The CMS Open Data workshop: Introduction

August 1st, 2022

Kati Lassila-Perini
CMS Data preservation and open access coordinator
Helsinki Institute of Physics (Finland)

Welcome!

On behalf of the CMS Open data team

The organizing team of this workshop:

Matt

Edgar

Tom

Kati

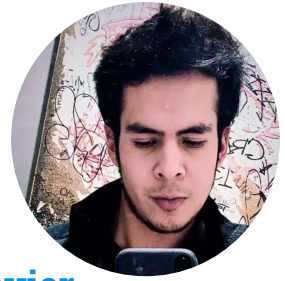
Clemens

Julie

Sezen

Achim

Jieun



Xavier

Dietrich

Kalin

Kharol

Andres

Sarah

Nitish

Pamela

The tutorial team 💙

Organizers and speakers

Edgar Carrera Jarrin (U. San Francisco de Quito, EC)

Julie Hogan ((Bethel University, Brown University, US)

Matthew Bellis (Siena College, US)

Kati Lassila-Perini (Helsinki Institute of Physics, FI)

Clemens Lange (PSI)

Jieun Yoo (UIC, US)

Achim Geiser (DESY)

Thomas McCauley (University of Notre Dame, US)

Xavier Alexander Tintin Gavilanes (Escuela Politecnica Nacional (EC))

Sezen Sekmen (Kyungpook National University, KR)

Helpers and contributing

Sarah Markham (Siena College, US)

Kalin Johnson (Brown University, US)

Dietrich Liko (Austrian Academy of Sciences, AT)

Kharol Chicaiza (EPN, EC)

Andres Luis Chicaiza (EPN, EC)

Pamela LLerena (EPN, EC)

Nitsih Dhringa (Panjab University, IN)

Daniela Merizalde (USFQ, EC)

Jose Ochoa (USFQ, EC)

Edison Patricio Jimenez (EPN, EC)

Enrique Piedra (EPN, EC)

David Mena (EPN, EC)

1.

Goals?

What do you expect?
What do we expect?

We made some assumptions

We think that you want to use CMS open data and simulation for physics research.

Therefore, we think you want to understand:

- the basic physics object usage (object access, id, corrections, how to write them out)
- how one can select events and access trigger information
- how to evaluate the luminosity
- the possibilities for large-scale data processing.

In addition, we think you will be interested in

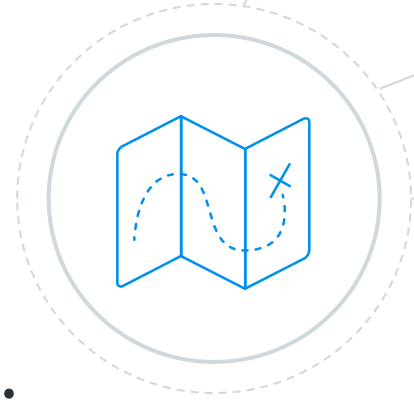
- how to put this all together in an analysis

But that's not all - we get something as well

We want to:

- build a community of users
- remind of https://opendata-forum.cern.ch/
- get understanding of the usage patterns and needs
- get feedback of what is missing in the documentation and tutorial material
- build a proper CMS open data user guide.

Ambitious goals ➟
Do we reach them?



Bear with us:

CMS Open data is always work in progress

2.

How to get there?

Workshop structure
Working methods

A set of mandatory pre-exercises

Pre-exercises

(**Mandatory exercises must be completed before the start of the workshop**)

(**Submission of work assignments is required as explained in the Orientation below**)

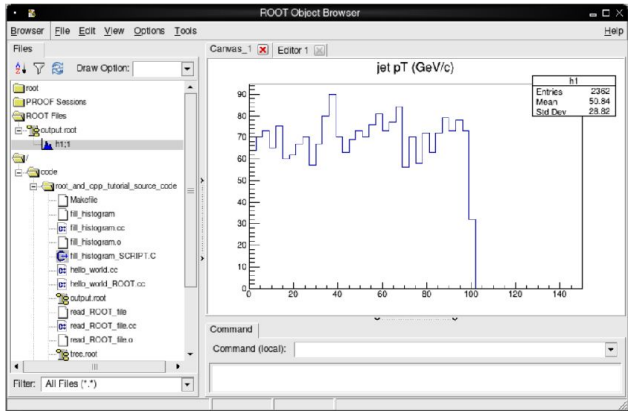
Mandatory 5 min	Orientation
Optional (external lesson)	The Unix Shell
Optional (external lesson)	Version Control with Git
Optional (external lesson)	Programming with Python
Mandatory 2h	Docker containers
Mandatory 2h	Dataset scouting
Mandatory 4h	ROOT with C++ and Python
Mandatory 2h	Intro to CMSSW
Mandatory 2h	Intro to cloud computing

Pre-exercises

- Importantly, to set up and test your working environment before the workshop
- To give some background information on the topics of this workshop
  - Many thanks for your questions!

