

Mathematical Research Data Initiative (MaRDI)

Sprecher: Prof. Michael Hintermüller, WIAS Berlin, michael.hintermueller@wias-berlin.de

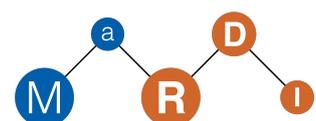
Kurze Beschreibung des geplanten Konsortiums Mathematische Forschungsdaten haben gewaltige Ausmaße angenommen. Sie sind komplex und vielseitig und durch die Anwendung von Mathematik in der interdisziplinären Forschung in der Wissenschaftslandschaft weit verbreitet. Das reicht von hochkomplexen Daten aus dem Scientific Computing bis hin zu mathematischen Informationsdatenbanken wie der Standardreferenz für spezielle Funktionen und Tabellen usw., wie sie z. B. vom US National Institute for Standards and Technology bereitgestellt und üblicherweise von Experten der verschiedenen Fachgebiete abgerufen werden. Das Datenvolumen und die Geschwindigkeit seines Anwachsens nehmen durch die Verbreitung von mathematischen Methoden in den Datenwissenschaften und zunehmende Rechnerleistung stetig zu. Dadurch entstehen zunehmend komplexere mathematische und rechnergestützte Modelle in verschiedenen Wissenschaftsfeldern.

Motiviert durch die Bedürfnisse und Anforderungen aus der mathematischen *Community*, aber auch aus anderen wissenschaftlichen Disziplinen, die quantitative Methoden verwenden, will MaRDI (**M**athematical **R**esearch **D**ata **I**nitiative), die Konsortialinitiative der mathematischen Wissenschaften, Standards setzen für zertifizierte mathematische Forschungsdaten und das Design nachvollziehbarer, bestätigbarer *Workflows* und will Dienstleistungen für die wissenschaftliche *Community* bereitstellen. Erklärtes Ziel ist es, die FAIR-Prinzipien auf dem gesamten Gebiet der Mathematik und ihrer Anwendungen einzuführen und umzusetzen.

Die drei durch die mathematische Forschung motivierten Säulen dieses Antrags basieren auf der Einteilung in *exakte und symbolische Daten* (Computer-Algebra und experimentelle Mathematik), *Gleitkommatdaten* (Scientific Computing) und *Daten mit Unsicherheiten* (Statistik). In der interdisziplinären mathematischen Forschung spielen *mathematische Modelle* eine zentrale Rolle. Dazu zählen auch Modelle des maschinellen Lernens. MaRDI will zertifizierte Daten und Softwareentwicklungen sowie nachvollziehbare (*confirmable*) *Workflows* in ihren jeweiligen Bereichen umsetzen und arbeitet auf die Entwicklung entsprechender Service-Prototypen hin. Außerdem gibt es wegen der interdisziplinären Kraft der Mathematik verschiedene *Use Cases* und Kooperationen mit anderen NFDI-Konsortialinitiativen, welche u.a. NFDI4Ing, NFDI4MSE, PAHN-PaN, NFDI4Chem, FAIRmat, NFDI-Neuro, und NFDI4Culture beinhalten. Die prototypischen Dienste sollen zu betriebsbereiten Diensten mit einem signifikanten Mehrwert für die aktuelle Forschung ausgeweitet werden oder neue Forschung ermöglichen. Ein weiterer wichtiger Bestandteil von MaRDI ist die Entwicklung von wesentlich verbesserten Diensten zur Informationsgewinnung, inkl. mathematischen Modellen als Forschungsdaten, ein disziplinenübergreifender mathematischer digitaler semantischer Atlas, Ontologien und Metadaten, wie z. B. eine Algorithmen-Metadaten-Bibliothek.

Beispiele mathematischer Forschungsdaten

- **Mathematische Dokumente** in PDF, LaTeX, XML, MathML etc.
- **Notebooks** z.B. in Jupyter oder Mathematica.
- **Domain-spezifische wissenschaftliche Software-Pakete und ihre Bibliotheken** wie R für Statistik, Octave, NumPy/SciPy oder Julia für Matrix-Berechnungen, CPLEX, Gurobi, Mosel



und SCIP für Integer Programming, DUNE, deal.II und Trilinos für numerische Simulationen.

