

HEP Software Foundation

# Training WG

Coordination meeting, June 18

Kilian Lieret<sup>1,2</sup>, Sudhir Malik<sup>3</sup>, Samuel Meehan<sup>4</sup>

<sup>1</sup>Ludwig Maximilian University of Munich

<sup>2</sup>Excellence Cluster Universe

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

<sup>4</sup>CERN





Project objective

Create a full HEP software curriculum  
built from standardized modules of  
hands-on training sessions



# Training modules

- Training module = webpage built from markdown files (using Jekyll)
- Using templates from <https://software-carpentry.org/> (SWC)
- Our modules are hosted at <https://github.com/hsf-training>
- Overview of modules relevant for HEP at <https://hepsoftwarefoundation.org/training/curriculum>

## Exit (light) Codes

### Overview

Teaching: 10 min

Exercises: 10 min

#### Questions

- What is an exit code?

#### Objectives

- Understand exit codes
- How to print exit codes
- How to set exit codes in a script
- How to ignore exit codes
- Create a script that terminates in success/error

As we enter the first episode of the Continuous Integration / Continuous Deployment (CI/CD) session, we learn how to exit.

### Start by Exiting

How does a general task know whether or not a script finished correctly or not? You could parse (`grep`) the output:

#### Code

```
> ls nonexistent-file
```

#### Output

```
ls: cannot access 'nonexistent-file': No such file or directory
```

But every command outputs something differently. Instead, scripts also have an (invisible) exit code:

#### Code

```
> ls nonexistent-file  
> echo $?
```

<https://hepsoftwarefoundation.org/training/curriculum>



## Beginner level

Module	Description	Status
The Unix Shell		<a href="#">Created &amp; Maintained by software carpentries</a>
Advanced Shell	E.g. SSH	
Version controlling with git		<a href="#">Created &amp; Maintained by software carpentries</a>
Advanced git		
Programming with python		<a href="#">Created &amp; Maintained by software carpentries</a>
Build systems: From <code>gcc</code> to <code>cmake</code>		
Basic <code>C++</code>		
Distributed file systems and grid computing		
<code>ROOT</code>		
A simple analysis	A simple analysis using CMS open data	<a href="#">Created by S. Wunsch; maintained by HSF</a>

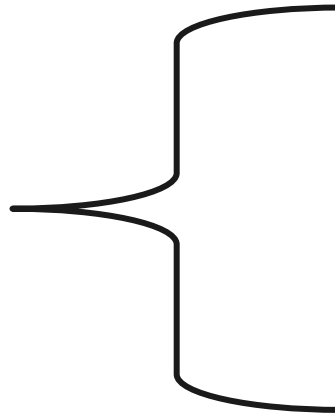


# Status

- Current modules:
  - Bash, Python, git (from SWC)
  - Simple CMS open data analysis (from S. Wunsch, maintained at [gh/hsf-training](#))
  - CI/CD (from G. Stark, maintained at [gh/hsf-training](#))
  - Docker (from M. Feickert)
  - Unit testing (from K. Huff)
- Want to expand number of modules, especially for HEP specific topics

