Proposal for Sabbatical Leave

Cyberinfrastructure for Mathematics Research and STEM Higher Education

Steven Clontz
Department of Mathematics and Statistics
University of South Alabama

2022 September

Abstract

NSF's Office of Advanced Cyberinfrastructure (OAC) "supports and coordinates the development, acquisition, and provision of state-of-the-art cyberinfrastructure resources, tools and services essential to the advancement and transformation of science and engineering", and the NSF has doubled-down on this commitment with the recent establishment of its "Technology, Innovation and Partnerships" directorate, the first new NSF directorate in over 30 years. In service of this mission, Dr. Clontz serves as co-PI on NSF DUE Award 2011807, responsible for the development of several technologies to support evidence-based pedagogies in undergraduate mathematics education, and his NSF Proposal 2230153 as PI to develop an Open-Source Ecosystem fostering technologies used in the development of STEM higher-education Open Educational Resources has been recommended for funding beginning in 2023.

This proposal requests to leverage the momentum from these projects by allowing Clontz to spend Spring 2024 collaborating with scholars across the country in the development of advanced free-and-open-source (FOSS) cyberinfrastructure for mathematics research and STEM higher education. The American Institute of Mathematics will serve as Clontz's host for this sabbatical leave. The deliverables of this sabbatical will include multiple software products benefiting mathematics research and STEM higher education, enhancement of the human infrastructure surrounding these products, and the continued development of Clontz's ongoing research in general topology.

1 Background

The term **cyberinfrastructure** traces its roots back to Presidential Decision Directive NSC-63 [27], and is commonly used today to refer to "computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible" [14]. What cyberinfrastructure "looks like" varies greatly between disciplines, but a constant quality of it is the utilization of modern technologies that make the jobs of scholars more efficient, thereby producing more and better advancements in STEM. As such, the National Science Foundation [20] and other federal agencies have made the development of cyberinfrastructure a key priority in supporting STEM scholarship.

The NSF has doubled-down on this commitment with the establishment of its "Technology, Innovation and Partnerships" directorate in Spring 2022, the first new NSF directorate in over 30 years [24]. Included in this directorate is the Pathways for Open Source Ecosystems solicitation [22] whose first call for proposals was due in May 2022.

1.1 Cyberinfrastructure in Mathematics Research

One of the features of many areas of mathematics, including the applicant's area of general topology, is the expansive breadth of open problems that still technically require just pencil and paper. Nonetheless, as technology advances, we are seeing more uses for tools beyond the blackboard for making advancements

in even abstract mathematics. One obvious example is the use of high-performance (or even conventional) computing to brute-force the exploration of finite mathematical spaces [4].

But perhaps underrated are the systems used by mathematicians for sharing and collaborating bleeding-edge advancements with the community of research. The mathematics written on paper or the chalkboard cannot travel far on their own; generally, mathematicians use software such as LaTeX to typeset proofs and formulas for distribution in PDF format. Then, rather than asking colleagues from across the country to physically visit their office computer, the researcher uses technologies such as email, a personal website, and/or a preprint server to post their work for dissemination. High-quality work must then be submitted to a journal to be vetted by peers, a process handled sometimes through email, and other times via non-trivial content management systems hosted in the cloud. These results are then published, but this content is now burned into inaccessible (in many senses of the word) PDFs locked behind paywalls. A typical strategy for finding recent developments in mathematics is to Google key words, but a search for "selection games", an area of mathematical game theory studied by the applicant, quickly reveals the limitations of this approach.

One approach to overcoming this limitation is the development of databases of mathematical objects. As of writing, the *Catalogue of Mathematical Datasets* [6] enumerates 84 such tools. One of these is the *π-Base* (a.k.a. pi-Base) community database of topological counterexamples [13], created by software engineer James Dabbs. Clontz joined the project in 2017 to serve as its lead mathematical editor. In the beginning the database was treated similar to a wiki, with unreviewed contributions made by several mathematicians, students, and other unknown contributors. To address this, changes were made to the software to enforce a level of peer review, and Clontz was awarded with an internal Faculty Development Council grant to hire a student to audit and expand the content of the database.

