Challenges and opportunities for system biology standards and tools in medical research

Matthias König[1,†], Anika Oellrich[2,†], Dagmar Waltemath[3,†,*], Richard JB Dobson[2,4], Tim JP Hubbard[5] and Olaf Wolkenhauer[3]

[1]Humboldt University Berlin, Institute for Theoretical Biology, D-10115 Berlin, Germany
[2]King's College London, IoPPN, London, SE5 8AF, UK
[3]University of Rostock, Department of Systems Biology and Bioinformatics, D-18051 Rostock, Germany
[4]UCL Institute of Health Informatics, Farr Institute of Health Informatics Research, University College London, London, WC1E 6BT, UK
[5]King's College London, Department of Medical & Molecular Genetics, London, SE1 9RT, UK
[†]Authors contributed equally.

ABSTRACT

Kinetic models are increasingly relevant in medical research. In systems biology, more than 10 years of experience with the development of standards and tools to construct and analyse kinetic models exists. This has supported the sharing of kinetic models, increased their reuse, and thereby has helped to reproduce and validate scientific results. Given the expertise and the existing infrastructure, it seems natural to consider the application and development of standards and tools to meet the requirements of medical scientists.

In this paper, we discuss challenges and opportunities for standards and tools from systems biology in medical research, and we put forward criteria for the safe use of simulations. We conclude that standards, tools and infrastructure need to be extended to ensure the quality, reliability and safety required when working with medical and patient data. This will foster the adaptation of modelling in the clinic, providing tools for improved diagnosis, prognosis and therapy.
Contact: dagmar.waltemath@uni-rostock.de

1 INTRODUCTION

In modern medicine, technologies complement conventional clinical data with molecular and genetic information. Patient-specific molecular profiling provides opportunities for earlier diagnosis, more accurate prognoses and optimised therapeutic decisions [1]. The data generated from these new technologies have led to a rise of computational approaches in medicine [2, 3, 4].

'Personalised Medicine' and 'Systems Medicine' are two terms that are frequently used to capture this trend for interdisciplinary approaches in which clinical research, molecular and cell biology, medical informatics, bioinformatics, biostatistics and systems biology approaches join forces. Personalised medicine uses marker-assisted diagnosis and targeted therapies derived from an individual's molecular profile and patient data [5]. Systems medicine aims to bring computational models closer to the clinic to shed light on the dynamic complexity of human physiology and disease [6].

In this context, the focus for systems biology approaches has been on the modelling of phenomena, where an understanding of processes (kinetics) is crucial. This includes the response of cells, tissues and organs to drugs, e. g. [7]; the simulation of disease progression, e. g. [8]; and the understanding of mechanisms, as opposed to just predicting outcomes, e. g [9].

With new technologies available to provide us with data to identify and characterise disease relevant components, there is an increasing demand for methodologies that enable us to study the interactions of molecular and cellular components in a patient. Arguably, the success of systems and personalised medicine relies then on the application of kinetic models in the clinic [10].

The construction of kinetic models for the clinic requires an integration of clinical and patient-specific molecular data with public databases like KEGG [11], UniProt [12], BLAST [13], Ensembl [14], ENCODE [15], GEO [16], STRING [17] to name only a few. This process effectively brings together the two worlds of basic research and clinical practice. For this union to succeed, ontologies will play a crucial role. Standards to encode information together with ontologies to unambiguously characterise domain knowledge, form the basis for the development of tools that can analyse kinetic models. These tools in turn support the sharing and reuse of models, which is also a means to validate results and generally improve reproducibility in medical research.

Here we illustrate open challenges that need to be overcome in future work to achieve trustworthy systems that can easily be integrated in any hospital environment or GP practice. We further outline criteria that need to be fulfilled to ensure patient and data safety, and to eliminate ethical concerns.

2 HURDLES IN THE APPLICATION OF SYSTEMS BIOLOGY STANDARDS AND TOOLS

A number of challenges need to be overcome before systems biology standards and tools can be applied in medical research. These challenges are further detailed in the following subsections.

2.1 Access to clinical data

Almost no clinical data sets are available for integration with models, neither are these data sets sufficiently documented in a formalised manner. Consequently, the selection process of clinical data for a given model (and *vice versa*) is hindered. This is partly due to patient data being sensitive, limiting its accessibility for analysis,

*to whom correspondence should be addressed

but mainly due to missing incentives, guidelines and requirements of providing data access upon publication of clinical studies.

