

MathRepo

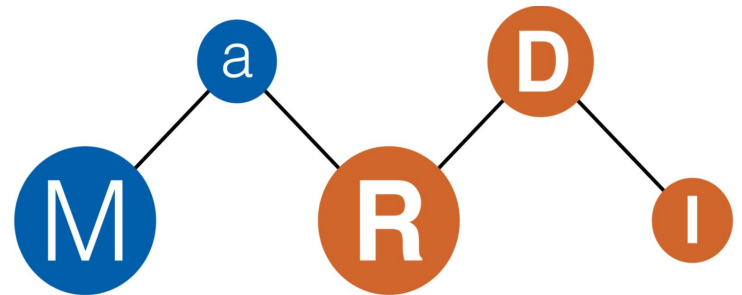
an example of a mathematical research-data repository

DR. CHRISTIANE GÖRGEN

DMV Jahrestagung
September 14, 2022



UNIVERSITÄT
LEIPZIG



National research-data infrastructure NFDI

Some facts



- 90 mio euros funding volume
- 20-30 consortia
- 5-10 years
- aim: make research data FAIR

- management guidelines, data formats, workflows, knowledge graphs, ...
- 'joint solutions for individual problems'
- 'culture change'

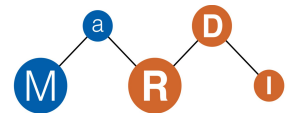
But: no hardware infrastructure.

→ **Where should you store your research data?**

Hartl, N., Wössner, E. & Sure-Vetter, Y. Nationale Forschungsdateninfrastruktur (NFDI). *Informatik Spektrum* **44**, 370–373 (2021). <https://doi.org/10.1007/s00287-021-01392-6>



UNIVERSITÄT
LEIPZIG

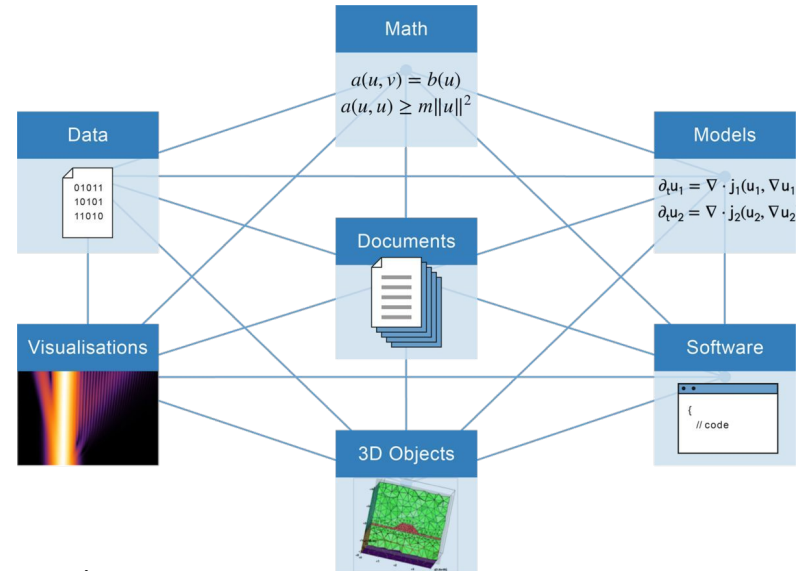


MaRDI and mathematical research data

- comes in many different forms
- hence many different formats
- with many different needs

some obvious challenges but also

- relevance?
- should be FAIR
- annotated with metadata (which?)

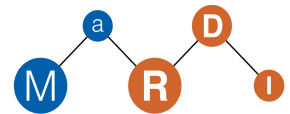


no island/decentralised solutions (no personal homepages)

→ but **good** repositories... what does good mean?



UNIVERSITÄT
LEIPZIG



What makes a ‘good’ repository?

- trustworthy storage (known institution?)
- appropriate audience
- long term

‘Researchers must be certain that data held in archives remain useful and meaningful into the future.’

