

INCOME – integrative, kollaborative modellierung in der systemmedizin

Per Hackathon in eine Kultur des *Data and Model Sharing*

von Nina Fischer, Wolfgang Müller, Dagmar Waltemath, Olaf Wolkenhauer und Jan Hasenauer

Die Systemmedizin ist ein interdisziplinärer Ansatz, bei dem Ärzt*innen und klinische Forscher*innen mit Expert*innen aus den Bereichen Biologie, Biostatistik, Informatik und Mathematik zusammenarbeiten, um Diagnostik, Prävention und Behandlung von Krankheiten zu verbessern. Dafür müssen großskalige, heterogene Datensätze experimentell erhoben und mittels ganzheitlicher Modelle zusammengeführt werden. In der Realität beschränken sich jedoch die meisten Projekte auf einzelne zelluläre Prozesse und entwickeln hierfür maßgeschneiderte Modelle. Im BMBF-geförderten e:Med Projekt INCOME haben wir mit gezielten Vernetzungsaktivitäten die Zusammenarbeit und den Austausch zwischen Forschungsgruppen gefördert. Forscher*innen und Entwickler*innen schufen gemeinsam in zahlreichen INCOME-Meetings eine Kultur des „Data and Model Sharing“ und entwickelten so Wege zur besseren technischen Verknüpfung und Wiederverwendbarkeit bestehender Simulationsmodelle. Die Meetings boten damit eine Plattform zur Stärkung der Community.

Mathematische Modelle werden in der Systemmedizin verwendet, um krankheitsbezogene biochemische Prozesse für Diagnostik, Prognostik und therapeutische Entscheidungen zu untersuchen. Diese Modelle können in offenen Standards wie der Systems Biology Markup Language (SBML) formuliert und in Modelldatenbanken wie BioModels hinterlegt werden. Standards wie die Simulation Experiment Description Language (SED-ML) zielen auf eine eindeutige Verknüpfung von Daten

und Modellen sowie die Beschreibung verschiedener Experimentalbedingungen ab. Mit der Entwicklung und Etablierung der Standards und Datenbanken haben sich die Verfügbarkeit und Wiederverwendbarkeit von Modellen und die Reproduzierbarkeit von Forschungsergebnissen massiv verbessert. Dennoch haben viele Forschungsgruppen und auch weit verbreitete Software Tools diese Standards noch nicht adaptiert und verwenden stattdessen eigene wenig standardisierte Formate, die unter anderem oft keine rigorose Annotation gestatten. Dementsprechend sind gemäß aktueller Erhebungen weiterhin nur ca. 50% der veröffentlichten Modellierungsstudien reproduzierbar (Tiwari *et al.*, 2021).

Neben Modellen selbst sind auch die zur Parametrisierung und Validierung verwendeten Datensätze oft schwer zugänglich. Sie werden nicht veröffentlicht, nicht ausreichend annotiert oder nicht eindeutig an das Modell geknüpft. Die genannten Faktoren stehen der Wiederverwendung und Erweiterung bestehender Modelle und Datensätze im Wege und kompromittieren somit den Erfolg von Forschungsprojekten.

Die stark begrenzte Wiederverwendbarkeit und der Zeitaufwand für die Entwicklung qualitativ hochwertiger Modelle sind zwei Gründe dafür, dass viele mechanistische Modelle nur einzelne Signalwege abbilden und Wechselwirkungen mit anderen Signalwegen oft ignorieren. Es ist an der Zeit, die



Abbildung 1: Impressionen aus zwei INCOME Veranstaltungen. INCOME 2019 in Berlin und INCOME 2021 virtuell. (Fotos: Jan Hasenauer).

bestehenden Infrastrukturen zu nutzen, Forschende mit den vorhandenen Ressourcen vertraut zu machen und Software-Entwickler*innen auf verfügbare Bibliotheken und Werkzeuge für die standard-konforme Modellierung explizit aufmerksam zu machen.

Wichtigste Ziele des INCOME Projektes

Ziele des Verbundprojekts INCOME waren

1. die Vernetzung von Forschungsgruppen zu verbessern und die Verwendung von Standards voranzutreiben,
2. eine integrierte Datenbank für Modelle und Daten zu etablieren, und
3. die kollaborative Entwicklung großskaliger Modelle zu vereinfachen.

