SHORT: Summer of Reproducibility: Building Global Capacity for Practical Reproducibility through Hands-On Mentorship

Stephanie Lieggi
slieggi@ucsc.edu
UC Santa Cruz
Santa Cruz, California, USA

Fraida Fund
ffund@nyu.edu
NYU Tandon School of Engineering
Brooklyn, New York, USA

Kate Keahey
keahey@mcs.anl.gov
University of Chicago
Chicago, Illinois, USA

Marc Richardson
mtrichardson@uchicago.edu
University of Chicago
Chicago, Illinois, USA

Abstract

Thanks to increasing awareness of the importance of reproducibility in computer science research, initiatives such as artifact review and badging have been introduced to encourage reproducible research in this field. However, making "practical reproducibility" truly widespread requires more than just incentives. It demands an increase in *capacity* for reproducible research among computer scientists - more tools, workflows, and exemplar artifacts, and more human resources trained in best practices for reproducibility. In this paper, we describe our experiences in the first two years of the Summer of Reproducibility (SoR), a mentoring program that seeks to build global capacity by enabling students around the world to work with expert mentors while producing reproducibility artifacts, tools, and education materials. We give an overview of the program, report preliminary outcomes, and discuss plans to evolve this program.

CCS Concepts

• **Information systems → Open source software**.

Keywords

Practical Reproducibility, Mentorship, Education, Capacity Building

ACM Reference Format:
Stephanie Lieggi, Fraida Fund, Kate Keahey, and Marc Richardson. 2025. SHORT: Summer of Reproducibility: Building Global Capacity for Practical Reproducibility through Hands-On Mentorship. In *ACM Conference on Reproducibility and Replicability (ACM REP '25), July 29–31, 2025, Vancouver, BC, Canada.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3736731.3746149

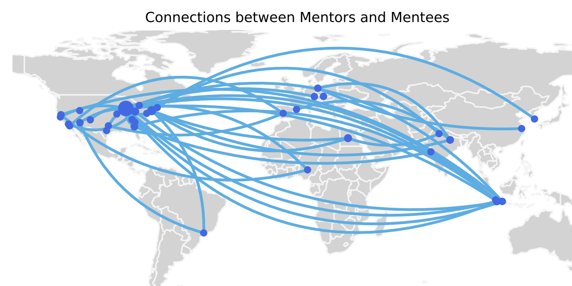

Figure 1: Map of mentor-mentee links in the 2023 and 2024 Summer of Reproducibility.

1 Introduction

With increasing recognition of the importance of reproducibility in computer science research, a number of initiatives have emerged to encourage, evaluate, and reward reproducible research practices. The most widespread is the introduction of artifact review and badging processes by academic conferences and journals. For instance, the ACM (Association for Computing Machinery) has implemented an artifact review and badging system that awards badges like Artifacts Evaluated, Artifacts Available (with sub-types Artifacts Evaluated - Functional and Artifacts Evaluated - Reusable), and Results Validated (with sub-types Results Reproduced and Results Replicated) to indicate varying levels of reproducibility [2]. There is an ongoing discussion [9–12, 30, 34, 37] about the limitations of artifact badging as a concept, problems in its implementation, and whether or not it has realized its goals, but meanwhile it remains the most widely adopted tool for advancing reproducibility.

However, efforts to incentivize the creation of reproducible research artifacts, like badges, are only one part of a healthy ecosystem for reproducible research. These efforts do not address *barriers* to reproducibility [13], such as lack of infrastructure and tools, or insufficient education and training. To make "practical reproducibility" [15]) widespread demands an increase in capacity for reproducible research among computer scientists by addressing these barriers at a global scale. While there have been efforts by individual instructors to introduce reproducibility training in computer science and computational science coursework [3, 6–8, 14, 20–22, 25, 28, 33, 35, 36], and of course individual PIs may train research assistants in reproducible research practices, these efforts are highly localized to the individual research group or educational institution.

