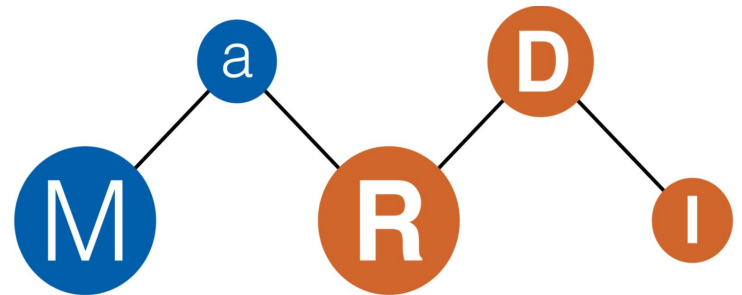# MathRepo

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, ...
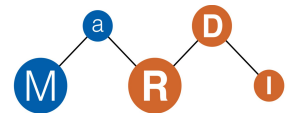- 'joint solutions for individual problems'
- 'culture change'

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
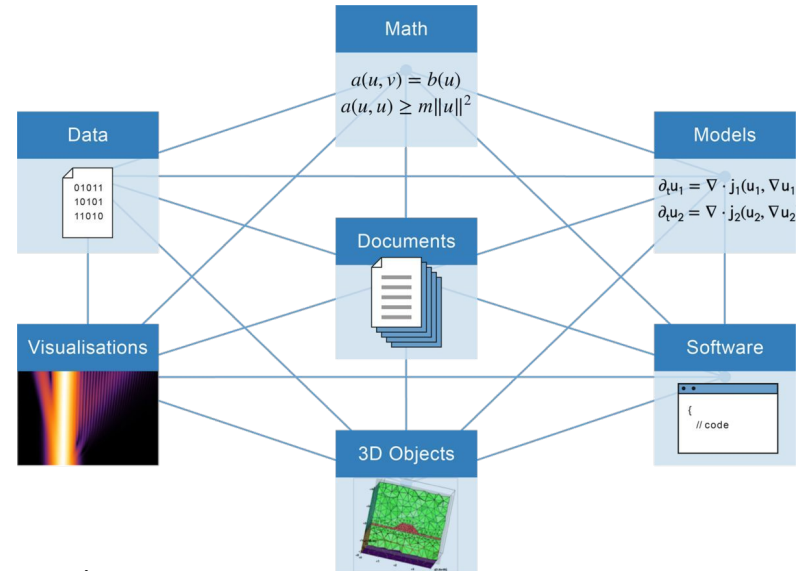
MaRDI and mathematical research data

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

some obvious challenges but also
- relevance?

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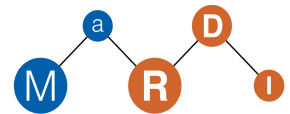- annotated with metadata (which?)

no island/decentralised solutions (no personal homepages)
→ but **good** repositories… what does good mean?

$$a(u, v) = b(u)$$
$$a(u, u) \geq m\|u\|^2$$

$$\partial_t u_1 = \nabla \cdot j_1(u_1, \nabla u_1)$$
$$\partial_t u_2 = \nabla \cdot j_2(u_2, \nabla u_2)$$

