# Training Activities and Experiences in the HSF.

## Delivering sustainable software training

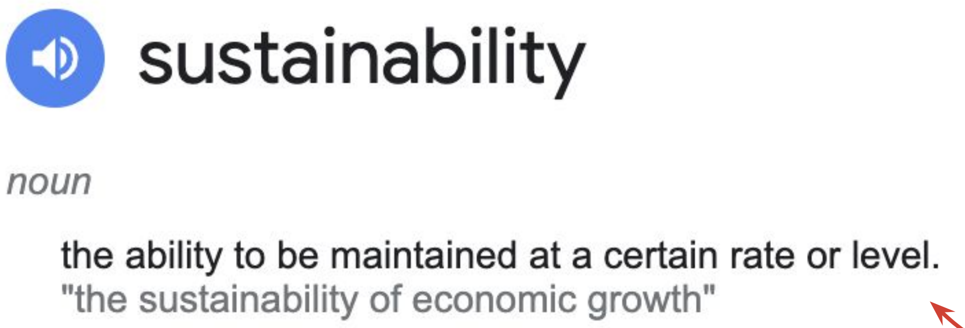
Michel Hernandez Villanueva (DESY),
and many more! (see slide 12)

On behalf of the HSF Training WG

**FH Sustainability Forum meeting**
June 5, 2023

# Definition

## Or “Why am I here today?”



sustainability

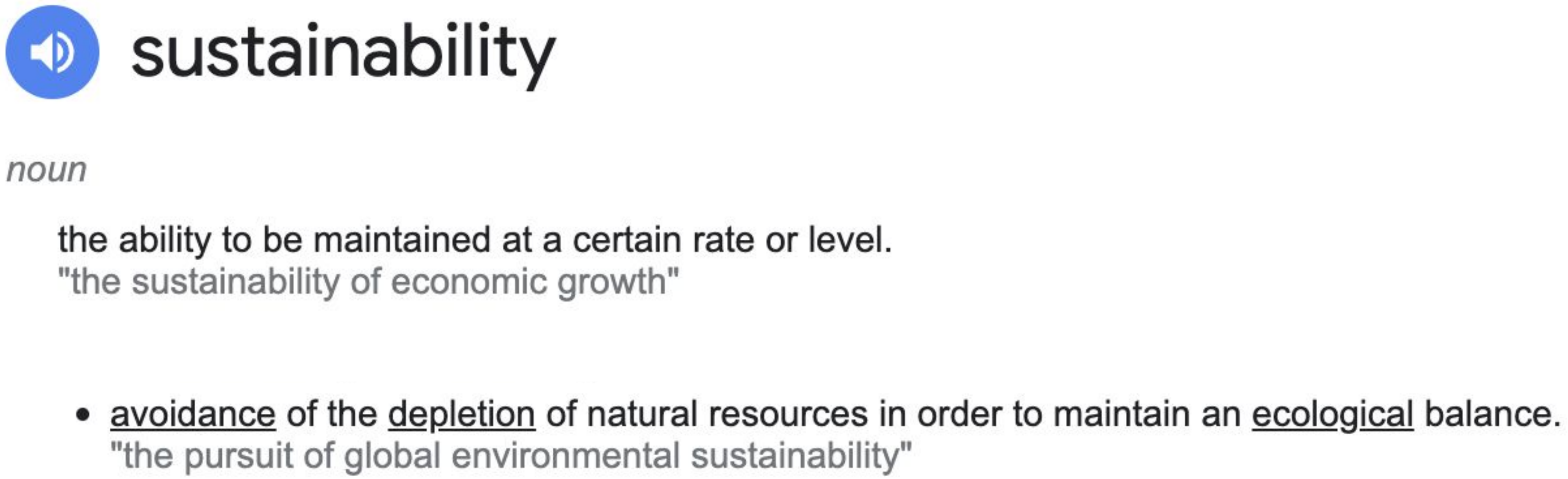
*noun*

the ability to be maintained at a certain rate or level.
"the sustainability of economic growth"

*What we mean by “sustainable training” in other talks*

# Definition

## Or “Why am I here today?”



sustainability

*noun*

the ability to be maintained at a certain rate or level.
"the sustainability of economic growth"

- avoidance of the depletion of natural resources in order to maintain an ecological balance.
  "the pursuit of global environmental sustainability"

*What we will also discuss today*

# Software Development in HEP

## As a key for a successful scientific program

- Scientific collaborations are **big and growing**.
  - O(1000) collaborators in hundreds of institutes around the world.

# Software Development in HEP

## As a key for a successful scientific program

- Scientific collaborations are **big and growing**
  - O(1000) collaborators in hundreds of institutes around the world
- High Energy Physics (HEP) and Nuclear Physics (NP) are **computationally intensive** and **data driven** fields
  - A full physics potential requires investment into the software used to collect, process, and analyse data
- **Developers with strong foundation** are critical resources in the success of the current and future experiments
  - The researchers must be brought up to date with new software technologies, concurrent programming, and artificial intelligence
  - They must maintain, improve, and sustain the software

# Training and Onboarding Initiatives in HEP

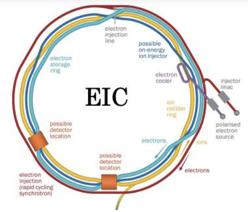
## How do experiments teach software?

