

ATLAS Open Data: Using Open Education Resources Effectively

Even S. Haaland

On behalf of the ATLAS Collaboration

University of Oslo (UiO)

EPS-HEP Conference 2019 (Ghent, Belgium)



The ATLAS Open Data project

Purposes:

- Provide public access to real and simulated data from the ATLAS experiment.
- Develop and produce educational tools, such as analysis code examples.

Target audiences:

- Mainly physics students at undergraduate and master level.
- But of course also anyone else that are interested!

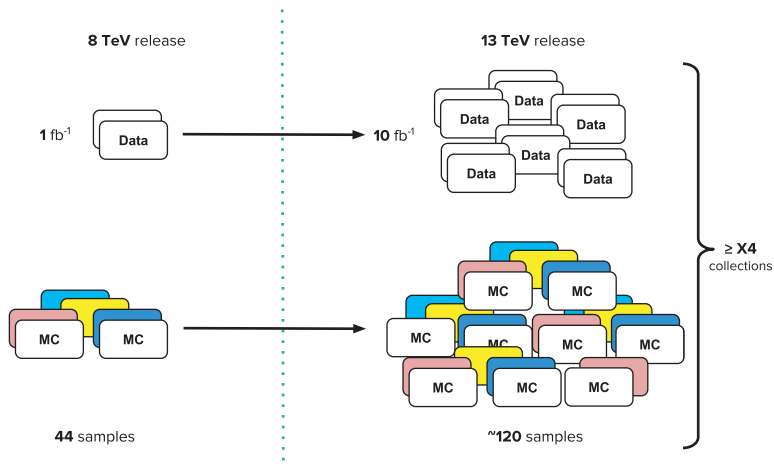
Aim:

- For universities to use ATLAS Open Data when teaching high energy physics or data analysis techniques.

ATLAS Open Data releases

Previous release (2016)

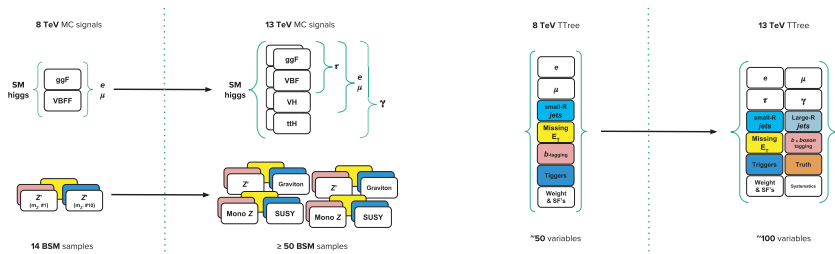
New release (2019)



Data format: **ROOT nTuples**

Going from 8 TeV to 13 TeV

Increasing the possibilities and flexibility of ATLAS Open Data by adding more data, more MC samples and more variables to the release.



Accessing ATLAS Open Data and Tools

Open Data portals (CERN and ATLAS):

<http://opendata.cern.ch>

<http://opendata.atlas.cern>

On the data portal you will find:

- Data samples
- Software and Jupyter Notebooks (C++/Python)
- JavaScript applications
- Virtual Machines (VMs)
- GitHub/GitLab repositories

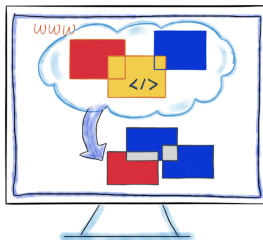
ATLAS Open Data portal

Data & Tools Repository

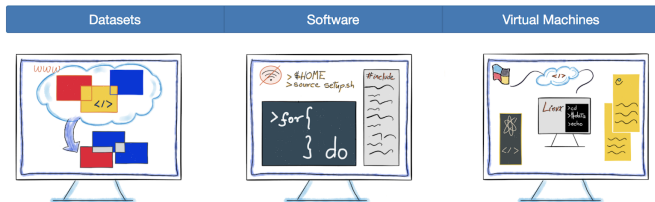
Here you have in a single place all the necessary pieces in order to start you physics analysis in a more complete way. Look into the data like an ATLAS particle physicist!

In this section, you can find where to download:

- The complete collection of available datasets
- The different analysis software
- The virtual machines to perform physics searches



Downloads



How can the ATLAS Open Data resources be used effectively?

- Motivation:

- ▶ Students should spend their time doing analysis, rather than fighting technical issues.
- ▶ The users should **not** be required to have computers with a specific operating system, very large storage capacity, “super-fast” internet, etc.

