ATLAS Open Data at 13TeV The journey to a fully educational HEP dataset

Kate Shaw, University of Sussex

On behalf of the ATLAS Collaboration

International Conference on High Energy Physics

29th July 2020

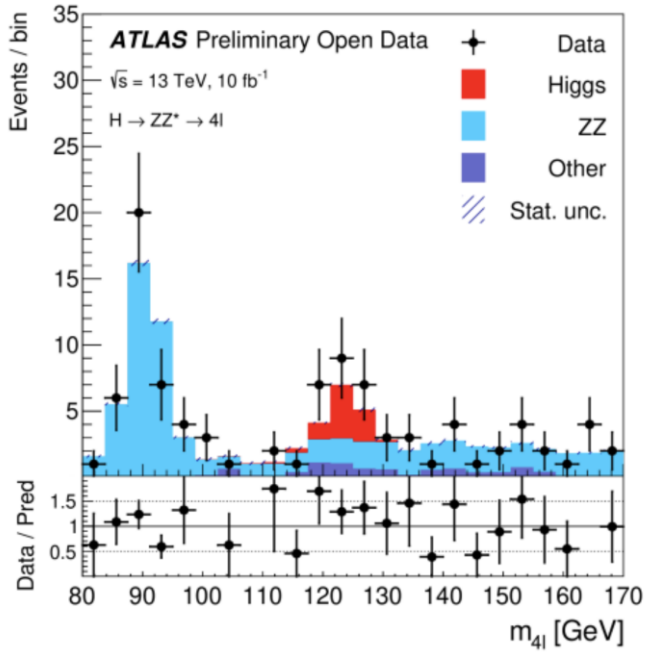
ATLAS EXPERIMENT

UNIVERSITY OF SUSSEX

ATLAS Open Data

- The ATLAS experiment is dedicated to providing our data to the public,
  - Specifically with physics undergraduates, graduates, teachers and lecturers, and high school students in mind

- We provide these proton-proton collision datasets within a comprehensive educational package to ensure usability at various levels, and for different educational objectives.

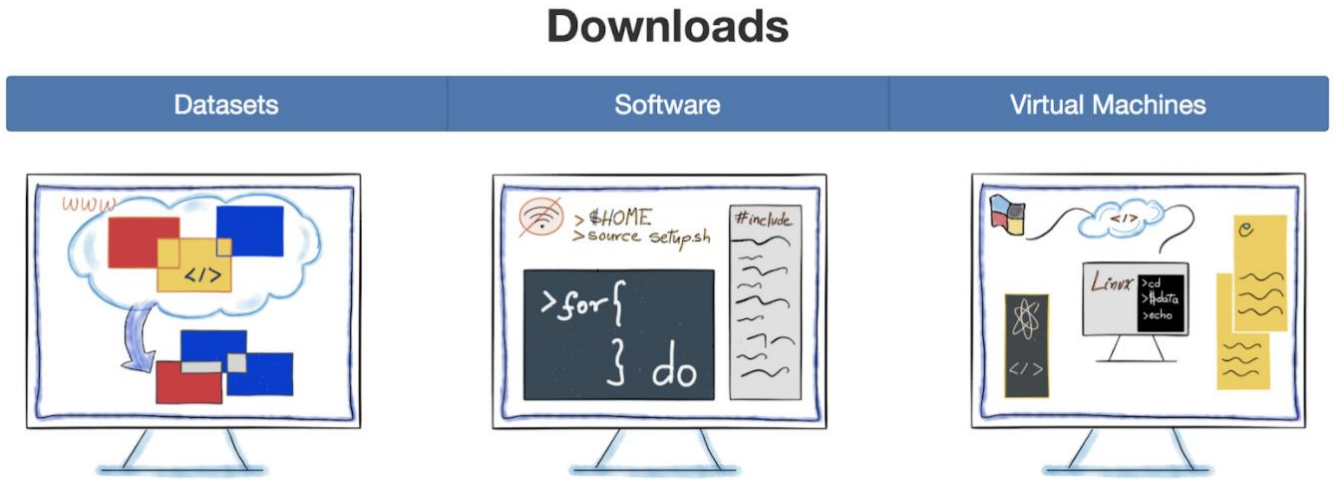


- Students get to analyse the data themselves to search for particles such as the Higgs, or SUSY!
- Using example code and an existing framework, advanced students can learn various analysis techniques, and gain an understanding of statistics and uncertainty.
- For less experienced students we provide tools without need for coding, and simple introductory notebooks to give students basic coding and analysis training.

