HeFDI Data Talk

Date	Topic	Presenter
02. December 2022	MaRDI4NFDI. Mathematical Research Data Initiative	Tabea Bacher (MaRDI; Max-Planck-Institut for Mathematics in the Sciences [MPI MiS])

Abstract:

Research data in mathematics come in many different forms. The Mathematical Research Data Initiative MaRDI, the consortium for mathematics in the National Research Data Infrastructure (NFDI), is working with the scientific community to develop an infrastructure for the FAIR handling of these diverse mathematical research data. This presentation describes the structure of the consortium, services under construction and current developments.

About the HeFDI Data Talks:

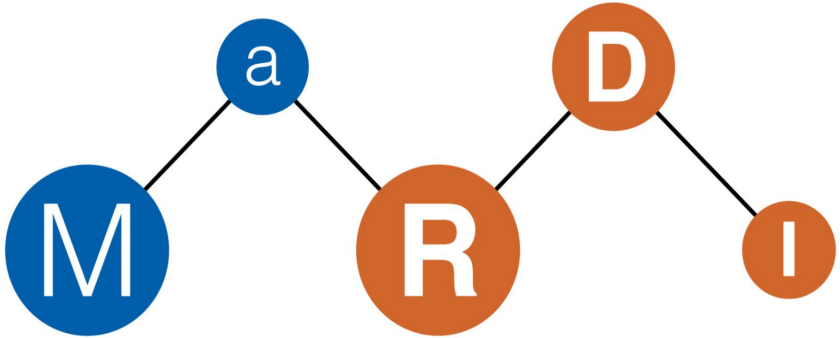
The HeFDI Data Talks are a bi-weekly open information and discussion event focused on data management in the context of science, in which relevant NFDI consortia as well as research data management services present themselves. The series discusses current topics and presents numerous – including local and regional – tools and services. The HeFDI Data Talks are an offer of the HeFDI Initiative (Landesinitiative HeFDI), which is funded by Hesse's Ministry for Science and Arts (HMWK).

