COMBINE 2020

[B] Break [D] Discussion [I] Invited Talk [L] Lightning talks [S] Social space [T] Talk

OCTOBER 5 • MONDAY

PINNED
01:00 – 01:45 | I | **CellDesigner: A modeling tool for biochemical networks** | Room 1

Speakers: Akira Funahashi

Abstract: Understanding the logic and dynamics of gene-regulatory and biochemical networks is a significant challenge of systems biology. To facilitate this research topic, we have developed a modeling/simulating tool called CellDesigner.
CellDesigner primarily has capabilities to visualize, model, and simulate gene-regulatory and biochemical networks. Two significant characteristics embedded in CellDesigner boost its usability to create/import/export models:
(1) Solidly defined and comprehensive graphical representation (Process Diagram) of network models and
(2) Systems Biology Markup Language (SBML) as a model-describing basis, which functions as inter-tool media to import/export SBML-based models.
CellDesigner also supports simulation and parameter scan, supported by integration with SBML ODE Solver, COPASI, and Simulation Core Library enabling users to simulate through our sophisticated graphical user interface. Users can also browse and modify existing models by referring to existing databases (ex. BioModels.net) directly through CellDesigner.
These enhancements make CellDesigner not only a modeling/simulation tool, but also an integrated analysis suite. CellDesigner is implemented in Java and thus supports various platforms (i.e., Windows, Linux, and macOS). CellDesigner version 4.4.2 is freely available via our web site http://celldesigner.org, and a new version of CellDesigner is under development, which will support spatial modeling.

Target audience: modelers

Standards: SBML, SBGN

01:45 – 02:00 | T | **A modular, thermodynamic approach for constructing large-scale kinetic models in systems biology** | Room 1

Speakers: Michael Pan

Abstract: Comprehensive large-scale mathematical models of biomolecular systems have the potential to direct future advances in health and biotechnology, but are currently difficult to develop. It is generally acknowledged that kinetic models should be constructed by reusing and coupling together existing models of smaller systems, i.e. in a modular fashion. However, models of biological systems are often expressed in different formalisms and therefore do not naturally interface with each other. In this talk, I argue that the conservation laws of physics provide a unified interface for models to communicate. The bond graph - a graphical, energy-based modelling framework that is well-established in the field of engineering - is introduced as a means of enabling this approach. The approach will be illustrated in the development of a model of the MAPK signalling cascade.

Target audience: kinetic modellers, tool developers

Standards: CellML, SBGN, SBML, SED-ML

02:00 – 02:15 | T | **Open Source and Sustainability** | Room 1

Speakers: Jacob Barhak

Abstract: In short: The presenter will discuss encountered issues and limitations with current open source practices that impact on sustainability. Those issues include: over interpretation, abandoned code, and potential conflict with external restrictions such as patents and government restrictions. Relation to government funding policies will be discussed. A possible solution to improve incentives and sustainability will be presented. Details: The presenter develops computational disease modeling tools that are both open source and proprietary. The presenter benefited from open source and currently owns the most validated diabetes model known and a COVID-19 model. Yet current practices make it harder to sustain research using current open source practices. A recent call by the modeling community with regards to COVID-19 modeling, calls to publish model code openly. However, making such a call is insufficient unless issues related to sustainability are addressed. Modeling practices should better fit the reality and deal with issues such as over interpretation of licenses, abandoned research, and other restrictions such as patents or government restrictions. Open source and proprietary development should work in harmony and provide proper incentives for development.
A simple solution will be suggested to allow both sustainable development of open source code and incentives for researchers to advance their research.

Target audience: All those interested with licenses and open source and potential legal issues

Standards: All those interested with licenses and open source and potential legal issues

02:15 – 02:30 | T | **Physiome - Make your model publications discoverable, reproducible, and reusable.** | Room 1

Speakers: Karin Lundengård

Abstract: Physiome is proud to show our first publications. We are a journal committed to reproducibility and reusability of mathematical models of physiological processes. When you publish in Physiome, the persistent identifier for your article is connected to a curated and permanent version of the model code. This makes the code necessary to run the model easily accessible by just downloading an omex archive file, to be reused as it is or as a module in a bigger model. Your publication is also connected to a primary paper published in a field specific journal, where the validation and scientific value of the model is discussed. A Physiome publication is a complement to your primary article that ensures reproducibility, reusability and discoverability of you model. The format encourages modularity that facilitates combination of different models to develop the next level of systems understanding. And all the models are in one place, easy to find and accessible.

Physiome is open access with a low Author Processing Charge. Physiome curators will help authors ensure that models and simulation experiments are made available using appropriate community standards prior to acceptance for publication in Physiome. When aspects of a computational modelling study are not able to be encoded in standard formats, Physiome editors will help authors ensure their work is as open, reproducible, and reusable as possible. We will help as much as is currently possible with model curation and annotation to ensure that the modeling results claimed in the primary paper are consistent with the published model.

Reproducibility and confirmation of results is crucial for useful science and should be incentivised. Yet, publication of it is often treated as a secondary result at best, which undermines the quality of our work. Waste no more valuable time and effort on trying to implement models from papers that lack information, or having your results lost because others cannot use them. Publish your models in Physiome and contribute to making science useful in society (and less frustrating for your colleagues).

For more information visit: https://journal.physiomeproject.org

Target audience: Anyone who wants to publish reproducible physiological models

Standards: CellML, COMBINE Archive, NeuroML, SBML, SED-ML

02:30 – 02:45 | D | **Discussion** | Room 1

02:45 – 03:00 | B | **Break** | Room 1

03:00 – 03:15 | T | **OpenCOR: how to enable reproducible science using community standards and tools** | Room 1

Speakers: Alan Garny

Abstract: OpenCOR (https://opencor.ws) is a cross-platform environment for organising, editing, simulating and analysing models that primarily consist of ordinary differential equations encoded in the CellML format.

It relies on community standards and tools such as CellML (https://cellml.org), SED-ML (https://sed-ml.org), the COMBINE archive format (http://co.mbine.org/standards/omex) and the Physiome Model Repository (PMR; https://models.physiomeproject.org).

It provides a means to create models, reuse existing models (e.g., from PMR), and collaborate on their development. Simulation experiments can be created and shared, ensuring reproducibility, which can also be achieved through Python.

Target audience: Modelers and tool developers

Standards: CellML, COMBINE Archive, SED-ML

03:15 – 03:30 | T | **CellML 2.0** | Room 1

Speakers: David Nickerson

Abstract: CellML 2.0 was released on April 17, 2020, and contains significant changes since the previous 1.1 version. Here, we will present a summary of the CellML format and the changes introduced in CellML 2.0.

The CellML format is based on the extensible markup language (XML), and is intended to enable the representation and dissemination of reproducible and reusable mathematical models of biology. The changes were motivated by community feedback on the 1.0 and 1.1 specifications, as well as a desire to separate the normative and informative aspects of the CellML specification.

Included are the introduction of a reset element which allows discontinuities to be modelled without ambiguity; the restriction of allowed MathML elements; the removal of comments and annotations; and the removal of the reaction element and its children.

The reset element is designed to allow discontinuities and discrete changes in variables based on conditional statements. This was possible in earlier versions using piecewise expressions, but the use of an order attribute within the resets means that the behaviour of conflicting switches can now be resolved uniquely. Possible applications include modelling a cell dividing as its size exceeds a threshold; the discrete switching of electrical stimuli; and re-initialising conditions during a simulation.

The normative specification is available on the CellML website at https://www.cellml.org/specifications/cellml_2.0, with an informative specification with a discussion of examples, explanations, and common mistakes also provided.

Target audience: -

Standards: CellML

03:30 – 03:45 | T | **Implementing OMEX metadata v1.1** | Room 1

Speakers: John Gennari

Abstract: Building on previous, community-driven efforts to standardize the representation of metadata (e.g., annotations) for computational models, we have developed and published the initial OMEX (Open Modeling and Exchange) metadata specification version 1.0. Our goal is to enable cross-library and cross-language model searching, composition, and reuse. Not only are there multiple modeling languages, but there are many modeling tools and simulation environments, each of which currently manipulates and provides for annotations in idiosyncratic and language-dependent ways. To solve this potential tower of Babel, the OMEX specification describes how tools should support reading and writing annotations, storing these in a separate file, so that annotations can be decoupled from specific languages and formats.

Recently, we have proposed updates to the OMEX metadata specification (v1.1) to further standardize model-level annotations and to improve the URI scheme. In particular, the new specification states that all OMEX archives will use URIs derived from “omex-library.org”. This address simply serves as a stub for all OMEX archives, thereby supporting improved portability of these archives, as they no longer use relative paths.

Here, we report on two implementations of this new v1.1 OMEX metadata specification. These are software libraries, designed to be used by other tool developers, rather than by modelers or computational biologists. The first is the C++ libOmexMeta library, which we have also made available as a Python package, “PyOmexMeta”. The second implementation, designed for Java developers, is an enhanced SemSim Java API, which undergirds the SemGen tool for model annotation and composition.

Both libraries read and write COMBINE Archives that conform to the OMEX metadata specification. They can create and manipulate annotations. In addition, both packages also support the ability to extract annotations and other metadata from existing SBML models, especially those from the BioModels repository. This capability will allow us to easily collect metadata from multiple SBML models into a single, queryable network of information. Our use of the standard “Resource Description Framework” (RDF) to serialize annotations allows for SPARQL queries across all OMEX annotations. A standardized approach to model annotation will make models more understandable, findable, interoperable and reusable, thus supporting the FAIR principles of scientific data management.

Target audience: -

Standards: CellML, COMBINE Archive, SBML

03:45 – 04:00 | D | **Discussion** | Room 1

PINNED
04:00 – 06:00 | S | **Social space/BREAK** | Room 1

PINNED
06:00 – 06:45 | I | **COVID-19 Disease Map: the key role of standards in community-driven development of systems biology disease models** | Room 1

Speakers: Marek Ostaszewski

Half a year ago, when the global threat from COVID-19 became apparent, we started an initiative to develop a knowledge repository of SARS-CoV-2 molecular mechanisms. COVID-19 Disease Map Project is a community-driven effort, involving biocurators, domain experts and bioinformaticians to develop and use systems biology diagrams for exploration of knowledge, generating new hypotheses and formulating predictions about drug targets. Because of its crowdsourcing nature it was critical for the project to establish standards and guidelines for the community to harmonise the content developed in WikiPathways, in Reactome and by using standalone diagram editors. Handling of systems biology layout supporting formats - SBML, SBGN and GPML - and ensuring their interoperability allowed to integrate the content and identify areas of overlap. This interoperability was also critical for bringing together the manually curated diagrams, and other COVID-19 related resources like interaction databases or text mining results. Finally, because of standard representations it was possible to make the content of the COVID-19 Disease Map accessible for other resources and useful in a range of different analytical pipelines, including network analysis or discrete modelling.

06:45 – 07:00 | T | **PEtab – Interoperable Specification of Parameter Estimation Problems in Systems Biology** | Room 1

Speakers: Daniel Weindl

Abstract: Reproducibility and reusability of the results of data-based modeling studies are essential. Yet, there has been – so far – no broadly supported format for the specification of parameter estimation problems in systems biology. Therefore, we developed PEtab, a format which facilitates the specification of parameter estimation problems using Systems Biology Markup Language (SBML) models and a set of tab-separated value files describing the observation model and experimental data as well as parameters to be estimated.

We already implemented PEtab support into eight well-established model simulation and parameter estimation toolboxes with hundreds of users in total. We provide a Python library for validation and modification of a PEtab problem and as well as example parameter estimation problems based on recent studies.

Specifications of PEtab, the PEtab Python library, as well as links to examples, and all supporting software tools are publicly available at https://github.com/PEtab-dev/PEtab.

Target audience: tool developers and modelers interested in parameter estimation

Standards: SED-ML, Parameter estimation

07:00 – 07:15 | T | **Spatial Model Editor** | Room 1

Speakers: Liam Keegan

Abstract: We present Spatial Model Editor, a graphical user interface tool for editing and simulating two dimensional spatial SBML models. The model geometry can be imported from segmented pixel images, or from existing spatial SBML models. The tool automatically identifies the contours of compartment boundaries in the image, and constructs a triangular mesh approximation to the geometry. A system of PDEs is constructed from the model, and solved on this triangular mesh using Finite Element methods with the dune-copasi solver. This solver, which is being developed as part of this project, is an extension of the DUNE PDE framework to solve multi-compartment reaction-diffusion equations - see https://gitlab.dune-project.org/copasi/dune-copasi for more information. A simple pixel based simulator is also included which uses the finite difference method (https://spatial-model-editor.readthedocs.io/en/latest/reference/pixel.html). In addition to the graphical user interface, there is also a python interface as well as a command line interface.

The tool is open source and is available for download for windows, linux and mac from https://github.com/lkeegan/spatial-model-editor

Target audience: -

Standards: SBML

07:15 – 07:30 | T | **pyABC: likelihood-free inference** | Room 1

Speakers: Emad Alamoodi

Abstract: Understanding the source of variation among and within populations is still a major question in system biology. Due to the continuous improvement in computational power, simulation-based methods have been intensively used to answer this question. Among these methods is Approximate Bayesian Computation (ABC). ABC is a likelihood-free inference approach that facilitates the approximation of the Bayesian posterior distribution, which problems, for which the evaluation of the likelihood is expensive or even infeasible. While ABC provides a theoretical solution, its usage is often hindered by its high computational demands.

Here, we present pyABC: a distributed and scalable ABC-Sequential Monte Carlo (ABC-SMC) framework. It implements a strategy for multi-core and distributed environments scaling to thousands of cores. The framework is easy to use and also enables advanced users to customize and experiment with many options of ABC-SMC schemes, such as acceptance threshold schedules, transition kernels, distance functions, and the use of complex noise models with stochastic acceptors without alteration of pyABC's source code. The code has been substantially improved over the last years and the methods are now used by various research groups from different research fields.

Target audience: modelers, parameter estimate, Bayesian inference, model selection

Standards: SBML

07:30 – 07:45 | D | **Discussion** | Room 1

07:45 – 08:00 | B | **Break** | Room 1

08:00 – 08:15 T **FAIRDOM: standard compliant data and model management** Room 1

Speakers: Olga Krebs

Abstract: Systems Biologists need a data management infrastructure that enables collaborating researchers to share and exchange information and data as and when it is produced, throughout the entire iterative cycle of experimentation and modelling. Data exchange and reuse rely on sufficient annotation, consistent metadata descriptions, and the use of standard exchange formats for models, data, and the experiments they are derived from. FAIRDOM offers integrated data management support for systems biology research projects within and across national and international consortia comprising a whole package of solutions. This is applied to large-scale research initiatives in which FAIRDOM members are responsible for the scientific data management, e.g. the German projects LiSyM (Liver Systems Medicine, https://lisym.org/), and MESI-STRAT (Systems Medicine of Metabolic-Signaling Networks, https://mesi-strat.eu/) , as well as the European research networks ERASysAPP (ERA-Net for Systems Biology Applications), and the Synthetic Biology Centres at Manchester (SynBioChem) and Edinburgh (SynthSys). FAIRDOM is an integral member of several infrastructure initiatives (e.g. ELIXIR, https://elixir-europe.org/ and de.NBI - German Network for Bioinformatics Infrastructure, https://www.denbi.de/)
The FAIRDOM data management concept consists of 4 major pillars:
1) Infrastructure backbone: The FAIRDOMHub/SEEK platform as registry and a commons for data, models, processes and resulting publications and presentations, at the same time yellow pages for projects, people and events
2) Terminology: Tailored use of controlled vocabularies and ontologies to describe the data
3) Modelling support: Seamless handling and simulation of models by integrated modelling platforms (JWS-Online,Cytoscape)
4) Social support: Data management advocates within the projects for gathering requirements and dissemination.

Target audience: COMBINE community, modelers, experimentalists, developers

Standards: CellML, COMBINE Archive, SBGN, SBML, SBOL and SBOL Visual, SED-ML, MIBBI, FAIRsharing

08:15 – 08:30 T **FAIR principles in literature-based kinetic modelling** Room 1

Speakers: Christoff Odendaal

Abstract: In view of the laborious nature of model building and validation, a key part of the effort of modular modelling and expansion, is avoiding duplicative work (Klipp et al., 2007). Progress has been made to standardise model reporting (e.g. SBML) and annotation (e.g. MIRIAM), and these standards are widely endorsed by systems biologists. Kinetic model-building from literature relies on making several modelling decisions based on a priori biochemical knowledge. The degree to which experimental conditions affect measurements are, however, often research questions in themselves, making it unrealistic for a modeller to exhaustively evaluate and recalculate the parameters in her model. At such points, parameter decisions must be made under substantial uncertainty.
Modelling decisions need not be final, however, and can be revised based on new information. We propose that a modelling method be developed for aligning the model-building process more with the FAIR (findable, accessible, interoperable, reusable) principles of Open Science. This would include, amongst others, an accompanying sheet of metadata for each model clearly stating not only model parameters and citations, but also all major assay conditions (pH, temperature, buffer composition, etc.) as well as explicit reasons why a certain values were preferred above others – the necessity of this is widely agreed upon, but it is still often not done. We also suggest that the author include a list of all consulted literature (also rejected literature) as well as a standard operating procedure that was used for the literature search. Such a metadata sheet would also include relations that were omitted due to unconvincing evidence as well as questions which arose in the process of model creation. This will give the modelling process more the character of a systematic review.
These measures would chart the searches that a modeller has performed to make that same course of enquiry transparent to other researchers. Furthermore, the logic of explicitly flagging unresolved questions which arose during the literature-search, would encourage complementary work by pointing directly to ways in which the model can be improved. This would allow experimentalists who are not as interested in deeply scrutinising entire models to find synergies with their work.
References: Klipp, E., Liebermeister, W., Helbig, A., Kowald, A., & Schaber, J. (2007). Nature biotechnology, 25(4), 390-391.

