# Weierstraß-Institut
## für Angewandte Analysis und Stochastik
### Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855


# Mathematical models: A research data category?


Thomas Koprucki, Karsten Tabelow



submitted: May 25, 2016



Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: thomas.koprucki@wias-berlin.de
karsten.tabelow@wias-berlin.de



No. 2267
Berlin 2016



2010 *Mathematics Subject Classification.* 68U35, 97M10.

*Key words and phrases.* Research data, mathematical modeling and simulation, mathematical knowledge management.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

**Abstract.** Mathematical modeling and simulation (MMS) has now been established as an essential part of the scientific work in many disciplines and application areas. It is common to categorize the involved numerical data and to some extend the corresponding scientific software as research data. Both have their origin in mathematical models. In this contribution we propose a holistic approach to research data in MMS by including the mathematical models and discuss the initial requirements for a conceptual data model for this field.

# 1 Introduction

In recent years the handling of research data as part of the scientific practice has created vivid discussions within the scientific community, at research institutions as well as in funding agencies. Specifically, the importance of research data and its storage in view of new digital technologies is emphasized by the recent adoption of the “DFG Guidelines on the Handling of Research Data” by the Deutsche Forschungsgemeinschaft [1], the Open Research Data Pilot within the EU Horizon 2020 program [2], or the development of principles for research data handling within the german scientific organizations, the Leibniz Association, the Max-Planck-, Helmholtz-, and Fraunhofer Society by the Priority Initiative “Digital Information”[1].

The importance of appropriate handling of research data is increasingly recognized in view of its rising amount. It is central part of the discussion on Open Data and a prerequisite of the scientific method. In the face of the emerging digital science agenda research data proves to be an essential foundation for scientific work. Driven by such considerations universities and scientific institutions started creating policies for the handling of research data. These include rules for the full data life-cycle including generation, storage, preparation for subsequent re-use, publication and curation of data. However, the nature of research data is as diverse as the scientific disciplines requiring specific discussions and concepts.

# 2 Research data in mathematical modeling and simulation

Mathematics is one of the foundations of today’s key technologies and science. Mathematical methodology is required for interdisciplinary modeling of a research problem, for its mathematical treatment and solution, and for the transfer of the results into practice. In the last decade *mathematical modeling and simulation* (MMS) has been established alongside experiment and theory and is now essential part of the scientific work in many disciplines and application areas.

[1] http://www.allianzinitiative.de/en/core-activities/research-data.html

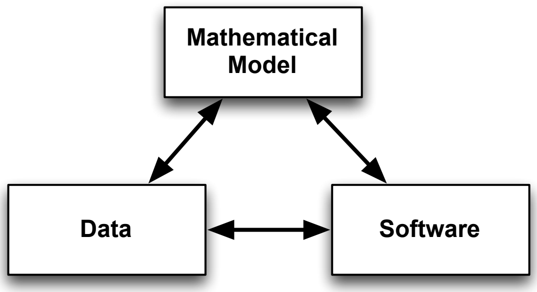


**Figure 1** – *This figure illustrates the three different components constituting research data in MMS: mathematical models, simulation software and numerical data (e.g., input parameters and data, output of simulation software).*

Research in the area of MMS is characterized by *mathematical models*, *scientific software* for their treatment, and *numerical data* related to computations (input, output, parameters), see Figure 1. Here, we propose to categorize these three parts as the research data in MMS as they are jointly required to understand and verify research results, or to build upon them.

Specifically, numerical data is generally regarded as research data in common sense and data repositories and information services such as DataCite [3] or RADAR [4, 5] exist or are emerging. Increasingly, software is categorized as reasearch data [1] and a world-leading information service on mathematical software, swMath [6], has already been developed.

Yet, communication in MMS suffers from the absence of a unified concept including mathematical models: instead of considering mathematical models as entities of their own class, which can be uniquely identified, cited and categorized, they are rather found as plain text with a mixture of mathematical notation and common language. This potentially leads to ambiguity, cites to different original work, incompleteness, and “re-invention of the wheel”.

A comprehensive approach to research data in MMS should cover all three aspects in a similar manner. While for software and numerical data the above mentioned services are reasonable starting points for the implementation of such a concept, a corresponding definition and service for mathematical models is missing. A similar system for computational models of biological processes has been introduced at the BioModels Database[2].

In this contribution we discuss the initial requirements for a conceptual data model for mathematical models, starting from a simple and widely-used example. This is a first step towards the creation of a semantic corpus for mathematical models in MMS, which can serve as a standardized access to mathematical models with cross-links to software, data repositories, and publications.

