1 Binding letter of intent as advance notification or non-binding letter of intent

☑ Binding letter of intent (required as advance notification for proposals in 2020)

☐ Non-binding letter of intent (anticipated submission in 2021)

2 Formal details

Planned name of the consortium Mathematical Research Data Initiative

Acronym of the planned consortium MaRDI

Applicant institution	Location
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS), Leibniz-Institut im Forschungsverbund Berlin e. V. Director: Michael Hintermüller	Mohrenstraße 39 10117 Berlin

Spokesperson	Institution, location
Michael Hintermüller email: michael.hintermueller@wias-berlin.de	WIAS Berlin Mohrenstraße 39 10117 Berlin

Co-applicant institutions	Location
Deutsche Mathematiker-Vereinigung e.V. (DMV) c/o WIAS President: Friedrich Götze, from January 2021 on: Ilka Agricola	Mohrenstraße 39 10117 Berlin
FIZ Karlsruhe – Leibniz Institut für Informationsinfrastruktur GmbH (FIZ) CEO: Sabine Brünger-Weilandt	Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V. für ihr Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (ITWM) Director: Anita Schöbel	Fraunhofer-Platz 1 67663 Kaiserslautern
Freie Universität Berlin (FUB) President: Günter M. Ziegler	Kaiserswerther Str. 16/18 14195 Berlin
Ludwig-Maximilians-Universität München (LMU) President: Bernd Huber	Geschwister-Scholl-Platz 1 80539 München
Mathematisches Forschungsinstitut Oberwolfach gGmbH (MFO) Director: Gerhard Huisken	Schwarzwaldstraße 9-11 77709 Oberwolfach-Walke

Co-applicant institutions	Location
Max-Planck-Gesellschaft zur Förderung der Wissenschaften e. V. für ihr Max-Planck-Institut für Dynamik komplexer technischer Systeme (MPI DCTS) Managing Director: Udo Reichl	Sandtorstraße 1 39106 Magdeburg
Max-Planck-Gesellschaft zur Förderung der Wissenschaften e. V. für ihr Max-Planck-Institut für Mathematik in den Naturwissenschaften (MPI MIS) Managing Director: Jürgen Jost	Inselstraße 22 04103 Leipzig
Technische Universität Berlin (TUB) President: Christian Thomsen	Straße des 17. Juni 135 10623 Berlin
Technische Universität Kaiserslautern (TUK) President: Arnd Poetzsch-Heffter	Gottlieb-Daimler-Straße 47 67653 Kaiserslautern
Technische Universität München (TUM) President: Thomas Hofmann	Arcisstraße 21 80333 München
Universität Stuttgart (USTUTT) Rector: Wolfram Ressel	Keplerstraße 7 70174 Stuttgart
Westfälische Wilhelms-Universität Münster (WWU) Rector: Johannes Wessels	Schloßplatz 2 48149 Münster
Zuse Institut Berlin (ZIB) President: Christof Schütte	Takustraße 7, 14195 Berlin

Co-spokesperson	e-mail	Institution, location
President of DMV (Jan 21 ++: Ilka Agricola)	Agricola@mathematik.uni-marburg.de	DMV Mohrenstraße 39 10117 Berlin
Peter Benner	Benner@mpi-magdeburg.mpg.de	MPI DCTS Sandtorstraße 1 39106 Magdeburg
Bernd Bischl	Bernd.Bischl@stat.uni-muenchen.de	LMU Geschwister-Scholl-Platz 1 80539 München
Wolfram Decker	Decker@mathematik.uni-kl.de	TUK Gottlieb-Daimler-Straße 47 67653 Kaiserslautern
Mathias Drton	Mathias.Drton@tum.de	TUM Boltzmannstr. 3 85748 Garching
Claus Fieker	Fieker@mathematik.uni-kl.de	TUK Gottlieb-Daimler-Straße 47 67653 Kaiserslautern

Co-spokesperson	e-mail	Institution, location
Dominik Göddeke	Dominik.Goeddeke@mathematik.uni-stuttgart.de	USTUTT Allmandring 5b 70569 Stuttgart
Michael Joswig	Joswig@math.tu-berlin.de	TUB Straße des 17. Juni 135 10623 Berlin
Stephan Klaus	Klaus@mfo.de	MFO Schwarzwaldstraße 9-11 77709 Oberwolfach-Walke
Mario Ohlberger	Mario.Ohlberger@uni-muenster.de	WWU Schloßplatz 2 48149 Münster
Harald Sack	Harald.Sack@fiz-Karlsruhe.de	FIZ Franklinstraße 11 10587 Berlin
Rainer Sinn	rsinn@zedat.fu-berlin.de	FUB Kaiserswerther Str. 16/18 14195 Berlin
Anita Schöbel	Anita.Schoebel@itwm.fraunhofer.de	ITWM Fraunhofer-Platz 1 67663 Kaiserslautern
Christof Schütte	Schuette@zib.de	ZIB Takustraße 7 14195 Berlin
Bernd Sturmfels	Bernd.Sturmfels@mis.mpg.de	MPI MIS Inselstraße 22 04103 Leipzig

