A Jupyter-based Interface Integrated with ROOT for ATLAS Open Data Analysis

Yixin WANG

Supervisors: Arturo Rodolfo SANCHEZ PINEDA, Leonid SERKIN

To introduce myself…

Name: Yixin Wang   **Country:** China

Where you study?
— Xi'an Jiaotong University, Xi'an

Field of Studies and Degree:
— Control Science and Technology, Master

What have you been doing at CERN?
— ATLAS Outreach Open Data & Tool group

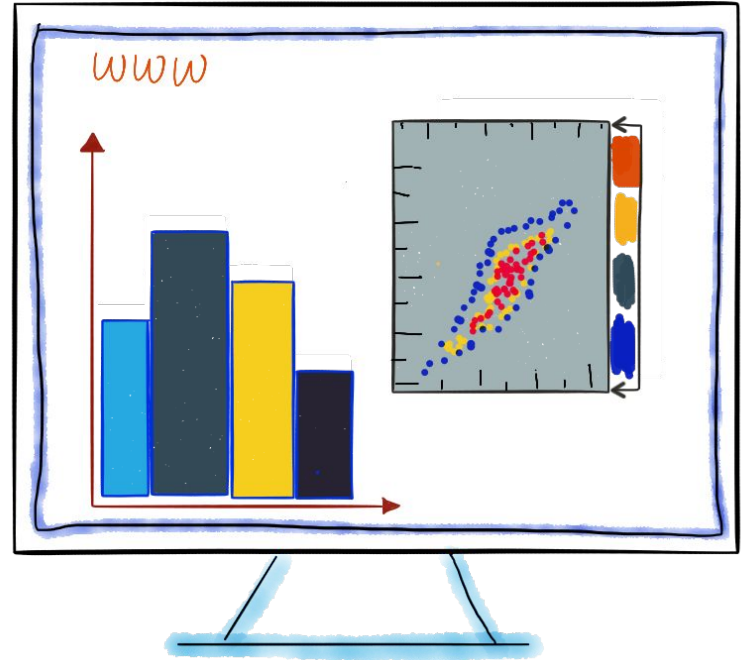**What are you thinking of doing after CERN?**
— Phd in CS

What you liked most in your experience as a summer student?
— Workshops!
E.g. Web application security penetration testing

Outline

- Overview of ATLAS Open Data

- ATLAS Open Data Analysis Tools

- Jupyter Notebooks for 13 TeV dataset

  - Notebook Analysis Example:

    H→γγ analysis

  - Analysis Framework Interface

- What I learned?

# Overview of ATLAS Open Data

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.