Virtual — Hybrid — In person

Belle II: Online book

LHCb: Starter Kit

CMS: Data Analysis Schools


EIC
electron injection line
possible on-energy ion injector
electron storage ring
electron cooler
injector linac
ion collider ring
polarised electron source
possible detector location
electrons
ions
possible detector location
electrons
electron injection (rapid cycling synchrotron)

Online tutorials

ATLAS EXPERIMENT: Software tutorials

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT: Synchronous tutorials “Carpentries-style”

“Software is different, but challenges are common”

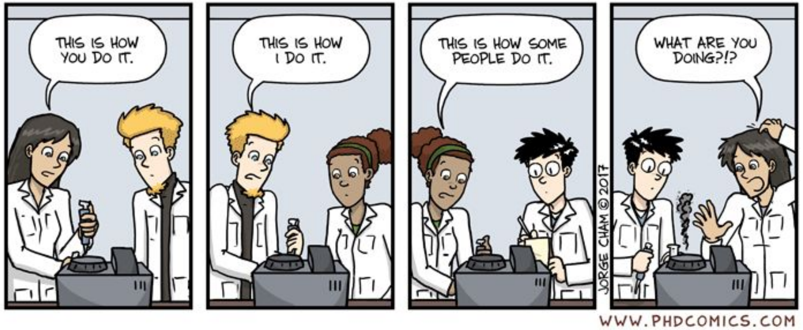
Reinsvold Hall, Allison (US Naval Academy), CHEP 2023

# HEP Software Training

## Why not leave everything to the experiments?

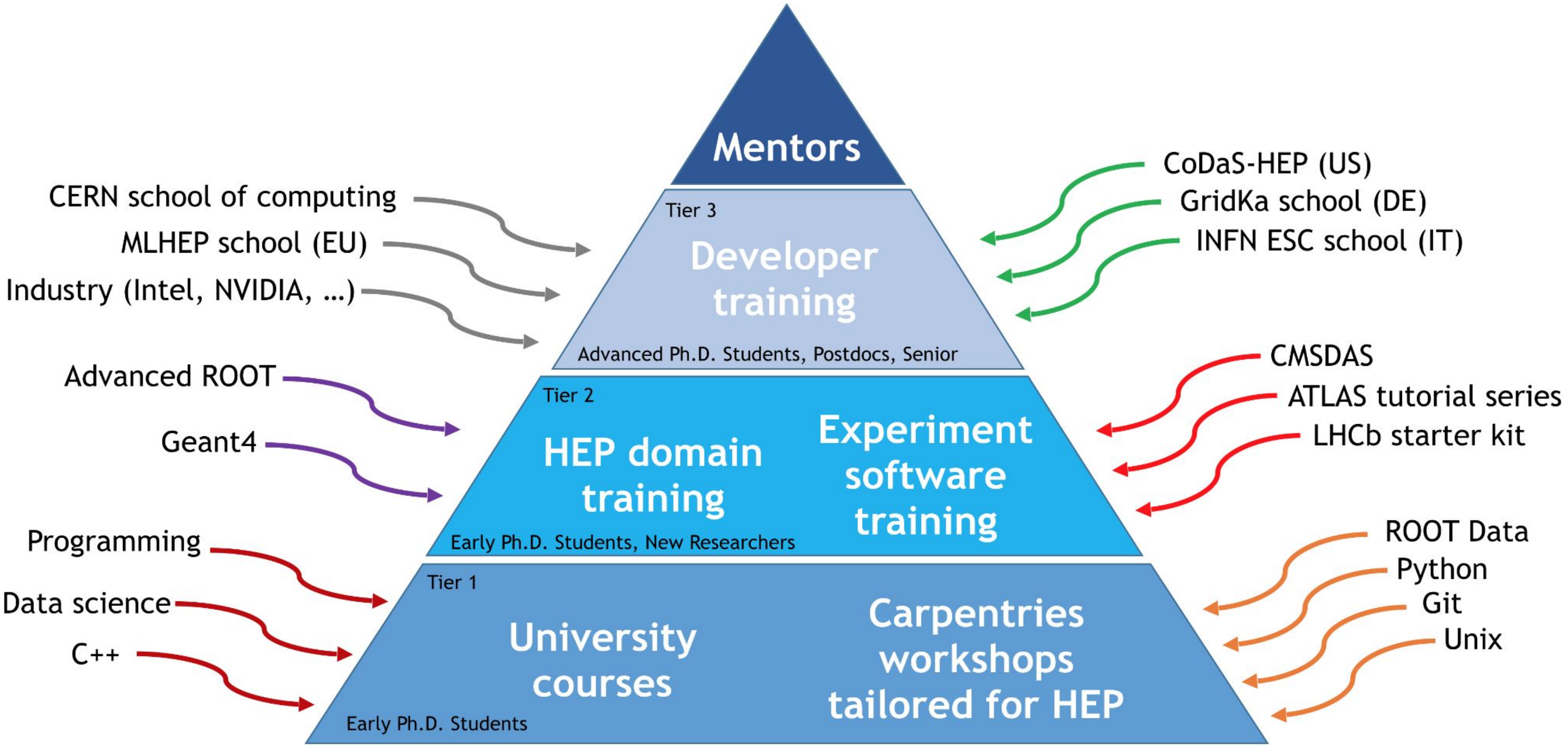
- O(10k) HEP people worldwide need to be trained in software engineering & computing
- Common challenges faced:
    - Most people developing code have non-permanent positions with contracts of 2 - 4 years
    - Training activities are undervalued in making career steps and by funding agencies
    - Material for training is a moving target as technology evolves (e.g., ML, GPUs, FPGAs, …)

- **This should be a community effort!**



# HEP Software Training

## We can cover more ground together!



https://iris-hep.org/ssc.html

# HEP Software Foundation

## (HSF)

- The role of the HEP Software Foundation, started in 2015, is to facilitate coordination and common efforts in software and computing across HEP in general

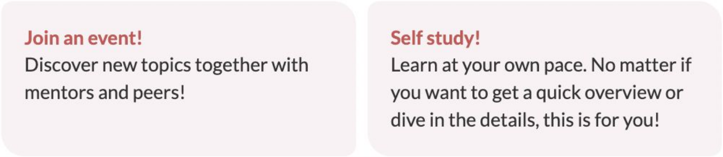
- The goal is to describe a global vision for software and computing for the current and future experiments
  - Working groups cover Training, Analysis, Generators, Simulation, Reconstruction and Software Triggers, etc.