DOI: https://doi.org/10.5281/zenodo.7505425

gefördert durch Hessisches Ministerium für Wissenschaft und Kunst

MATHEMATICAL RESEARCH DATA INITIATIVE

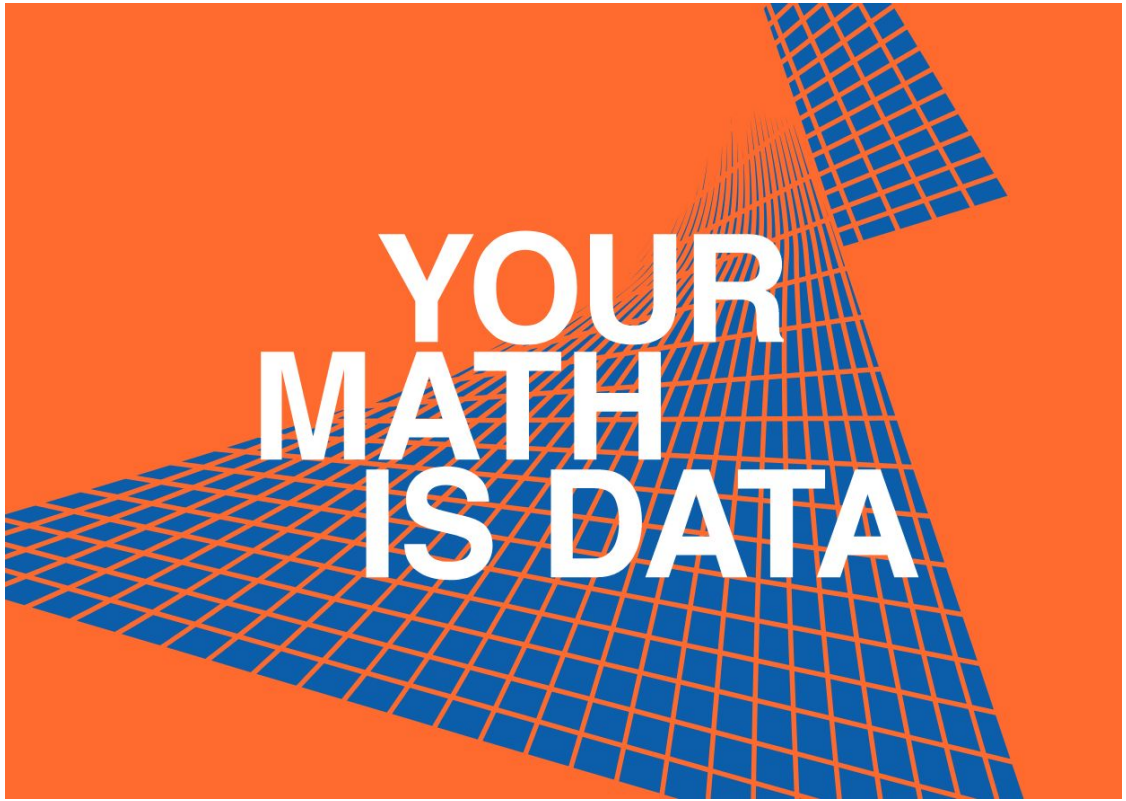
Leipzig, 2022-12-02



Tabea Bacher
MPI Mathematics in the Sciences

MAX-PLANCK-INSTITUT
FÜR MATHEMATIK
IN DEN NATURWISSENSCHAFTEN

What is mathematical research data?



What is research data?

“The recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”

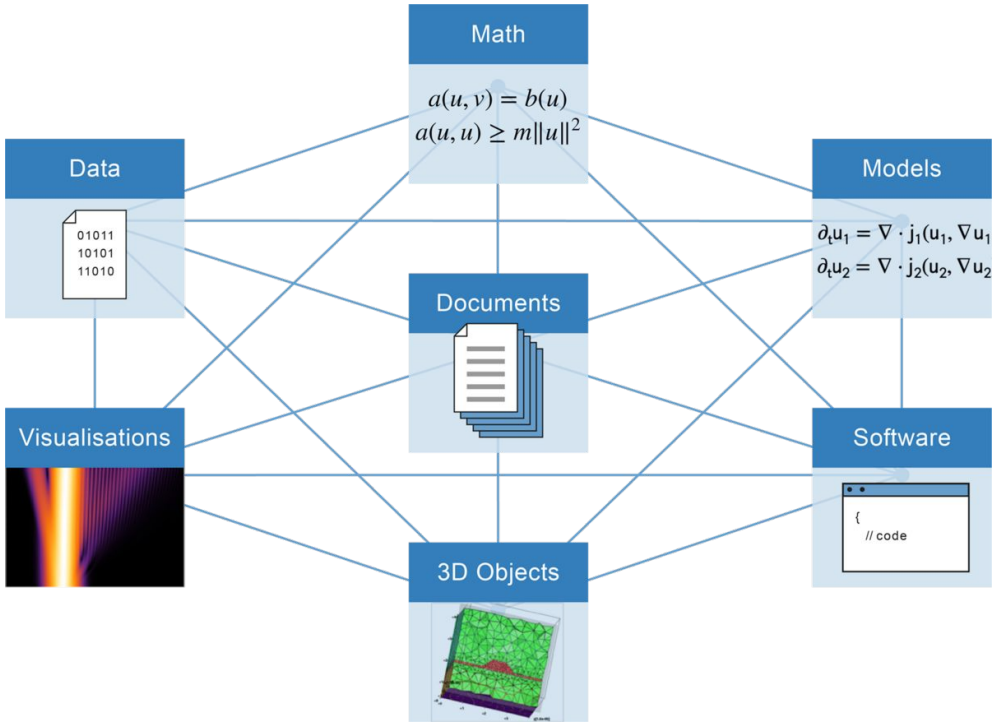
“Alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihre Ergebnisse sind.”