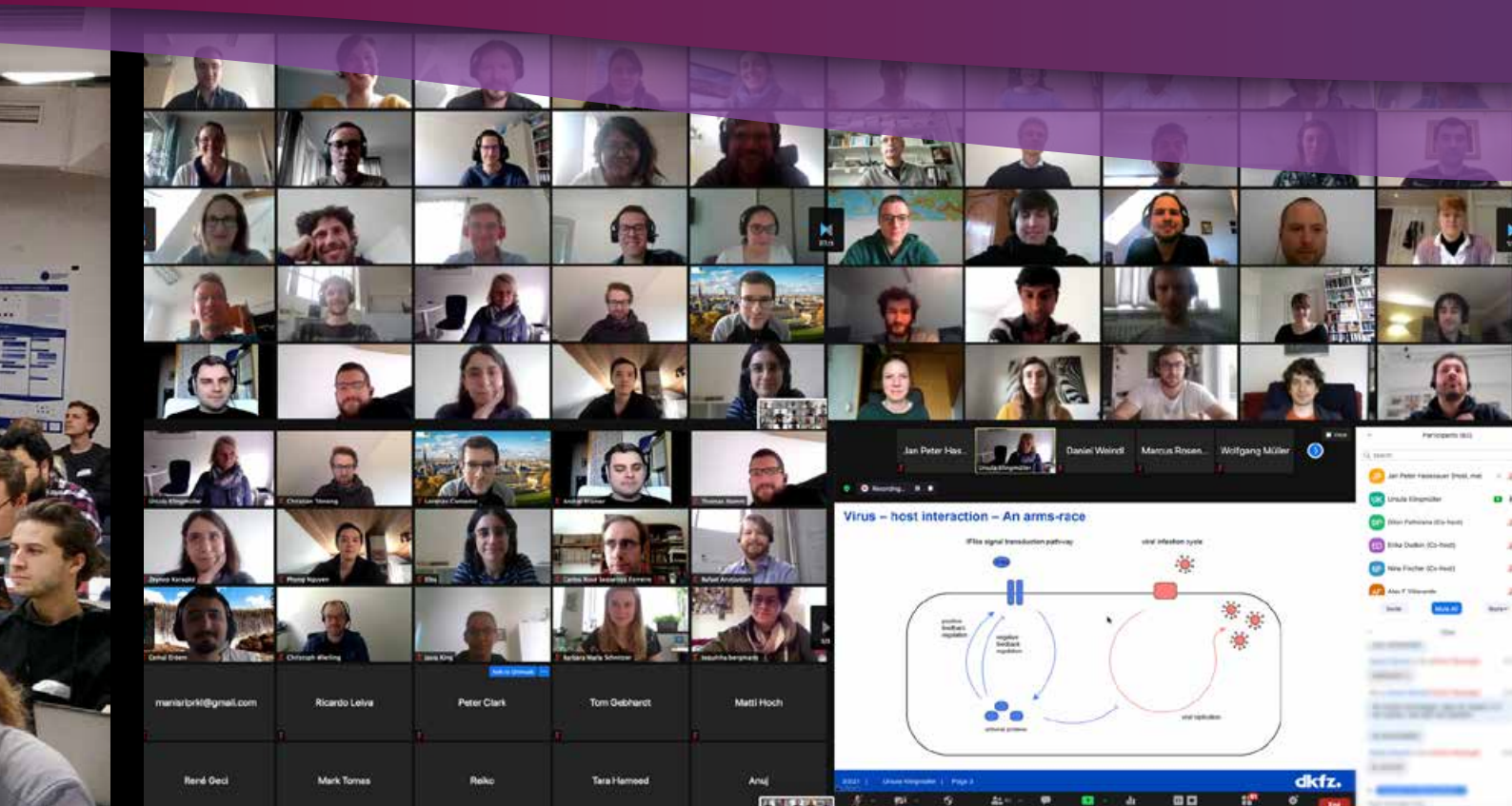
An der Umsetzung dieser Ziele wurde an mehreren Standorten in Deutschland gearbeitet. INCOME (2017 – 2021) wurde von Jan Hasenauer (Helmholtz Zentrum München und Uni Bonn) koordiniert und mit Wolfgang Müller (HITS Heidelberg), Olaf Wolkenhauer (Uni Rostock) und Dagmar Waltemath (Universitätsmedizin Greifswald, vormals Uni Rostock) bearbeitet.

