FAIR enabling practices in Mathematics research data

Volker Mehrmann, President of the European Mathematical Society (EMS)
In the past, the mathematical community has not paid much attention to creation and sharing of data, exceptions were classical logarithm tables or geometric models.

Today it is routine to use mathematical datasets in the Gigabyte range, both Human curated and machine produced data.

There is a huge variety of Open and Closed Source examples, e.g. (LMFDB) the Modular forms Data Base or (GAP) the Small Groups Library, Mathematical Modeling libraries used by languages like modelica, or subroutine libraries like the SLICOT library in control.
FAIR principles in Math.

• Wide agreement that mathematics data sets should be a common resource and open and freely available. An this includes the software that produce the data.

• However, data are produced, published, and maintained with virtually no systematic attention to the FAIR principles. [Wil 2016]

• In fact, often the sharing of data is an afterthought — see [Ber 2019] for an overview of mathematical datasets and their “FAIRreadiness”.


Current mathematical data is largely unFAIR
Mathematical data sets

- The strength of mathematics to abstractify different real objects with the same formalism is a weakness when it comes to FAIR principles, because the data always have to be associated with a detailed semantic.
- E.g. $Ax=b$, can be a linear system, a partial differential equation, a data fitting procedure, a statistical approach.
- Currently there is no good and systematic way to represent, formulas, formal proofs, programs, graphs, diagrams.
- Different data types: Record data, array data, linked data, knowledge graphs and metadata
Challenges for FAIR Math.

• In mathematics and related sciences, the data of interest are **non-uniform and highly structured entities** that are hard to find and reuse individually.

• **Representation and modeling of mathematical data** is much more difficult than anticipated.

• **No available standard for associating complex semantics and provenance data.** This effectively impedes most reuse in practice.

• Often mathematical datasets are so large that determining the identifier of the sought-after object is harder than recreating the object itself.
Current initiatives in the European Math. community

- National initiatives like MARDI in Germany
- EU DML
- EMS Publishing House (S2O)
- zbMath
- swMath
- Encyclopedia Mathematics
- arxiv, S2O publishing at EMS Press
- DLMF
What should be done for Math within EOSC

• Create **semantics aware** data structures.
• Assure **semantic interoperability**.
• **Deep FAIR services** that process individual entries of a dataset and their individual structure.
• Improve **Symbolic Representation Languages**.
• **Semantic Web and Knowledge Graphs**.
• **Databases for Concrete Data**.
• **General Data Sharing Infrastructures (GitHub)**
Conclusion and Outlook

• Particular needs in the roll-out of Open Science: to preserve the historical corpus, and to provide technical solutions for math. data and expressions.

• The mathematical community in Europe has been at the forefront of Open Science: our knowledge and experience is of use to the EOSC.

• Mathematics has been in worldwide development for millennia, and Europe is a key hub in its preservation and expansion.

• EOSC should be a vehicle for the uniting of European Science.

• A jargon-minimum approach to EOSC, keeping all sciences fully engaged, and listening to active scientists, is the way forward.

• There should be a way of overcoming financial hurdles to the future involvement of scientific societies including the EMS, in the EOSC.

• Support of Projects like FAIRMAT in EOSC.