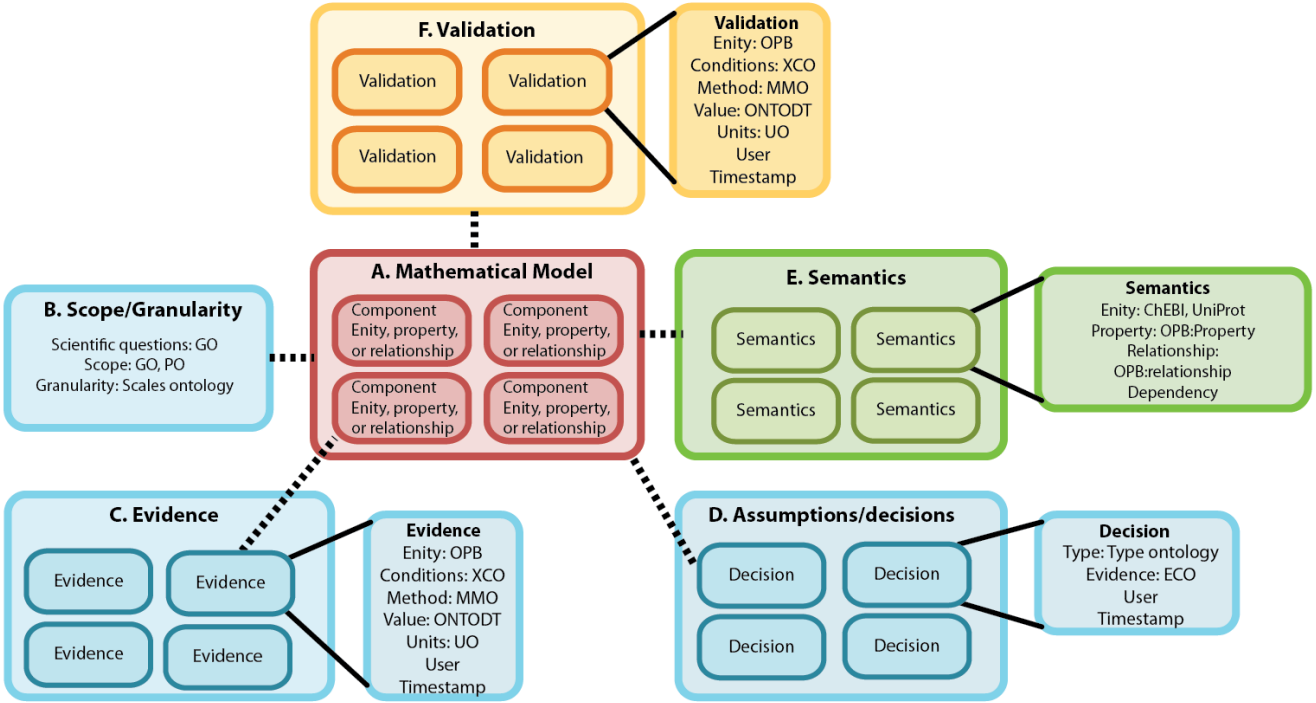


10.17608/k6.auckland.7770065

TR&D 2: Enhanced semantic and provenance annotation to facilitate scalable modeling

TR&D 2: Informatics Support

UNIVERSITY of WASHINGTON



TR&D 2 will develop tools for annotating the meaning and provenance of models as well as annotating simulation results, model behavior and model validation. This will include developing the schema and ontologies for describing the provenance, simulation data and validation.

