

# Software Training

With a focus on cross-experiment training

Kilian Lieret<sup>1,2</sup>, Wouter Deconinck<sup>8</sup>, Peter Elmer<sup>1,2</sup>, Sudhir Malik<sup>3</sup>, Alexander Moreno<sup>4</sup>, Stefan Roiser<sup>5</sup>, Jason Vaetch<sup>6</sup>, Michel Hernandez Villanueva<sup>7</sup>  
**for the IRIS-HEP/HSF Training Group**

<sup>1</sup>Princeton University

<sup>2</sup>IRIS-HEP

<sup>3</sup>University of Puerto Rico

<sup>4</sup>Universidad Antonio Nariño

<sup>5</sup>CERN

<sup>6</sup>Universität Göttingen

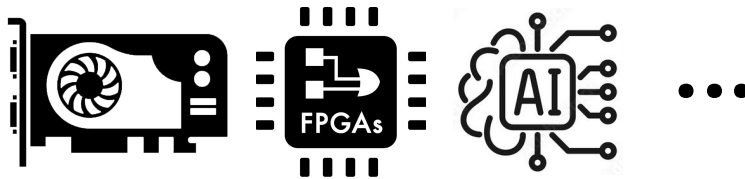
<sup>7</sup>DESY

<sup>8</sup>University of Manitoba



# The need for software training

- As we collect more and more data and perform increasingly complex analyses, our **software is mission critical**  
→ need to follow **industry standards & best practices**
- The broader **data analysis ecosystem is evolving** faster than ever, but these changes are driven by industry → we **must keep pace**



- (Almost) all scientists write software but few have formal software education  
→ **almost every scientist needs to be trained**

# Much is being done already...

## Experiment onboarding

Recent [talk at CHEP](#) from the HSF Data Analysis and HSF Training WGs (proceedings in progress)

Training and onboarding initiatives in HEP



Allison Reinsvold Hall	(US Naval Academy)
Gabriele Benelli	(Brown University)
Ben Carlson	(Westmont College)
Claire David	(York University/FNAL)
Jonathan Davies	(U. of Manchester)
Wouter Deconinck	(U. of Manitoba)
Stephan Hageboeck	(CERN)
Kilian Lieret	(Princeton University)
Lera Lukashenko	(Nikhef)
Sudhir Malik	(U. of Puerto Rico)
Andy Morris	(Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France)
Heidi Schellman	(Oregon State University)
Nicole Skidmore	(U. of Manchester)
Jason Veatch	(Cal State, East Bay)
Michel Hernandez Villanueva	(DESY)

Common **challenges** for all initiatives:

- Supporting **self-study** and **events** (virtual, in-person, hybrid)
- **Maintaining** up-to-date materials (unit testing and regular updates)
- **Consolidate** different resources (Sphinx, doxygen, twiki, forums, ...)
- **Interactivity** & active learning
- **Sustaining** educator workforce

No two experiments have the same strategy; every experiment excels at certain aspects; not the focus of this talk

