



# Software Training in HEP

Sudhir Malik<sup>1</sup> · Samuel Meehan<sup>2</sup> · Kilian Lieret<sup>3</sup> · Meirin Oan Evans<sup>4</sup> · Michel H. Villanueva<sup>5</sup> · Daniel S. Katz<sup>6</sup> · Graeme A. Stewart<sup>2</sup> · Peter Elmer<sup>7</sup> · Sizar Aziz<sup>8</sup> · Matthew Bellis<sup>18</sup> · Riccardo Maria Bianchi<sup>25</sup> · Gianluca Bianco<sup>30,31</sup> · Johan Sebastian Bonilla<sup>23</sup> · Angela Burger<sup>24</sup> · Jackson Burzynski<sup>27</sup> · David Chamont<sup>8</sup> · Matthew Feickert<sup>6</sup> · Philipp Gadow<sup>12</sup> · Bernhard Manfred Gruber<sup>2,34,35</sup> · Daniel Guest<sup>15</sup> · Stephan Hageboeck<sup>2</sup> · Lukas Heinrich<sup>2</sup> · Maximilian M. Horzela<sup>16</sup> · Marc Huwiler<sup>26</sup> · Clemens Lange<sup>2</sup> · Konstantin Lehmann<sup>17</sup> · Ke Li<sup>19</sup> · Devdatta Majumder<sup>28</sup> · Judita Mamužić<sup>10</sup> · Kevin Nelson<sup>22</sup> · Robin Newhouse<sup>13</sup> · Emery Nibigira<sup>14</sup> · Scarlet Norberg<sup>1</sup> · Arturo Sánchez Pineda<sup>11</sup> · Mason Proffitt<sup>19</sup> · Brendan Regnery<sup>23</sup> · Amber Roepe<sup>33</sup> · Stefan Roiser<sup>2</sup> · Henry Schreiner<sup>7</sup> · Oksana Shadura<sup>21</sup> · Giordon Stark<sup>9</sup> · Stephen Nicholas Swatman<sup>2,20</sup> · Savannah Thais<sup>7</sup> · Andrea Valassi<sup>2</sup> · Stefan Wunsch<sup>2,16</sup> · David Yakobovitch<sup>32</sup> · Siqi Yuan<sup>29</sup>

Received: 1 August 2021 / Accepted: 20 September 2021 / Published online: 8 October 2021  
© The Author(s) 2021

## Abstract

The long-term sustainability of the high-energy physics (HEP) research software ecosystem is essential to the field. With new facilities and upgrades coming online throughout the 2020s, this will only become increasingly important. Meeting the sustainability challenge requires a workforce with a combination of HEP domain knowledge and advanced software skills. The required software skills fall into three broad groups. The first is fundamental and generic software engineering (e.g., Unix, version control, C++, and continuous integration). The second is knowledge of domain-specific HEP packages and practices (e.g., the ROOT data format and analysis framework). The third is more advanced knowledge involving specialized techniques, including parallel programming, machine learning and data science tools, and techniques to maintain software projects at all scales. This paper discusses the collective software training program in HEP led by the HEP Software Foundation (HSF) and the Institute for Research and Innovation in Software in HEP (IRIS-HEP). The program equips participants with an array of software skills that serve as ingredients for the solution of HEP computing challenges. Beyond serving the community by ensuring that members are able to pursue research goals, the program serves individuals by providing intellectual capital and transferable skills important to careers in the realm of software and computing, inside or outside HEP.

**Keywords** Software · Training · HEP

## Introduction

Particle physics in the coming decades will continue to explore the fundamental workings of the universe. This requires upgrading existing major facilities such as the Large Hadron Collider (LHC) to the High Luminosity LHC [1] and building new facilities like the Long-Baseline Neutrino Facility (LBNF) [2] and Deep Underground Neutrino Experiment (DUNE) [3], among many others. To realize the full physics potential of this work, an equivalent investment must be made into the software required to collect, process,

and analyze the deluge of the data recorded. Recent efforts such as the HSF [4] and IRIS-HEP [5] are facilitating cooperation and common efforts in HEP software and computing worldwide to develop state-of-the-art software cyberinfrastructure required to meet the challenges of the upcoming HEP experiments' data-intensive scientific research. The rapid evolution of computing technology with a concomitant increase in the complexity of software algorithms for analysis requires developers to acquire a broad portfolio of programming skills to enable future discoveries.