Target audience: Modellers

Standards: SBML, SED-ML, MIRIAM

08:30 – 08:45	T	**Stochastic Modelling and Prediction of the COVID-19** *Speakers: Joab Odhiambo* **Abstract:** Since the inception of the novel Covid-19 in December in China, the spread has been massive leading World Health Organization to declare it a world pandemic. While epicenter of COVID-19 was Wuhan city in China mainland, Italy has been affected most due to the high number of recorded deaths as at 1st April, 2020 at the same time USA recording the highest number of virus reported cases. In addition, the spread has been experienced in many developing African countries including Kenya. While the Kenyan government have had plans for those who have tested positive through self-quarantine beds at Mbagathi Hospital, lack of a proper mathematical model that can be used to model and predict the spread of Covid-19 for adequate response security has been one of the main concerns for the government. Many mathematical models have been proposed for proper modeling and forecasting, but this paper will focus on using a generalized linear regression that can detect linear relationship between the risk factors. The paper intents to model and forecast the confirmed Covid-19 cases in Kenya as a Compound Poisson process where the parameter follows a generalized linear regression that is influenced by the number of daily contact persons and daily flights with the already confirmed cases of the virus. Ultimately, this paper should assist the government in proper resource allocation to deal with pandemic in terms of available of bed capacities, public awareness campaigns and virus testing kits not only in the virus hotbed within Nairobi but also in the other 47 Kenyan counties.	Room 1
08:45 – 09:00	D	**Discussion**	Room 1
PINNED 09:00 – 10:00	S	**Social space/BREAK**	Room 1
10:00 – 10:15	T	**Automated inference of Boolean models from molecular interaction maps using CaSQ** *Speakers: Anna Niarakis* **Abstract:** Motivation: Molecular interaction maps have emerged as a meaningful way of representing biological mechanisms in a comprehensive and systematic manner. However, their static nature provides limited insights to the emerging behaviour of the described biological system under different conditions. Computational modelling provides the means to study dynamic properties through in silico simulations and perturbations. We aim to bridge the gap between static and dynamic representations of biological systems with CaSQ, a software tool that infers Boolean rules based on the topology and semantics of molecular interaction maps built with CellDesigner. Results: We developed CaSQ by defining conversion rules and logical formulas for inferred Boolean models according to the topology and the annotations of the starting molecular interaction maps. We used CaSQ to produce executable files of existing molecular maps that differ in size, complexity and the use of Systems Biology Graphical Notation (SBGN) standards. We also compared, where possible, the manually built logical models corresponding to a molecular map to the ones inferred by CaSQ. The tool is able to process large and complex maps built with CellDesigner (either following SBGN standards or not) and produce Boolean models in a standard output format, Systems Biology Marked Up Language-qualitative (SBML-qual), that can be further analyzed using popular modelling tools. References, annotations and layout of the CellDesigner molecular map are retained in the obtained model, facilitating interoperability and model reusability. Availability and implementation: The present tool is available online: https://lifeware.inria.fr/~soliman/post/casq/ and distributed as a Python package under the GNU GPLv3 license. The code can be accessed here: https://gitlab.inria.fr/soliman/casq. **Target audience:** SBML and SBGN communities, modelers, molecular map curators **Standards:** SBGN, SBML, SBML-qual	Room 1

10:15 – 10:30 | T | **MEWpy: A Metabolic Engineering Workbench for Constraint-Based Strain Optimization** | Room 1

Speakers: Vítor Pereira

Abstract: Constraint-Based Modeling (CBM) provides tools for the integrative analysis of molecular systems and quantitative prediction of physicochemical and biochemical phenotypic states. In particular, when coupled with optimization strategies, CBM based methods can be applied to find metabolic engineering designs that improve microbial production performance. Recent reviews cover the growing integration of transcriptomics, proteomics, and genomics with Genome-Scale Metabolic Models (GEMs) to improve the characterization of cell physiology and contribute to a better understanding of organisms' metabolism. Some illustrative approaches are the models of Metabolism and Expression (ME-models), Expression and Thermodynamics Flux (ETFL), and the GECKO toolbox, which integrate omics for enhanced phenotype predictions while making available a python implementation.
In this context, and given the lack of integrative tools for the increasing number of modeling approaches, we propose MEWpy, an integrated Metabolic Engineering Workbench written in Python, that offers methods to explore different classes of constraint-based models able to introduce metabolic constraints (reactions and genes), enzymatic constraints (GECKO and sMOMENT methods) and regulatory constraints (OptRAM and OptORF). MEWpy makes available panoply of both phenotype simulation and strain optimization strategies for optimized microbial production, being compatible with state-of-the-art metabolic modeling packages. Metaheuristics such as Evolutionary Algorithms, Local Search, and Simulated Annealing, also including multi objective optimization methods, drive the optimization procedures across the search space to indicate the best set of enzymes, genes, or reactions, to under/overexpress or delete to maximize the production of target compounds.
By enabling out of the box analysis and comparison of solutions obtained from different modeling approaches, all sharing a common API, MEWpy can become an essential tool for the development of microbial cells as producers of recombinant proteins and natural products. Also, MEWpy is coupled with a strain optimization database that gathers solutions, combinatorial modifications, retrieved from the literature and resulting from optimizations. Still under development and with new features being constantly added, MEWpy is already available for the community at https://github.com/biosystemsUM/mewpy.

Target audience: SBML community, modelers, tool developers

Standards: SBML

10:30 – 10:45 | T | **Troppo, a Python Framework for Tissue-Specific Reconstruction and Phenotype Prediction Using Omics Data** | Room 1

Speakers: Vítor Vieira

Abstract: Cancer is still one of the leading causes of death in the 21st century, responsible for more than one third of them. Due to its heterogeneous nature, personalized approaches such as constraint-based metabolic modelling can be particularly useful to gain insights on altered metabolic states and discover novel therapies. Recent advances in high throughput molecular biology techniques coupled with state-of-the-art computational methods for reconstruction and analysis of genome-scale metabolic models (GSMMs) are shaping the future of personalized medicine [1]. The integration of omics data with constraint-based GSMMs using context-specific model reconstruction (CSMR) algorithms has already been explored as an interesting tool providing new insights into the mechanism of several diseases [2]. One of the main drawbacks of the current set of CSMR tools is their availability. Although mostly open-source, many are only available in commercial platforms such as MATLAB.
Here we present troppo (Tissue-specific RecOnstruction and Phenotype Prediction using Omics data, https://github.com/BioSystemsUM/troppo), an open-source modular Python framework for tissue-specific reconstructions, implementing a pipeline involving data-preprocessing, reconstruction, metabolic task evaluation and gapfilling. The tool was validated with a breast cancer case study comprising the reconstruction of 320 metabolic models of the MCF7 breast cancer cell line using transcriptomics data present in the Cancer Cell Line Encyclopedia (CCLE). We demonstrate the wide range of features in our framework by assessing a large variety of omics preprocessing strategies, two distinct reconstruction algorithms as well as in-house gapfilling methods to ensure model consistency under well defined growth media.
[1] Bordbar, A., et al., (2014). Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics, 15(2), 107–120. https://doi.org/10.1038/nrg3643
[2] Folger, O., et al., (2011). Predicting selective drug targets in cancer through metabolic networks. Molecular Systems Biology, 7. https://doi.org/10.1038/msb.2011.35

Target audience: Computational biologists interested in: contextualizing and integrating omics data in metabolic models; building tissue-specific models for mammalian cells

Standards: SBML

10:45 – 11:00 | D | **Discussion** | Room 1

11:00 – 11:15 | B | **Break** | Room 1

11:15 – 11:30 T **VSM: the intuitive, general-purpose curation technology** Room 1

Speakers: Steven Vercruysse

Abstract: For every new curation project, a new curation platform needs to be developed, and for every new requirement that then emerges, its interface and database need to be updated.
To alleviate this, we developed a universal curation interface that enables scientists to easily capture any kind of information in computable form. VSM (Visual Syntax Method) allows scientists to formulate knowledge in a simple but powerful 'sentence'-like format, where all terms are linked to ontology identifiers. The syntax of how terms relate to each other is defined by a small set of intuitive rules and connectors, usable no matter how long or complex the sentence is. VSM also supports 'template' sentences where users only need to fill out a number of terms+IDs, accessible via autocomplete; and any 'sentence' can still be extended without changes to the interface. This gives users an intuitive and flexible tool to capture semantics-based information on any topic, and with any amount of context details. See https://vsm.github.io .
We present how VSM works, and how a large software development effort has recently made VSM available as an open-source web-component on GitHub. VSM's main interface is called a 'vsm-box', and is built on a collection of supporting modules in the 'vsm' GitHub project. The software is highly customizable: the 'vsm-dictionary' module provides a scaffold for connecting a vsm-box to any controlled vocabulary provider (e.g. BioPortal); the 'vsm-autocomplete' module allows customization of the content of autocomplete-items; and much more. The code is diligently covered with automated tests. The extensive documentation (via https://vsm.github.io) describes how to configure and embed a vsm-box in new projects, and enables community contributions to the VSM project itself. We encourage the community to make use of this open-source technology, and look forward to assisting new projects of VSM-based knowledge curation

Target audience: modelers, tool developers, curators, knowledge engineers

Standards: BioPAX, CellML, COMBINE Archive, NeuroML, SBGN, SBML, SBOL and SBOL Visual, SED-ML, MI2CAST

11:30 – 11:45 T **Synthetic Biology Curation Tools (SYNBICT)** Room 1

Speakers: Nicholas Roehner

Abstract: A significant hindrance to biological design is the absence of tools for annotating DNA sequence features and inferring whether these features participate in biological networks. Existing tools tend to focus on either sequence annotation or network inference alone and they generally do not support standards for representing biological designs. Without tools that bridge the gap between sequence annotation and network inference, the curation of biological designs remains a manual task that typically must be performed by a subject matter expert. Through the development of tools to automate design curation, the synthetic biology community can support workflows in which newly built strains and constructs are automatically compared to their intended design to identify potential implementation flaws. Application of curation tools to design databases can also functionally enrich their contents and help document the reuse of existing designs. Towards these goals, we have developed SYNBICT (Synthetic Biology Curation Tools), a Python application with three primary modules: annotation of sequences with biological components, inference of interactions between components, and generation of truth table representations of their higher-order function. SYNBICT supports the Synthetic Biology Open Language (SBOL) standard for representing biological designs and also permits input files belonging to bioinformatics formats such as FASTA and GenBank. We have applied SYNBICT to quality control of genetic circuit designs produced as part of the DARPA Synergistic Discovery and Design (SD2) program, and also to enrich databases such as the iGEM Registry of Standard Biological Parts.

Target audience: -

Standards: SBOL and SBOL Visual

11:45 – 12:00 | T | **Modelling the Relationship between Design and Assembly** | Room 1

Speakers: Alexis Casas

Abstract: As part of its work on automated assembly pipelines at the London DNA Foundry the Kitney Lab has identified a bottleneck between the design phase and the assembly phase of a project.
We present here the case for the introduction of a distinction between virtual, functional (design) units and physical assembly parts and the development of a simple data model to ease the integration of assembly software with existing repositories.
The assembly process itself requires the introduction of a new class of assembly-specific objects.
Assembly is not a simple concatenation operation of functional units. Methods such as Golden Gate or BASIC require the insertion of specific flanking sites between the units in order to drive the reaction towards a designed conformation - the flankers annealing into short sequences called scars. Finally, the scars are unlikely to be inert, since upstream and downstream sequences influence the properties of adjacent units such as the strength of promoters or RBS.
We define a Design Unit (DU) as a virtual unit of information encoding one or several combined functions. The design process consists of combining DUs into a higher level construct. DUs can be simple or composite-parts, e.g. Transcription Unit or a multi-input gate.
An Assembly Unit (AU) is a physical unit of DNA , practically made of three blocs: a core unit (≡DU, conserved during assembly) and prefix and suffix elements. An AU has two functions:
storage unit in a storage vector and physical unit involved in the molecular reaction of the assembly.
The distinction between DU and AU mirrors that of an object (DU) and its serialized object (AU) in software engineering. AUs and DUs are similarly related : an AU is associated to one DU and one only, and relationships between DUs are transferred to corresponding AUs - for instance, the sharing of a property among a collection of DUs (e.g function) or their ranking (e.g by strength).
We have started incorporating the DU/AU distinction into our automated assembly pipeline ; DNA CAD (Computer Aided Design) software utilise DUs, while DNA CAM (Computer Aided Manufacturing) software use AUs. A dedicated internal database is being built to link libraries of AUs and DUs, so that for a given design and a specified assembly method, the system returns the availability of the required AUs and helps us plan construction.
Target audience: synthetic biologists, DNA tool developers, genomic foundries
Standards: SBOL and SBOL Visual

12:00 – 12:15 | D | **Discussion** | Room 1

PINNED
12:15 – 13:00 | I | **Information and data standards used at Ginkgo Bioworks** | Room 1

Speakers: Ariel Hecht

Abstract: Ginkgo Bioworks uses and develops information and data standards to enable more effective coordination of labor and reuse of information. I will discuss four standards currently in use at Ginkgo, three of which were developed in-house over the past few years. We use SBOL Visual in Loom, our new DNA design and ordering tool; we've developed standards for tech transferring partner processes in-house; we've developed standards for storing and retrieving strain data; and we've developed standards for coordinating screen development and execution between multiple stakeholders across the company. Discussion will focus on motivations for choosing and using these standards, and for the ones that we have developed, the philosophies that guided the development of the standards and how they have impacted workflows.

PINNED
13:00 – 14:00 | S | **Social space/BREAK** | Room 1

14:00 – 14:15 | T | **Creating SBOL Designs with Excel** | Room 1

Speakers: Isabel Marleen Pötzsch

Abstract: SynBioHub is a repository of genetic components and designs and can represent a powerful way to share and curate synthetic biology projects. To do so effectively, it should invite and fulfil the needs of bioinformaticians and mathematicians while at the same time bridging the gap to wet lab experimentalists. For many biologists, however, creating genetic designs with the SBOL data standard can be difficult. To address this difficulty, we aim to allow submission to SynBioHub directly from spreadsheets, which are very user-friendly and part of the existing workflows in many labs. Using SynBioHub's plugin functionality and pySBOL2, it was possible to program two plugins that enable users to submit spreadsheets based on templates that are converted into valid SBOL files. The first of the two plugins parses library parts and the second interprets composite build requests from existing library parts that are downloaded from SynBioHub. These can be used in tandem or in separate workflows. This can simplify the process of data curation for publication as well as data sharing more generally. Next steps for this tool include the deployment of the tools to the user and a spreadsheet download plugin to complete the workflow.
Target audience: -
Standards: SBOL and SBOL Visual

14:15 – 14:30	T	**SABIO-RK: Curation and Visualization of Reaction Kinetics Data** *Speakers: Ulrike Wittig* **Abstract:** The SABIO-RK database (http://sabiork.h-its.org) contains manually curated data for biochemical reactions and their kinetic properties. Data in SABIO-RK are mainly manually extracted from literature but also uploaded from laboratories or other kinetic data resources. The annotations to controlled vocabularies, ontologies, and external database identifiers as well as the support of standard data exchange formats (e.g. SBML) allow the interlinkage of SABIO-RK with many different other databases and the integration in data workflows of several modelling and simulation tools. The focus of the database is both to support users working on computational modelling to create models of biochemical reaction networks and to allow experimentalists to gain further knowledge about enzymatic activities and reaction properties. To improve the usability of the web interface we are developing different ways of visualization to get a fast and flexible overview and navigate through the database content. Visualizations interact with the database search to improve the search functionalities and to refine queries. Allowing the user to focus and custom define the important aspects of the search can give new insights into the data. Besides increasing the usability the visualization supports the database curators in finding errors and inconsistencies within the database. It summarizes and clusters the data based on different scientific questions and therefore helps curators to identify outliers in the database and to easier detect curation errors. **Target audience:** - **Standards:** SBML	Room 1
14:30 – 14:45	T	**EnzymeML – an SBML-based data exchange format for biocatalysis and enzymology** *Speakers: Juergen Pleiss* **Abstract:** EnzymeML is a data exchange format that facilitates the comprehensive transfer of enzymatic data by describing reaction conditions, time course of substrate and product concentrations, the kinetic model, and the estimated kinetic constants. EnzymeML is based on standards such as SBML, MathML, and COMBINE Archive, which were extended by implementing the STRENDA Guidelines. EnzymeML was designed as a container to store and transfer data between experimental platforms, modelling platforms, and publication platforms. EnzymeML makes enzymatic data findable, accessible, interoperable, and reusable according to the FAIR data principles. An API based on Python and Java libraries and an application-specific thin API layer support the integration of applications. The feasibility of a seamless data flow using EnzymeML is demonstrated by creating an EnzymeML document on an enzymatic reaction from a spreadsheet or from a STRENDA DB entry, kinetic modelling of the reaction data by COPASI, and upload of the results to SABIO-RK. **Target audience:** - **Standards:** COMBINE Archive, SBML	Room 1
14:45 – 15:00	D	**Discussion**	Room 1
15:00 – 15:15	B	**Break**	Room 1
15:15 – 15:30	T	**Stochastic Differential Equations and their Application in Systems Biology** *Speakers: Stefan Hoops* **Abstract:** Stochastic differential equations (SDEs) present a modeling frame work suitable for simulating 'noisy' models in Systems Biology. In particular the Chemical Langevin Equation (CLE) can be represented as an SDE. We introduce a general physically correct (mass conservations are guaranteed) subset of SDEs. Furthermore we show how this subset is implementation in COPASI and discuss our solution for encoding 'noisy' models in SBML. This solution uses application specific annotations of SBase and is meant as precursor to a general solution. **Target audience:** SBML community, modelers, tool developers **Standards:** SBML	Room 1

15:30 – 15:45 | T | **Towards in silico approaches for personalized medicine – Recommendations for verifying and validating predictive computational models in EU collaborative research** | Room 1

Speakers: Catherine Collin

Abstract: Harmonization of data integration is the key to standardization efforts in personalized medicine. While standardization of models themselves is undesirable within a research context, where new models are created and tested in line with research progress – harmonization and/or standardization of input data is both feasible and necessary.

However, we argue that model validation should receive more attention, and other measures should be implemented, such that validation of models within personalised medicine becomes easier. While this is an evident necessity within the context of models implemented as medical devices, which are regulated by European Medicines Agencies and national competent authorities, we argue that model validation should be a higher priority at research level also, facilitating assessment by peers and by medical doctors – who themselves should receive better training in assessment of research using in silico models.

Acceptance by doctors and the relevant medical specialties is a key hurdle for in silico models in personalised medicine. Any medical product - device, algorithm or drug - has to prove itself safe and effective to be permitted for use by regulators; however, it has also to be accepted by medical experts as being a good choice, and be recommended within clinical specialties.