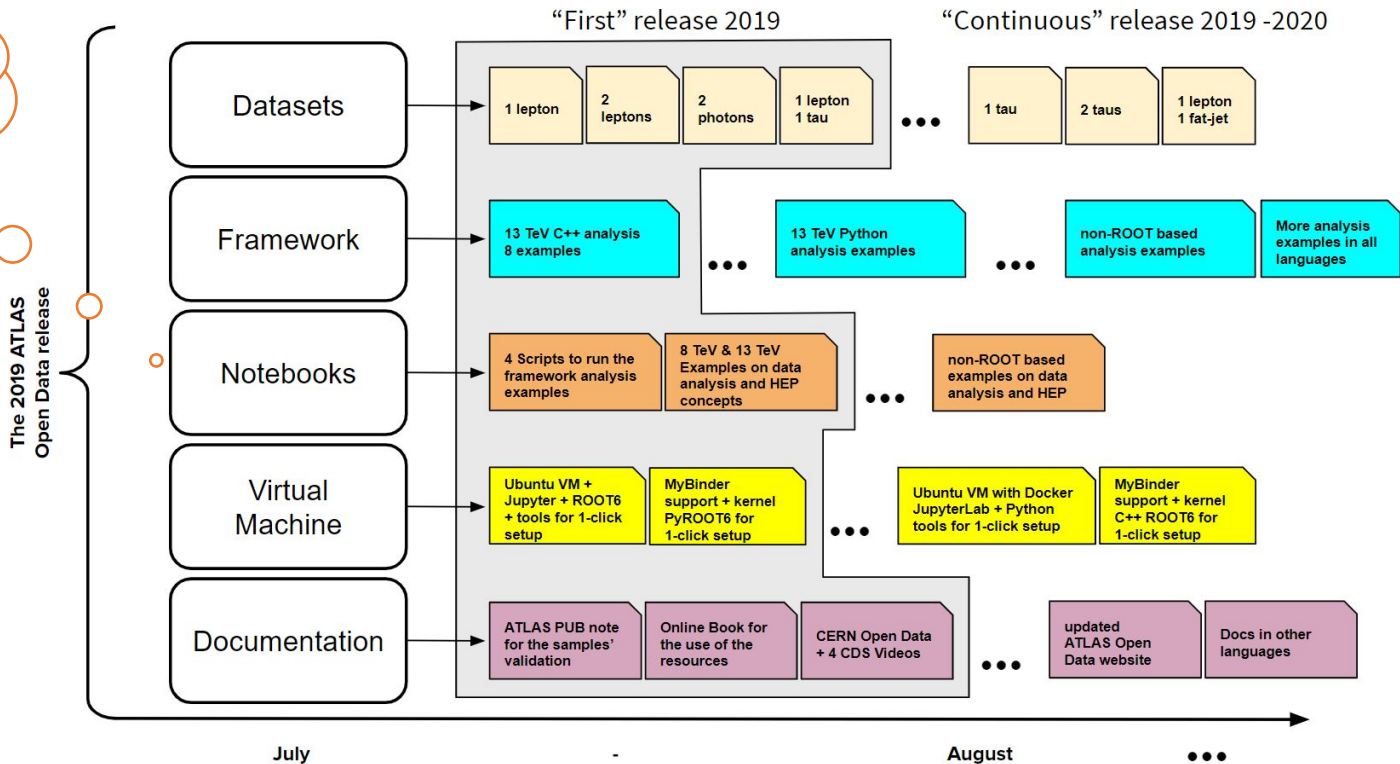- But of course also anyone else that are interested!

Aim:

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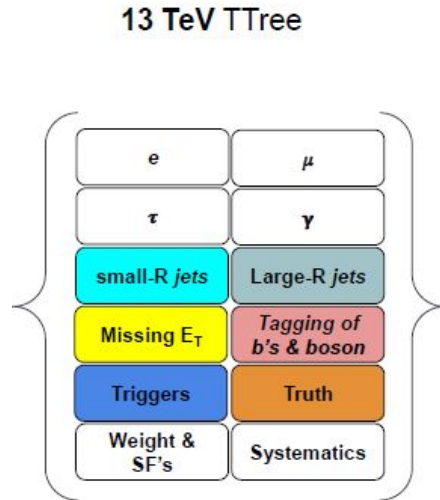