The Summer of Reproducibility (SoR) program seeks to address this gap by funding students around the world to work with expert

This work is licensed under a Creative Commons Attribution 4.0 International License.
ACM REP '25, Vancouver, BC, Canada
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1958-5/2025/07
https://doi.org/10.1145/3736731.3746149

mentors to produce reproducibility artifacts, tools, and education materials [19, 24]. It is, to our knowledge, the first *global* initiative to build capacity for reproducible research in computer science through a distributed mentorship model. In its first two years, SoR supported 41 students working with 35 unique mentors from 10 different countries around the world (see Fig. 1) – with another 15 students currently enrolled in the 2025 cohort – across disciplines like machine learning, genomics, and computer networks. In this paper:

- We describe a distributed mentoring model for increasing global capacity for practical reproducibility.
- We report on outcomes from its first two years.
- We describe plans to evolve this program based on lessons learned.

2 Context

SoR is a hands-on mentorship program, funded by the NSF-supported REPETO project [27], aimed at broadening the reach of reproducibility practices in computer science. It takes a "science without borders" approach, focusing on involving the international academic community as well as young researchers in overall practical reproducibility efforts. SoR is based on the highly successful Google Summer of Code (GSoC), which has been supporting open source projects around the globe since 2005. While originally envisioned as a way to avoid "academic backsliding" in computer science students – and thus improving the talent pipeline to companies like Google – GSoC became a powerful avenue for building communities for open source projects while at the same time training the next generation of technologists [31]. GSoC – fully funded through the Google Open Source Office – supports on average 1,100 contributors (typically undergraduates) working with around 200 open source projects. [32]

The UC Santa Cruz Center for Research in Open Source Software (CROSS), a partner of the REPETO project, have been a mentor organization for GSoC since 2018. Having seen the extent to which its affiliated projects and their communities were buoyed by GSoC, and how students gained from the experience, SoR followed the GSoC model to provide a viable pathway for growing the practical reproducibility community of practice at a global level, and creating new tools, artifacts, and educational materials for this community [19, 24]. Projects cover a wide range of topics with students building tools and datasets to facilitate reproducible experimentation, creating and evaluating reproducible artifacts, developing best practices for artifact packaging, and creating open educational resources to promote reproducibility practices. Students, who receive a competitive stipend, acquire valuable skills while at the same time producing materials that are of value to the wider scientific community.

3 Overview of program

Copying the distributed approach of GSoC – which allows a small team of administrators in the Google Open Source office to support over a thousand students a year – the SoR team empowers mentors from any university or research institution to expand their research by working with dedicated students from around the world. These students, in turn, build much-needed skills in reproducibility and gain a deeper understanding of the need of practical reproducibility in all scientific sectors. The annual program timeline is illustrated in Fig. 2. In the first phase, which typically takes place in January-February, the SoR organization releases a call for mentors. SoR mentors — who may be faculty, researchers, or industry professionals — propose "projects ideas" based on unique research and real-world problems related to reproducibility. These project ideas – which tell a story of what the mentor wants a student to do over the summer – are placed on a central website, allowing students and other interested newcomers to discover them. The SoR organization includes this page in their application to be a GSoC mentor organization.

In late February/early March, GSoC publishes a list participating mentor organizations with links to their project pages. This commences the proposal development period, in which prospective mentees discover projects of interest and develop proposals for them. By leveraging participation in GSoC, the SoR benefits from its global reach, allowing SoR project ideas to be seen by thousands of students and early career professionals from all over the world. The SoR program then acts as a clearinghouse, providing a place for mentors and mentees to get matched.

The distributed project selection process is a critical aspect of the SoR. The ultimate selection of the mentees is left to the mentors. Mentees are encouraged to reach out to mentors early in the proposal writing process; this interaction allows for a highly iterative drafting process for the proposals that also provides a chance for mentors to get to know interested mentees ahead of the application deadline. The mentors are able to test applicants' soft and hard skills – not just their understanding of important scientific concepts or programming languages, but also their ability to take critiques, interact in a group and communicate with the project team. Their interaction with applicants prior to the submission deadline typically plays as much a part of the final decision as what is written in final proposal submitted by the applicant. While this process requires a significant amount of time for the mentor during this phase of the program, it also typically results in more successful mentor-mentee matches. The success rate for SoR projects – defined as mentees completing the summer project to the satisfaction of the mentor and meeting agreed upon milestones – is close to 100% in the first two years. While difficult to quantify, the interaction prior to the application has been cited by many mentors as the likely reason for this high level of success.

