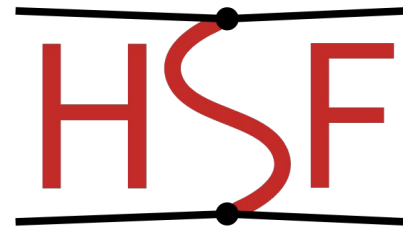




HSF: Retrospective and Future

Graeme Stewart, for the HSF Core Coordination Team

HEP Software Foundation - The Challenges*



High Energy Physics Software Foundation

- Changing hardware landscape making efficient software more difficult to implement
 - Multi-core processing was with us; many experiments had not yet fully adapted
 - GPUs were around, primitive and difficult to use in those days, but clear potential
- An exciting physics programme ahead, e.g., High-Luminosity LHC, but one that we knew would put tremendous pressure on software and computing
 - Event complexity and rate would jump
- The need to train a new generation of software experts
 - And give them recognition and a career path (RSE discussion had begun)
- No slack in the system for duplicated solutions and multiple rediscoveries of the same things, given the severe lack of experts

How it all began...

HEP Software Collaboration meeting

3-4 Apr 2014
CERN
Europe/Zurich timezone

Enter your search term

Overview

Timetable

Contribution List

Author List

Registration

Participant List

Accommodation

Visa

Videoconference

Current HEP software is the result of 20 years of development and now must evolve to meet the challenges posed by new experimental programmes, including future upgrades of the LHC. In addition, the computing landscape is evolving rapidly and we need to exploit all the expertise available in our community, and in other scientific disciplines, in order to meet the technical challenges we are facing. During the last 2 years, a Concurrency Forum has been meeting to promote discussions within the HEP community regarding various independent R&D efforts. Formalising this activity into a Collaboration would be a natural next step to eventually transform the R&D efforts into joint projects to build software libraries of general interest. A formal Collaboration would also provide a framework for attracting effort and external funding, as well as providing a structure for the community to set priorities and goals for the work. This should be a HEP community-wide effort, but should be open enough that eventually it could form the basis for collaboration with other sciences. The short note attached describes some of the thinking behind this initiative.

The current agenda is still very preliminary.

VIDEO CONFERENCE: see under "Video Services"

Vidyo public room ongoing until Fri 4/Apr at 18:00 : HEP_Software_Collaboration_meeting
Extension 9297652

 **Starts** 3 Apr 2014, 08:00
Ends 4 Apr 2014, 18:00
Europe/Zurich

 [Ian Bird](#)
[John Harvey](#)

 CERN
503/1-001 - Council Chamber
[Go to map](#)

 **Background information**
[Open_Scientific_Software_Collaboration.pdf](#)

- HEP Software Foundation hadn't even been christened
- However, this was the first meeting where we discussed the ideas that would lead to the HSF
- Happy Birthday!



Well, did it really begin earlier?

- The [Concurrency Forum](#) was an antecedent activity
- Brought people from many different experiments together to look at technology for a changing hardware landscape: multi-core, wide vector registers, non-x86 CPUs, even GPUs
- This work identified Intel's Threaded Building Blocks as a good choice

TBB investigation for SuperB

Implement a small system with the Intel Threading Building Blocks (TBB), which is a library offering a rich approach to expressing parallelism in a C++ program

Shibboleth login

[CERN Login](#)

Main menu

- [Home](#)
- [Demonstrators](#)
- [Projects](#)
- [Benchmarks](#)
- [Meetings](#)
- [Events](#)
- [Resources](#)
- [TechLab](#)

Software engineering is moving towards a paradigm shift in order to accommodate new CPU architectures with many cores, in which concurrency will play a more fundamental role in programming languages and libraries. HEP software will also need to accommodate the new hardware architectures by introducing parallelism whenever possible in order to make efficient use of all the available cores. This implies the development of new models and specialized software frameworks to assist scientists in developing their software algorithms and applications in order to benefit from all the available cores.

This web is an attempt to share knowledge among interested parties that should work together to develop 'demonstrators' and agree minimally on technology so that they can share code and compare results.

- And all of this builds on a long tradition of collaborative software
- CERNLIB, ROOT, Geant4, Generators, etc.

What do we really want to do?

This is a followup of the HEP software collaboration [initial workshop](#) that took place at CERN in April. The main goal of the workshop is to refine the next steps for building the [HEP Software Foundation](#) (HSF) that is being organized to facilitate coordination and common efforts in high energy physics (HEP) software and computing internationally.