It is critical that all stakeholders across HEP make a major effort to provide a strong foundation for new researchers entering the field. The researchers must be brought up to date with new software technologies, concurrent programming, and artificial intelligence, and must maintain,

✉ Sudhir Malik  
sudhir.malik@upr.edu

Extended author information available on the last page of the article

improve, and sustain the existing HEP software. However, young researchers graduating from universities worldwide currently do not receive adequate preparation in modern computing practices to respond to the growing needs related to the above experimental challenges. A community white paper [6] outlined the initiatives to address training needs and issues that need to be taken into account for these to be successful. In the last two years, the HSF Training working group, together with IRIS-HEP and FIRST-HEP [7] and partnering with The Carpentries [8], has begun development of a software training program. The efforts of this group have been focused on two specific goals: (1) developing material for an introductory HEP software curriculum, and (2) teaching this curriculum to HEP scientists. Thus far, over 1000 people in HEP and related computing areas have been trained. This paper describes the activities, the curriculum, and future directions of HEP software training.

## Organization

The HSF Training working group, which is led by three co-conveners, engages with different experimental collaborations and initiatives such as IRIS-HEP, FIRST-HEP, and The Carpentries. The training group has weekly public meetings [9] to plan and assess progress. This is where ideas and proposals are discussed and events are planned. These meetings are held remotely using Zoom and live notes are maintained for anyone unable to join. Training events are announced via several email lists, with registration and timetables organized using Indico [10].

The style and pedagogy of the training is heavily inspired by The Carpentries. The training is student-centric, suitable for self-study, and experiment agnostic, with reusable study material that is open source and open access, and hosted in the HSF's Training repositories on GitHub [11]. We encourage participants to provide feedback and suggestions for improvement by opening issues in these repositories or to directly help with the development by opening pull requests. In most cases the training material is in the form of a website that is built from files written in the easy-to-learn Markdown language. The website is automatically built using the static site generator Jekyll [12] via GitHub Pages [13] and an adapted and extended template from The Carpentries [14, 15]. Thus, the entry barrier to contribute to the material is fairly low, as only basic knowledge of git is required (and in most cases, all necessary steps can be performed via the GitHub web interface). All lessons are listed in the HSF Training Center [16], which provides an overview of the available training modules and serves as an entry point for anyone wishing to learn.

Based on our experiences, we have also formalized the procedure used to organize a training event and have

compiled our knowledge in a compact guide [17]. As organization is all about dividing work, we distinguish between three relevant roles at our events:

- *Instructors* are subject-matter experts who develop training material and then teach it, either in person, in recorded live sessions, or by recording videos before the event. Instructors are the primary academic drivers of the program at large and provide guidance to mentors and students alike. They gain experience in curriculum design with a focus on optimizing pedagogy for all learning styles.
- *Mentors* work closely with participants, for example, by conducting small group mentoring sessions with ideally only five students per mentor. They optimize the learning environment for individual participants and help them persevere. They are critical to the success of any event and through participation as a mentor not only serve the community, but develop pedagogical communication skills that are transferable to other aspects of their research/teaching portfolio.
- *Facilitators* take care of organizational aspects. They are responsible for putting together all of the pieces of the puzzle to successfully execute the full event while serving as the primary point of reference for participants to communicate. They take on a dynamic responsibility beyond the “core content” of the training event itself, and they also learn the essential “soft skills” necessary to be a leader in the academic community and beyond.

All three groups are collectively referred to as *educators*. While some of our members are allowed to use a fraction of their regular working hours for our teaching activities, most of the work is done on a voluntary basis. As creating training material and teaching requires a lot of commitment and time, it is therefore of great importance to acknowledge the efforts of everyone involved. Currently this is mostly achieved by listing helping community members on the pages of the relevant training and on a central community page [18].