The result of this support was successful, and today the pi-Base is commonly cited on mathematics discussion boards used by students and researchers alike [28] [29]. However, it also exposed critical inefficiencies in the user experience (UX) of both contributing to the database and reviewing contributions.

In [7], Buzzard points to the π-Base as an important example of how semantic search helps researchers more quickly and thoroughly query the literature than standard search engines. Furthermore, he points to the increasing use of in computer-verified proof techniques as another emerging element of cyberinfrastructure in mathematics research. In particular, the Lean Prover [19] from Microsoft Research and the `mathlib` [17] library of mathematics written in Lean are becoming the de facto standard for formalization of mathematical results that can be verified by computer.

1.2 Cyberinfrastructure in STEM Higher Education

Logistics are frequently a limiting factor in the adoption of evidence-based practices in instruction, particularly in undergraduate mathematics education [25]. Often, faculty are willing, if not eager, to change instruction in ways that benefit students, but do not have the resources to implement such change.

Likewise, the authors of [26] observed the limitations of educational software that technically works, but isn't designed for platforms that are readily in the hands of students and instructors. For example, while CalcPlot3D has always been free and open-source software, it was originally limited in its reach due to being written in Java. By rewriting the application in (the similarly-named but unrelated programming language) Javascript, students and instructors were no longer required to be at a computer station with a Java runtime installed, but could instead utilize the program from any device with a web browser.

Partially supported by NSF DUE 2011807, Clontz has developed two software applications to support mathematics instruction and Team-Based Inquiry Learning (TBIL), a flavor of Team-Based Learning that was the focus of the University's most recent Quality Enhancement Project. The first is the CheckIt Platform [8], allowing instructors to write minimal code to generate randomized mathematics exercises that can be automatically exported not only to LaTeX/PDF for printing, but also published to the web as practice exercises, and to LMSes including Canvas, D2L, and Moodle. The second is Scratchee [10], a virtualization of the Instant Feedback Assessment Technique (IF-AT) [12] integral to TBIL. In addition, Clontz serves as collaborator on the PreTeXt project [5], developing a user-friendly platform for authoring the PreTeXt markup language that produces both PDF and accessible HTML documents (including textbooks and research manuscripts) from the same source, including this proposal [11]. Furthermore, docuemnts authored in PreTeXt can also be automatically published as Braille [2], an uncommon feature for commercial textbooks, much less free Open Educational Resources (OER), providing access to mathematics often out of reach to blind

students.

In addition to supporting mathematics instruction directly, the CheckIt Platform is also being used to support Research in Undergraduate Mathematics Education (RUME). Exercises on the platform are designed to assess particular learning outcomes; in order to measure the effectiveness of instruction as part of DUE 2011807, CheckIt-generated assessments will be used at several campuses across the country. This allows instructors to administer as many versions of each exercise as needed for logistical purposes, while still ensuring that each version of the exercise measures exactly the same learning outcome.

1.3 Free and Open-Source Software (FOSS)

The focus of this project is to produce **Free and Open-Source Software** that will benefit scholars, instructors, and students of mathematics. Frequently, the NSF requires that software products it funds be FOSS. It's worth clarifying what is meant by this.

Open-source software is most easily defined. All code written as part of this project will be made available to the public via Clontz's GitHub [9] (or other publicly available repositories as appropriate). This means that anyone will be able to obtain a copy of any software developed during this sabbatical, use this software to benefit their research or instruction, and contribute corrections or improvements to the codebase to benefit others.

The word **free** in FOSS does the heaviest lifting. Primarily, it means that this software will be explicitly licensed for free use and adaptation by anyone who wishes, removing legal barriers that might prevent its adoption by other researchers or instructors.

But for the purposes of this project, "free" also implies that the software, whenever possible, will be developed mindfully to avoid dependencies on non-free infrastructures. For example, technically the Canvas Learning Management System is FOSS software [16]; however, that does not mean that it can truly be adopted without cost. Maintainance of a learning management system server incurs both technology costs and personhour costs, which is why many campuses, including the University, simply pay Instructure to provide the service rather than utilize its FOSS directly. Technical debt can never be completely avoided; however, by making smart design decisions in the development of software packages that aren't intended to turn a profit, this debt can be kept minimal. To this end, most of the software produced will either be written in HTML/Javascript, which can be freely hosted and run in any modern web browser, or will produce static such HTML/JS products for dissemination.