Participants	Location
European Mathematical Society Department of Mathematics and Statistics	P.O.Box 68 00014 University of Helsinki Finland
Exzellenzcluster MATH$^+$ (EXC 2046)	Technische Universität Berlin Sekr. MA 2-2 Strasse des 17. Juni 136 10623 Berlin
Exzellenzcluster ”Data-integrated Simulation Science” (EXC 2075)	Universität Stuttgart SC SimTech Pfaffenwaldring 5a 70569 Stuttgart

Participants	Location
Exzellenzcluster ”STRUCTURES: A unifying approach to emergent phenomena in the physical world, mathematics, and complex data” (EXC 2181)	Ruprecht-Karls-Universität Heidelberg Institut für Theoretische Physik Philosophenweg 16 69120 Heidelberg
Fraunhofer-Verbund IUK-Technologie	Anna-Louisa-Karsch-Straße 2 10178 Berlin
Gesellschaft für Angewandte Mathematik und Mechanik e. V. (GAMM)	Institut für Statik und Dynamik der Tragwerke Fakultät Bauingenieurwesen 01062 Dresden
Gesellschaft für Operations Research (GOR)	Kackertstraße 7 52072 Aachen
IMAGINARY gGmbH	Mittenwalder Str. 48 10961 Berlin
Bettina Eick Technische Universität Braunschweig	Universitätsplatz 2 38106 Braunschweig
Michael Kohlhase Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)	Schloßplatz 4 91054 Erlangen

3 Objectives, work programme and research environment

Research area of the proposed consortium (according to the DFG classification system)
Mathematics 312-01

Concise summary of the planned consortium's main objectives and task areas Mathematical research data has become vast, it is complex, and multifaceted, and, through the successful application of mathematics in interdisciplinary research, it is wide spread in the scientific landscape. It ranges from highly complex data from scientific computing to information bases such as the standard reference data for special functions, tables etc. as provided, e.g., by the US National Institute for Standards and Technology and routinely consulted by experts from various disciplines. The volume of data and its creation velocity increases dynamically with the nowadays rapid unfolding of mathematics in data science and with the ever-increasing computing power. The latter leads to increasingly complex mathematical and computational models arising in various scientific fields. Even more, mathematical research data is extremely diverse as it stems from many different disciplines (such as physics, chemistry, engineering sciences, humanities and life sciences) which use mathematical methods and solution approaches. This is due to the power of mathematical abstraction which yields structures that can be applied not only within mathematics but also in other disciplines. We hence aim at developing a research data infrastructure not only useful in mathematics but which will have significant impact also in other fields.

Motivated by the needs and requests from the mathematical community, but also from other scientific disciplines that utilize quantitative methods, MaRDI (**Ma**thematical **R**esearch **D**ata **I**nitiative), *the* consortial initiative of mathematical sciences, aims to set standards for certified mathematical research data, the design of confirmable workflows, and it provides services for the scientific community. The designated goal is to realize the FAIR principles across the field of mathematics and its applications. Thereby, the consortium will address the following objectives:

1. **Interoperable mathematical research data**: The goal is to develop interoperability of mathematical research data (from symbolic, numerical or runtime time data, equations, functions, to mathematical models and model hierarchies). This data conception should not only be useful within matheamtics, but also suitable for scientific disciplines that process mathematical data.

2. **Secure confirmable and reproducible results**: Mathematics has the unique property that its results can not only be made plausible by theory and experiment, but rigorously proven. However, complexity has increased vastly, and reproducibility of results has become an issue. In particular, MaRDI will focus on the interlinkage of mathematical results, software, and mathematical data with the aim to guarantee confirmability and reproducibility.

3. **Establish confirmable workflows**: In mathematics, the confirmability of scientific results is primarily associated with proofs, however, since the advent of computer-aided mathematics and scientific computing, the entire workflow needs to be documented, including types and versions of software used, program code, intermediate results, and many more. MaRDI aims to establish tools and standards for making this workflow transparent and reliable.

4. **Enable the development of mathematical services**: Services need to be developed that make the creation of FAIR research data easy and attractive for the user. A central element in this service orientation is the creation of an easy to access agile digital portal that gathers the majority of MaRDI services.