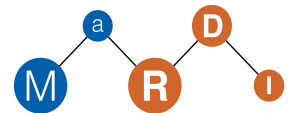
- information about services
- data protection and curation
- access/license/policy statements
- persistent identifier (citability)
- external certification



CoreTrustSeal Standards and Certification Board (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 <https://doi.org/10.5281/zenodo.3638211>

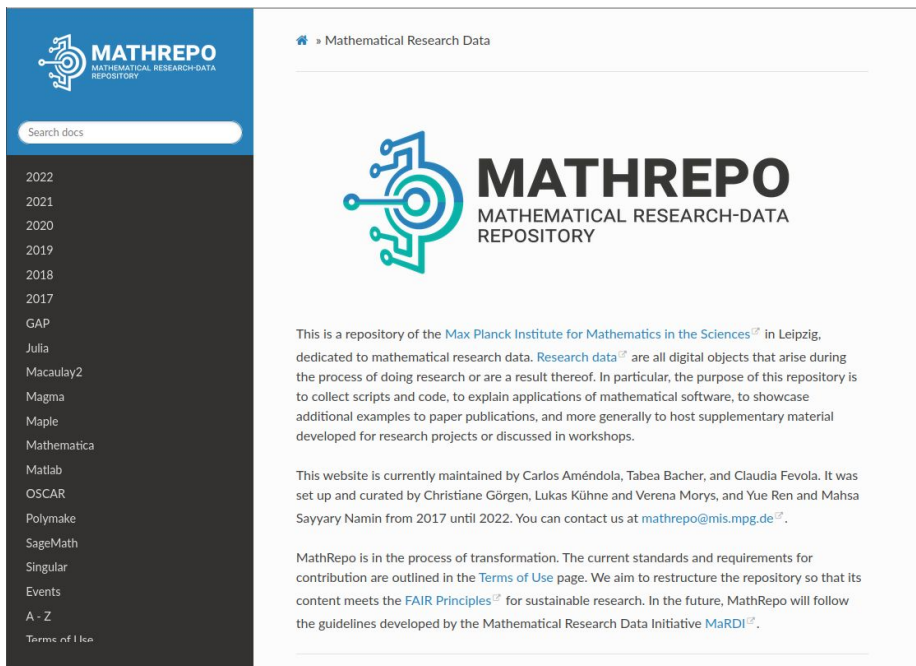


UNIVERSITÄT
LEIPZIG




One maths-specific solution

<http://mathrepo.mis.mpg.de>



» Mathematical Research Data



MATHREPO
MATHEMATICAL RESEARCH-DATA
REPOSITORY

This is a repository of the [Max Planck Institute for Mathematics in the Sciences](#)¹ in Leipzig, dedicated to mathematical research data. [Research data](#)² are all digital objects that arise during the process of doing research or are a result thereof. In particular, the purpose of this repository is to collect scripts and code, to explain applications of mathematical software, to showcase additional examples to paper publications, and more generally to host supplementary material developed for research projects or discussed in workshops.

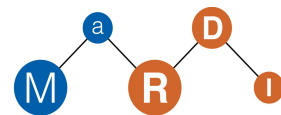
This website is currently maintained by Carlos Améndola, Tabea Bacher, and Claudia Fevola. It was set up and curated by Christiane Görden, Lukas Kühne and Verena Morys, and Yue Ren and Mahsa Sayyary Namin from 2017 until 2022. You can contact us at mathrepo@mis.mpg.de³.

MathRepo is in the process of transformation. The current standards and requirements for contribution are outlined in the [Terms of Use](#) page. We aim to restructure the repository so that its content meets the [FAIR Principles](#)⁴ for sustainable research. In the future, MathRepo will follow the guidelines developed by the Mathematical Research Data Initiative [MaRDI](#)⁵.