Finally, Blueprint workshops [19] and hackathons [20] are organized to brainstorm new training events, develop content, and discuss improvements. The travel cost for educators and video captioning of training material have been supported by IRIS-HEP and FIRST-HEP.

## Curriculum

An initial survey of the software and training needs of the HEP community was conducted in February of 2019 [21]. This was followed by the development of “prototype” course modules and pilot training events from which feedback from participants was solicited.

Based on the surveys and the experiences gathered at the events, the course structure was extended into a full curriculum consisting of a variety of training modules. Each training module is independent from the others (except for some clearly marked requirements), so that students can prioritize certain skills before others. This is especially important in academia because students are often expected to work directly towards scientific results with minimal time given for acquiring software knowledge or best practices.

The most basic skill set (Unix shell, Python, and git) is covered by modules directly developed by the Software Carpentry [22]. A large module that covers the basics of modern C++ is currently in development and other modules focusing on development in C++ such as CMake have already been taught with great success.

This is complemented by a series of broader software engineering topics, such as continuous integration and deployment using both GitHub Actions and GitLab CI as examples. These modules are also particularly relevant for analysis preservation, for which modules covering domain-specific software such as REANA [23] are in development.

A lesson on machine learning and a lesson specifically targeting machine learning with GPUs started a data analysis techniques curriculum section. Similarly, important are HEP-specific tools, especially the ROOT data analysis framework [24] and packages such as uproot [25] that enhance its interoperability with non-HEP-specific packages.

Finally, development is ongoing of modules that cover advanced topics that are important for students striving to become core developers, such as code documentation, performance optimization and parallel programming.

The module list [16] and the material evolves continuously depending on input from participants and person-power available; as it is open source, any interested stakeholder can contribute.

## Training Events

During the initial period of training, 150 people received “introductory” software skills training at Fermilab (FNAL), Argonne National Lab (ANL), Lawrence Berkeley Lab (LBNL), and CERN [26–29]. National labs are the hub of the HEP community and provide an environment where it is easier to reach a diverse population of participants with good infrastructure for in-person training. At the CoDaS-HEP school [30], over 50 people participated in the advanced “computing bootcamp” software training. These training events were in-person.

However, the COVID-19 pandemic necessitated a rapid adjustment to virtual platforms, which evolved throughout the course of 2020 as we gained experience. The events that we had to pivot to use a virtual environment include

training on continuous integration and deployment [31, 32], Docker [33], machine learning on GPUs [34] and C++ [35] (organized together with SIDIS [36]).

To date, nearly 100 educators have taught over 1000 participants in about a dozen training events. Valuable lessons have been learned regarding in-person and virtual training. There is very clear and detailed guidance for anyone willing to host, request or organize a training while staying aligned with the approach, philosophy, and code of conduct of the HSF Training group so as to make the tools and techniques that are developed persistent, reusable, and broadly accessible [17].

While in-person events offer more opportunities for active and efficient engagement of participants and community building, they are generally more exclusive: participants need sufficient funding and extra preparation time to arrange travel to the venue. Hosts have to book specially arranged/equipped rooms with multiple projectors and screens to simultaneously show teaching materials and slides. The space constraints typically limit the number of participants to a few dozen and a long lead time is required for the logistics. Our in-person events have been managed by about five educators, which is necessary for the “hands-on” aspect to be successful. These educators also need to make a large time commitment; they cannot just present their material and leave. Virtual events have a broader reach of participant attendance that is much higher compared to in-person events and enable a considerably more equitable service to the community. Since the teaching materials are fully preserved via lesson creation and YouTube videos beforehand, an inability to attend during the scheduled time does not considerably degrade learning. Finally, these video materials are captioned to be inclusive of those with hearing impairments. Captioning videos for a week-long event (~\$50/day) is considerably more economical than the cost of a hired sign language interpreter (~\$1000/day).