Finally, NSF's recognition of the critical role FOSS products play in the cyberinfrastructure of STEM research is witnessed by solicitations such as its new Pathways for Open-Source Ecosystems solicitation mentioned earlier. Clontz's $266K one-year Phase I proposal as PI to establish an Open-Source Ecosystem for the PreTeXt community has been recommended for funding by an NSF program officer as part of the inaugural round of awards, and will lead to a $1.5M two-year Phase II proposal to be submitted in Fall 2023.

2 Proposed Activities

Sabbatical leave is requested for Spring 2024. This timing is requested to complement Clontz's anticipated NSF-funded release from teaching duties in Fall 2023.

The American Institute of Mathematics (AIM) located in San Jose, California will serve as host for this sabbatical, connecting Clontz with other likeminded scholars engaged in the work of developing cyberinfrastructure for mathematics research and education. To support these activities, AIM has offered Clontz office space for the duration of the sabbatical, as well as funding to cover travel to San Jose and local expenses.

2.1 Cyberinfrastructure for Mathematics Research

A current shortcoming of the pi-Base application is lack of tooling to preview work on the actual web application (or a facsimile thereof) by both contributors and referees. Instead, numerical IDs for each object in the database must be manually looked up and compared across several files, and this process allows plenty of room for human error. The main development goal for the pi-Base is to add functionality via GitHub

Actions that will automatically provide an online preview of how individual contributions will change the overall database, which can be used by both contributors and referees to catch such mistakes before they are published to the public. Such features will also assist new contributors in making improvements by providing a more graphical user interface compared to the current editing of text files.

To begin to make these improvements, Clontz will collaborate with the pi-Base's lead developer James Dabbs for one week at the beginning of the sabbatical for a focused development sprint. Note that the pi-Base is developed using the same technologies (particularly, the Svelte Javascript framework) as Clontz's CheckIt and Scratchee projects, and is used by Dabbs in his daily work as a software engineer. As such, this meeting will have cascading effects across the rest of the work performed throughout the sabbatical, by providing time and support for Clontz to pick up more advanced Javascript programming techniques fueled by best industry practices. In addition, this training will enable Clontz to serve not only as the pi-Base's lead mathematical editor, but also as a co-developer of the platform itself, ensuring that an academic researcher is capable of maintaining the platform in case of future technical issues.

Additionally, Clontz will use time granted by sabbatical to become fluent in the Lean proof assistant language, and contribute his expertise in general topology to the `mathlib` library. Collaborating with Jim Fowler at Ohio State University, Clontz will also work to integrate the π-Base and `mathlib` databases to connect computer-verified results in mathematics to an easily-accessible semantic search enginer for use by researchers.

2.2 Cyberinfrastructure for STEM Higher Education

Throughout 2023, Clontz will be working on the development of an Open Source Ecosystem for PreTeXt software products used to develop Open Educational Resources for STEM higher education classrooms. In particular, Clontz's NSF-funded teaching release in Fall 2023 will allow him to hold virtual office hours for community members, iterate on security infrastructure for PreTeXt products, and develop a Software-as-a-Service solution for authoring, building, and deploying PreTeXT documents in the cloud. A $1.5M Phase II proposal will be submitted to the NSF around October 2023; sabbatical release in Spring 2024 will facilitate the continued development of this Open-Source Ecosystem while this proposal is under review.

After preliminary development for two years for Clontz's personal use, the CheckIt Platform received its first public release in June 2021. Two dozen instructors participated in an introductory authoring workshop, and as of August 2022 randomized exercise banks have be published or are in development for calculus, linear algebra, differential equations, introductory statistics, quantitative reasoning, introduction to proofs, and mathematics for liberal arts courses, and were been used in classrooms across the country throughout 2021 and 2022. This quick evidence of productivity witnesses the utility of the platform, and it has also generated many requests for new features and bugfixes to be worked on during this sabbatical.

