Institute for Research and Innovation in Software
for High Energy Physics (IRIS-HEP)

Defining a "Training Challenge"

Peter Elmer (Princeton)

http://iris-hep.org
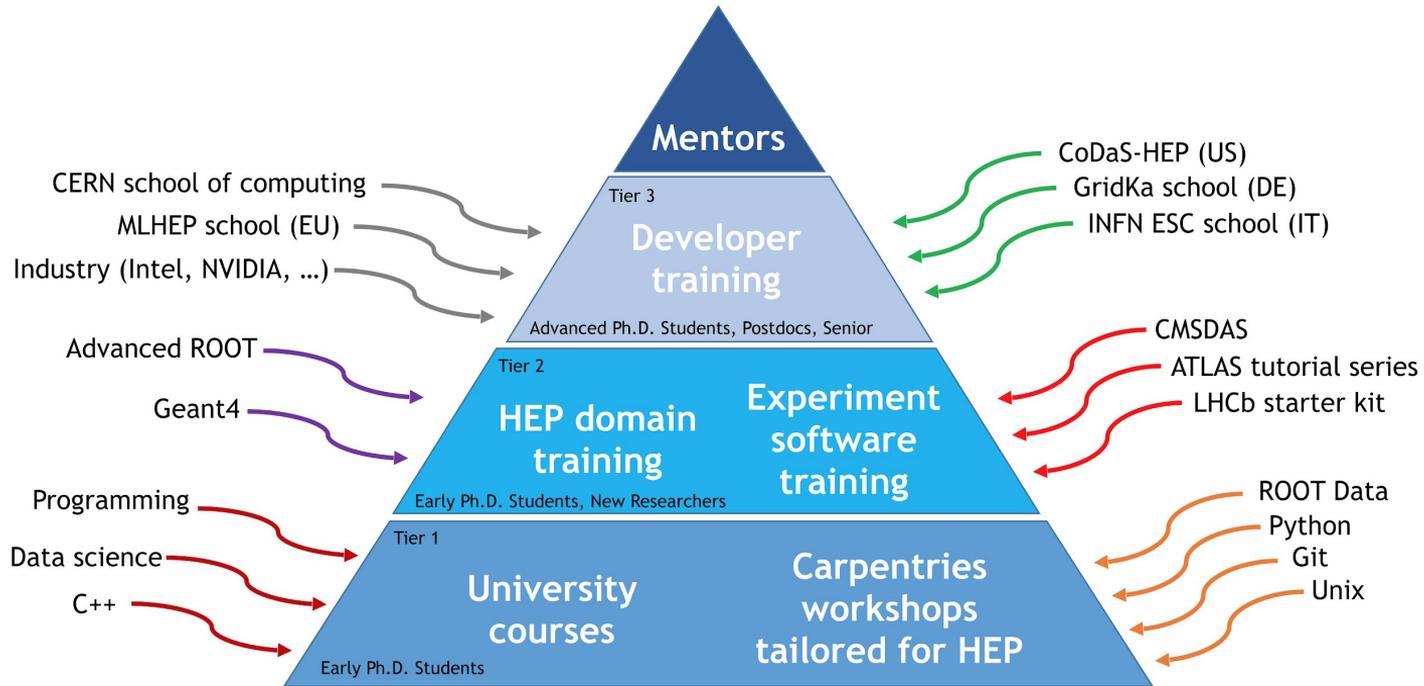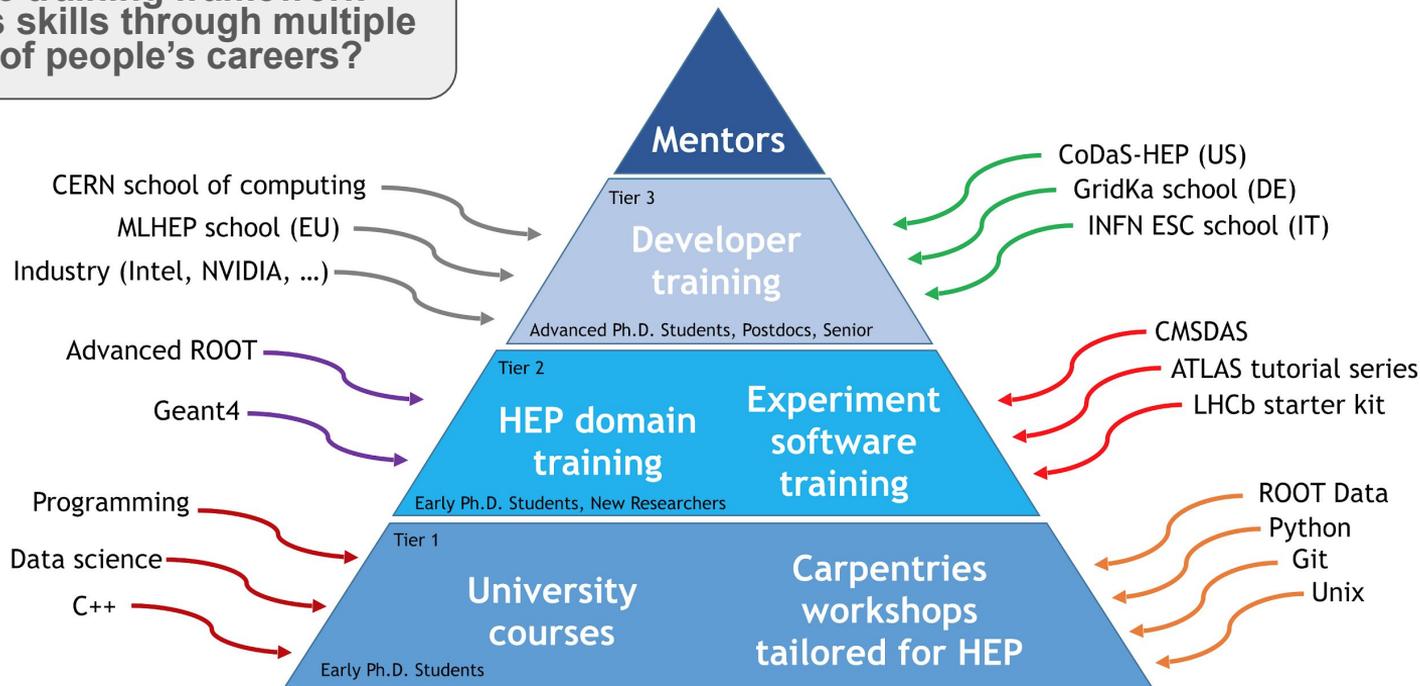
Original HEP Software Foundation Training Vision

