Train to Sustain

Sudhir Malik[1,*], *Kilian* Lieret[2], *Peter* Elmer[2], *Michel* Hernandez Villanueva[3], and *Stefan* Roiser[4]

[1]University of Puerto Rico Mayagüez, Mayagüez, PR 00681, U.S.A.
[2]Princeton University, Princeton, NJ 08544, U.S.A.
[3]Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany
[4]Conseil Européen pour la Recherche Nucléaire (CERN), Geneva, Switzerland

Abstract. The HSF/IRIS-HEP Software Training group provides software training skills to new researchers in High Energy Physics (HEP) and related communities. These skills are essential to produce high-quality and sustainable software needed to do the research. Given the thousands of users in the community, sustainability, though challenging, is the centerpiece of its approach. The training modules are open source and collaborative. Different tools and platforms, like GitHub, enable technical continuity, collaboration and nurture the sense to develop software that is reproducible and reusable. This contribution describes these efforts and its broader impacts.

1 Introduction

The current landscape of typical experimental collaborations in HEP involves hundreds to thousand of users spread across different continents, time zones and cultures. These diverse collaborations are coordinated global scientific research endeavours, like the ones at CERN or Fermilab. They exhibit coalitions of diverse physics interests held together by a common interest in a detector, like CMS [1] or ATLAS [2], and accelerator, like the LHC [3], that can potentially address scientific goals. CMS and ATLAS collaborations, together, and DUNE collaboration [4] have about 10K and over 1K users, respectively. These scientific endeavours are accompanied by daunting challenges. The management of the data volume at the scale of hundreds of petabytes requires distributed computing resources from data simulation and processing to physics analysis. Instrumentation, assembly and operations of HEP detectors requires years of expertise and involvement. Analysing and processing the data for detector performance and physics searches requires computing and software skills that take years to learn. In addition, there are emerging technologies, novel techniques and disruptive changes (COVID-19, computing architecture, ideas) that require adaptive response. While every HEP experiment has its own specific design and needs, a lot of know-how and preparation is common. The 2021 US Community Study on the Future of Particle Physics (also called Snowmass 2021 [5]) has strongly emphasised the need for training in the areas of software, instrumentation and accelerator in order to prepare a diverse workforce. This is essential to mitigate the above challenges, meet the future needs and guarantee its success

*e-mail: sudhir.malik@upr.edu

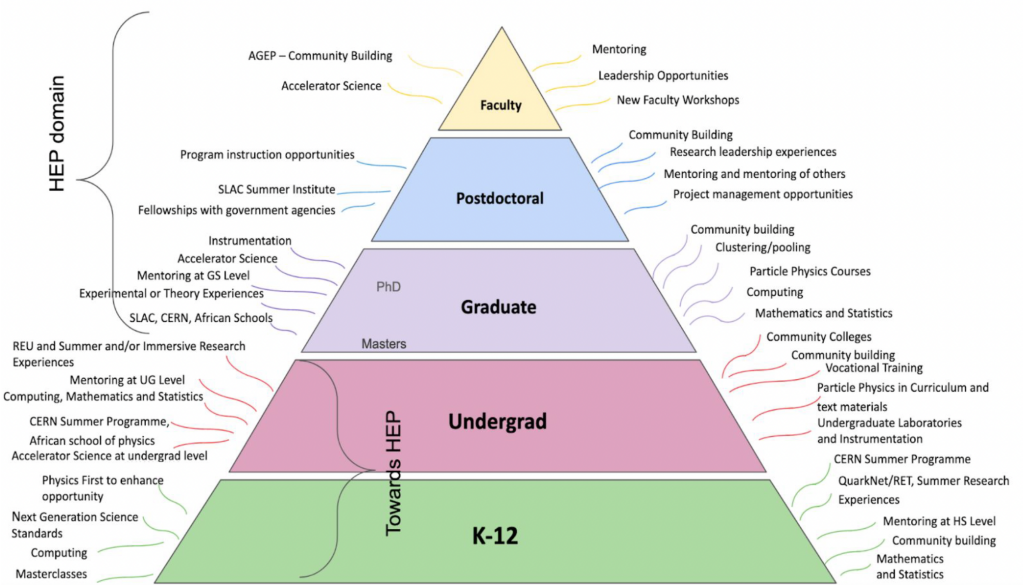

Figure 1. Progression of HEP Education and Training.

and also lead to careers in HEP or other STEM areas. A robust software underpins all components of HEP facilities - physics analysis, data processing, data collection from detectors and running an accelerator. Figure 1 shows the training pyramid epitomizing the current training efforts that start with the K-12 level onward. This paper describes the software training efforts led jointly by the IRIS-HEP [6] and the HSF [7].