[2]https://www.ebi.ac.uk/biomodels-main/

## 3 Mathematical models: the heat transport problem

As an example for mathematical modeling we consider the heat transport problem. Modeling and simulation of heat transport is a common task in many technical applications ranging from large heat exchangers to heating effects in small semiconductor devices. For our discussions on a formalization concept for mathematical models we outline the corresponding description of the heat transport model and its ingredients as plain text, which might be similarly found in typical publications on this topic or in software documentations.

**Heat transport model.** We describe heat conduction by Fourier's law

$$q = -\lambda \nabla T, \tag{1}$$

where $q$ denotes the *heat flux*, $\lambda$ represents the *heat conductivity* and $\nabla T$ is the *temperature* gradient. In a bounded spatial domain $\Omega$ the time evolution of the *temperature distribution* $T(x,t)$ is then governed by the *heat flow equation*:

$$\frac{\partial}{\partial t}(C(x)T(x,t)) - \nabla \cdot (\lambda(x)\nabla T(x,t)) = f(x) \text{ in } \Omega, \tag{2}$$

with a *heat source* $f(x)$, the *heat capacity* $C(x)$ of the material, and boundary conditions

$$-\nu \cdot q = \nu \cdot (\lambda(x)\nabla T(x,t)) = \kappa(x)(T(x,t) - T_a(x)) \text{ on } \partial\Omega, \tag{3}$$

where $\nu$ denotes the outer normal vector, $\kappa$ the *heat transfer coefficient* to the environment and $T_a$ is the *ambient temperature*. In studies of time-dependent heating phenomena the time evolution of the temperature and thus the heat flow is given by the solution of the boundary value problem (1)-(2) with the initial value $T_0(x) = T(x, t = 0)$. In contrast, one is often only interested in the stationary heat paths, e.g., in studying the heat flow from a device. Then it suffices to solve the *stationary heat equation*

$$-\nabla \cdot (\lambda(x)\nabla T(x,t)) = f(x) \text{ in } \Omega \tag{4}$$

subject to the boundary conditions (2).

## 4 Towards a conceptual data model for mathematical models

The recognition of mathematical models as part of research data in MMS can be established by the creation of a semantic digital corpus of mathematical models. An information service for the registration and retrieval of mathematical models is then necessary for the adoption of the approach by the MMS community and for navigation, indexing and searching the model corpus.

The creation of such a corpus cannot just rely on a plain text description as above, instead one has to develop a normal or canonical form. A similar normalization is common in general mathematical texts where definitions, lemmas, theorems, proofs, corollaries, propositions help to structure the content. Similar to the approach of the semantical annotation of mathematical texts a normal form for mathematical models needs to be represented in a modeling-oriented markup-language, which can be based on LaTeX or MathML. In contrast to a pure plain text description such a mark-up can be used to generate relations between the entities of the formal description.

The encoded entities should contain the main characteristics of the model, such as the equation, the domain, boundary conditions, material laws and constitute a signature for the mathematical model. However, the complexity of the task is far above simple one-to-one mappings as it is possible, e.g., for special functions. A mathematical model is an abstract notion relying on a mathematical equation combined with semantic binding. Despite the fact that typically multiple notations for the same equation exist, the task is further complicated by the non-trivial question which entities are to be considered as atoms of the description. For example, the definition of the heat flux in the heat transport model can be itself considered as a model. The same applies to the material laws such as the heat conductivity or the heat capacity where a constant or linear dependence can be described by a single parameter. Finally, the replacement of the boundary conditions of the heat transfer (2) on parts of the boundary by a model for heat radiation ($T^4$-law) leads to further variants of the original model with specific properties.

A data model for mathematical models must reflect a sufficient level of complexity of the formal description to cover a large number of models while avoiding unnecessary duplications in their encoding. It is a-priori not clear whether such a description exists. The problem can be mitigated by appropriate relations between different entries of the model corpus. In its final form an information service for mathematical model should not only include models characterized by partial differential equations, but also statistical or discrete models, as well as systems of ordinary differential equations and many more.

Beside the plethora stemming from different specializations of a certain model as introduced above two further dimensions of a data model are essential which we introduce as *math bindings* and *application bindings*.

**Math bindings.** The mathematical notation of a specific model is everything but unique. Even for the non-dimensionalized heat equation with constant coefficients ($\lambda(x) = const.$, $C(x) = const.$) there exists a whole diversity of possible mathematical notations such as a notation with Nabla calculus as above, a representation in Cartesian coordinates, simplification to a Laplacian, a notation with $div$- and $grad$- operators, weak formulations, or formulations as a gradient flow.