10.17608/k6.auckland.7770065

TR&D 2: Goals

Improved semantic annotation

- Ontology-based composite annotations
- Tools that support common annotation formats (COMBINE Archives)
- Annotation that describes model provenance & modeling assumptions
- Annotation that can describe data as well as models

Tools that use these annotations

- Semantic search for relevant models
- Automatic data-to-model matching
- Model merging, model visualization, model modularization

10.17608/k6.auckland.7770065

TR&D 2: progress

Completion of Java API for annotation

- Available with release 4.2 of the SemGen software (Dec ’18)
- Read/write of COMBINE Archive format

Proof-of-concept demonstration of annotation API

- With Antimony/Tellurium: Kyle Medley
- Begun communication with Alan Garny for use by OpenCOR

Meetings with Auckland team

- Train and coordinate annotation efforts

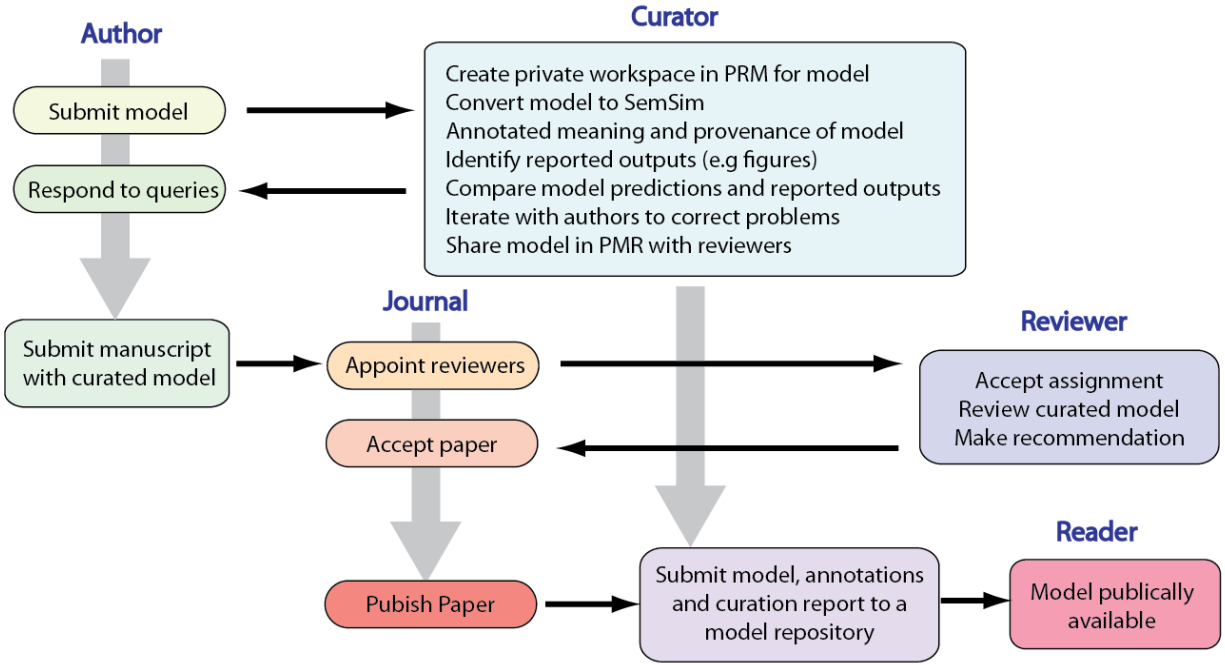
10.17608/k6.auckland.7770065

Model curation service to enhance model reuse and composition

Curation service: goals

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Manuscripts received by journals will be curated to make sure that any author supplied code will faithfully reproduce the results presented in the manuscript.



10.17608/k6.auckland.7770065

Curation service: journal pilots

- Physiome: agreed
- Biophysical Journal: agreed
- Mathematical Biosciences: agreed
- Bulletin of Mathematical Biology: agreed
- BMC Systems Biology: journal closing down!
- PLoS Computational Biology: agreed
- Molecular Systems Biology: potential
- Cell Systems: declined
- Other suggestions?

10.17608/k6.auckland.7770065

Curation service: in practice

Anand Rampadarath
mathematician!

Karin Lundengård
biologist!