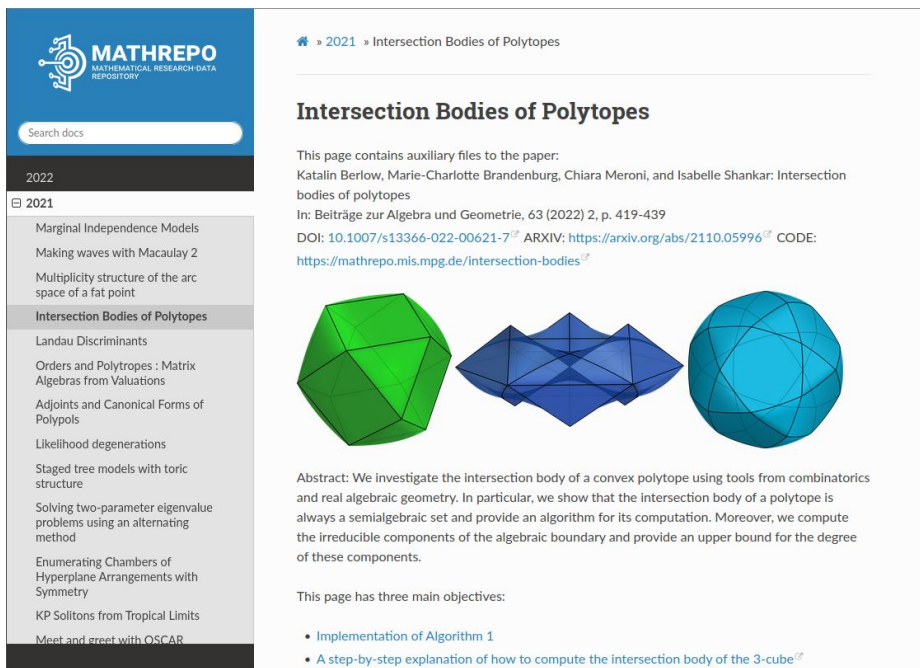
- established 2017 at MPI MiS
- hosted by the MPG
- underlying gitlab
- maintained by postdocs
- some 50 contributions
- maths foremostly non-linear algebra
- sortable alphabetically, by year, and by programming language
- three main types: code for computations, additional examples to papers, material for workshops/teaching



UNIVERSITÄT
LEIPZIG



A typical contribution




The screenshot shows the MathRepo website interface. At the top left is the MathRepo logo and a search bar. A sidebar on the left lists various mathematical topics, with 'Intersection Bodies of Polytopes' highlighted. The main content area displays the paper title, authors (Katalin Berlow, Marie-Charlotte Brandenburg, Chiara Meroni, and Isabelle Shankar), and publication details. Below the text are three 3D polytope visualizations: a green cube, a blue complex polytope, and a blue sphere with internal lines. An abstract and a list of objectives are provided at the bottom of the page.

» 2021 » Intersection Bodies of Polytopes

Intersection Bodies of Polytopes

This page contains auxiliary files to the paper:
Katalin Berlow, Marie-Charlotte Brandenburg, Chiara Meroni, and Isabelle Shankar: Intersection bodies of polytopes
In: Beiträge zur Algebra und Geometrie, 63 (2022) 2, p. 419-439
DOI: [10.1007/s13366-022-00621-7](https://doi.org/10.1007/s13366-022-00621-7) ARXIV: <https://arxiv.org/abs/2110.05996> CODE: <https://mathrepo.mis.mpg.de/intersection-bodies>



Abstract: We investigate the intersection body of a convex polytope using tools from combinatorics and real algebraic geometry. In particular, we show that the intersection body of a polytope is always a semialgebraic set and provide an algorithm for its computation. Moreover, we compute the irreducible components of the algebraic boundary and provide an upper bound for the degree of these components.

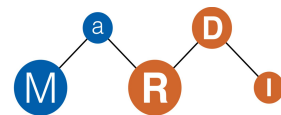
This page has three main objectives:

- [Implementation of Algorithm 1](#)
- [A step-by-step explanation of how to compute the intersection body of the 3-cube](#)

- supplementary material to a paper
- doi/arxiv link to paper, authors' names, dates, software and version
- subpages with downloadable code
- and interactive visualisations
- extra explanatory text
- blog-style entry
- MathRepo page cited in paper as URL



UNIVERSITÄT
LEIPZIG



My most recent research project

Staged tree models with toric structure

- Macaulay2 code and examples
- Mathematica code
- Mathematica discussion and examples