Mentors review the submitted proposals and indicate to the SoR organization adminstrators which mentees they would like to work with. The primary criteria mentors use to select mentees is simple: will the mentee's work provide a tangible benefit or solution to the proposed project? While the SoR team does a final review to prioritize projects that most meet the reproducibility related goals of the program, it is the mentor's choice that decides which mentee earn a slot. Then, mentees receive offers from the SoR organization. Mentees are offered stipends according to their level of education (undergraduate, graduate, and PhD) as well as their country of residence, using a purchasing power parity guide developed by GSoC for determining rates for mentees. The SoR stipends are structured as fellowships for mentees – so not an internship – and specifically focuses on their independent project that allows them to develop expertise through participating in research activities with the mentors.

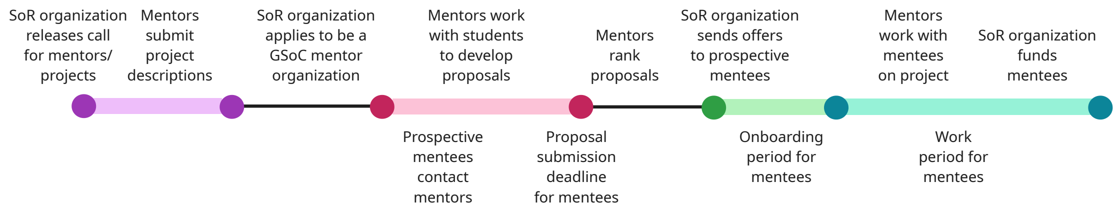

Figure 2: Program timeline, showing the roles of the organization, the mentors, and the contributors (mentees).

Matched mentors and mentees begin an "onboarding" period in May, then work on the project in earnest typically over June, July, and August. During this time, the administration of the program is managed by the SoR project team, allowing mentors to focus solely on their research and the work of the mentees. The mentees are required to submit three assessment reports (published as blogs on the SoR website) at the beginning, middle, and end of the project detailing their approaches, goals, and accomplishments. The final of these blogs includes references to the mentee's deliverables (papers, code bases, datasets, and other artifacts) as well as their final outcomes. Mentors also provide feedback to the program administration at the middle and end of the project period regarding their satisfaction with the mentee's work and the project progress.

A unique aspect of the GSoC model, which the SoR emulates, its ability to work relatively seamlessly across borders. As noted above, the SoR is able to attract students and early career professionals from across the globe to work with our mentors, and while most of our mentors are based in US universities, the program has also included mentors from the EU each year. This is illustrated in Fig. 1, which shows mentor-mentee links from the first two years of the program, based on the location of their academic affiliation.

4 Preliminary outcomes

In 2023, the program's inaugural year 36 students applied to 13 separate reproducibility projects working with a total of 18 mentors. In this first year, SoR was able to support 18 SoR students, including 15 undergraduate and 3 graduate students. The total number of mentors includes both individual mentors and mentor teams. Additionally, mentor and mentor teams are allowed to post multiple project ideas, allowing them to create small cohorts working on different aspects of their projects over the summer. SoR's first year saw student participants from seven countries: 6 were from the U.S., 5 were from Indonesia, 3 from India, and one each from Brazil, China, Egypt, and Nigeria. SoR 2024 saw an increase in applications and interested mentors; the program received 50 applications to work on 19 separate SoR projects with a combined total of 30 mentors (10 projects had two co-mentors). The program was able to support 23 SoR students - 12 undergraduates and 11 graduate students. As in 2023, the SoR 2024 program continued its global reach while at the same time bolstering the capacity of U.S. students. In 2024, 12 students were from US universities; 4 students were from India; 3 from Indonesia; and one each from Austria, China, Egypt, and Germany.

The SoR projects varied in disciplines, but their benefit to reproducibility could be broken generally into five categories. Some invited contributions to reproducibility tools or technical platforms; some relate to reproducible methodologies or workflows; some were for packaging specific high-value reproducible artifacts; some supported artifact evaluation initiatives; and others developed open educational resources for reproducibility training [24]. A selection of exemplar projects is described in Appendix A.