Training, Networking und Community Building

INCOME hat insgesamt fünf Veranstaltungen durchgeführt: drei Konferenzen mit integrierten Hackathons, und zwei separate Hackathons (siehe Abbildung 1).

Die Konferenzen informierten über aktuelle Ergebnisse im Feld der Modellbildung, der Methodenentwicklung und der Anwendung. Eine Vielzahl von Vorträgen hat zudem der nächsten Generation von Systembiologen*innen und Systemmediziner*innen Standards nähergebracht und deren Nutzen aufgezeigt. Podiumsdiskussionen sowie Gruppen- und Einzelgespräche wurden dazu genutzt, offene Probleme anzusprechen und Lösungsansätze zu erarbeiten. Die Konferenzen hatten einen hohen Vernetzungscharakter und ermöglichten somit Nachwuchswissenschaftler*innen einen Austausch mit erfahrenen Wissenschaftler*innen der Community.

Die Hackathons boten ergänzend zu den Konferenzen Raum und Zeit für das Arbeiten an gemeinsamen Projekten. Um die Verwendung von Standards deutlich zu vereinfachen und sowohl Reproduzierbarkeit als auch Wiederverwendbarkeit zu verbessern, wurde Support von Standards in vielen verfügbaren Software Tools angeboten (z.B. Copasi, Data2Dynamics, Dmod und pyPESTO). Darüber hinaus haben Entwickler*innen sich Zeit genommen, neue Software vorzustellen und Nutzer*innen bei Problemen zu beraten. Im Gegenzug erhielten Sie direktes Feedback zu eigenen Tools und konnten die Weiterentwicklung an den Bedürfnissen der User ausrichten.



Die Veranstaltungen lockten Teilnehmer aus über 20 Ländern an. Diese Internationalität und die ebenfalls vorhandene Interdisziplinarität haben maßgeblich zur Entwicklung einer starken Community an der Schnittstelle von Modellierung, Methodenentwicklung und Standardisierung beigetragen. Die kontinuierlich steigende Teilnehmerzahl der Veranstaltungen spiegelt das wachsende Interesse hieran und allgemein an einer intensiven Zusammenarbeit wider (siehe Abbildung 2). Viele Aspekte der INCOME Veranstaltungen fanden direkt Eingang in die Entwicklungen des globalen Standardisierungsnetzwerks COMBINE (www.combine.org).

Lücken schließen

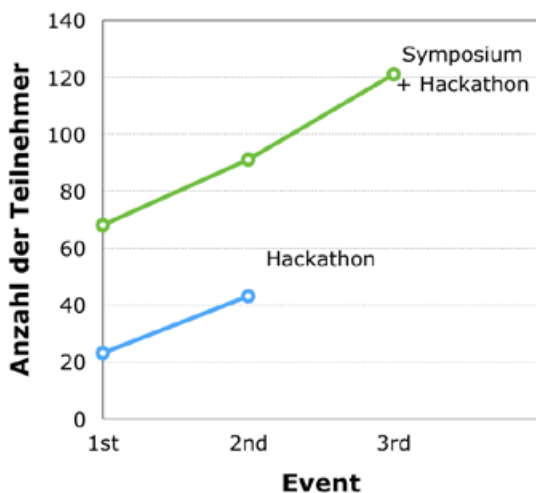
INCOME hat neben der Implementierung von Standards in existierende Workflows auch offene Fragen identifiziert und adressiert. So wurde beispielsweise bereits bei der ersten Veranstaltung im Oktober 2018 erkannt, dass es keine passende Lösung gibt, um Parametrisierungsprobleme zu formulieren. Um diese Lücke zu schließen, wurde während der Hackathons, aber auch in der Zeit zwischen den Treffen, das Parameter Estimation table format (PETab) entwickelt (siehe Abbildung 3). PETab erlaubt eine Standardisierung auf einer neuen Ebene, und verbessert damit die Wiederverwendbarkeit von Daten und Modellen für neue Studien, z.B. die Entwicklung ganzheitlicher Modelle.

An der Entwicklung von PETab hat eine Vielzahl von Forschungsgruppen mitgearbeitet und so ist der neue Standard bereits nach kürzester Zeit in acht Software Tools für Tausende von Nutzern zugänglich (Schmiester *et al.*, 2021). Dies wäre ohne die Ausbildung einer starken, thematisch fokussierten Community und die gemeinsamen Hackathons nicht möglich gewesen.