# Much is being done already...

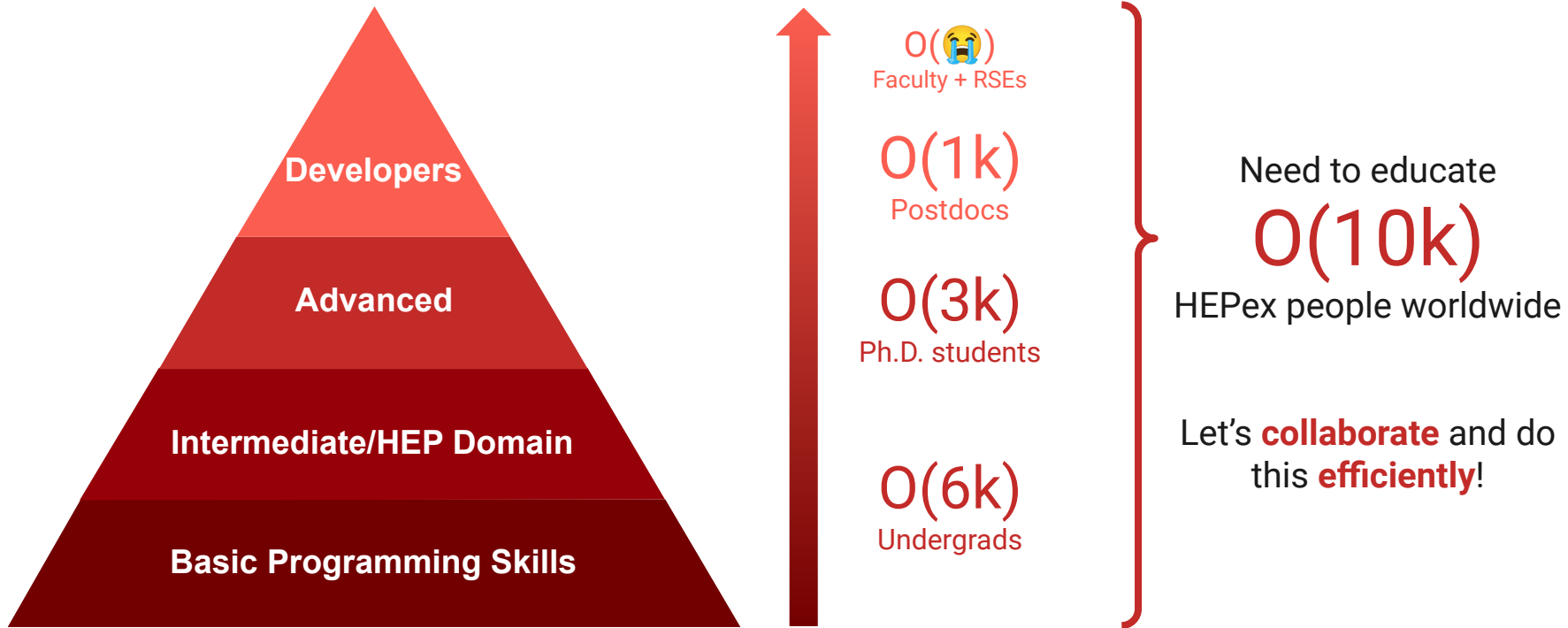
## Cross-experiment training schools

**Unsurpassed** for hands-on training and networking, but **limited #participants** and **self-study support**



... and more

# Much is being done already... but we need to scale!



We need a **unified**, **scalable**, and **sustainable** software training framework



## We need a **unified**, **scalable**, and **sustainable** software training framework

### Unified

- Material and events should be **centrally listed** & **discoverable**
- Concentrate efforts by developing **cross-experiment** content
- A **community** must guide, support, and coordinate

### Scalable

- Material must be teachable by **multiple instructors**
- **Self-study** must not be an afterthought

### Sustainable

- Material must be **open source** and **maintained collaboratively**
- **Incentives & recognition** important motivators

The **IRIS-HEP/HSF Training** group is building a community around these principles

# A unified Training Center for HEP

HEP specific tools

<p><b>Scikit-HEP</b></p> <p>A collection of packages for particle physics analyses in Python.</p> <p>Start learning now!</p> <p>Contribute!</p>	<p><b>ROOT</b></p> <p>The most famous data analysis framework used in HEP.</p> <p>Start learning now!</p> <p>Contribute!</p>	<p><b>UnROOT</b></p> <p>Open ROOT files in Julia!</p> <p>* Status: Beta testing</p> <p>Start learning now!</p> <p>Contribute!</p>
<p><b>Reproducible analyses with REANA</b></p> <p>Run containerised data analysis pipelines on remote compute clouds.</p> <p>Start learning now!</p> <p>Contribute!</p>	<p><b>Particle physics methods</b></p> <p>Learn about ROOT, RooFit, machine learning with TMVA, and physics simulations.</p> <p>* Status: Beta testing</p> <p>Start learning now!</p> <p>Contribute!</p>	

Analysis preservation

<p><b>Version controlling with git</b></p> <p>Track code changes, undo mistakes, collaborate. This module is a must.</p> <p>Start learning now!</p> <p>Contribute!</p>	<p><b>CI/CD (gitlab)</b></p> <p>Continuous integration and deployment with gitlab.</p> <p>Start learning now!</p> <p>Watch the videos!</p> <p>Contribute!</p>	<p><b>CI/CD (github)</b></p> <p>Continuous integration and deployment with github actions.</p> <p>Start learning now!</p> <p>Watch the videos!</p> <p>Contribute!</p>
<p><b>Docker</b></p> <p>Introduction to the docker container image system.</p> <p>Start learning now!</p> <p>Watch the videos!</p> <p>Contribute!</p>	<p><b>Singularity</b></p> <p>Introduction to containerization with Singularity/Apptainer.</p> <p>Start learning now!</p> <p>Watch the videos!</p> <p>Contribute!</p>	<p><b>Unit testing</b></p> <p>Unit testing in python.</p> <p>* Status: Beta testing</p> <p>Start learning now!</p> <p>Contribute!</p>

Reproducible analyses with REANA

Run containerised data analysis pipelines on remote compute clouds.