- SLAC Workshop in January 2015
 - Reframe the goals as “bottom up”
 - From this the famous maxim of do-ocracy was spawned (Torre)
 - Much more appropriate for community software activities
- Many community input papers
 - Key lessons learned from past experience
 - Lots of discussion on software projects
 - How could they be HSF projects and what would that mean?
 - What could the HSF offer projects (like hosting, CI, etc.)
 - Forming of nascent activity areas where we could identify common problems
 - E.g., the issue of packaging started a useful [review](#) of options both in and out of the field
 - HSF developed [licensing guidelines](#) that helped the LHC experiments open source their software (recommendation shifting from GPL to Apache 2)

HSF Technical Notes (Zenodo)



Community White Paper

- HL-LHC was identified as a step-change for software
 - Just as the experiments were planning their detector upgrades we needed a “software upgrade” to go alongside
- In July 2016 HSF was charged by the WLCG to write a community white paper to understand how...
 - to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
 - to enable new approaches to computing and software that could radically extend the physics reach of the detectors
 - to ensure the long term sustainability of the software through the lifetime of the HL- LHC
- Process of engagement of many stakeholders from experiments, software projects and labs to achieve this

CWP Process



- [Kick-off workshop](#) 23-26 January 2017, San Diego
- Groups held workshops and meetings in the subsequent months
 - Broadening the range of participation, often with non-HEP experts participated
- [Workshop in Annecy](#) 26-30 June started to draw the process to a close
 - 13 Working Groups had made good progress on their chapters
- Both workshops involved ~100 people, mainly US and EU
 - Total number of people involved in the writing process was about 250
 - Many others commenting
- Finally CWP was published at the end of 2017 [[1712.06982](#)] as well as later in [Computing and Software for Big Science](#)
 - *Aside:* CSBS was a specific outcome from HSF discussions and the need for a recognised place for academic publication of key software papers

CWP Outcome

- 310 supporting authors, from 124 institutions
- Covered the broad range of areas that were important for the field
- Clearly had a very significant impact inside and outside of HEP
 - Has now been cited 189 times
 - The whole CWP process, including the final paper, **significantly raised the profile of software**
 - Had a direct positive impact on the funding of IRIS-HEP (NSF) and SWIFT-HEP (STFC)

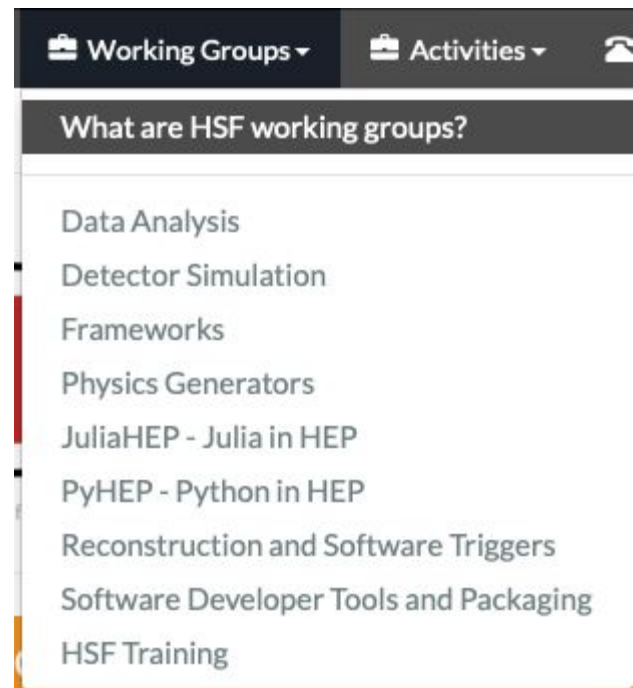


Contents

1	Introduction	2
2	Software and Computing Challenges	5
3	Programme of Work	11
3.1	Physics Generators	11
3.2	Detector Simulation	15
3.3	Software Trigger and Event Reconstruction	23
3.4	Data Analysis and Interpretation	27
3.5	Machine Learning	31
3.6	Data Organisation, Management and Access	36
3.7	Facilities and Distributed Computing	41
3.8	Data-Flow Processing Framework	44
3.9	Conditions Data	47
3.10	Visualisation	50
3.11	Software Development, Deployment, Validation and Verification	53
3.12	Data and Software Preservation	57
3.13	Security	60
4	Training and Careers	65
4.1	Training Challenges	65
4.2	Possible Directions for Training	66
4.3	Career Support and Recognition	68
5	Conclusions	68
	Appendix A List of Workshops	71
	Appendix B Glossary	73
	References	79

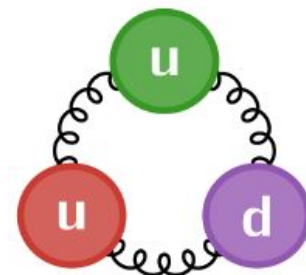
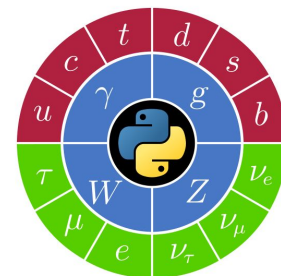
Catalysing Community Activities

