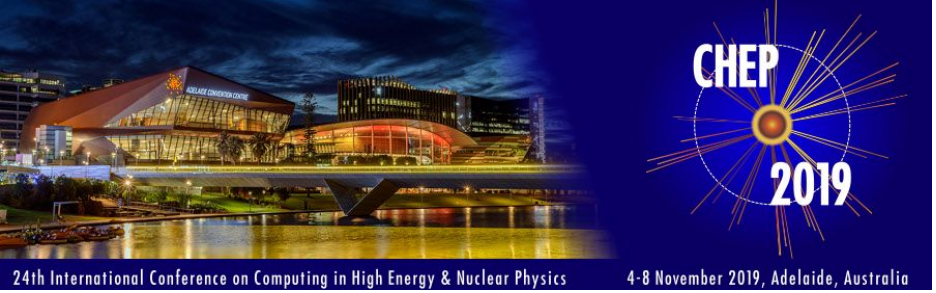


Enabling HEP software development for the 2020s - a coherent vision for training

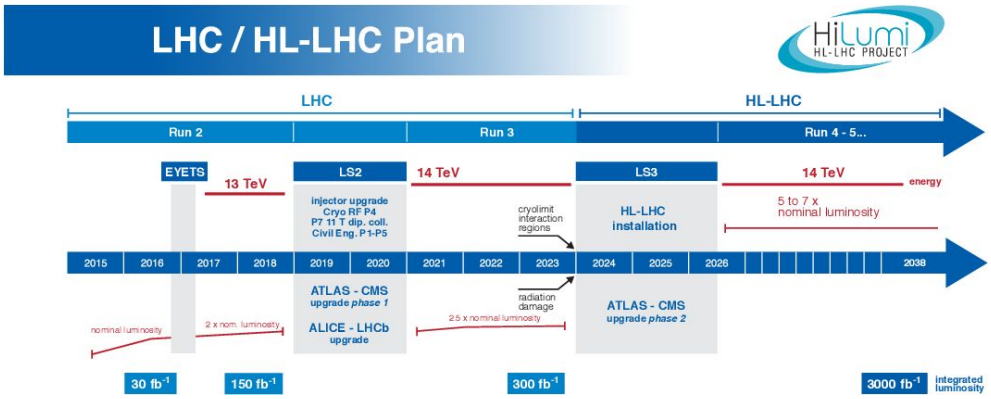
Peter Elmer, Jim Pivarski, Henry Schreiner - Princeton University
Sam Meehan, Graeme Stewart - CERN
Sudhir Malik - University of Puerto Rico - Mayaguez
+ The Carpentries (https://carpentries.org)
+ many others involved in specific workshops and schools!

Software training is key in HEP

Software skills are now integral to being a successful experimental HEP physicist, and a key transferable skillset for career evolution for people trained in HEP.

The software and computing systems are key subsystems of our experiments, and arguably the lifetime costs are larger than many others.

Maximizing science from the hardware investments increasingly relies critically on software.



We know we have medium/long term challenges both from planned facilities (HL-LHC, DUNE, etc.) and from technology evolution. But this also means we can take a long term view to how we develop software capabilities.

Building a Training Program

Although software training is now key in many research fields, individual universities do not uniformly provide such training today prior to a Ph.D. student beginning their research career.

In addition, there are many domain-specific aspects that add challenges in the universities. No “one size fits all” computational methods class is possible, e.g. partial differential equations are not so relevant for experimental HEP, whereas MC methods (random number generators, etc.) are. On top of this we have additional HEP- and experiment-specific software environments and challenges.

A possible solution in HEP is to exploit our large-scale community structure and organize these within our research domain: at the level of experiments or larger.

The main issues then become (1) **sustainability** and (2) **scalability**.

Training challenges

Not all funding agencies, institutions and funded projects have the same priority for training and education (e.g. DOE vs NSF in the US) relative to other goals like building/operating experiments, physics analysis, etc. Training activities are not always valued relative to other activities in making career steps.