ATLAS Open Data

- Collision data, with Standard Model MC simulations and analysis examples:
  - 1 fb^{-1} of 8 TeV collision data released in 2016 (ATL-OREACH-PUB-2016-001)
  - 10 fb^{-1} of 13 TeV collision data released in 2020 (ATL-OREACH-PUB-2020-001)
    - Additional BSM MC simulations
    - Many new analyses added

- All ATLAS Open Data datasets and MC available on CERN Open Data portal
- Analysis examples in C++ and Python, various options to access data and tools including Jupyter Notebooks and Virtual Machines

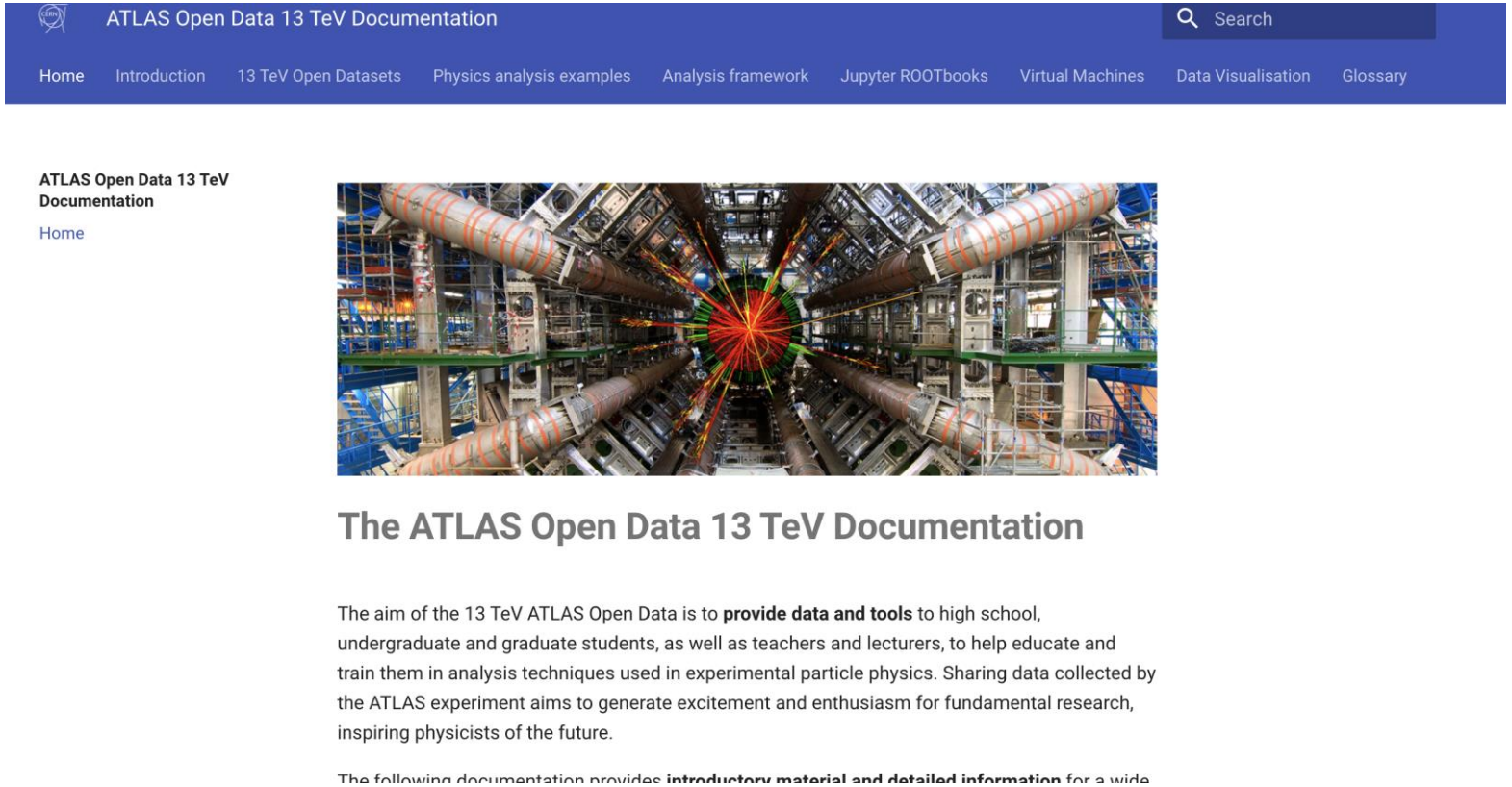


ATLAS Open Data

- **Educational objectives may include:**
  - **Basics of programming in C++ / Python**