Clinical data sets are required for testing as well as prediction purposes. While a reoccurring complaint is the lack of suitable data sets to test a model with, this problem is hard to overcome given that patient data needs to be secured over unauthorised access at all times or anonymised in a proper manner. Some efforts such as the 100,000 genome project conducted by Genomics England[1] and the openEHR[2] project aim to provide access to structured, semantically annotated clinical data for research purposes. However, the amount of available data is still too limited to test models and computational simulations reliably.

In practise, most research data are neither shared nor recycled outside the original project team [18]. Models are instead being developed and used within a single clinic, e. g. by collaborative projects that incorporate clinical research groups and computational biology groups located in the same institution. In these settings, however, modelling has already been applied successfully, for example to study melanoma resistance to immunotherapy [19].

In addition to relevant clinical data being accessible, it must be represented in a way that it can be integrated and interpreted by both humans and machines. This requires a dialogue not only between healthcare providers and researchers, but also with staff recording the data and policy makers regulating patient data records.

2.2 Good quality models and documentation

Currently, the majority of published models are not available in standard formats, and the model quality is not sufficiently documented. In many cases, even the computational code underlying a model is inaccessible. Without the ability of reproducing the models, they cannot be exploited for clinical use. In addition, available models are not fully annotated, i.e. the description of model components and parameters are missing, hindering interpretation and integration with other models and clinical data. Model provenance information is not kept, leading to misinterpretations and even irreproducibility of the original findings.

Major efforts are required to reproduce the results reported in a publication of a computational model. Ongoing efforts such as curation processes in BioModels [20], or the provision of fully reproducible archives of virtual experiments in the Physiome Model Repository [21] improve this situation. However, curation is very slow due to the manual labour involved and seldom performed after a model has been published. Moreover, we lack concerted efforts for model validation, annotation, and conversion into computable formats.

2.3 Standardised representation of models and data

The systems biology community developed a set of interoperable standards for modeling, including the Systems Biology Markup Language (SBML), CellML, Synthetic Biology Markup Language (SBOL), NeuroML, Simulation Experiment Description Markup Language (SED-ML), or BioPax [22]. These standards, however, are specific to the computational biology community. As a consequence, sharing and/or integrating models across communities can be challenging, as different standards are used for the representation and annotation of the data.

Additionally, there is no consensus on which ontologies should be used for data and model representation and to which degree of detail models and data need to be annotated, creating further obstacles to integrate models for simulation purposes. Extensive cross-domain initiatives need to be built and are required to take decisions on ontologies and standards that are not only convenient for model developers, curators and researchers, but that are also practical (from an implementation and cost point of view) in a clinical application scenario.

2.4 Validated predictions in a clinical context

A major hurdle for the translation of computational models into medical research is the difficulty to proof the efficiency and predictive value of the model. Every recommendation determined by a clinical decision support system needs to be in line with the policies for medical care providers as issued by the health authorities in the respective country. In order to proof health economic efficiency, extensive, potentially double-blinded, clinical trials are required that compare model-based treatment decisions with unsupported decisions by clinical staff. These clinical trials have to span over all areas of clinical application, i.e. cover different types of diseases as well as ranges of treatments and patients in differing health conditions to assess clinical safety. Every *in silico* model provides an estimation of pathological processes and therefore naturally contains errors. These errors can potentially lead to wrong treatment decisions, which is why great care needs to be taken when transporting systems biology models, standards and tools into clinical practice.

2.5 Detailed documentation of virtual experiments

Finally, it is challenging to describe the virtual experiments that can be applied to a model. We argued before that promoting the reuse of such virtual experiments would vastly improve the usefulness and relevance of computational models in biomedical endeavours and large scale biomedical research projects such as the *Virtual Physiological Human* [23]. While models are being published and made available, they lack descriptions of the simulation settings. These simulation settings are required in order to reproduce and verify the results released with the model. They indicate how a model can be sensibly applied in medical research, and what further work is needed to ensure patient safety. A standard for the encoding of simulation setups is SED-ML [24]. It defines the models to be used in a virtual experiment, together with possible parametrisations and simulation setups. However, SED-ML to date encodes only for a subset of experiments performed in clinical research. Further extensions are needed in the standard itself. Similarly important is extended software support.

