A FAIR File Format for Mathematical Software

Antony Della Vecchia

Joint work with M. Joswig and B. Lorenz

Technische Universität Berlin

Towards a digital infrastructure for mathematical research
2023-09-25

Overview

- MaRDI and the FAIR principles
- History of Files in Mathematical Software
- Technicalities with Algebraic Data
- Current Status of Prototype
- The File Format Specification
- Future Work

MaRDI and the FAIR Principles

- **Ma**thematics **R**esearch **D**ata Initiative.
- Develop a mathematical research data infrastructure.
- Set standards for confirmable workflows and certified mathematical research data.
- Provide services to both the mathematical and wider scientific community.
- **F**indable **A**ccessible **I**nteroperable **R**eusable [M. D. Wilkinson et al. 2016]

The Importance of Files for Computer Algebra

- People often have a preferred software system.

The Importance of Files for Computer Algebra

- People often have a preferred software system.
- Computations can be expensive.

The Importance of Files for Computer Algebra

- People often have a preferred software system.
- Computations can be expensive.
- Software changes often and requires maintenance.

The Importance of Files for Computer Algebra

- People often have a preferred software system.
- Computations can be expensive.
- Software changes often and requires maintenance.
- Data is often more valuable than the software.

The Importance of Files for Computer Algebra

- People often have a preferred software system.
- Computations can be expensive.
- Software changes often and requires maintenance.
- Data is often more valuable than the software.
- Verification of results is at most as computationally expensive.

The Importance of Files for Computer Algebra

- People often have a preferred software system.
- Computations can be expensive.
- Software changes often and requires maintenance.
- Data is often more valuable than the software.
- Verification of results is at most as computationally expensive.

Storing Computer Algebra Data

- It's common to have multiple perspectives on an object in mathematics.
- While storing mathematical data a choice of perspective must me made.
- Such a choice might not be describeable in an email.

Say we want to store:

$$p = 2y^3z^4 + (\mathbf{a} + 3)z^2 + 5\mathbf{a}y + 1$$

Storing Computer Algebra Data

- It's common to have multiple perspectives on an object in mathematics.
- While storing mathematical data a choice of perspective must me made.
- Such a choice might not be describeable in an email.

Say we want to store:

$$p = 2y^3z^4 + (\mathbf{a} + 3)z^2 + 5\mathbf{a}y + 1$$

- Some technicalities with the coefficients.
- Is y considered a coefficient of z?
- What is **a**?
- How can we guarantee the objects behave as expected on load?

History of File Formats

- The LP file format, the MPS file format. IBM [1970s]
- `Mathematica` Notebooks. Wolfram Mathematica [1988]
- `OpenMath` (tree structure). Mike Dewar [2000]
- `IPython 0.12` Interactive Browser Notebooks (Jupyter) [2011]
- `polymake` File Format. E. Gawrilow, S. Hampe, and M. Joswig [2016]

OSCAR Demo

Tree Structure

$2y^3z^4$

$+$

$(a+3)z^2$

$+$

$5ay$

$+$

1

Example File Serialized with OSCAR

Current Status and Features of OSCAR Serialization

- Over 100* registered types.
- Can store sessions over multiple files.
- Parameter Overriding.
- Serialization extensible from outside OSCAR due to Julia multiple dispatch.
- Option to attach metadata (name and ORCID for author).
- Upgrade scripts.

Schema

- A schema defines a structure for data.

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

Schema

- A schema defines a structure for data.
- Schema languages. (RELAX NG [2002], JSON Schema [2022])

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

Schema

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

- A schema defines a structure for data.
- Schema languages. (RELAX NG [2002], JSON Schema [2022])
- Is possible to define recursive structure.

Schema

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

- A schema defines a structure for data.
- Schema languages. (RELAX NG [2002], JSON Schema [2022])
- Is possible to define recursive structure.
- Schemata allow data to be validated before loading.

Schema

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

- A schema defines a structure for data.
- Schema languages. (RELAX NG [2002], JSON Schema [2022])
- Is possible to define recursive structure.
- Schemata allow data to be validated before loading.
- Adds structure to document based databases.

Schema

Figure:
`https://www.pexels.com/photo/plastic-shape-shorter-toy-11030155/`

- A schema defines a structure for data.
- Schema languages. (RELAX NG [2002], JSON Schema [2022])
- Is possible to define recursive structure.
- Schemata allow data to be validated before loading.
- Adds structure to document based databases.
- `PolyDB`, Paffenholz [2017]

File Format Specification

On Going Work

- Add Functionality for most `OSCAR` types.
- Minimal example loaders in other software systems.
- Aim to be software independant.
- Setup small databases with collaborators using the File Format.

References

Thank You!

You can find more information here

`https://arxiv.org/abs/2309.00465`