

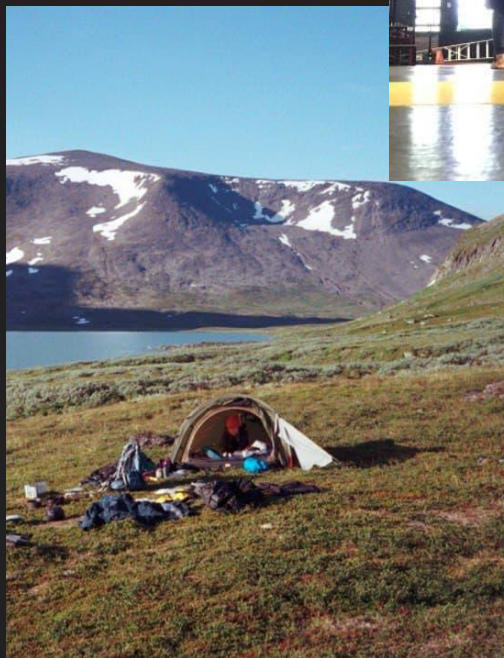
# CMS OPEN DATA

Workshop hosted by  
Roxane Theriault and Veera Juntunen

1. Introduction – about us

# Veera Juntunen

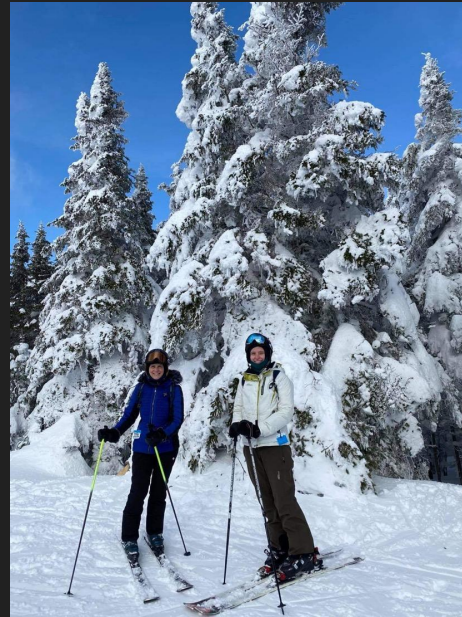
Summer trainee at HIP / CERN  
MSc student, molecular & atmospheric physics  
Finland, Oulu



1. Introduction – about us

# Roxane Theriault

CERN summer student  
MSc. student starting this fall  
Quebec, Canada



# Schedule

10 min	Introduction
10 min	What is open data
10 min	Open data from CMS
10 min	Open data in high school education
30 min	Examples
40 min	Hands on
10 min	Wrap up

## 1. Introduction

# Use of Open Data from the CMS Experiment - Summer Student Workshop

6 August 2021

Europe/Zurich timezone



Overview

Registration

Participant List

Videoconference

Contact

✉ [summer.student.info@c...](mailto:summer.student.info@c...)

✉ [NMS.SummerStudent@...](mailto:NMS.SummerStudent@...)

In this workshop, we will be presenting ways to work with open data, in particular open data released from the CMS experiment in teaching at a high school level. Among other topics, we will discuss teaching applications and show an example using the Higgs boson. This will be done using Jupyter Notebooks, but no prior knowledge of them or Python is required.

**Registration:** Though not mandatory, participants are highly encouraged to register for the event. Note that there is no limit for registrations, all students are welcome to join the virtual workshop.



**Starts** 6 Aug 2021, 15:00

**Ends** 6 Aug 2021, 17:00

Europe/Zurich



Roxane Theriault  
Veera Emilia Juntunen



Hands-on materials

- 🔗 ADL-exercises
- 🔗 Animations example
- 🔗 Creating sound from data
- 🔗 Normfit Transverse Momentum+Pseudo...
- 🔗 Overlaid histograms
- 🔗 Shakespeare
- 🔗 Statistics
- 🔗 Text classification



Zoom-link and materials

- 🔗 Feedback
- 🔗 Zoom-link

# Who are you?

## 2. What is open data?

# What is open data?

- Data provided free of charge and available to all
- Usually in a format like csv or xml
- Important to make the distinction between data and information



## 2. What is open data?

# Why do we need it?

- To preserve data and make it available to a wider audience
- Increases reproducibility of research and promotes discoveries
- Prevents duplication and loss of research
- Promotes collaboration
- Helps students learn data analysis skills
- Can help boost performance<sup>1</sup>

<sup>1</sup>Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415.



