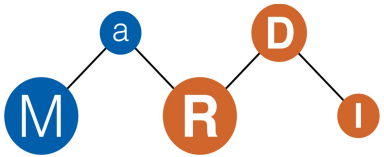
FAIRness of the mathematical research-data repository MathRepo

Dr. Christiane Görgen

Mathematical Data Consultant in the NFDI consortium MaRDI

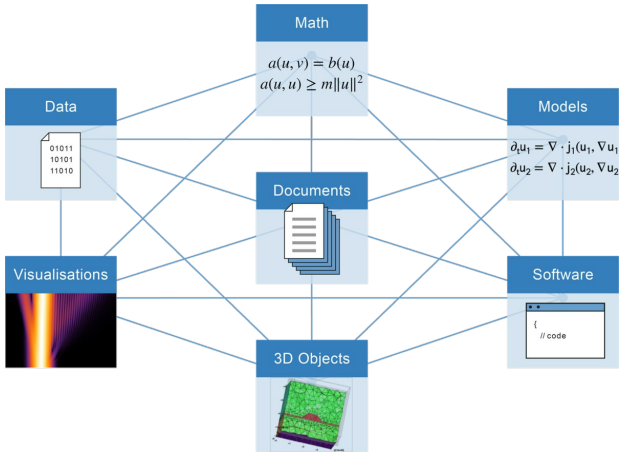
Universität Leipzig, Mathematisches Institut

Future Opportunities for Software in Research – May 13, 2022



What is research data in mathematics?

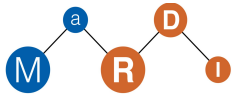
"The recorded factual material commonly accepted in the scientific community as necessary to validate research findings."



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”.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

WIKIDATA

zenodo

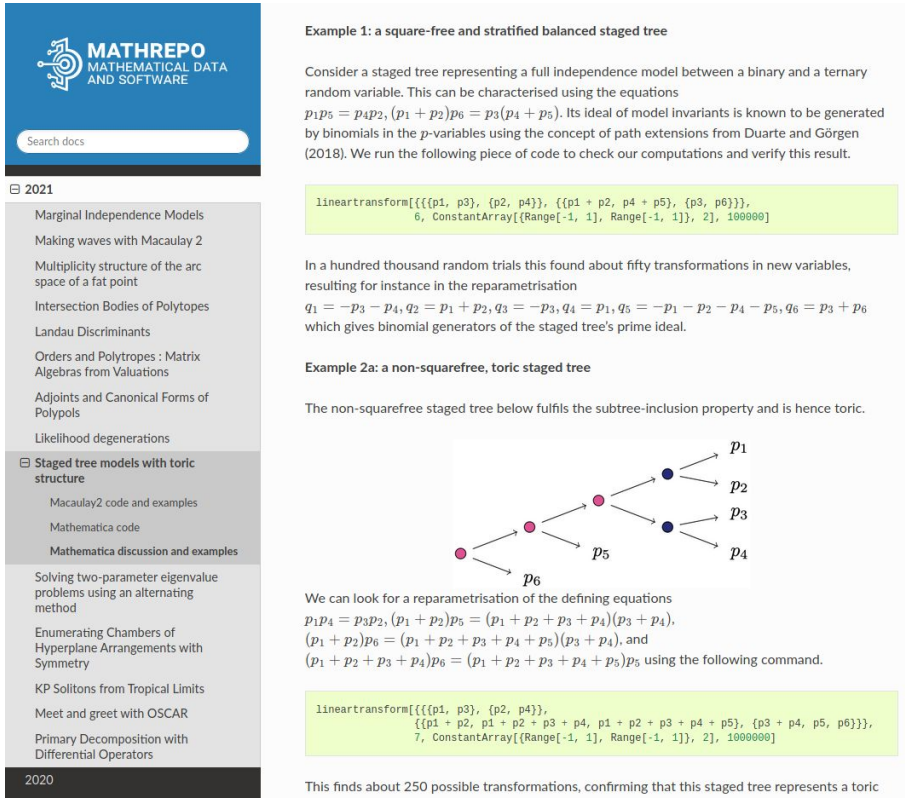
arXiv.org

https://mathrepo.mis.mpg.de

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

MATHREPO
MATHEMATICAL DATA AND SOFTWARE

Advantages of MathRepo



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

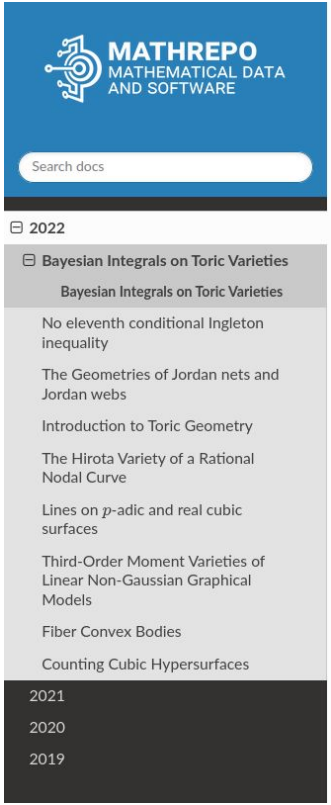
- content non-linear algebra and neighbouring fields: extra examples, conjectures, code, teaching material, computational proof of results
- locally quite broad coverage of maths topics
- more than fourty pages, employing over ten different programming languages