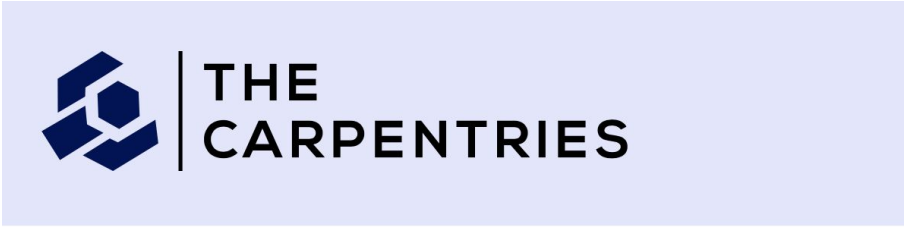
Despite this many individuals do get enthusiastic about training others, but often only in specific career phases and often as a side “hobby” project. How do the activities then scale up?

Technology evolution means that training materials need to evolve, too. Separating “local” specifics (e.g. computing environments or experiment-specific bits) from generally usable material is important, but doesn’t always happen.

Are training materials a common good or an individual product? Even if individuals *do* want to contribute to a common good, how do they do so?

Inspirations

https://carpentries.org



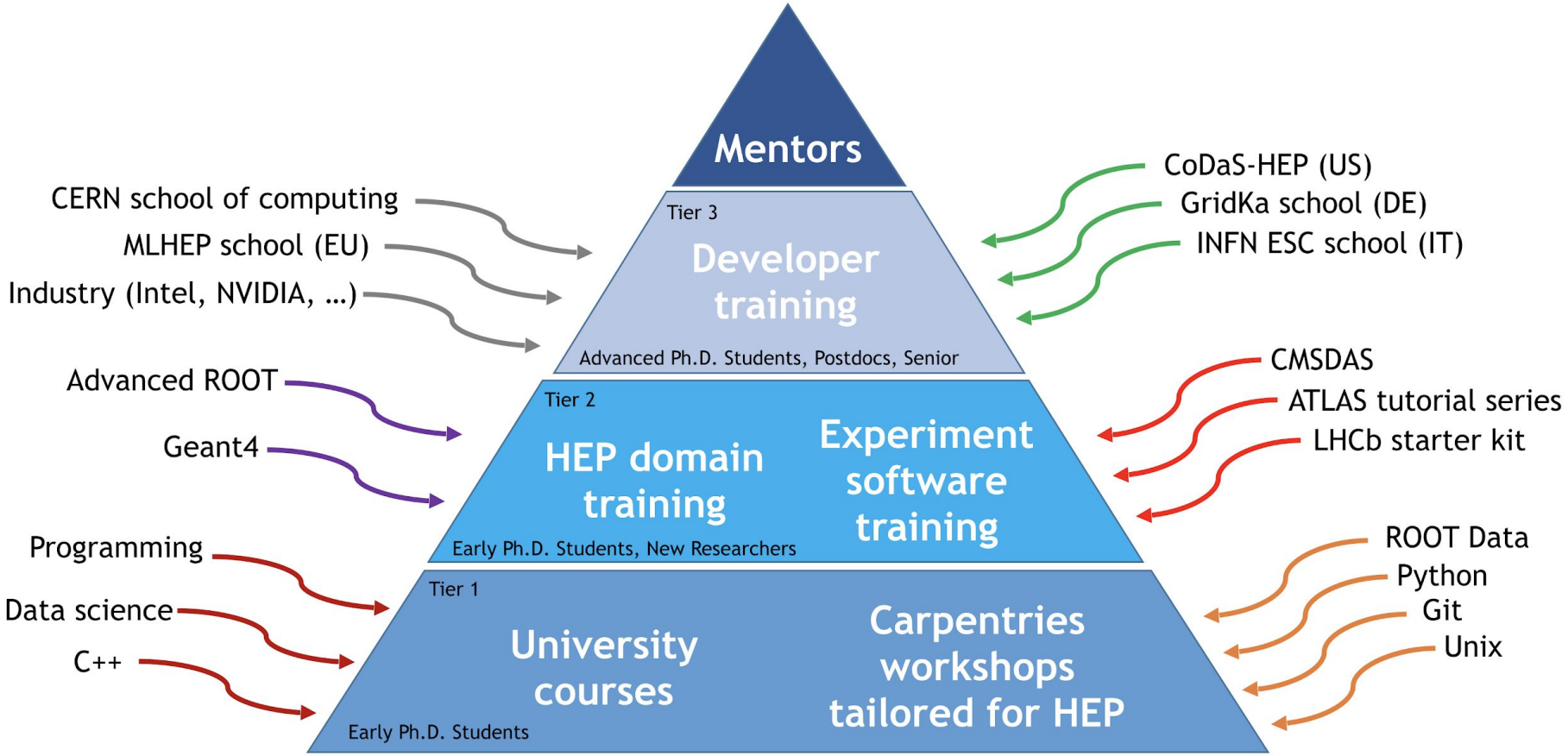
About The Carpentries Curricula

- Data Carpentry: Ecology
- Data Carpentry: Genomics