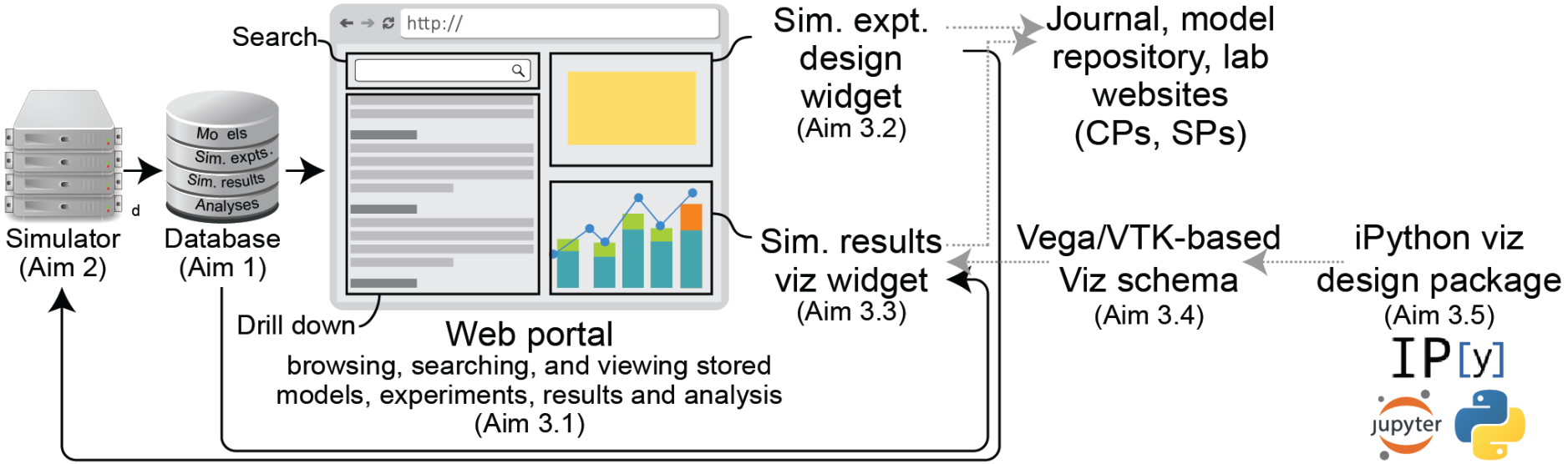
10.17608/k6.auckland.7770065

TR&D 3: Online scalable simulation, analysis, and visualization

TR&D 3: Simulation

UCONN HEALTH

Note: We do not intend to write new simulators. We will use existing third-party simulation software.



TR&D 3 will develop tools for reproducibly simulating and analyzing models online. This will include (1) web-based tools for designing simulation experiments and visualizing simulation results, (2) a universal simulator for simulating biomodels and (3) a database for organizing and storing simulation results.

10.17608/k6.auckland.7770065

TR&D 3: progress

- Simulation: Infrastructure Design
  - Databases:
    - PostgreSQL database for models, data, simulation, provenance.
    - MongoDB for noSQL (logging information)
  - RESTful API servers:
    - Separate servers with different access permissions and features
    - Different classes of API (data, solver, execution)
  - Registry of solvers:
    - Algorithm capabilities
    - Task capabilities
  - Container infrastructure:
    - Container registry – Docker
    - Container orchestration – Kubernetes (seven Virtual Machines)
  - Job manager (+ own database + API)
    - Slurm/XDMod workload manager – job, *not* workflow manager
  - Compute and storage resources
    - Local: dedicated partition on cluster, HDF5 storage
    - AWS (overhead involved, adds to cost; needed for portability)
- Standards: Language and Tool Development
  - Python support for SBML render and layout extensions
  - New high-level, human-readable API for SED-ML

10.17608/k6.auckland.7770065

Technology integration

Testing and documentation

Encourage a more systematic approach to modeling, treating modeling more as an engineering discipline especially when developing larger models.