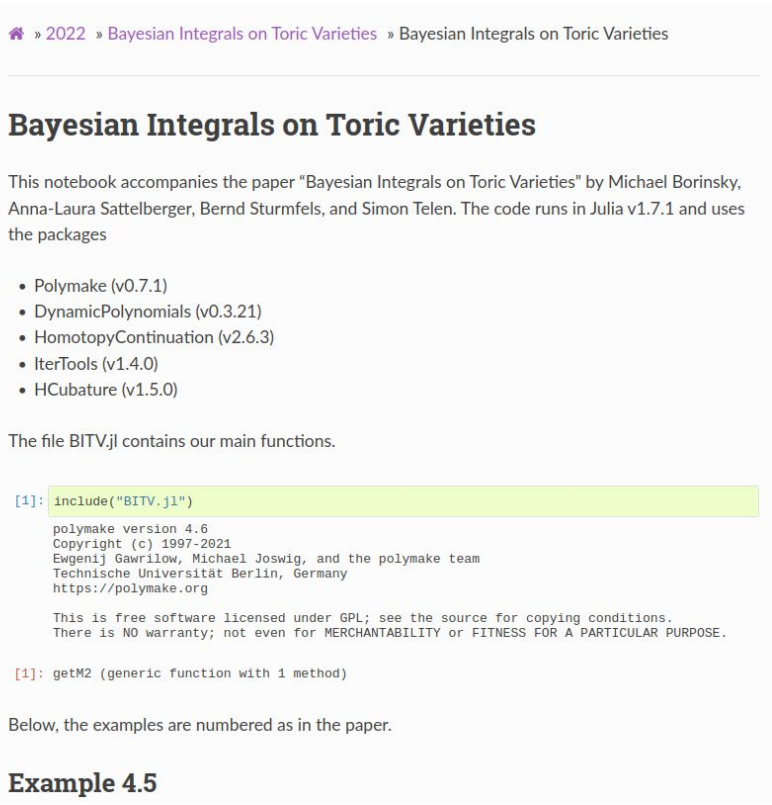
Ch. Görgen, A. Maraj, and L. Nicklasson (2021). Staged tree models with toric structure. Preprint on arxiv:2107.04516 [math.AC].

Software and code on MathRepo

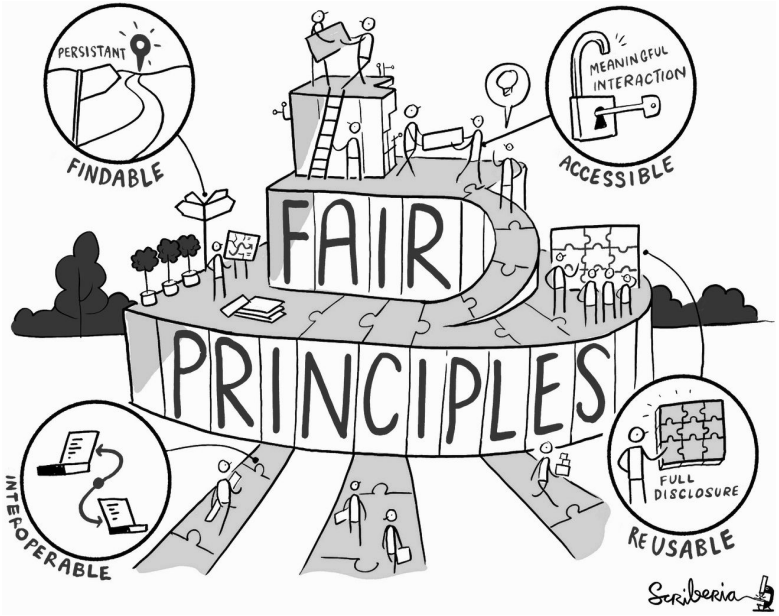
- GAP
- Julia
- Macaulay2
- Magma
- Maple
- Mathematica
- Matlab
- OSCAR
- Polymake
- SageMath
- Singular

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





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. Görgen (2022). The mathematical research-data repository MathRepo. Submitted. Preprint on arxiv:2202.04022 [math.HO].