⇒ Need to provide solutions that ensure effective use of the resources.

- Projects/ideas:

- ▶ Data analysis frameworks in Python and C++, with example analyses
- ▶ Jupyter Notebooks (Python and C++)
- ▶ VMs (including as Servers)
- ▶ Electron.js applications
- ▶ C++ and Python hybrid framework

How can the ATLAS Open Data resources be used effectively?

- Motivation:

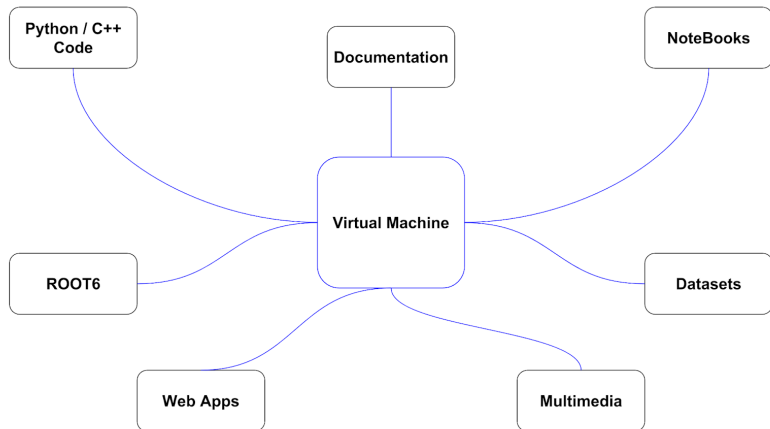
- ▶ Students should spend their time doing analysis, rather than fighting technical issues.
- ▶ The users should **not** be required to have computers with a specific operating system, very large storage capacity, “super-fast” internet, etc.

⇒ Need to provide solutions that ensure effective use of the resources.

- Projects/ideas:

- ▶ Data analysis frameworks in Python and C++, with example analyses
- ▶ Jupyter Notebooks (Python and C++)
- ▶ **VMs (including as Servers)**
- ▶ Electron.js applications
- ▶ **C++ and Python hybrid framework**

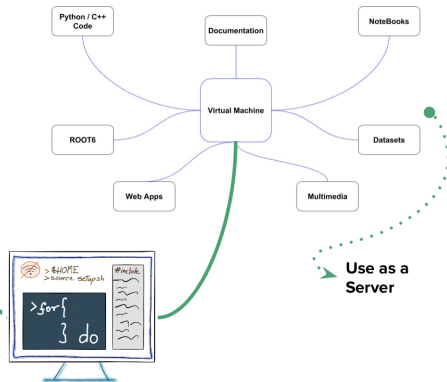
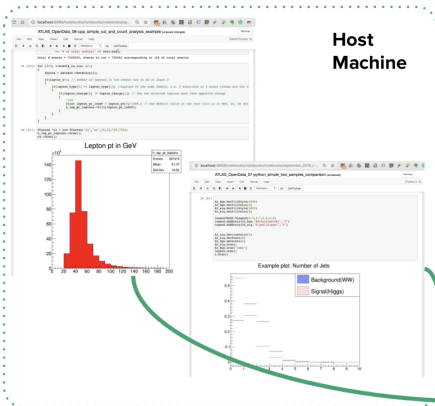
Anatomy of ATLAS Open Data Virtual Machines:



Virtual Machines as Servers

- New generation of VMs (2019), running only as servers: require less space and graphic resources.
- Apps (notebooks, terminal, etc.) are opened in host machine browser while running the VM.
- We will release VMs running Docker Containers able to execute JupyterLab from host machine's browser. (Summer student project)

Virtual Machines as Servers II



C++/Python Hybrid Framework

- **Idea:** making a framework that is a mixture of C++ and Python.
- Want to take advantage of the speed of C++ and the user-friendliness of Python.
- Back end code in C++: processing ntuples
- Front end code in Python: producing plots/results
- (Summer student project)

Summary & Outlook

- Release of ATLAS Open Data at 13 TeV **this summer!**
 - ▶ Dynamical release: software and samples will be updated/added continually.
- Want to make the resources easily accessible by providing:
 - ▶ Software frameworks and examples
 - ▶ Jupyter notebooks
 - ▶ VMs
 - ▶ Graphical applications

Stay tuned!