

## An Interdisciplinary Perspective on Managing Process Model Simulation Data

Clemens Schreiber<sup>1</sup>, Gunther Schiefer<sup>1</sup>, and Andreas Oberweis<sup>1</sup>

**Abstract:** Managing process model simulation data in a findable, accessible, interoperable, and reusable way faces several challenges. We discuss three main challenges of simulation data documentation, which we encountered in an interdisciplinary research project on research data management. These challenges refer to (1) the definition of a discipline-specific metadata model, (2) the development of user support to ensure metadata quality and (3) the interlinking of the simulation data across research areas. Based on the identified challenges and the reported insights gained from the research project, we provide novel ideas for the development of process model simulation tools, enabling a better documentation of simulation data in the future.

**Keywords:** Research Data Management, Process Model Simulation

### Introduction

Process model simulation is an important approach for the analysis of processes and information systems [Au18; De97; KSR23; PV21]. It is for example used to analyse the validity [De97] and the performance [KSR23; PV21] of a modelled system, or to compare and evaluate process mining algorithms [Au18]. To share simulation data with the research community, it should be documented in a systematic manner to ensure that it is findable, accessible, interoperable, and reusable, in accordance with the FAIR-principles promoted by the NFDI ([nfdi.de/fair-data-spaces](http://nfdi.de/fair-data-spaces)). However, handling process model simulation data in a FAIR way faces some challenges. In this extended abstract we want to point out three main challenges in handling simulation data, which we observed during the interdisciplinary research project MoMaF (Science Data Center for Molecular Materials Research) [A123]. The goal of MoMaF is to facilitate the research data management across different research disciplines, with a focus on experimental and simulation data. This is mainly achieved by the development of research data management software, which can provide discipline-specific user support (for an overview see [A123]). The encountered challenges during the development are of generic nature and can be summarized as follows: (1) Defining a discipline-specific metadata model, (2) ensuring metadata quality, and (3) interlinking research data across different research areas. In the following sections we will describe these generic challenges in more detail and discuss their relevance for the management of process model simulation data. Our overall goal is to provide a road-map for the development of research data management tools, which can facilitate the management of process model simulation data in the future.

---

<sup>1</sup> Karlsruher Institut für Technologie, AIFB., Karlsruhe, Germany, [Clemens.Schreiber@kit.edu](mailto:Clemens.Schreiber@kit.edu);  
[Gunther.Schiefer@kit.edu](mailto:Gunther.Schiefer@kit.edu); [Andreas.Oberweis@kit.edu](mailto:Andreas.Oberweis@kit.edu)

This work is licensed under Creative Commons Attribution 4.0 International License <http://creativecommons.org/licenses/by/4.0/>, <https://doi.org/10.18420/modellierung2024-ws-011>

### **Challenge 1: Defining a discipline-specific metadata model**

The structured documentation of simulation data requires the definition of a metadata model, which supports all FAIR principles. The documentation should thereby contain generic as well as discipline-specific metadata [SI20]. However, the variety of existing metadata standards was a significant challenge in the MoMaF project, since there do not exist commonly accepted metadata standards for the involved research disciplines. To identify the researcher's requirements for the documentation of research data, a workshop with researchers from different disciplines was conducted [Ig21]. Additionally, the selection of standards was influenced by the requirements of public research data repositories, which might require a specific metadata standard (an overview of repository requirements can be found at [re3data.org](https://re3data.org)).

On the website [fairsharing.org](https://fairsharing.org) there does not yet exist a metadata standard for process model simulation data. However, the EngMeta metadata model proposed in [SI20] provides a good starting point. The model is designed to document simulation data in the field of computation engineering involving big computing systems and consists of four metadata types: descriptive, process, technical, and subject specific metadata. While this model covers well the documentation of the simulation output, it does not cover all relevant aspects of the simulated process model. This is partially addressed in [SW22], where some suggestions are made to specify the metadata for discrete event system simulation. However, especially information regarding the applied modeling language and the model itself is still missing.

### **Challenge 2: Ensuring metadata quality**

Once a set of required metadata is defined, the next step is to ensure the metadata quality. Metadata quality has many different dimensions, such as completeness, accuracy, and provenance. One way to improve the metadata quality is to (partially) automate the documentation work [SI20; SW22]. An important goal of the MoMaF project was to integrate such automation mechanisms in existing electronic laboratory notebooks (ELN) for the documentation of laboratory experiments ([chemotion.net](https://chemotion.net)), and simulations of phase separation processes ([kadi4mat.iam.kit.edu](https://kadi4mat.iam.kit.edu)). The ELNs enable better guidance for researchers and to continuously track metadata during the research process. They also allow to restrict the vocabulary for certain metadata, based on discipline specific ontologies. Additionally, we applied generative AI, i.e., large language models, to automatically generate textual descriptions of the research data.