Solving two-parameter eigenvalue problems using an alternating method

Enumerating Chambers of Hyperplane Arrangements with Symmetry

KP Solitons from Tropical Limits

Meet and greet with OSCAR

Primary Decomposition with Differential Operators

The Gaussian entropy map in valued fields

2020

2019

2018

2017

GAP

Julia

Macaulay2

Magma

Maple

Mathematica

Matlab

OSCAR

» 2021 » Staged tree models with toric structure

Staged tree models with toric structure

This page presents the code cited in the final section of the article:
Christiane Görgen, Aida Maraj, and Lisa Nicklasson: Staged tree models with toric structure
In: Journal of symbolic computation, 113 (2022), p. 242-268
DOI: [10.1016/j.jsc.2022.04.006](https://doi.org/10.1016/j.jsc.2022.04.006) ARXIV: <https://arxiv.org/abs/2107.04516> CODE: <https://mathrepo.mis.mpg.de/StagedTreesWithToricStructures>

The theoretical results of this work explore the algebraic geometry of discrete statistical models for asymmetric conditional independence relationships that can be represented by staged trees. Background literature and information on the applications of these models is cited in the article. Sometimes, as explored in previous work and enhanced in various new technical results, these statistical models are toric varieties intersected with the probability simplex. Whilst in special cases the toric structure is apparent from the tree graph representation itself, oftentimes it is not. We here present two computational approaches to complement our theoretical work.

The first approach is based on the software **Macaulay2**. It provides a simple tool which allows the user to input a staged-tree parametrisation and output its ideal of model invariants. The given staged tree has toric structure whenever this ideal is generated by binomials.

- [Macaulay2 code and examples](#)

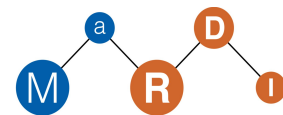
The second approach uses the software **Mathematica**. It provides functions which randomly generate transformations of a given model specification and try out whether these give rise to binomial generators as above. The code is presented in the first subpage below and discussed and applied in examples listed in the second.

- [Mathematica code](#)

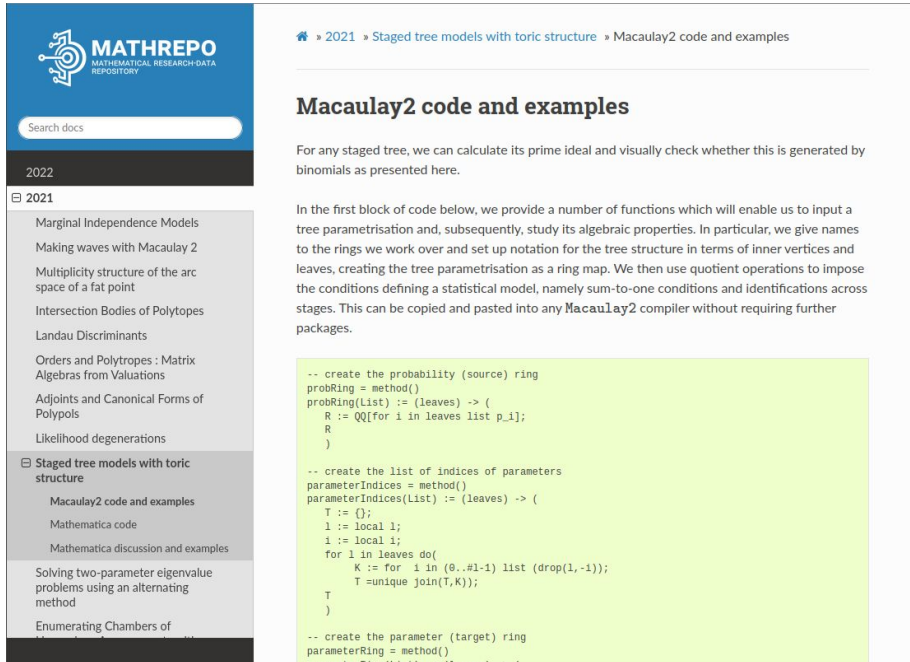
- Algebraic statistics: connect discrete statistical models with algebraic varieties, study their properties
- ‘pure maths’ paper
- supplementary code
 - proofs
 - extra examples
 - conjectures
- relevant links to publication
- expository text (not the abstract)



