

FAIRness of the mathematical research-data repository MathRepo

Dr. Christiane Görgen

Mathematical Data Consultant in the NFDI consortium MaRDI

Universität Leipzig, Mathematisches Institut

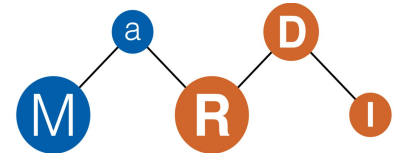
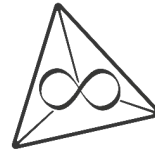
Future Opportunities for Software in Research – May 13, 2022

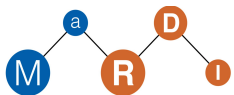


MATHREPO
MATHEMATICAL DATA
AND SOFTWARE



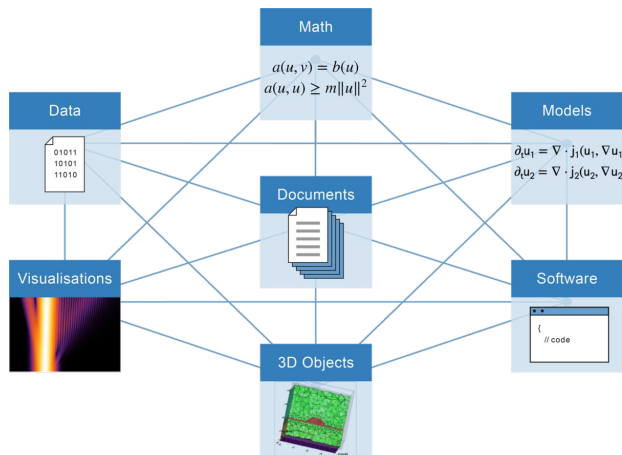
MAX PLANCK INSTITUTE
FOR **MATHEMATICS**
IN THE SCIENCES





What is research data in mathematics?

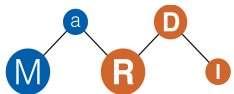
“The recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”



This is much broader than data alone!

- 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
- ...

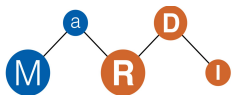
Software and code is often implicit in the process of doing research.



What is the problem with research data in mathematics?

- software/code/computations not valued as a contribution to science
- often many different types of software used for one single project
- very often implicit in the final maths (paper) publication
- Computer algebra: missing quality badges and unclear peer-review processes
- Scientific computing: implementations, state of the art, and publications not available in one place
- Statistics: missing benchmarks for learning problems
- non-standardised workflows in interdisciplinary projects, proprietary software vs. open source
- extra material often found (or not found!) on individual webpages

Currently no common standards, a lot of word-of-mouth knowledge and sometimes big hurdles to build on other people's research!



One solution: repositories

- No established repository for maths.
- Often domain-specific: journals for software packages,...
- No “one size fits all”.

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

- established 2017
- hosted by Max Planck society
- maintainers are scientists
- underlying gitlab

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES



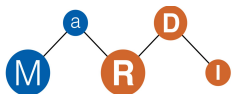
WIKIDATA

zenodo

arXiv.org



MATHREPO
MATHEMATICAL DATA
AND SOFTWARE



Advantages of MathRepo

MATHREPO
 MATHEMATICAL DATA
 AND SOFTWARE

Search docs

☰ 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 Polypols
- 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 Hyperplane Arrangements with Symmetry
- KP Solitons from Tropical Limits
- Meet and greet with OSCAR
- Primary Decomposition with Differential Operators

