

DIE ROLLE DER GAMM BEIM AUFBAU EINER NATIONALEN FORSCHUNGSDATEN-INFRASTRUKTUR

VON PETER BENNER UND STEFAN DIEBELS

Am 6. Juni 2019 wurde von der DFG die Ausschreibung 2019 für die Förderung von Konsortien zum Aufbau einer Nationale Forschungsdateninfrastruktur (NFDI) veröffentlicht¹. Die GAMM unterstützt dabei insbesondere zwei der geplanten Fachkonsortien in den Bereichen Mathematik und Materialwissenschaften. Wir möchten hier einen kurzen Überblick über Struktur, Ziele und Aufgaben der NFDI, sowie der beiden genannten Konsortien, geben.

Viele populärwissenschaftliche Veröffentlichungen der letzten 10-15 Jahre in den computergestützten Ingenieur- und Naturwissenschaften („CSE“), also auch zu vielen GAMM-relevanten Themen, beginnen mit der Feststellung, dass sich neben Theorie und Experiment die Computersimulation als dritte Säule des wissenschaftlichen Erkenntnisgewinns etabliert hat. Seit neuestem spricht man manchmal auch von einer vierten Säule, nämlich den „Daten“. Big Data und Methoden des maschinellen Lernens gehören sicher zu den am meisten gehypten Begriffen, mit denen Wissenschaft derzeit in Verbindung gebracht wird. Dabei spielen Daten natürlich unterschiedlichste Rollen und werden auf unterschiedlichste Weise generiert und weiterverarbeitet. In der Forschung gibt es dabei zum einen empirisch gewonnene Datensätze, auf der anderen Seite – und das ist keinesfalls neu – aus Experimenten (im Labor oder in silico) entstehende Forschungsdaten. Der Umgang mit diesen Datensätzen ist in den verschiedenen wissenschaftlichen Communities bislang sehr unterschiedlich – dies geht von systematischer Auswertung und dauerhafter, öffentlicher Verfügbarkeit bis hin zum Löschen der Daten, wenn diese einmal (ggf. für eine Veröffentlichung) ausgewertet wurden, um Speicherplatz zu sparen.

Das Ziel der NFDI ist es nun, dass „Datenbestände in einem aus der Wissenschaft getriebenen Prozess systematisch erschlossen, langfristig gesichert und entlang der

FAIR-Prinzipien über Disziplinen- und Ländergrenzen hinaus zugänglich gemacht werden“ [aus dem ersten Absatz der o.g. DFG Ausschreibung – die FAIR Prinzipien erläutern wir weiter unten]. Dazu stellt das BMBF im Endausbau jährlich 90 Millionen Euro zur Verfügung, wovon 5 Millionen Euro in den Aufbau und Betrieb eines Direktorats inkl. einer Geschäftsstelle fließen sollen. Es stehen also 85 Millionen Euro (inkl. Programmpauschale) zur Förderung von bis zu 30, aus den wissenschaftlichen Communities zu entwickelnde, Konsortien zur Verfügung, die in den drei Ausschreibungsrunden 2019, 2020 und 2021 Förderung beantragen können. Das Förderprogramm selbst wird von der DFG durchgeführt. Die GAMM unterstützt den Aufbau der NFDI, u.a. durch Ernennung der beiden Autoren zu ihren NFDI Vertretern in den weiter unten beschriebenen Konsortien. Allerdings ist z.Zt. die Rolle der Fachgesellschaften innerhalb der Konsortien und der NFDI noch keineswegs endgültig geklärt.