The disadvantage of virtual events, however, is that it is difficult for educators and participants to interact closely—you just cannot recreate the in-person environment on Zoom. Educators and participants have to plan and act upon their spread across time zones in the best possible way. It is also challenging to keep everyone engaged and on the same page due to the pervasive culture of “multi-tasking” within HEP. Due to this issue, although initial registrations for these events are very high, the actual attendance is typically only 50% of those who have registered. The online experience is more prone to be distracted by other professional duties. However, it should be noted that this does not mean that there is a lesser degree of learning occurring at the training event. Tools such as Mattermost, discord, and Slack have been effectively deployed for asynchronous communication, both during and after the event.

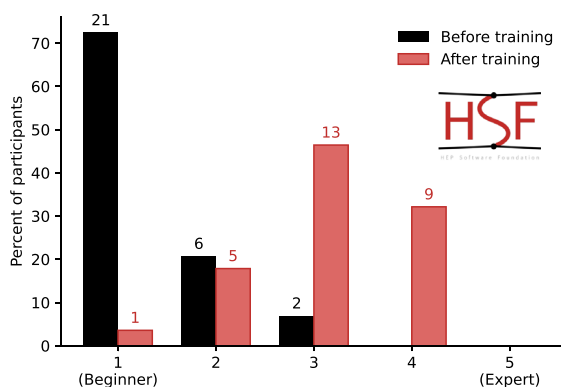
In general, devoting full time to training is always challenging. Though there is widespread desire to engage in training, there is an institutional culture that prioritizes immediate research activity over dedicated professional development, even though the latter will lead to higher productivity in the long term.

## Feedback

Feedback is required for us to evaluate if we are effectively facilitating learning and to ensure the success of future training. Every training has a pre- and post-survey to collect feedback from the participants. This includes a set of baseline questions pertaining to demographics and questions to assess the quality and method of training. These questions can be adapted to the nature and topic of each training event. In addition, we organize a “post-mortem” session among the educators to internally discuss the successes and failures. This typically occurs after completion of the results of the post- (and pre-) workshop surveys, which guide the discussion. Finally, a short presentation about the training experience is presented at the HSF Training weekly meeting and/or at the HSF all-working-groups planning meeting.

Figure 1 shows feedback on a training event involving containerization with Docker [37]: clearly the training made a difference. However, we are aware that this type of “learning evaluation” does not fully encompass the impact of our training. It only probes the perceived and self-reported learning of a skill. Instead, what is needed is a survey that is conducted sufficiently later to understand how well the learned skill is being applied in the context of research.

How do you rate your knowledge and abilities when using Docker?



**Fig. 1** The self-reported pre- and post-training level of knowledge on the topic of Docker (a software container technology)

## Community

The solutions to future computing challenges require a large workforce trained in a wide range of software skills. To train this workforce, we rely on an active community whose members are enthusiastic and motivated to teach. Our members include people with various roles and backgrounds in HEP, such as experimental physicists from different collaborations, as well as software engineers from different institutes. As we scale our training activities, we also have members from nuclear physics and computer science as well. Members of The Carpentries teach part of our very basic curriculum by an agreement via membership subscription through IRIS-HEP. The overall diversity of the background of the instructors and mentors adds great value to the training. Each educator brings their own flavor of experience from a different computing environment with a common goal of creating, teaching, and sustaining a common set of software skills.

As the success of our mission depends crucially on the motivation and participation of the community, we cultivate a strong sense of community ownership and pay special attention to acknowledge contributions of all kinds. We also encourage the participants in our training events to remain active or become more active, share feedback, and in particular, to sign up to be a mentor in one of the next iterations of the same training module. If former participants do not yet feel confident about their mentoring skills, we offer to match them with a more senior mentor. In the same way, we encourage mentors to become instructors or facilitators and to become more and more active in our organization. By actively engaging participants and educators throughout the training community, we can sustain and nurture a culture of intentional learning and grow our community in an organic fashion [38].