- If these are not familiar to you, you're in trouble!

```
Setting up CMSSW_7_6_7
WARNING: There already exists /code/CMSSW_7_6_7 area for SCRAM_ARCH slc6_amd64_gcc493.
CMSSW should now be available.
This is a standalone image for CMSSW_7_6_7 slc6_amd64_gcc493.
(/code/CMSSW_7_6_7/src) ls
```



- In that case: try to catch up, we do our best to guide you through, but the priority is given to the ongoing exercises.

Schedule

Monday

14:30-14:50	Welcome and Intro	K. Lassila-Perini
14:50-15:50	Physics Objects: Intro and POET	M. Bellis, E. Carrera, K. Lassila-Perini
15:50-16:30	Physics Objects: Electrons	M. Bellis, E. Carrera, K. Lassila-Perini
16:30-17:00	Break	
17:00-17:40	Physics Objects: Muons	M. Bellis, E. Carrera, K. Lassila-Perini
17:40-18:30	Physics Objects: Jets	M. Bellis, E. Carrera, J. Hogan, K. Lassila-Perini

Tuesday

14:30-15:30	Trigger	E. Carrera
15:30-16:30	Luminosity	J. Yoo
16:30-17:00	Break	
17:00-18:30	Analysis example with Run 1 data	A. Geiser

Wednesday

14:30-15:10	Simplified Run 2 analysis: Intro	E. Carrera, T. McCauley
15:10-16:30	Simplified Run 2 analysis: Analysis chain	E. Carrera, T. McCauley
16:30-17:00	Break	
17:00-18:30	Simplified Run 2 analysis: Systematics and Stats	E. Carrera, T. McCauley

Thursday

14:30-16:30	Cloud Computing	C. Lange, K. Lassila-Perini, X. Tintin
16:30-17:00	Break	
17:00-18:30	Run 2 analysis with ADL	S. Sekmen
18:30-18:45	Wrap-up and feedback	

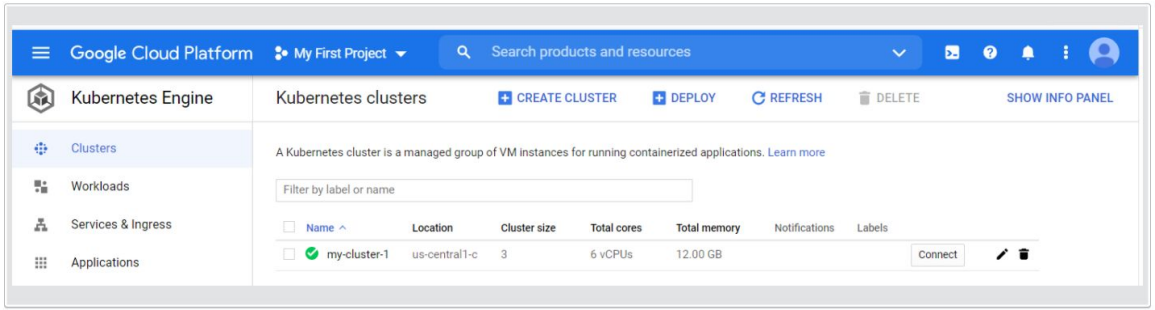
Full 4 days of work ahead of us!

