Open-source cyberinfrastructure supporting mathematics research

organized by
Robert Beezer, Steven Clontz, and David Lowry-Duda

Workshop Summary

Workshop Summary

Technology is integral to many areas of modern mathematics. Computer algebra systems like Sagemath, Maple, Mathematica, and others inform research through computation. Databases such as the LMFDB or the π-base collect mathematical objects and their properties. Theorem proving software, such as Lean, supports mathematical reasoning itself. The tools we use to author scholar documents (LaTeX, PreTeXt) and the platforms we use to collaborate (the arXiv, github, overleaf, MathOverflow, and many more) all play essential roles.

This workshop was devoted to building capacity for open-source software supporting mathematics research. Participants gathered to share knowledge of existing software solutions, to clarify goals for future cyberinfrastructure, and to establish communication channels to facilitate later development. We believe that we, as a community, are best-equipped to make the tools that will enable us to flourish.

The workshop participants came from several different backgrounds, including established mathematicians, young researchers, educators, and software developers working in industry. In the mornings, we had talks that aimed to introduce and familiarize participants with different areas of software and how they interact with mathematics. These included talks on the arXiv, mathematical databases, math formalization software, and talks on what makes digital communities effective; in addition there were a series of lightning talks for quick, focused introductions of particular products and problems by the participants.

On the first day of the workshop, we identified 10 broad problems we wanted to investigate. This led to several working groups for the rest of the week.

Working Group Summaries

Formal Math Interoperability with CAS.

This group looked at how formal math software and computer algebra systems could interact. For example, formal theorem provers, such as Lean, might benefit from fast computation from a CAS such as Sagemath, and later verify the result. Conversely, Sagemath might be able to verify the correctness of a result or an algorithm implementation.

This group specifically looked into two examples.

The first was to determine if a number a was a primitive root mod a prime p in Lean, taking advantage of Sagemath's implementation of integer factorization. There is a demo, including running sage code from Lean and verifying the algorithm, at https://github.com/aim-cyber-workshop-2023/lean-sage/github.com/aim-cyber-workshop-2023/lean-sage/.

The second was to incorporate the Algebrite CAS to validate simple propositions in the Lurch proof processor. They made a functional proof of concept.

It was remarkable that two separate, working integrations were made over just a couple of days.

Databases of Visualizations.

This group looked at creating an online database of mathematical visualizations. They focused on diagrams, not "data visualization" or "mathematical art". The purpose is to see what diagrams are possible, how they were created, whether the tools to make them are usable, and whether the visualizations themselves are licensed for reuse — and all in as easy a way as possible. Implicitly, this database could also encourage best practices by way of inclusion.

By the end of the week, they built a proof-of-concept prototype.

Database of Databases.

Continuing the momentum from the Dagstuhl seminar Automated Mathematics: Integrating Proofs, Algorithms and Data, this working group made a major step towards a sustainable, collaborative catalog of existing datasets in mathematics. The group designed and partially implemented a new setup, making it as easy as possible to update the contents of the catalog while avoiding the need for active server maintenance. The group also produced an initial index of potential datasets that are included in papers and appendices on Arxiv.

The catalog frontend is available at https://mathbases.org. In addition to the catalog, Mathbases aims to provide resources for mathematicians interested in creating new mathematical databases, and we are tracking progress using github.

Canonical Concepts and Alignment.

This working group studied the general problem of recognizing the same/similar concepts across different pieces of mathematical literature, computer algebra systems, and technologies.

Mathematics is a huge web of concepts, definitions and results, deeply interrelated. While learning to prove results is very hard and students need all the help they can get, learning concepts and how they relate to already known concepts is not so hard and even seasoned mathematicians could do with some help. Mathematics is very specialized at present and people learn on an "as you need" basis.

Thus laying out basic concepts and results, a canon of mathematical ideas would be useful, relating these ideas using mathematical English is the way we propose doing it, using Wikidata/Wikipedia as a basic repository of concepts. Can we use NLP/AI to identify fundamental concepts from mathematical literature?

During the workshop week, the group worked on concepts from linear algebra by aligning two open-source books. The goal is to cover, eventually, all areas of math, using open-source books and mathematical wikis already extant, but building "bottom up", specific areas.

Knowledge Tracing.

This working group examined how to model a student's understanding as they interact with new material. Specifically, how can we improve our understanding of students' mastery of mathematical concepts through automated assessment?

This is a known problem and there are existing tools that try to address this (namely ALEKS). The group began by identifying problems and successes in ALEKS. Some of the notes from this conversation can be read at https://github.com/aim-cyber-workshop-2023/.github/wiki/Dunking-on- ALEKS –(Knowledge-Tracing)github.com/aim-cyber-workshop-2023/.github/wiki/Dunking-on-∼ALEKS∼–(Knowledge-Tracing). The group also brainstormed what an ideal interface and process would look like.

Math in the Future.

This working group tried to identify what aspects of research can be improved with the technology we have today, and what might change in the near future? More broadly, can we identify a desirable future of math, and can we imagine the tools and processes to get us there?

The major points of discussion can be read at https://github.com/orgs/aim-cyber-workshop-2023/discussions/17. We expect that this sort of discussion, and the topics we discussed, will be the source of important innovation for years to come.

Developer On-Ramps.

This group examined what inspires new users to have successful experiences with new mathematical tools; further, how could the community of mathematical software users and developers facilitate positive experiences?

This group met for two days and some of their notes can be read at https://github.com/aim-cyber-workshop-2023/.github/wiki/Notes-for-On-Ramps-Groupgithub.com/aim-cyber-workshop-2023/.github/wiki/Notes-for-On-Ramps-Group. Much of the discussion centered around brainstorming aspects that drive new users to try new tools, anchors that negatively drag people back, and boons that propel people forward.

The ideas from this group informed a group that met on the fourth day, working to draft **Community Values**.

Community Values.

This group imagined what a central community that fosters mathematical cyberinfrastructure would look like. They noted that successful organizations have a guiding set of principles that informs decisions and behavior. What core *community values* should our community have? We decided on the following.

- We are a community of mathematics software developers and their advocates, dedicated to the advancement and accessibility of mathematics.

- We create, develop, and maintain digital products for mathematics researchers, educators, and learners. These products include software, datasets, documentation, and infrastructure.
- Our community empowers this work by sharing knowledge and best practices.
- We recognize and value this work as scholarly activity, and provide artifacts of these efforts recognized by both academic and industry stakeholders.
- We ensure our products are accessible to everyone who needs them.
- We provide opportunities for professional development at all career stages and particularly welcome and support new contributors.
- Our community is open to everyone. We promote, encourage, and improve diversity in our ecosystem by maintaining an inclusive environment that treats all members equitably.

code4math, Organization and Governance.

Building on the other activities and working groups from this workshop, on the last day a group met to discuss *creating* an organization that supports the community of people using and developing mathematical cyberinfrastructure.

We call this community `code4math`, the *Consortium of Digital Ecosystems for Mathematics*. This community has a website, http://code4math.org/code4math.org, a zulip chat server, and a steering council formed from participants at this workshop.

Future Plans

This was an extremely productive week. Several of the working groups initiated new collaborations and have plans to continue working in the future. And a new organization, `code4math`, was formed to foster more conversation and progress towards a better future for mathematicians and mathematics.