Wie bereits oben angesprochen, sollen Forschungsdaten, aber auch empirische Datensätze, die zu Forschungszwecken verwendet werden (können), gemäß den FAIR Prinzipien erschlossen, systematisiert und weitest möglich zugänglich gemacht werden. Daten sollen dabei

Findable,
Accessible,
Interoperable,
Re-usable

sein. Erste „best practice“ Studien, diese Prinzipien – zunächst oft unter dem Begriff RRR (Replicability, Reproducibility, Reusability) untersucht – im Bereich der CSE zu etablieren, finden sich z.B. in [1-4]. Hauptaufgabe der Konsortien wird es sein müssen, zunächst zu analysieren,

¹ Siehe https://www.dfg.de/foerderung/info_wissenschaft/info_wissenschaft_19_37/index.html

was der Datenbegriff in den einzelnen Fächern überhaupt bedeutet, wie die FAIR Prinzipien darauf anzuwenden sind und welche (Infra-)Strukturen geschaffen werden müssen, um den FAIR Prinzipien folgend die jeweiligen Datenbestände langfristig möglichst barrierefrei nutzbar zu machen. Dabei ist jedoch auch wichtig, in Zusammenarbeit der unterschiedlichen Konsortien Gemeinsamkeiten herauszuarbeiten, diese dann fachübergreifend zu nutzen sowie fachspezifische Gegebenheiten zu berücksichtigen. Hierbei ist es erforderlich, dass diese Prozesse aus der Wissenschaft heraus getragen und gesteuert werden, denn die erarbeiteten Konzepte und Regeln werden sicher mittel- bis langfristig auch zu Standards im Reporting von Forschungsergebnissen führen. Bereits bestehende Forschungsdaten-Infrastrukturen sollen dabei möglichst integriert und in die Prozesse eingebunden werden, um auf das dort bereits vielfältig existierende Wissen, nicht zuletzt auch im Bereich der notwendigen informationstechnischen Methoden und Strukturen (Software!) zurückgreifen zu können.

Mathematical Research Data Infrastructure - MaRDI

Der Begriff Daten hat in der Mathematik unterschiedlichste Konnotationen. In erster Linie haben Mathematiker bis zum Ende des 20. Jahrhunderts ihre Informationen aus Artikeln in mathematischen Fachzeitschriften bezogen. Um diesen großen und weiter exponentiell wachsenden „Datenbestand“, der teilweise bis ins 17. Jahrhundert zurückgeht, einfacher zugänglich zu machen, haben sich bereits in der ersten Hälfte des 20. Jahrhunderts zwei Servicedienste etabliert, die Kurzzusammenfassungen der mathematischen Literatur zur Verfügung stellen und insbesondere seit deren Digitalisierung die Suche nach mathematischer Fachliteratur erheblich vereinfachen. Einerseits ist dies das 1931 gegründete und heutzutage gemeinsam von der EMS², der Heidelberger Akademie der Wissenschaften und vom FIZ Karlsruhe (Leibniz-Institut für Informationsinfrastruktur) heraus-gegebene „Zentralblatt für Mathematik und ihre Grenzgebiete“ (heute: zbMATH), andererseits MathSciNet, basierend auf den 1940 gegründeten Mathematical Reviews von der AMS³. Mit dem FIZ, das die Dateninfrastruktur für zbMATH bereitstellt, steht damit ein kanonischer Partner sowie mit dem zbMATH ein bereits digitalisierter, mit Metadaten versehener und systematisch durchsuchbarer Datensatz als eine Säule für die NFDI in der Mathematik zur Verfügung. Natürlich ist die zbMATH-Datenbank nur ein kleiner Teil der inzwischen in der Mathematik zur Verfügung stehender Teil der Forschungsdaten. Dies reicht von Sammlungen mathematischer Objekte, wie z.B. den „kleinen