Material available from the schedule and from the indico agenda
Each lesson has a dedicated Mattermost channel in CMSODWS2022

CMS analysis on a cloud environment

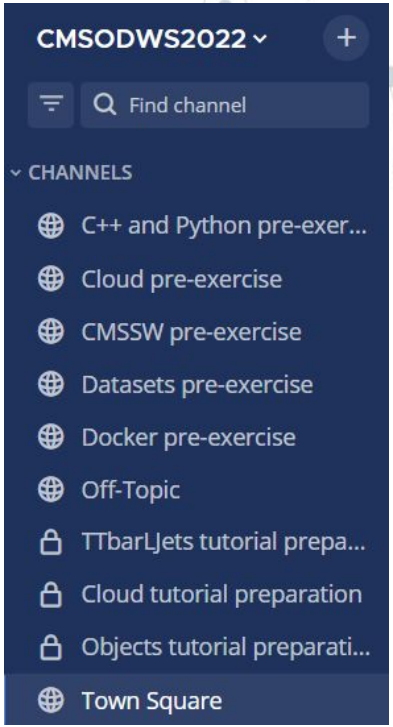
- You will have the opportunity to learn how to run a CMS open data processing job in real scale on commercial cloud environment.
- If you want to participate, let us know in the survey at the latest on Tuesday.
- It will be hands-on and you will get a temporary account
  - We've got resources for it through CERN IT projects.
- We'll be using Kubernetes engine on Google Cloud Platform

- Don't miss it!



Getting help - live

- In mattermost, choose the channel corresponding to the lesson.
- Do not hesitate to ask!
  - But check if the same question has already been asked.
- Cut and paste the command and the error message
  - If needed, use `some code in line`
  - or ```block of code or output```
  - shift-return for a line break in a message
- Reload the tutorial page every now and then for updates.
- During live lessons
  - In the meeting room, use the mic.
  - In zoom:
    - Quick remarks (e.g. zoom problems) in zoom chat
    - Use “Raise hand” for voice questions



Getting help - live

- You are many we are really happy about that.
  - 130 registrations
  - 38 of you have already gone through the pre-exercises.
- We are there to help you!

- Please read the instructions carefully
  - We are not professional technical writers and some details may not be that well exposed in the text.
  - Suggestions for improvements are most welcome.

Ask! Ask! Ask!

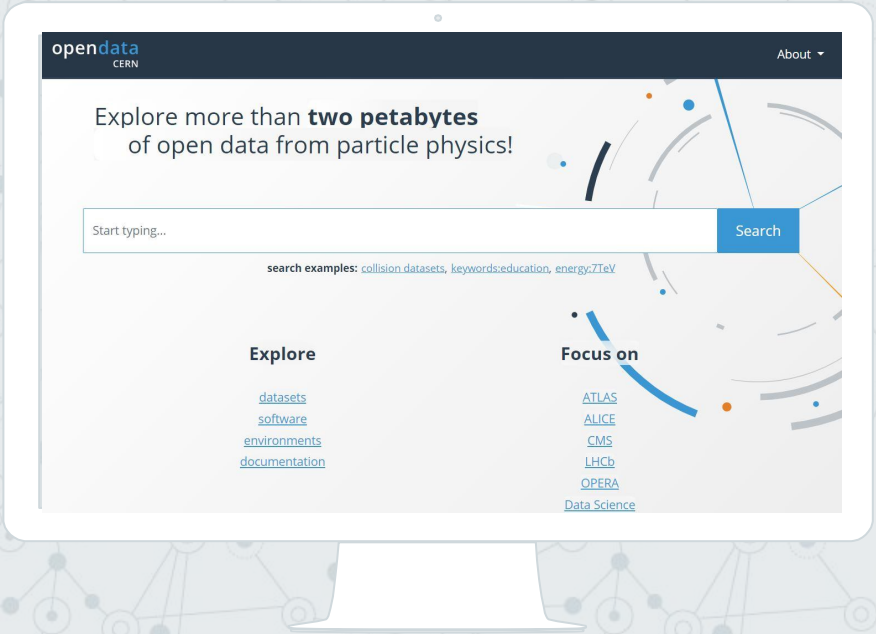
The excellence in teaching is not pushing forward those who know the most,

but taking care that no one is left behind.

3.