Signaling the extent to which our mentors appreciated the contributions of their SoR mentees, the program saw 13 mentors return for a second year, and one 2023 mentee who returned to mentor in 2024. As we prepare for the 2025 program, we are again seeing a similar number of returning mentors to the program. In their final evaluations, mentors have been overwhelmingly positive about their experiences; many SoR student summer efforts have led to follow-on work, including: journal and conference papers, follow-on projects and new research proposals; workflow improvements; education materials; and documentation and other materials to assist benchmarking and reproducibility efforts. SoR related papers have been accepted at conferences like ACM REP and EuroSys, as well as published in eScience. Many mentors are attracted to SoR because of the program's ability to recruit mentees from around the world and the extent to which the mentee's work has had a significant impact on their reproducibility related research. Specific qualities that mentors have highlighted from the first two year of the program included:

Mentees' efforts make notable positive impact on mentor's work and further research and education in the field:

- "[The student's project will] reduce the gap between domain scientists ... and computer system scientists/engineers by performing reproducible benchmarks that will provide a clear understanding of genomic workflow performance and execution infrastructure."
- "This project developed educational materials that will be used immediately in the upcoming fall semester, and beyond."
- "The contributor's improvements will be merged into the main repo in time for the Fall 2023 semester. I know of at least one university that plans to use this material this Fall."

Enthusiastic mentees working on a project of their own creation (with guidance from mentors):

- "This program is excellent for providing the opportunity to work with smart, creative, and productive young scientists and engineers. … I also learned a lot from my student."
- "[The SoR student was] enthusiastic and self-motivated.…. SoR provided us a great platform to connect with great students and have them join our projects and make progress."
- "[The SoR student] is self-motivated and is able to solve most problems independently. He is also careful and created detailed

docs and scripts to easily reproduce our evaluations. I really appreciate his work."

Global reach of program allows for creating a global co-hort of reproducibility-engaged students:

- "An interesting aspect of this project was the chance to collaborate with talented students from around the world."
- "The SoR can help me find good students to push my project and provide unique financial support to the students."

Unique method for funding students and furthering research:

- "While funding opportunities for students exist, shaping soft funding into paid research positions requires foresight and effort... [SoR] makes this process significantly easier."
- "[SoR] offers rewarding opportunities for undergraduate students to learn and contribute to our research projects, providing them with valuable hands-on experience that goes beyond what they can gain from coursework alone."
- "SoR funded the student who helped us create detailed docs and scripts to easily reproduce our project evaluations. This not only eases our own future work, but also helps us promote our work as other people can easily reproduce our results."

Mentees have been similarly positive about the program, based on surveys during and after the first two program years. They value the skills and experience they gained, and see them as having substantial impact on their career and educational goals, with many undergraduates noting that the experience prompted them to consider graduate studies for the first time:

- "The SOR was an incredibly impactful experience [that provided] a critical launchpad for my dissertation research, deepened my professional network, and expanded my methodological perspective on reproducibility."
- "I was able to deepen my understanding of the theory, state of the art, and challenges of reproducibility during the project... [It] has enabled me to secure two job opportunities that I may not have been eligible for without the experience."
- "[The SoR was] a really notable experience, the funding helped me focus on research without worrying about expenses. I was also accepted into several Ph.D. programs this cycle. I am sure the program helped bolster my application strength."

While it is still early to measure broader impacts beyond the mentors and mentees directly participating in the program, there are signs of a compounding "network effect" which can have a wider reach. One anecdotal example follows: a SoR 2023 mentor worked with a mentee to develop open educational resources related to reproducibility in machine learning (see Appendix A.3), and then used these materials to teach a graduate class at their own institution. For SoR 2024, mentor and mentee developed a project together as co-mentors, and collaborated with two new student mentees; the materials developed by this collaboration were then used in other university courses, collectively reaching 350 students in three separate courses at two universities on different continents.

5 Next steps

We have identified a number of lessons from the first two years that help us better understand how to better promote reproducibility and ensure that the SoR model remains scalable and sustainable:

Improving the management of student interest: SoR has benefited from the GSoC connection, but mentors have struggled with filtering prospective mentees, some of whom are more interested in a conventional GSoC software engineering project and do not necessarily have a strong interest in research. The SoR's success may now allow it to bring in interested students without direct linkage with GSoC. Beginning in 2026, the SoR will pilot a stand-alone effort that will not rely on the GSoC timeline and draw, but instead tap the networks of ACM REP and other reproducibility and open science focused organizations to attract students. This should allow mentors to better manage initial interest, while providing them with a strong and more focused pool of candidates.

Providing opportunities for students to highlight their work: Highlighting mentee work is a critical way to show the value of the SoR program to potential sponsors. While students blog about their work on the parent organization's website, the SoR administrators aim to increase the reach of student work to a wider audience. For example, the SoR could help with organizing poster sessions and panels at academic conferences that highlights the work of SoR and other students working in reproducibility. Moving forward, the SoR team plans to provide training and support students presenting their work in academic and professional settings. Students will be encouraged to package their work as something that could be included in or the basis of a paper or presentation, and then be provided with opportunities to get feedback on their work as they prepare to submit or present.