- Post-CWP we had a whole number of areas where we knew the community would be active
- Experiments were ramping up their investment into software improvements and software R&D
- We decided that the HSF could help by broadening the range of working groups, that would help with information flow between developers, different projects and experiments
 - To that end, we launched a whole new series of working groups, corresponding to significant areas
 - Open nomination process and an HSF search committee to make appointments
 - These were intended to occupy a development space before finished items of work that might be presented in the usual conferences or as publications



Language Working Groups

- HSF Working Groups were particularly successful where they filled in a gap in current activities
- [PyHEP](#) catalysed activities around the growing interest in data science tools in HEP in 2018 (pre-CHEP workshop)
 - Massively popular PyHEP online workshops (hitting ~500 attendees)
 - [Next edition 1-4 July](#)
 - Now also organising PyHEP.dev as more developer focused events
 - [Next event 26-30 August](#), Aachen
 - The group also helped shape strategy, via LHCC inputs
- [JuliaHEP](#) is much smaller, but it's a new activity that is helping focus efforts for HEP in Julia
 - One successful JuliaHEP event, [next edition 30 September - 3 October](#), CERN



From Generators to Analysis

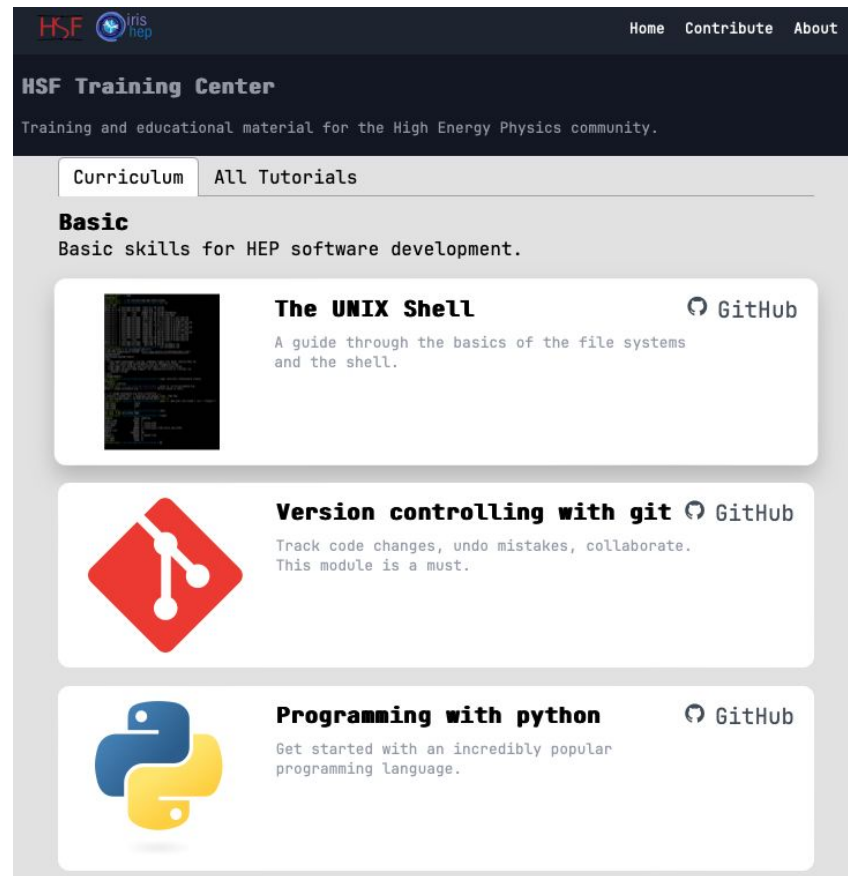
- HSF Working Group on [event generators](#) started from [Physics Event Generator Computing Workshop](#) in 2018
 - This was a first for HEP and brought to the fore a growing concern about the computing needs for event generators
 - Followed by a number of very active years in the working group
 - Also helping to raise the issue at the LHCC and strategic level
- [Data Analysis Working Group](#) (DAWG) catalysed discussions around analysis topics
 - Also birthing several papers on this topic, specifically on *Metadata* and *Training and Onboarding*
- Very much related to the community workshops on the HEP Analysis Ecosystem
 - The [Amsterdam Edition](#) 2017, [Adelaide Discussions](#) 2019, The [Orsay Variation](#) 2022
 - Ergonomics, workflows, facilities all discussed
- Analysis Facilities are still active topic for which we have the AFF Report [[2404.02100](#)] and a [session this week](#)



HSF Training

- HSF Training working group has now trained more than 2000 students in its various training courses
 - Huge thanks to all of those involved - from HSF, IRIS-HEP, ROOT, CERN, Fermilab, ...
- Lots of material developed and curated
 - Broadening involvement of younger colleagues in delivering and tutoring the material

Training session at this workshop!



