1st Conference on Research Data Infrastructure
Harmonizing RDM
https://doi.org/10.52825/CoRDI.v1i.255

Published: 07 Sept. 2023

Building Ontologies and Knowledge Graphs for Mathematics and its Applications

Björn Schembera[1] [https://orcid.org/0000-0003-2860-6621], Frank Wübbeling[2] [https://orcid.org/0000-0002-2375-2008], Thomas Koprucki[3] [https://orcid.org/0000-0001-6235-9412], Christine Biedinger[4] [https://orcid.org/0009-0002-5082-8386], Marco Reidelbach[5] [https://orcid.org/0000-0002-1919-1834], Burkhard Schmidt[3] [https://orcid.org/0000-0002-9658-499X], Dominik Göddeke[1] [https://orcid.org/0000-0002-1552-497X], and Jochen Fiedler[4] [https://orcid.org/0000-0002-9176-780X]

[1]Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Stuttgart, Germany

[2]Institute of Applied Mathematics: Analysis and Numerics, University of Münster, Münster, Germany

[3]Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

[4]Fraunhofer Institute for Industrial Mathematics, Kaiserslautern, Germany

[5]Mathematics of Complex Systems, Zuse Institute Berlin, Berlin, Germany

Abstract: Ontologies and knowledge graphs for mathematical algorithms and models are presented, that have been developed by the Mathematical Research Data Initiative. This enables FAIR data handling in mathematics and the applied disciplines. Moreover, challenges of harmonization during the ontology development are discussed.

Keywords: Research Data Management, Mathematics, Ontologies, Knowledge Graphs

1 Introduction

Mathematical research data is vast, complex and multifaceted, and its typical appearances are formulae, models, data (numerical, symbolic, tabular), software and documents [1]. It emerges within the mathematical core sciences but also in the applied sciences, such as engineering, physics or even digital humanities. Given this, the Mathematical Research Data Initiative (MaRDI) is being established as the NFDI consortium of mathematics [2], [3]. Its mission is to develop a robust research data infrastructure for mathematics and beyond. It is here that MaRDI is setting up semantic technology (metadata, ontologies, knowledge graphs) and data infrastructures (portals, repositories) [4] for relevant information assets such as algorithms, models, problems, software or data to foster compliance with the FAIR principles [5].

MaRDI aims to support researchers solving real world problems by translating them into mathematical ones, applying adequate algorithms and transferring back the results (see figure 1). Here, several questions arise, e.g. existence of a mathematical model, availability of solving algorithms, input data or model validity. A non-negligible amount of time is required for finding existing models, algorithms and solvers.

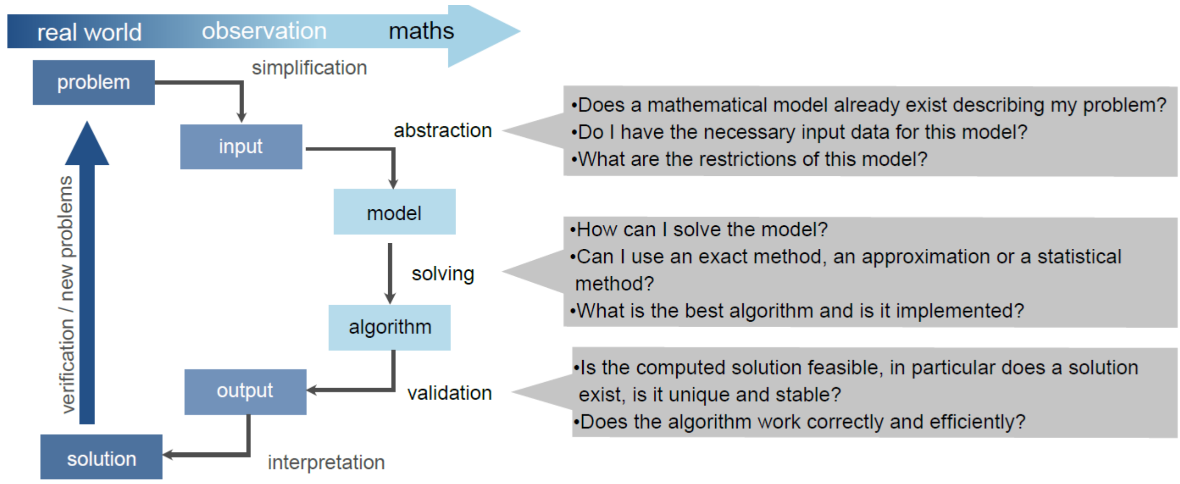

Figure 1. Typical modeling-simulation-optimization workflow considered in MaRDI and resulting competency questions [6].