- **Computer Algebra Systeme** wie SageMath, SINGULAR, GAP, polymake, Pari/GP, OSCAR und die in ihnen enthaltenen Daten-Sammlungen.
- **Programme und Skripte** in den obigen Systemen oder in Systemen, die nicht in der mathematischen Community entwickelt werden; aber auch Eingabedaten für diese Systeme wie Algorithmenparameter, Gitter, mathematische Objekte in Sammlungen, die Definition eines neuronalen Netzes als Graph im maschinellen Lernen etc.
- **Data Frames und Simulationsdaten**, meist Serien von Zuständen eines Systems, diskretisierte Felder, allgemeiner sehr große aber strukturierte Daten als Ausgabe einer Simulation oder als experimentelles Ergebnis, gespeichert in etablierten Datenformaten (z.B. HDF5).
- **Formalisierte Mathematik**: Coq, HOL, Isabelle, Mizar, NASA PVS library.
- **Sammlungen mathematischer Objekte**, z.B. L-Functions und modular forms database (LMFDB), Online Encyclopedia of Integer Sequences (OEIS), Class Group Database, ATLAS of Finite Group Representations, Manifold Atlas, GAP Small Groups Library.
- Beschreibung **mathematischer Modelle** in mathematischen **Modellierungssprachen**, z.B. Modelica für Komponenten-basierte Modellierung komplexer Systeme, Systems Biology Markup Language (SBML) für biologischer Prozesse, SPICE zur Modellierung elektronischer Schaltkreise und AIMMS oder LINGO als Modellierungssprache für ganzzahlige Optimierung.

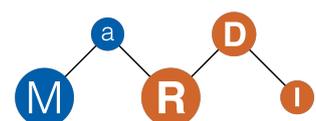
MaRDI Services MaRDI will ein digitales Serviceportal als einzigen, leicht zu erreichenden “One Stop“-Kontakt entwickeln, wo die wissenschaftliche *Community* die Dienste von MaRDI abrufen und abfragen kann. Das Portal soll agil entwickelt und permanent eingerichtet werden.

MaRDI setzt dabei auf

- verteilte Repositorien, Langzeit-Archivierung, Erzeugung von stabilen persistenten Identifiern (PIDs), Autorisierung und Autorisierungs-Mechanismen;
- Verarbeitung von Datenströmen und verteiltes Rechnen (Cloud-Computing), Erzeugung mathematischer Knowledge Graphen
- fortgeschrittene mathematische Suchfunktionen wie Formelsuche, Auswahl und Filterung nach mathematischen Eigenschaften.

Die nachhaltige Umsetzung der Ergebnisse von MaRDI erfordert eine *Community*, die eine von den FAIR-Vorgaben bestimmte Datenkultur und ebensolche Forschungs-Workflows einhält. Dafür will MaRDI *Collaborative Platforms* einrichten, denen eine wichtige Rolle bei der Wissensverbreitung, der Fortsetzung des wissenschaftlichen Diskurses und der Sicherstellung der Qualitätskontrolle zukommen soll.

Spezielle Anforderungen an das Forschungsdatenmanagement Das zentrale Anliegen von MaRDI ist die Etablierung der FAIR-Prinzipien für mathematische Forschungsdaten, was aufgrund ihrer Heterogenität eine Herausforderung darstellt. Neben klassischen Forschungsdaten (z.B. numerische Simulationsdaten) umfassen diese auch Algorithmen, Diskretisierungen sowie mathematische



Objekte (z.B. Polytope) oder Gleichungen, Modelle und mathematische Dokumente. Von besonderer Komplexität sind etwa mathematische Dokumente und (Jupyter-)Notebooks, weil diese Text, Gleichungen, Berechnungen und mathematische Objekte gemischt enthalten. Ziel der MaRDI-Plattform ist die Entwicklung von Mehrwertdiensten für die Forschung auf der Basis dieser komplexen Datenstrukturen. Zur effizienten Gestaltung plant MaRDI eine Ebene zwischen Anwendung und den Datenspeichern. Diese Zwischenebene stellt die Interoperabilität sicher und wird ausgehend von bestehenden Standards und Beschreibungssprachen aufgebaut.

Erfahrungen und Expertise MaRDI verfügt über Expertise im Bereich der Informations- und Dateninfrastrukturen sowie im mathematischen Wissensmanagement. Beispiele sind das Repositoryum RADAR zur Langzeit-Archivierung von Forschungsdaten, das weltweit umfangreichste mathematische Informationsportal zbMATH, die Software-Datenbank swMATH und MathWebSearch für mathematische Formeln. Darüberhinaus kann MaRDI Erfahrungen aus dem EU-H2020 Projekt OpenDreamKit nutzen. Die dort entwickelte E-Infrastruktur für kollaboratives Arbeiten, das Einbinden von spezieller mathematischer Software in die Jupyter Notebook-Architektur und das Bereitstellen von Software-Containern zum Cloud-Computing oder zur Integration in JupyterLab haben zu einer deutlichen Steigerung der Nutzerfreundlichkeit geführt und ermöglichen die Durchführung reproduzierbarer Computereperimente.

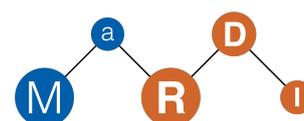
Schnittstellen zur gesamten NFDI Mathematische Forschungsdaten sind außerordentlich vielfältig, da sie aus vielen verschiedenen Disziplinen (wie Physik, Chemie, Ingenieurs-, Lebens- und Geisteswissenschaften) stammen, die mathematische Methoden und Lösungsansätze benutzen. Der Grund dafür ist die Kraft der mathematischen Abstraktion, die Strukturen hervorbringt, die nicht nur in der Mathematik, sondern auch in anderen Disziplinen angewendet werden können. Darum soll eine Forschungsdaten-Infrastruktur geschaffen werden, die nicht nur der Mathematik von Nutzen sein, sondern auch einen bedeutenden Einfluss auf andere Wissenschaftsgebiete haben soll.