Supporting students interested in pursuing reproducibility career pathways: Mentees have noted that participation in the SoR inspired them to consider academic and professional paths related to reproducibility. While still small, the SoR mentee alumni co-hort is growing and could develop into an active community of practice around practical reproducibility. The open source nature of the projects fosters longer-term collaboration. This community can be leveraged to support program alumni as they undertake efforts that help bring reproducibility into the mainstream of computer science and other scientific fields.

Finding resources and funding to continue SoR efforts: The SoR is currently funded from a three-year NSF grant. To be sustainable, the project will need to diversified its sponsorship – including government agencies, foundations and industry interested in open science and workforce development. Efforts are currently underway to undertake outreach to these new sponsors as the organizers move in to year four of the program.

Moving forward, the SoR will aim to be not just a platform for collecting repeatable artifacts and educational curricula and mainstreaming reproducibility best practices, but also an avenue for workforce and scientific development that transforms the way students do science. The SoR will help promote the practice of regularly examining the results produced by others in the new generation of technologists, helping researchers think about reproducibility as a primary concern from the inception of their work.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2226406, Grant No. 2226407, and Grant No. 2226408.

References

[1] Jordan Ash and Ryan P Adams. 2020. On Warm-Starting Neural Network Training. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., Red Hook, NY, USA, 3884–3894. https://proceedings.neurips.cc/paper_files/paper/2020/file/288cd2567953f06e460a33951f55daaf-Paper.pdf

[2] Association for Computing Machinery. 2020. Artifact Review and Badging Version 1.1. https://www.acm.org/publications/policies/artifact-review-and-badging-current Accessed: 2025-04-04.

[3] Richard Ball, Norm Medeiros, Nicholas W Bussberg, and Aneta Piekut. 2022. An Invitation to Teaching Reproducible Research: Lessons from a Symposium. *Journal of Statistics and Data Science Education* 30, 3 (2022), 209–218.

[4] Chameleon Cloud. 2025. Chameleon Cloud Portal. https://chameleoncloud.org/ Accessed: 2025-07-11.

[5] Chameleon Cloud. 2025. Trovi: A Collection of Shared Artifacts. https://trovi.chameleoncloud.org/dashboard/ Accessed: 2025-07-11.

[6] Thomas Cokelaer, Sarah Cohen-Boulakia, and Frédéric Lemoine. 2023. Reprohackathons: promoting reproducibility in bioinformatics through training. *Bioinformatics* 39, Supplement_1 (2023), i11–i20.

[7] Fraida Fund. 2023. We Need More Reproducibility Content Across the Computer Science Curriculum. In *Proceedings of the 2023 ACM Conference on Reproducibility and Replicability* (Santa Cruz, CA, USA) *(ACM REP '23)*. Association for Computing Machinery, New York, NY, USA, 97–101. doi:10.1145/3589806.3600033

[8] Yayu Gao, Chengwei Zhang, Xiaojun Hei, and Guohui Zhong. 2019. Learning networking by reproducing research results in an ns-3 simulation networking laboratory course. In *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*. IEEE, IEEE, New York, NY, USA, 1–6.

[9] Quentin Guilloteau, Florina Ciorba, Millian Poquet, Dorian Goepp, and Olivier Richard. 2024. Longevity of Artifacts in Leading Parallel and Distributed Systems Conferences: a Review of the State of the Practice in 2023. In *Proceedings of the 2nd ACM Conference on Reproducibility and Replicability* (Rennes, France) *(ACM REP '24)*. Association for Computing Machinery, New York, NY, USA, 121–133. doi:10.1145/3641525.3663631

[10] Quentin Guilloteau, Millian Poquet, Jonas H Müller Korndörfer, and Florina M Ciorba. 2024. Artifact Evaluations as Authors and Reviewers : Lessons, Questions, and Frustrations. In *Community Workshop on Practical Reproducibility in HPC*. Publisher, Atlanta (GA), United States, xx–yy. https://hal.science/hal-04764265

[11] Ben Hermann. 2022. What Has Artifact Evaluation Ever Done for Us? *IEEE Security & Privacy* 20, 5 (2022), 96–99. doi:10.1109/MSEC.2022.3184234