  - **Basic ideas in particle physics and usage of ROOT and histograms**
  - **Training in more advanced analysis techniques used in experimental particle physics**
  - **Understanding of Big Data**
  - **Training in machine learning**

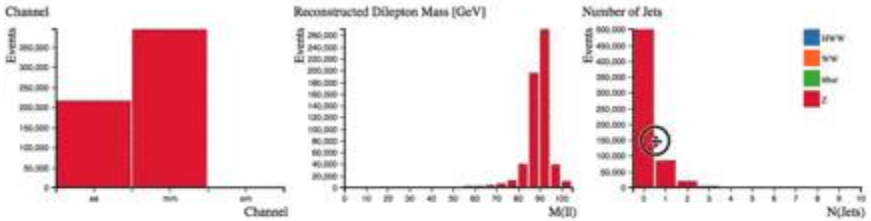
ATLAS Open Data

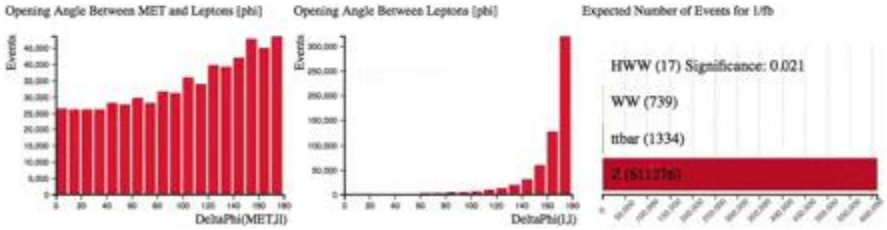
- **Full documentation on our website** https://atlas.cern/resources/opendata



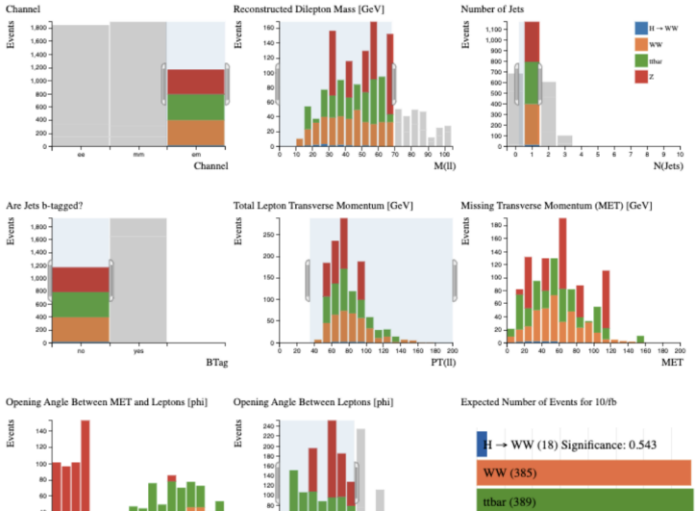
ATLAS Open Data

- Histogram Analyser
  - Perfect for beginners
  - Analyse data and find the Higgs without coding!
  - Full 8 TeV workbook developed
  - More advanced13 TeV documentation