3 CRITERIA FOR REUSABLE SIMULATION MODULES AND SEMANTIC DATA

The reproducibility and reusability of models and model-based results have been discussed in several assays over the past years [10, 25, 26]. One conclusion of these assays is that the reusability of simulation models needs to be ensured, before computational models can be considered for predictive processes in the clinic. Four

[1] https://www.genomicsengland.co.uk/the-100000-genomes-project/

[2] http://www.openehr.org

important aspects that determine reusability are discussed in the following subsections.

3.1 Semantic annotation to biomedical ontologies

An essential step to ensure reusability of models is a thorough semantic annotation to biomedical ontologies. An ontology formally defines concepts and relations between concepts in a knowledge domain [27]. Semantic annotation describes the process of linking the entities and processes of a model to terms in relevant ontologies. This "upgrade" from a syntactical description of model entities to a semantic description is an established procedure in systems biology. It allows researchers and tools alike to describe the data used in experimental studies and models. A semantic description enables not only integration of different types of data but also reasoning over the data, thus connecting data items (or models) to existing knowledge. Systems biology established a system for semantic annotations of models, using RDF together with a set of standardised relationships [20] and resources identifiers [28]. Recently, composite annotations have been proposed as a means to provide exact descriptions of the model entities [29].

In order to implement models in the clinic, the systems biology data must be linked to biomedical data, biomedical measurements and personalised patient data. An integration on the syntactical level is not expressive enough to allow for automatisation, but integration on the semantic level holds the promise of overcoming this limitation. Figure 1 illustrates the necessary steps for the semantic integration of patient data, computational models, and external data for the benefit of patients and clinical staff.

Integration on the semantic level requires high-quality, reliable mappings of ontologies across and within domains. Ongoing efforts such as 100,000 Genome project conducted by Genomics England aim to represent any clinical data gathered from patients with rare diseases or cancer, not only for treatment but also for research purposes. However, while great care was taken to choose established ontologies, such as Human Phenotype Ontology (HPO) [30], the choice of ontologies is specific to the aims of the project and may not be replicated by other data collections. Another ongoing effort is the openEHR initiative that also focuses on providing access to clinical data by rendering it in computable formats using archetypes, templates and bindings to ontologies.

Many biomedical ontologies are maintained in online portals, such as BioPortal or the Open Biomedical Ontologies (OBO) Foundry web page, which provide search interfaces, web services, version control, and mappings between ontologies [31, 32]. However, different ontologies are used for a semantic representation due to differences in the medical systems used in different countries. This requires reliable mappings between these ontologies. One effort addressing the mapping between terminologies and ontologies is the Unified Medical Language Systems (UMLS) [33], which to date harmonises over 150 terminologies and ontologies[3]. For example, the Human Phenotype Ontology [34], the International Classification of Diseases [4] and SNOMED CT [35] are all integrated in UMLS. So whenever data is represented with a semantic annotation from one of these ontologies, it can be easily transferred to the others. It is important to be aware that the transfer to other ontologies is restricted by the quality of the mapping, in this case UMLS.

Another set of ontologies to consider for this endeavour are those encoding the versioning of models and the provenance of data encoded in the model. For example, PROV-O [36] is an ontology of provenance terms that has been used in a range of applications and thus can be expected to be easily adaptable to attach provenance to model data. Another effort going into this direction is the Ontology of Biomedical AssociatioN (OBAN), used for provenance information on disease-phenotype associations text mined through EuropePMC[5] [37]. While there are a lot of ongoing efforts and projects, further work is needed to allow for the integration of computational models with a variety of independent data resources for the purpose of clinical applications. Furthermore, works in the direction of mappings and similarity measures for terms within and across bio-ontologies should be taken into account [38], for example, to determine the similarity of data sets thatare annotated to different ontologies.

3.2 Generation of safe simulation modules

Reusability depends on the availability of all model-related data [10]. For studies performed by medical researchers, it is particularly important to provide full documentation of safe parameter ranges or test case scenarios. This requires tailor-made standards for the reporting. The data description must ensure that unique interpretations of simulation modules are possible without a modelling background.