UNIVERSITÄT
LEIPZIG



Providing code for proofs



The screenshot shows the MATHREPO website interface. At the top left is the MATHREPO logo (Mathematical Research Data Repository). Below it is a search bar and a navigation menu for the year 2021. The main content area is titled 'Macaulay2 code and examples' and includes a breadcrumb trail: » 2021 » Staged tree models with toric structure » Macaulay2 code and examples. The text below the title states: 'For any staged tree, we can calculate its prime ideal and visually check whether this is generated by binomials as presented here.' This is followed by a paragraph explaining the code's purpose: 'In the first block of code below, we provide a number of functions which will enable us to input a tree parametrisation and, subsequently, study its algebraic properties. In particular, we give names to the rings we work over and set up notation for the tree structure in terms of inner vertices and leaves, creating the tree parametrisation as a ring map. We then use quotient operations to impose the conditions defining a statistical model, namely sum-to-one conditions and identifications across stages. This can be copied and pasted into any Macaulay2 compiler without requiring further packages.' A code block follows, containing Macaulay2 code for creating a probability ring, parameter indices, and a target ring.

```
-- create the probability (source) ring
probRing = method()
probRing(List) := (leaves) -> (
  R := QQ[for i in leaves list p_i];
  R
)

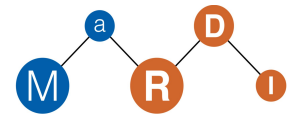
-- create the list of indices of parameters
parameterIndices = method()
parameterIndices(List) := (leaves) -> (
  T := {};
  l := local i;
  i := local i;
  for l in leaves do(
    K := for i in (0..#l-1) list (drop(l,-1));
    T = unique join(T,K);
  )
)

-- create the parameter (target) ring
parameterRing = method()
```


- how did we come up with (re-)parametrisations we found?
- function specification, extra libraries, M2 version,...
- extra example plus picture
- copy&paste code from the screen
- verify our results and reuse them!



UNIVERSITÄT
LEIPZIG



Providing code for conjectures



MATHREPO
MATHEMATICAL RESEARCH DATA
REPOSITORY

Search docs

2022

☰ 2021

- Marginal Independence Models
- Making waves with Macaulay 2
- Multiplicity structure of the arc space of a fat point
- Intersection Bodies of Polytopes
- Landau Discriminants
- Orders and Polytopes : Matrix Algebras from Valuations
- Adjoints and Canonical Forms of Polytops
- Likelihood degenerations
- ☰ Staged tree models with toric structure
 - Macaulay2 code and examples
 - Mathematica code
 - Mathematica discussion and examples
- Solving two-parameter eigenvalue problems using an alternating method
- Enumerating Chambers of

Example 1: a square-free and stratified balanced staged tree

Consider a staged tree representing a full independence model between a binary and a ternary random variable. This can be characterised using the equations

$$p_1 p_5 = p_4 p_2, (p_1 + p_2) p_6 = p_3 (p_4 + p_5).$$

Its ideal of model invariants is known to be generated by binomials in the p -variables using the concept of path extensions from Duarte and G\"{o}rger (2018). We run the following piece of code to check our computations and verify this result.

```
lineartransform[{{p1, p3}, {p2, p4}}, {{p1 + p2, p4 + p5}, {p3, p6}},
6, ConstantArray[Range[-1, 1], Range[-1, 1]], 2], 100000]
```

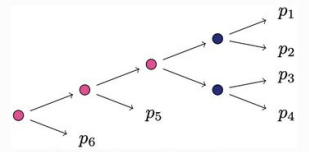
In a hundred thousand random trials this found about fifty transformations in new variables, resulting for instance in the reparametrisation

$$q_1 = -p_3 - p_4, q_2 = p_1 + p_2, q_3 = -p_3, q_4 = p_1, q_5 = -p_1 - p_2 - p_4 - p_5, q_6 = p_3 + p_6$$

which gives binomial generators of the staged tree's prime ideal.

Example 2a: a non-squarefree, toric staged tree

The non-squarefree staged tree below fulfils the subtree-inclusion property and is hence toric.