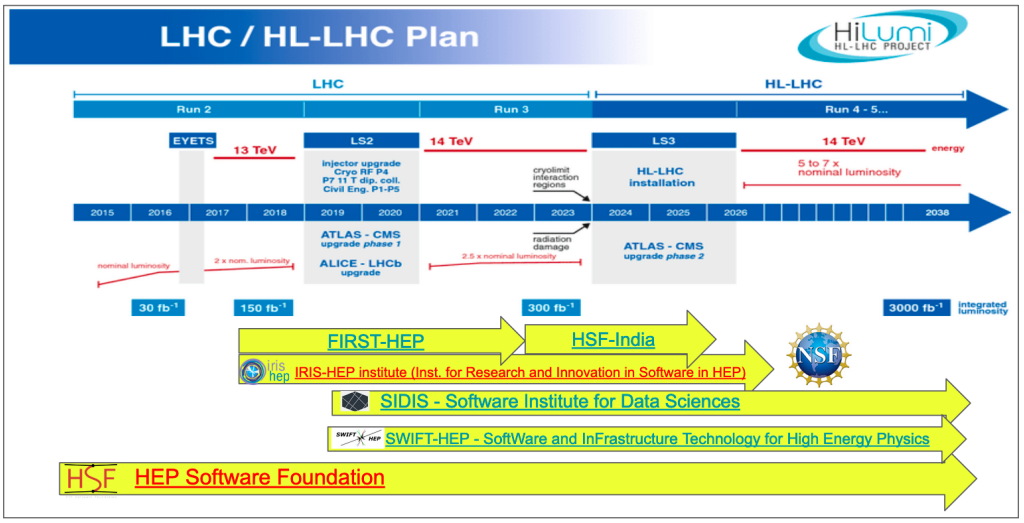

Figure 2. LHC/HL-LHC timeline and corresponding training efforts.

2 Training Paradigm

The challenges mentioned above necessitate an investment in organised hands-on software training that is accessible to everyone, builds a future workforce and leads to corresponding careers in HEP or other STEM areas. The current paradigm is that students trained as physicists need also be proficient data analysts. However, there are no standard software curricula for HEP students and not everyone can attend university-offered software courses. In many cases HEP students don't receive any programming training and it mostly happens on the job as the need arises. The HSF/IRIS-HEP is leading training efforts with several other players with a goal to provide a unified, scalable, and sustainable software training framework that is powered by the HEP community. Figure 2 shows the LHC and the HL-LHC driven timeline of the HSF/IRIS-HEP training efforts along with its synergistic players like the SIDIS [8], SWIFT-HEP [9] and HSF-India [10]. In addition, IRIS-HEP is collaborating with QuarkNet [11] to access its network of teachers to give them opportunities to improve coding skills via "Coding Camps" [12] and explore how these could be embedded in instructional materials implemented in classrooms; and to help foster the coding skills of their students. This would facilitate the creation of a robust pipeline of students pursuing STEM careers equipped with software awareness. The broad spectrum of the above software training program covers mentors, scientists, graduate and undergraduate students and high school teachers.

3 Software Training

The HSF/IRIS-HEP Software training group has developed a program [13–15] that serves as a training hub for new researchers in HEP and related nuclear, neutrino, astro and theory communities including scientists, postdocs, graduate and undergraduate students. The skills offered range from basic core software needed by everyone to the advanced training required by specialists in software and computing. These skills are essential to produce high-quality and sustainable software needed to do the research and solve future challenge as emphasised above. The training content is open source with sustainability is the centerpiece of its approach. It is available via a combination of GitHub [16], Websites, Indico [17] and video recordings. The open content enables technical continuity, collaboration and nurture the sense to develop software that is reproducible and reusable. The training group efforts were very impactful in the Snowmass 2021 process where it was emphasised making training integral of HEP future to prepare an inclusiveness workforce in HEP.

4 Scalability and Sustainability

The long term scalability and sustainability of the research software ecosystem is important for HEP as HL-LHC and other facilities of the 2020s will be relevant through at least the 2030s. Meeting this challenge requires a workforce with a combination of HEP domain knowledge and advanced software skills. The software training model meet the needs of the present community as well as allow future users to shape activities to meet their needs. The foundational aspect of this is to be open and easily accessible. This basic tenet is reflected in the organisation itself where all activities are open to everyone in the community. There are open weekly meetings to discuss the planning of events, updates and suggestions. There are monthly hackathons to improve the training material and create new content. The details of these, beyond the meetings, can be followed up in a Slack channel and it is also available from the HSF Working Group homepage [13]. The mentors are recognised with their profile displayed on the website. Presentations by the mentors at conferences are strongly

encouraged. In addition, publications in journals and conference proceedings recognise our efforts more formally [18]. There are also guidelines [19] for any one who wishes to organise training in their local community and needs support for it. The training modules are built on each other with enough verbosity for self study or self organised training workshops.