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

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

CORE TRUST SEAL

zenodo
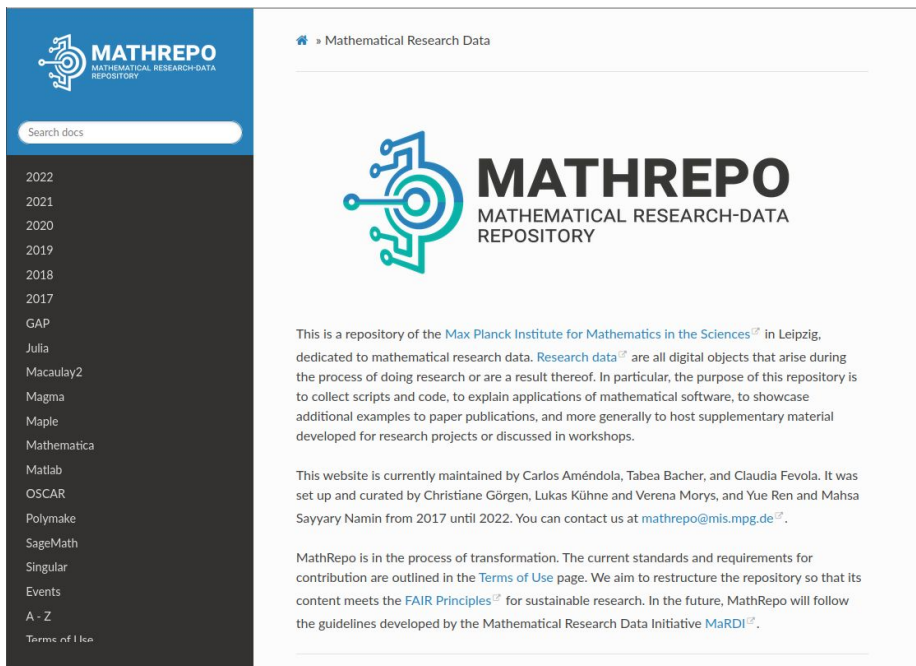
UNIVERSITÄT LEIPZIG
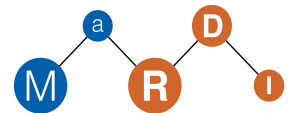
One maths-specific solution

http://mathrepo.mis.mpg.de

- 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

A typical contribution

- 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
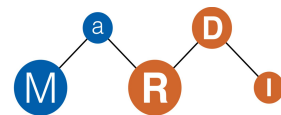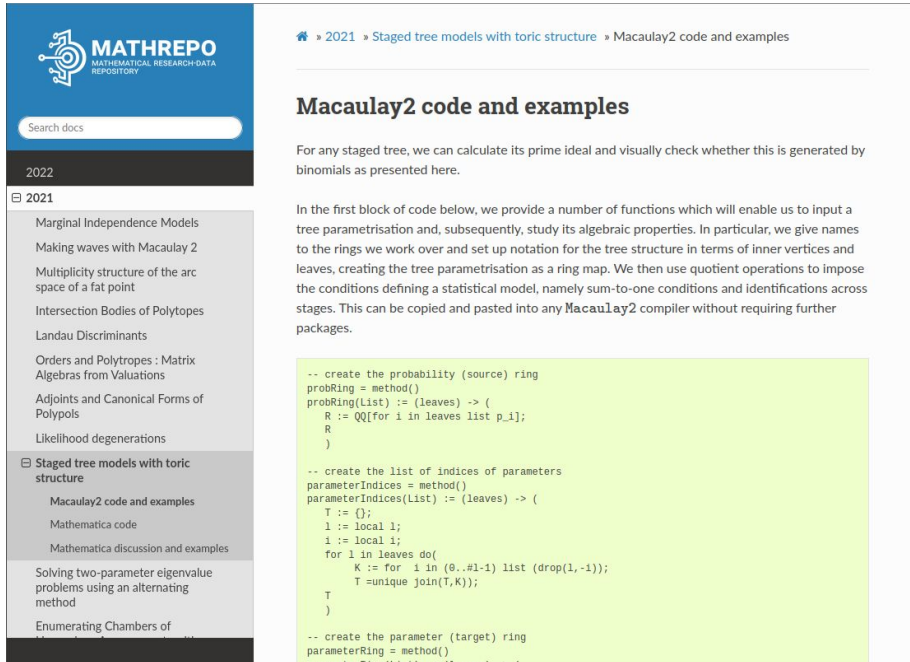
My most recent research project

- 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)
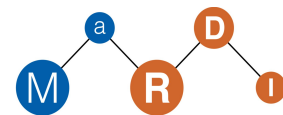
Providing code for proofs

- 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!
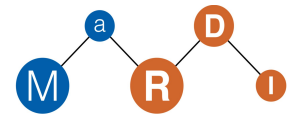
Providing code for conjectures

- 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
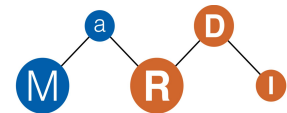
How FAIR is this?