Educators not only provide an invaluable service to the HEP community, but they also get the opportunity to develop and sharpen their pedagogical skills and enhance their professional portfolio. About two-thirds of the HEP workforce eventually work outside of HEP, such as in the software industry and in data science. The training makes a meaningful difference in the preparation for such careers in terms of software knowledge and experience, and enhances the employability for both the educators and participants. The skills taught and learned, like Python, machine learning, and data analysis, align with the needs of the software industry and strengthen the job profile of a physicist to work in industry. At the same time, recognizing the importance of software skills within HEP may hopefully help to provide more incentives and clearer career paths in academia to those who want to pursue their career within HEP, or in other scientific research fields.

In particular, strengthening the research software engineer career path [39, 40] could significantly help retain the expertise within the HEP community.

## Sustainability

Sustainable software [41] is essential for HEP. A sustainable training program [42] is key to pursuing this goal. While continuing the existing work, it will be essential to spread the training events and training expertise geographically to keep the costs low and move to an online training model to reduce financial burdens that accompany in-person training. In parallel, it is important that as the curriculum grows, it begins to include material specifically aimed at making software sustainable.

Training should be structured so that a minimal set of people are needed for maintenance and costs per event are minimized. Growing the community is an important aspect of sustaining the workforce. Providing recognition and possible financial incentives can keep the community vibrant and motivated. The community should recognize and appreciate the broader value of our software training, which prepares a workforce to solve computing challenges that are essential to advance our field and society at large.

To lead software training across HEP and related communities over the long run, we need a core team whose main focus is to support the overall mission of HEP software training. To scale up training efforts, we need to build mentorship and leadership at the local and regional level supported by the core team. Specifically, while we have started the following set of activities, we need to scale up by:

- Engaging more HEP labs, institutes, and universities in this endeavor.
- Promoting equity, diversity, inclusion, and accessibility in participation across HEP communities and being mindful of under-resourced institutions in different geographical regions.
- Establishing a mechanism to get feedback from our communities and improve the training.
- Ensuring that our core team and volunteers are afforded opportunities to grow professionally and have career paths.
- Exploring ways to manage a financial support model to share costs in the long term.

## Broader Impacts

HSF-led training is multilayered, with a basic HEP software curriculum progressing to HEP-specific physics tools. Integrated with this is a growing outreach program that is

essential to building an influx of software workforce and training young minds, catching them early in their educational development. For example, several outreach events are organized on introducing Python programming to K-12 teachers [43] under IRIS-HEP and FIRST-HEP. The teachers can turn this into a classroom experience for their students where physics, astronomy, and math courses can have problem solving components that integrate programming with Python. In outreach events, the teachers analyze and interpret physics data with Python using Google Colab [44], which allows them to work directly in the web browser without requiring any additional setup. Workshops teaching the basics of machine learning to school teachers are also organized [45]. We plan to scale this experience by partnering with other stakeholders in HEP outreach, for example, Quarknet [46], which already has a well developed network of teachers and schools taking part in HEP outreach programs.

## Summary

HSF and IRIS-HEP are creating software training and ensuring sustainability of software in HEP for years to come. The training material is open source and open access, shared publicly via GitHub. This allows anyone to join the discussion and make contributions by proposing changes, thereby continuously improving the available material. This process is guided by continual feedback solicited from the participants of the training events. Finally, we have established a growing community of educators to broadly promote a culture within HEP that goes beyond valuing software skills, but also values the teaching of those skills to others. In doing so, we aim to foster a more active, inclusive, and diverse scientific community. By leading software training across HEP and related communities, we will be able to meet the challenges in the field and beyond.

**Acknowledgements** We would like to thank all members of our community for their contributions, big or small. Your voluntary work helps prepare the next generation of HEP scientists for the big challenges that lie ahead!

**Funding** This work is supported in part by National Science Foundation Cooperative Agreement OAC-1836650 and grants OAC-1829707 and OAC-1829729.