5. **Establish next-generation peer review**: Correctness of computational methods, program code, and software cannot be ensured by traditional means of peer review. MaRDI aims to establish standards and requirements for ensuring the correctness of these research results.

6. **Standardize semantic relations**: Mathematical research data is used in many disciplines and is the foundation of cross-disciplinary methodologies, mathematical modeling and simulation (MMS) as well as statistics and data science. MaRDI establishes a common ground between the disciplines concerning standards for models, software and other data. These standards link semantically related research data and thereby guarantee findability.

7. **Culture development and community embedding**: The long-term goal is the recognition of FAIR mathematical research data as an accomplishment in its own right. This includes establishing standards, services and workflows for mathematical research data by embedding into the mathematical community, training of mathematicians in good practices of data handling and by broad dissemination through conferences and publications.

The three research motivated pillars of this proposal are based on the division of mathematical research data in *exact and symbolic data* (**T1:** Computer Algebra), *floating point data* (**T2:** Scientific Computing), and *data with uncertainty* (**T3:** Statistics). These task areas target certified data and software developments as well as confirmable workflows in their respective realm, and they work toward respective service prototypes. Mathematical research data also include mathematical models as entities with model for machine learning as special case. Moreover, the interdisciplinary power of mathematics leads to several use cases and cooperations with other disciplines and their respective NFDI consortial initiatives (**T4:** Cooperation with other disciplines).

The prototypical services will be expanded into full services which provide added value to current research. Another important MaRDI component is the development of significantly enhanced information retrieval services, including mathematical models as research data, a cross-disciplinary mathematical digital semantic atlas, ontologies, a mathematical knowledge graph and metadata. MaRDI will develop a digital service portal as a one-stop unique and easy to access contact point for the scientific community to retrieve and consult MaRDI services (**T5:** The MaRDI-Portal). This Portal will be developed in an agile fashion and installed permanently.

The sustainable realization of the findings of MaRDI requires a community which adheres to a FAIR data culture and FAIR research workflows. For this, MaRDI will build and enable collaborative platforms which play an important role in disseminating knowledge, pursuing the scientific discourse and ensuring quality control (**T6:** Data Culture and Community Integration).

In the long term MaRDI aims at securing and advancing the FAIR principles in the mathematical sciences and in disciplines relying on mathematically quantified research. Thereby, it will contain agile task areas reflecting the latest challenges in mathematical research data. These areas will advance

their agenda towards introducing prototypes which can then be transferred into services within the MaRDI portal. The digital portal is designed to become a permanent one-stop contact point for the scientific community and within the NFDI for services concerning mathematical research data.

Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives (Most important)

The **Digital Library of Mathematical Functions** provides machine-readable semantics that allows for formulae search and interactive display of additional metadata. This includes links to definitions for the symbols and identifiers used in the formula, references to proofs and sketches of proofs when proofs are not available on the literature, as well as hyperlinks to related concepts.

MathWebSearch is a mathematical formula search engine for the semantically querying mathematical documents with content MathML markup, e.g., generated from LaTeX. The system has been in active use in mathematical information systems like zbMATH and arXiv.org.

The **MORwiki** has, by 2019, collected sixty interactive wiki article pages on model order reduction written by more than fifty contributors. At the heart of the MORwiki are three main sections: Benchmarks, methods (algorithms) and software.

The **Open Machine Learning project** is an open science project to build an open, organized, online ecosystem for machine learning. It builds open source tools to discover and share open data from any domain, connects to various popular ML frameworks and supports reproducibility and sharing of results.

The **OSCAR** project develops a comprehensive open source computer algebra system for computations in algebra, geometry, and number theory. In particular, the emphasis is on supporting complex computations which require a high level of integration of tools from different mathematical areas.

polymake deals with polytopes, polyhedra and fans as well as simplicial complexes, matroids, graphs, tropical hypersurfaces, toric varieties and other objects. Two key design features are particularly relevant to MaRDI: the extendible polymake type system is serialized and formalized; there is a built-in interface to the database **polyDB**.

pyMOR is an open source software library for building model order reduction applications with the Python programming language. Tools and computing infrastructure are developed that enable cloud-based scientific computing with pyMOR and other web services.

RADAR is a not-for-profit and discipline-agnostic research data repository, which guarantees the availability of published research data for at least 25 years. It provides a generic and interoperable metadata schema which can be complemented with discipline-specific information. The repository has a Core Trust Seal certification and is listed on re3data.

The **Small Groups library** provides access to descriptions of the groups of small order. This includes, e.g., all 423 164 062 groups of order at most 2000 (except 1024).