8 TeV release

13 TeV release

1 fb⁻¹ Data → 10 fb⁻¹

x 7 collections

44 samples

~120 samples
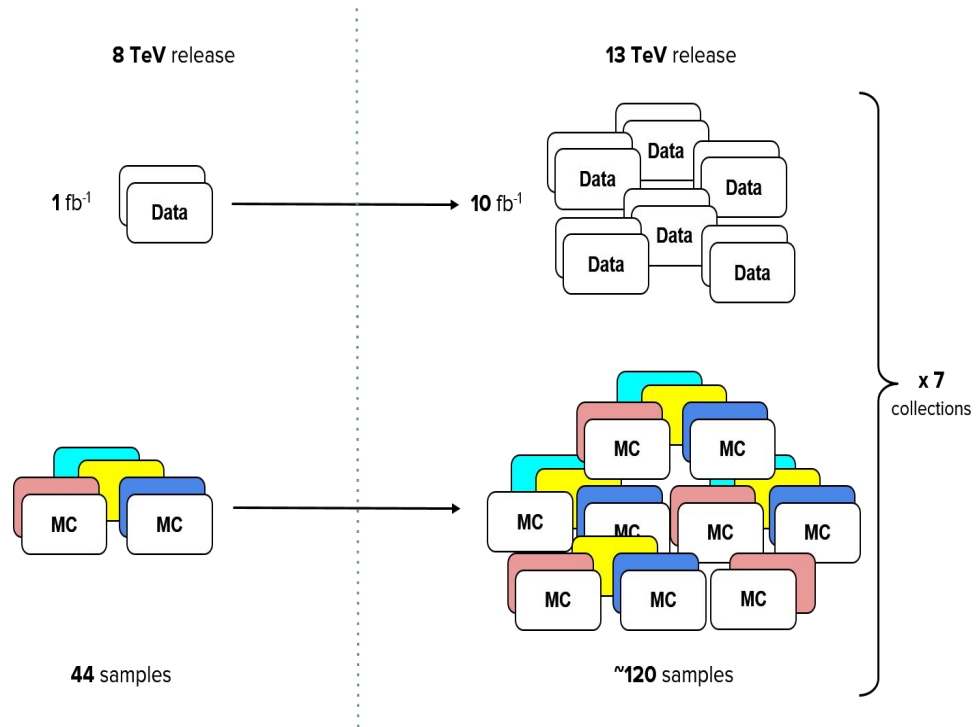
ATLAS Open Data Analysis Tools for 13 TeV dataset

Jupyter Notebooks for 13 TeV dataset
- Analysis Example: $H \rightarrow \gamma\gamma$ analysis (Python and C++)

13 TeV TTree

e	μ
τ	γ
small-R *jets*	Large-R *jets*
Missing E_T	*Tagging of b's & boson*
Triggers	Truth
Weight & SF's	Systematics

~100 variables

Hgg Analysis

NEW!

Based on:

PRD 98 (2018) 052005 — 11th January 2019

Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector

The ATLAS Collaboration

Simplified selection:

- 2 photons with pT > 25 GeV
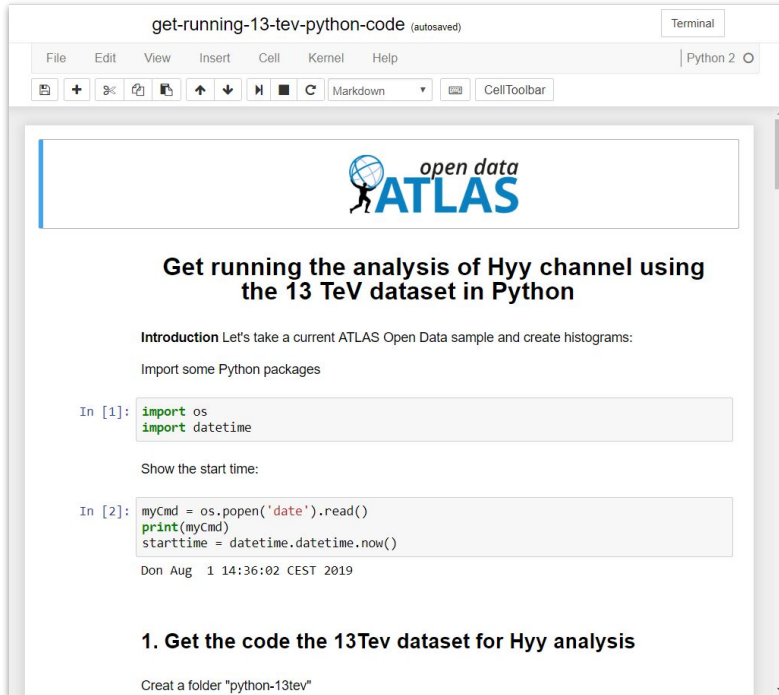- $E_T / m(\gamma\gamma) > 0.35$
- 105 GeV < $m(\gamma\gamma)$ < 160 GeV

Figures to reproduce:
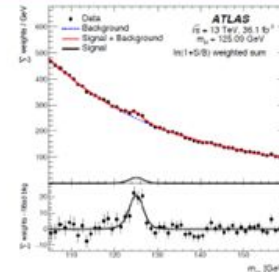
https://nbviewer.jupyter.org/github/veritasalice/CERN_Summer_Student_demo_2019/blob/master/notebooks/atlas/ATLAS_OpenData_01-python_Hyy_channel_analysis_example.ipynb

Jupyter Notebooks for 13 TeV dataset
- Interface of Framework (Python and C++)

https://nbviewer.jupyter.org/github/veritasalice/CERN_Summer_Student_demo_2019/blob/master/notebooks/atlas/nbviewer-get-running-13-tev-python.ipynb

What I learned?