An alpha version of the Scratchee app for creating and sharing virtualized scratch cards for IF-AT was released in late July 2021. This prototype was created in response to developments related to Clontz's NSF grant for supporting Team-Based Inquiry Learning; while physical IF-AT cards are not incredibly expensive, they introduce logistical friction for implementing TBIL (ordering the cards, waiting for shipment, aligning multiple-choice questions with the predetermined correct responses printed on the card). Multiple TBIL instructors have used Scratchee throughout 2021 and 2022; their experiences will be used to determine the necessary enhancements required for a public release of the platform for the broader Team-Based Learning community.

2.3 External Grant Activity

The primary external mechanism used to fund proposed activities related to the infrastructure of STEM higher education is the NSF POSE solicitation. As noted, a $1.5M two-year Phase II proposal will be submitted in Fall 2023, to begin following the requested Spring 2024 sabbatical.

Another relevant solicitation is Research Experiences for Undergraduates (REU) [23]. In particular, the research data used as the basis for the pi-Base was originally developed as part of an NSF-funded REU [15], which resulted in the reference text *Counterexamples in Topology*. In this spirit, a new REU program would allow students to extend this work using the pi-Base, expanding its coverage of the literature and making the platform more useful by researchers. Additionally, unlike some traditional mathematics REUs,

students would gain knowledge of the cyberinfrastructure powering the pi-Base platform, and relevants skills that would extend to careers across STEM. This proposal is anticipated to be submitted in Summer 2024, developed as part of a Spring 2024 sabbatical.

2.4 Additional Activites

Sabbatical leave from instruction is requested to facilitate progress on the above activities while allowing Clontz to continue his active line of traditional mathematics research on the game-theoretic characterization of topological properties. Briefly, a **topological space** models data points along with neighborhoods that describe nearby data; for example, the Euclidean line uses real numbers $x \in \mathbb{R}$ as points, and intervals $(x - r, x + r)$ of radius $r > 0$ as neighborhoods. At the time of writing, Clontz is investigating how to characterize topological dimension (as simple examples, a square is two-dimensional and a cube is three-dimensional) in terms of which player in a certain game defined for each topological space has a winning strategy. Recent results by Babinkostova and Scheepers published in the Proceedings of the Americian Mathematical Society have established three such games [3]; however, these results are only guaranteed when a **metric** measuring distance between points as non-negative real numbers is known to exist (for example, $|x - y|$ is the distance from x to y on the Euclidean line of real numbers). When such a measurement is not guaranteed, the two common definitions of dimension no longer are equivalent, making the study of non-metrizable spaces difficult. This also illustrates the utility of a more robust pi-Base, as finding literature on dimension in the context of non-metrizable spaces is very difficult when only using search engines and chasing citations, so this work will also be incorporated into the pi-Base as it is completed to assist future scholars.

3 Summary of Anticipated Outcomes

In summary, sabbatical leave will provide Clontz support in the development of multiple software packages in support of mathematics research and instruction, enhancement of the technological and human infrastructure surrounding these products, the preparation of several NSF proposals, and continued success in mathematics research related to general and set-theoretic topology. Additionally, collaboration with the American Institute of Mathematics, which hosts over twenty workshops involving hundreds of mathematicians annually, will provide Clontz with networking opportunities that will further his career and bring positive exposure to the University.

References

[1] American Institute of Mathematics, *American Institute of Mathematics*. https://aimath.org/.

[2] American Institute of Mathematics, *Math that feels good: Creating learning resources for blind students*. https://aimath.org/aimnews/braille_full/.

[3] L. Babinkostova & M. Scheepers, *Countable dimensionality, a game and the Haver property*. Proceedings of the American Mathematical Society. https://www.ams.org/proc/0000-000-00/S0002-9939-2021-15492-0/S0002-9939-2021-15492-0.pdf.

[4] D. H. Bailey, D. Broadhurst, Y. Hida, Xiaoye S. Li, & B. Thompson, *High Performance Computing Meets Experimental Mathematics*. Proceedings of the 2002 ACM/IEEE Conference on Supercomputing, 2002, **SC '02**. https://doi.org/10.1109/SC.2002.10060.