The Horizon2020 funded Coordinating and Support Action “EU-STANDS4PM” joined forces to examine to what extent existing standards or standards under development for both, format and semantics, can be used to link clinical healthcare data to computational models that build on these data. As all requirements should be equally understood and fulfilled by users, it is important to define them uniformly in an international context. To achieve this the conclusion of our work shall be also discussed in international standardization and technical committees, especially in the case of standards that are still being drawn up, and new standardization projects shall be initiated where necessary.

We present an overview of recommendations for standardization of data integration as well as recommendations for standardization of model validation within a collaborative research context, such that health-related data can be optimally used for translational research and personalized medicine across Europe.

Target audience: -

Standards: -

15:45 – 16:00 | T | **Automated Extraction of Implicit Molecular Structure from Reaction Network Models** | Room 1

Speakers: Ali Sinan Saglam

Abstract: Most mathematical models of biological processes are encoded as reaction network models (RNMs). One problem with RNMs is that they do not explicitly encode species composition information (e.g. a binding site or a phosphorylation site on a protein) and how these subunits transform during reactions (e.g. binding, phosphorylation). This makes it harder to interpret, analyze, and re-use these models. One solution to this problem is to use rule-based modelling (RBMs), where the model building blocks are structured objects that encode species composition (e.g. a protein with binding sites and/or phosphorylation sites) and reactions are between the subunits of these building blocks (e.g. binding between two proteins encoded as a reaction between two binding sites). This approach makes it easier to understand what these biological species contain and how they interact with each other while also enabling more automated approaches to model analysis. To get the benefits of RBM encoding using the rich set of existing RNMs, we created atomizer, a tool that utilizes stoichiometric and lexical analysis as well as annotation information and user input to recover the implicit assumptions made in RNMs and re-encode them as RBMs. Atomization works by using the species and reactions in a network to determine the basic set of building blocks along with their internal structure. Each species in the network is transformed into a complex comprised of these basic building blocks with bonds linking specific components and component states tracking transformations (e.g., phosphorylation). Atomization facilitates comparison of models of the same biological process by showing which building blocks (molecules) and which components (sites within molecules) are included in a model. As an example, we compared two models of IL-6 signal transduction taken from the curated BioModels database (BMD), models 151 and 543, encoded in SBML. To investigate the differences between the two models, we visualized the atomized versions using contact maps and other diagrams developed for RBMs. The contact map quickly reveals that model 151 is a subset of model 543 and the State Transition Diagram, a new form of RBM visualization that we developed, shows differences in how binding to a central scaffold molecule are modeled. We will also report the results of a comprehensive analysis of atomization of BMD models and a comparative analysis with a test set of RBMs constructed manually.

Target audience: SBML community, modellers

Standards: SBML

16:00 – 16:15 | D | **Discussion** | Room 1

PINNED
16:15 – 17:00 | I | **TBA** | Room 1

Speakers: Nathan Hillson

TBA

PINNED 17:00 – 18:00	S	**Social space/BREAK**	Room 1
PINNED 18:00 – 18:15		**Opening session**	Room 1
PINNED 18:15 – 19:00	I	**Novel technologies for systematically building and simulating whole-cell models** *Speakers: Jonathan Karr*	Room 1
19:00 – 19:30	L	**Lightning talks (1)** *Speakers: Paul Stapor, Mudasir Shaikh, Michael Blinov, Gonzalo Vidal, Adrien Rougny, Joseph Hellerstein* **Paul Stapor** *AMICI: High-Performance Sensitivity Analysis for Large Ordinary Differential Equation Models* **Mudasir Shaikh** *GeneTech - Detailed SBOL representation of genetic logic circuits along with improved designing capability* **Michael Blinov** *ModelBricks - Annotated reproducible modeling blocks* **Gonzalo Vidal** *LOICA: Logical Operators for Intelligent Cell Algorithms* **Adrien Rougny** *SBGN bricks as a tool to describe recurring concepts in molecular networks* **Joseph Hellerstein** SBStoat: Simple and Scalable Parameter Fitting for Kinetic Models in Systems Biology	Room 1
19:30 – 19:45	D	**Discussion**	Room 1
19:45 – 20:00	B	**Break**	Room 1
20:00 – 20:30	L	**Lightning talks (2)** *Speakers: Malik-Sheriff, Rahuman S., Eirini Tsirvouli, Hugh Sorby, Adel Heydarabadipour, Joab Odhiambo* **Malik-Sheriff, Rahuman S.** *COVID-archives: Rapid curation and dissemination of COVID19 model collection in BioModels* **Eirini Tsirvouli** *Logical modeling of the cPLA2 involvement in psoriatic keratinocyte phenotypes* **Hugh Sorby** *libCellML* **Adel Heydarabadipour** *TBA* **Joab Odhiambo** *TBA*	Room 1
20:30 – 21:00	D	**Wrap-ups** // **Discussion**	Room 1
PINNED 21:00 – 21:45	I	**Putting energy into systems biology: biophysical models of cell systems for understanding, simulation and design** *Speakers: Edmund Crampin* **Abstract:** Mathematical modelling is used in systems biology in order to help understand how cells work. In synthetic biology, mathematical models are used for design: to modify existing biochemical networks or to engineer entirely new ones with new desirable functions. Several different approaches have been used to mathematically model cellular systems. Most of these completely ignore energy. This is, at face value, puzzling, as energy is central to biology in general and to cell biology in particular. In this talk I will provide a lot of motivation, a little technical detail, and a few examples to make the case for an energy-based approach to modelling in systems and synthetic biology. This involves some new ideas and rediscovery of some old ones which have, by and large, been overlooked in the field. I will show that an energy-based approach has some desirable consequences for large-scale modelling, in particular, models developed in this manner are modular, hierarchical and are well suited to reuse and reproducibility. **Target audience:** modelers **Standards:** CellML, SBML, SBOL	Room 1

21:45 – 22:00 T **A thermodynamic Model of EGFR and ERK Signaling explains Adaptive and Genetic Resistance in Melanoma** Room 1

Speakers: Fabian Fröhlich

Abstract: Allosteric interactions are at the core of many signal transduction processes and provide robustness and enable context dependency for the underlying molecular mechanisms. This is prominently captured by paradoxical activation, a clinically observed phenomenon where RAF inhibitors inhibit tumor growth in BRAF mutant cancers, but promote tumor growth in BRAF wild-type cancers. Energy based formalisms to describe such allosteric effects in kinetic models have been developed, but approaches to enable intelligibility of and address the computational complexity associated with such large, multi-scale models are currently missing.

Here we demonstrate the use of a programmatic, thermodynamic, energy-balanced rule-based formalism in PySB to describe allosteric interactions. We tackle the numerical challenges of large kinetic models by using and extending state of the art high performance computing simulation and calibration tools. To address the conceptual challenge of rendering large kinetic models intelligible, we introduce a novel approach to causally separate intertwined signaling channels.

We apply these methods to an ordinary differential equation model of adaptive resistance in melanoma (EGFR and ERK pathways, >1k state variables, >10k reactions), accounting for paradoxical activation.

We trained the model on absolute proteomic and phospho-proteomic as well as time-resolved immunofluorescence data, both in dose-response to small molecule inhibitors. We deconvolve oncogenic and physiological causal paths to derive simple explanations for complex dose-response relationships, explain how synergy and antagonism can arise without direct drug interaction and establish a link between adaptive and genetic resistance in melanoma.

Target audience: modelers, tool developers

Standards: SBML

22:00 – 22:15 T **The Systems Biology Graphical Notation: a standardised representation of biological maps** Room 1

Speakers: Michael Blinov

Abstract: Visualization of biological processes plays an essential role in life science research. Over time, diverse forms of diagrammatic representations, akin to circuit diagrams, have evolved without well-defined semantics potentially leading to ambiguous network interpretations and difficult programmatic processing.

The Systems Biology Graphical Notation (SBGN) standard aims to reduce ambiguity in the visual representation of biomolecular networks. It provides specific sets of well-defined symbols for various types of biological concepts. SBGN comprises three complementary languages: Process Description (PD), Entity Relationship (ER), and Activity Flow (AF). The XML-based SBGN Markup Language (SBGN-ML) facilitates convenient storage and exchange of SBGN maps, supported by the library libSBGN. The SBGN project is an ongoing open community-driven effort coordinated and maintained by an elected international editorial board. All documents and source code are freely available at http://sbgn.org.

Target audience: all communities

Standards: SBGN, SBML

22:15 – 22:30	T	**BioSimulators: a registry of containerized biosimulation tools with standard interfaces that enhance the reuse of biomodels** *Speakers: Bilal Shaikh* **Abstract:** Motivation & background: More predictive models could advance biology, medicine, and bioengineering. Building such models will likely require teams who can share and reuse models. Several standards, including CellML, KiSAO, NeuroML, OMEX, SBML, SBO, and SED-ML, facilitate model sharing. To support multiple formalisms such as logical, FBA, and kinetic modeling, these standards include numerous features such as delays, events, and objectives. Hundreds of simulators support many of these features. However, it remains challenging to reuse many models. The incomplete support for these features among simulators and the lack of centralized documentation of the features supported by each simulator often make it difficult to find a simulator for a model. Furthermore, simulators frequently become unavailable when projects terminate, old versions of tools that support old versions of features are often hard to obtain, many tools are cumbersome to install, and it takes significant effort to learn a different interface for each tool. Results: To facilitate model reuse, we are developing Biosimulators, a registry of standardized containers of biosimulation tools. BioSimulators will help researchers find simulators by tracking the formalisms, algorithms, and formats that they support. BioSimulators will also help researchers use these tools by encapsulating them into consistent command-line interfaces inside containers. By facilitating reuse, we anticipate that the registry will promote more predictive models. Methods: To assemble the registry, we have developed a standard command-line interface and container structure for biosimulation tools; a standard format for specifying the capabilities (supported formalisms, algorithms, and formats) of a simulator; and a tool for verifying the capabilities of a container; and a website for finding containers with specific capabilities. We have used these resources to develop several containers for logical, FBA, ODE, stochastic, network-free, and DAE tools. Future directions: We aim to work with the community to develop additional containers that support other formalisms, algorithms, and formats. We plan to expand the testing tool to verify each container's capabilities more rigorously. To make it even easier to reuse models, we are using the registry to develop a simple web application for executing simulations. Availability: BioSimulators is openly available at https://biosimulators.org along with examples and documentation. **Target audience:** Biomodeling community; Investigators who want to quickly try out and reuse published models or access older versions of simulation tools; Simulation software developers who want to easily distribute their simulation tools or compare the accuracy or performance of their tools with other tools; Peer reviewers and editors who want to evaluate models pre-publication. **Standards:** CellML, COMBINE Archive, NeuroML, SBML, SED-ML, BioContainers, EDAM, KiSAO, SBO	Room 1
22:30 – 22:45	D	**Discussion**	Room 1
22:45 – 23:00	B	**Break**	Room 1
23:00 – 23:15	T	**Datanator: an integrated database of molecular data for quantitatively modeling cellular behavior** *Speakers: Lian Zhouyang* **Abstract:** Integrative research about multiple biochemical subsystems has significant potential to help advance biology, bioengineering, and medicine. However, it is difficult to obtain the diverse data needed for integrative research. To facilitate biochemical research, we developed Datanator (https://datanator.info), an integrated database and set of tools for finding of multiple types of molecular data about specific molecules and reactions in specific organisms and environments, as well as data about chemically-similar molecules and reactions in phylogenetically-similar organisms in similar environments. Currently, Datanator includes metabolite concentrations, RNA modifications and half-lives, protein abundances and modifications, and reaction rate constants about a broad range of organisms. Going forward, we aim to launch a community initiative to curate additional data. Datanator also provides tools for filtering, visualizing, and exporting these data clouds. We believe that Datanator can facilitate a wide range of research from integrative mechanistic models, such as whole-cell models, to comparative data-driven analyses of multiple organisms. **Target audience:** modelers, tool developers **Standards:** BioPAX, CellML, SBML, SED-ML	Room 1

23:15 – 23:30 | T | **SBViper: Verification Testing of Kinetics Models in Systems Biology** | Room 1

Speakers: Joseph L Hellerstein

Abstract: Over the last decade, kinetics models in systems biology have grown in size from small tens of reactions to hundreds and thousands. This growth in complexity has increased interest in testing model correctness.

There are two broad approaches to testing modeling correctness. Validation testing seeks to assess the accuracy and utility of the models. This addresses questions such as: Does the model accurately predict values of new experimental results? Does the model provide helpful insights, such as chemical pathways to target? Validation tests are constructed in a manner that is specific to the model and how it will be used.

Verification testing checks for errors in the model implementation and/or the way in which the model is executed. To illustrate, consider the following incorrect implementation of a kinetics model of a two species linear pathway with mass action kinetics:

R1: S1 -> S2; k1*S1

R2: S2 -> S3; k2*S1

(The kinetics expression follows the semi-colon.) We see that there is a typographical error in the kinetics expression for R2 in that S1 appears instead of S2. This incorrectly implemented reaction network is still a linear pathway, but it does not have the desired kinetics. As model complexity grows, it will become increasingly common to introduce such errors and increasingly difficult to discover them.

This project is developing tools of verification testing of kinetics models. While there has been substantial work with verification of constraint-based models (e.g., mass balance analysis and detection of blocked reactions), we are unaware of work to date that addresses kinetics models. We are developing codes in three areas:

1. Testing model specifications. These tests do not require running simulation codes. An example is verification of kinetics expressions (e.g., detecting undefined chemical species).

2. Verifying simulation results. These tests analyze the results of simulation runs, such as whether the concentration of a chemical species converges to a particular value. The purpose of these checks is to verify that the model operates as the modeler intended.

3. An infrastructure for test execution and reuse. We find that some tests are applicable to many kinetics models. We have created an infrastructure that supports test reuse.

Target audience: SBML community, modelers, tool developers

Standards: SBML, SED-ML

23:30 – 23:45 | T | **pyPESTO: A python package for Parameter Estimation and Uncertainty Quantification** | Room 1

Speakers: Jakob Vanhoefer

Abstract: Fitting a mathematical model to experimental data is an integral step in systems biology. A likelihood or posterior distribution quantifies the probability of the observed data under a given model parameterization. Parameter estimates are commonly derived by maximizing the likelihood or posterior distribution. This is typically done via numerical optimization, which is computationally demanding.

Uncertainty quantification accesses uncertainties in parameter estimates and model predictions due to noisy data. Established methods for uncertainty quantification include profile- or sampling-based methods.

Here, we present pyPESTO, a python package for optimization and uncertainty quantification. pyPESTO provides an interface to the model simulation tool AMICI and hence is tailored, but not restricted, to Ordinary Differential Equation models in the SBML and PEtab standard. For optimization, pyPESTO interfaces gradient and Hessian based optimizers from SciPy, dlib, and the interior point Optimizer IpOpt, as well as gradient free Particle Swarm optimizers (pyswarm).

For uncertainty quantification pyPESTO implements the profile likelihood method as well as various sampling algorithms like adaptive Metropolis, adaptive parallel tempering and a No-U-Turn-Sampler (via an interface to pymc3). pyPESTO offers functionality to parallelize multi-start optimization and sampling and store optimization results. Furthermore, pyPESTO provides visualizations of parameter estimation results, profile likelihood estimations and sampling results. Further extensions e.g. to model selection and experimental design are currently under development.

Target audience: modelers, SBML community, persons interested in parameter estimation

Standards: SBML

23:45 – 00:00 | D | **Discussion** | Room 1

COMBINE 2020

[B] Break [D] Discussion [I] Invited Talk [S] Social space [T] Talk

OCTOBER 6 • TUESDAY

PINNED 00:00 – 01:00	S	**Social space/BREAK**	Room 1
01:00 – 01:15	T	**(Replay) Automated inference of Boolean models from molecular interaction maps using CaSQ** *Speakers: Anna Niarakis* **Abstract:** Motivation: Molecular interaction maps have emerged as a meaningful way of representing biological mechanisms in a comprehensive and systematic manner. However, their static nature provides limited insights to the emerging behaviour of the described biological system under different conditions. Computational modelling provides the means to study dynamic properties through in silico simulations and perturbations. We aim to bridge the gap between static and dynamic representations of biological systems with CaSQ, a software tool that infers Boolean rules based on the topology and semantics of molecular interaction maps built with CellDesigner. Results: We developed CaSQ by defining conversion rules and logical formulas for inferred Boolean models according to the topology and the annotations of the starting molecular interaction maps. We used CaSQ to produce executable files of existing molecular maps that differ in size, complexity and the use of Systems Biology Graphical Notation (SBGN) standards. We also compared, where possible, the manually built logical models corresponding to a molecular map to the ones inferred by CaSQ. The tool is able to process large and complex maps built with CellDesigner (either following SBGN standards or not) and produce Boolean models in a standard output format, Systems Biology Marked Up Language-qualitative (SBML-qual), that can be further analyzed using popular modelling tools. References, annotations and layout of the CellDesigner molecular map are retained in the obtained model, facilitating interoperability and model reusability. Availability and implementation: The present tool is available online: https://lifeware.inria.fr/~soliman/post/casq/ and distributed as a Python package under the GNU GPLv3 license. The code can be accessed here: https://gitlab.inria.fr/soliman/casq. **Target audience:** SBML and SBGN communities, modelers, molecular map curators **Standards:** SBGN, SBML, SBML-qual	Room 1
01:15 – 01:30	T	**(Replay) MEWpy: A Metabolic Engineering Workbench for Constraint-Based Strain Optimization** *Speakers: Vítor Pereira* **Abstract:** Constraint-Based Modeling (CBM) provides tools for the integrative analysis of molecular systems and quantitative prediction of physicochemical and biochemical phenotypic states. In particular, when coupled with optimization strategies, CBM based methods can be applied to find metabolic engineering designs that improve microbial production performance. Recent reviews cover the growing integration of transcriptomics, proteomics, and genomics with Genome-Scale Metabolic Models (GEMs) to improve the characterization of cell physiology and contribute to a better understanding of organisms' metabolism. Some illustrative approaches are the models of Metabolism and Expression (ME-models), Expression and Thermodynamics Flux (ETFL), and the GECKO toolbox, which integrate omics for enhanced phenotype predictions while making available a python implementation. In this context, and given the lack of integrative tools for the increasing number of modeling approaches, we propose MEWpy, an integrated Metabolic Engineering Workbench written in Python, that offers methods to explore different classes of constraint-based models able to introduce metabolic constraints (reactions and genes), enzymatic constraints (GECKO and sMOMENT methods) and regulatory constraints (OptRAM and OptORF). MEWpy makes available panoply of both phenotype simulation and strain optimization strategies for optimized microbial production, being compatible with state-of-the-art metabolic modeling packages. Metaheuristics such as Evolutionary Algorithms, Local Search, and Simulated Annealing, also including multi objective optimization methods, drive the optimization procedures across the search space to indicate the best set of enzymes, genes, or reactions, to under/overexpress or delete to maximize the production of target compounds. By enabling out of the box analysis and comparison of solutions obtained from different modeling approaches, all sharing a common API, MEWpy can become an essential tool for the development of microbial cells as producers of recombinant proteins and natural products. Also, MEWpy is coupled with a strain optimization database that gathers solutions, combinatorial modifications, retrieved from the literature and resulting from optimizations. Still under development and with new features being constantly added, MEWpy is already available for the community at https://github.com/biosystemsUM/mewpy. **Target audience:** SBML community, modelers, tool developers **Standards:** SBML	Room 1

