

Contributions to the reuse and reproducibility of computational biology models

Ron Henkel^{1,*}, Dagmar Waltemath¹

¹University Medicine Greifswald, Departement Medical Informatics, Institute for Community Medicine, Walther-Rathenau-Straße 48, 17475 Greifswald, Germany

Abstract

Research that can not be reproduced is of no scientific value - this also holds true for computational models in systems biology. This poster elucidates the identification of challenges and suggested solutions for model retrieval and ranking, model version control and model storage. Furthermore, it describes how models should be managed, stored and retrieved within a knowledge graph in order to foster model reproducibility and reuse.

Keywords

Graph Database, Systems Biology, Computational Models, Knowledge Graph

The ability to reproduce research is at the heart of science. Results can only be scientifically relevant if it can be reproduced [1], providing that experimental settings are coherently reported. This also holds true for computational research, such as models developed to gain deeper insights into biological systems as it is done in the life sciences and systems biology in particular [2].

With the rapidly increasing number of available models and new ones being developed [3], crucial tasks for any researcher in this field are to find models of relevance, to keep track of changes in a model of interest, and to ultimately run the model in order to reproduce scientific findings from in a publication [4]. However, without a sophisticated storage and organisation of models, those tasks are in the best case tedious, thus rendering model reuse and reproducibility of modeling results nearly impossible.

This poster describes how to improve the reuse and reproducibility of computational biology models by suggesting and implementing innovative and efficient means for model storage [5], model retrieval [6] and comparison [7, 8], and model provenance [9]. The model management concepts developed and implemented, provide a graph-based model storage and a sophisticated model retrieval framework, two crucial steps towards a FAIR model management [10]. In addition, as the storage and retrieval concept also include model meta-information from a variety of domains, the thereby created knowledge graph is able to provide results for domain-spanning queries. Having such a knowledge graph at hand offers a variety of new research

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

*Corresponding author.

✉ ron.henkel@uni-greifswald.de (R. Henkel); dagmar.waltemath@uni-greifswald.de (D. Waltemath)

🌐 <https://www.mezizin.uni-greifswald.de/medizininformatik/> (R. Henkel);

<https://www.mezizin.uni-greifswald.de/medizininformatik/> (D. Waltemath)

🆔 0000-0001-6211-2719 (R. Henkel); 0000-0002-5886-5563 (D. Waltemath)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

possibilities, e.g., for model analysis [11] and data exploration [12].

With a scientifically broader view, reusing concepts and implementations developed in context of systems biology in medical sciences seems to be the next logical step. Such a transition will connect different domains such as systems biology, systems medicine and medical informatics and might serve as a way to bridge the gaps between those scientific disciplines [13].

References

- [1] O. Wolkenhauer, Why model?, *Frontiers in physiology* 5 (2014) 21.
- [2] R. A. McDougal, A. S. Bulanova, W. W. Lytton, Reproducibility in computational neuroscience models and simulations, *IEEE Transactions on Biomedical Engineering* 63 (2016) 2021–2035.
- [3] R. S. Malik-Sheriff, M. Glont, T. V. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan, et al., Biomodels—15 years of sharing computational models in life science, *Nucleic acids research* 48 (2020) D407–D415.
- [4] D. Waltemath, O. Wolkenhauer, How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine, *IEEE Transactions on Biomedical Engineering* 63 (2016) 1999–2006.
- [5] R. Henkel, O. Wolkenhauer, D. Waltemath, Combining computational models, semantic annotations and simulation experiments in a graph database, *Database* 2015 (2015) bau130.
- [6] R. Henkel, L. Endler, A. Peters, N. Le Novère, D. Waltemath, Ranked retrieval of computational biology models, *BMC bioinformatics* 11 (2010) 423.
- [7] R. Henkel, R. Hoehndorf, T. Kacprowski, C. Knüpfer, W. Liebermeister, D. Waltemath, Notions of similarity for systems biology models, *Briefings in Bioinformatics* (2016) bbw090.
- [8] M. Scharm, O. Wolkenhauer, D. Waltemath, An algorithm to detect and communicate the differences in computational models describing biological systems, *Bioinformatics* 32 (2015) 563–570.
- [9] D. Waltemath, R. Henkel, R. Hälke, M. Scharm, O. Wolkenhauer, Improving the reuse of computational models through version control, *Bioinformatics* 29 (2013) 742–748.
- [10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, *Scientific data* 3 (2016).
- [11] R. Alm, D. Waltemath, M. Wolfien, O. Wolkenhauer, R. Henkel, Annotation-based feature extraction from sets of sbml models, *Journal of biomedical semantics* 6 (2015) 20.
- [12] F. Lambusch, D. Waltemath, O. Wolkenhauer, K. Sandkuhl, C. Rosenke, R. Henkel, Identifying frequent patterns in biochemical reaction networks: a workflow, *Database* 2018 (2018).
- [13] L. Gütebier, T. Bleimehl, R. Henkel, J. Munro, S. Müller, A. Morgner, J. Laenge, A. Pachauer, A. Erdl, J. Weimar, K. Walther Langendorf, V. Vialard, T. Liebig, M. Preusse, D. Waltemath, A. Jarasch, CovidGraph: a graph to fight COVID-19, *Bioinformatics* 38 (2022) 4843–4845. doi:10.1093/bioinformatics/btac592.