In this context, a simulation module encapsulates a computational model that has been tested, documented, annotated, and certified to meet safety requirements. A module suitable for inclusion into a diagnostic tool needs to provide extensive documentation and safe, standardised software interfaces (e. g. for resetting simulation parameters or accessing and interpreting simulation results; see more details section 3.4). The documentation of a model is clearly defined in a Minimum Information guideline (MIRIAM) [39]. We argue that the documentation of a simulation module for medical research needs to be extended to also cover information on applicable virtual experiments, allowed applications, and conditions under which the data are applicable in simulations. For example, a scientist using the module has to be able to assess which parameters can safely be changed and what are the expected sensitivities. An important component of a module is test data for model evaluation consisting of simulation input and output data allowing to evaluate predictive error, sensitivity and specificity of the module.

Furthermore, a potential user also requires access to the tests with which the parameter ranges and prediction outcomes have been assessed during model development. The documentation released with a simulation module should further detail how simulation results are to be correctly interpreted. This is particularly relevant for the classification of results in terms of quantiles within patient cohorts. In order to verify whether a module is safe for use, information detailing the history, developer(s), input data and test results is strictly necessary. Only if this information is provided one can evaluate if the latest version of a module is safe for application and how the changes made over time have affected the error rates of predictions as well as edge-cases in simulation scenarios.

Systems biology offers tools for model version control (e. g., [40]). However, we note that the potential of model provenance has