[12] Ben Hermann, Stefan Winter, and Janet Siegmund. 2020. Community expectations for research artifacts and evaluation processes. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Virtual Event, USA) *(ESEC/FSE 2020)*. Association for Computing Machinery, New York, NY, USA, 469–480. doi:10.1145/3368089.3409767

[13] Serge PJM Horbach, Nicki Lisa Cole, Simone Kopeinik, Barbara Leitner, Tony Ross-Hellauer, and Joeri Tijdink. 2025. How to get there from here? Barriers and enablers on the road towards reproducibility in research. *Journal* 1, 1 (2025), 1–33.

[14] Nestoras Karathanasis, Daniel Hwang, Vibol Heng, Rimal Abhimannyu, Phillip Slogoff-Sevilla, Gina Buchel, Victoria Frisbie, Peiyao Li, Dafni Kryoneriti, and Isidore Rigoutsos. 2022. Reproducibility efforts as a teaching tool: A pilot study. *PLOS Computational Biology* 18, 11 (11 2022), 1–11. doi:10.1371/journal.pcbi.1010615

[15] Kate Keahey, Jason Anderson, Mark Powers, and Adam Cooper. 2023. Three pillars of practical reproducibility. In *2023 IEEE 19th International Conference on e-Science (e-Science)*. IEEE, IEEE, Piscataway, NJ, USA, 1–6.

[16] Kate Keahey, Jason Anderson, Zhuo Zhen, Pierre Riteau, Paul Ruth, Dan Stanzione, Mert Cevik, Jacob Colleran, Haryadi S Gunawi, Cody Hammock, et al. 2020. Lessons learned from the chameleon testbed. In *2020 USENIX annual technical conference (USENIX ATC 20)*. USENIX Association, Berkeley, CA, USA, 219–233.

[17] Alexander Kolesnikov, Alexey Dosovitskiy, Dirk Weissenborn, Georg Heigold, Jakob Uszkoreit, Lucas Beyer, Matthias Minderer, Mostafa Dehghani, Neil Houlsby, Sylvain Gelly, Thomas Unterthiner, and Xiaohua Zhai. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR '20')*. N/A, N/A, N/A.

[18] Klaus Kraßnitzer. 2025. AutoAppendix: Towards one-click Reproduction of Computational Artifacts. *arXiv preprint arXiv:2504.00876* N/A, N/A (2025), –. preprint.

[19] Stephanie Lieggi and Carlos Maltzahn. 2023. Summer of Reproducibility: Exposing Students to Research Advancing Practical Reproducibility. *Better Scientific Software (BSSw)* December, N/A (2023), N/A. https://bssw.io/blog_posts/summer-of-reproducibility-exposing-students-to-research-advancing-practical-reproducibility

[20] Ana Lucic, Maurits Bleeker, Sami Jullien, Samarth Bhargav, and Maarten de Rijke. 2022. Reproducibility as a Mechanism for Teaching Fairness, Accountability, Confidentiality, and Transparency in Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 11 (6 2022), 12792–12800. doi:10.1609/aaai.v36i11.21558

[21] Wolfgang Mauerer, Stefan Klessinger, and Stefanie Scherzinger. 2023. Beyond the Badge: Reproducibility Engineering as a Lifetime Skill. In *Proceedings of the 4th International Workshop on Software Engineering Education for the Next Generation* (Pittsburgh, Pennsylvania) *(SEENG '22)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3528231.3528359

[22] K Jarrod Millman, Matthew Brett, Ross Barnowski, and Jean-Baptiste Poline. 2018. Teaching computational reproducibility for neuroimaging. *Frontiers in Neuroscience* 12 (2018), 727.

[23] Leonardo Murta, Vanessa Braganholo, Fernando Chirigati, David Koop, and Juliana Freire. 2015. noWorkflow: Capturing and Analyzing Provenance of Scripts. In *Provenance and Annotation of Data and Processes*, Bertram Ludäscher and Beth Plale (Eds.). Springer International Publishing, Cham, 71–83.

[24] UCSC Open Source Program Office. 2022. The Open Source Research Experience Program. https://ucsc-ospo.github.io/osre/ Accessed: 2025-04-04.