# Virtual Analysis Preservation Training

3 (full) days in February @ CERN



[1] GitLab Pipelines - June

[2] Docker - July

[3] Workflows/REANA - August

<https://indico.cern.ch/event/854880/>

# Virtual Pipelines Training - June 2-4

<https://indico.cern.ch/event/904759>

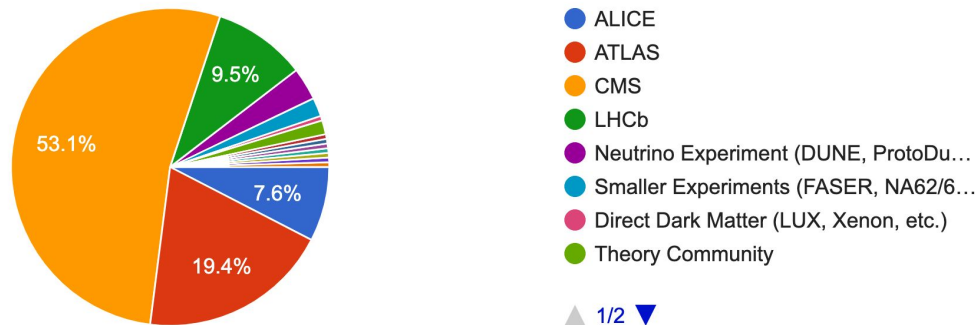
- June 2 @ 15h00 : Kickoff with description of virtual workshop logistics
- June 3 : Individual work
  - Standalone lessons - - [CMS OpenData “payload”](#) and [CI/CD Pipelines](#)
    - With videos recorded on trainings - [CMS OpenData “payload”](#) and [CI/CD Pipelines](#)
  - Supported via use of Mattermost ([Awesome WS](#))
    - Success! Good interaction, some issues with registering non-CERN folk.
  - Supported via Google doc “FAQ” page ([Link to Page](#))
    - Kind of a flop - unclear if it was concept or implementation.
- June 4 @ 2 hour blocks from 8h00-22h00
  - Gather one mentor with 5 participants - [arranged ahead of time here](#)
    - 30 people signed up per slot → only 10-15 showed up (need to find a way to get more commitment)
  - Using “breakout rooms”
    - In retrospect this was not the best solution - required constant host moderation (@LeonoraVesterbacka)

# Who attended?

- 250 people registered (most within one week of opening)
  - ~40 people did \*NOT\* have CERN computing accounts → gitlab.com works fine for this subject matter
- Retention
  - 200 had some participation - based on pre-survey and June 4 signup (“I got busy at the last minute”)
  - 100 watched videos - based on you-tube views (“Not enough time between start and hands-on”)
  - <100 attended the Hands-On sessions (“I had a last minute meeting” / “I did not have time to finish the videos”)

## What is your experimental collaboration?

211 responses



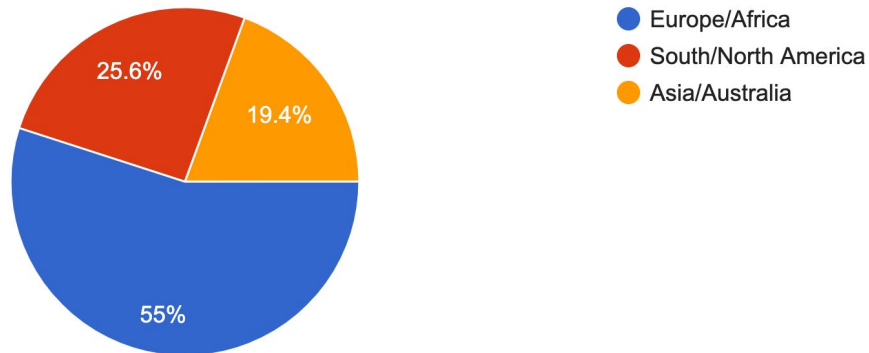


# Who attended?

- 250 people registered (most within one week of opening)
  - 200 had some participation - based on pre-survey and June 4 signup (“I got busy at the last minute”)
  - 100 watched videos - based on you-tube views (“Not enough time between start and hands-on”)
  - <100 attended the Hands-On sessions (“I had a last minute meeting” / “I did not have time to finish the videos”)

## Physical location

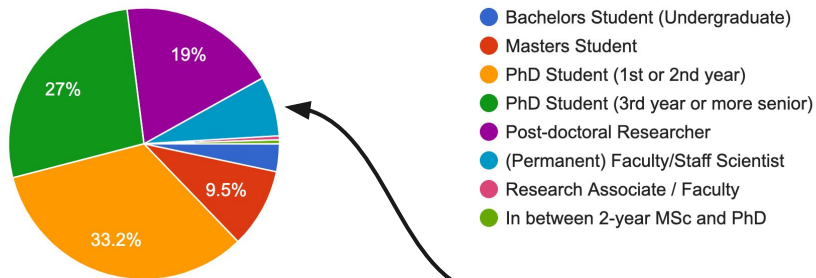
211 responses



# Who attended?

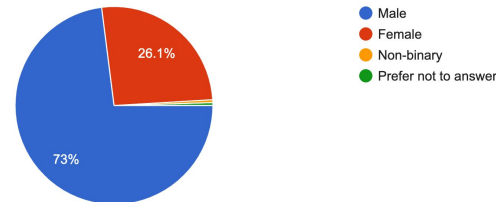
- Demographically more diverse than previous trainings (in my experience)
- Mentors ([Link to Bios](#)) : Mostly male (~80%), one URM
  - This is important because teachers can serve as role models