## 2. What is open data?

# Where can we find it?

- Google “open data” or “open data portal”
- Sites like the CERN Open Data Portal and World Bank
- Always question the legitimacy of the source and data
  - How reliable is the information from different institutes or states?
  - Is it honestly and neutrally presented, or chosen to support a particular agenda?

## 2. What is open data?

# Where can we find it?

Gapminder: <https://www.gapminder.org/data/>

World Bank: <https://databank.worldbank.org/home>

CERN Open Data Portal: <http://opendata.cern.ch/>

Zenodo: <https://zenodo.org/>

Our World in Data: <https://ourworldindata.org/>

Solar Influence Data Center: <http://sidc.oma.be/silso/datafiles>

Awesome Public Datasets:

<https://github.com/awesomedata/awesome-public-datasets>

### 3. Open data from CMS

# CMS aka the Compact Muon Solenoid

LHC is 27 km long particle accelerator, collisions at 4 locations around its ring. At one of these points is CMS, designed to observe any new physics phenomena from LHC.

#### Facts and stats about CMS:

- 14 000-tonne detector
- 15 x 21 meters
- It has the most powerful solenoid magnet ever made

#### Its jobs:

- Bend particles
- Identify tracks
- Measure energy
- Detect Muons

### 3. Open data from CMS

# CMS aka the Compact Muon Solenoid

What they found:  
Higgs boson

Undiscovered questions:  
What is the universe made of?  
Do we really live in only 3D?  
How did matter form?  
What and where is the antimatter?  
Are there more particles left to find?



### 3. Open data from CMS

# CERN open data portal

<http://opendata.cern.ch/>

- Needed software and documentation
- 4 level system from LHC
- Citable
- Education & research

Level 1: more info on published results

Level 2: simplified data formats for outreach and analysis training

Level 3: reconstructed collision data and simulated data with software

Level 4: basic raw data with software

## 4. Open data in education

# CMS open data in high school education

For educational use, CMS releases data in three different levels:

Beginner – visualise collisions

Intermediate – Make histograms with collision data

Advanced – Dive deeper in the data

In our case we are interested in csv-files, which have been extracted from the original datasets.

## 4. Open data in education

# Making material

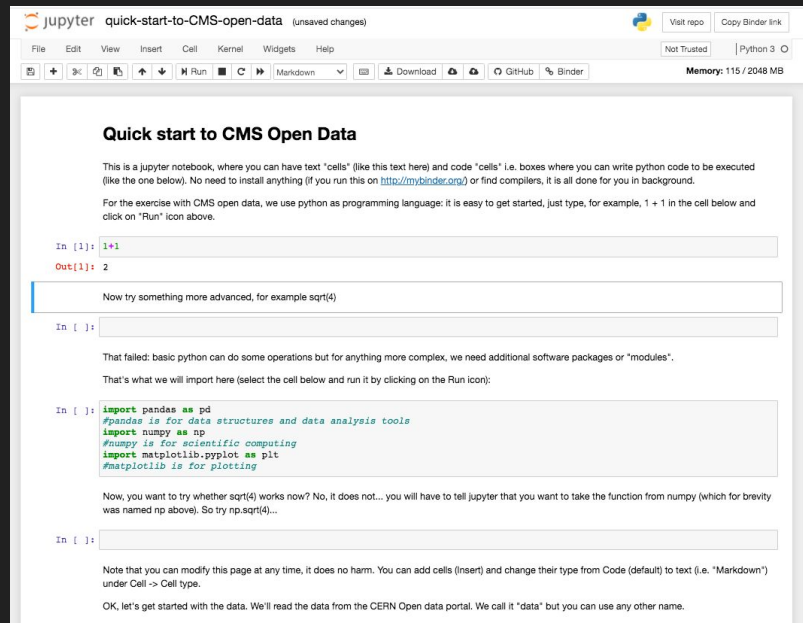
- Ready made material sets
  - High school curricula
  - .CSV
  - Different subjects, beginner / advanced, short / long
  - Jupyter Notebooks / Anaconda Navigator and Python
- Courses for teachers

## 4. Open data in education

# Jupyter Notebook

Interactive environment that allows you to create documents with code, equations, visualizations and text.