01:30 – 01:45	T	**(Replay) Troppo, a Python Framework for Tissue-Specific Reconstruction and Phenotype Prediction Using Omics Data** *Speakers: Vítor Vieira* **Abstract:** Cancer is still one of the leading causes of death in the 21st century, responsible for more than one third of them. Due to its heterogeneous nature, personalized approaches such as constraint-based metabolic modelling can be particularly useful to gain insights on altered metabolic states and discover novel therapies. Recent advances in high throughput molecular biology techniques coupled with state-of-the-art computational methods for reconstruction and analysis of genome-scale metabolic models (GSMMs) are shaping the future of personalized medicine [1]. The integration of omics data with constraint-based GSMMs using context-specific model reconstruction (CSMR) algorithms has already been explored as an interesting tool providing new insights into the mechanism of several diseases [2]. One of the main drawbacks of the current set of CSMR tools is their availability. Although mostly open-source, many are only available in commercial platforms such as MATLAB. Here we present troppo (Tissue-specific RecOnstruction and Phenotype Prediction using Omics data, https://github.com/BioSystemsUM/troppo), an open-source modular Python framework for tissue-specific reconstructions, implementing a pipeline involving data-preprocessing, reconstruction, metabolic task evaluation and gapfilling. The tool was validated with a breast cancer case study comprising the reconstruction of 320 metabolic models of the MCF7 breast cancer cell line using transcriptomics data present in the Cancer Cell Line Encyclopedia (CCLE). We demonstrate the wide range of features in our framework by assessing a large variety of omics preprocessing strategies, two distinct reconstruction algorithms as well as in-house gapfilling methods to ensure model consistency under well defined growth media. [1] Bordbar, A., et al., (2014). Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics, 15(2), 107–120. https://doi.org/10.1038/nrg3643 [2] Folger, O., et al., (2011). Predicting selective drug targets in cancer through metabolic networks. Molecular Systems Biology, 7. https://doi.org/10.1038/msb.2011.35 **Target audience:** Computational biologists interested in: contextualizing and integrating omics data in metabolic models; building tissue-specific models for mammalian cells **Standards:** SBML	Room 1
01:45 – 02:00	D	**Discussion**	Room 1
02:00 – 02:15	B	**Break**	Room 1
02:15 – 02:30	T	**(Replay) VSM: the intuitive, general-purpose curation technology** *Speakers: Steven Vercruysse* **Abstract:** For every new curation project, a new curation platform needs to be developed, and for every new requirement that then emerges, its interface and database need to be updated. To alleviate this, we developed a universal curation interface that enables scientists to easily capture any kind of information in computable form. VSM (Visual Syntax Method) allows scientists to formulate knowledge in a simple but powerful 'sentence'-like format, where all terms are linked to ontology identifiers. The syntax of how terms relate to each other is defined by a small set of intuitive rules and connectors, usable no matter how long or complex the sentence is. VSM also supports 'template' sentences where users only need to fill out a number of terms+IDs, accessible via autocomplete; and any 'sentence' can still be extended without changes to the interface. This gives users an intuitive and flexible tool to capture semantics-based information on any topic, and with any amount of context details. See https://vsm.github.io. We present how VSM works, and how a large software development effort has recently made VSM available as an open-source web-component on GitHub. VSM's main interface is called a 'vsm-box', and is built on a collection of supporting modules in the 'vsm' GitHub project. The software is highly customizable: the 'vsm-dictionary' module provides a scaffold for connecting a vsm-box to any controlled vocabulary provider (e.g. BioPortal); the 'vsm-autocomplete' module allows customization of the content of autocomplete-items; and much more. The code is diligently covered with automated tests. The extensive documentation (via https://vsm.github.io) describes how to configure and embed a vsm-box in new projects, and enables community contributions to the VSM project itself. We encourage the community to make use of this open-source technology, and look forward to assisting new projects of VSM-based knowledge curation **Target audience:** modelers, tool developers, curators, knowledge engineers **Standards**: BioPAX, CellML, COMBINE Archive, NeuroML, SBGN, SBML, SBOL and SBOL Visual, SED-ML, MI2CAST	Room 1

02:30 – 02:45 | T | **(Replay) Synthetic Biology Curation Tools (SYNBICT)** | Room 1

Speakers: Nicholas Roehner

Abstract: A significant hindrance to biological design is the absence of tools for annotating DNA sequence features and inferring whether these features participate in biological networks. Existing tools tend to focus on either sequence annotation or network inference alone and they generally do not support standards for representing biological designs. Without tools that bridge the gap between sequence annotation and network inference, the curation of biological designs remains a manual task that typically must be performed by a subject matter expert. Through the development of tools to automate design curation, the synthetic biology community can support workflows in which newly built strains and constructs are automatically compared to their intended design to identify potential implementation flaws. Application of curation tools to design databases can also functionally enrich their contents and help document the reuse of existing designs. Towards these goals, we have developed SYNBICT (Synthetic Biology Curation Tools), a Python application with three primary modules: annotation of sequences with biological components, inference of interactions between components, and generation of truth table representations of their higher-order function. SYNBICT supports the Synthetic Biology Open Language (SBOL) standard for representing biological designs and also permits input files belonging to bioinformatics formats such as FASTA and GenBank. We have applied SYNBICT to quality control of genetic circuit designs produced as part of the DARPA Synergistic Discovery and Design (SD2) program, and also to enrich databases such as the iGEM Registry of Standard Biological Parts.

Target audience: -

Standards: SBOL and SBOL Visual

02:45 – 03:00 | T | **(Replay) Modelling the Relationship between Design and Assembly** | Room 1

Speakers: Alexis Casas

Abstract: As part of its work on automated assembly pipelines at the London DNA Foundry the Kitney Lab has identified a bottleneck between the design phase and the assembly phase of a project.

We present here the case for the introduction of a distinction between virtual, functional (design) units and physical assembly parts and the development of a simple data model to ease the integration of assembly software with existing repositories.

The assembly process itself requires the introduction of a new class of assembly-specific objects.

Assembly is not a simple concatenation operation of functional units. Methods such as Golden Gate or BASIC require the insertion of specific flanking sites between the units in order to drive the reaction towards a designed conformation - the flankers annealing into short sequences called scars. Finally, the scars are unlikely to be inert, since upstream and downstream sequences influence the properties of adjacent units such as the strength of promoters or RBS.

We define a Design Unit (DU) as a virtual unit of information encoding one or several combined functions. The design process consists of combining DUs into a higher level construct. DUs can be simple or composite-parts, e.g. Transcription Unit or a multi-input gate.

An Assembly Unit (AU) is a physical unit of DNA , practically made of three blocs: a core unit (≡DU, conserved during assembly) and prefix and suffix elements. An AU has two functions:

storage unit in a storage vector and physical unit involved in the molecular reaction of the assembly.

The distinction between DU and AU mirrors that of an object (DU) and its serialized object (AU) in software engineering. AUs and DUs are similarly related : an AU is associated to one DU and one only, and relationships between DUs are transferred to corresponding AUs - for instance, the sharing of a property among a collection of DUs (e.g function) or their ranking (e.g by strength).

We have started incorporating the DU/AU distinction into our automated assembly pipeline ; DNA CAD (Computer Aided Design) software utilise DUs, while DNA CAM (Computer Aided Manufacturing) software use AUs. A dedicated internal database is being built to link libraries of AUs and DUs, so that for a given design and a specified assembly method, the system returns the availability of the required AUs and helps us plan construction.

Target audience: synthetic biologists, DNA tool developers, genomic foundries

Standards: SBOL and SBOL Visual

03:00 – 03:15 | D | **Discussion** | Room 1

PINNED

03:15 – 04:00 | I | **(Replay + Live Discussion) Information and data standards used at Ginkgo Bioworks** | Room 1

Speakers: Ariel Hecht

Abstract: Ginkgo Bioworks uses and develops information and data standards to enable more effective coordination of labor and reuse of information. I will discuss four standards currently in use at Ginkgo, three of which were developed in-house over the past few years. We use SBOL Visual in Loom, our new DNA design and ordering tool; we've developed standards for tech transferring partner processes in-house; we've developed standards for storing and retrieving strain data; and we've developed standards for coordinating screen development and execution between multiple stakeholders across the company. Discussion will focus on motivations for choosing and using these standards, and for the ones that we have developed, the philosophies that guided the development of the standards and how they have impacted workflows.

PINNED 04:00 – 06:00	S	**Social space/BREAK**	Room 1
06:00 – 06:15	T	**(Replay) Creating SBOL Designs with Excel** *Speakers: Isabel Marleen Pötzsch* **Abstract:** SynBioHub is a repository of genetic components and designs and can represent a powerful way to share and curate synthetic biology projects. To do so effectively, it should invite and fulfil the needs of bioinformaticians and mathematicians while at the same time bridging the gap to wet lab experimentalists. For many biologists, however, creating genetic designs with the SBOL data standard can be difficult. To address this difficulty, we aim to allow submission to SynBioHub directly from spreadsheets, which are very user-friendly and part of the existing workflows in many labs. Using SynBioHub's plugin functionality and pySBOL2, it was possible to program two plugins that enable users to submit spreadsheets based on templates that are converted into valid SBOL files. The first of the two plugins parses library parts and the second interprets composite build requests from existing library parts that are downloaded from SynBioHub. These can be used in tandem or in separate workflows. This can simplify the process of data curation for publication as well as data sharing more generally. Next steps for this tool include the deployment of the tools to the user and a spreadsheet download plugin to complete the workflow. **Target audience:** - **Standards:** SBOL and SBOL Visual	Room 1
06:15 – 06:30	T	**(Replay) SABIO-RK: Curation and Visualization of Reaction Kinetics Data** *Speakers: Ulrike Wittig* **Abstract:** The SABIO-RK database (http://sabiork.h-its.org) contains manually curated data for biochemical reactions and their kinetic properties. Data in SABIO-RK are mainly manually extracted from literature but also uploaded from laboratories or other kinetic data resources. The annotations to controlled vocabularies, ontologies, and external database identifiers as well as the support of standard data exchange formats (e.g. SBML) allow the interlinkage of SABIO-RK with many different other databases and the integration in data workflows of several modelling and simulation tools. The focus of the database is both to support users working on computational modelling to create models of biochemical reaction networks and to allow experimentalists to gain further knowledge about enzymatic activities and reaction properties. To improve the usability of the web interface we are developing different ways of visualization to get a fast and flexible overview and navigate through the database content. Visualizations interact with the database search to improve the search functionalities and to refine queries. Allowing the user to focus and custom define the important aspects of the search can give new insights into the data. Besides increasing the usability the visualization supports the database curators in finding errors and inconsistencies within the database. It summarizes and clusters the data based on different scientific questions and therefore helps curators to identify outliers in the database and to easier detect curation errors. **Target audience:** - **Standards:** SBML	Room 1
06:30 – 06:45	T	**(Replay) EnzymeML – an SBML-based data exchange format for biocatalysis and enzymology** *Speakers: Juergen Pleiss* **Abstract:** EnzymeML is a data exchange format that facilitates the comprehensive transfer of enzymatic data by describing reaction conditions, time course of substrate and product concentrations, the kinetic model, and the estimated kinetic constants. EnzymeML is based on standards such as SBML, MathML, and COMBINE Archive, which were extended by implementing the STRENDA Guidelines. EnzymeML was designed as a container to store and transfer data between experimental platforms, modelling platforms, and publication platforms. EnzymeML makes enzymatic data findable, accessible, interoperable, and reusable according to the FAIR data principles. An API based on Python and Java libraries and an application-specific thin API layer support the integration of applications. The feasibility of a seamless data flow using EnzymeML is demonstrated by creating an EnzymeML document on an enzymatic reaction from a spreadsheet or from a STRENDA DB entry, kinetic modelling of the reaction data by COPASI, and upload of the results to SABIO-RK. **Target audience:** - **Standards:** COMBINE Archive, SBML	Room 1
06:45 – 07:00	D	**Discussion**	Room 1
07:00 – 07:15	B	**Break**	Room 1

07:15 – 07:30 T **(Replay) Stochastic Differential Equations and their Application in Systems Biology** Room 1

Speakers: Stefan Hoops

Abstract: Stochastic differential equations (SDEs) present a modeling frame work suitable for simulating 'noisy' models in Systems Biology. In particular the Chemical Langevin Equation (CLE) can be represented as an SDE. We introduce a general physically correct (mass conservations are guaranteed) subset of SDEs. Furthermore we show how this subset is implementation in COPASI and discuss our solution for encoding 'noisy' models in SBML. This solution uses application specific annotations of SBase and is meant as precursor to a general solution.

Target audience: SBML community, modelers, tool developers

Standards: SBML

07:30 – 07:45 T **(Replay) Towards in silico approaches for personalized medicine – Recommendations for verifying and validating predictive computational models in EU collaborative research** Room 1

Speakers: Catherine Collin

Abstract: Harmonization of data integration is the key to standardization efforts in personalized medicine. While standardization of models themselves is undesirable within a research context, where new models are created and tested in line with research progress – harmonization and/or standardization of input data is both feasible and necessary.

However, we argue that model validation should receive more attention, and other measures should be implemented, such that validation of models within personalised medicine becomes easier. While this is an evident necessity within the context of models implemented as medical devices, which are regulated by European Medicines Agencies and national competent authorities, we argue that model validation should be a higher priority at research level also, facilitating assessment by peers and by medical doctors – who themselves should receive better training in assessment of research using in silico models.

Acceptance by doctors and the relevant medical specialties is a key hurdle for in silico models in personalised medicine. Any medical product - device, algorithm or drug - has to prove itself safe and effective to be permitted for use by regulators; however, it has also to be accepted by medical experts as being a good choice, and be recommended within clinical specialties.

The Horizon2020 funded Coordinating and Support Action “EU-STANDS4PM” joined forces to examine to what extent existing standards or standards under development for both, format and semantics, can be used to link clinical healthcare data to computational models that build on these data. As all requirements should be equally understood and fulfilled by users, it is important to define them uniformly in an international context. To achieve this the conclusion of our work shall be also discussed in international standardization and technical committees, especially in the case of standards that are still being drawn up, and new standardization projects shall be initiated where necessary.

We present an overview of recommendations for standardization of data integration as well as recommendations for standardization of model validation within a collaborative research context, such that health-related data can be optimally used for translational research and personalized medicine across Europe.

Target audience: -

Standards: -

07:45 – 08:00 | T | **(Replay) Automated Extraction of Implicit Molecular Structure from Reaction Network Models** | Room 1

Speakers: Ali Sinan Saglam

Abstract: Most mathematical models of biological processes are encoded as reaction network models (RNMs). One problem with RNMs is that they do not explicitly encode species composition information (e.g. a binding site or a phosphorylation site on a protein) and how these subunits transform during reactions (e.g. binding, phosphorylation). This makes it harder to interpret, analyze, and re-use these models. One solution to this problem is to use rule-based modelling (RBMs), where the model building blocks are structured objects that encode species composition (e.g. a protein with binding sites and/or phosphorylation sites) and reactions are between the subunits of these building blocks (e.g. binding between two proteins encoded as a reaction between two binding sites). This approach makes it easier to understand what these biological species contain and how they interact with each other while also enabling more automated approaches to model analysis. To get the benefits of RBM encoding using the rich set of existing RNMs, we created atomizer, a tool that utilizes stoichiometric and lexical analysis as well as annotation information and user input to recover the implicit assumptions made in RNMs and re-encode them as RBMs. Atomization works by using the species and reactions in a network to determine the basic set of building blocks along with their internal structure. Each species in the network is transformed into a complex comprised of these basic building blocks with bonds linking specific components and component states tracking transformations (e.g., phosphorylation). Atomization facilitates comparison of models of the same biological process by showing which building blocks (molecules) and which components (sites within molecules) are included in a model. As an example, we compared two models of IL-6 signal transduction taken from the curated BioModels database (BMD), models 151 and 543, encoded in SBML. To investigate the differences between the two models, we visualized the atomized versions using contact maps and other diagrams developed for RBMs. The contact map quickly reveals that model 151 is a subset of model 543 and the State Transition Diagram, a new form of RBM visualization that we developed, shows differences in how binding to a central scaffold molecule are modeled. We will also report the results of a comprehensive analysis of atomization of BMD models and a comparative analysis with a test set of RBMs constructed manually.

Target audience: SBML community, modellers

Standards: SBML

08:00 – 08:15 | D | **Discussion** | Room 1

PINNED
08:15 – 09:00 | I | **(Replay + Live Discussion) TBA** | Room 1

Speakers: Nathan Hillson

TBA

PINNED
09:00 – 10:00 | S | **Social space/BREAK** | Room 1

PINNED
10:00 – 10:45 | I | **(Replay + Live Discussion) Putting energy into systems biology: biophysical models of cell systems for understanding, simulation and design** | Room 1

Speakers: Edmund Crampin

Abstract: Mathematical modelling is used in systems biology in order to help understand how cells work. In synthetic biology, mathematical models are used for design: to modify existing biochemical networks or to engineer entirely new ones with new desirable functions. Several different approaches have been used to mathematically model cellular systems. Most of these completely ignore energy. This is, at face value, puzzling, as energy is central to biology in general and to cell biology in particular.

In this talk I will provide a lot of motivation, a little technical detail, and a few examples to make the case for an energy-based approach to modelling in systems and synthetic biology. This involves some new ideas and rediscovery of some old ones which have, by and large, been overlooked in the field. I will show that an energy-based approach has some desirable consequences for large-scale modelling, in particular, models developed in this manner are modular, hierarchical and are well suited to reuse and reproducibility.