This is much broader than data alone!

https://www.ukri.org/about-us/epsrc/our-policies-and-standards/policy-framework-on-research-data/scope-and-benefits/

https://www.forschungsdaten.info/themen/informieren-und-planen/was-sind-forschungsdaten/ und Forschungsdaten Definition: Kindling, Maxi und Schirmbacher, Peter: „Die digitale Forschungswelt“ als Gegenstand der Forschung. Information – Wissenschaft – Praxis 64 (2013): S. 130. doi.org/10.1515/iwp-2013-001

Research data in mathematics



- mathematical documents: papers, proofs, formulae,...
- notebooks, domain-specific research-software packages and libraries, computer algebra systems, programmes, scripts
- simulation data
- formalised mathematics
- collections of mathematical objects
- mathematical models
- …

Research data in mathematics

- “in contrast, for instance, to the life sciences, where older results can be overruled by new evidence, mathematical results that have been proven true remain true indefinitely.” *
- other disciplines using mathematical research data brings responsibility to preserve results in a sustainable manner *

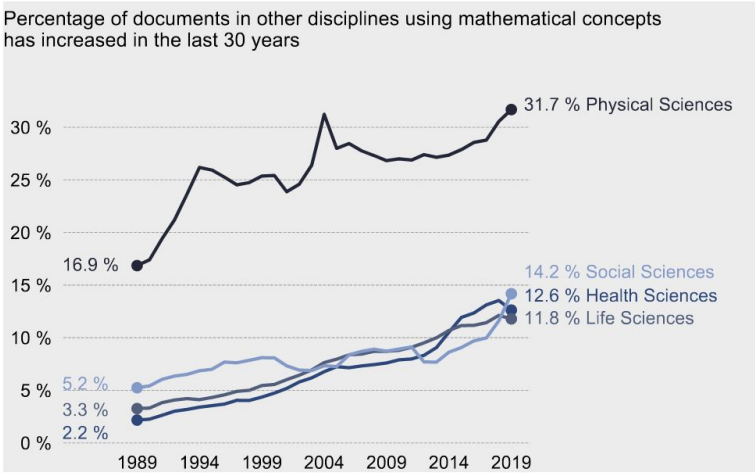
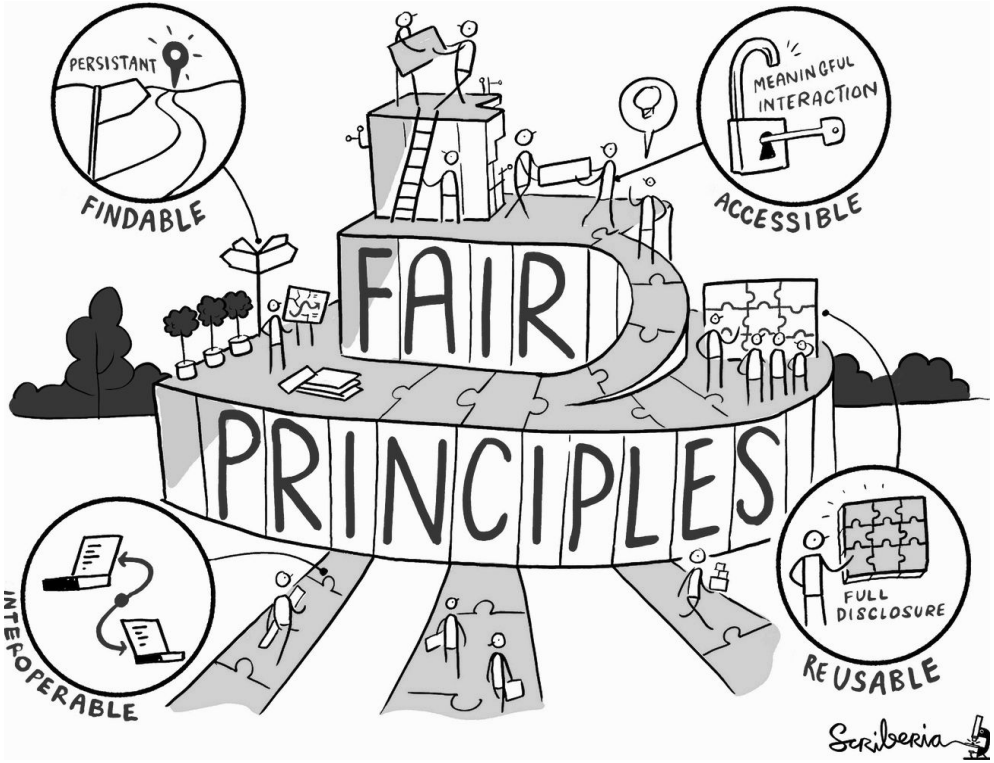