[25] Joel Ostblom and Tiffany Timbers. 2022. Opinionated Practices for Teaching Reproducibility: Motivation, Guided Instruction and Practice. *Journal of Statistics and Data Science Education* 30, 3 (2022), 241–250. doi:10.1080/26939169.2022.2074922

[26] Joao Felipe Pimentel, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. 2017. noWorkflow: a tool for collecting, analyzing, and managing provenance from python scripts. *Proceedings of the VLDB Endowment* 10, 12 (2017), N/A.

[27] REPETO Project. 2023. REPETO: Building a Network for Practical Reproducibility in Experimental Computer Science. https://repeto.cs.uchicago.edu Accessed: 2025-04-04.

[28] Melissa L. Rethlefsen, Hannah F. Norton, Sarah L. Meyer, Katherine A. MacWilkinson, Plato L. Smith II, and Hao Ye. 2022. Interdisciplinary Approaches and Strategies from Research Reproducibility 2020: Educating for Reproducibility. *Journal of Statistics and Data Science Education* 30, 3 (2022), 219–227. doi:10.1080/26939169.2022.2104767

[29] Mohamed Saeed and Fraida Fund. 2024. Scaffolding Reproducibility for the Machine Learning Classroom. In *Proceedings of the 2nd ACM Conference on Reproducibility and Replicability* (Rennes, France) *(ACM REP '24)*. Association for Computing Machinery, New York, NY, USA, N/A. doi:10.1145/3641525.3663631

[30] Damien Saucez, Luigi Iannone, and Olivier Bonaventure. 2019. Evaluating the artifacts of SIGCOMM papers. *SIGCOMM Comput. Commun. Rev.* 49, 2 (May 2019), 44–47. doi:10.1145/3336937.3336944

[31] Jefferson De Oliveira Silva, Igor Scaliante Wiese, Daniel M. German, Igor Fabio Steinmacher, and Marco Aurélio Gerosa. 2017. How Long and How Much: What to Expect from Summer of Code Participants?. In *2017 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, Piscataway, NJ, USA, 69–79. doi:10.1109/ICSME.2017.81

[32] Google Open Source. 2009. The History of GSoC. https://google.github.io/gsocguides/mentor/the-history-of-gsoc Accessed: 2025-04-04.

[33] Lars Vilhuber, Hyuk Harry Son, Meredith Welch, David N. Wasser, and Michael Darisse. 2022. Teaching for Large-Scale Reproducibility Verification. *Journal of Statistics and Data Science Education* 30, 3 (2022), 274–281. doi:10.1080/26939169.2022.2074582

[34] Stefan Winter, Christopher S. Timperley, Ben Hermann, Jürgen Cito, Jonathan Bell, Michael Hilton, and Dirk Beyer. 2022. A retrospective study of one decade of artifact evaluations. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Singapore, Singapore) *(ESEC/FSE 2022)*. Association for Computing Machinery, New York, NY, USA, 145–156. doi:10.1145/3540250.3549172

[35] Lisa Yan and Nick McKeown. 2017. Learning networking by reproducing research results. *ACM SIGCOMM Computer Communication Review* 47, 2 (2017), 19–26.

[36] Burak Yildiz, Hayley Hung, Jesse H Krijthe, Cynthia CS Liem, Marco Loog, Gosia Migut, Frans A Oliehoek, Annibale Panichella, Przemysław Pawełczak, Stjepan Picek, et al. 2021. ReproducedPapers.org: Openly teaching and structuring machine learning reproducibility. In *Reproducible Research in Pattern Recognition: Third International Workshop, RRPR 2021, Virtual Event, January 11, 2021, Revised Selected Papers*. Springer International Publishing, Cham, Switzerland, 3–11.

[37] Noa Zilberman and Andrew W. Moore. 2020. Thoughts about Artifact Badging. *SIGCOMM Comput. Commun. Rev.* 50, 2 (May 2020), 60–63. doi:10.1145/3402413.3402422

A Summer of Reproducibility Projects

In this appendix, we describe several exemplar projects in three areas (reproducibility tools, artifact evaluation, and education around reproducibility) from the 2023 and 2024 Summer of Reproducibility, and their preliminary outcomes.

A.1 Project: noWorkflow

Project description. The noWorkflow software project [23, 26] is an open-source tool that systematically and transparently collects provenance from Python scripts, with the goal of allowing scientists to benefit from provenance data analysis even when they don't use a workflow system. As part of SoR 2023, mentors João Felipe Pimentel (Northern Arizona University) and Juliana Freire (New York University) proposed a project inviting contributions to this reproducibility tool, and matched with mentee Jesse Lima (Sao Paulo University).