Academic level  
211 responses

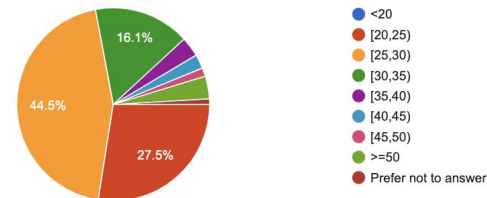


**15 “senior folk” → unique issues and role in our broader community : Sam/Lukas would like to hold a “VIP training”**

Gender  
211 responses



Age [years]  
211 responses

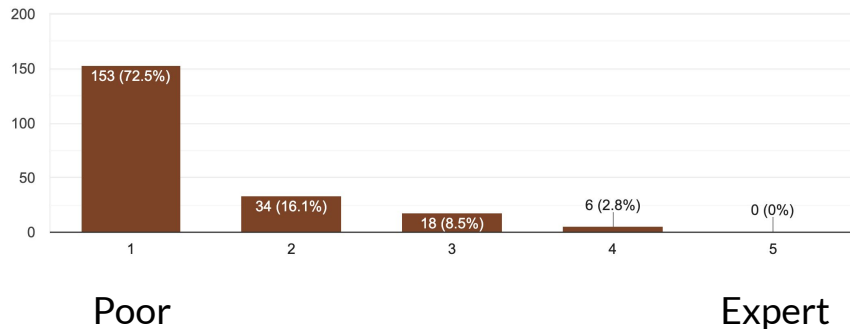


# What did they know?



- Individuals who attend are self-selecting novices
  - They don't know what we will teach
- and \*want\* to learn
  - They appreciate the importance of the topic (not just "curious")

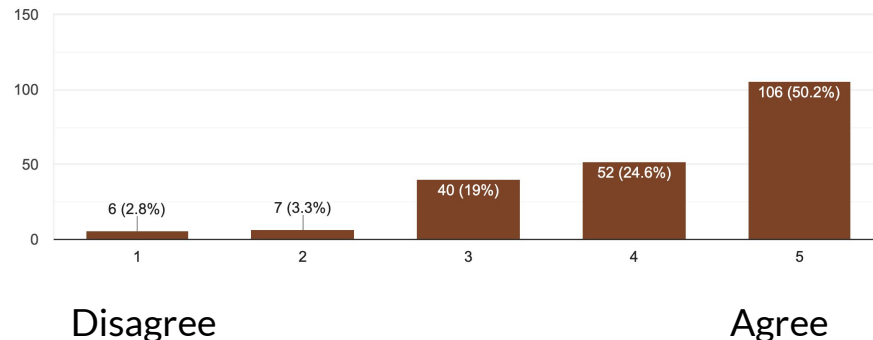
How advanced is your knowledge and abilities when using GitLab Pipelines?  
211 responses



I have used continuous integration and automated testing when developing code in the past.  
211 responses



Automated tests are a good thing to develop for your code.  
211 responses



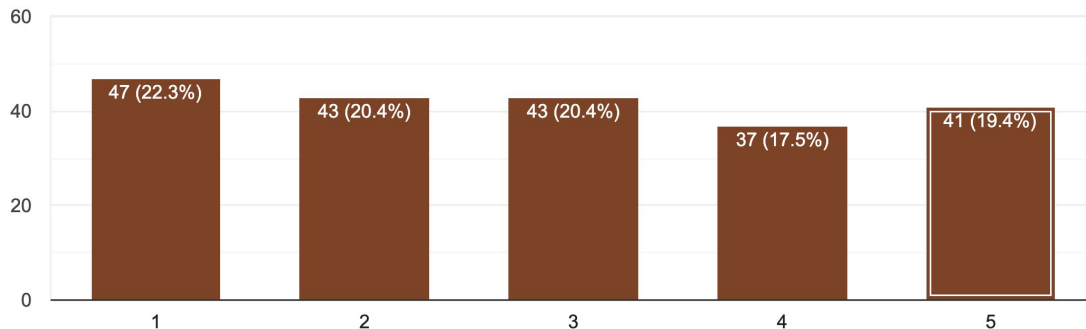
# The reach of virtual



- Two-fold increase of individuals who were able to “attend”

How likely would you have been to attend this bootcamp/workshop had it been held in person at CERN with no external connection?