Ressourcen schaffen

INCOME hat dazu beigetragen, Community Ressourcen zu schaffen, um die Systembiologie und die Systemmedizin als Forschungsgebiete voranzubringen. So wurde ein zuvor etabliertes Repository für Datensätze und Modelle (Hass *et al.*, 2019) in PETab überführt und kontinuierlich erweitert. Dies erleichterte nicht nur die Wiederverwendung, sondern ermöglicht auch ein realistisches Benchmarking von Methoden. Eine Untersuchung zur Skalierbarkeit und Robustheit von existierenden numerischen Lösungsverfahren wurde bereits abgeschlossen (Städter *et al.*, 2021), und mehrere weitere Studien zu Optimierungs- und Stichprobengenerierungsverfahren sollten

A Entwicklung der Teilnehmeranzahl



B Zusammensetzung der Teilnehmer

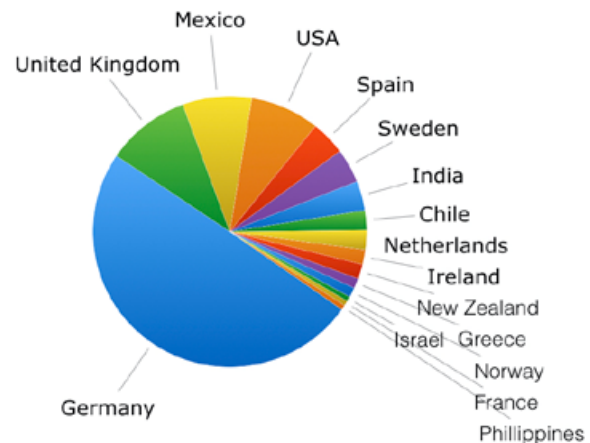


Abbildung 2: A Statistik aus Anmeldedaten (Anzahl der Teilnehmer) der 5 Veranstaltungen von 2018-2021, B Statistik aus Anmeldedaten der virtuellen Veranstaltung 2021. (Quelle: Jan Hasenauer).

zeitnah verfügbar sein. Darüber hinaus hat das INCOME Konsortium als Teil einer großen, weltumspannenden Community unter der Leitung von Marek Ostaszewski (Luxembourg Centre for Systems Biomedicine) an einem Modell für SARS-CoV-2 Infektionen gearbeitet (Ostaszewski *et al.*, 2020). Die daraus resultierende COVID-19 Disease Map ist wohl die mit Abstand umfangreichste Beschreibung der molekularen und zellulären Prozesse während der Infektion.

Alle geschaffenen Ressourcen sind frei verfügbar und können auf Grund ihres hohen Grades an Standardisierung einfach in weiteren Projekten genutzt werden. Dies ermöglicht einen Wissenstransfer und vereinfacht eine ganzheitliche Betrachtung von Krankheiten.

Aber wo geht es jetzt hin?

INCOME hat allen Konsortiumsmitgliedern aufgezeigt, wieviel mehr man als Teil einer starken Community erreichen kann, und wie wichtig im Umkehrschluss Community Building ist. Die Trainings während der Konferenzen, der Austausch und die gemeinsame Arbeit während der Hackathons, sowie die zahl-

reichen Diskussionen haben dazu beigetragen, Modell-, Software- und Standardentwicklung näher zueinander zu bringen. Dies war ein Prozess. Einzelne Treffen wären nicht ausreichend gewesen und es ergab sich ein substanzieller Mehrwert durch den wiederholten Austausch, wiederkehrende Teilnehmer, die wiederum in ihren Forscherkreisen für die Meetings geworben hatten. Wir gehen fest davon aus, dass das Wissen, welches durch die INCOME Veranstaltungen in einzelne Forschungseinrichtungen getragen worden ist und wird, die Wiederverwendbarkeit von Modellen deutlich verbessert hat und weiterhin steigern wird. Künftig sollen noch viel mehr Forscher*innen erreicht werden! Wir planen daher die Veranstaltungsreihe auch über das Projektende hinaus fortzusetzen (und arbeiten an der Realisierbarkeit).

Wir möchten uns bei allen Teilnehmer*innen der Konferenzen und Hackathons bedanken, die den Erfolg von INCOME möglich gemacht haben.