**Data Availability Statement** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of Interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- High-luminosity LHC. <https://home.cern/science/accelerators/high-luminosity-lhc>. Accessed 6 Oct 2021
- Papadimitriou V, Ammigan K, au2 JAJ, Anderson KE, Andrews R, Bocean V, Crowley CF, Eddy N, Hartsell BD, Hays S, Hurh P, Hylen J, Johnstone JA, Kasper P, Kobilarcik T, Krafczyk GE, Lundberg B, Marchionni A, Mokhov NV, Moore CD, Pushka D, Rakhno I, Reitzner SD, Schlabach P, Sidorov V, Stefanik AM, Tariq S, Valerio L, Vaziri K, Velev G, Vogel G, Williams K, Zwaska RM, Densham C (2017) Design of the LBNF beamline. <https://arxiv.org/abs/1704.04471>
- Deep Underground Neutrino Experiment. <https://bnf-dune.fnal.gov/>. Accessed 6 Oct 2021
- High Energy Physics Software Foundation. <https://hepsoftwarefoundation.org/>. Accessed 6 Oct 2021
- Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP). <https://iris-hep.org>. Accessed 6 Oct 2021
- HEP Software Foundation, Berzano D, Bianchi RM, Elmer P, Gleyzer SV, Harvey J, Jones R, Jouvin M, Katz DS, Malik S, Menasce D, Neubauer M, Psihas F, Navarro AP, Stewart GA, Tunnell C, Vasel JA, Wang SJ (2019) HEP Software Foundation community white paper working group—training, staffing and careers. <https://arxiv.org/abs/1807.02875>
- Framework for Integrated Research Software Training in High Energy Physics. <https://first-hep.org>. Accessed 6 Oct 2021
- The Carpentries. <https://carpentries.org/>. Accessed 6 Oct 2021
- HSF Training and Careers Working Group meetings. <https://indico.cern.ch/category/10294/>. Accessed 6 Oct 2021
- HSF training events. <https://indico.cern.ch/category/11386/>. Accessed 6 Oct 2021
- HSF training and educational material GitHub organization. <https://github.com/hsf-training>. Accessed 6 Oct 2021
- Jekyll: Transform your plain text into static websites and blogs. <https://jekyllrb.com/>. Accessed 6 Oct 2021
- GitHub Pages: Websites for you and your projects. <https://pages.github.com/>. Accessed 6 Oct 2021
- HSF Training module template repository. <https://github.com/hsf-training/hsf-styles>. Accessed 6 Oct 2021
- The Carpentries training module template repository. <https://github.com/carpentries/styles>. Accessed 6 Oct 2021
- Hsf training center. <https://hepsoftwarefoundation.org/training/curriculum.html>. Accessed 6 Oct 2021
- How to host an HSF training event. <https://hepsoftwarefoundation.org/training/howto-event.html>. Accessed 6 Oct 2021
- The HSF training community. <https://hepsoftwarefoundation.org/training/community.html>. Accessed 6 Oct 2021
- IRIS-HEP and HSF training blueprint meeting (2020). <https://indico.cern.ch/event/889665/>. Accessed 6 Oct 2021
- The HSF training hackathon (2021). <https://indico.cern.ch/event/997485/>. Accessed 6 Oct 2021
- Lange D (2019) Selected results from HSF training survey. In: 2019 Joint HSF/OSG/WLCG Workshop HOW2019. [https://indico.cern.ch/event/759388/contributions/3315848/attachments/1816082/2968198/training\\_how2019.pdf](https://indico.cern.ch/event/759388/contributions/3315848/attachments/1816082/2968198/training_how2019.pdf). Accessed 6 Oct 2021
- Software Carpentry—Teaching Basic Lab Skills for research computing. <https://software-carpentry.org/lessons/>. Accessed 6 Oct 2021
- Reproducible Research Data Analysis Platform (REANA). <https://reana.io/>. Accessed 6 Oct 2021
- ROOT: analyzing petabytes of data, scientifically (2021). <https://root.cern/>. Accessed 6 Oct 2021
- Getting started with uproot (2021). <https://uproot.readthedocs.io/>. Accessed 6 Oct 2021
- Software carpentry workshop (Fermilab) (2019). <https://indico.fnal.gov/event/20233/>. Accessed 6 Oct 2021
- FIRST-HEP/ATLAS training (Argonne) (2019). <https://indico.cern.ch/event/827231/>. Accessed 6 Oct 2021
- FIRST-HEP/ATLAS training (LBNL) (2019). <https://indico.cern.ch/event/827232/>. Accessed 6 Oct 2021
- Software carpentry workshop (CERN) (2019). <https://indico.cern.ch/event/834411/>. Accessed 6 Oct 2021
- Computational and Data Science Training for High Energy Physics. <http://codas-hep.org/>. Accessed 6 Oct 2021
- Virtual pipelines training with GitLab (2020). <https://indico.cern.ch/event/904759/>. Accessed 6 Oct 2021
- Virtual pipelines training with GitHub (2020). <https://indico.cern.ch/event/1001128/>. Accessed 6 Oct 2021
- Virtual Docker training (2020). <https://indico.cern.ch/event/934651/>. Accessed 6 Oct 2021
- Machine learning on GPU training (2020). <https://indico.cern.ch/event/958112/>. Accessed 6 Oct 2021
- 1st HEP C++ course and hands-on training (2020). <https://indico.cern.ch/event/946584/>. Accessed 6 Oct 2021
- Software Institute for Data-Intensive Sciences. <https://sidis.web.cern.ch/>. Accessed 6 Oct 2021
- HSF virtual Docker training (2020). <https://indico.cern.ch/event/934651/>. Accessed 6 Oct 2021
- Lieret K (2021) Community building. In: HSF WLCG Virtual Workshop. <https://indico.cern.ch/event/941278/contributions/4084356/>. Accessed 6 Oct 2021
- Katz DS, McHenry K, Reinking C, Haines R (2019) Research software development & management in universities: case studies from Manchester's RSDS Group, Illinois' NCSA, and Notre Dame's CRC. <https://doi.org/10.1109/SE4Science.2019.00009>
- Building a career path for research software engineers (2021). <https://iris-hep.org/2021/05/12/career-path-rse.html>. Accessed 6 Oct 2021
- Katz DS, Malik S, Neubauer MS, Stewart GA, Assamagan KA, Becker EA, Chue Hong NP, Cosden IA, Meehan S, Moysé EJW, Price-Whelan AM, Sexton-Kennedy E, Evans MO, Feickert M, Lange C, Lieret K, Quick R, Sánchez Pineda A, Tunnell C (2020) Software sustainability & high energy physics. <https://doi.org/10.5281/zenodo.4095837>.
- Malik S, Thais S, Villanueva M, Lieret K, Stark G, Nibigira EN, Evans MO, David C (2021) Software training and sustainable HEP. In: Sustainable HEP workshop. <https://indico.cern.ch/event/1004432/contributions/4377762/>. Accessed 6 Oct 2021
- Data analysis for STEM teachers (2020). <https://indico.cern.ch/event/927162/>. Accessed 6 Oct 2021
- Google Colaboratory. <https://colab.research.google.com/>. Accessed 6 Oct 2021
- Machine learning basics for STEM teachers (2021). <https://indico.cern.ch/event/998732/>. Accessed 6 Oct 2021
- Quarknet. <https://quarknet.org/>. Accessed 6 Oct 2021