[3] `https://www.nlm.nih.gov/pubs/factsheets/umls.html`, accessed 14 June 2016

[4] `http://www.who.int/classifications/icd/en/`

[5] `http://europepmc.org/`

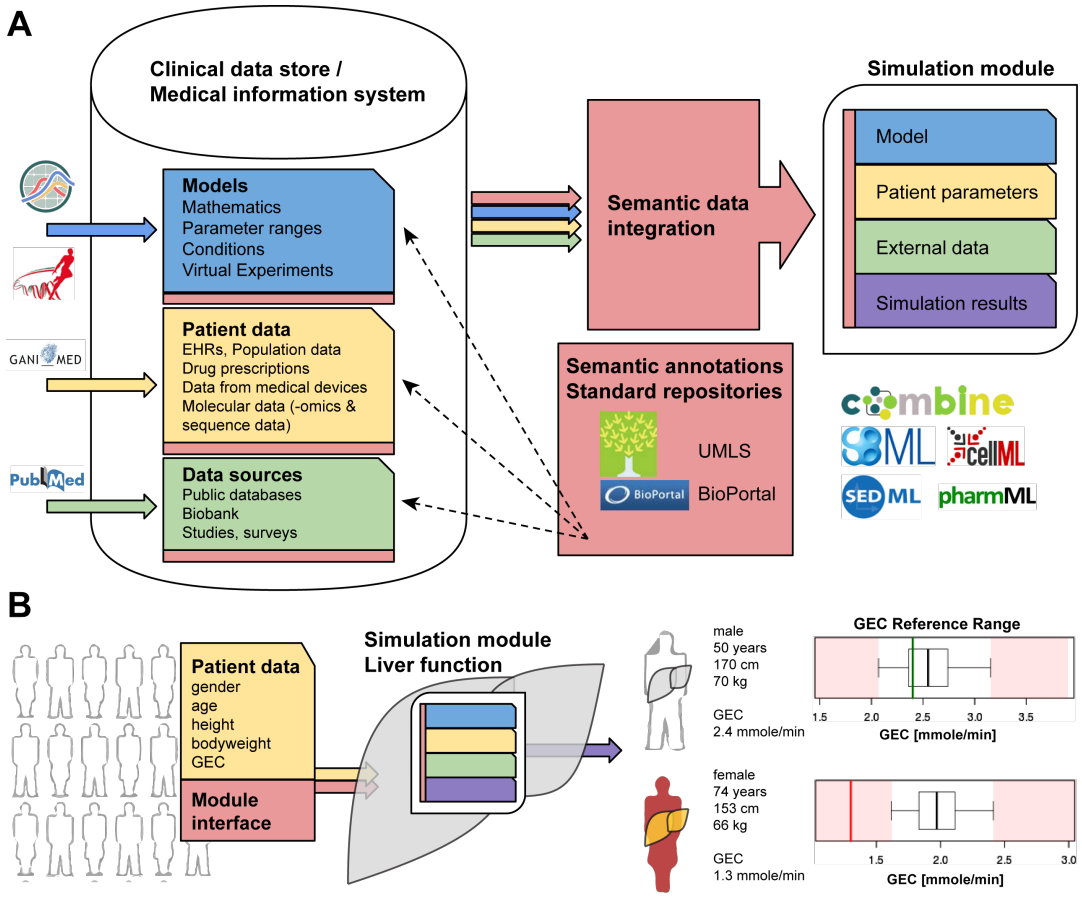

Figure 1. A) Illustration of the integration process of computational models and data from different sources. The integration strongly relies on the availability and detail of the ontologies used for the semantic annotations. User interfaces need to provide access to the simulation modules, but restrict the change of parameters to ranges that are safe w.r.t. a clinical application. Systems Biology Markup Language (SBML) and CellML are standards used to encode models, i.e. the entities and interactions, in a computable format. Electronic Health Records (EHRs) refers to any data recorded in a hospital or GP practice. B) Example workflow for the application of a simulation module. Semantically annotated patient data is used as input to the simulation module based on the defined module interface. The module performs individual predictions and risk estimation based on the input data which can be evaluated within the context of the reference ranges of the module. A proof-of-principle is available at `https://www.livermetabolism.com/gec_app/`.

not yet been fully explored, and the description of model parameters as well as a model's quality (in terms of applicability and reliability) is so far neither satisfactory nor standardised.

3.3 Testing procedures to ensure safety

Due to the sheer amount of data necessary to model the physiology of a human being, the development of future diagnostic tools will rely on previously developed, standardised simulation modules and on thorough semantic annotation. Before models and consequently modules can be consulted in medical predictions they need to be tested thoroughly. This is, in theory, possible for a subset of models in systems biology, such as those contained in the curated branch of BioModels. All curated models have been tested to reproduce the behavior stated in the reference publication.

For a module to be considered safe, the encapsulated model predictions must be medically reliable, i.e. they must not only capture the underlying disease mechanisms but also adapt to the uniqueness of each individual patient. This requirement entails that the error rate for predictions needs to be very small and under no circumstances can exceptions lead to failure in the intermediate computation. Due to the diversity of data that is included into a model, physical units, error ranges and data mappings have to be handled with special care. It is crucial that the patient-specific data to be simulated with the module matches the requirements of model parameters such that a reliable prediction can be ensured.

For this purpose, standardised tests need to be in place and continuously be passed throughout development. The electrophysiology web lab [41] is one example of a web-based tool to check the reliability of models relating to the physiology of the heart. It features a set of published models in CellML format, and applies to them several virtual experiments. The tests check how each model reproduces the expected behavior of a real heart under a variety of conditions.

3.4 Standardised and secure software interfaces

In order to apply modules in clinical practice, standardised software interfaces are required that enable the safe simulation of models (e.g.

through restricted parameter ranges), validation of input parameters, support for allometric scaling (of parameters like organ sizes or blood flow), and the evaluation of simulation results in terms of confidence intervals.

It is not unlikely that a model used through a diagnostic tool is administered by a clinician, nurse or other medical staff. The simulation mode must hence include a safe mode in which only defined properties of the model/module can be adapted. However, these defined properties need to cover, at the same time, the uniqueness of each patient so that the simulation can be truly personalised. An adaptation of the above web lab can help to provide clinicians with an overview of possible behaviors of a system given different sets of patient data and clinical investigations. Moreover, tool and model developers have to safeguard the data that is used as input to the computational model so that patient data cannot be used for other purposes than the treatment of this patient. Otherwise obtaining consent from patients to employ their data for medical purposes will be impossible. There is an arguable potential that the models could be improved over time as the patient data in itself can help tweaking model parameters but this would have to be covered by each patient's consent.

4 CONCLUSION

With kinetic models being increasingly used and reused for the prediction of disease risks, the monitoring of disease progression, or for drug development, the quality and reliability of models becomes a major concern. In this situation, medical research can benefit from the experiences in systems biology, by incorporating existing standards, tools and infrastructure. Standards and standard-compliant tools increase the exchangeability of models, and enable researchers to reproduce published results. As computational models can be readily parameterised with individual patient and cohort data, they are well-suited for personalisation. Moreover, the models can be embedded in pharmacokinetics and pharmacodynamics applications used during drug development.