We can look for a reparametrisation of the defining equations

$$p_1 p_4 = p_3 p_2, (p_1 + p_2) p_5 = (p_1 + p_2 + p_3 + p_4)(p_3 + p_4),$$

$$(p_1 + p_2) p_6 = (p_1 + p_2 + p_3 + p_4 + p_5)(p_3 + p_4),$$

and

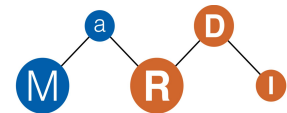
$$(p_1 + p_2 + p_3 + p_4) p_6 = (p_1 + p_2 + p_3 + p_4 + p_5) p_5$$

using the following command.

- how did we come up with conjectures?
- randomised component, yes/no existence questions
- separate page explaining individual functions in the code
- lots of examples to play around with
- all tools available to continue our conjectures, come up with your own



UNIVERSITÄT
LEIPZIG



How FAIR is this?

Findable

- ✓ URL stable for the foreseeable future
- ✓ git and MR findable in google search e.g.
- ✓ known academic institution / group
- ✗ no DOI
- ✗ no indexing or field classification

Interoperable

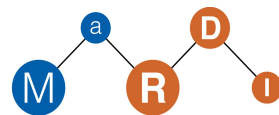
- ✓ maths with explanatory text, link to articles
- ✓ code annotated with software name, version, date, hardware setup
- ✗ very case specific, reproducibility might fail

Accessible

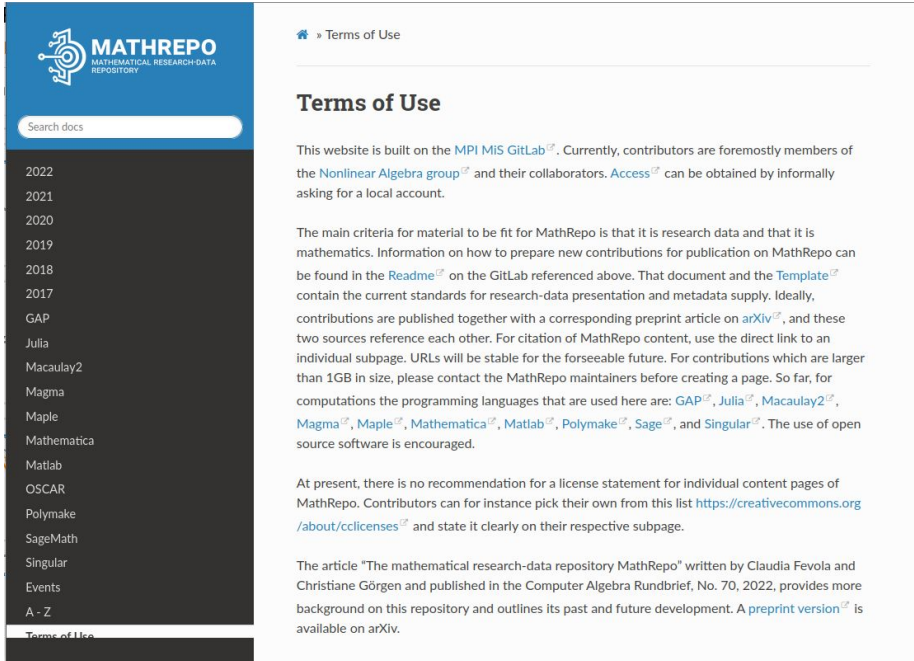
- ✓ no pay walls
- ✓ no sign-up process for readers
- ✓ low barrier for new contributions
- ✗ no automated retrieval of (meta) data

Reusable

- ✓ author names for contact
- ✓ often a lot of detail provided
- ✗ no license information



Recent improvements



MATHREPO
MATHEMATICAL RESEARCH-DATA
REPOSITORY

Search docs

2022
2021
2020
2019
2018
2017
GAP
Julia
Macaulay2
Magma
Maple
Mathematica
Matlab
OSCAR
Polymake
SageMath
Singular
Events
A - Z
Terms of Use

» Terms of Use

Terms of Use

This website is built on the [MPI MIS GitLab](#). Currently, contributors are foremostly members of the [Nonlinear Algebra group](#) and their collaborators. [Access](#) can be obtained by informally asking for a local account.