- Number of Jets = 1
- no b-tagged jets
- electron-muon channel only
- Reconstructed Dilepton Mass < 70 GeV
- Total Lepton Transverse Momentum > 30 GeV
- Opening angle between leptons < 80

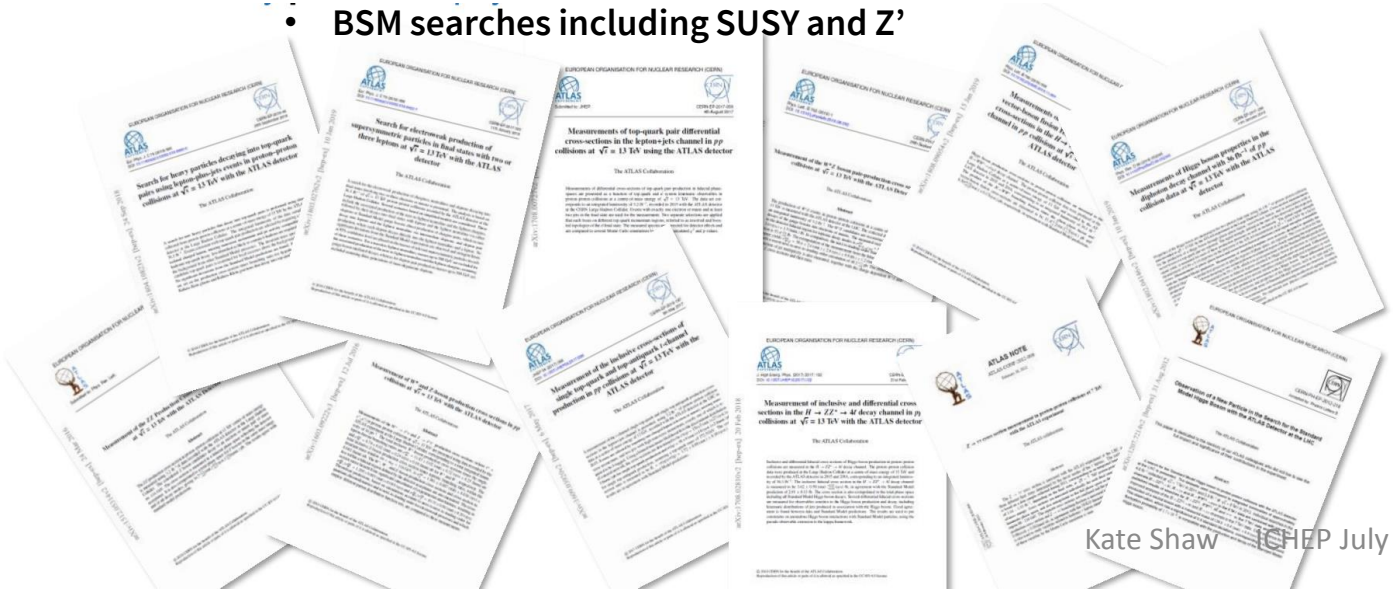


ATLAS Open Data

- 8 TeV analysis workbook and 13 TeV documentation
- Analysis framework includes
  - C++ based framework
  - Python uproot based framework
  - PyROOT-based framework

How to use:
- Download and run on your laptop
- Download Virtual Machine
- Use online Jupyter Notebooks

- Learn-by-doing: 12 examples of physics analysis
  - SM including single top, ttbar, W and Z
  - Higgs searches in diboson and gamma gamma
  - BSM searches including SUSY and Z'