- The HSF's role is one of an information conduit and meeting point
  - Report on interesting and common work being done
  - Forum for technical comments and discussion
  - Encourage cooperation across experiments and regions
  - Motivate the publication of summary documents or papers for reference

# HSF Software Training

## Organization

- Established 4 years ago
- Develops material for an introductory software curriculum
  - And teaches this curriculum to scientists
- Focuses on common software material across HEP
  - From basic core software skills
  - To advanced training required in software and computing
- Remote weekly public meetings (via Zoom) to plan and assess progress
  - Led by four co-conveners
  - Proposals are discussed and events are planned
- Engages with different experimental collaborations and initiatives
  - IRIS-HEP, FIRST-HEP, and The Carpentries



**Join an event!**
Discover new topics together with mentors and peers!

**Self study!**
Learn at your own pace. No matter if you want to get a quick overview or dive in the details, this is for you!

https://hepsoftwarefoundation.org/training

# HSF Software Training

## Principles

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

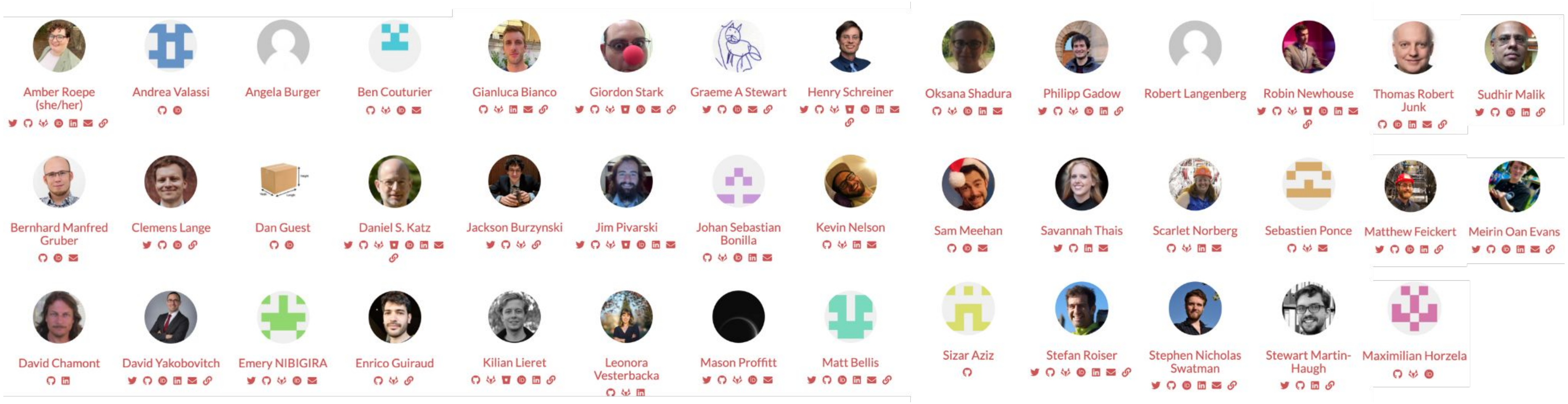
**Sustainable**

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

# HSF Software Training

## The community

- An active community of members supporting training on voluntary basis
  - Coming from multiple collaborations, adding value to the training from different environments
- Profile of each tutor that contributes is included in the HSF training page
  - Public recognition of their capability, skills and contribution



Amber Roepe (she/her)
Andrea Valassi
Angela Burger
Ben Couturier
Gianluca Bianco
Giordon Stark
Graeme A Stewart
Henry Schreiner
Oksana Shadura
Philipp Gadow
Robert Langenberg
Robin Newhouse
Thomas Robert Junk
Sudhir Malik

Bernhard Manfred Gruber
Clemens Lange
Dan Guest
Daniel S. Katz
Jackson Burzynski
Jim Pivarski
Johan Sebastian Bonilla
Kevin Nelson
Sam Meehan
Savannah Thais
Scarlet Norberg
Sebastien Ponce
Matthew Feickert
Meirin Oan Evans

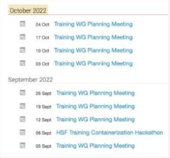
David Chamont
David Yakobovitch
Emery NIBIGIRA
Enrico Guiraud
Kilian Lieret
Leonora Vesterbacka
Mason Proffitt
Matt Bellis
Sizar Aziz
Stefan Roiser
Stephen Nicholas Swatman
Stewart Martin-Haugh
Maximilian Horzela

https://hepsoftwarefoundation.org/training/community.html

# HSF Software Training Platform

## We can cover more ground together!

### Weekly meetings



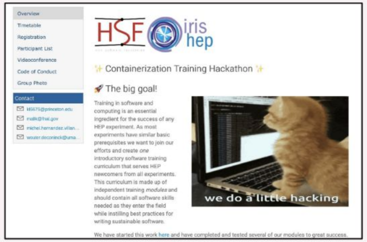
October 2022

- 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

### Monthly Hackathons



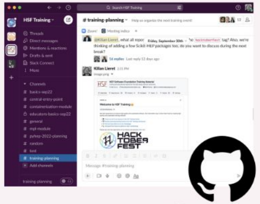
✨ Containerization Training Hackathon ✨

🚀 The big goal!

Training in software and computing is an essential ingredient for the success of any HEP experiment. As most experiments have similar basic prerequisites we want to join our efforts and create *one* introductory software training curriculum that serves HEP newcomers from all experiments. This curriculum is made up of independent training *modules* and should contain all software skills needed as they enter the field while instilling best practices for writing sustainable software.

We have started this work here and have completed and tested several of our modules to great success.