Figure 1: Percentage of peer-reviewed publications using mathematical concepts compared to the total number in each subject area excluding mathematics itself based on a Scopus query using mathematical keywords. For details see [SG]. MaRDI-Proposal https://zenodo.org/record/6552436

*T.Boege, R. Fritze, C. Görgen et al. (2022) Research-Data Management Planning in the German Mathematical Community. arXiv:2211.12071 [math.HO]

FAIR research data



Mark Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gaby Appleton, et al. The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3(160018), 2016.

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, et al. FAIR principles: Interpretations and implementation considerations. Data Intelligence, 2(1-2):10–29, 2020.

FAIR research data in mathematics?

Status quo:

- results in papers depend on software; the paper is peer-reviewed, the software not
- knowledge about algorithms (implementations, state of the art, publications) not available in one place
- missing benchmarks to compare algorithms and methods
- non-standardized workflows in interdisciplinary mathematics
- research data which was promised in papers and stored on long-gone personal homepages
- …

A lot of implicit knowledge and sometimes big hurdles to build on other people's research!

MaRDI – the Mathematical Research Data Initiative

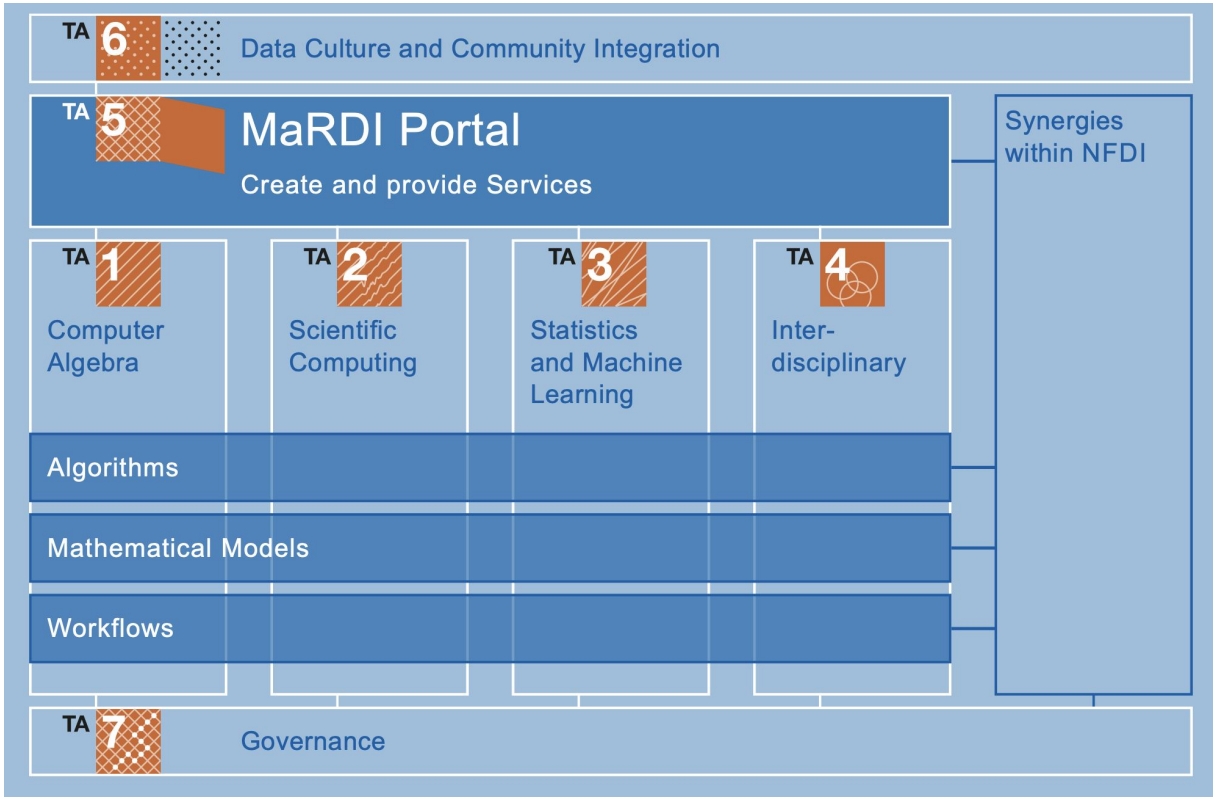
- 1 out of up to 30 NFDI consortia
- the one consortium of mathematics
- 15 institutions and partners
- kickoff November 2021
- 28 (full-time equivalent) employees
- funding over five years