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Sudhir Malik<sup>1</sup>  · Samuel Meehan<sup>2</sup> · Kilian Lieret<sup>3</sup> · Meirin Oan Evans<sup>4</sup> · Michel H. Villanueva<sup>5</sup> · Daniel S. Katz<sup>6</sup> · Graeme A. Stewart<sup>2</sup> · Peter Elmer<sup>7</sup> · Sizar Aziz<sup>8</sup> · Matthew Bellis<sup>18</sup> · Riccardo Maria Bianchi<sup>25</sup> · Gianluca Bianco<sup>30,31</sup> · Johan Sebastian Bonilla<sup>23</sup> · Angela Burger<sup>24</sup> · Jackson Burzynski<sup>27</sup> · David Chamont<sup>8</sup> · Matthew Feickert<sup>6</sup> · Philipp Gadow<sup>12</sup> · Bernhard Manfred Gruber<sup>2,34,35</sup> · Daniel Guest<sup>15</sup> · Stephan Hageboeck<sup>2</sup> · Lukas Heinrich<sup>2</sup> · Maximilian M. Horzela<sup>16</sup> · Marc Huwiler<sup>26</sup> · Clemens Lange<sup>2</sup> · Konstantin Lehmann<sup>17</sup> · Ke Li<sup>19</sup> · Devdatta Majumder<sup>28</sup> · Judita Mamučić<sup>10</sup> · Kevin Nelson<sup>22</sup> · Robin Newhouse<sup>13</sup> · Emery Nibigira<sup>14</sup> · Scarlet Norberg<sup>1</sup> · Arturo Sánchez Pineda<sup>11</sup> · Mason Proffitt<sup>19</sup> · Brendan Regnery<sup>23</sup> · Amber Roepe<sup>33</sup> · Stefan Roiser<sup>2</sup> · Henry Schreiner<sup>7</sup> · Oksana Shadura<sup>21</sup> · Giordon Stark<sup>9</sup> · Stephen Nicholas Swatman<sup>2,20</sup> · Savannah Thais<sup>7</sup> · Andrea Valassi<sup>2</sup> · Stefan Wunsch<sup>2,16</sup> · David Yakobovitch<sup>32</sup> · Siqi Yuan<sup>29</sup>