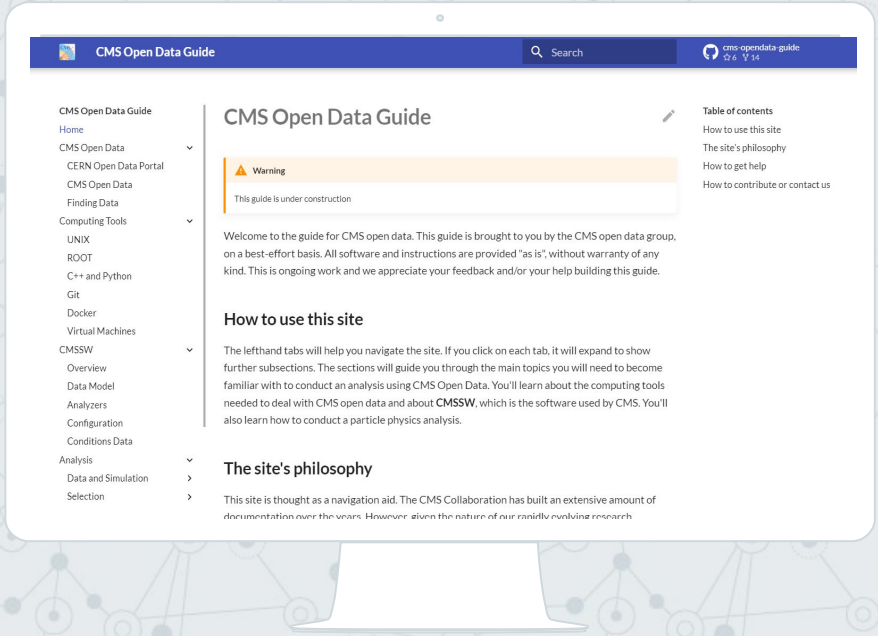
How to get help after?

Information sources
Communication



CERN Open data portal

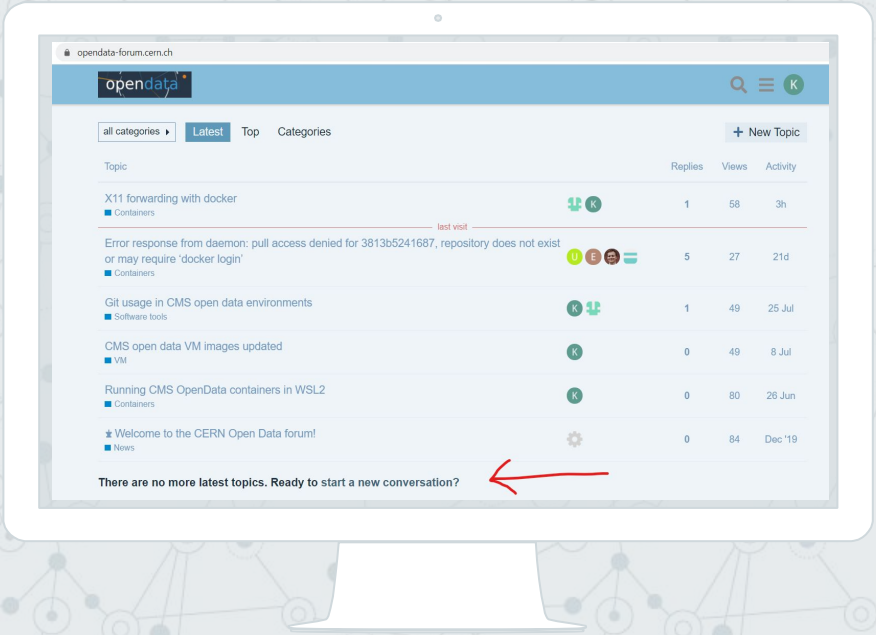
Serves the data, associated analysis artefacts, usage examples



CMS Open data guide

Work in progress, will be completed with the material in this tutorial.

Do you want to help?



CERN Open data forum

Feel free to post questions! Feel free to reply as well!

Most frequently asked questions at this workshop will be added.

Other sources of information

- Open data portal support mail: opendata-support@cern.ch
  - Technical issues
  - Questions to limited audience
- CMS WorkBook and SWGuide
  - Careful: instructions might not correspond to the CMSSW version needed for open data
- CMSSW source code
  - Keep in mind the versioning,
    - for 2011-2012 open data use CMSSW_5_3_X as tag.
    - for 2015 data use CMSSW_7_6_X as tag.

4.

Now, let’s get to work!

Enjoy the workshop!
We’ll love to hear feedback from you on Thursday.

Thanks!

Any questions?

Find us in mattermost

Credits

Thanks to my colleagues

- in the DPOA group in CMS
  - all organizers and contributors
- in the CERN Data preservation services
  - CERN Open data portal team, and many other services that we rely on

And great thanks to all CMS open data users!

And thanks to SlidesCarnival for this free presentation template