Can be used in online → perfect for educational use!



The screenshot shows a Jupyter Notebook interface with the title "quick-start-to-CMS-open-data (unsaved changes)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a status bar showing "Memory: 115 / 2048 MB".

The notebook content is titled "Quick start to CMS Open Data" and contains the following text and code:

This is a jupyter notebook, where you can have text "cells" (like this text here) and code "cells" i.e. boxes where you can write python code to be executed (like the one below). No need to install anything (if you run this on <http://mybinder.org>) or find compilers, it is all done for you in background.

For the exercise with CMS open data, we use python as programming language: it is easy to get started, just type, for example, `1 + 1` in the cell below and click on "Run" icon above.

```
In [1]: 1+1
```

Out[1]: 2

Now try something more advanced, for example `sqrt(4)`

```
In [ ]:
```

That failed: basic python can do some operations but for anything more complex, we need additional software packages or "modules".

That's what we will import here (select the cell below and run it by clicking on the Run icon):

```
In [ ]: import pandas as pd
#pandas is for data structures and data analysis tools
import numpy as np
#numpy is for scientific computing
import matplotlib.pyplot as plt
#matplotlib is for plotting
```

Now, you want to try whether `sqrt(4)` works now? No, it does not... you will have to tell jupyter that you want to take the function from numpy (which for brevity was named `np` above). So try `np.sqrt(4)`...

```
In [ ]:
```

Note that you can modify this page at any time, it does no harm. You can add cells (insert) and change their type from Code (default) to text (i.e. "Markdown") under Cell -> Cell type.

OK, let's get started with the data. We'll read the data from the CERN Open data portal. We call it "data" but you can use any other name.



## 4. Open data in education

# Sharing material

- GitHub
- MyBinder
  - Open and execute your Notebooks from GitHub
  - Students don't have to download any software
- Google Colab
  - Google account needed

## 4. Open data in education

# Sharing material



### Build and launch a repository

GitHub repository name or URL

GitHub ▾

GitHub repository name or URL

Git ref (branch, tag, or commit)

HEAD

Path to a notebook file (optional)

Path to a notebook file (optional)

File ▾

launch

Copy the URL below and share your Binder with others:

Fill in the fields to see a URL for sharing your Binder.



Expand to see the text below, paste it into your README to show a binder badge:  launch binder



Kahoot!

## 4. Examples

# Jupyter Notebooks

## 4. Examples

# Finding the Higgs Boson

## 4. Examples

# Questions about the examples

## 5. Exercises

# Exercises

<b>Name</b>	<b>Topics</b>	<b>Difficulty</b>
Overlaid Histograms	Particle physics, plotting	Basic
Statistics	Particle physics, statistics	Basic
Animations example	Particle physics, animations	Intermediate
Shakespeare	Literature, word processing	Intermediate
Creating Sound from Data	Particle physics, waves	Intermediate
Text Classification	Word processing	Advanced
Normfit Transverse Momentum+Pseudorapidity	Particle physics, plotting, fitting	Advanced

# Exercises with ADL

ADL-exercises → binder → CMS-OD-<example name>.ipynp

Name	Topics	Difficulty
CMS-OD-ZZ4L	Simple ZZ → 4L analysis	Basic
CMS-OD-HZZ4L	Simple H → ZZ → 4L analysis	Intermediate
CMS-OD-12350-Htautau	H → tautau analysis	Advanced



## 6. Conclusion

# Thank you for joining us!

Questions? Comments? Feedback?

Contact info:    [roxane.theriault2@mail.mcgill.ca](mailto:roxane.theriault2@mail.mcgill.ca)  
                      [veera.emilia.juntunen@cern.ch](mailto:veera.emilia.juntunen@cern.ch)