Target audience: modelers

Standards: CellML, SBML, SBOL

10:45 – 11:00 T **(Replay) A thermodynamic Model of EGFR and ERK Signaling explains Adaptive and Genetic Resistance in Melanoma** Room 1

Speakers: Fabian Fröhlich

Abstract: Allosteric interactions are at the core of many signal transduction processes and provide robustness and enable context dependency for the underlying molecular mechanisms. This is prominently captured by paradoxical activation, a clinically observed phenomenon where RAF inhibitors inhibit tumor growth in BRAF mutant cancers, but promote tumor growth in BRAF wild-type cancers. Energy based formalisms to describe such allosteric effects in kinetic models have been developed, but approaches to enable intelligibility of and address the computational complexity associated with such large, multi-scale models are currently missing.

Here we demonstrate the use of a programmatic, thermodynamic, energy-balanced rule-based formalism in PySB to describe allosteric interactions. We tackle the numerical challenges of large kinetic models by using and extending state of the art high performance computing simulation and calibration tools. To address the conceptual challenge of rendering large kinetic models intelligible, we introduce a novel approach to causally separate intertwined signaling channels.

We apply these methods to an ordinary differential equation model of adaptive resistance in melanoma (EGFR and ERK pathways, >1k state variables, >10k reactions), accounting for paradoxical activation.

We trained the model on absolute proteomic and phospho-proteomic as well as time-resolved immunofluorescence data, both in dose-response to small molecule inhibitors. We deconvolve oncogenic and physiological causal paths to derive simple explanations for complex dose-response relationships, explain how synergy and antagonism can arise without direct drug interaction and establish a link between adaptive and genetic resistance in melanoma.

Target audience: modelers, tool developers

Standards: SBML

11:00 – 11:15 T **(Replay) The Systems Biology Graphical Notation: a standardised representation of biological maps** Room 1

Speakers: Michael Blinov

Abstract: Visualization of biological processes plays an essential role in life science research. Over time, diverse forms of diagrammatic representations, akin to circuit diagrams, have evolved without well-defined semantics potentially leading to ambiguous network interpretations and difficult programmatic processing.

The Systems Biology Graphical Notation (SBGN) standard aims to reduce ambiguity in the visual representation of biomolecular networks. It provides specific sets of well-defined symbols for various types of biological concepts. SBGN comprises three complementary languages: Process Description (PD), Entity Relationship (ER), and Activity Flow (AF). The XML-based SBGN Markup Language (SBGN-ML) facilitates convenient storage and exchange of SBGN maps, supported by the library libSBGN. The SBGN project is an ongoing open community-driven effort coordinated and maintained by an elected international editorial board. All documents and source code are freely available at http://sbgn.org.

Target audience: all communities

Standards: SBGN, SBML

11:15 – 11:30 T **(Replay) BioSimulators: a registry of containerized biosimulation tools with standard interfaces that enhance the reuse of biomodels** Room 1

Speakers: Bilal Shaikh

Abstract: Motivation & background: More predictive models could advance biology, medicine, and bioengineering. Building such models will likely require teams who can share and reuse models. Several standards, including CellML, KiSAO, NeuroML, OMEX, SBML, SBO, and SED-ML, facilitate model sharing. To support multiple formalisms such as logical, FBA, and kinetic modeling, these standards include numerous features such as delays, events, and objectives. Hundreds of simulators support many of these features. However, it remains challenging to reuse many models. The incomplete support for these features among simulators and the lack of centralized documentation of the features supported by each simulator often make it difficult to find a simulator for a model. Furthermore, simulators frequently become unavailable when projects terminate, old versions of tools that support old versions of features are often hard to obtain, many tools are cumbersome to install, and it takes significant effort to learn a different interface for each tool.
Results: To facilitate model reuse, we are developing Biosimulators, a registry of standardized containers of biosimulation tools. BioSimulators will help researchers find simulators by tracking the formalisms, algorithms, and formats that they support. BioSimulators will also help researchers use these tools by encapsulating them into consistent command-line interfaces inside containers. By facilitating reuse, we anticipate that the registry will promote more predictive models.
Methods: To assemble the registry, we have developed a standard command-line interface and container structure for biosimulation tools; a standard format for specifying the capabilities (supported formalisms, algorithms, and formats) of a simulator; and a tool for verifying the capabilities of a container; and a website for finding containers with specific capabilities. We have used these resources to develop several containers for logical, FBA, ODE, stochastic, network-free, and DAE tools.
Future directions: We aim to work with the community to develop additional containers that support other formalisms, algorithms, and formats. We plan to expand the testing tool to verify each container's capabilities more rigorously. To make it even easier to reuse models, we are using the registry to develop a simple web application for executing simulations.
Availability: BioSimulators is openly available at https://biosimulators.org along with examples and documentation.
Target audience: Biomodeling community; Investigators who want to quickly try out and reuse published models or access older versions of simulation tools; Simulation software developers who want to easily distribute their simulation tools or compare the accuracy or performance of their tools with other tools; Peer reviewers and editors who want to evaluate models pre-publication.
Standards: CellML, COMBINE Archive, NeuroML, SBML, SED-ML, BioContainers, EDAM, KiSAO, SBO

11:30 – 11:45 D **Discussion** Room 1

11:45 – 12:00 B **Break** Room 1

12:00 – 12:15 T **(Replay) Datanator: an integrated database of molecular data for quantitatively modeling cellular behavior** Room 1

Speakers: Lian Zhouyang

Abstract: Integrative research about multiple biochemical subsystems has significant potential to help advance biology, bioengineering, and medicine. However, it is difficult to obtain the diverse data needed for integrative research. To facilitate biochemical research, we developed Datanator (https://datanator.info), an integrated database and set of tools for finding of multiple types of molecular data about specific molecules and reactions in specific organisms and environments, as well as data about chemically-similar molecules and reactions in phylogenetically-similar organisms in similar environments. Currently, Datanator includes metabolite concentrations, RNA modifications and half-lives, protein abundances and modifications, and reaction rate constants about a broad range of organisms. Going forward, we aim to launch a community initiative to curate additional data. Datanator also provides tools for filtering, visualizing, and exporting these data clouds. We believe that Datanator can facilitate a wide range of research from integrative mechanistic models, such as whole-cell models, to comparative data-driven analyses of multiple organisms.
Target audience: modelers, tool developers
Standards: BioPAX, CellML, SBML, SED-ML

12:15 – 12:30 | T | **(Replay) SBViper: Verification Testing of Kinetics Models in Systems Biology** | Room 1

Speakers: Joseph L Hellerstein

Abstract: Over the last decade, kinetics models in systems biology have grown in size from small tens of reactions to hundreds and thousands. This growth in complexity has increased interest in testing model correctness.

There are two broad approaches to testing modeling correctness. Validation testing seeks to assess the accuracy and utility of the models. This addresses questions such as: Does the model accurately predict values of new experimental results? Does the model provide helpful insights, such as chemical pathways to target? Validation tests are constructed in a manner that is specific to the model and how it will be used.

Verification testing checks for errors in the model implementation and/or the way in which the model is executed. To illustrate, consider the following incorrect implementation of a kinetics model of a two species linear pathway with mass action kinetics:

R1: S1 -> S2; k1*S1

R2: S2 -> S3; k2*S1

(The kinetics expression follows the semi-colon.) We see that there is a typographical error in the kinetics expression for R2 in that S1 appears instead of S2. This incorrectly implemented reaction network is still a linear pathway, but it does not have the desired kinetics. As model complexity grows, it will become increasingly common to introduce such errors and increasingly difficult to discover them.

This project is developing tools of verification testing of kinetics models. While there has been substantial work with verification of constraint-based models (e.g., mass balance analysis and detection of blocked reactions), we are unaware of work to date that addresses kinetics models. We are developing codes in three areas:

1. Testing model specifications. These tests do not require running simulation codes. An example is verification of kinetics expressions (e.g., detecting undefined chemical species).

2. Verifying simulation results. These tests analyze the results of simulation runs, such as whether the concentration of a chemical species converges to a particular value. The purpose of these checks is to verify that the model operates as the modeler intended.

3. An infrastructure for test execution and reuse. We find that some tests are applicable to many kinetics models. We have created an infrastructure that supports test reuse.

Target audience: SBML community, modelers, tool developers

Standards: SBML, SED-ML

12:30 – 12:45 | T | **(Replay) pyPESTO: A python package for Parameter Estimation and Uncertainty Quantification** | Room 1

Speakers: Jakob Vanhoefer

Abstract: Fitting a mathematical model to experimental data is an integral step in systems biology. A likelihood or posterior distribution quantifies the probability of the observed data under a given model parameterization. Parameter estimates are commonly derived by maximizing the likelihood or posterior distribution. This is typically done via numerical optimization, which is computationally demanding.

Uncertainty quantification accesses uncertainties in parameter estimates and model predictions due to noisy data. Established methods for uncertainty quantification include profile- or sampling-based methods.

Here, we present pyPESTO, a python package for optimization and uncertainty quantification. pyPESTO provides an interface to the model simulation tool AMICI and hence is tailored, but not restricted, to Ordinary Differential Equation models in the SBML and PEtab standard. For optimization, pyPESTO interfaces gradient and Hessian based optimizers from SciPy, dlib, and the interior point Optimizer IpOpt, as well as gradient free Particle Swarm optimizers (pyswarm).

For uncertainty quantification pyPESTO implements the profile likelihood method as well as various sampling algorithms like adaptive Metropolis, adaptive parallel tempering and a No-U-Turn-Sampler (via an interface to pymc3). pyPESTO offers functionality to parallelize multi-start optimization and sampling and store optimization results. Furthermore, pyPESTO provides visualizations of parameter estimation results, profile likelihood estimations and sampling results. Further extensions e.g. to model selection and experimental design are currently under development.

Target audience: modelers, SBML community, persons interested in parameter estimation

Standards: SBML

12:45 – 13:00 | D | **Discussion** | Room 1

PINNED

13:00 – 14:00 | S | **Social space/BREAK** | Room 1

PINNED

14:00 – 14:45

I **(Replay + Live Discussion) CellDesigner: A modeling tool for biochemical networks**

Room 1

Speakers: Akira Funahashi

Abstract: Understanding the logic and dynamics of gene-regulatory and biochemical networks is a significant challenge of systems biology. To facilitate this research topic, we have developed a modeling/simulating tool called CellDesigner.
CellDesigner primarily has capabilities to visualize, model, and simulate gene-regulatory and biochemical networks. Two significant characteristics embedded in CellDesigner boost its usability to create/import/export models:
(1) Solidly defined and comprehensive graphical representation (Process Diagram) of network models and
(2) Systems Biology Markup Language (SBML) as a model-describing basis, which functions as inter-tool media to import/export SBML-based models.
CellDesigner also supports simulation and parameter scan, supported by integration with SBML ODE Solver, COPASI, and Simulation Core Library enabling users to simulate through our sophisticated graphical user interface. Users can also browse and modify existing models by referring to existing databases (ex. BioModels.net) directly through CellDesigner.
These enhancements make CellDesigner not only a modeling/simulation tool, but also an integrated analysis suite. CellDesigner is implemented in Java and thus supports various platforms (i.e., Windows, Linux, and macOS). CellDesigner version 4.4.2 is freely available via our web site http://celldesigner.org, and a new version of CellDesigner is under development, which will support spatial modeling.

Target audience: modelers

Standards: SBML, SBGN

14:45 – 15:00

T **(Replay) A modular, thermodynamic approach for constructing large-scale kinetic models in systems biology**

Room 1

Speakers: Michael Pan

Abstract: Comprehensive large-scale mathematical models of biomolecular systems have the potential to direct future advances in health and biotechnology, but are currently difficult to develop. It is generally acknowledged that kinetic models should be constructed by reusing and coupling together existing models of smaller systems, i.e. in a modular fashion. However, models of biological systems are often expressed in different formalisms and therefore do not naturally interface with each other. In this talk, I argue that the conservation laws of physics provide a unified interface for models to communicate. The bond graph - a graphical, energy-based modelling framework that is well-established in the field of engineering - is introduced as a means of enabling this approach. The approach will be illustrated in the development of a model of the MAPK signalling cascade.

Target audience: kinetic modellers, tool developers

Standards: CellML, SBGN, SBML, SED-ML

15:00 – 15:15

T **(Replay) Open Source and Sustainability**

Room 1

Speakers: Jacob Barhak

Abstract: In short: The presenter will discuss encountered issues and limitations with current open source practices that impact on sustainability. Those issues include: over interpretation, abandoned code, and potential conflict with external restrictions such as patents and government restrictions. Relation to government funding policies will be discussed. A possible solution to improve incentives and sustainability will be presented. Details: The presenter develops computational disease modeling tools that are both open source and proprietary. The presenter benefited from open source and currently owns the most validated diabetes model known and a COVID-19 model. Yet current practices make it harder to sustain research using current open source practices. A recent call by the modeling community with regards to COVID-19 modeling, calls to publish model code openly. However, making such a call is insufficient unless issues related to sustainability are addressed. Modeling practices should better fit the reality and deal with issues such as over interpretation of licenses, abandoned research, and other restrictions such as patents or government restrictions. Open source and proprietary development should work in harmony and provide proper incentives for development.
A simple solution will be suggested to allow both sustainable development of open source code and incentives for researchers to advance their research.

Target audience: All those interested with licenses and open source and potential legal issues

Standards: All those interested with licenses and open source and potential legal issues

15:15 – 15:30	T	**(Replay) Physiome - Make your model publications discoverable, reproducible, and reusable.** *Speakers: Karin Lundengård* **Abstract:** Physiome is proud to show our first publications. We are a journal committed to reproducibility and reusability of mathematical models of physiological processes. When you publish in Physiome, the persistent identifier for your article is connected to a curated and permanent version of the model code. This makes the code necessary to run the model easily accessible by just downloading an omex archive file, to be reused as it is or as a module in a bigger model. Your publication is also connected to a primary paper published in a field specific journal, where the validation and scientific value of the model is discussed. A Physiome publication is a complement to your primary article that ensures reproducibility, reusability and discoverability of you model. The format encourages modularity that facilitates combination of different models to develop the next level of systems understanding. And all the models are in one place, easy to find and accessible. Physiome is open access with a low Author Processing Charge. Physiome curators will help authors ensure that models and simulation experiments are made available using appropriate community standards prior to acceptance for publication in Physiome. When aspects of a computational modelling study are not able to be encoded in standard formats, Physiome editors will help authors ensure their work is as open, reproducible, and reusable as possible. We will help as much as is currently possible with model curation and annotation to ensure that the modeling results claimed in the primary paper are consistent with the published model. Reproducibility and confirmation of results is crucial for useful science and should be incentivised. Yet, publication of it is often treated as a secondary result at best, which undermines the quality of our work. Waste no more valuable time and effort on trying to implement models from papers that lack information, or having your results lost because others cannot use them. Publish your models in Physiome and contribute to making science useful in society (and less frustrating for your colleagues). For more information visit: https://journal.physiomeproject.org **Target audience:** Anyone who wants to publish reproducible physiological models **Standards:** CellML, COMBINE Archive, NeuroML, SBML, SED-ML	Room 1
15:30 – 15:45	D	**Discussion**	Room 1
15:45 – 16:00	B	**Break**	Room 1
16:00 – 16:15	T	**(Replay) OpenCOR: how to enable reproducible science using community standards and tools** *Speakers: Alan Garny* **Abstract:** OpenCOR (https://opencor.ws) is a cross-platform environment for organising, editing, simulating and analysing models that primarily consist of ordinary differential equations encoded in the CellML format. It relies on community standards and tools such as CellML (https://cellml.org), SED-ML (https://sed-ml.org), the COMBINE archive format (http://co.mbine.org/standards/omex) and the Physiome Model Repository (PMR; https://models.physiomeproject.org). It provides a means to create models, reuse existing models (e.g., from PMR), and collaborate on their development. Simulation experiments can be created and shared, ensuring reproducibility, which can also be achieved through Python. **Target audience:** Modelers and tool developers **Standards:** CellML, COMBINE Archive, SED-ML	Room 1
16:15 – 16:30	T	**(Replay) CellML 2.0** *Speakers: David Nickerson* **Abstract:** CellML 2.0 was released on April 17, 2020, and contains significant changes since the previous 1.1 version. Here, we will present a summary of the CellML format and the changes introduced in CellML 2.0. The CellML format is based on the extensible markup language (XML), and is intended to enable the representation and dissemination of reproducible and reusable mathematical models of biology. The changes were motivated by community feedback on the 1.0 and 1.1 specifications, as well as a desire to separate the normative and informative aspects of the CellML specification. Included are the introduction of a reset element which allows discontinuities to be modelled without ambiguity; the restriction of allowed MathML elements; the removal of comments and annotations; and the removal of the reaction element and its children. The reset element is designed to allow discontinuities and discrete changes in variables based on conditional statements. This was possible in earlier versions using piecewise expressions, but the use of an order attribute within the resets means that the behaviour of conflicting switches can now be resolved uniquely. Possible applications include modelling a cell dividing as its size exceeds a threshold; the discrete switching of electrical stimuli; and re-initialising conditions during a simulation. The normative specification is available on the CellML website at https://www.cellml.org/specifications/cellml_2.0, with an informative specification with a discussion of examples, explanations, and common mistakes also provided. **Target audience:** - **Standards:** CellML	Room 1

16:30 – 16:45 | T | **(Replay) Implementing OMEX metadata v1.1** | Room 1

Speakers: John Gennari

Abstract: Building on previous, community-driven efforts to standardize the representation of metadata (e.g., annotations) for computational models, we have developed and published the initial OMEX (Open Modeling and Exchange) metadata specification version 1.0. Our goal is to enable cross-library and cross-language model searching, composition, and reuse. Not only are there multiple modeling languages, but there are many modeling tools and simulation environments, each of which currently manipulates and provides for annotations in idiosyncratic and language-dependent ways. To solve this potential tower of Babel, the OMEX specification describes how tools should support reading and writing annotations, storing these in a separate file, so that annotations can be decoupled from specific languages and formats.

Recently, we have proposed updates to the OMEX metadata specification (v1.1) to further standardize model-level annotations and to improve the URI scheme. In particular, the new specification states that all OMEX archives will use URIs derived from “omex-library.org”. This address simply serves as a stub for all OMEX archives, thereby supporting improved portability of these archives, as they no longer use relative paths.