Funded by the Deutsche Forschungsgemeinschaft (DFG),
Project number 460135501.
NFDI 29/1 "MaRDI – Mathematische Forschungsdateninitiative"

Task areas inside the consortium

“MaRDI will have a unique twofold function within the mathematics community – as a quality-controlled mathematical research-data library and as a digital service portal at the same time.”

Ilka Agricola, president of the German mathematical union, 2021

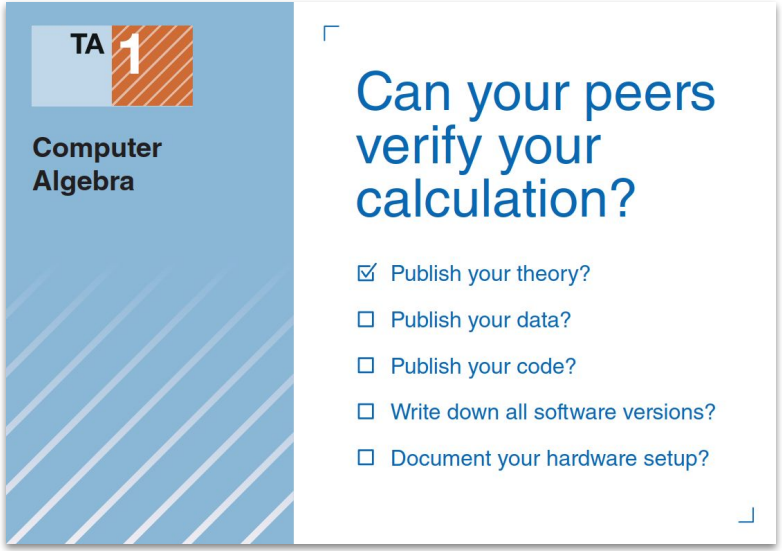


TA1: Computer Algebra

Exact and symbolic data

Services under development:

- **Confirmable workflows for computer algebra**
  - Best practices, guidelines, checklists
  - https://portal.mardi4nfdi.de/wiki/Portal/T1/guidelines/authors
- **Technical support for publishers and journals for a refereeing process for software and datasets**



TA1: Computer Algebra

Title: Paper about mathematics
Author(s): Claus Fieker, Max Horn
Reviewer: Jeroen Hanselman
Date: March 3, 2022

Technical review

Basic Info

Files provided

- ☐ Source Code
- ☐ Notebook
- ☐ Examples
- ☐ Docker file/VM
- ☐ Documentation
- ☐ Computed data
- ☐ Files that verify computed data

Programming languages:	Julia version 1.7.1
Standard software used:	Oscar version 0.7.2-DEV
Version reviewed:	MyMath Program v1.1.2
Downloaded from:	github.com/JHanselman

Importance of software in the paper

The results of the paper depend heavily on computations.
Score: ●●●●●

Reproducibility (Installation)

License:	+ Yes, Open Source, Creative Commons v4.0
Availability:	+ Code is available on the author's Github
Ease of installation:	+ It takes less than 5 minutes to install the program and let it run.
Dependence on other packages:	+ The software uses less than 3 other packages.
Score:	●●●●●