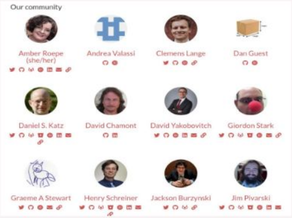
### Platforms



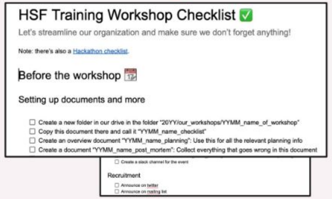
@Kilian Lieret, what all repo[...] the hacktoberfest tag? Also, we're thinking of adding a few Scikit-HEP packages too, do you want to discuss during the next break?

Kilian Lieret

### Community pages



### How-to guides



HSF Training Workshop Checklist ✅

Let's streamline our organization and make sure we don't forget anything!

Note: there's also a Hackathon checklist.

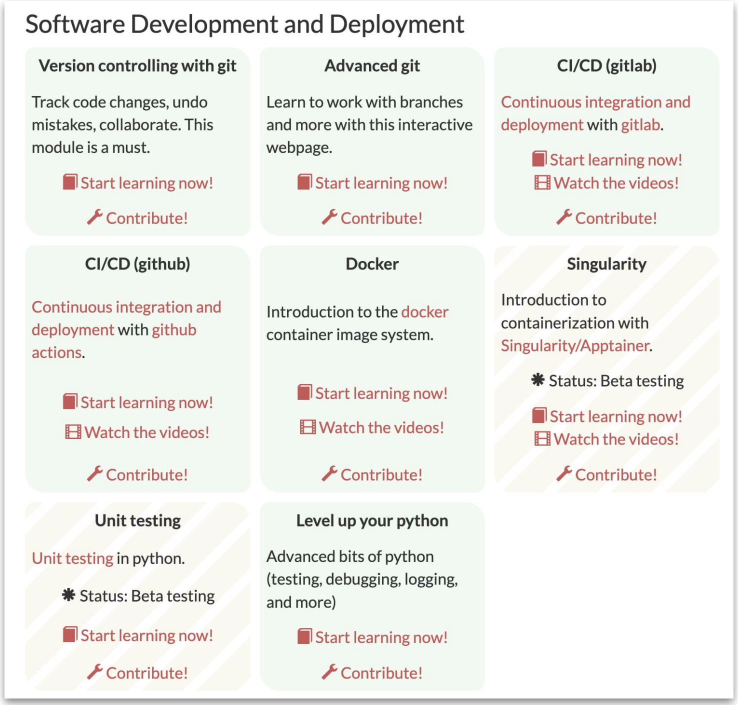
Before the workshop 📅

Setting up documents and more

- □ Create a new folder in our drive in the folder "20YY/our_workshops/YYMM_name_of_workshop"
- □ Copy this document there and call it "YYMM_name_checklist"
- □ Create an overview document "YYMM_name_planning": Use this for all the relevant planning info
- □ Create a document "YYMM_name_post_mortem": Collect everything that goes wrong in this document
- □ Create a slack channel for the event

Recruitment

- □ Announce on twitter
- □ Announce on mailing list



Software Development and Deployment

**Version controlling with git**

Track code changes, undo mistakes, collaborate. This module is a must.

Start learning now!

Contribute!

**Advanced git**

Learn to work with branches and more with this interactive webpage.

Start learning now!

Contribute!

**CI/CD (gitlab)**

Continuous integration and deployment with gitlab.

Start learning now!

Watch the videos!

Contribute!

**CI/CD (github)**

Continuous integration and deployment with github actions.

Start learning now!

Watch the videos!

Contribute!

**Docker**

Introduction to the docker container image system.

Start learning now!

Watch the videos!

Contribute!

**Singularity**

Introduction to containerization with Singularity/Apptainer.

✱ Status: Beta testing

Start learning now!

Watch the videos!

Contribute!

**Unit testing**

Unit testing in python.

✱ Status: Beta testing

Start learning now!

Contribute!

**Level up your python**

Advanced bits of python (testing, debugging, logging, and more)

Start learning now!

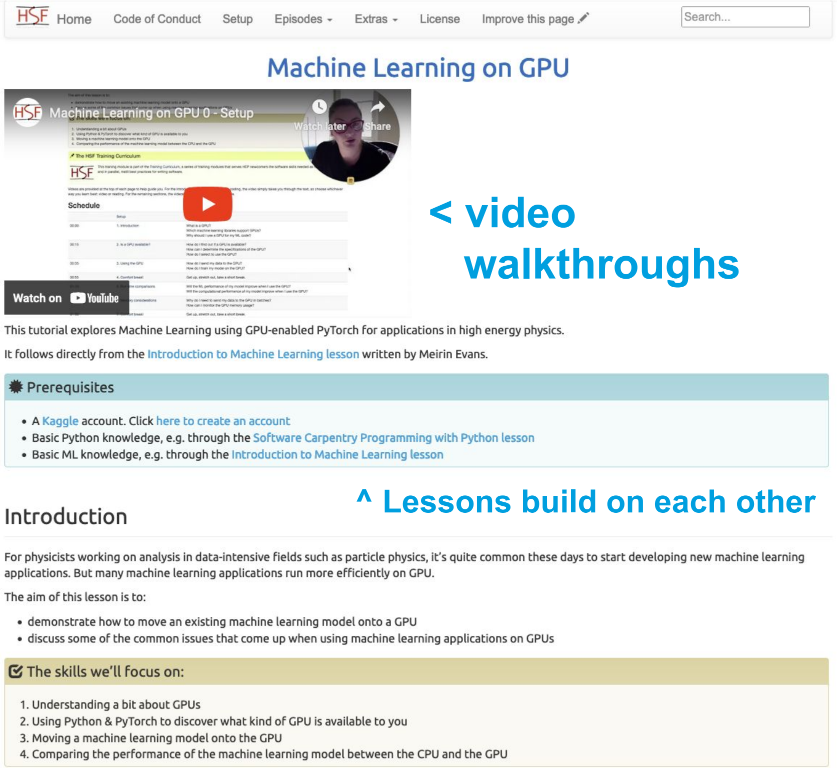
Contribute!

https://hepsoftwarefoundation.org/training/

# HSF Software Training Center

## For HEP newcomers

- HSF Training Center currently lists 21 training modules
  - Basics: Bash, Git, Python, Matplotlib
  - HEP basics: ROOT, Uproot
  - Software development, ML, devops, etc
- Goal: Training Center as a focal point for all HEP Training activities
  - Free and experiment-agnostic
- Embrace the framework of The Carpentries
  - Built from markdown files
  - Source at GitHub (**Anyone can contribute!**)
  - Verbose and self-study ready



Machine Learning on GPU

This tutorial explores Machine Learning using GPU-enabled PyTorch for applications in high energy physics.