5 Broader Impacts

IRIS-HEP outreach activities are focused on high school teachers with a goal to tap, grow and diversify the talent pipeline from K-12 students for future cyberinfrastructure. IRIS-HEP co-organised several outreach workshops with FIRST-HEP. A list of outreach activities is given in Reference [20]. A recent onset of partnership between QuarkNet and the IRIS-HEP has led to the organisation of workshops called "Coding Camp" that enhances and extends the coding skills of high school physics teachers. The partnership has enabled a quick reach to a wider network of STEM teachers associated with QuarkNet for years. This has potential to create a robust pipeline of students pursuing STEM careers equipped with software awareness. Besides this, IRIS-HEP and FIRST-HEP organises an annual CoDaS-HEP [21] school that provides graduate students from HEP, Computer Science and related fields, a broad introduction to critical skills like parallel programming, machine learning, data science tools as well as an overview of applications to High Energy Physics. A new 10-week internship proposal called "U.S. CMS SPRINT – A U.S. CMS Scholars Program for Research INTernship" [22], funded under the DOE HEP-RENEW program to address the barriers faced by the underrepresented populations, has its software training component derived from the HSF/IRIS-HEP training curriculum.

6 Conclusions

The HSF/IRIS-HEP Software Training program is mature and trying to address the challenges presented by the current paradigm in HEP where the landscape lacks a common organised effort. It is providing training at all levels - mentors, scientists, graduate and undergraduate students and K-12 schools teachers (and their students indirectly). It is enabling preparation of a workforce that is not only equipped to solve challenges presented by the software and computing needs of the HL-LHC but other STEM areas as well, including career building. The organisation sustainability and scalability is the centerpiece of its approach. The training content is open source. It impacted the Snowmass 2021 planning process that recognised the role of software training as an integral part of HEP in preparing an inclusiveness workforce. The broader impacts of its work have been far reaching. It is training K-12 teachers in coding skills that can improve classroom curriculum and reaching out to teachers community across the network established by QuarkNet.

We thank NSF grants OAC-1836650, OAC-1829707 and OAC-1829729 for support of the training program.

References

[1] *CMS Experiment*, https://home.cern/science/experiments/cms
[2] *ATLAS Experiment*, https://atlas.cern/
[3] *Large Hadron Collider*, https://www.home.cern/science/accelerators/large-hadron-collider
[4] *DUNE Experiment*, https://www.dunescience.org/
[5] *Snowmass 2021*, https://snowmass21.org/

[6] *IRIS-HEP*, https://iris-hep.org/
[7] *HEP Software Foundation*, https://hepsoftwarefoundation.org/
[8] *SOFTWARE INSTITUTE FOR DATA-INTENSIVE SCIENCES* , https://sidis.web.cern.ch/
[9] *SoftWare and InFrastructure Technology for High Energy Physics*, https://swift.hep.ac.uk/
[10] *HSF-India*, https://research-software-collaborations.org/about/team
[11] *QuarkNet*, https://quarknet.org/
[12] *Coding Camp 3-day*, https://indico.cern.ch/event/1305233/
[13] *HSF training*, https://hepsoftwarefoundation.org/workinggroups/training.html
[14] S. Malik et al, *Software Training in HEP*, https://link.springer.com/article/10.1007/s41781-021-00069-9
[15] Michel H. Villanueva, Sudhir Malik and Meirin Oan Evans, *Software Training in High Energy Physics*, https://iopscience.iop.org/article/10.1088/1742-6596/2438/1/012063
[16] *HEP Software Foundation Training Material*, https://github.com/hsf-training
[17] *Training Events*, https://indico.cern.ch/category/11386/
[18] *HSF Training Talks & Papers*, https://hepsoftwarefoundation.org/training/resources
[19] *Guidelines for training material*, https://hepsoftwarefoundation.org/training/module-guidelines.html
[20] *IRIS-HEP Outreach*, https://indico.cern.ch/category/17100/
[21] *Computational and Data Science Training for High Energy Physics*, https://codas-hep.org/
[22] *RENEW-HEP: U.S. CMS SPRINT - A U.S. CMS Scholars Program for Research INTernship (See page 24)*, https://science.osti.gov/-/media/Initiatives/pdf/renew/RENEW_Public_Abstracts-FY22.pdf