



# Data Analysis with Real Data

CMS open data with jupyter notebook

# What is Data?

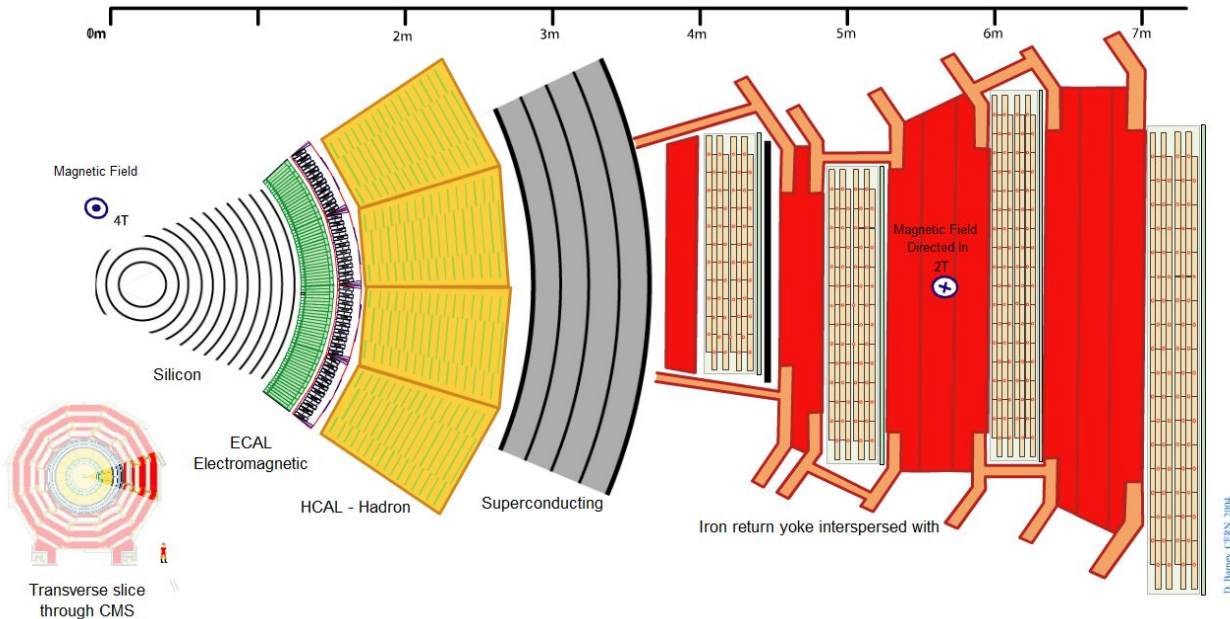
- Data collected from a particle detector for high energy particle physics
- In fact the format of data is highly correlated with the setup of the experiment and the target physics
- It is highly recommended that you understand the experiment!

# Particles and their detection

- What particles can we detect:
  - **Electron**: A stable elementary particle belonging to the *fermion* family of particles. It has an electrical charge of -1, while its antiparticle **positron** has an electrical charge of +1. An electron has a mass of approximately  $0.5 \text{ MeV } /c^2$ .
  - **Photon**: A stable elementary particle belonging to the *boson* family of particles. A photon is massless with no electrical charge. It is the carrier of the electromagnetic force. It's represented by Greek letter  $\gamma$  (gamma).
  - **Hadron**: A “heavy” composite particle made of two or more quarks. They can carry a charge or be neutral. For example, protons and neutrons belong to this category.
  - **Muon**: An elementary particle. It has an electrical charge of -1. Muon is a lepton with properties that are similar to those of an electron but 200 times more mass. It is represented by Greek letter  $\mu$  (mu). Muon's antiparticle **antimuon** has an electrical charge of +1.

# A little about CMS

- How can we detect particles?
  - There are different kinds of particle detectors but the basic principles are the same.
  - The Compact Muon Solenoid (CMS) is a detector that uses a huge magnet to bend the paths of particles produced in proton-proton collisions.