<sup>1</sup> University of Puerto Rico, Mayaguez, PR, USA

<sup>2</sup> CERN, Geneva, Switzerland

<sup>3</sup> Ludwig Maximilian University of Munich, Munich, Germany

<sup>4</sup> University of Sussex, Brighton, UK

<sup>5</sup> University of Mississippi, Oxford, MS, USA

<sup>6</sup> University of Illinois at Urbana-Champaign, Champaign, IL, USA

<sup>7</sup> Princeton University, Princeton, NJ, USA

<sup>8</sup> IJCLab, Université Paris-Saclay, CNRS/IN2P3, Orsay, France

<sup>9</sup> Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA, USA

<sup>10</sup> Instituto de Física Corpuscular/Consejo Superior de Investigaciones Científicas, University of Valencia, Valencia, Spain

<sup>11</sup> LAPP, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy, France

<sup>12</sup> Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany

<sup>13</sup> University of British Columbia, Vancouver, Canada

<sup>14</sup> Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

<sup>15</sup> Humboldt University of Berlin, Berlin, Germany

<sup>16</sup> Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

<sup>17</sup> Simon Fraser University, Burnaby, BC, Canada

<sup>18</sup> Siena College, Loudonville, NY, USA

<sup>19</sup> University of Washington, Seattle, WA, USA

<sup>20</sup> University of Amsterdam, Amsterdam, The Netherlands

<sup>21</sup> University of Nebraska, Lincoln, NE, USA

<sup>22</sup> University of Michigan, Ann Arbor, MI, USA

<sup>23</sup> University of California Davis, Davis, CA, USA

<sup>24</sup> Oklahoma State University, Stillwater, OK 74078, USA

<sup>25</sup> University of Pittsburgh, Pittsburgh, PA, USA

<sup>26</sup> Universität Zürich, Zurich, Switzerland

<sup>27</sup> University of Massachusetts, Amherst, MA, USA

<sup>28</sup> Institut Ruđer Bošković, Zagreb, Croatia

<sup>29</sup> Boston University, Boston, USA

<sup>30</sup> University of Bologna, Bologna, Italy

<sup>31</sup> Istituto Nazionale di Fisica Nucleare (INFN), Rome, Italy

<sup>32</sup> SingleStore, New York, USA

<sup>33</sup> University of Oklahoma, Norman, OK, USA

<sup>34</sup> Center for Advanced Systems Understanding, Saxony, Germany

<sup>35</sup> Technische Universität Dresden, Dresden, Germany