Start learning now!

Contribute!

**25 modules**  
**0(500) views/month**

- HSF Training Center currently lists **25 training modules** of various authors
- Want to become **focal point for all HEP Training activities**
- Central list of training events (everyone can add)

- 7 Aug - 11 Aug 2023 - Open Science Grid User School 2023
- 17 Jul - 21 Jul 2023 - CODAS-HEP 2023
- 13 Jul - 14 Jul 2023 - HSF/IRIS-HEP Software Basics Training (Virtual) **HSF**
- 12 Jul - 12 Jul 2023 - HSF/IRIS-HEP Python for Analysis Training (Virtual) **HSF**
- 18 May - 19 May 2023 - HSF/IRIS-HEP Software Basics Training (Virtual) **HSF**
- 6 Mar - 10 Mar 2023 - 6th HEP C++ Course and Hands-on Training - The Essentials **HSF**
- 16 Jan - 20 Jan 2023 - Analysis Preservation Workshop **HSF**
- 11 Oct - 13 Oct 2022 - 5th HEP C++ Course and Hands-on Training - Advanced C++ **HSF**
- 3 Oct - 8 Oct 2022 - ESC22 EFFICIENT SCIENTIFIC COMPUTING
- 28 Sep - 30 Sep 2022 - HSF/IRIS-HEP Software Basics Training **HSF**
- 4 Sep - 17 Sep 2022 - CERN School of Computing 2022
- 13 Jul - 15 Jul 2022 - Software Carpentry (Virtual) **HSF**
- 19 Jun - 25 Jun 2022 - Thematic CERN School of Computing on "Security of Research Computing Infrastructures"
- 1 May - 7 May 2022 - Thematic CERN School of Computing on "Scientific Software for Heterogeneous Architectures"
- 21 Apr - 22 Apr 2022 - Matplotlib Training (Virtual) **HSF**
- 28 Mar - 30 Mar 2022 - Software Carpentry (Virtual) **HSF**
- 15 Mar - 17 Mar 2022 - 4th HEP C++ Course and Hands-on Training - The Essentials
- 13 Dec - 15 Dec 2021 - Software Carpentry (Virtual) **HSF**
- 10 Oct - 16 Oct 2021 - Thematic CERN School of Computing (ICSC autumn 2021)
- 25 Aug - 27 Aug 2021 - Software Carpentry (Virtual) **HSF**
- 17 Jul - 22 Jul 2021 - 2021 Intensity Frontier Computing Summer School
- 14 Jun - 18 Jun 2021 - Thematic CERN School of Computing (ICSC spring 2021 - online)
- 16 Feb - 20 Feb 2021 - GitHub CI/CD Training **HSF**

# Building a community

## Weekly meetings

October 2022	
24 Oct	Training WG Planning Meeting
17 Oct	Training WG Planning Meeting
10 Oct	Training WG Planning Meeting
03 Oct	Training WG Planning Meeting
September 2022	
26 Sept	Training WG Planning Meeting
19 Sept	Training WG Planning Meeting
12 Sept	Training WG Planning Meeting
06 Sept	HSF Training Containerization Hackathon
05 Sept	Training WG Planning Meeting

## Increasing our reach



## Platforms



## Community pages

## Monthly Hackathons

## Recognition

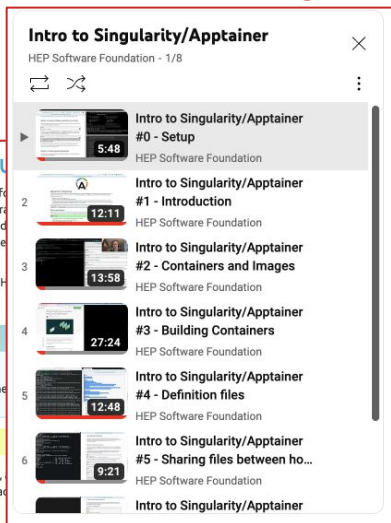
## How-to guides



# Scalable & sustainable trainings

Example 1: HSF Analysis preservation training (5 individual modules)