- Work in team with Github

- Write analysis code in C++ and Python

- Write Shell scripts

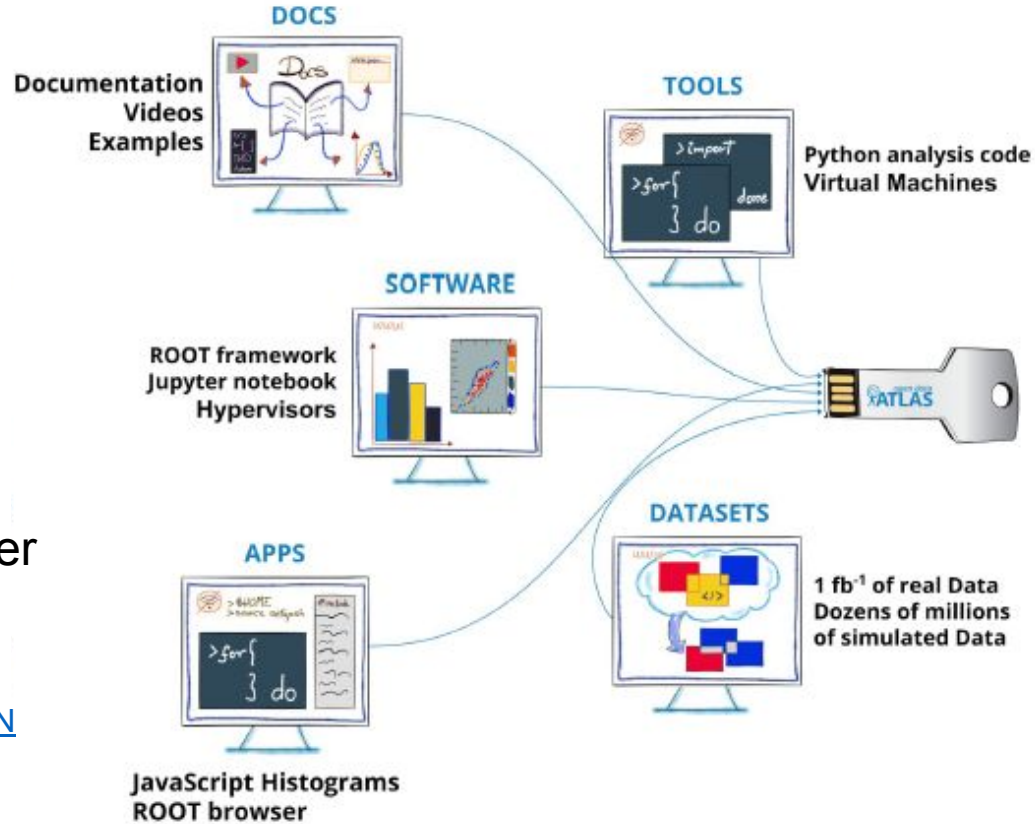- Documentation with Mkdocs

- Virtual Machine and MyBinder

- Of course, ROOT!

https://github.com/veritasalice/CERN_Summer_Student_demo_2019

http://opendata.atlas.cern/

DOCS
Documentation
Videos
Examples

TOOLS
Python analysis code
Virtual Machines

SOFTWARE
ROOT framework
Jupyter notebook
Hypervisors

ATLAS

DATASETS
1 fb^{-1} of real Data
Dozens of millions
of simulated Data

APPS
JavaScript Histograms
ROOT browser

Thank you for your attention!