- Data Carpentry: Geospatial
- Data Carpentry: Social Sciences
- Library Carpentry
- Software Carpentry (All Workshops)
- Software Carpentry (Plotting and Programming in Python)
- Software Carpentry (Programming with Python)
- Software Carpentry (Programming with R)
- Software Carpentry (R for Reproducible Scientific Analysis)

https://lhcb.github.io/starterkit/

Key insight: thinking of training as a community building exercise. And not only for the “student” participants, but also for the “instructors”.

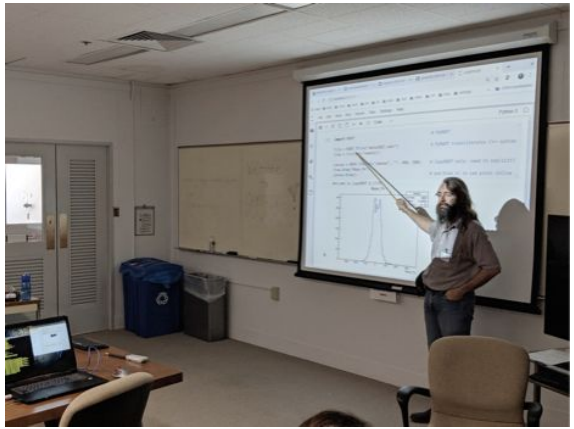
HEP Software Training



Software Carpentry Workshop (FNAL, 1-2 Apr, 2019)

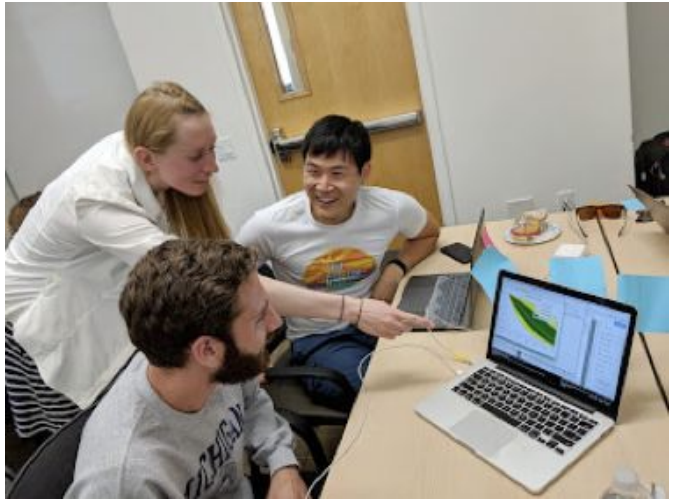
https://indico.fnal.gov/event/20233/

FIRST-HEP/US-ATLAS Training (ANL, 10 Jun, 2019)