## Video walkthroughs



## Bundling 5 Modules

**Introduction to Singularity/Apptainer**

Apptainer (formerly known as Singularity) is a free and open-source container platform called "containers" in a simple, portable, fast, and secure manner. It performs operations on container platforms are available, but Apptainer is designed to bring containers and Computing (HPC) use cases. Using Apptainer, developers can work in reproducible environments can be easily copied and executed on other platforms.

This is an introduction to Singularity/Apptainer, its motivations and applications in HPC. Based on the [Apptainer user guide](#).

**Prerequisites**

- Basic knowledge of the Unix Shell, e.g., from the [carpentry course](#).
- Access to a computing system with Apptainer/Singularity available. It can either be installed on a local machine or accessed via a remote environment.

**HSF Software Training**

This training module is part of the HSF Software Training Center, which provides a central location for all HSF training materials. It is designed to be self-paced and self-study ready, and includes best practices for learning and sharing.

**Clear prerequisites**

Setup

Download files required for the lesson

00:00 1. Introduction

What issues motivated the creation of Apptainer/Singularity?

What are the differences between Docker, Singularity and Apptainer?

What are the differences between Docker, Singularity and Apptainer?

What are the differences between Docker, Singularity and Apptainer?

What are the differences between Docker, Singularity and Apptainer?

Many of our modules embrace the framework of [The Carpentries](#)

- Built from **markdown** files (easy & maintainable!)
- Rendered as a **webpage** with [Jekyll](#)
- Verbose and **self-study ready**
- Events use combination of self study with videos + live lectures + small-group virtual mentoring sessions

# Scalable & sustainable trainings

## Example 2: HSF C++ course

● TeX ☆ 95 🗄 Apache-2.0 🔑 40 ⌚ 50 🔗 5 Updated 7 minutes ago

Intro base OO More exp Tool conc py

### 548 slides, 695 pages, > 1k commits

## HEP C++ course

B. Gruber, S. Hageboeck, S. Ponce  
sebastien.ponce@cern.ch

CERN 8 events till now

- 8th HEP C++ Course and Hands-on Training (essentials), Manchester, August 2023
- 7th HEP C++ Course and Hands-on Training (essentials), JLAB, May 2023
- 6th HEP C++ Course and Hands-on Training (essentials), CERN, March 2023
- 5th HEP C++ Course and Hands-on Training (advanced), CERN, October 2022
- 4th HEP C++ Course and Hands-on Training (essentials), CERN, March 2022
- 3rd HEP C++ Course and Hands-on Training, CERN, August 2021
- 2nd HEP C++ Course and Hands-on Training, virtual, January 2021
- 1st HEP C++ Course and Hands-on Training (2020 October)

B. C

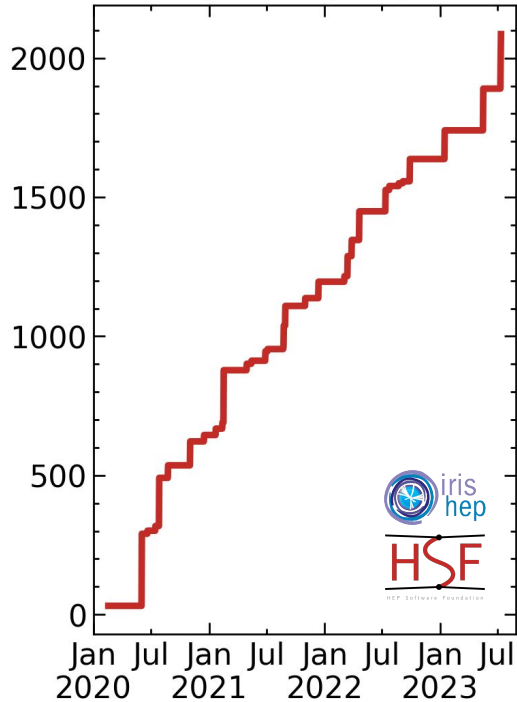
- Slides generated from **LaTeX** source
- Has been taught **in-person, virtual and hybrid**
- **Life lectures and exercise sessions**
- Full **videos** available from events

Originally developed by **S. Ponce**, now community effort driven by B. Gruber, S. Hageboeck et. al.