But even for small models where there are clinical implications, the following broad desirable attributes should be considered:

a) Documentation (TR&D 1, 2, 3)
b) Uncertainty Quantification (TR&D 3)
c) Reusable (TR&D 1, 2)
d) Exchangeable (TR&D 1, 2)
e) Stress-Tested (TR&D 3)

This dovetails with existing efforts such as the Credible Practice of Modeling & Simulation in Healthcare at IMAG who have the Ten Simple Rules.

10.17608/k6.auckland.7770065

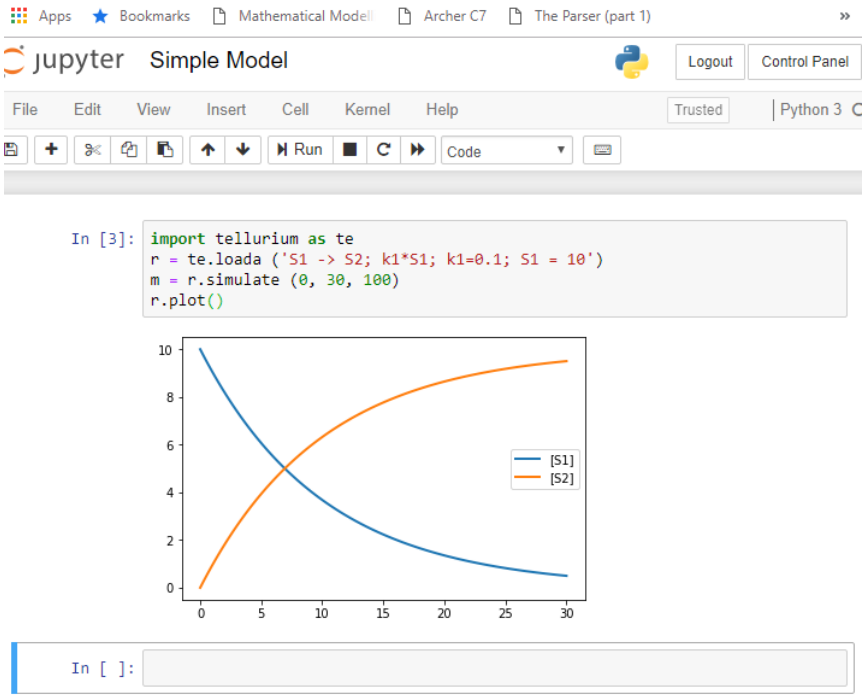
Training and dissemination

Online course

- Course 0: Introduction to Biomodeling
- Lesson 0: Introduction to the Course
- Lesson 1: Introduction to Modeling
- Lesson 2: Model Elements
- Lesson 3: Cellular Networks
- Lesson 4: Differential Equations
- Lesson 5: Mass-Action Kinetics
- Lesson 6: Differential Equations Modeling
- Lesson 7: Steady State and Stability (1)
- Lesson 8: Steady State and Stability (2)
- Lesson 9: Enzyme Kinetics (1)
- Lesson 10: Enzyme Kinetics (2)

10.17608/k6.auckland.7770065

JupyterHub examples server



http://jupyterhub.reproduciblebiomodels.org

10.17608/k6.auckland.7770065

Center for Reproducible Biomodeling

Cell Modeling Seminar

Nov 6: Sheriff Rahuman

Project Leader, European Bioinformatics Inst

Leveraging public data repositories for cell modeling

Dec 4: Bill Hlavacek

Scientist, Los Alamos National Laboratory

Formalizing and leveraging qualitative observations of system behavior in model development