This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-1836650) and a separate project FIRST-HEP (OAC-1829707, OAC-1829729, http://first-hep.org) which are working towards implementing this model.

Original HEP Software Foundation Training Vision

How to build a sustainable and scalable training framework that grows skills through multiple stages of people's careers?

Mentors

Tier 3
Developer training
Advanced Ph.D. Students, Postdocs, Senior

CERN school of computing
MLHEP school (EU)
Industry (Intel, NVIDIA, ...)

CoDaS-HEP (US)
GridKa school (DE)
INFN ESC school (IT)

Tier 2
HEP domain training
Experiment software training
Early Ph.D. Students, New Researchers

Advanced ROOT
Geant4

CMSDAS
ATLAS tutorial series
LHCb starter kit

Tier 1
University courses
Carpentries workshops tailored for HEP
Early Ph.D. Students

Programming
Data science
C++

ROOT Data
Python
Git
Unix

Growth of the Training project 2018-2021

We have made significant progress in the past few years. The HSF Training Working Group has brought together the HEP community and has focused efforts not only on running a number of training events (despite COVID over the past year), but also the development of a first draft of a common curriculum.

https://arxiv.org/abs/2103.00659

The effort encompasses both funded projects such as IRIS-HEP, FIRST-HEP (recently joined by SWIFT-HEP in the UK), but also a large number of enthusiastic volunteers. We propose to build on and expand the effort in the coming 3 years by defining a clear target in the form of a community **"Training Challenge"**.