Here, we report on two implementations of this new v1.1 OMEX metadata specification. These are software libraries, designed to be used by other tool developers, rather than by modelers or computational biologists. The first is the C++ libOmexMeta library, which we have also made available as a Python package, “PyOmexMeta”. The second implementation, designed for Java developers, is an enhanced SemSim Java API, which undergirds the SemGen tool for model annotation and composition.

Both libraries read and write COMBINE Archives that conform to the OMEX metadata specification. They can create and manipulate annotations. In addition, both packages also support the ability to extract annotations and other metadata from existing SBML models, especially those from the BioModels repository. This capability will allow us to easily collect metadata from multiple SBML models into a single, queryable network of information. Our use of the standard “Resource Description Framework” (RDF) to serialize annotations allows for SPARQL queries across all OMEX annotations. A standardized approach to model annotation will make models more understandable, findable, interoperable and reusable, thus supporting the FAIR principles of scientific data management.

Target audience: -

Standards: CellML, COMBINE Archive, SBML

16:45 – 17:00 | D | **Discussion** | Room 1

PINNED
17:00 – 18:00 | S | **Social space/BREAK** | Room 1

PINNED
18:00 – 18:45 | I | **Reproducibility initiatives in computational biology** | Room 1

Speakers: Jason Papin

Abstract: There has been much discussion in the scientific literature on a crisis of reproducibility in science. It has been reported that the percentage of studies that are reproducible is as low as 10% or less, depending on the discipline (Baker. Nature News. 2016). This inability to reproduce scientific findings from a given paper has been attributed to a lack of clarity in the methods and inherent variability in the biological system being studied (Ioannidis. PLOS Medicine. 2005).

Reproducibility in computational biology research is certainly a problem, yet perhaps a challenge that our field can uniquely tackle. A lack of reproducibility in computational biology research can be attributed to many factors but incomplete or erroneous descriptions of the simulations (e.g., which software version was used), incomplete documentation on how to run simulations, or simply failing to post the relevant computer code needed to run a given simulation are common issues that occur.

Many tools have emerged that we can leverage to make computational biology research more reproducible (e.g., http://co.mbine.org/ and https://normsys.h-its.org/) and there exist articles that propose best practices such as Ten Simple Rules for Reproducible Computational Research (Sandve et al. PLOS Computational Biology. 2013) or Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks (Rule et al. PLOS Computational Biology. 2019).

Many challenges for reproducibility in computational biology research remain. This presentation will discuss recent initiatives at PLOS Computational Biology to address some of these challenges.

Target audience: -

Standards: BioPAX, CellML, COMBINE Archive, NeuroML, SBGN, SBML, SBOL and SBOL Visual, SED-ML

18:45 – 19:00	T	**LibSBML 2020 and beyond** *Speakers: Sarah Keating* **Abstract:** Libraries such as libSBML, libSEDML and libSBGN2 (among others) provide the backbone for software written and used by many researchers in the Computational Biology field. The Research Software Development Group at University College London have been funded to begin the development of a maintenance framework to facilitate continued community contributions to the development of libSBML. In this talk we will describe our transition to GitHub and our plans for Continuous Integration testing and automated building of the many language bindings available with libSBML. The grant also provides for extension of Deviser to allow for easier development/maintenance of other libraries for XML-based standards. We will explain our plans to extend and simplify other libraries that have been developed using Deviser. **Target audience:** Tool developers **Standards:** COMBINE Archive, SBGN, SBML, SED-ML	Room 1
19:00 – 19:15	T	**pySBOL3: A Python library for SBOL 3** *Speakers: Tom Mitchell* **Abstract:** Python is a commonly used programming language for biological work. pySBOL3 is a Python library that implements the SBOL 3 specification, enabling practitioners to read, manipulate, and write SBOL3 compliant files. pySBOL3 can be used to translate from other data standards into SBOL 3, or convert SBOL 3 information to other formats. In this talk, we will describe the architecture and use of pySBOL3, leaning heavily on code examples. **Target audience:** SBOL Community **Standards:** SBOL and SBOL Visual	Room 1
19:15 – 19:30	T	**Latest developments with the Python modeling package: Tellurium** *Speakers: Herbert Sauro* **Abstract:** Python has become one of the primary software tools used across many scientific disciplines, including systems biology to carry out computations, data manipulation, analysis, and visualization. Using Jupyter notebooks, Python is also becoming an important tool for distributing computational research in a repeatable manner. This allows simulations, for example, of signaling networks, to be very easily made available to other researchers where they can readily repeat the work with minimal effort. In our research we are using Python together with Tellurium to build and distribute models of the EGF pathway and P. putida metabolism. Tellurium is a python package specifically design for systems biology simulations and includes support for the high-performance simulator libRoadRunner as well as many of the community exchange standards such as SBML, SED-ML and Combine archives. This allows models generated by our tooling to be readily used by other tools that support these standards. Models that used standards such as SBML can be easily uploaded to model repositories such as Biomodels. We will report on latest developments, including but not limited to a direct API and a new plugin system. **Target audience:** Beginners in modeling **Standards:** COMBINE Archive, SBML, SED-ML	Room 1
19:30 – 19:45	D	**Discussion**	Room 1
19:45 – 20:00	B	**Break**	Room 1
20:00 – 20:15	T	**Integration of sequence visualization into SynBioHub** *Speakers: Linhao Meng* **Abstract:** As a design repository for Synthetic Biology, SynBioHub enables genetic components and designs to be uploaded and shared. To extend its capability for genetic wet-lab workers who need to interrogate the information for specific DNA ‘parts’ prior to using them or know about specific function by the annotations of the DNA sequence, we created a sequence view visualization plugin engine for SynBioHub. It is composed of three different viewers, respectively, an SBOL Visual viewer using VisBOL, a plasmid viewer, and a sequence viewer. This tool can enhance the functional annotations of DNA sequences, which are coordinated with the SBOL Visual standard. **Target audience:** tool developers **Standards:** SBOL and SBOL Visual	Room 1
20:15 – 20:30	T	**BpForms and BcForms: a toolkit for concretely describing non-canonical polymers and complexes to facilitate global biochemical networks** *Speakers: Paul F Lang*	Room 1

20:30 – 20:45 | T | **Quantitative study of spike propagation in a one-dimensional strand of detrusor smooth muscle cell** | Room 1

Speakers: Chitaranjan Mahapatra

Abstract: Urinary incontinence is associated with enhanced spontaneous phasic contractions of the detrusor smooth muscle (DSM). It is suggested that the spontaneously evoked action potentials (sAPs) in DSM cells initiate and modulate the contractions. One of the key properties of APs is their non-attenuating propagation along lengths of cable-like structures such as axons and muscle cells. DSM, like some other smooth muscles, is known to exhibit one-dimensional cable-like behaviour when uniformly polarized at a plane. In syncytial tissues such as smooth muscle, the presence of gap junctions between cells underpins intercellular electrical communication.

To ascertain whether our computational AP would exhibit this property, we constructed a one-dimensional cable model of DSM by linking five cells end-to-end, the electrical connectivity being provided via gap junctions. The computational AP is simulated from a previously designed biophysically constrained single DSM cell, which consists nine ion channels [1]. The peak amplitude of the propagated APs is higher than the evoked control AP due to more charge dissipation to neighbouring segments in both directions for the evoked one. The convex-upward foot of the AP at stimulus injection point is converted into a concave-upward foot in the propagated APs, as expected from theory owing to the effect of cable properties. Given these findings, it can be hypothesized that some proportion of the varied AP shapes in DSM cells mentioned in the foregoing sections could be explained on the basis of whether, in experimental recordings, APs were recorded at or close to the locus of neurotransmitter action or at a distance from the locus.

Our exploration of spike propagation by its incorporation into a 1-D model is a preliminary one. Because the smooth muscle of detrusor, akin to many other smooth muscles, forms a 3-D syncytium of cells, a biophysically realistic 3-D model is essential for a more physiologically realistic investigation.

1. Mahapatra, Chitaranjan, Keith L. Brain, and Rohit Manchanda. "A biophysically constrained computational model of the action potential of mouse urinary bladder smooth muscle." PloS one 13, no. 7 (2018): e0200712.

Target audience: -

Standards: CellML, NeuroML

20:45 – 21:00 | D | **Discussion** | Room 1

PINNED
21:00 – 21:45 | I | **(Replay + Live Discussion) COVID-19 Disease Map: the key role of standards in community-driven development of systems biology disease models** | Room 1

Speakers: Marek Ostaszewski

TBA

21:45 – 22:00 | T | **(Replay) PEtab – Interoperable Specification of Parameter Estimation Problems in Systems Biology** | Room 1

Speakers: Daniel Weindl

Abstract: Reproducibility and reusability of the results of data-based modeling studies are essential. Yet, there has been – so far – no broadly supported format for the specification of parameter estimation problems in systems biology. Therefore, we developed PEtab, a format which facilitates the specification of parameter estimation problems using Systems Biology Markup Language (SBML) models and a set of tab-separated value files describing the observation model and experimental data as well as parameters to be estimated.

We already implemented PEtab support into eight well-established model simulation and parameter estimation toolboxes with hundreds of users in total. We provide a Python library for validation and modification of a PEtab problem and as well as example parameter estimation problems based on recent studies.

Specifications of PEtab, the PEtab Python library, as well as links to examples, and all supporting software tools are publicly available at https://github.com/PEtab-dev/PEtab.

Target audience: tool developers and modelers interested in parameter estimation

Standards: SED-ML, Parameter estimation

22:00 – 22:15	T	**(Replay) Spatial Model Editor** *Speakers: Liam Keegan* **Abstract:** We present Spatial Model Editor, a graphical user interface tool for editing and simulating two dimensional spatial SBML models. The model geometry can be imported from segmented pixel images, or from existing spatial SBML models. The tool automatically identifies the contours of compartment boundaries in the image, and constructs a triangular mesh approximation to the geometry. A system of PDEs is constructed from the model, and solved on this triangular mesh using Finite Element methods with the dune-copasi solver. This solver, which is being developed as part of this project, is an extension of the DUNE PDE framework to solve multi-compartment reaction-diffusion equations - see https://gitlab.dune-project.org/copasi/dune-copasi for more information. A simple pixel based simulator is also included which uses the finite difference method (https://spatial-model-editor.readthedocs.io/en/latest/reference/pixel.html). In addition to the graphical user interface, there is also a python interface as well as a command line interface. The tool is open source and is available for download for windows, linux and mac from https://github.com/lkeegan/spatial-model-editor **Target audience:** - **Standards:** SBML	Room 1
22:15 – 22:30	T	**(Replay) pyABC: likelihood-free inference** *Speakers: Emad Alamoodi* **Abstract**: Understanding the source of variation among and within populations is still a major question in system biology. Due to the continuous improvement in computational power, simulation-based methods have been intensively used to answer this question. Among these methods is Approximate Bayesian Computation (ABC). ABC is a likelihood-free inference approach that facilitates the approximation of the Bayesian posterior distribution, which problems, for which the evaluation of the likelihood is expensive or even infeasible. While ABC provides a theoretical solution, its usage is often hindered by its high computational demands. Here, we present pyABC: a distributed and scalable ABC-Sequential Monte Carlo (ABC-SMC) framework. It implements a strategy for multi-core and distributed environments scaling to thousands of cores. The framework is easy to use and also enables advanced users to customize and experiment with many options of ABC-SMC schemes, such as acceptance threshold schedules, transition kernels, distance functions, and the use of complex noise models with stochastic acceptors without alteration of pyABC's source code. The code has been substantially improved over the last years and the methods are now used by various research groups from different research fields. **Target audience:** modelers, parameter estimate, Bayesian inference, model selection **Standards**: SBML	Room 1
22:30 – 22:45	D	**Discussion**	Room 1
22:45 – 23:00	B	**Break**	Room 1
23:00 – 23:15	T	**(Replay) FAIRDOM: standard compliant data and model management** *Speakers: Olga Krebs* **Abstract:** Systems Biologists need a data management infrastructure that enables collaborating researchers to share and exchange information and data as and when it is produced, throughout the entire iterative cycle of experimentation and modelling. Data exchange and reuse rely on sufficient annotation, consistent metadata descriptions, and the use of standard exchange formats for models, data, and the experiments they are derived from. FAIRDOM offers integrated data management support for systems biology research projects within and across national and international consortia comprising a whole package of solutions. This is applied to large-scale research initiatives in which FAIRDOM members are responsible for the scientific data management, e.g. the German projects LiSyM (Liver Systems Medicine, https://lisym.org/), and MESI-STRAT (Systems Medicine of Metabolic-Signaling Networks, https://mesi-strat.eu/) , as well as the European research networks ERASysAPP (ERA-Net for Systems Biology Applications), and the Synthetic Biology Centres at Manchester (SynBioChem) and Edinburgh (SynthSys). FAIRDOM is an integral member of several infrastructure initiatives (e.g. ELIXIR, https://elixir-europe.org/ and de.NBI - German Network for Bioinformatics Infrastructure, https://www.denbi.de/) The FAIRDOM data management concept consists of 4 major pillars: 1) Infrastructure backbone: The FAIRDOMHub/SEEK platform as registry and a commons for data, models, processes and resulting publications and presentations, at the same time yellow pages for projects, people and events 2) Terminology: Tailored use of controlled vocabularies and ontologies to describe the data 3) Modelling support: Seamless handling and simulation of models by integrated modelling platforms (JWS-Online,Cytoscape) 4) Social support: Data management advocates within the projects for gathering requirements and dissemination. **Target audience:** COMBINE community, modelers, experimentalists, developers **Standards:** CellML, COMBINE Archive, SBGN, SBML, SBOL and SBOL Visual, SED-ML, MIBBI, FAIRsharing	Room 1

23:15 – 23:30 | T | **(Replay) FAIR principles in literature-based kinetic modelling** | Room 1

Speakers: Christoff Odendaal

Abstract: In view of the laborious nature of model building and validation, a key part of the effort of modular modelling and expansion, is avoiding duplicative work (Klipp et al., 2007). Progress has been made to standardise model reporting (e.g. SBML) and annotation (e.g. MIRIAM), and these standards are widely endorsed by systems biologists. Kinetic model-building from literature relies on making several modelling decisions based on a priori biochemical knowledge. The degree to which experimental conditions affect measurements are, however, often research questions in themselves, making it unrealistic for a modeller to exhaustively evaluate and recalculate the parameters in her model. At such points, parameter decisions must be made under substantial uncertainty.

Modelling decisions need not be final, however, and can be revised based on new information. We propose that a modelling method be developed for aligning the model-building process more with the FAIR (findable, accessible, interoperable, reusable) principles of Open Science. This would include, amongst others, an accompanying sheet of metadata for each model clearly stating not only model parameters and citations, but also all major assay conditions (pH, temperature, buffer composition, etc.) as well as explicit reasons why a certain values were preferred above others – the necessity of this is widely agreed upon, but it is still often not done. We also suggest that the author include a list of all consulted literature (also rejected literature) as well as a standard operating procedure that was used for the literature search. Such a metadata sheet would also include relations that were omitted due to unconvincing evidence as well as questions which arose in the process of model creation. This will give the modelling process more the character of a systematic review.

These measures would chart the searches that a modeller has performed to make that same course of enquiry transparent to other researchers. Furthermore, the logic of explicitly flagging unresolved questions which arose during the literature-search, would encourage complementary work by pointing directly to ways in which the model can be improved. This olgawould allow experimentalists who are not as interested in deeply scrutinising entire models to find synergies with their work.

References: Klipp, E., Liebermeister, W., Helbig, A., Kowald, A., & Schaber, J. (2007). Nature biotechnology, 25(4), 390-391.

Target audience: Modellers

Standards: SBML, SED-ML, MIRIAM

23:30 – 23:45 | T | **(Replay) Stochastic Modelling and Prediction of the COVID-19** | Room 1

Speakers: Joab Odhiambo

Abstract: Since the inception of the novel Covid-19 in December in China, the spread has been massive leading World Health Organization to declare it a world pandemic. While epicenter of COVID-19 was Wuhan city in China mainland, Italy has been affected most due to the high number of recorded deaths as at 1st April, 2020 at the same time USA recording the highest number of virus reported cases. In addition, the spread has been experienced in many developing African countries including Kenya. While the Kenyan government have had plans for those who have tested positive through self-quarantine beds at Mbagathi Hospital, lack of a proper mathematical model that can be used to model and predict the spread of Covid-19 for adequate response security has been one of the main concerns for the government. Many mathematical models have been proposed for proper modeling and forecasting, but this paper will focus on using a generalized linear regression that can detect linear relationship between the risk factors. The paper intents to model and forecast the confirmed Covid-19 cases in Kenya as a Compound Poisson process where the parameter follows a generalized linear regression that is influenced by the number of daily contact persons and daily flights with the already confirmed cases of the virus. Ultimately, this paper should assist the government in proper resource allocation to deal with pandemic in terms of available of bed capacities, public awareness campaigns and virus testing kits not only in the virus hotbed within Nairobi but also in the other 47 Kenyan counties.