Mathematik kann präzise Formulierungen der Probleme und Theorien aus anderen Disziplinen zur Verfügung stellen. Das beinhaltet die verwendeten mathematischen Objekte mittels Mathematik als verbindender Sprache. Dazu gehören Modelle, Parameter, Daten und numerische Algorithmen. Es ist wichtig, diese mathematischen Objekte mit ihrem domain-spezifischen Kontext und Semantik zu verknüpfen. Dasselbe mathematische Objekt oder Modell kann in völlig verschiedenen Kontexten mit unterschiedlichen semantischen Bedeutungen auftreten.

Querschnittsthemen

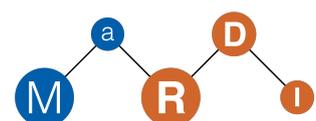
- Cloud-basierte Datenverarbeitung (MaRDI)
- Knowledge Graph und semantische Technologien (MaRDI)
- Förderierte Repositorien
- Meta-Daten für wissenschaftliche Software, Workflows, Computer-Experimente (MaRDI)

Erwartungen an die NFDI-Konferenz MaRDI möchte auf der Konferenz erfahren, wie die NFDI als Ganzes aufgebaut werden soll und wie sich die Konsortien darin eingliedern werden. Außerdem wird der Kontakt zu den relevanten neuen Partnern der diesjährigen Antragsrunde gesucht werden.



Vorgesehene Mitglieder des Konsortiums

Co-Sprecher/in	Zugehörige Institution
Michael Joswig E-Mail: joswig@math.tu-berlin.de	Deutsche Mathematiker-Vereinigung e.V. (DMV) c/o WIAS Mohrenstraße 39 10117 Berlin
Olaf Teschke Position: Leiter Abteilung Mathematik am FIZ E-Mail: olaf.teschke@fiz-karlsruhe.de	FIZ Karlsruhe – Leibniz Institut für Informationsinfrastruktur GmbH (FIZ) Franklinstraße 11 10587 Berlin
Anita Schöbel Position: Leiterin des ITWM E-Mail: anita.schoebel@itwm.fraunhofer.de	Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (ITWM) Fraunhofer-Platz 1 67663 Kaiserslautern
Michael Kohlhase Position: Professor für Wissensrepräsentation und -verarbeitung E-Mail: michael.kohlhase@fau.de	Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) Schloßplatz 4 91054 Erlangen
Stephan Klaus Position: wissenschaftlicher Administrator MFO E-Mail: klaus@mfo.de	Mathematisches Forschungsinstitut Oberwolfach gGmbH (MFO) Schwarzwaldstraße 9-11 77709 Oberwolfach-Walke
Peter Benner Position: geschäftsführender Direktor MPI DCTS E-Mail: benner@mpi-magdeburg.mpg.de	Max-Planck-Institut für Dynamik komplexer technischer Systeme (MPI DCTS) Sandtorstraße 1 39106 Magdeburg
Bernd Sturmfels Position: Direktor MPI MIS E-Mail: bernd@mis.mpg.de	Max-Planck-Institut für Mathematik in den Naturwissenschaften (MPI MIS) Inselstraße 22 04103 Leipzig
Michael Joswig Position: Einstein-Professor für Diskrete Mathematik/Geometrie E-Mail: joswig@math.tu-berlin.de	Technische Universität Berlin (TUB) Straße des 17. Juni 135 10623 Berlin
Wolfram Decker Position: Professor, Arbeitsgruppe Algebra, Geometrie und Computeralgebra E-Mail: decker@mathematik.uni-kl.de	Technische Universität Kaiserslautern (TUK) Gottlieb-Daimler-Straße 47 67653 Kaiserslautern
Dominik Göddeke Position: Lehrstuhl Mathematische Methoden für komplexe Simulation der Naturwissenschaft und Technik E-Mail: dominik.goeddeke@mathematik.uni-stuttgart.de	Universität Stuttgart (USTUTT) Allmandring 5b 70569 Stuttgart



Co-Sprecher/in	Zugehörige Institution
Mario Ohlberger Position: Professur für Angewandte Mathematik, insbesondere Numerik E-Mail: mario.ohlberger@uni-muenster.de	Westfälische Wilhelms-Universität Münster (WWU) Schloßplatz 2 48149 Münster
Christof Schütte Position: Präsident ZIB E-Mail: schuette@zib.de	Zuse Institut Berlin (ZIB) Takustraße 7 14195 Berlin
Claus Fieker Position: Professor, Arbeitsgruppe Algebra, Geometrie und Computeralgebra E-Mail: fieker@mathematik.uni-kl.de	Technische Universität Kaiserslautern (TUK) Gottlieb-Daimler-Straße 47 67653 Kaiserslautern