2020

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\"orgen (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], 1000000]
```

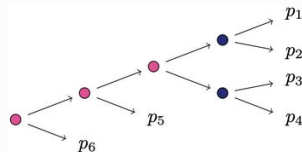
In a hundred thousand random trials this found about fifty transformations in new variables, resulting for instance in the reparametrisation

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

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.

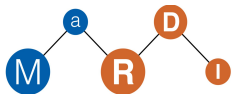
```
lineartransform[{{{p1, p3}, {p2, p4}},
{{p1 + p2, p1 + p2 + p3 + p4, p1 + p2 + p3 + p4 + p5}, {p3 + p4, p5, p6}}},
7, ConstantArray[{Range[-1, 1], Range[-1, 1]}, 2], 1000000]
```

This finds about 250 possible transformations, confirming that this staged tree represents a toric

- usability key concern
- blog-style entries possible
- no hard criteria for content
- metadata very broad

- content non-linear algebra and neighbouring fields: extra examples, conjectures, code, teaching material, computational proof of results
- locally quite broad coverage of maths topics
- more than fourty pages, employing over ten different programming languages

Ch. G\"orgen, A. Maraj, and L. Nicklasson (2021). Staged tree models with toric structure. Preprint on [arxiv:2107.04516](https://arxiv.org/abs/2107.04516) [math.AC].



Software and code on MathRepo

- GAP
- Julia
- Macaulay2
- Magma
- Maple
- Mathematica
- Matlab
- OSCAR
- Polymake
- SageMath
- Singular

https://mathrepo.mis.mpg.de/BayesianIntegrals/BITV_notebook.html

MATHREPO
MATHEMATICAL DATA
AND SOFTWARE

Search docs

2022

Bayesian Integrals on Toric Varieties

Bayesian Integrals on Toric Varieties

- No eleventh conditional Ingleton inequality
- The Geometries of Jordan nets and Jordan webs
- Introduction to Toric Geometry
- The Hirota Variety of a Rational Nodal Curve
- Lines on p -adic and real cubic surfaces
- Third-Order Moment Varieties of Linear Non-Gaussian Graphical Models
- Fiber Convex Bodies
- Counting Cubic Hypersurfaces

2021

2020

2019

» 2022 » Bayesian Integrals on Toric Varieties » Bayesian Integrals on Toric Varieties

Bayesian Integrals on Toric Varieties

This notebook accompanies the paper "Bayesian Integrals on Toric Varieties" by Michael Borinsky, Anna-Laura Sattelberger, Bernd Sturmfels, and Simon Telen. The code runs in Julia v1.7.1 and uses the packages

- Polymake (v0.7.1)
- DynamicPolynomials (v0.3.21)
- HomotopyContinuation (v2.6.3)
- lterTools (v1.4.0)
- HCubature (v1.5.0)

The file BITV.jl contains our main functions.

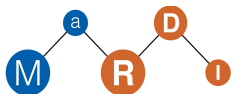
```
[1]: include("BITV.jl")  
  
polymake version 4.6  
Copyright (c) 1997-2021  
Ewgenij Gawrilow, Michael Joswig, and the polymake team  
Technische Universität Berlin, Germany  
https://polymake.org
```

This is free software licensed under GPL; see the source for copying conditions.
There is NO warranty; not even for MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

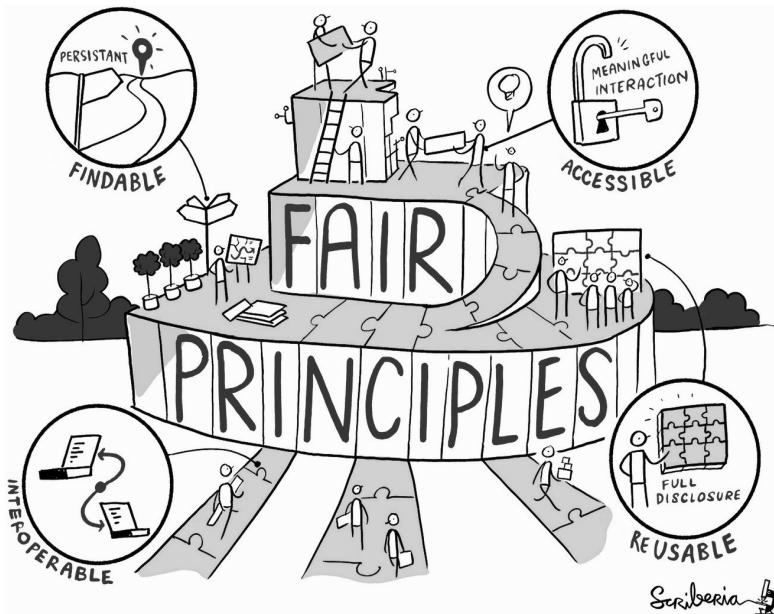
```
[1]: getM2 (generic function with 1 method)
```

Below, the examples are numbered as in the paper.

Example 4.5



Is the content on MathRepo FAIR?



- Findable: stored in a central location, labelled by metadata such as topic, author, date, system setup e.g.; findable in google search but no index search e.g. and no DOI
- Accessible: (meta-)data available for everyone but licences and restrictions not clearly stated and no API
- Interoperable: provided with information on how to make computations run easily, linked to resources, interactive notebooks; no dedicated environments
- Reusable: equipped with all important usage information?

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.

[C. Fevola](#) and Ch. Görden (2022). The mathematical research-data repository MathRepo. Submitted. Preprint on [arxiv:2202.04022 \[math.HO\]](https://arxiv.org/abs/2202.04022).