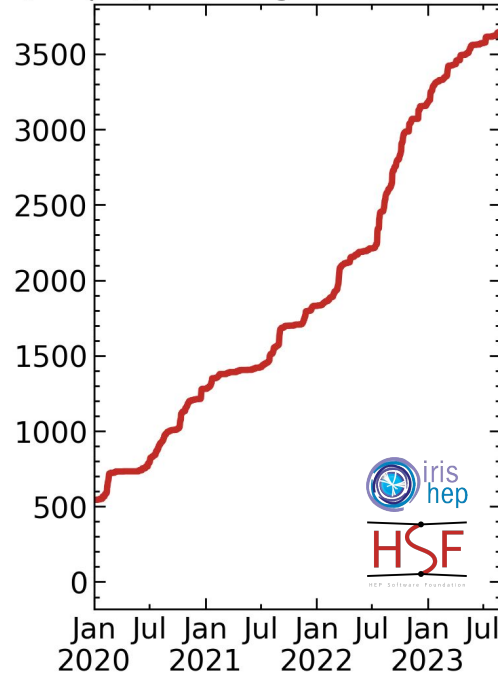
Abhishek L	Attila Krasznahorkay	Bernhard Manfred Gruber	David Chamont	Enrico Guiraud	Graeme A Stewart	Kilian Lieret
Sebastien Ponce	Stefan Roiser	Stephan Hageboeck	bcouturi			

# We scale!

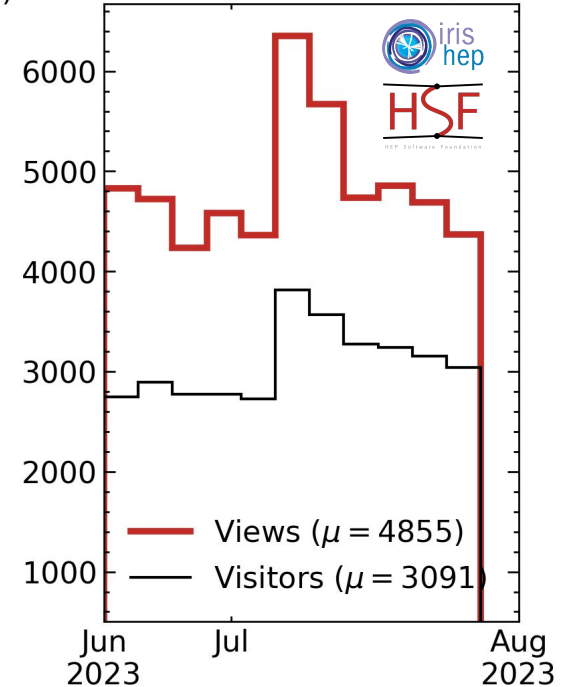
Cumulative registrations  
(total=2,086)



Material additions/revisions  
(unique; excluding SWC; total=3,645)



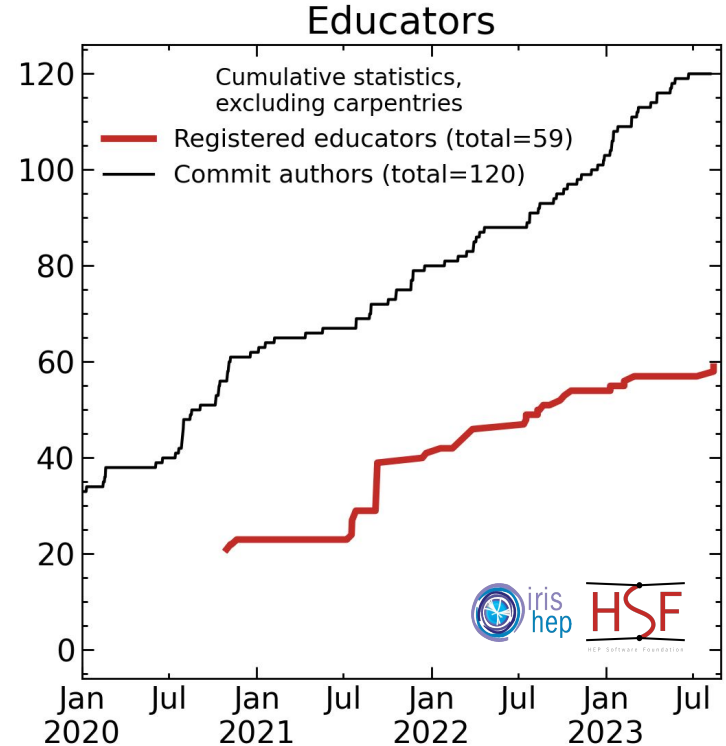
Weekly visitors and views  
(subset of trainings)