Jan 8: Stephen Larson

Co-founder, OpenWorm & CEO, MetaCell

OpenWorm: building a whole animal simulation

Feb 5: Barbara Bakker

Professor, Univ Medical Center Groningen

Computational models and network-based drug design for metabolic disease

Mar 5: Andrew Hessel

CEO, Humane Genomics

GP-write and the future of engineering living organisms: a personal perspective

Apr 2: Jacky Snoep

Professor, University of Stellenbosch

Data and model management using the FAIRDOMHub: from experiment to model simulation and publication

Details

Day/Time
First Tuesday of each month
3pm EDT (7pm UTC)

Replay with second discussion
First Thursday of each month
9am IST (3:30am UTC)

Location
Online using Zoom webinar
http://bit.ly/2SB1mSk

Format
25-min presentation
35-min Q&A

Archive
Online @ YouTube
http://bit.ly/2yTmOtN

More info

reproduciblebiomodels.org/seminar

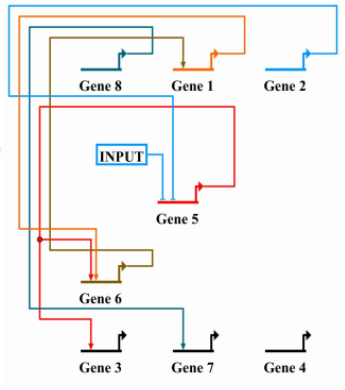
Questions

Yosef Roth & Veronica Porubsky
seminar@reproduciblebiomodels.org

https://reproduciblebiomodels.org/seminar

10.17608/k6.auckland.7770065

Modeling game



Strategies - Granger Causality Test

Regulator

Regulated	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Gene 6	Gene 7	Gene 8
Gene 1	3	0.471073	0.171779	1.15774	0.351971	2.00698	1.0356	1.61553
Gene 2	1.79798	3	1.89804	0.00345586	1.72799	0.832344	1.23155e-09	0.00380074
Gene 3	2.44281	2.61155e-08	3	0.584077	1.58147e-07	1.37904	0.000800385	0.0282527
Gene 4	0.217466	0.457343	0.140683	3	1.18991	5.31902e-10	0.953515	0.925021
Gene 5	1.21131	3.73519e-17	1.43189	0.667769	3	1.03219	0.000698253	0.0556494
Gene 6	0.266648	1.52821	0.774729	1.38176	0.396075	3	0.907911	0.588501
Gene 7	1.3032	1.50701	0.469524	1.58661e-05	0.206547	0.0675559	3	1.25601e-05
Gene 8	0.0746103	0.037321	0.694628	6.38787e-14	0.305837	9.85709e-08	1.47101	3

Known connection, true positive — Minimum (by row) — Not minimum (by row)

*Summate results of Time Delay = 20s, 40s, 60s

QUESTIONS RESPONSES 38

BIOEN 498/599: Experiment Request Form

Buy your experimental data here!

You may only order one type of perturbation/data collection method at a time.

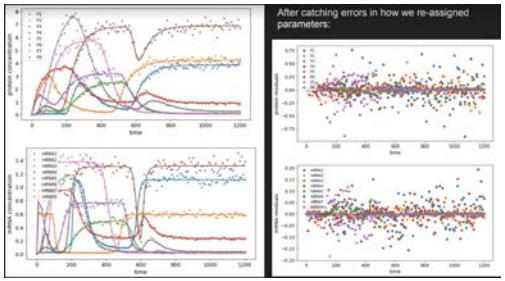
This form is automatically collecting email addresses for UW users. Change settings

Team Name *

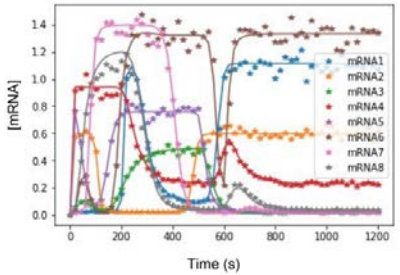
- Quintus

Data Collection Type *

- Mass Spectrometry (all proteins) - Low Resolution (every 20)
- Mass Spectrometry (all proteins) - High Resolution (every 10 mins)
- RNA Sequencing (all mRNA) - Low Resolution (every 20 mins)
- RNA Sequencing (all mRNA) - High Resolution (every 10 mins)
- Fluorescence Tagging (up to 3 proteins - high resolution time course) - every 10 mins