Preliminary outcomes. Jesse's goal was to better support provenance capture with noWorkflow in typical data science and machine learning experiments. Over the course of the summer, Jesse identified several challenges in the use of noWorkflow for data science and machine learning, notably:
- Typical pattern of non-linear execution in Jupyter Notebooks.
- Use of complex data types (e.g., matrices, tensors).
- Compatibility issues with newer Jupyter and IPython versions.

and he implemented several improvements in noWorkflow for Jupyter notebook users. Jesse also created a use case/tutorial demonstrating the use of noWorkflow when developing a machine learning model for credit card fraud detection. These contributions have since been merged into the noWorkflow code base (PR 153).

A.2 Project: AutoAppendix

Project description. "AutoAppendix" was a SoR 2024 project proposed by mentor Sascha Hunold (TU Wien) for artifact evaluation, with the following project description:

> The SC Conference Series, a leading forum on High Performance Computing (HPC), supports scientific rigor through an enhanced reproducibility of accepted papers. To that end, all manuscripts submitted to the SC Technical Papers program must contain an Artifact Description. Authors of accepted papers may request reproducibility badges, for which an Appendix describing the Artifact Evaluation is required.
> In recent years, Chameleon has facilitated SC's reproducibility initiative by enabling authors to develop and share computational, reproducible artifacts through the Chameleon cloud. The Chameleon platform helps authors and reviewers to easily share computational artifacts, which are included in the papers' artifact appendices.
> The proposed project aims to assess all AD/AE appendices submitted for reproducibility badge requests. This evaluation will focus on AD/AE appendices that utilized the Chameleon cloud as the execution platform, examining their potential for automation. Our aim is to evaluate the feasibility of fully automating various components of the appendices. Students will engage directly with the chairs of the SC24 Reproducibility Initiative in this effort.

The project was matched to mentee Klaus Kraßnitzer (TU Wien).

Preliminary outcomes. During the summer, Klaus reviewed the 45 SC24 papers seeking the "Results Replicated" badge, and from those, selected 18 that would be easily replicable on Chameleon Cloud [4, 16]. For these 18 artifacts, the results of Klaus's replication are detailed in a technical report [18], along with suggestions for authors and best practices identified as a result of the experience. The project also developed several reproducible research artifacts for the Trovi [5] artifact sharing portal.

A.3 Project: Using Reproducibility in Machine Learning Education

Project description. "Using Reproducibility in Machine Learning Education" was a SoR 2023 project proposed by mentor Fraida Fund (New York University) to advance education related to reproducibility, with the following project description:

> The computer science and engineering classroom is as essential part of the reproducibility "ecosystem" - because of broad reach and potential for big impact, and because for many students, the classroom is their first exposure to research in their field. For machine learning in particular, reproducibility is an important element of the research culture, and can be a valuable part of any introductory or advanced courses in the field. These projects will develop highly interactive open educational resources, that may be adopted by instructors of graduate or undergraduate machine learning courses to incorporate more instruction about reproducibility and reproducible research.

After the proposal development period, the project was matched to mentee Mohamed Saeed (Alexandria University, Egypt).

Preliminary outcomes. During the summer, Mohamed prepared two interactive learning modules for use in machine learning courses:

Warm starting neural networks reproduces the result in [1], which shows that a model that is trained on some data, then later trained on more data (warm start), has worse performance than a model that is trained from scratch on all of the data. This paper was selected because students can understand it with only a basic machine learning background (i.e. without knowledge of modern neural network architectures). The learning materials teach students how to:
- Read a paper and identify claims and the supporting experimental evidence.
- Use provided code to replicate experiments and validate claims.
- Consider the computational cost associated with reproducing results.

Vision transformer is based on [17], which introduces the vision transformer. This paper was selected because it was a recent and transformative work in computer vision. The learning materials teach students how to:
- Determine when data and/or pre-trained models are available, and which claims can or cannot be validated as a result.
- Use pre-trained models to replicate experiments and validate claims.

All of the materials developed are publicly available in the Chameleon Trovi artifact repository [15]. They have been adopted for classroom use by the mentor, and Mohamed presented a poster [29] at ACM REP the following year.