Reproducibility (Records of setup)

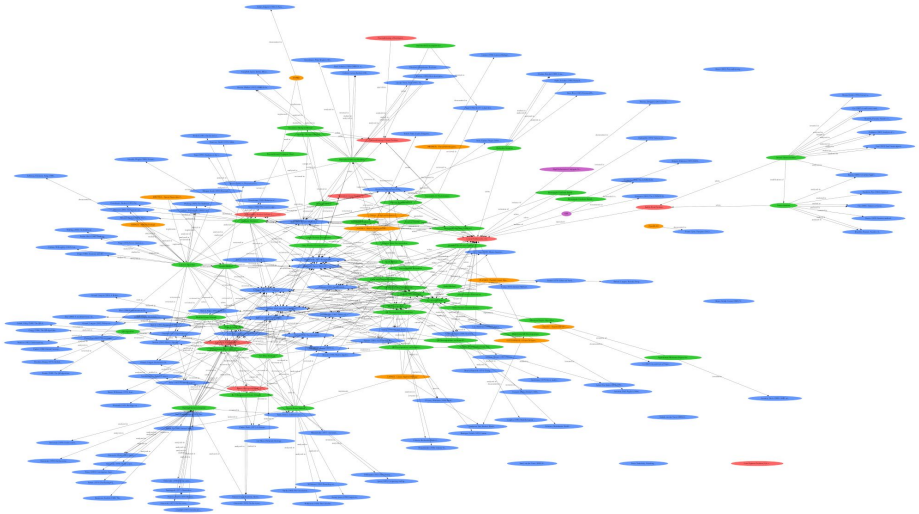
Specification of CPU:	+ Yes
Specification of Memory:	+ Yes
Specification of OS/software used:	+ Yes, including version numbers of all software involved.
References and citation:	− The software depends on software that was not referenced in the paper.
Score:	●●●●○

TA2: Scientific Computing

Floating point data

Services under development:

- **Benchmark Framework: MaRDI Mark**
  - standardized way to compare algorithms
- **Knowledge Graph of Numerical Algorithms**
  - The F in FAIR



TA2: Scientific Computing

https://algodata.mardi4nfdi.de DEMO

TA3: Statistics and Machine Learning

Data with uncertainty

Services under development:

- **Library of Curated Benchmark Datasets**
  - to illustrate and test new methods
  - with rich metadata and well-selected
- **Library of Statistical Analyses**
  - play the role of demos
  - link to literature describing the considered methods and software

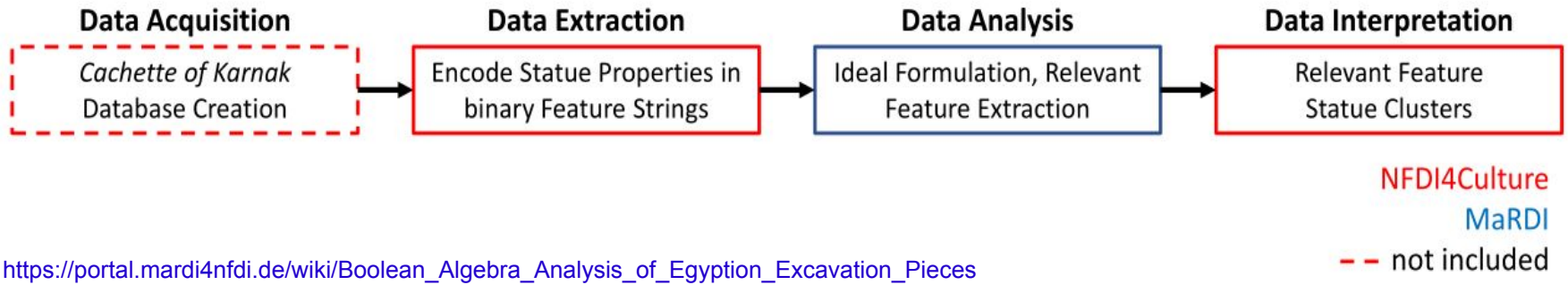
TA4: Cooperation with Other Disciplines

Data from other disciplines analyzed using mathematical methods

case studies with other disciplines

TA4: Cooperation with Other Disciplines

An example for an easy workflow:



https://portal.mardi4nfdi.de/wiki/Boolean_Algebra_Analysis_of_Egyption_Excavation_Pieces