From a process model simulation perspective there already exist a variety of simulation tools, such as CPN ([cpntools.org](https://cpntools.org)) and ProM ([promtools.org](https://promtools.org)), but their support for the documentation of the simulation runs is rather limited. Moreover, there does not exist any support based on ontologies or generative AI.

### **Challenge 3: Interlinking research data across research areas**

Finally, the metadata should allow for an integrated and synchronized view on research data, across different research areas [AI23]. This is particularly challenging, since different research areas have different requirements regarding research data documentation. At MoMaF we worked together with research groups coming from two different areas: organic

chemistry and applied material sciences. While the research methods used in both groups differ (laboratory experiments vs. computational simulations), they are both interested in the properties of materials on a molecular as well as on a macroscopic scale. To support a better exchange between these areas, interfaces were defined to allow for an exchange of data between the two ELNs (for an overview see [AI23]).

The research on process model simulation is also often connected to other research areas, e.g., when it comes to the development of digital twins for a specific domain [PV21]. Additionally, process model simulation is only an initial step in information system development. Further steps involve system implementation and testing, as well as instantiation and observation. Data in these stages is typically considered separately and stored in different systems. This makes it difficult to get a coherent view on the relevant data over the whole system life cycle and to define interfaces between usage in different stages. For this reason, the data for running the processes and the data about the running processes together with the execution environment and the context information for all life cycle stages should be stored in a uniform schema within a suitable research data infrastructure.

## Conclusion

Based on our observations from MoMaF (an interdisciplinary research project on scientific data management) we identify three main important challenges for managing research data, which also appear to be relevant for the management of process model simulation data. These challenges are (1) the definition of a discipline-specific metadata model, (2) the development of user support to ensure metadata quality and (3) the interlinking of simulation data across different research areas. Although there already exist partial solutions for the identified issues (e.g., based on the automated generation of metadata for simulations as proposed in [SI20] and [SW22]), specific solutions for process model simulation are yet missing. In this regard we provide novel ideas for the development of simulation tools, allowing for a better documentation of process model simulation data in the future.

## References

- [AI23] Altschuh, P.; Bräse, S.; Hartmann, T.; Jaeger, D.; Jung, N.; Koepe, A.; Nestler, B.; Schiefer, G.; Schreiber, C.; et al.: An interdisciplinary approach to data management. In: E-Science-Tage 2023, Heidelberg. 2023.
- [Au18] Augusto, A.; Conforti, R.; Dumas, M.; La Rosa, M.; et al.: Automated discovery of process models from event logs: Review and benchmark. *IEEE Transactions on Knowledge and Data Engineering* 31 (4), 2018.
- [De97] Desel, J.; Oberweis, A.; Zimmer, T.; Zimmermann, G.: Validation of information system models: Petri nets and test case generation. In: *Proc. IEEE International Conference on Systems, Man, and Cybernetics*. Vol. 4, IEEE, 1997.
- [Ig21] Iglezakis, D.; Fuhrmans, M.; Arndt, S.; Demandt, E.; et al.: Interoperabilität von Metadaten innerhalb der NFDI: Konsortienübergreifender Metadaten-Workshop am 2./3. Juli 2020. *Bausteine Forschungsdatenmanagement* (2), 2021.
- [KSR23] Kinkelin, N.; Schreiber, C.; Reimold, J.: Combining process mining and simulation to guide users towards process improvement in mpmX. In: *Proc. of CEUR Workshop on ICPM Demo Track - FORTHCOMING*. 2023.

- [PV21] Park, G.; Van Der Aalst, W.M.: Realizing a digital twin of an organization using action-oriented process mining. In: Proc. ICPM. IEEE, 2021.
- [SI20] Schembera, B.; Iglezakis, D.: EngMeta: Metadata for Computational Engineering. Int. Journal of Metadata, Semantics and Ontologies 14 (1), 2020.
- [SW22] St-Aubin, B.; Wainer, G.: Metadata Specification for DEVS simulation models. In: Proc. IEEE/ACM 26th DS-RT. 2022.