Beginner level

Module	Description	Status	Authors	Repo	Site/Material
The Unix Shell	Introduction to the unix command line/shell	✅	authors		
SSH	Introduction to the Secure Shell (SSH)	α	authors		
Version controlling with git		✅	authors		
Advanced git		α	authors		
Programming with python		✅	authors		
HEP C++ Course		✅	Sebastien Ponce		
Basic Modern C++		α	authors		
Build systems: cmake		✅	authors		
Distributed file systems and grid computing					
ROOT					
uproot	Reading and writing ROOT files without having to install ROOT.	β	authors		
A simple analysis	A simple analysis using CMS open data	✅	authors		
Unit testing	Unit testing in python	β	authors		
Matplotlib for HEP		α	authors		

Intermediate

Module	Description	Status	Authors	Repo	Site/Material
Parallel programming					
Docker	Introduction to the docker container image system	✅	authors		
Workflows & reproducability	E.g. yadage and reana				
Machine learning		✅	authors		
Machine learning on GPU		✅	authors		
CI/CD	Continous integration and deployment with gitlab	✅	authors		
CI/CD github	Continous integration and deployment with github actions	β	authors		

Advanced

Module	Description	Status	Authors	Repo	Site/Material
Documentation	sphinx, doxygen, etc.				
Event generation and MC	pythia, sherpa, madgraph, etc.				
alpaka	alpaka is a header-only C++ abstraction library for accelerator development	α	authors		

Training Challenge

We are now working to define, with the larger community, a series of specific goals for the period 2021-2023 in four categories and to work with the community to achieve them.

Scalability - We aim for sufficient scalability in the training activities such that all students and postdocs can receive training in both the introductory material and the more advanced material. In the steady state we expect a required scale approximately equal to the number of incoming students each year.

Sustainability - We aim to develop community processes by which both the instructors involved in training activities, and the training materials themselves, are continually renewed and meet the other two goals.

Training Scope - We aim for a curriculum (introductory, intermediate, advanced) that broadly meets the needs of the community and evolves over time as needed.

Diversity and Inclusion - The participation in the training should be representative of our community and (as we engage earlier in the pipeline) should work to represent the society at large

Scalability

What is the actual scale at which we are aiming? (In 2022, in 2023, for "Beginner" level events, for "Intermediate events", etc.) How many events does that imply?

Are these events in-person / online / hybrid? (What are the tradeoffs? See Ben's talk...)

What kind of events are these? Standalone? Connected to other types of experiment events? (New collaborator "induction" events?) Cross-experiment?

How do we actually reach "all incoming collaborators"? (Broadcast on mailing lists? Proper recognition by the experiments as part of "induction" or registering as a collaborator?) Which experiments want to be actively involved?

How can we measure the impact? Can we create/adopt standardized surveys so that we can compare across events, experiments, years? (e.g. Carpentries)

Sustainability and Training Scope

The cost (in time and effort) to run an increasing number of events cannot all fall on a small group of people. We also have to prepare for specific individuals to come and go as trainers. What is the metric we will use here?

How do we renew and *incentivize* new trainers to continually join the team?

- One obvious model is former "students" participating as trainers.
- Another is incentives like funding for travel to events or "Training Fellows". Can we cost this and get some sponsors for events in 2022 and 2023?
- A third might be some formal "training the trainers" activity (e.g. such as offered by the Carpentries) that people can put on their CVs.

What mechanism is used to update the material itself and/or add new material?

Diversity and Inclusion

Do we have an inclusive environment in which everyone can and feels welcome to participate in the training activities? If not, what should we be doing to create it?

Although most of these training activities will likely aim at graduate students and postdocs, the material would of course work for many undergraduates. Should we have some additional goals to work with students "farther up the pipeline" (and perhaps also diversify the pool of institutions/countries involved)?

Here metrics are important: if we can't measure it, we can't improve it. Can we standardize on a set of surveys regarding demographics information and use those consistently? (We probably shouldn't be inventing our own here from scratch, others have thought through these things.)

"Call to Action"

I hope we will have a productive discussion today, and perhaps new ideas will arise.

We will probably not nail down answers to all of those questions today, including a fully concrete "Training Challenge" set of goals and plans to achieve.

However the best outcome will be if come out of this meeting is we get a solid strawman picture down as well as concrete actionable items that specific individuals will follow up on.

Then a 2nd iteration of this meeting in a month or two might be able to nail down a reference "2 pager" with specific information on the evolving goals and the plan to achieve them.

[Also are there opportunities for funding to realize our goals?]