Fitted model - Cross validation with RNA-seq



10.17608/k6.auckland.7770065

Conference and seminar talks

Conferences

- COBRA conference
- COMBINE
- GP-Write Meeting
- ICSB
- ISMB
- ISSB Siena Summer School
- World Congress of Biomechanics
- VPH Conference

Seminars

- Mount Sinai, NY
- New York University, NY
- NIH, MD
- Pacific Northwest National Lab, WA
- SUNY Downstate, NY
- University of Washington, WA

10.17608/k6.auckland.7770065

@

AUCKLAND
BIOENGINEERING INSTITUTE

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Te Whare Wānanga o Tāmaki Makaurau

Where does this fit in with other ABI projects…

- Physiome/VPH
  - Journal
  - PMR
  - Aotearoa Fellowship
- SPARC
  - DRC is proof that these issues are being taken seriously by NIH leadership!

10.17608/k6.auckland.7770065

Where does this fit in with other ABI projects…

- All modelling projects trying to publish…

Curation service: journal pilots

- Physiome: agreed
- Biophysical Journal: agreed
- Mathematical Biosciences: agreed
- Bulletin of Mathematical Biology: agreed
- BMC Systems Biology: journal closing down!
- PLoS Computational Biology: agreed
- Molecular Systems Biology: potential
- Cell Systems: declined
- Other suggestions?

Want to join in?

cellML

Search Site

Home | About CellML | Getting started | Tools | Models | Specifications | Community | Log in

You are here: Home / Community / Events / Workshops / 13th International CellML Workshop

13th International CellML Workshop

The 13th International CellML Workshop will be held 6 & 7 May 2019 at the Goldie Estate on Waiheke Island, Auckland, New Zealand.

The annual International CellML Workshop is an opportunity for scientists to present and discuss ongoing work related to physiological modelling and simulation with particular focus on model and data exchange & reuse, archiving, and versioning and related software support.

While the primary emphasis of this workshop is on CellML and related technologies, work from across the VPH & Physiome Project and computational systems biology communities are covered (e.g., SED-ML, Physiome Repository, COMBINE, reproducibility and model curation, etc.). In addition to participants from the Auckland Bioengineering Institute, the Medical Technologies Centre of Research Excellence, the Maurice Wilkins Centre, and the University of Auckland, previous workshops have attracted delegates from the UK, USA, Japan, Norway, France, Poland, Belgium, Italy, and Germany for both in-person and remote presentations.

This year the workshop will feature the following topics.

- libCellML developments and plans (documentation; GitHub).
- SPARC MAPcore.
- Model curation.
- Model and data annotation.

Location

The 13th International CellML Workshop will be held at the Goldie Estate, Waiheke Island, Auckland, New Zealand. Daily transport from downtown Auckland to the venue will be arranged each day of the workshop.

Programme

A preliminary programme will be forthcoming. Please note that the actual schedule is subject to change.

News

Announcing the 13th International CellML Workshop
2019-01-21

HARMONY 2019
2019-01-21

Announcing the 12th International CellML Workshop
2018-03-01

Announcing the 11th International CellML Workshop
2017-03-01

More...

Recent changes

Announcing the 13th International CellML Workshop

HARMONY 2019

13th International CellML Workshop

NIH Common Fund

13th International CellML Workshop

More...

May 6 & 7

Goldie Estate, Waiheke Island

Stay tuned for registration announcement!

FREE!

https://www.cellml.org/community/events/workshop/2019

10.17608/k6.auckland.7770065

Help is available!
















10.17608/k6.auckland.7770065

Thank you!