23:45 – 00:00 | D | **Discussion** | Room 1

COMBINE 2020

[R] Breakout [S] Social space [U] Tutorial

OCTOBER 7 • WEDNESDAY

PINNED 00:00 – 01:00	S	**Social space/BREAK**	Room 1
PINNED 04:00 – 06:00	S	**Social space/BREAK**	Room 1
06:00 – 09:00	R	**SED-ML L1V4** *Speakers: Matthias König* **Abstract:** Discussing tool support and remaining issues for the SED-ML L1V4 release. **Schedule:** Short introduction SED-ML Short introduction SED-ML L1V4 (plotting & parameter fitting) Presentations: Tool support L1V4 Hacking specification **Target audience: SED-ML community, modelers, tool developers** **Standards:** SED-ML	Room 1
PINNED 09:00 – 10:00	S	**Social space/BREAK**	Room 1
10:00 – 12:00	R	**Standards for curating flux balance constraint (FBC) models** *Speakers: Malik-Sheriff, Rahuman S.* **Abstract:** Reproducibility of systems biology models is currently an important topic of discussion. A curated model should be able to faithfully reproduce the analysis results. In the case of constraint-based modelling, flux values - the commonly reported results in manuscripts performing flux balance analysis - cannot be easily reproduced as these values are not unique solutions. To address this, an FBC curation standard work group was initiated in Harmony 2020. Following our discussion, we propose the potential use of FROG analysis with numerically reproducible results as a community standard to assess the reproducibility of FBC models. BioModels will employs these guidelines to test the reproducibility and curate deposited FBC models. **Target audience**: Constraint-based modelling groups, FBC **Standards**: COMBINE Archive, SBML, SED-ML	Room 1
10:00 – 13:00	R	**SBOL Libraries Roundtriping and Testing** **Abstract:** During this breakout, we will be experimenting with the prototype SBOL3 libraries written in Python, Java, and Javascript. In particular, we will ensure that serialization produced by each library can be successfully read by the other libraries, and vice versa. We will also work on common testing infrastructure for all the libraries. **Target audience**: SBOL **Standards**: SBOL	Room SBOL
12:00 – 13:00	R	**What is the best way to add thermodynamic information to an SBML model?** *Speakers: Moritz Beber* **Abstract**: The session is intended to briefly discuss the pros and cons of different methods to add thermodynamic information to SBML models. By thermodynamic information we mean formation energies, standard Gibbs free energy, conditions within compartments, and covariance. Some possible suggestions include: using the upcoming key-value annotation within SBML, making use of the array extension, and/or creating a COMBINE archive. **Target audience:** SBML community and modelers **Standards:** COMBINE Archive, SBML, SED-ML	Room 1
PINNED 13:00 – 14:00	S	**Social space/BREAK**	Room 1

14:00 – 17:00 R **BioSimulations** Room 2

Abstract: More predictive models could advance biology, medicine, and bioengineering. Building such models will likely require teams who can share and reuse models. Several standards such as KiSAO, OMEX, SBML, SBO, and SED-ML help researchers share models and simulations. To support multiple formalisms such as logical modeling and FBA, these standards include numerous features such as delays and objectives. Hundreds of simulators support many of these features. However, it remains difficult to reuse many models. In particular, the incomplete support for these features among simulators and the lack of centralized documentation of the features supported by each simulator often make it difficult to find a simulator for a model, especially for non-experts. Furthermore, simulators frequently become unavailable when projects terminate, old versions of tools that support old versions of features are often hard to obtain, many tools are cumbersome to install, and it takes significant effort to learn a different interface for each tool. To make it easier to reuse models, we are developing BioSimulations, a web application for sharing and reusing models, simulations, and visualizations of their results. To enable BioSimulations, we are developing BioSimulators, a collection of containerization simulators that provide standardized interfaces and metadata. In turn, to develop BioSimulators, we have prototyped a standard command-line interface and container structure for simulation tools and a standard format for specifying their capabilities (supported formalisms, algorithms, frameworks). These resources build upon several standards including COMBINE, EDAM, KiSAO, SBO, and SED-ML. To date, we have developed containers for several popular simulators.

Session overview:

This session will provide a forum for discussion about these prototypes for standards. The goals of the session will be to identify the community's needs, solicit feedback and input on the proposed formats, and prioritize the next steps of the development of BioSimulations and BioSimulators.

Schedule:

Intro (10 min)

Overview of BioSimulations & BioSimulators (5 min): Organizers

Overview of proposed standards (15 min): Organizers

Interface to simulators

Structure for containers

Format for specifying the capabilities of simulators

Discussion about interface for simulators (1 hr)

Discussion about format for the capabilities of simulators (1 hr)

Discussion about next steps (30 min)

Target audience: Investigators who would like to shape the direction of BioSimulations or contribute to its development. Simulation software developers who want to make their tools easier to discover and easier to install. Modelers who want to help make it easier to find appropriate simulation tools for a given model.

Standards: CellML, COMBINE Archive, NeuroML, SBML, SED-ML, BioContainers, EDAM, KiSAO, SBO

14:00 – 17:00 R **SBOL 3.0.1** Room SBOL

Abstract: During this breakout, we will be going over several recent updates to SBOL 3 that fix some issues that we discovered as we began implementation of the SBOL 3 libraries.

Target audience: SBOL

Standards: SBOL

14:00 – 17:00 | R | **Towards in silico approaches for personalized medicine** | Room 1

Speakers: Catherine Collin

Abstract: Harmonization of data integration is the key to standardization efforts in personalized medicine. While standardization of models themselves is undesirable within a research context, where new models are created and tested in line with research progress – harmonization and/or standardization of input data is both feasible and necessary.

However, we argue that model validation should receive more attention, and other measures should be implemented, such that validation of models within personalised medicine becomes easier. While this is an evident necessity within the context of models implemented as medical devices, which are regulated by European Medicines Agencies and national competent authorities, we argue that model validation should be a higher priority at research level also, facilitating assessment by peers and by medical doctors – who themselves should receive better training in assessment of research using in silico models.

Acceptance by doctors and the relevant medical specialties is a key hurdle for in silico models in personalised medicine. Any medical product - device, algorithm or drug - has to prove itself safe and effective to be permitted for use by regulators; however, it has also to be accepted by medical experts as being a good choice, and be recommended within clinical specialties.

The Horizon2020 funded Coordinating and Support Action “EU-STANDS4PM” joined forces to examine to what extent existing standards or standards under development for both, format and semantics, can be used to link clinical healthcare data to computational models that build on these data. As all requirements should be equally understood and fulfilled by users, it is important to define them uniformly in an international context. To achieve this the conclusion of our work shall be also discussed in international standardization and technical committees, especially in the case of standards that are still being drawn up, and new standardization projects shall be initiated where necessary.

We present an overview of recommendations for standardization of data integration as well as recommendations for standardization of model validation within a collaborative research context, such that health-related data can be optimally used for translational research and personalized medicine across Europe.

Target audience: modellers, data managers, tool developers, standard developers, clinical data creators

Standards: all

PINNED
17:00 – 18:00 | S | **Social space/BREAK** | Room 1

18:00 – 21:00 | R | **Systems Biology Graphical Notations** | Room 2

Speakers: Michael Blinov

Abstract:

* Requests from the community
* Handling annotations
* Suggestions for specs readability
* The fate of SBGN-ER and SBGN-AF, introducing hybrid maps
* Revisit the comments (some I think from Adrien) from when we wrote the SBGNML paper
* Improving of community collaboration, e.g. enabling template for suggestions
* Community publications, e.g. tools review
* Incoming SBGN elections.

Target audience: SBGN community, tool developers and modelers

Standards: BioPAX, COMBINE Archive, SBGN, SBML

18:00 – 21:00 | U | **Building, exploring and sharing rule-based models of cellular signaling pathways (I/II)** | Room 1

Speakers: Martin Meier-Schellersheim

Abstract: We will provide a practical introduction into creating models of cellular signaling pathways using Simmune’s visual modeler and Simmune’s Python interface. Whereas the visual modeling approach is biologically intuitive, using iconographic symbols for molecules and their interactions, the Python interface allows modelers to set up simulations with repetitive elements, such as stepwise varying stimulation of cells with chemokines or cytokines. The tutorial will also demonstrate how to explore the behavior of the models through systematic parameter scans. Finally, we show how models created with Simmune can be exported into SBML, including the new SBML3 standard for multi-state, multi-component and multi-compartment biochemical species.

Target audience: Cell biologists, beginners in modeling, SBML community

Standards: SBML

18:00 – 21:00 U **SBOL Version 3: Simplified Data Exchange for Bioengineering** Room SBOL

Speakers: Jacob Beal

Abstract: The Synthetic Biology Open Language (SBOL) allows knowledge about biological designs to be captured using a machine-tractable, ontology-backed representation that is built using Semantic Web technologies. While early versions of SBOL focused only on the description of DNA-based components and their sub-components, SBOL can now be used to represent knowledge across multiple scales and throughout the entire synthetic biology workflow, from the specification of a single molecule or DNA fragment through to multicellular systems containing multiple interacting genetic circuits.

The third major iteration of the SBOL standard, SBOL3, is an effort to streamline and simplify the underlying data model with a focus on real-world applications, based on experience from the deployment of SBOL in a variety of scientific and industrial settings. This tutorial will provide the following to attendees:

- general introduction to SBOL
- explanation of the new SBOL3 data model in detail
- comparison SBOL3 and SBOL2,
- hands-on experience with libraries implementing SBOL3.

Target audience: Modelers, tool developers, and anyone with interest in SBOL

Standards: SBOL and SBOL Visual

21:00 – 23:00 R **SBML Layout and Render Extensions** Room 1

Speakers: Herbert Sauro

Abstract: A discussion period to dicuss the future and current developments in standards for drawing biochemical neworks including adhoc efforts such as Escher.

Target audience: Modelers, SBML community, Tool developers

Standards: SBGN, SBML, Render and Layout Extension

COMBINE 2020

[R] Breakout [S] Social space [U] Tutorial

OCTOBER 8 • THURSDAY

PINNED 00:00 – 01:00	S	**Social space/BREAK**	Room 1
01:00 – 03:00	U	**Using Python HoloViz Technologies to Create Interactive Presentations** *Speakers: James Bednar, Jacob Barhak* **Abstract:** The HoloViz project provides a set of Python libraries for high-level visualization of complex datasets. They are particularly useful for handling big data and multi-dimensional data that is common in machine-learning applications. HoloViz technologies support multiple graphical engine backends and integrate seamlessly with flexible development and deployment environments like Jupyter notebooks and modern web browsers. The visualization outputs are interactive, with features such as widgets like sliders or selection boxes or hover tools to inspect data, while not requiring any JavaScript, HTML, CSS, or other web-technology expertise. This tutorial will focus on two HoloViz libraries: HoloViews: high level interface providing plots (heat maps, histograms, spikes, etc.) in many spatial and temporal combinations, with or without widgets for selecting along dimensions Panel: simple application and dashboard creation from images, plots, Markdown, LaTeX, and other elements into one HTML page incorporating interactive tabs and widgets. During the tutorial an interactive presentation will be constructed to show the attendees how to construct their own interactive poster / presentation. Sample References: • HoloViz web site: https://holoviz.org • HoloViz on Github: https://github.com/holoviz/holoviz • Jacob Barhak, Joshua Schertz, Visualizing Machine Learning of Units of Measure using PyViz, PyData Austin 2019, 6-7 December 2019, Galvanize Austin. Presentation: https://jacob-barhak.github.io/Presentation_PyData_Austin_2019.html. Video: https://youtu.be/KS-sRpUvnD0. **Target audience:** Modelers and python users **Standards:** -	Room 1
PINNED 04:00 – 06:00	S	**Social space/BREAK**	Room 1
06:00 – 08:00	U	**COPASI - an update on recently added functionality** *Speakers: Pedro Mendes* **Abstract:** COPASI is now a well-established and popular simulator. In recent years a number of new features have been added that are not well known, for example stochastic differential equations. This tutorial will highlight these features, and also discuss them in the context of SBML and SED-ML. **Target audience:** Modelers, SBML community **Standards:** SBML, SED-ML	Room 1

08:00 – 09:00	U	**Newt: view, design and analyze pathways in SBGN and more** *Speakers: Ugur Dogrusoz* **Abstract:** This tutorial will introduce Newt, a free, web based, open source viewer and editor for pathways in Systems Biological Graphical Notation (SBGN) and Simple Interaction Format (SIF). It was written with a series of libraries and extensions based on Cytoscape.js with utmost customization in mind. What distinguishes Newt from other viewers and editors for biological maps can be summarized as: - Rich and refined, yet easy-to-use web based UI - Convenient construction and annotation of pathways from scratch as well as viewing and editing existing maps - Full support for compound structures (including automatic layout) to properly represent compartments, molecular complexes, and sub-maps - Semantic validation and guided fix for SBGN PD maps - State-of-the-art complexity management capabilities through hide-show or highlight parts of a map and collapse-expand compound structures - Advanced diagramming through interactive move, resize, and styling of map objects including color schemes, re-routing and re-connection of interactions, and grid & alignment guideline support - Support for experiment data overlay on maps - Facilities for querying, viewing, and editing pathways in Pathway Commons - Conversion to and from CellDesigner and SBML file formats - Launching with a remote SBGN model specified as a URL (example) or as a URI (example) - Support for SBGN bricks **Target audience:** Beginners in modeling **Standards:** SBGN	Room 1
PINNED 09:00 – 10:00	S	**Social space/BREAK**	Room 1
10:00 – 13:00	R	**SBOL Validation** *Speakers: Christopher Myers* **Abstract:** During this breakout session, we will be developing the validation rules for SBOL 3. **Target audience**: SBOL **Standards**: SBOL	Room SBOL

12:00 – 13:00

R **Discussion about centralizing data for calibrating and validating models and enabling standards (1/2)**

Speakers: Jonathan Karr

Room 1

Motivation and background: Models promise to help scientists understand behavior, help physicians treat disease, and help engineers design biomachines. Realizing such models will likely require substantial, multidimensional data. For example, whole-cell models will likely require data about the sequences of chromosomes; the abundances, modifications, localizations, and half-lives of RNAs and proteins; and the fluxes of reactions. This data is becoming available. This data is also becoming easier to find due to efforts such as OmicsDI. However, it remains difficult to assemble the diverse data that is often required for modeling because much of our existing data is siloed into separate databases for different kinds of data or ad hoc supplementary tables. These sources also support different formats, identifiers, and units. In addition, much of our existing data is difficult to interpret due to the lack of precise metadata about its meaning (e.g., precise subunit composition of the protein complex associated with a kcat measurement). Furthermore, across the community there is duplicate effort to use similar data.

To make it easier to obtain data for biochemical modeling, we have prototyped a notation for precisely describing macromolecules, a simple spreadsheet-based format for reusable datasets, and an integrated database of several types of molecular data. Because there are many challenges to assembling the data needed for modeling, we believe that we must address these issues as a community. Rather than each researcher creating small, siloed datasets, we think there's an opportunity to pool our effort to create central resources that are more useful to everyone.

Session overview: We will discuss a collaborative effort to integrate the data needed for biochemical modeling. We will discuss the community's needs for data, how to create a central resource of data, and the standards needed to enable this. Our goal will be to sketch a roadmap for a working group that will meet online.

Session schedule:

- Intro of participants (10 m)
- Overview of goals (10 m)
  Need for data
  Best existing solutions
  Limitations of the state-of-the-art
  Open challenges
- Proposed roadmap (10 m)
  Standards & ontologies for representing measurements
  Standards for datasets
  Integrated database
- Discussion
  Needed data & metadata (1 h)
  Needed ways to access data (1 h)
  Standards & ontologies (1 h)
  Incentives for investigators to contribute (30 m)
  Database (1 h)
- Next steps (1 h)

Target audience:

- Modelers who would like more systematic access to data
- Data generators who would like to make their data easier to discover, reuse, and compose
- Developers of standards for experimental data

Related standards and repositories: OmicsDI, ISAtab, RightField

S **Social space/BREAK**

Room 1

PINNED

13:00 – 14:00

14:00 – 17:00 R **Discussion about centralizing data for calibrating and validating models and enabling standards (2/2)**

Speakers: Jonathan Karr

Room 1

Abstract: Models promise to help scientists understand behavior, help physicians treat disease, and help engineers design biomachines. Realizing such models will likely require substantial, multidimensional data. For example, whole-cell models will likely require data about the sequences of chromosomes; the abundances, modifications, localizations, and half-lives of RNAs and proteins; and the fluxes of reactions. This data is becoming available. This data is also becoming easier to find due to efforts such as OmicsDI. However, it remains difficult to assemble the diverse data that is often required for modeling because much of our existing data is siloed into separate databases for different kinds of data or ad hoc supplementary tables. These sources also support different formats, identifiers, and units. In addition, much of our existing data is difficult to interpret due to the lack of precise metadata about its meaning (e.g., precise subunit composition of the protein complex associated with a kcat measurement). Furthermore, across the community there is duplicate effort to use similar data.

To make it easier to obtain data for biochemical modeling, we have prototyped a notation for precisely describing macromolecules, a simple spreadsheet-based format for reusable datasets, and an integrated database of several types of molecular data. Because there are many challenges to assembling the data needed for modeling, we believe that we must address these issues as a community. Rather than each researcher creating small, siloed datasets, we think there's an opportunity to pool our effort to create central resources that are more useful to everyone.

Session overview: We will discuss a collaborative effort to integrate the data needed for biochemical modeling. We will discuss the community's needs for data, how to create a central resource of data, and the standards needed to enable this. Our goal will be to sketch a roadmap for a working group that will meet online.

Schedule:

Intro of participants (10 m)
Overview of goals (10 m)
Need for data
Best existing solutions
Limitations of the state-of-the-art
Open challenges
Proposed roadmap (10 m)
Standards & ontologies for representing measurements
Standards for datasets
Integrated database
Discussion
Needed data & metadata (1 h)
Needed ways to access data (1 h)
Standards & ontologies (1 h)
Incentives for investigators to contribute (30 m)
Database (1 h)
Next steps (1 h)

Target audience: Modelers who would like more systematic access to data; Data generators who would like to make their data easier to discover, reuse, and compose; Developers of standards for experimental data

Standards: OmicsDI, ISAtab, RightField

14:00 – 17:00 R **ELIXIR Systems Biology Focus Group**

Room 2

Speakers: John Hancock

Abstract: As systems models become more complex and the need to share and integrate different kinds of model, including large-scale models, becomes more increasingly important, the question of what large-scale infrastructure is needed for systems biology becomes increasingly important. ELIXIR is the European Infrastructure for life science data (www.elixir-europe.org). It has recently established a Focus Group to evaluate the need for infrastructure oriented towards systems biology and what that might need to look like taking into account experiences from ISBE. The session will provide an overview of the Focus Group, some perspectives on what might be needed, and an opportunity to discuss with members of the Focus Group and input into its deliberations.

Please see the **program** for more details.

Target audience: Modellers, tool developers, standards communities

Standards: BioPAX, CellML, COMBINE Archive, NeuroML, SBGN, SBML, SBOL and SBOL Visual, SED-ML

14:00 – 17:00 R **SBOL Visual Parametric SVG**

Room SBOL

Speakers: Christopher Myers

Abstract: During this breakout session, we will be working on the SBOL Visual parametric SVG representation, as well as libraries supporting this representation.