[5] Rob Beezer, *PreTeXt*. https://pretextbook.org/.

[6] Katja Berčič, *Catalogue of Mathematical Datasets*. https://mathdb.mathhub.info/.

[7] Kevin Buzzard, *What is the point of computers? A question for pure mathematicians*. arXiv, 2021, https://arxiv.org/abs/2112.11598.

[8] Steven Clontz, *Checkit*. https://checkit.clontz.org/.

[9] Steven Clontz, *GitHub Repositories*. https://github.com/StevenClontz/.

[10] Steven Clontz, *Scratchee.* https://scratchee.clontz.org/.

[11] Steven Clontz, *StevenClontz/sabbatical-2024-spring.* GitHub. https://github.com/StevenClontz/sabbatical-2024-spring.

[12] S.H. Cotner, B.A. Fall, S.M. Wick, J.D. Walker, & P.M. Baepler, *Rapid feedback assessment methods: Can we improve engagement and preparation for exams in large-enrollment courses?.* Journal of Science Education and Technology, 2008, **17**, 5. 437–443.

[13] James Dabbs, π-*Base.* https://topology.pi-base.org/.

[14] EDUCAUSE Campus Cyberinfrastructure Working Group and Coalition for Academic Scientific Computation, *Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies*, 2009. https://library.educause.edu/-/media/files/library/2009/4/epo0906-pdf.pdf.

[15] Deanna Haunsperger & Stephen Kennedy, *The Idea Man: An Interview with Lynn Steen.* Mathematics Magazine, 2015, **88**, 3 163–176. https://doi.org/10.4169/math.mag.88.3.163.

[16] Instructure, *Canvas LMS on GitHub.* https://github.com/instructure/canvas-lms.

[17] Lean Community, *mathlib.* https://github.com/leanprover-community/mathlib.

[18] Andrew Lewis, Shiladitya Chaudhury, Steven Clontz, Julie Estis, & Christopher Parrish, *Award Abstract #2011807: Transforming Lower Division Undergraduate Mathematics Through Team-Based Inquiry Learning.* National Science Foundation. https://nsf.gov/awardsearch/showAward?AWD_ID=2011807.

[19] Microsoft Research, *Lean.* https://leanprover.github.io/.

[20] National Science Foundation, *About Advanced Cyberinfrastructure.* https://www.nsf.gov/cise/oac/about.jsp.

[21] National Science Foundation, *Improving Undergraduate STEM Education: Education and Human Resources (IUSE: EHR).* https://www.nsf.gov/pubs/2021/nsf21579/nsf21579.htm.

[22] National Science Foundation, *Pathways to Enable Open-Source Ecosystems (POSE).* https://beta.nsf.gov/funding/opportunities/pathways-enable-open-source-ecosystems-pose.

[23] National Science Foundation, *Research Experiences for Undergraduates (REU).* https://www.nsf.gov/pubs/2022/nsf22601/nsf22601.htm.

[24] National Science Foundation, *NSF establishes new Directorate for Technology, Innovation and Partnerships.* https://beta.nsf.gov/news/nsf-establishes-new-directorate-technology-innovation-and-partnerships.

[25] S.E. Shadle, A. Marker, & B. Earl, *Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments.* IJ STEM Ed, 2017, **4**, 8. https://doi.org/10.1186/s40594-017-0062-7.

[26] M. VanDieren, D. Moore-Russo, & P. Seeburger, *Technological Pedagogical Content Knowledge for Meaningful Learning and Instrumental Orchestrations: A Case Study of a Cross Product Exploration Using CalcPlot3D..* Teaching and Learning Mathematics Online, 2020, 267–294.

[27] The White House, *Presidential Decision Directive NSC-63*, 1998. https://fas.org/irp/offdocs/pdd/pdd-63.htm.

[28] *Search results for url topology.jdabbs.com*, Math.Stackexchange. https://math.stackexchange.com/search?q=url%3A%22topology.jdabbs.com%22.

[29] *Search results for url topology.jdabbs.com*, MathOverflow. https://mathoverflow.net/search?q=url%3A%22topology.jdabbs.com%22.