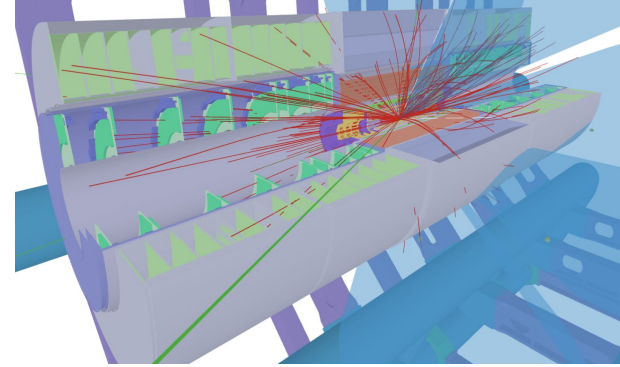
The screenshot shows the HSF Training Center website. At the top, there are logos for HSF and IRIS HEP, and navigation links for Home, Contribute, and About. The main heading is "HSF Training Center" with the subtitle "Training and educational material for the High Energy Physics community." Below this, there are two tabs: "Curriculum" (selected) and "All Tutorials". The content is organized under a "Basic" section, described as "Basic skills for HEP software development." There are three tutorial cards:

- The UNIX Shell** (with a GitHub icon): "A guide through the basics of the file systems and the shell." The icon shows a terminal window with code.
- Version controlling with git** (with a GitHub icon): "Track code changes, undo mistakes, collaborate. This module is a must." The icon is the Git logo.
- Programming with python** (with a GitHub icon): "Get started with an incredibly popular programming language." The icon is the Python logo.

New HSF Training Centre Page!

General Forums and Other Communities

- Compute and Accelerator Forum has been a very successful meeting series
 - Still going strong after four years!
 - Initial focus on compute accelerators
 - Project presentations
 - Facility talks
 - Vendor roadmaps
 - Now generalising to other interesting topics, e.g., D language, garbage collectors
- Software and Computing Roundtable was very active and may become so again
 - Big input from the nuclear physics community, JLab and BNL
 - Was a very successful part of our engagement with the nuclear physics community
 - Engagement with ePIC and EIC colleagues is strong in a number of areas (WG conveners)
 - We also built bridges with Belle II and DUNE (WG conveners)



Software Projects

- The HSF didn't spawn software projects in the way that we expected
 - That said, a number of projects were developed with close links to the HSF, e.g., Phoenix Event Display, prmon monitor, PODIO
- We did organise useful community “reviews” around a number of topics
 - GeantV and Geant on GPU (AdePT, Celeritas)
 - DUNE framework requirements
- And one incredibly useful exercise was done in the area of conditions data
 - Group of experts from multiple experiments looked the problem
 - Wrote a requirements document from that
 - BNL NPPS group then provided an implementation that is in production for sPHENIX and being evaluated by others
 - This has become a de facto **reference implementation**
- HSF, together with CERN has also organised GSoC for many years
 - Hundreds of students have participated in our projects, useful work and a talent pipeline!

HSF - Where now?

- We started to think in the Core Coordination Team how to make the HSF even more effective in the coming years
- This is the discussion we have tomorrow, but highlight ideas are...
 - Restructure the Core Coordination Team to be more active and have a better spread of responsibilities
 - Setup a new Advisory Group to help gather outside input to help guide activities
 - Reorient the relationship with projects to recognise the great work being done in the field, while projects retain full control

Summary

- “I can’t believe how fast they grow up...”
- The HSF has been active for 10 years now
 - Thanks to the voluntary efforts of many, many people 🙌
- We did a great deal at the strategic level, which paid off
 - CWP
 - LHCC inputs (WLCG software liaison representatives)
 - European Strategy and Snowmass Process
- We did many practical things as well
 - Strong role in communications and advice in the working groups
 - Workshops on important focused topics
 - Papers summarising experience of the community and problem framing
 - Reviews and advice to projects and experiments
 - Google Summer of Code
- This is definitely *work*, but it pays off and there are plenty of opportunities to contribute to our future activities



Backup

References

- [HSF Workshops](#) on Indico
- [Concurrency Forum Website](#)
- The HEP Software Foundation Community [[2205.08193](#)]

HSF Objectives in 2015

From [Pere Mato's presentation](#)

- Share expertise
- Raise awareness of existing software and solutions
- Catalyze new common projects
- Promote commonality and collaboration in new developments to make the most of limited resources
- Aid developers and users in creating, discovering, using and sustaining common software
- Support training career development for software and computing specialists
- Provide a framework for attracting effort and support to S&C common projects
- Provide a structure for the community to set priorities and goals for the work
- Facilitate wider connections; while the HSF is a HEP community effort, it should be open enough to form the basis for collaboration with other sciences