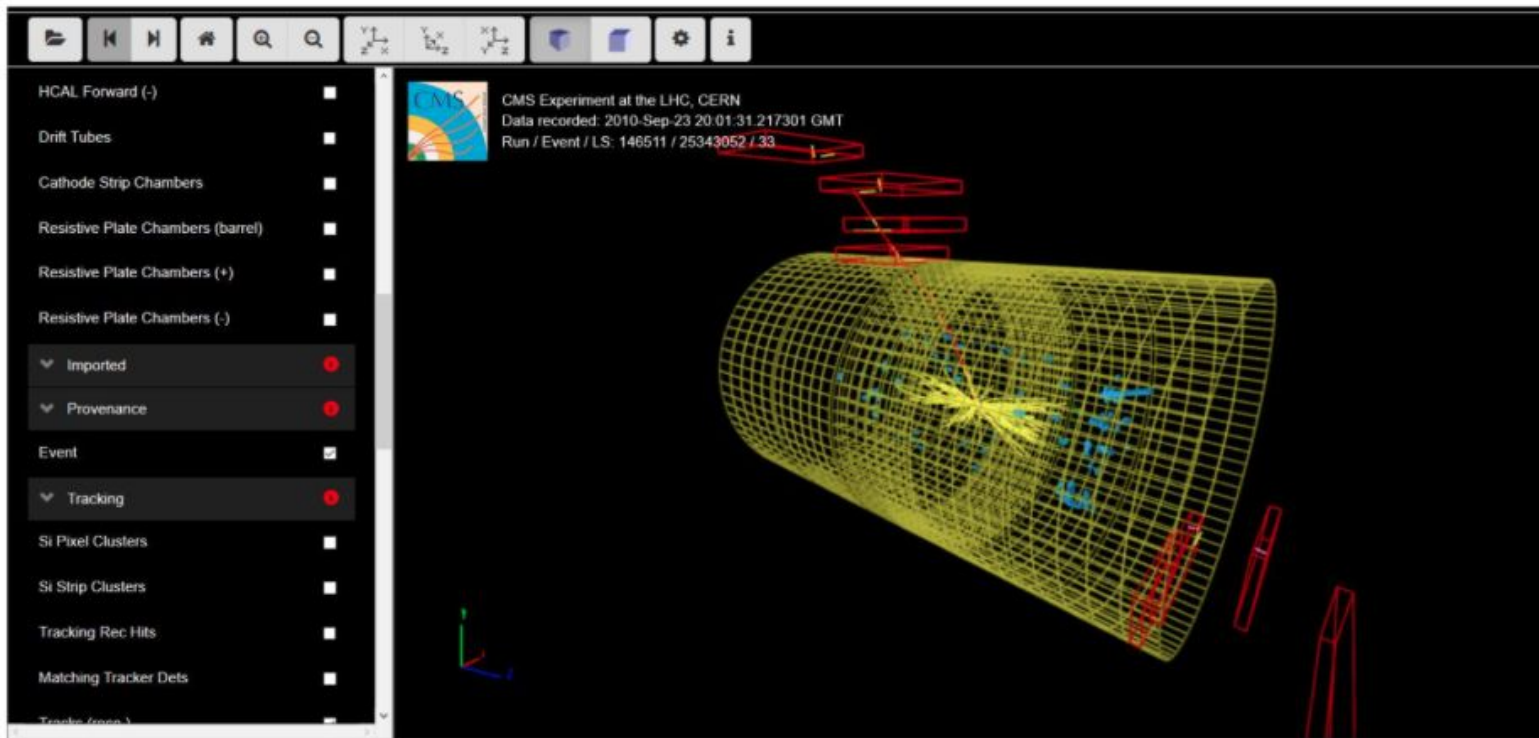


# CERN Open Data

- CERN has a project to open its own data to general public. One can use the data to do research or education
- You may find some more details here: <http://opendata.cern.ch>
- The data from major four experiments are available online. You may be able to download them, together with the paired virtual machine (for the corresponding software, for example CMSSW) for user to perform their own analysis at home.
- CMS Open Data in the form of CSV files can be used in programming applications such as Jupyter notebooks.

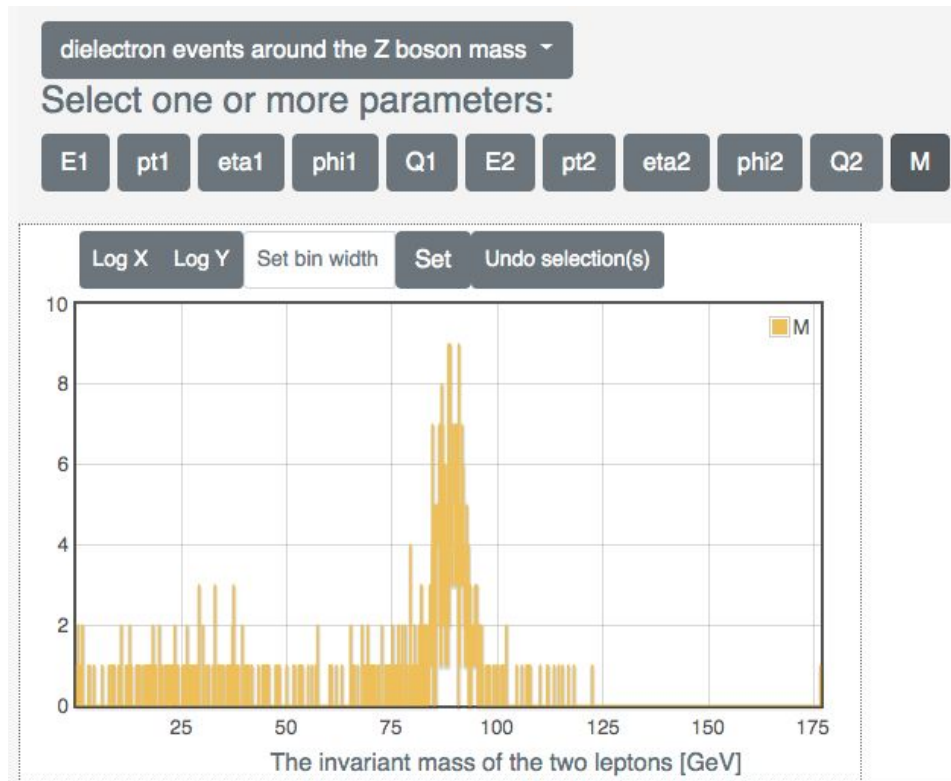
# First play: Look into CMS data online

- There is an online version (=simplified) of CMS event display available.
  - See: <http://opendata.cern.ch/visualise/events/cms>



# Second play: Look into the already prepared histograms

- Some of the typical histograms prepared with CMS data are available online as well: <http://opendata.cern.ch/visualise/histograms>



# Online Open Data Analysis

- You can explore CERN open datasets and physics analyses in the form of CSV files can be used in programming applications such as Jupyter notebooks
  - CMS: <https://github.com/cms-opendata-education/cms-jupyter-materials-english>
  - Atlas: <https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata>
- For today: <https://github.com/ssilvado/notebooks-cms-opendata>



# Exercises

You can find the exercises in the end of the notebook:

<https://github.com/ssilvado/notebooks-cms-opendata/blob/master/Exercises-with-open-data/7-Pseudorapidity-resolution.ipynb>