2 Knowledge Graphs in Mathematics

Knowledge graphs are a well-established semantic technology for representing knowledge in a graph-structured data model, i.e. an ontology. Generally, they represent relations between objects as semantic triples, e.g. *(algorithm, solves, computable model)*, where *solves* is the relation describing that a specific algorithm acts as a solver for a computable model. Taking line planning in public transport as an example (cf. LinTim [7]–[9]), different algorithms for optimization objectives like costs or passenger satisfaction are applicable. The integer program minimizing costs (the computable model), can then be solved by a branch-and-bound algorithm.

In the following, we present two ontologies for algorithms and models. These are linked to additional ontologies, e.g. for benchmarking and statistics.

2.1 A Knowledge Graph for Algorithms

Algorithms are a basic building block of applied mathematics. Algorithms solve problems, are implemented in software, and tested by benchmarks. Additionally, they are documented, invented, etc. in Publications.

The curated AlgoData knowledge graph [10] seeks to formalize the relations between the objects. It allows to answer questions of the form

- Which algorithms solve my problem?
- Which implementations are available? Is there a common analysis?
- Which benchmarks should I use for my problem?

The ontology has been kept as minimal as possible, with object classes algorithm, problem, software, benchmark and publication, and a total of 16 relations (see figure 2).

Initial graphs have been created in Control Configuration Selection, Model Order Reduction and Linear Algebra, for a total of 150 algorithms, documented in 825 publications. The current state is available on the AlgoData site [11]. The interface includes documentation of the ontology, keyword search, and a guided query tailored to the ontology. Technically, the frontend is based on a Apache Jena Fuseki with OWL reasoner and SPARQL interface to a Django server. A SPARQL endpoint can be provided.

Currently, the fundamental algorithms, problems etc. are provided by experts in the field. We envision editorial boards for each subdiscipline that keep this data up to date by accepting (or denying) proposals, made through the MaRDI portal. We have given major attention to ensuring that the processes for editors and users cause as little work as possible. A preview of the editorial process is available on the beta web site [12].

