

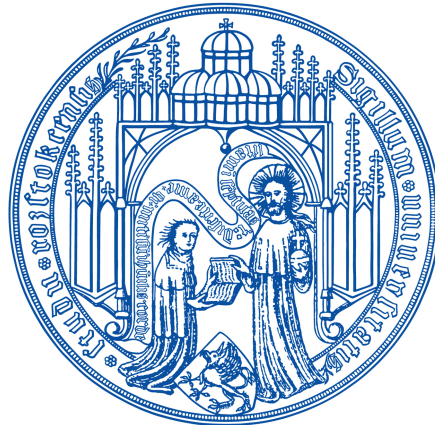
PhD Thesis

Contributions to the reuse and reproducibility of computational biology models

by

Dipl.-Inf. Ron Henkel

Dissertation zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.)



Universität Rostock
Fakultät für Informatik und Elektrotechnik
Institut für Informatik

Submitted by: Dipl.-Inf. Ron Henkel
Date of Birth: 1982-09-19 in Waren (Müritz)

Rostock, November 18, 2022

Dissertation Summary

This thesis addresses reproducibility and reusability of systems biology models and elaborates concepts of model management to support a FAIR (findable, accessible, interoperable, reusable) handling of systems biology models.

To this end, this thesis focuses on six key research questions:

1. How to establish a baseline for model management?
2. How can heterogeneous, semi-structured model data be retrieved and ranked?
3. How to trace a model over its lifetime?
4. How to store a model and its accompanying meta-data?
5. What model parts are important to elaborate on model similarity?
6. What becomes possible once the aforementioned research questions are answered?

To answer these questions, this thesis describes the development of concepts and implementations for model storage, model search, model version control, and model similarity analysis. These developments are put in perspective regarding the state of the art at the time of development and regarding the current state of the art, respectively. In summary, this thesis provides an overview of the means and measures of model management and describes in its individual chapters solutions for various aspects of a FAIR model management. Finally, the approaches presented here are taken up and extended in the context of the current COVID-19 research.

Zusammenfassung der Dissertation

Die hier vorliegende Arbeit beschäftigt sich mit der Reproduzierbarkeit und Wiederverwendbarkeit systembiologischer Modelle. Die Arbeit behandelt Konzepte des Modellmanagements um eine FAIRe (auffindbar, zugreifbar, interoperabel, wiederverwendbar) Handhabung von systembiologischen Modellen zu unterstützen.

Im Fokus dieser Arbeit stehen sechs Forschungsfragen:

1. Welche Maßnahmen sind nötig um eine Grundlage für Modellmanagement zu etablieren?
2. Wie können heterogene, semi-strukturierte Modelle mittels Suchanfrage gefunden und bewertet werden?
3. Wie kann die Weiterentwicklung eines Modells über die Zeit verfolgt werden?
4. Wie können, neben dem Modell, auch die zugehörigen Metadaten gespeichert werden?
5. Welche Modellkomponenten sind wichtig um eine Aussage über Modellähnlichkeit treffen zu können?
6. Welche weiteren Optionen werden verfügbar, wenn Konzepte und Technologien zur Beantwortung der vorherigen Fragen existieren?

Zur Beantwortung dieser Fragen werden in dieser Arbeit Konzepte und Umsetzungen zur Modellspeicherung, Modellsuche, Modellversionierung und zur Bewertung von Modellähnlichkeiten entwickelt. Diese Entwicklungen werden in den Wissensstand zur Entwicklungszeit und vergleichend zum aktuellen Wissensstand eingeordnet. Zusammenfassend vermittelt diese Arbeit somit ein Bild des aktuellen Standes des Modellmanagements und beschreibt in ihren einzelnen Kapiteln Lösungsmöglichkeiten für verschiedene Aspekte eines FAIRen Modellmanagements. Abschließend werden die hier vorgestellten Ansätze aufgegriffen und im Kontext der aktuellen COVID-19 Forschung erweitert.