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

Accessible

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

Interoperable

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

Reusable

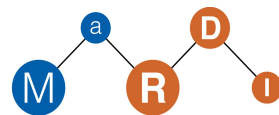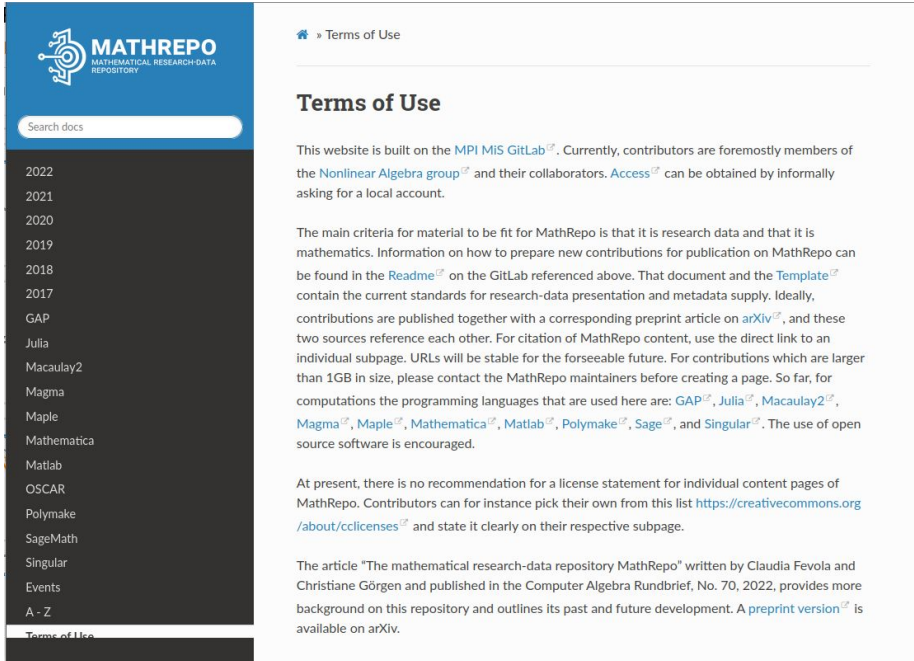✅ author names for contact
✅ often a lot of detail provided
❌ no license information

UNIVERSITÄT
LEIPZIG

Recent improvements

- 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
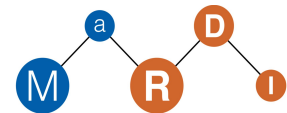
How to FAIR your research data

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

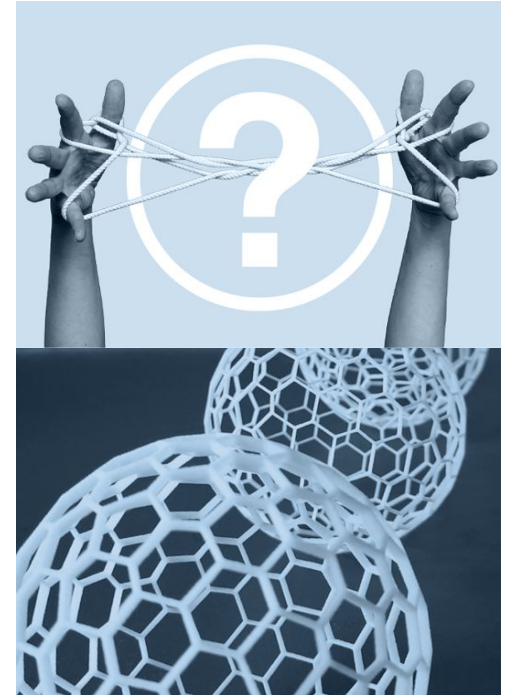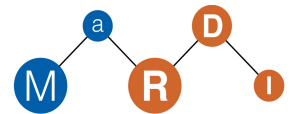Welcome to the MaRDI Community Hub. We support you in all questions concerning mathematical research data. Get in touch with our team, we are happy to help. Data culture and community integration is the responsibility of Task Area 6.

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