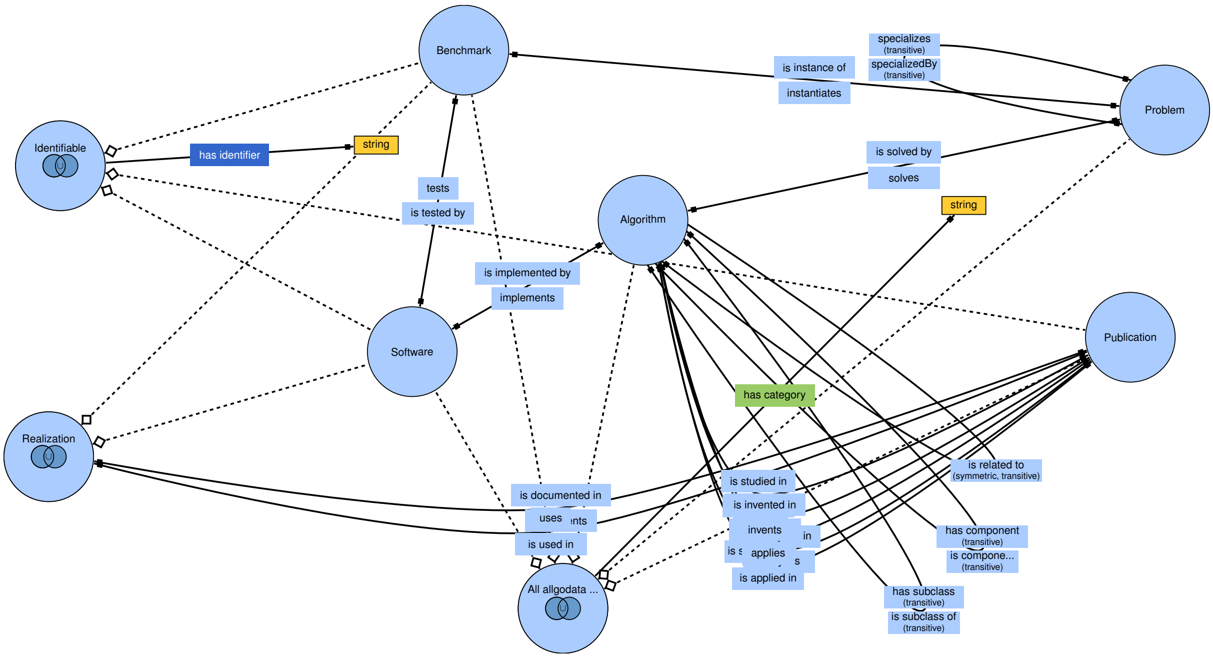

Figure 2. AlgoData ontology, visualized in VOWL

2.2 A Knowledge Graph for Models

Models are at the intersection of mathematics and multitude of other disciplines, serving as vital tools for understanding, predicting, and analyzing complex phenomena across various domains. Their importance stems from their ability to simplify real-world systems and represent them using mathematical equations which makes them computable. Because of their universal nature, models within different domains share common properties, such that developing or refining a model in one field can often lead to advancements and insights in others. By identifying these shared properties, researchers can bridge gaps between disciplines, fostering interdisciplinary collaboration and promoting the growth of knowledge in multiple areas simultaneously. Within MaRDI, the development of an ontology for mathematical models is advanced along case studies from manifold fields of the applied sciences, such as engineering or materials science. Through analysis of the case studies and their according workflows, it is possible to carve out relations, interdependencies and details of the underlying mathematical models. These are displayed on the MaRDI portal as a Wiki [13]. As of April 2023, the preliminary MaRDI model ontology includes the 8 classes (*model*, *computable model*, *application problem*, *application domain*, *equation*, *law*, *quantity*, *term*) and is shown in figure 3. The ontology can be seen as an extension of Model Pathway Diagrams [14], which are based on quantities (represented by terms) which are connected by laws (represented by equations). The ontology has been developed along the case studies, and first models, such as Navier-Stokes or diffusion from the mechanics domain, or Bayes from the stochastics domain have been integrated for testing purposes.

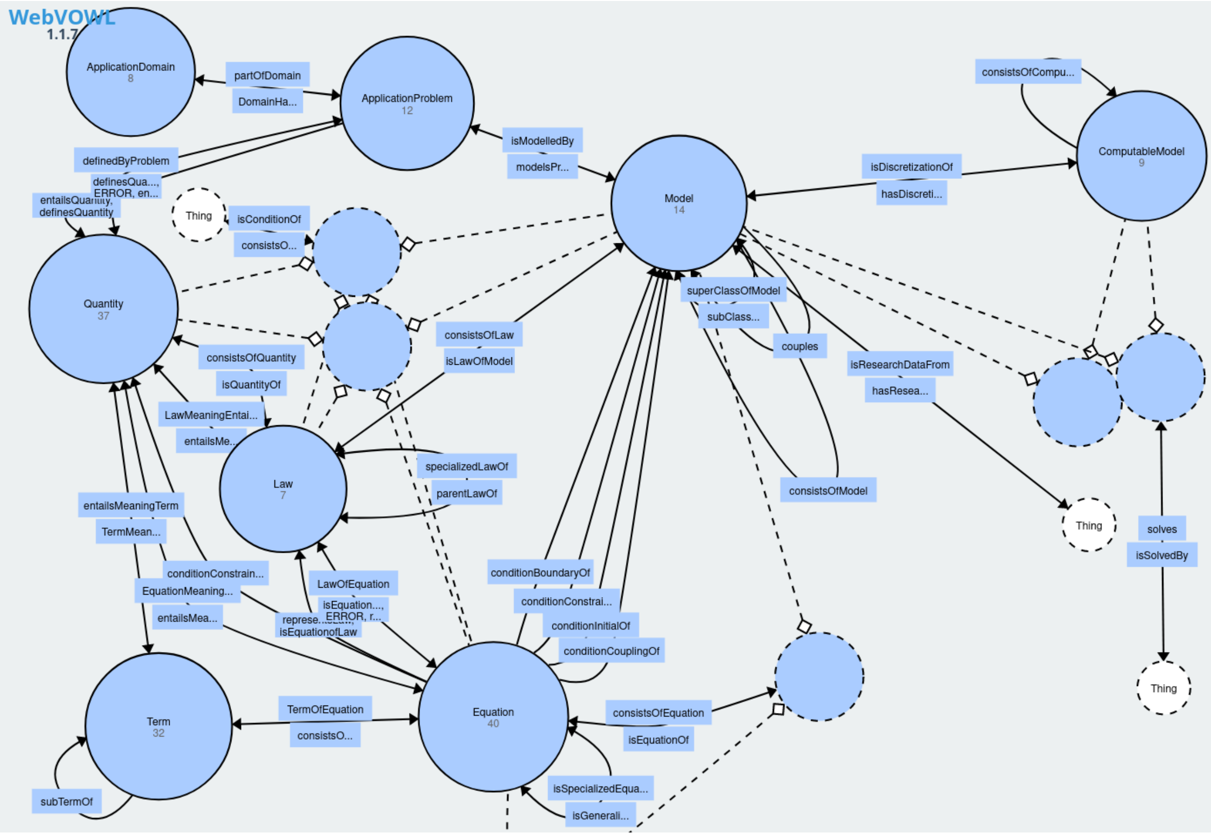

Figure 3. MaRDI Model Ontology classes and their relations. Visualisation done with WebVOWL.