It follows directly from the Introduction to Machine Learning lesson written by Meirin Evans.

Prerequisites

- A Kaggle account. Click here to create an account
- Basic Python knowledge, e.g. through the Software Carpentry Programming with Python lesson
- Basic ML knowledge, e.g. through the Introduction to Machine Learning lesson

Introduction

For physicists working on analysis in data-intensive fields such as particle physics, it's quite common these days to start developing new machine learning applications. But many machine learning applications run more efficiently on GPU.

The aim of this lesson is to:

- demonstrate how to move an existing machine learning model onto a GPU
- discuss some of the common issues that come up when using machine learning applications on GPUs

The skills we'll focus on:

1. Understanding a bit about GPUs
2. Using Python & PyTorch to discover what kind of GPU is available to you
3. Moving a machine learning model onto the GPU
4. Comparing the performance of the machine learning model between the CPU and the GPU

< video walkthroughs

^ Lessons build on each other

^ Enough verbosity for self-study

# HSF Software Training Center

## For HEP newcomers

- **The big picture: scientists with skills for delivering high-quality code**

- We must aim to train our community with the best practices for sustainable software development
  A few examples:
  - Continuous Integration
  - Testing, testing, testing
  - Reproducibility, preservation
  - Project development methodologies
  - Green coding practices: efficient algorithms and data structures, reduce memory consumption and network traffic…

- Large impact at computing centers in the long term!

# HSF Software Training Center

## For HEP newcomers

- **The big picture: scientists with skills for delivering high-quality code**

- We must aim to train our community with the best practices for sustainable software development
  A few examples:
  - Continuous Integration
  - Testing, testing, testing
  - Reproducibility, preservation
  - Project development methodologies
  - Green coding practices: efficient algorithms and data structures, reduce memory consumption and network traffic…

↕

We are halfway on this list.

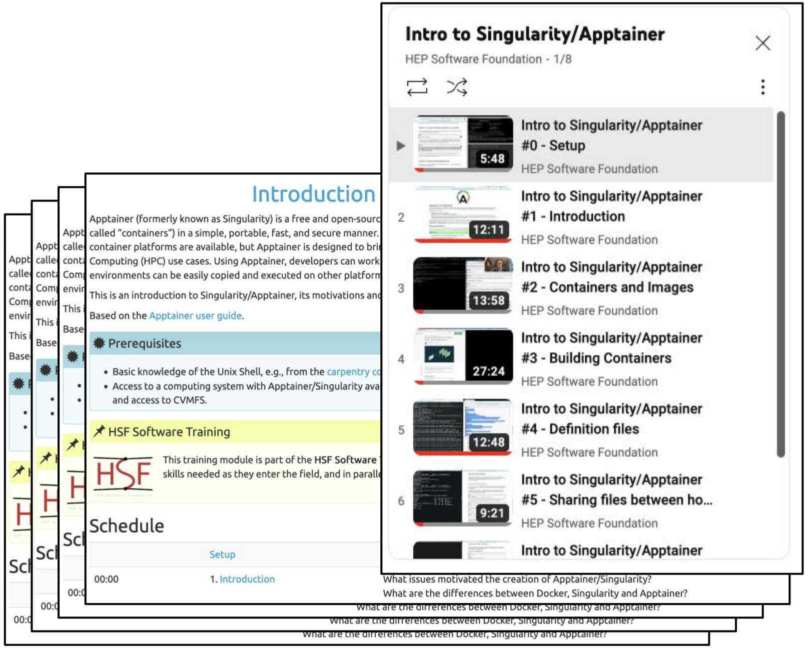
**Reaching the bottom needs support from the HEP community (For example, keeping communication with this forum 😄)**

- Large impact at computing centers in the long term!

# Analysis Preservation

## An example on how-to “train to sustain”

- **“Preservation reduce the resource footprint of our analyses”** [Yves Kemp et al. "Sustainable computing in HEP"]
- Last year, we developed modules to teach how to consider analysis preservation right from the beginning



Introduction

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

This is an introduction to Singularity/Apptainer, its motivations and

Based on the Apptainer user guide.

Prerequisites

- Basic knowledge of the Unix Shell, e.g., from the carpentry co
- Access to a computing system with Apptainer/Singularity ava and access to CVMFS.