Working Program:

- documentation and analysis of interdisciplinary workflows
  - https://portal.mardi4nfdi.de/wiki/Category:Workflow
- standardization of mathematical descriptions across disciplines
- develop ontology and align to other TAs knowledge graphs

TA5: The MaRDI Portal

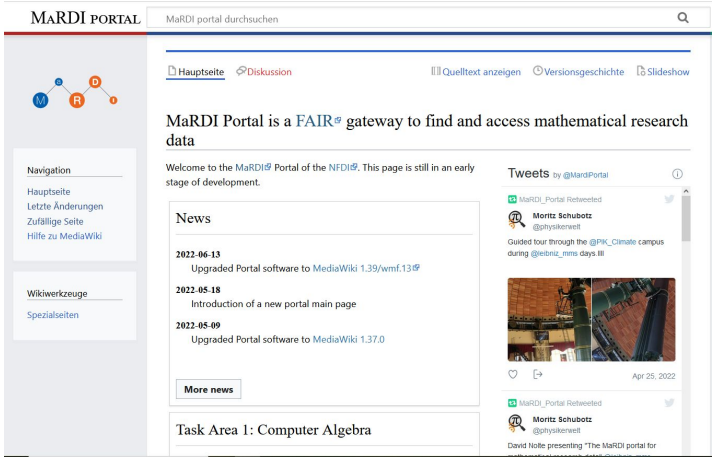
Vision:

- a one-stop contact point for mathematical research data for the scientific community
- portal points to MaRDI services developed in other TAs
- planned to rely on wikidata - in agreement with other NDFI consortia
- ultimate goal: One NFDI Portal

Status quo:

- still in an early stage of development
- integrating external databases
- first implementation of a formula search

portal.mardi4nfdi.de



TA6: Data Culture and Community Integration

UNIVERSITÄT LEIPZIG

MAX-PLANCK-INSTITUT FÜR MATHEMATIK IN DEN NATURWISSENSCHAFTEN

IMAGINARY open mathematics

Mathematisches Forschungsinstitut Oberwolfach

DMV

Raise awareness, build a community, dissemination and training

Target groups: mathematicians (from any field), information specialists, general public

- interactive Talks on Mathematical Research Data
- survey, publications
- presence at conferences and workshops
- design of outreach material [newsletter, interview series, movies]
- Help desk: RDMPs, bring your data

www.mardi4nfdi.de/community

TA7: Governance and Consortium Management

- build up of the internal consortial infrastructure
- participation in the creation of an NFDI Basisdienste-Konsortium
- connect to other consortia in the NFDI

MaRDI up until now

- joint article on RDMPs in Mathematics
- joint publications in DMV Mitteilungen and the GAMM Newsletter: www.mardi4nfdi.de/resources/publications
- several community workshops and events in 2022

- quarterly Newsletter with user stories and interviews
  - subscribe: https://t1p.de/ewmt6
- Making MaRDI series interviews MaRDI employees in their work and FAIR research data

MaRDI in the future

- MaRDI-Workshop on Data in Discrete Math, March 2023 in Leipzig
  - https://www.mis.mpg.de/calendar/conferences/2023/dataindiscretemath.html
- DMV Jahrestagung, September 2023 in Ilmenau
- MaRDI-Workshop “MaRDI meets Libraries” 2023, tba
- summer school 2024

www.mardi4nfdi.de

Get in Contact

Tabea Bacher

MaRDI Dissemination Coordinator

bacher@mardi4nfdi.de

Max Planck Institute for Mathematics in the Sciences,

Leipzig

Phone: 0341 9959 705

Newsletter https://t1p.de/ewmt6

www.mardi4nfdi.de