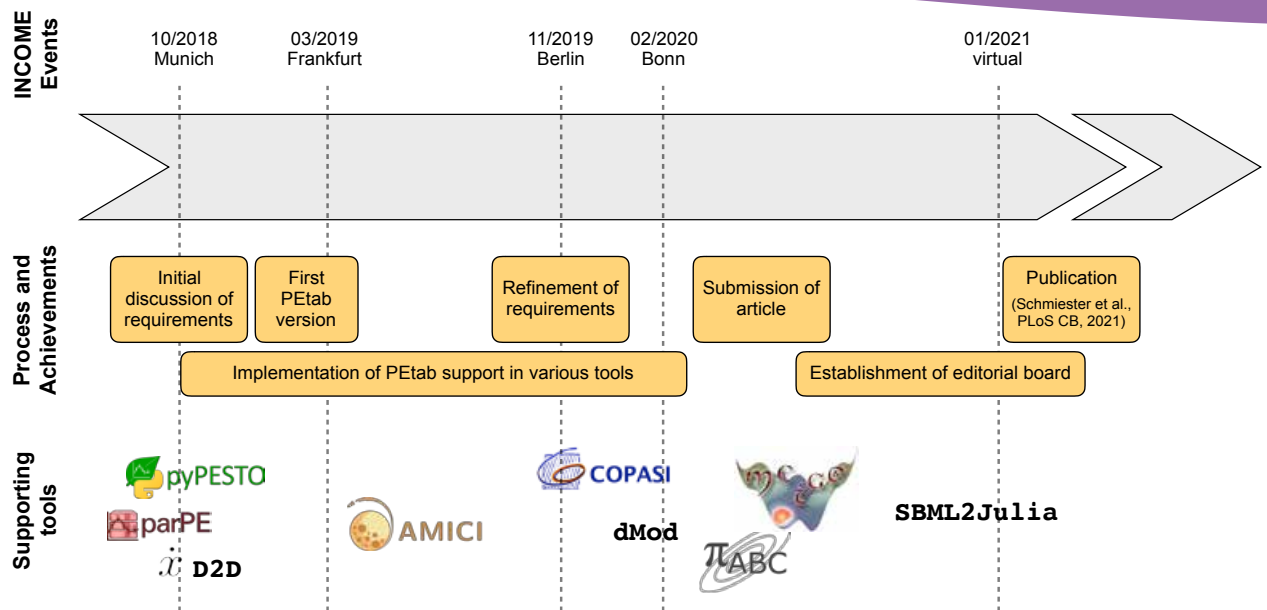


Abbildung 3: Entwicklung des PETab Standards im Rahmen der INCOME Konferenzen (Quelle: Jan Hasenauer).

Referenzen:

- Tiwari *et al.* (2021). Reproducibility in systems biology modeling. *Molecular Systems Biology*, 17:e9982.
- Schmiester *et al.* (2021). PETab - Interoperable specification of parameter estimation problems in systems biology. *PLOS Computational Biology*, 17(1):e1008646.
- Hass *et al.* (2019). Benchmark problems for dynamic modeling of intracellular processes. *Bioinformatics*, 35(17):3073-3082.
- Städter *et al.* (2021). Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, 11:2696.
- Ostaszewski *et al.* (2020). COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Scientific Data*, 1:136.

Kontakt:



Prof. Dr. Jan Hasenauer
 Life and Medica Sciences (LIMES) Institute
 University of Bonn
 jan.hasenauer@uni-bonn.de

<https://www.mathematics-and-life-sciences.uni-bonn.de/de>



E:MED VERNETZUNGSFONDS

Die **e:Med Vernetzungsfonds-Projekte** zeigen, wie in Hackathons, Workshops und Summer Schools wertvolle Ressourcen wie Software-Tools, Repositorien und Standards erarbeitet werden können. Ein Erfolg dieser Vorhaben sind konkrete translationale Anwendungen in klinischen Studien, wie zu Magenkrebs, Brustkrebs und entzündlichen Darmerkrankungen. Sechs Verbünde vernetzen jeweils gezielt mehrere e:Med Vorhaben. Die Wissenschaftler*innen bringen ihre Expertise zu relevanten Querschnittsthemen ein und erzeugen so durch ihre interdisziplinäre Zusammenarbeit in den Bereichen Klinik, Grundlagen, Bioinformatik und Modellierung zusätzlich entscheidenden Mehrwert. Diese Initiative hat das BMBF im Rahmen seiner Förderinitiative e:Med mit Mitteln aus einem Vernetzungsfonds ermöglicht.

www.sys-med.de