The **swMATH** database, maintained by FIZ and ZIB, establishes a connection between scientific publications and mathematical software. It provides software metadata and semantic information such as links to the home pages, the Internet Archive, licensing terms, versions, MSC classifications, authors, as well as software usage and citations in publications.

zbMATH, edited by the EMS, FIZ, and the Heidelberg Academy of Sciences, is the world's most comprehensive and longest-running abstracting and reviewing service in mathematics, which covers the complete research literature since 1868 by the effort of currently more than 7,000 mathematicians worldwide. zbMATH has recently been transformed to an information system providing open services, data and API.

Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

MaRDI and NFDIMatWerk will collaborate with respect to the development and usage of representations of mathematical objects, in particular with respect to mathematical multi-scale modelling and as a solid foundation for the ontologies developed at NFDIMatWerk. We will connect experimental and simulation data of real materials from NFDIMatWerk with mathematical models and meta data in order to established confirmable workflows. Ideally, this will be done for use cases that build upon best practise examples from NFDIMatWerk.

MaRDI and PUNCH4NFDI will collaborate with respect to data integration and annotation with meta data in order to make experimental data from PUNCH4NFDI accessible for data analysis with mathematical tools, such as machine learning or statistical data analysis. We will also work on showcases to use mathematical models and confirmable workflows for data based simulation in PUNCH4NFDI. The consortia will also collaborate with respect to building up a cross cutting platform for sustainable development of research software.

MaRDI and FAIRmat will collaborate on metadata. FAIRmat comprises the NOMAD repository, where computational data in materials science, resulting, e.g., from density functional theory (DFT) are collected. It is planned to harmonize metadata generation for Use Cases developed in both consortia (e.g., on battery materials).

MaRDI, NFDI4Ing, and NFDI4Chem will closely collaborate, including but not limited to reproducible science, the sharing of mathematical models, the generation and description of input data sets from experiments and measurements, and simulation software developed for analysis and quantifiable predictions.

MaRDI and NFDI4RSE will collaborate in the area of software carpentries and the curation of the swMATH database. MaRDI will contribute to the pool of carpentry instructors and develop course programs for mathematical research software engineering with a focus on confirmable workflows and data APIs.

MaRDI and BERD@NFDI aim at advancing machine learning approaches and at enabling researchers to successfully apply them to specific research questions. While MaRDI contributes its expertise in algorithms, their implementation and empirical benchmarks, BERD@NFDI focuses on the application of these tools to data and research questions in economics and social sciences. Another field of collaboration is the assessment of data quality in heterogeneous and unstructured data.

MaRDI and NFDI-Neuro will collaborate on signal processing and image analysis. Further collaborations are currently evaluated with NFDI4Cat, NFDI4MobileTech, and NFDI4hpc.

4 Cross-cutting topics

Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.

- Authentication and Authorisation Infrastructure (AAI)
- Cloud-based Data Processing
- Knowledge Graph and Semantic Technologies
- Federated Repositories
- Metadata for Scientific Software, Workflows, Computer-Experiments
- Journal Integration / Next-Generation Peer Review
- Author Identification (Data, Software, Publications)
- Data Culture and Training

Please indicate which of these cross-cutting topics your consortium could contribute to and how.

MaRDI will provide the MaRDI-Portal **T5** as a one-stop unique and easy to access contact point for the scientific community to retrieve and consult specific MaRDI services, which will be distributed among MaRDI partners. The heterogeneous nature of the involved mathematical data requires viable Federated Repository infrastructure involving sophisticated data descriptions.

The reproducibility crisis and the ever-growing amount of data and use of computational methods in science as well as mathematics has lead to the urgent need of establishing Next-Generation Peer Review and provide guidelines for Journal Integration. MaRDI will contribute to this topic in collaboration with a number of mathematical scientific journals especially in **T1** and **T3** as prototype before working on distribution widely within the mathematical publication community.

Metadata for Scientific Software, Workflows, Computer-Experiments is key for the reproducibility of scientific results especially for computed-assisted mathematics. Therefore, MaRDI will dedicate a lot of effort in the definition of suitable metadata for the related research data objects. This will not only allow for the reproducibility of scientific workflows but also for the re-usabilty of established models and processing pipelines in **T1-4**.

The Knowledge Graph and Semantic Technologies are not only important for the distribution of mathematical knowledge within the mathematical community but also within the other disciplines that rely on mathematical objects, models, workflows, and software. MaRDI will therefore build-up an agile knowledge graph to provide reliable knowledge sources and prevent scientific project from re-invention of the wheel. The MaRDI partners already established a number of services that will be integrated into the knowledge graph and that will be build upon.

Last, but not least, a cross-cutting concern is the training of the research community in the NFDI standards, data sets, services, and workflows to establish a mathematical research Data Culture **T6**.