HSF Software Training

This training module is part of the HSF Software
skills needed as they enter the field, and in paralle

Schedule

| | Setup | |
|---|---|---|
| 00:00 | 1. Introduction | What issues motivated the creation of Apptainer/Singularity?<br>What are the differences between Docker, Singularity and Apptainer? |

Intro to Singularity/Apptainer

HEP Software Foundation - 1/8

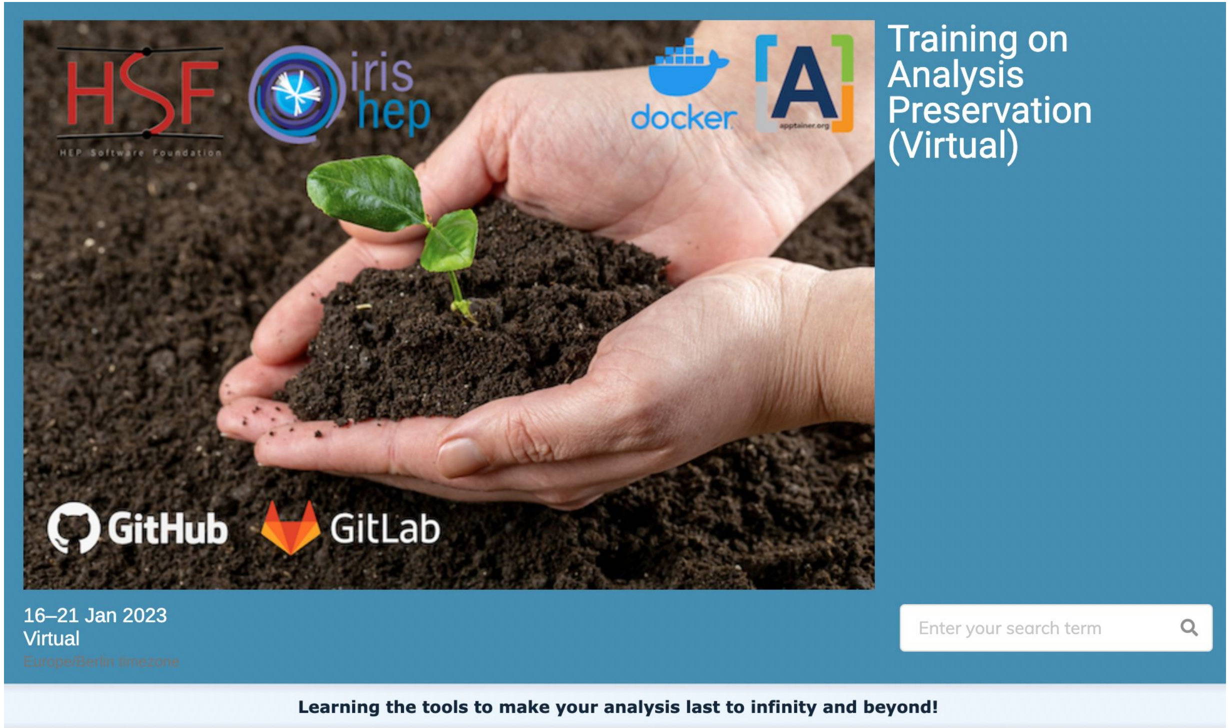
| | | |
|---|---|---|
| ▶ | 5:48 | Intro to Singularity/Apptainer #0 - Setup<br>HEP Software Foundation |
| 2 | 12:11 | Intro to Singularity/Apptainer #1 - Introduction<br>HEP Software Foundation |
| 3 | 13:58 | Intro to Singularity/Apptainer #2 - Containers and Images<br>HEP Software Foundation |
| 4 | 27:24 | Intro to Singularity/Apptainer #3 - Building Containers<br>HEP Software Foundation |
| 5 | 12:48 | Intro to Singularity/Apptainer #4 - Definition files<br>HEP Software Foundation |
| 6 | 9:21 | Intro to Singularity/Apptainer #5 - Sharing files between ho...<br>HEP Software Foundation |
| | | Intro to Singularity/Apptainer |

- Developed by the HEP community during Containerization & Analysis Preservation Hackathons
- Teaching Docker, Singularity/Apptainer, CI/CD with github/gitlab, REANA (soon)
- Using CMS Open Data
- Emphasis on **self study with videos**
- Material can be used in training events

# Analysis Preservation

## Virtual workshop

- A week
- 100 participants
- A good example on how the interest of the community can drive towards training events related to environmental sustainability



Training on Analysis Preservation (Virtual)

16–21 Jan 2023
Virtual
Europe/Berlin timezone

Enter your search term

**Learning the tools to make your analysis last to infinity and beyond!**

https://indico.cern.ch/event/1219810/

# Training Events

## In person

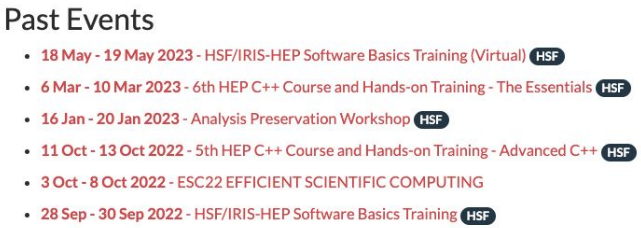
- HSF Training software tutorials through 2020:
  - In-person participation only
  - Approximately 35 participants per workshop

- Impact on ecological and social sustainability:
  - Travel limits the accessibility to research groups with access to sufficient funding
  - Typical carbon footprint ~0.5 t CO2e / person:
    - Intra-continental travel: 0.4 t CO2e per person
    - Hotel stays: ~25 kg CO2e per person per night
  - Compare with estimated average EU (US) annual carbon footprint of 7 (16) t CO2e per person
  - **A workshop can increase one's footprint by 5% to 10%**

# Training Events

## Virtual

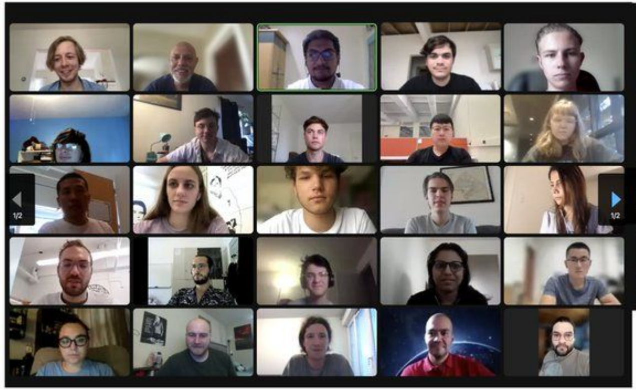
- Holding Virtual events since 2020.
  - COVID-motivated, but this training modality is here to stay
- 18 online software trainings, **1300+ participants trained**
  - Logistics are easier. Recordings available
  - Minimum environmental impact