The main criteria for material to be fit for MathRepo is that it is research data and that it is mathematics. Information on how to prepare new contributions for publication on MathRepo can be found in the [Readme](#) on the GitLab referenced above. That document and the [Template](#) contain the current standards for research-data presentation and metadata supply. Ideally, contributions are published together with a corresponding preprint article on [arXiv](#), and these two sources reference each other. For citation of MathRepo content, use the direct link to an individual subpage. URLs will be stable for the foreseeable future. For contributions which are larger than 1GB in size, please contact the MathRepo maintainers before creating a page. So far, for computations the programming languages that are used here are: [GAP](#), [Julia](#), [Macaulay2](#), [Magma](#), [Maple](#), [Mathematica](#), [Matlab](#), [Polymake](#), [Sage](#), and [Singular](#). The use of open source software is encouraged.

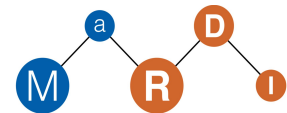
At present, there is no recommendation for a license statement for individual content pages of MathRepo. Contributors can for instance pick their own from this list <https://creativecommons.org/about/cclicenses> and state it clearly on their respective subpage.

The article "The mathematical research-data repository MathRepo" written by Claudia Fevola and Christiane Gorgen and published in the Computer Algebra Rundbrief, No. 70, 2022, provides more background on this repository and outlines its past and future development. A [preprint version](#) is available on arXiv.

- added terms of use page
 - contributor community
 - gitlab access
 - how to / standards
 - thoughts on licenses
 - historical / future comments
- discussed maintainer handover
- adapted logo
- provide workshops and training, also in a MaRDI context



UNIVERSITÄT
LEIPZIG



How to FAIR your research data

current MaRDI recommendations

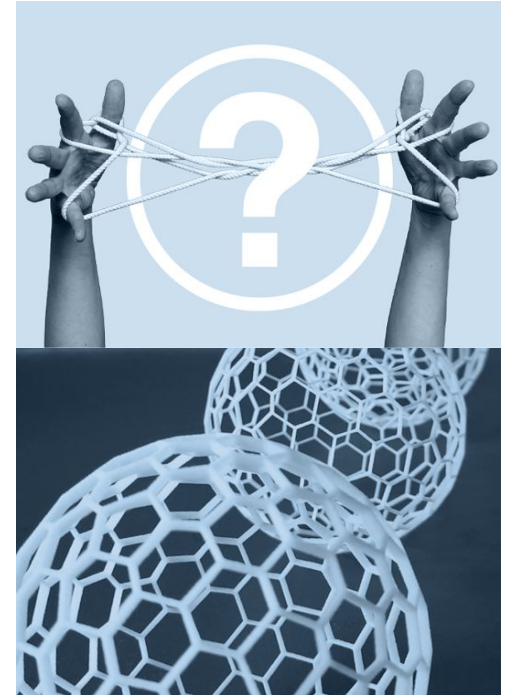
- research data on Zenodo
- metadata on Wikidata

→ make integration into the portal possible

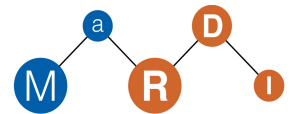
two user stories in the MaRDI Newsletter

<https://www.mardi4nfdi.de/community/newsletter>

Subscribe: t1p.de/ewmt6



UNIVERSITÄT
LEIPZIG



Take home message

Mathematical Research Data Initiative

The NFDI Consortium of Mathematics



Help Desk

Welcome to the MaRDI Community Hub. We support you in all questions concerning mathematical research data. Get in touch with our team, we are happy to help. Data culture and community integration is the responsibility of [Task Area 6](#).



Dr. Christiane Görgen
MaRDI Mathematical Data Consultant
+49 341 9732183
goergen@mardi4nfdi.de

Christiane is MaRDI's mathematical data consultant and

- for RDMPs:
 - where do you put your data?
 - why?
 - choose trustworthy repositories
- in maths there's no one-size-fits-all solution
- MR is one local solution
- MaRDI provides plenty of opportunity to discuss improvements



UNIVERSITÄT
LEIPZIG