Target audience: SBOL visual

Standards: SBOL visual

PINNED 17:00 – 18:00	S	**Social space/BREAK**	Room 1
18:00 – 19:00	U	**Equilibrator for Metabolic Network Analysis: Thermodynamic Profiling and Enzyme-Cost Minimization** *Speakers: Moritz Beber* **Abstract:** This session will briefly introduce the theory behind thermodynamic and enzyme cost analysis of metabolic pathways. It will focus on the tools that implement this theory and make it accessible for broad use via a web-based interface and a Python package. For thermodynamic analysis, we will first learn how equilibrium constants are estimated using the component-contribution method. Then, we will see how Max-min Driving Force analysis is solved mathematically and what are the motivations behind it. The eQuilibrator website is a user-friendly interface for obtaining equilibrium constants and performing this type of pathway analysis. In parallel, we will show how to perform the same kind of operations programmatically in Python. In the second part of this session, we will learn about Enzyme Cost Minimization (ECM) and how to solve it using convex optimization. We will then introduce an online interface based on NEOS (Network-Enabled Optimization System) that can perform ECM for any given pathway, and furthermore can run a full enzyme-cost optimization for a metabolic network model with kinetic parameters based on Elementary Flux Modes. **Target audience:** modelers interested in better predictions or metabolic engineering **Standards:** COMBINE Archive, SBML, SED-ML, SBtab, ObjTables	Room 1
18:00 – 21:00	U	**Building, exploring and sharing rule-based models of cellular signaling pathways (II/II)** *Speakers: Martin Meier-Schellersheim* **Abstract:** We will provide a practical introduction into creating models of cellular signaling pathways using Simmune's visual modeler and Simmune's Python interface. Whereas the visual modeling approach is biologically intuitive, using iconographic symbols for molecules and their interactions, the Python interface allows modelers to set up simulations with repetitive elements, such as stepwise varying stimulation of cells with chemokines or cytokines. The tutorial will also demonstrate how to explore the behavior of the models through systematic parameter scans. Finally, we show how models created with Simmune can be exported into SBML, including the new SBML3 standard for multi-state, multi-component and multi-compartment biochemical species. **Target audience:** Cell biologists, beginners in modeling, SBML community **Standards:** SBML	Room 2
19:00 – 21:00	U	**MAGINE: From time-series multi-omics to cellular mechanism of action** *Speakers: Alex Lubbock* **Abstract:** Recent technological advances have made it feasible to collect multi-condition transcriptome and proteome time-courses of cellular response to perturbation. The increasing size and complexity of these datasets impedes mechanism of action discovery due to challenges in data management, analysis, visualization, and interpretation. In this tutorial, we will introduce MAGINE (https://github.com/lolab-vu/magine), a software framework to explore complex time-course multi-omics datasets and build mechanistic hypotheses of dynamic cellular response. MAGINE combines data management, enrichment, and network analysis and visualization within an interactive, Jupyter notebook-based environment. To gain the most from this tutorial, we recommend participants have experience using the Python programming language for data analysis. Experience using the "pandas" library is helpful but not required. During the tutorial, we will demonstrate how measurements from HL-60 cellular response to bendamustine treatment can be used to build a mechanistic, multi-resolution description of cellular commitment to fate. Using MAGINE's interactive data analysis and visualization capabilities, we will show how to construct a systems-level description of signal execution from cellular DNA-damage response, to cell cycle arrest, and eventual commitment to apoptosis, mediated by over 2000 biochemical species. This will start with loading and organizing the data in MAGINE. We proceed to construct a molecular interaction network of significantly- changed species from the dataset utilizing large databases such as KEGG for prior knowledge of connectivity. This network is then coarse-grained into an Annotated Gene-set Network (AGN), which incorporates enrichment analysis to combine molecular species dynamics with ontology terms. As part of the enrichment analysis, MAGINE automatically selects the appropriate level of granularity for gene ontology terms, removing redundant or similar terms while also avoiding overly broad ones. Participants will learn how to analyze large, multi-sample, multi-time point, -omics datasets from several platforms (RNA-seq, proteomics) within the integrated MAGINE environment. MAGINE is open source software, so participants will then have the skills required to apply MAGINE to their own data, or other publicly-available datasets. **Target audience:** Modelers and (computational) biologists wishing to gain experience analyzing -omics data **Standards:** -	Room 1

19:00 – 21:00 U **SBOL Visual: Communicating engineered biological designs with diagrams** Room SBOL

Speakers: Thomas Gorochowski

Abstract: The Synthetic Biology Open Language (SBOL) Visual aims to allow for the unambiguous communication of biological design information using diagrams and has already begun to be adopted by many leading groups and journals across bioengineering. Since its original establishment, the scope of SBOL Visual has grown to now include the means to capture information covering sequence features in DNA, other nucleic acids (e.g. RNAs) and molecules (e.g. proteins) as well as the functional relationships and interactions between these components. Recent advances also allow for complex design to be represented in a hierarchical manner, enabling varying levels of details to hide unnecessary specifics and simplify the communication of complex engineered biological systems. This tutorial session aims to help you get started with SBOL Visual and provide:
– an introduction to SBOL Visual, its aims and capabilities;
– hands-on experience drawing SBOL Visual compliant diagrams;
– an update on recent advances to incorporate parametric SVG for greater machine-readable customisations;
– a sneak-peak at paraSBOLv, a Python library for creating SBOL Visual diagrams using code.

Target audience: SBOL community, beginners to SBOL Visual, Any bioengineers

Standards: SBOL and SBOL Visual

21:00 – 23:00 R **SED-ML Breakout Session** Room 1

Speakers: Herbert Sauro, Matthias König

Abstract: A discussion of the current and future developments of SED-ML

Target audience: SED-ML community, modelers, tool developers

Standards: SED-ML

23:00 – 00:00 R **Model Annotation & the OMEX Metadata Spec v1.1** Room 1

Speakers: John Gennari

Abstract: We will begin this session with a review of the basic ideas of Harmonizing Annotations, as well as the scope and goals of the OMEX metadata specification. One purpose of this session is to prepare for a formal vote to ratify version 1.1 (Version 1.0 has been ratified; version 1.1 has been distributed, but not yet voted on.)
Subject to audience interest, I would like to focus on three “open” questions for v1.1 development of the OMEX metadata specification:
• Discuss of the use of “omex-library.org” as a placeholder for building non-relative URI names. This is one of the big changes in the proposed 1.1 specification.
• Discussion of the use and appropriateness of Dublin Core “creator” predicate for specifying model authors & model curators. This is issue is also tied to our recommendation for the use of OrcIDs to identify researchers / curators.
• Discussion of completeness (or sufficiency) of the listed model-level annotations: Creator, Created, isDescribedBy, isDerivedFrom, Description, and hasTaxon.
I would welcome other annotation-related topics for discussion. Time and interest permitting, we could also review and demonstrate the Python and Java implementations of the OMEX metadata specification.

Target audience: Modelers, annotators, repository curators

Standards: CellML, COMBINE Archive, SBML, SED-ML

COMBINE 2020

[R] Breakout [D] Discussion [S] Social space [U] Tutorial

OCTOBER 9 • FRIDAY

PINNED 00:00 – 01:00	S	**Social space/BREAK**	Room 1
01:00 – 04:00	U	**Center for Reproducible Biomedical Modeling Tutorial** *Speakers: Veronica Porubsky* **Abstract:** The Center for Reproducible Biomedical Modeling will host a tutorial to demonstrate how to build a biochemical model reproducibly, using COMBINE community standards and tools which support them. Participants will leave with a better understanding of the importance of reproducible methods across a model building and simulation workflow, and will be introduced to resources where they may seek assistance from developers and experts within the COMBINE community to assist with their model building challenges, especially those challenges which inhibit reproducibility. Participants will gain practical experience in the import and export of SBML models into standard-compatible tools, data aggregation and curation, writing model equations and defining parameters, annotating models, parameter fitting, writing simulation experiments which follow the MIASE guidelines, simulating models using SED-ML, model verification and validation. The tutorial will cover an entire modeling workflow using a single model and set of compatible tools - primarily Tellurium and associated utilities - to show how reproducibility can be practically implemented, but alternative tools which could be used to achieve each task will also be mentioned. Participants will have the opportunity to validate that they have independently reproduced the discussed model and simulation results at the end of the tutorial. **Target audience:** SBML community, modelers **Standards:** CellML, COMBINE Archive, SBML, SED-ML	Room 1
PINNED 04:00 – 06:00	S	**Social space/BREAK**	Room 1
06:00 – 09:00	U	**libCellML: How to get started** *Speakers: Keri Moyle* **Abstract:** This tutorial will introduce the use of libCellML to compose, validate, and simulate the Hodgkin-Huxley 1952 model using Python and/or C++. The prerequisite for this tutorial in Python is Python 3.7 (minimum). Prerequisites for this tutorial in C++ are CMake (3.10.2 minimum) and a C++ toolchain (with C++17 support minimum). **Target audience:** tool developers, modellers **Standards:** CellML, SED-ML	Room 1
PINNED 09:00 – 10:00	S	**Social space/BREAK**	Room 1
10:00 – 13:00	R	**SBOL 3 Examples and Use Cases** *Speakers: Christopher Myers* **Abstract:** During this breakout, we will working SBOL 3 examples and use cases. **Target audience**: SBOL **Standards**: SBOL	Room SBOL

11:00 – 13:00 | R | **Reproducibility in Systems Biology Modelling** | Room 1

Speakers: Malik-Sheriff, Rahuman S.

The reproducibility crisis has emerged as an important concern across many fields of science including life science, since many published results failed to reproduce [1-3]. Systems biology modelling, which involves mathematical representation of biological processes to study complex system behaviour, was expected to be least affected by this crisis. While lack of reproducibility of experiment results and computational analysis could be a repercussion of several compounded factors [4-7], it was not fully understood why systems biology models with well defined mathematical expressions fail to reproduce and how prevalent it is. Hence, we systematically attempted to reproduce 455 kinetic models of biological processes published in peer-reviewed research articles from 152 journals; which is collectively a work of about 1400 scientists from 49 countries. This is one of the largest studies ever performed to independently assess reproducibility of scientific results across any fields. Our investigation revealed that about half (49%) of the models could not be reproduced using the information provided in the published manuscripts. With further effort, an additional 12% of the models could be reproduced either by empirical correction or support from authors. The other 37% models remained non-reproducible due to missing parameter values, missing initial concentration, inconsistent model structure, or a combination of these factors. Among the corresponding authors of the non-reproducible model we contacted, less than 30% responded. Our analysis revealed that models published in journals across several fields of life science failed to reproduce, revealing a common problem in the peer-review process. Hence, we proposed an 8-point reproducibility scorecard that can be used by authors, reviewers and journal editors to assess each model and address the reproducibility crisis [8]. In this breakout session, we would like to discuss in details about our reproducibility scorecard.

Baker 2016, Nature
Fanelli 2018, PNAS
Goodman 2016, Sci Trans Med
Errington et al. 2014 ELife
Kulkarni et al. 2018, BMC Bioinformatics
Papin et al. 2020, PLOS Computational Biology;
Pusztai et al. 2013 Nature Reviews Clinical Oncology
Tiwari et al. 2020 BioRxiv

Target audience: Modellers concerned about reproducibility (in principle all)
Standards: CellML, COMBINE Archive, NeuroML, SBML, SBOL and SBOL Visual, SED-ML

PINNED
13:00 – 14:00 | S | **Social space/BREAK** | Room 1

14:00 – 16:00 | R | **ModeleXchange - Are We Ready Yet?** | Room 1

Speakers: Henning Hermjakob

Abstract: Model repositories play an essential role in making models easily sharable and accessible. Currently, searching for a model requires users to visit multiple repositories and look for a model of their interest. In model curation, multiple repositories might independently try to reproduce a model and fail, a huge waste of time and public resources. Initiated at COMBINE 2019 and continued in multiple sessions at HARMONY 2020, key model repositories have initiated the ModeleXchange collaboration, aiming to co-ordinate curation efforts, provide a common metadata-based search interface, and overall better connect the global efforts in model repository provision. We will review the current status of ModeleXchange and discuss next steps and priorities.
Target audience: Model database providers and stakeholders
Standards: CellML, COMBINE Archive, SBML, SBOL and SBOL Visual

14:00 – 17:00 | R | **SBOL Visual Workflow and Ontology** | Room SBOL

Speakers: Christopher Myers

Abstract: During this breakout, we will be working on the workflow for extending SBOL Visual, and the SBOL Visual Ontology.
Target audience: SBOL visual
Standards: SBOL visual

14:00 – 17:00 | R | **Workshop on standardised neuronal network specifications** | Room 2

Speakers: Padraig Gleeson

Abstract: Many biophysical neuronal network simulations contain descriptions of structured connectivity between groups of neuronal elements (e.g. cells, populations, brain areas). These can be specified in many ways, often in formats quite closely tied to the simulator's internal data model. Extracting these network descriptions, to allow the network structure to be visualised, analysed, shared, and importantly compared with other network models and anatomical data, is a key part of maximising their usefulness as scientific entities.
While procedural creation of networks is common across all simulators (e.g. using Python scripts), there is still a need for a consensus on a declarative, compact, high level format for describing the properties of the networks (i.e. a reusable template for generating instances of the network) which encapsulates the essential biologically relevant features of the modelled system.

This workshop will gather together developers of simulators and interoperability formats to:
1) Gather information on existing (internal) formats being used by simulators for defining network structures
2) Discuss initiatives to make cross simulator formats for exchanging network specifications
3) Decide on future plans to minimise repetition of work, ensure interoperability and maximise usability and accessibility of the formats for end users

This workshop is organised as part of the INCF Special Interest Group (SIG) on Standardised Representations of Network Structures.

Target audience: Computational neuroscientists, neuronal modellers
Standards: NeuroML

More info: https://neuralensemble.github.io/Networks_SIG/COMBINE2020

16:00 – 17:00 | R | **Improving annotation and COMBINE archives: Problems, open research ideas & task brainstorming** | Room 1

Speakers: John Gennari

Abstract: This session aims to be of broad interest for those working in model reuse and sharing. We are seeking ideas for effective demonstration of the value of model annotation. As an example, we could demonstrate cross-library search and retrieval of appropriate models. This might require:
• Automatic and semi-automatic annotation efforts.
• Large-scale "OMEX-ication" of BioModels and CellML repositories.
• Improved consistency of annotation norms (e.g. which ontologies to use)
• Improved version control and model provenance information ("isDerivedFrom")
• Improving/extending the BioModel.net collection of qualifiers
As a specific example for the last point, should we consider adding the qualifiers "hasReactant" and "hasProduct" to the list of sanctioned BioModels.net qualifiers? More generally, should we consider "hasSourceParticipant" and "hasSinkParticipant"?
Finally, we should consider how a large collection of OMEX archives might differ from (or be the same as) the goals of the ModelExchange effort.
Target audience: COMBINE community
Standards: BioPAX, CellML, COMBINE Archive, NeuroML, SBGN, SBML, SBOL and SBOL Visual, SED-ML

PINNED
17:00 – 18:00 | S | **Social space/BREAK** | Room 1

18:00 – 19:00 | U | **FlapJack** | Room 1

Speakers: Gonzalo Vidal

Abstract: Characterization is fundamental to the design, build, test, learn (DBTL) cycle for engineering synthetic genetic circuits. Components must be described in such a way as to account for their behavior in a range of contexts. Measurements and associated metadata, including part composition, constitute the test phase of the DBTL cycle.With current robotic assays these data may consist of thousands of DNAs, measured in hundreds of conditions, in multiple assays, potentially performed in different labs and using different techniques. In order to inform the learn phase this large volume of data must be filtered, collated, and analyzed. In general this process consists of parameterizing models of component function in different contexts, and combining them to predict behaviors of novel circuits. More sophisticated approaches based on machine learning or deep learning require labels (metadata) and features (parameters) in order to train neural networks to classify or predict circuit behaviors. Tools to store,organize, share, and analyze large volumes of measurement and metadata are therefore essential to linking the test phase to the build and learn phases, closing the loop of the DBTL cycle. In this workshop we present such a system, implemented as a web app with a backend database, a frontend that provides sophisticated filtering, plotting and analysis tools, and a Python API for integration with external build and learn software. All measurements are associated to circuit part composition via SBOL. We will teach how to use our tool from the front end to API by characterizing and plotting relevant information of a broad range of genetic components and circuits according to part composition and context.

Target audience: biologists, data scientists, SBOL community, synthetic biologists

Standards: SBOL and SBOL Visual

18:00 – 19:00 | U | **PySB: a mathematical framework for modeling biochemical reactions as python programs** | Room 2

Speakers: Samantha Beik

Abstract: Models of cellular processes have moved beyond writing a set of ODEs to a complex pipeline that involves model building, parameter calibration, model analysis, and subsequent release of all details. Traditional methods to encode

models as mathematical equations to describe the behavior of cellular processes are difficult to revise, extend, and share. Python Systems Biology (PySB) is a modeling framework to build, maintain, and extend models by leveraging software practices and tools. In PySB, biological interactions are encoded as rules, and these in turn abstracted as functional objects to describe a system of biochemical reactions as a native Python program. In addition, PySB allows users to access thousands of tools for model calibration, analysis, and dissemination using a Literate Programming paradigm that facilitates model development. PySB is open source and easily extendable, making it ideal for many applications to different types of biological modeling.

The workshop should last ~3 hours including demonstration and practice sessions. It will be broken into three parts: The first part will be the introduction providing a background on modeling, rule-based models, and how to represent mathematical equations into PySB; the second part will go into detail about the PySB framework, demonstrating how calibration and analysis can be used in the Python environment; Finally, we will demonstrate the use of Jupyter notebooks for model dissemination as an interactive document. Attendees will go through the tutorial we have developed and work with them answering questions for the remainder of the workshop.

Target audience: Those with interest in programming and/or modeling

Standards: BioPAX, SBGN, SBML

19:00 – 21:00 | U | **Modelling with VCell** | Room 1

Speakers: Michael Blinov

Abstract: We will demonstrate the whole suite of VCell features, including deterministic (compartmental ODE or reaction-diffusion-advection PDE with support for 2D kinematics), stochastic (SSA solvers), spatial stochastic (reaction-diffusion with Smoldyn), hybrid deterministic/stochastic and network-free agent based simulations, as well as geometries from 2D or 3D microscope images or from idealized analytical expressions and modeling of membrane flux, lateral membrane diffusion and electrophysiology. The developers of VCell will also be available for the discussion of individual questions from the participants.

Target audience: modelers, beginners in modeling, tool developers

Standards: BioPAX, COMBINE Archive, SBGN, SBML, SED-ML

PINNED
21:00 – 21:30 | D | **Closing session** | Room 1