Gruppen“ oder geometrischer Objekte, die teilweise bislang nur handschriftlich zusammengetragen wurden, bis zu umfangreichen Simulationsdaten im Bereich CSE. Im ersten Bereich könnte MaRDI die Digitalisierung vorantreiben und solche Sammlungen systematisch der mathematischen Welt zur Verfügung stellen, wobei eine entsprechende Infrastruktur unterstützend dazu kommen muss. Im Bereich CSE kann die angewandte Mathematik nicht alleine agieren, da es hier immer um das Zusammenspiel mit unterschiedlichen Wissenschaftsbereichen geht. Hier sollen anhand von sogenannten „Use Cases“ und Benchmark-Studien „best practice“ Szenarien geschaffen werden, die als Vorbild zur Umsetzung der FAIR bzw. RRR Prinzipien dienen können. Dies soll weitestgehend im Zusammenspiel mit anderen NFDI Konsortien aus den entsprechenden Anwendungsdisziplinen erfolgen, z.B. mit dem im Folgenden vorgestellten NFDI4MSE Konsortium. Das MaRDI Konsortium, bestehend aus derzeit 14 Partnern inkl. FIZ, ZIB und MFO, unter der Führung des WIAS⁴ in Berlin wird u.a. von der EMS, der GAMM, der DMV⁵ und der GOR⁶ sowie den vier mathematischen DFG-Exzellenzclustern MATH+ in Berlin, MathematicsMünster, HCM⁷ Bonn und Structures in Heidelberg unterstützt.

NFDI für Materials Science and Engineering – NFDI4MSE

Eines der Ziele der "Materials Science and Engineering" (MSE) ist die Entwicklung, Charakterisierung und Optimierung neuer Materialien und Werkstoffe. Dabei ist die Erhebung, Analyse und die Verknüpfung von Werkstoffdaten ein wichtiger Baustein. Ein besonderer Aspekt ist dabei die intrinsische Mikrostruktur der Werkstoffe, die es erforderlich macht, physikalische oder auch chemische Prozesse auf unterschiedlichen Längen- und Zeitskalen zu berücksichtigen. Das Konsortium NFDI4MSE hat sich zum Ziel gesetzt, eine digitale Plattform zu etablieren, die es Wissenschaftlern ermöglicht, diese Daten nachhaltig zu nutzen. Zum vollständigen Verständnis des Materialverhaltens muss es dabei möglich sein, Informationen über die Historie des Werkstoffs, von der Werkstoffherstellung über unterschiedliche Verarbeitungsschritte bis zur Probenherstellung, zu erfassen, da die aktuelle Mikrostruktur und damit die Eigenschaften von diesen Schritten abhängen.

Um dieses Ziel zu erreichen, plant NFDI4MSE unter anderem die Entwicklung und Umsetzung von Datenformaten und Workflows, die das Arbeiten und dem Umgang mit hierarchischen Werkstoffdaten erleichtern. Einerseits sollen diese Formate den Austausch von Daten zwischen Wissenschaftlern erleichtern, andererseits müssen sie eine Datenhoheit, z.B. über eine Regelung von Zugriffsrechten

2 EMS – European Mathematical Society

3 AMS – American Mathematical Society

4 WIAS - Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin

5 DMV – Deutsche Mathematiker-Vereinigung e.V.

6 GOR – Gesellschaft für Operations Research e.V.

7 HCM - Hausdorff Center for Mathematics, Bonn

und Nutzungsbedingungen, garantieren. Eine Ontologie, die eine Verknüpfung zwischen unterschiedlichen Daten sowie das gezielte Auffinden von Daten ermöglicht, soll unter enger Einbeziehung der Fachcommunity entwickelt und umgesetzt werden. Auf dieser Basis können dann Auswertetools und Analyseverfahren weiterentwickelt und zugänglich gemacht werden und zu einer digitalen Darstellung der Materialien genutzt werden. Diese Grundlagen gestatten es auch Modellierern, zur Validierung ihrer Modelle gezielt auf Datensätze zuzugreifen und so auch ohne eigene Expertise im Bereich der Charakterisierung die Qualität der entwickelten Modelle zu überprüfen. Alle Maßnahmen der NFDI4MSE sollen skalierbar sein, so dass sowohl einzelne Wissenschaftler als auch größere Forschungsgruppen und -konsortien davon profitieren werden. Eine enge Verknüpfung mit geeigneten, existierenden oder neu beantragten Forschungsverbänden wie z.B. Exlnis, SFBs und SPPs dienen dabei als zusätzliche Multiplikatoren.

Unterstützt wird NFDI4MSE auch durch die Fachgesellschaften VDI⁸, DVM⁹, GAMM, DGM¹⁰.