3 Conclusion

In many areas of science, novel findings are acquired by processing and analyzing data. In general, these steps can be regarded as data transformations. Our mission in MaRDI is to structure and classify the possible mathematical data transformations and make them findable and accessible within mathematics and beyond.

To achieve this, we have presented ontologies and knowledge graphs for mathematical algorithms and models. In the future we will also integrate statistical algorithms and machine learning models. However, harmonization is one of the biggest challenges, not only between, but also within disciplines. As an example, the notion of a seemingly simple term like "problems" is controversial: in the models ontology, it represents the application problem, whereas in AlgoData, it stands for the pure mathematical problem. This could be solved by more precise labels and definitions. We see that especially across disciplines, harmonization requires ample coordination efforts, which can be provided, for example, through the NFDI.

Declarations

Data availability statement

The data in the knowledge graph for algorithms can be accessed via algodata.mardi4nfdi.de/. The data in the knowledge graph for models is as of now just for testing and will be published later.

Competing interests

The authors declare that they have no competing interests.

Funding

All authors are supported by MaRDI, funded by the Deutsche Forschungsgemeinschaft (DFG), project number 460135501, NFDI 29/1 "MaRDI – Mathematische Forschungsdateninitiative".

References

[1] T. Boege, R. Fritze, C. Görgen, *et al.*, "Data Management Planning in the German Mathematical Community," *arXiv preprint arXiv:2211.12071*, 2022.

[2] C. Görgen and R. Sinn, "Mathematik in der Nationalen Forschungsdateninfrastruktur," *Mitteilungen der Deutschen Mathematiker-Vereinigung*, vol. 29, no. 3, pp. 122–123, 2021.

[3] P. Benner, M. Burger, D. Göddeke, *et al.*, "Die mathematische Forschungsdateninitiative in der NFDI: MaRDI (Mathematical Research Data Initiative)," *GAMM Rundbrief*, no. 1, pp. 40–43, 2022.

[4] M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, and B. Schembera, *Data Technology in Materials Modelling*. Springer Nature, 2021.

[5] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[6] The MaRDI consortium, *MaRDI: Mathematical Research Data Initiative Proposal*, May 2022. DOI: 10.5281/zenodo.6552436. [Online]. Available: https://doi.org/10.5281/zenodo.6552436.

[7] A. Schiewe, S. Albert, P. Schiewe, A. Schöbel, F. Spühler, and M. Stinzendörfer, *Documentation for LinTim 2022.08*, coursematerial, 2022. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-69236.

[8] A. Schiewe, S. Albert, V. Grafe, P. Schiewe, A. Schöbel, and F. Spühler, *LinTim - integrated optimization in public transportation.* [Online]. Available: https://www.lintim.net.

[9] M. Goerigk, S. Michael, and A. Schöbel, "Evaluating line concepts using travel times and robustness: Simulations with the LinTim toolbox," *Public Transport*, vol. 5, pp. 267–284, 2013. DOI: https://doi.org/10.1007/s12469-013-0072-x.

[10] C. Himpe, H. Kleikamp, R. Fritze, and S. Rave, *MaRDI Task Area 2 - Scientific Computing @ WWU Münster. AlgoData - Algorithm Knowledge Graph - Ontology (Version 0.1)*, 2022. [Online]. Available: https://mardi4nfdi.de/algodata/0.1.

[11] The MaRDI consortium, *AlgoData*, https://algodata.mardi4nfdi.de/, [Online; accessed 24-April-2023], 2023.

[12] The MaRDI consortium, *AlgoData Editorial Access Beta*, https://beta.m1.mardi.ovh/, [Online; accessed 24-April-2023], 2023.

[13] The MaRDI consortium, *MaRDI TA4 Portal*, https://portal.mardi4nfdi.de/wiki/Portal/TA4, [Online; accessed 24-April-2023], 2023.

[14] T. Koprucki, M. Kohlhase, K. Tabelow, D. Müller, and F. Rabe, "Model pathway diagrams for the representation of mathematical models," *Optical and Quantum Electronics*, vol. 50, pp. 1–9, 2018.