Past Events

- 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

- But also disadvantages:
  - Lower engagement, distractions
  - Meaningful interactions harder



HSF iris hep THE CARPENTRIES

# Training Events

## Back to in person?

- Discussion on how to quantify how effective **online vs in-person events** are
  - 3 eigenvectors: knowledge exchange, co-creation, community building
  - We need standardise metrics to use throughout the courses

- How to get the advantages of in-person interaction without the environmental impact? Several ideas, discussion in progress
  - Self-organized training events (we provide all the material)
  - Events allocated with major conferences
  - Regional training events

- Or the other way around: **improve interaction in online events**
  - Breaking-Ice sessions at the beginning of the event may bring back the joy of the events when they are organized online

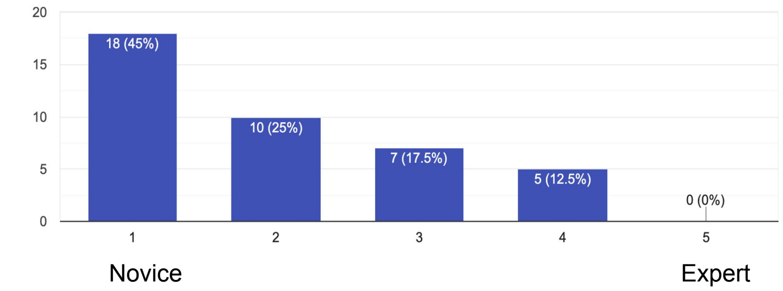
# Training Events

## Measuring how effective are online events

- Each event, anonymized pre and post-surveys are circulated with the students
  - Pre-survey: Demographics, How much do you know?
  - Post-survey: How much do you know **now**? What can we do better next time?
- We also do our best collecting information of people dropping the event

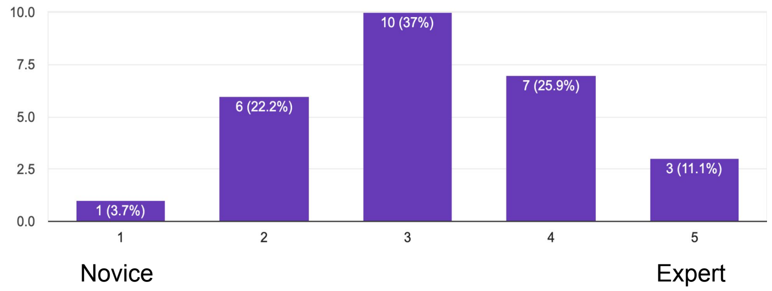
How confident are you in your knowledge and abilities when using Git?

40 responses



How confident are you in your knowledge and abilities when using Git after the workshop?

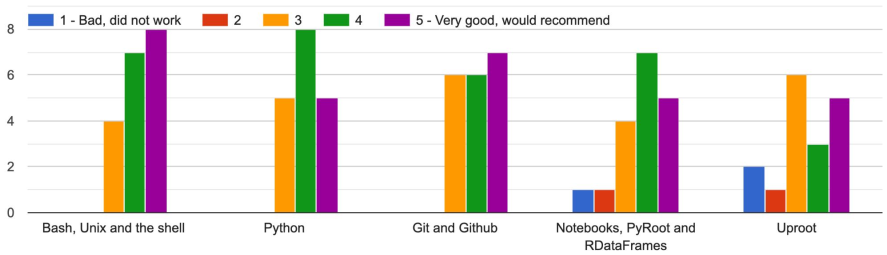
27 responses



# Surveys

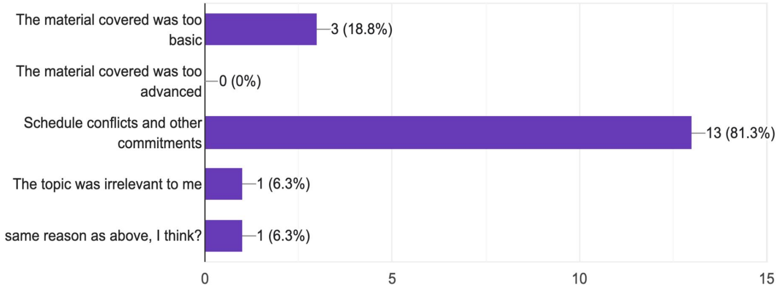
## Some examples

Please rate how successfully each topic was covered (1 - 5)



If you only attended FEW sessions, why did you skip the other sessions?

16 responses



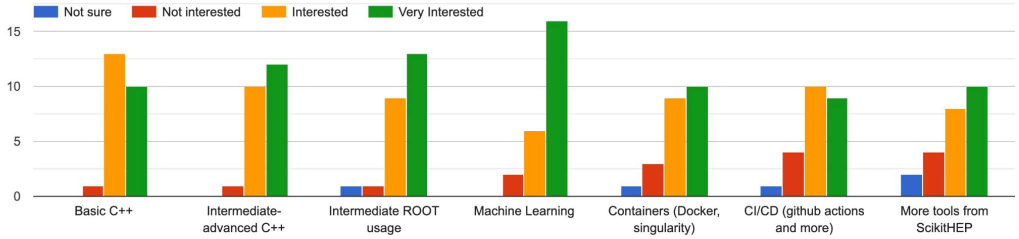
How can we improve our bash/shell training?

3 responses

How can we improve our python training?