# Sustainability

Training is mission-critical and costs significant resources → Efforts must be incentivized:

- **Acknowledged:** HSF credits contributors both on our community page and on individual lessons, but could more be done?
- **Encouraged:** Supervisors must motivate and push their students/ employees to contribute
- **Rewarded:**
  - **Service work credits** should be given for all contributions (including cross-experiment training)
  - **Hiring discussions** should acknowledge and push for contributions



# Conclusion

- Software training efforts are **central**
- **Experiment-specific** training initiatives differ vastly between experiments (let's learn from each other's strengths!)
- **Experiment-independent** topics are the foundation on which to build; we need to scale up training in this area
- The HSF Training WG built a **community of educators** around training material that is **open source, verbose** and **sustainably maintainable**
- Scaled to **> 2000 learners** and more could be done
- Our **training center** aims to be a central entry point to all relevant material
- We **need buy-in from the experiments** to grow and sustain our efforts:
  - Need instructors & mentors
  - We welcome & support new material
  - Cross-linking our training center/ material will increase readership
- Contributions to training must be
  - **encouraged** (by supervisors)
  - **rewarded** (service work credits, hiring, ...)



# Thank you!



@hepssoftfound



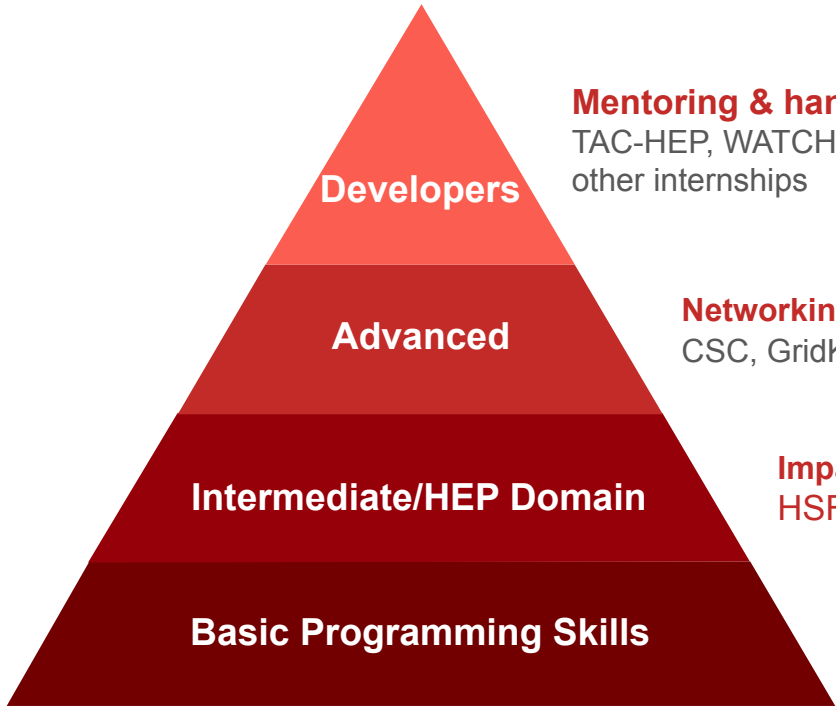
@hsf-training



hepssoftwarefoundation.org



# Much is being done already. [WIP] WILL BE REPLACED BY COMPLETELY DIFFERENT SLIDE



**Mentoring & hands-on experience:**  
TAC-HEP, WATCHEP, IRIS-HEP, GSoC,  
other internships

**Networking & catch up with recent developments:**  
CSC, GridKA, CoDaS-HEP, MLHEP, INFN ESC, ...

**Impart best practices & HEP domain knowledge:**  
HSF Intermediate Training, experiment StarterKits, ...

**Democratize science:**  
HSF software basic training, University courses,  
Self-study, ...