211 responses

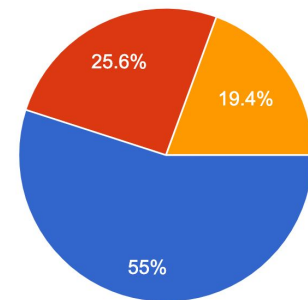


I definitely would **not** have come

I definitely would have come

Physical location

211 responses



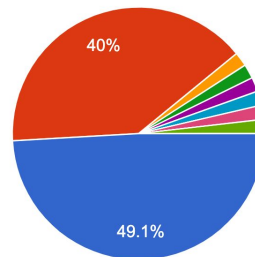
- Europe/Africa
- South/North America
- Asia/Australia

# Were captions worth it?

- Yes
  - Very few of the individuals in attendance explicitly brought deaf/HoH to our attention - based on registration
  - Videos persist → larger reach
- Cost: ~\$1/minute
  - 5 hours for 300 USD
  - Technical quality “great” (Giordon Stark)
  - [www.rev.com](http://www.rev.com)
- LHCP is in the process of working to improve auto-captioning for HEP
  - More details here - [snowmass21](#)
  - Contact Giovanni Marchiori <[Giovanni.Marchiori@cern.ch](mailto:Giovanni.Marchiori@cern.ch)> if you are interested in contributing
- Suggestion: funding requests from supporting agencies should include budget for captioning and related things to be more inclusive

This workshop would not have been as successful without captioned videos. [Read carefully to be sure you understand the double negative.]

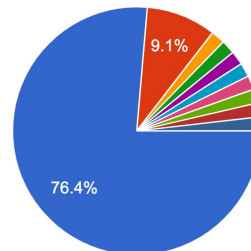
55 responses



- True - the workshop would \*not\* have been successful without captions
- False - the workshop could have been successful without captions
- I'm not sure it wouldn't be successful,...
- the workshop could have been succes...
- I haven't gone through the material.
- Again for me personally, didn't really...
- Not sure the videos are crucial (I most...
- As I said, I didn't notice them, so from...

I appreciated having captions on the videos.

55 responses



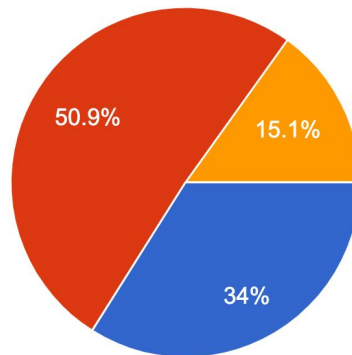
- True
- False
- It was nice to have the captions but th...
- didn't use it
- I did not read the captions but it is alw...
- I'd rather have the full text.
- I haven't gone through the material.
- Didn't need/use them

# Building a Community

- We've got between [10-30] new "mentors" to help with training in the future

Would you be interested in us reaching out to you to help teach/mentor at an event like this in the future?

53 responses



- No thank you.
- Yes, potentially, but I think I would need help learning more and am not sure if I am ready yet.
- Yes, definitely. Let's do this!

# Upcoming Trainings



- July : Virtual Docker
- August : Virtual REANA
  - Need to start talking with Tibor and IT folks
- In “development” : GPU+ML
  - Allesandro Forti @ Manchester
  - Enough people willing to tutor and develop material
  - How to make GPU’s persistently accessible?
    - Could use : <https://www.atlas-ml.org/> or <https://colab.research.google.com/>
- Reminder : Wishlist of trainings here - [Link to Blueprint Plan](#)
  - Keen on developing material for one? Know someone who is keen?



**Thank you.**