However, before modeling can be fully incorporated into medical workflows, additional requirements should be met. Among these are further standards to represent the provenance of a model and to document valid parameter ranges under certain conditions. Furthermore, solutions for high-quality annotation of models and for the curation of data need to be developed. Other challenges, like the representation of uncertainties, restricted model changes and personalisation are yet unsolved and have to be addressed in future research. A specific focus of future works should be on the definition of a minimal semantic interface that patient data has to fulfill for a model to be applicable, i. e., a minimal set of semantically encoded data the model requires as input. For instance, in the case of a regression model, all independent variables of the model must exist.

Finally, models used in the clinic need to fulfill safety requirements and adhere to data privacy guidelines. For example, at no point would it be acceptable to mix data from several patients and give a patient or other unauthorised staff access to patients' data.

We conclude that previous work from systems biology research can be reused to establish an infrastructure for reusable models in the clinic. However, the existing infrastructure needs to be evaluated thoroughly, and it needs to be extended to meet clinical standards when working with patient data.

ACKNOWLEDGEMENT

MK is supported by the Federal Ministry of Education and Research (BMBF, Germany) within the research network Systems Medicine of the Liver (LiSyM) (grant number 031L0054). AO and RJBD would like to acknowledge NIHR Biomedical Research Centre for Mental Health, the Biomedical Research Unit for Dementia at the South London, the Maudsley NHS Foundation Trust and Kings College London. RJBD's work is also supported by awards to establish the Farr Institute of Health Informatics Research, London, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1). TJPH would like to acknowledge Kings College London and the NIHR Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and the NIHR Biomedical Research Centre for Mental Health. DW is funded through the BMBF e:Bio program (grant no. 0316194). The authors acknowledge support through CaSyM, the EC FP7 coordinating action *Coordinating Systems Medicine across Europe*.

REFERENCES

[1]Leroy Hood, James R Heath, Michael E Phelps, and Biaoyang Lin. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306(5696):640–643, 2004.

[2]Raimond L Winslow, Natalia Trayanova, Donald Geman, and Michael I Miller. Computational medicine: translating models to clinical care. *Science Translational Medicine*, 4(158):158rv11–158rv11, 2012.

[3]Partha Pratim Chakraborty, Sujoy Ghosh, and Sanjay Kalra. Online risk engines and scoring tools in endocrinology. *Indian Journal of Endocrinology and Metabolism*, 17(Suppl 3):S601, 2013.

[4]M Benson. Clinical implications of omics and systems medicine: focus on predictive and individualized treatment. *Journal of internal medicine*, 2015.

[5]Geoffrey S Ginsburg and Jeanette J McCarthy. Personalized medicine: revolutionizing drug discovery and patient care. *TRENDS in Biotechnology*, 19(12):491–496, 2001.

[6]Olaf Wolkenhauer, Charles Auffray, Robert Jaster, et al. The road from systems biology to systems medicine. *Pediatric research*, 73(4-2):502–507, 2013.

[7]William E Evans and Mary V Relling. Pharmacogenomics: translating functional genomics into rational therapeutics. *science*, 286(5439):487–491, 1999.

[8]K Romero, K Ito, JA Rogers, D Polhamus, et al. The future is now: Model-based clinical trial design for alzheimer's disease. *Clinical Pharmacology & Therapeutics*, 97(3):210–214, 2015.

[9]Jessica Nasica-Labouze, Phuong H Nguyen, Fabio Sterpone, et al. Amyloid β protein and alzheimer's disease: When computer simulations complement experimental studies. *Chemical Reviews*, 115(9):3518–3563, 2015.

[10]Dagmar Waltemath and Olaf Wolkenhauer. How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine. *IEEE Transactions on Biomedical Engineering*, June 2016.

[11]Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[12]Michele Magrane, UniProt Consortium, et al. Uniprot knowledgebase: a hub of integrated protein data. *Database*, 2011:bar009, 2011.

[13]Stephen F Altschul, Warren Gish, Webb Miller, et al. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.

[14]Fiona Cunningham, M Ridwan Amode, Daniel Barrell, et al. Ensembl 2015. *Nucleic Acids Research*, 43(D1):D662–D669, 2015.

[15]Kate R Rosenbloom, Cricket A Sloan, Venkat S Malladi, et al. Encode data in the ucsc genome browser: year 5 update. *Nucleic Acids Research*, 41(D1):D56–D63, 2013.