Weitere Informationen unter: <https://www.nfdi4mse.de>

8 VDI – Verein Deutscher Ingenieure e.V.

9 DVM – Deutscher Verband für Materialforschung und -prüfung e.V.

10 DGM – Deutsche Gesellschaft für Materialkunde e.V.

Literatur

- [1] J. Fehr, J. Heiland, C. Himpe, J. Saak: Best Practices for Replicability, Repeatability and Reusability of Computer-Based Experiments Exemplified by Model Reduction Software. *AIMS Mathematics* 1(3):261-281, 2016.
- [2] T. Koprucki, K. Tabelow: Mathematical models: A research data category? In Proceedings of "Mathematical Software - ICMS 2016: 5th International Conference", Berlin, Germany, July 11-14, 2016, G.-M. Greuel, Th. Koch, P. Paule, A. Sommese (Hrsg.), Lecture Notes in Computer Science, Springer International Publishing AG Switzerland, Cham, S. 423-428, 2016.
- [3] M. Kohlhase, T. Koprucki, D. Müller, K. Tabelow: Mathematical models as research data via flexiformal theory graphs. In Proceedings of "Intelligent Computer Mathematics: 10th International Conference, CICM 2017, Edinburgh, UK, July 17-21, 2017", H. Geuvers, M. England, O. Hasan, F. Rabe, O. Teschke (Hrsg.), Springer International Publishing, Cham, S. 224-238, 2017.
- [4] B. Drees, A. Kraft, T. Koprucki: Reproducible research through persistently linked and visualized data. *Opt. Quantum Electron.*, 50:59/1-59/10, 2018.



Peter Benner schloss sein Mathematik-Studium an der RWTH Aachen in 1993 mit dem Diplom ab, promovierte 1997 bei Volker Mehrmann an der TU Chemnitz-Zwickau und habilitierte sich 2001 in Mathematik an der Universität Bremen. Nach Stationen an der TU Hamburg-Harburg und der TU Berlin folgte er 2003 dem Ruf auf eine C4-Professur für „Mathematik in Industrie und Technik“ an der TU Chemnitz. 2010 wurde er zum Wissenschaftlichen Mitglied und Direktor am Max-Planck-Institut für Dynamik komplexer technischer Systeme in Magdeburg ernannt. Seit 2011 ist er auch Honorarprofessor an der Otto-von-Guericke-Universität Magdeburg. Seine Arbeitsgruppe an der TU Chemnitz führt er seit 2010 in reduziertem Umfang fort. Die Forschungsschwerpunkte liegen im Bereich der Numerischen Linearen und Multilinearen Algebra, insbesondere Eigenwertprobleme, Vorkonditionierung, Matrixgleichungen und Niedrigrang-Matrix-/Tensorfaktorisierungen, sowie der Modellreduktion, Regelung und Steuerung dynamischer Systeme. Besondere Aufmerksamkeit legt er auch auf die Implementierung der entwickelten Methoden auf modernen Rechnerarchitekturen sowie der Bereitstellung der zugehörigen mathematischen Software, z.B. SLICOT, MORLab und M.E.S.S..



Stefan Diebels studierte von 1984 bis 1989 an der Technischen Hochschule Darmstadt, der jetzigen TU Darmstadt, Maschinenbau und Mechanik. Anschließend war er Doktorand bzw. Postdoc am Institut für Mechanik der Technischen Hochschule Darmstadt. Die Promotion erfolgte 1992 mit einem Thema aus der Strömungsmechanik. Von 1995 bis 2002 war er Akademischer Rat am Institut für Mechanik (Bauwesen) der Universität Stuttgart, wo er sich 2000 für das Fach Mechanik habilitierte. Seit 2002 ist er Professor für Technische Mechanik an der Universität des Saarlandes. Die Arbeitsgebiete sind die Materialmodellierung von Polymeren und von Werkstoffen mit Mikrostruktur sowie die experimentelle Mechanik.