1 response

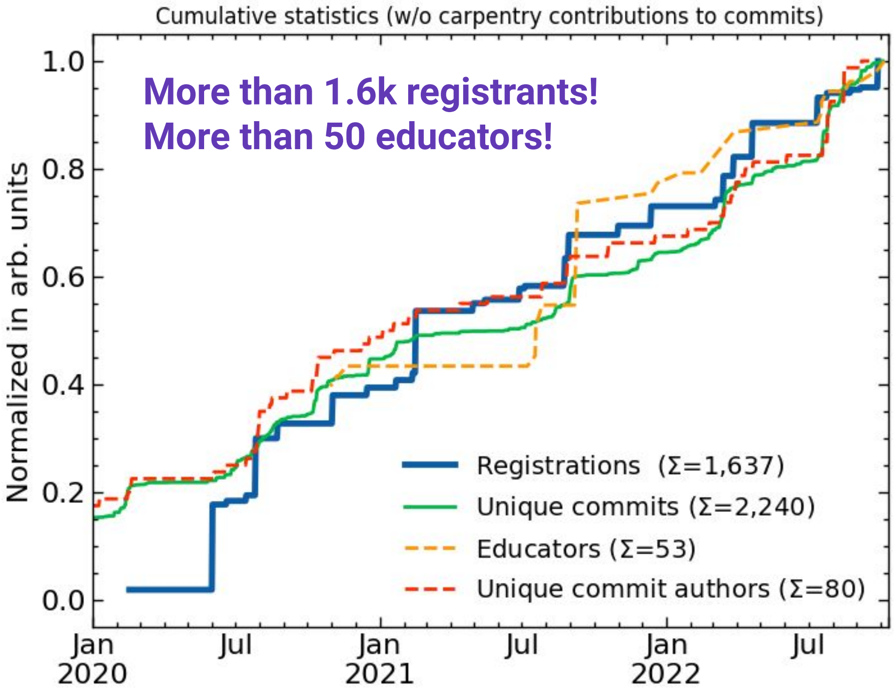
Please rate your interest in attending future tutorials on the following topics:



# Training Challenge

## Scaling up

- Proposal to expand the effort in the long-term, defining a clear target in form of a Training Challenge.
    - **Scalability**: What is the number of students to reach? How many events does imply?
    - **Sustainability**: How to incentivize new trainers to continually join?
    - … **and Sustainability:** How to minimize the environmental impact, delivering effective training?
    - **Diversity and inclusion**: Everyone feels welcome to participate? How to standardize metrics?
- Active discussion happening right now.



# Summary

## And how to collaborate

- The HSF Training is a community-driven effort, covering the software training requisites for a sustainable operation of physics experiments

- We have consolidated Software Training events virtually held
  - Discussion on how to sustain & scale up is relevant and happening right now

- We have included a training event teaching Analysis Preservation: containerization & CI/CD with open data
  - **Extend to more topics related to sustainability depends on the motivation of the community**

- Public weekly meetings: Mondays at 4pm CEST
  - https://indico.cern.ch/category/10294/
  - Everyone is welcome to join!

- **Reach us also via the channels shown in our webpage.**

# Join us!

@hepsoftfound

@hsf-training

hepsoftwarefoundation.org

indico

slack

# Backup

# Training Events

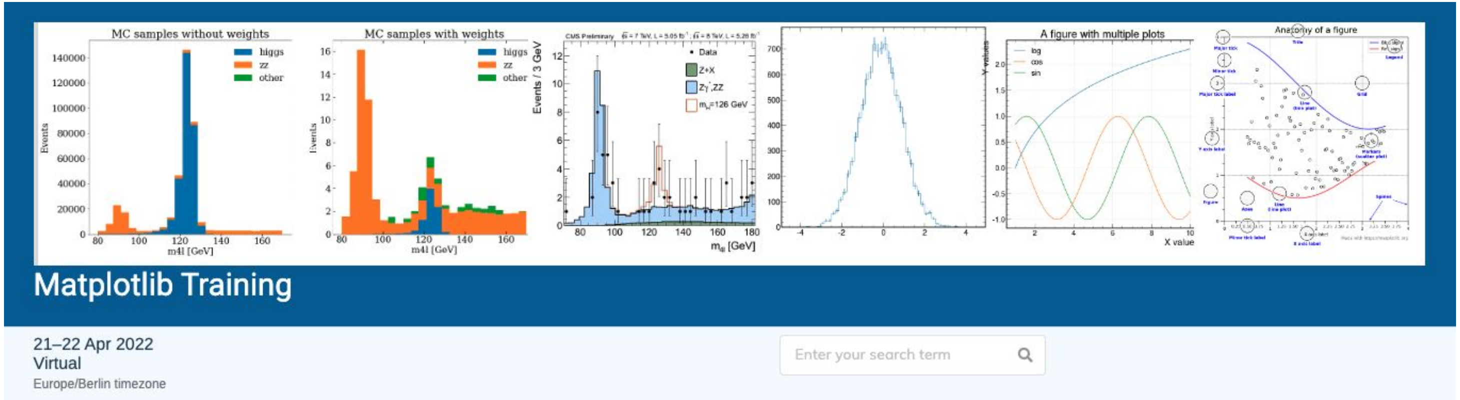
## Software Training Basics

- Since Aug 2021, in collaboration with The Carpentries, we have established a training event for newcomers 3 times per year.
  - Agenda of events in 2022: here, here and here.
- Three days event organized as follows:
  - Two days of fast-track to competency with software fundamentals: Bash, Git, Python.
  - One day dedicated to HEP Software:
    ROOT (data analysis framework), Scikit-HEP (data analysis in Python)
- A limit 80 students per event.
  - Instructors for each topic.
  - 5 mentors in average, helping in breakout rooms.
- Material and recordings are preserved on the page of the event.

# Intermediate Training

## Development and organization

- Continuously organizing training meetings and hackathons for extending/improving material.
- Established training events in C++ and Matplotlib.
- Current topics in development: Docker / Apptainer, CI/CD in GitHub and GitLab.
  - During hackathons, aiming to prepare an Analysis Preservation Training event.
- New ideas are always welcome.



Matplotlib Training

21–22 Apr 2022
Virtual
Europe/Berlin timezone