[16]Tanya Barrett, Stephen E Wilhite, Pierre Ledoux, et al. Ncbi geo: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995, 2013.

[17]Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, et al. String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, page gku1003, 2014.

[18]Taavi Tillmann, Alexander R Gibson, Gregory Scott, et al. Systems medicine 2.0: potential benefits of combining electronic health care records with systems science models. *Journal of Medical Internet Research*, 17(3):e64, 2015.

[19]Guido Santos, Svetoslav Nikolov, Xin Lai, et al. Model-based genotype-phenotype mapping used to investigate gene signatures of immune sensitivity and resistance in melanoma micrometastasis. *Scientific Reports*, 6, 2016.

[20]Chen Li, Marco Donizelli, Nicolas Rodriguez, et al. Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4(1):92, 2010.

[21]Tommy Yu, Catherine M Lloyd, David P Nickerson, et al. The physiome model repository 2. *Bioinformatics*, 27(5):743–744, 2011.

[22]Falk Schreiber, Gary D Bader, Martin Golebiewski, et al. Specifications of standards in systems and synthetic biology. *J. Int. Bioinformatics*, 12(258.10):2390, 2015.

[23]Jonathan Cooper, Jon Olav Vik, and Dagmar Waltemath. A call for virtual experiments: accelerating the scientific process. *Progress in biophysics and molecular biology*, 117(1):99–106, 2015.

[24]Dagmar Waltemath, Richard Adams, Frank T Bergmann, et al. Reproducible computational biology experiments with sed-ml-the simulation experiment description markup language. *BMC systems biology*, 5(1):1, 2011.

[25]Leonard P Freedman, Iain M Cockburn, and Timothy S Simcoe. The economics of reproducibility in preclinical research. *PLOS Biology*, 13(6):e1002165, 2015.

[26]Shareen A Iqbal, Joshua D Wallach, Muin J Khoury, et al. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLOS Biology*, 14(1):e1002333, 2016.

[27]Victoria Uren, Philipp Cimiano, José Iria, et al. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: science, services and agents on the World Wide Web*, 4(1):14–28, 2006.

[28]Nick Juty, Nicolas Le Novère, and Camille Laibe. Identifiers. org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Research*, 40(D1):D580–D586, 2012.

[29]John H Gennari, Maxwell L Neal, Michal Galdzicki, and Daniel L Cook. Multiple ontologies in action: composite annotations for biosimulation models. *Journal of Biomedical Informatics*, 44(1):146–154, 2011.

[30]Sebastian Köhler, Sandra C Doelken, Christopher J Mungall, et al. The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research*, page gkt1026, 2013.

[31]Manuel Salvadores, Paul R Alexander, Mark A Musen, and Natalya F Noy. BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semantic Web*, 4(3):277–284, 2013.

[32]Barry Smith, Michael Ashburner, Cornelius Rosse, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, 2007.

[33]Olivier Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(suppl 1):D267–D270, 2004.

[34]Sebastian Köhler, Sandra C Doelken, Christopher J Mungall, et al. The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Research*, 42(D1):D966–D974, 2014.

[35]Kevin Donnelly. SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in Health Technology and Informatics*, 121:279, 2006.

[36]Timothy Lebo, Satya Sahoo, Deborah McGuinness, et al. Prov-o: The prov ontology. *W3C Recommendation*, 30, 2013.

[37]Sirarat Sarntivijai, Drashtti Vasant, Gary Saunders, et al. Linking rare and common disease: mapping clinical disease-phenotypes to ontologies in therapeutic target validation. *Journal of Biomedical Semantics*, 7(8), 2016.

[38]Michael Hartung, Anika Gross, Toralf Kirsten, and Erhard Rahm. Effective composition of mappings for matching biomedical ontologies. In *Extended Semantic Web Conference*, pages 176–190. Springer, 2012.

[39]Nicolas Le Novère, Andrew Finney, Michael Hucka, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology*, 23(12):1509–1515, 2005.

[40]Martin Scharm, Olaf Wolkenhauer, and Dagmar Waltemath. An algorithm to detect and communicate the differences in computational models describing biological systems. *Bioinformatics*, page btv484, 2015.

[41]Jonathan Cooper, Martin Scharm, and Gary R Mirams. The Cardiac Electrophysiology Web Lab. *Biophysical Journal*, 110(2):292–300, 2016.