https://indico.cern.ch/event/827231/

FIRST-HEP/US-ATLAS Training (LBNL, 19-21 Aug, 2019)



https://indico.cern.ch/event/816946/

Mentoring

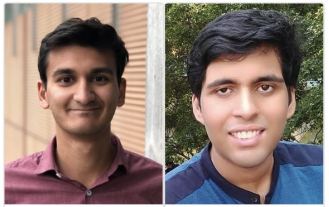
We have extensive networks for interacting with students, in particular at the undergrad level: CERN summer students, Openlab summer students, Google Summer of Code, etc.

Plus some national connections: DOE/INFN, Lithuanian ERASMUS students at CERN, US REU and IRES students, etc.

These have been the initial contact for a number of people now working in S&C in HEP, in particular (but not always) with a CS background.

Project focused: bring students into contact with “mentors” to work on a specific, pre-defined project.

Current IRIS-HEP Fellows



Raghav Kansal
University of California, San Diego

IRIS-HEP Fellow

Jun-Aug 2019

Pratyush (Reik) Das
Institute of Engineering & Management (Kolkata)

IRIS-HEP Fellow
Jun-Sep 2019

Key Insight: this is also an opportunity to provide incentivized and explicit paths forward for enthusiastic students from the more advanced training schools (ESC/Bertinoro, CoDaS-HEP, MLHEP, etc.)

CoDaS-HEP School

CoDaS-HEP 2017

CoDaS-HEP 2018

The CoDaS-HEP school aims to provide a broad introduction to these critical skills as well as an overview of applications High Energy Physics. Specific topics to be covered at the school include:

- Parallel Programming
- Big Data Tools and Techniques
- Machine Learning - Technology and Methods
- Practical skills: performance evaluation, use of git for version control

The program includes both lectures and practical hands-on exercises.

CoDaS-HEP 2019

http://codas-hep.org

Lessons Learned

As a community we have all of the ingredients to build a larger, coherent vision of how we train and engage people through multiple stages of their early career.

We need to continue to discuss and work together to build this coherent, federated vision of how the pieces fit together.

Training should not be an ad-hoc enterprise. We are a “research *and* education” enterprise, we need to focus on leveraging our international community structure to enhance the **sustainability and scalability** of our activities to maximize the impact. (It is useful, but not maximally impactful, to execute a training session in one or another “local” environment.)

[There are many more specific lessons about how to effectively execute training activities. No time in this talk for those, but definitely can be part of a collective effort!]

Next Steps

Later this fall or in early 2020, we would like to do a checkpoint on the curricula we are using (including also the LHCb/Alice efforts) and attempt to define some new training modules with The Carpentries that are appropriate for natural sciences and/or engineering students.

Then during 2020 we continue to run similar workshops using the new curriculum and look more closely how to scale them up and integrate them with planned experiment meetings and experiment-specific training activities.

Sometime during 2020, we would also like to hold a workshop or BoF session on the more advanced training schools (ESC, CSC, CoDaS-HEP, …) and a community-wide, federated, approach to facilitating student projects (“mentoring”)

(Then we take Berlin… and/or “Profit!”)

Acknowledgments

This talk is mostly about the overall training vision, but we would like to acknowledge that the specific training activities described as examples in this talk have been supported by:

- The US National Science Foundation through grants OAC-1829707 and OAC-1829729 (FIRST-HEP), Cooperative Agreement OAC-1836650 (IRIS-HEP)
- The US-ATLAS Operations Program Analysis Support Centers
- Fermi National Accelerator Laboratory, Argonne National Laboratory and Lawrence Berkeley National Laboratory (soon CERN!)
- The Carpentries
- Princeton University, the Princeton Institute for Computational Science and Engineering, the University of Puerto Rico at Mayaguez and the University of Chicago