It is common to classify linear, second order partial differential equations as elliptic, parabolic or hyperbolic. For instance the transient heat equation (1) is mathematically classified as a *parabolic partial differential equation*, whereas the stationary heat flow problem (4) constitutes a partial differential equation of *elliptic* type. The classification provides useful hints for their mathematical treatment and for the characterization of their solutions. The mathematically precise formulation of the model equations relates the assumptions on the data of the problem, e.g., regularity of coefficient functions or smoothness of the domain and its boundaries, to mathematical theory.

Furthermore, for the numerical solution of the model equations different computational methods can be used. This introduces another aspect related to the model description and the utilized software, which might also be regarded as a math binding.

**Application bindings.** The universality of mathematical models allows for transferring models from one application area to a different context. For example the heat flow model above can be re-interpreted as model of diffusion processes of particles. In this case the quantities get a new semantic meaning together with a new notation: the temperature $T$ is the particle density $u$, the heat conductivity $\lambda$ becomes the diffusion coefficient $D$.

A second aspect is the usage of models as building blocks to describe coupled phenomena like in thermistor models, which couple thermal and electric transport, or heat treatment of steel which couples heat transport with phase transitions and elasticity. In these cases coefficient functions are defined by solutions of supplemental differential equations, e.g., for thermistors the Joule heat generated by a current flow enters the heat equation (1) as a source term $f(x)$.

Both aspects are key features of mathematical modeling which are related to the abstraction given by the mathematical language. They are the basis for the strength of mathematical modeling and for the success of MMS as a third discipline between theory and experiment.

**Connection to software and data.** The application of a mathematical model, such as the heat transport model, to a specific technical problem requires a mapping of mathematical objects such as coefficient functions to *properties*, or more precisely *material parameters*, of the involved *materials* and boundaries. In our example these are the heat conductivity $\lambda(x)$, the heat transfer coefficient $\kappa(x)$ and the heat capacity $C(x)$. The numerical solution of the heat transport problem requires the approximation of the continuous problem (1) by discretization methods. This involves a *geometric description* of the simulation domain by suitable meshes. Typically, the simulation results in numerical values of the temperature distribution $T$ on the numerical mesh constituting the output of the simulation software. Correspondingly, *initial values* and the *material data* are the input for the software. Both, simulation results and input data, constitute the numerical data part of the research data in MMS, see Figure 1. Certain mathematical objects occurring in mathematical model have a semantical binding, namely of $T$ being the temperature, $\lambda$ being the heat conductivity etc., but they also link input and output of the utilized software and its interpretation.

# 5 Conclusions

We proposed to categorize mathematical models, scientific software and numerical data as the research data in MMS requiring suitable information services for their management and handling. For numerical data and scientific software the awareness of this fact in the MMS community is growing and suitable concepts and information services are emerging. However, for the category of mathematical models, a corresponding definition and service is missing. We highlighted the initial conceptual requirements for the definition of a suitable data model and its difficulties on the basis of the heat transport model. A unified approach to research data management that not only includes numerical data and scientific software but also mathematical models can help to enhance future MMS publications by making them more concise. A digital corpus of mathematical models together with a suitable information service is necessary to reduce the additional effort for the authors. On success its creation will be an important contribution of applied math to the digital science agenda. Furthermore, it has the potential to reduce today's language barriers between disciplines and requires an interdisciplinary effort.


## Acknowledgments

The authors are grateful for many fruitful discussions with W. Sperber who helped to shape their knowledge on the topic.


## References


[1] Deutsche Forschungsgemeinschaft: DFG Guidelines on the Handling of Research Data. Adopted by the Senate of the DFG at September 30, 2015.

[2] European Commission: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.1, 15 February 2016.

[3] Brase, J.: DataCite - a global registration agency for research data. Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO '09. Fourth International Conference on, Beijing, 257-261 (2009)

[4] Razum, M., Neumann, J., Hahn, M.: RADAR – Ein Forschungsdaten-Repositorium als Dienstleistung für die Wissenschaft. Zeitschrift für Bibliothekswesen und Bibliographie, 61, 18–27 (2014)

[5] Kraft, A.: RADAR – A repository for long tail data. Proceedings of the IATUL Conferences. Paper 1 (2015)

[6] Greuel, G.-M., Sperber, W.: swMATH – an information service for mathematical software. In: H. Hong (ed.) et al., Mathematical software – ICMS 2014. Lecture Notes in Computer Science 8592, 691-701 (2014).