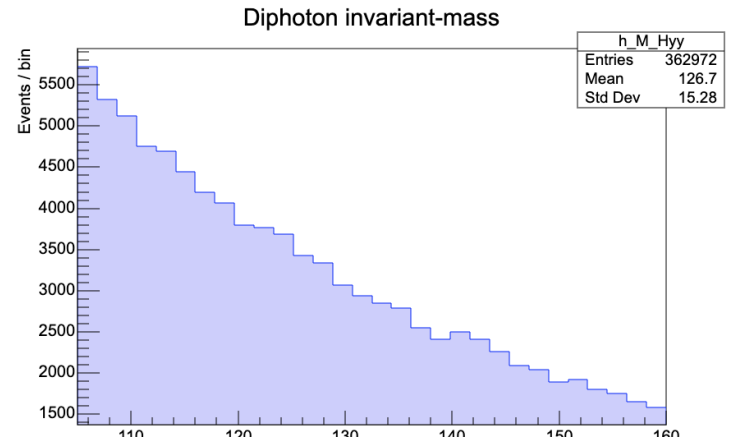
```
                float m_yy      = sqrt( 2 * Photon_1.Pt() * Photon_2.Pt() *
                                  (cosh( Photon_1.Eta() - Photon_2.Eta()) - cos(dPhi_yy)
                //Calculation of the Invariant Mass using TLorentz vectors
                TLorentzVector Photon_12 = Photon_1 + Photon_2;
                float mass_inv_GeV = Photon_12.M()/1000.;

                h_M_Hyy->Fill(mass_inv_GeV);
            }
        }
    } // end TrigPhoton request
}
std::cout << "* Analysed a total of: " << nentries << " in this sample." << std::endl;
```

```
* Total number of entries to analyse: 7798424
* Analysed a total of: 7798424 in this sample.
```

Final plot

```
In [10]: TCanvas *cz = new TCanvas("cz","cz",10,10,900,600);
         TText tz; tz.SetTextFont(42); tz.SetTextAlign(21);
         h_M_Hyy->Draw();
         cz->Draw();
```



Inspired and following as closely as possible the procedures and selections taken in already published physics results

ATLAS Open Data

Hyy Analysis

Based on:

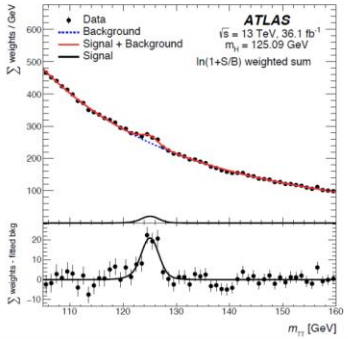
PRD 98 (2018) 052005 11th January 2019

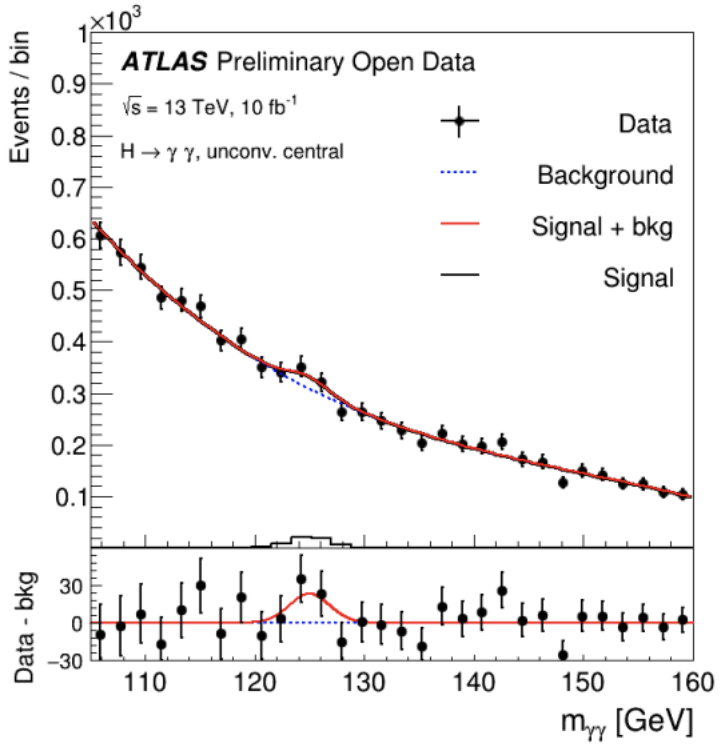
Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Simplified selection:

- 2 photons with pT > 35, 25 GeV
- E_T / m(γγ) > 0.35(0.25)
- 105 GeV < m(γγ) < 160 GeV

Figures to reproduce:

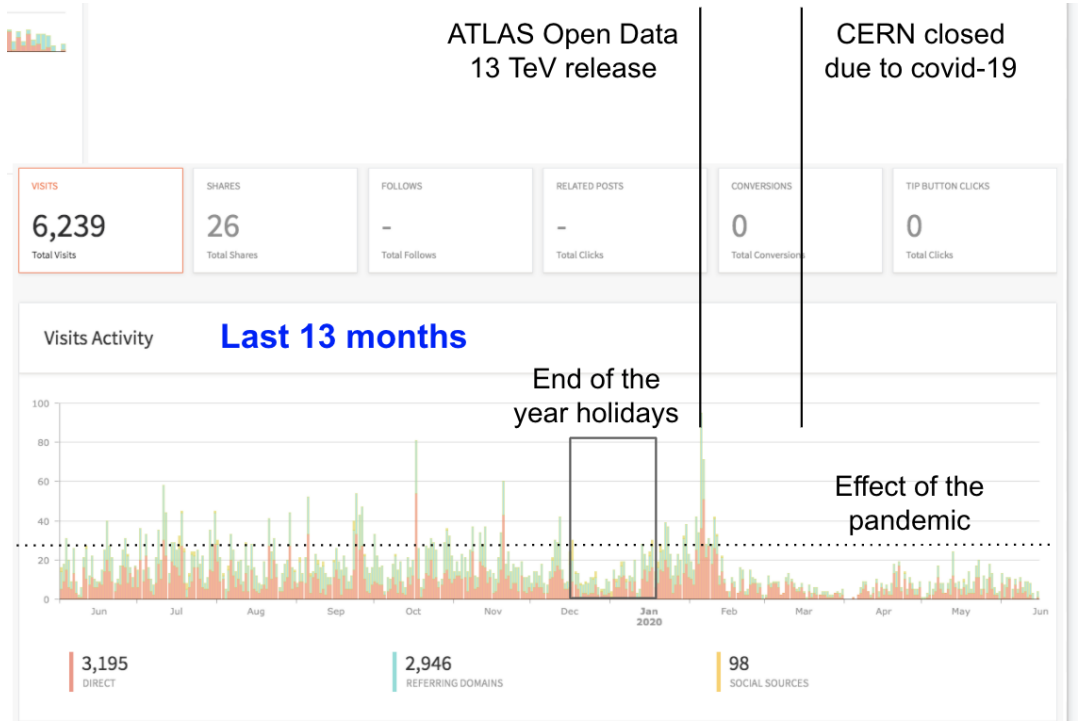




ATLAS Open Data

- Widley used in curriculums of multiple universities in Belgium, Canada, Colombia, Greece, Germany, Norway, Poland, Portugal, Spain, Sweden, Switzerland, UK, USA, Venezuela and many others
- Integrated into lab courses, used in BSc, MPhys projects
- Used in workshops & schools worldwide for training undergraduates in physics analysis, machine learning, coding and big data
- Topics include machine learning, computing skills, coding, and of course experimental particle physics

Visits to our Educational website



http://opendata.atlas.cern/visualisations/

ATLAS Open Data

How did we develop the ATLAS open data idea?

- Its an important part of any publicly funded experiment to release data
- We explicitly wanted it to be useful! Along with datasets a big effort giving to the educational resources, thinking about different levels of physics knowledge
- We wanted advanced university students to do high-level physics, giving them options of resources depending on computational skills and resources
- While also allowing beginners to start analysing data without even having coding, and develop workbooks easing them into basic coding and histograms
- We worked with different levels of students in our development of the project to test and improve

ATLAS Open Data

- High school and undergrad students worked on writing and testing the analysis frameworks and notebooks

- Summer Students @CERN
  - 2016: Tea Band (Montenegro) Anthony Abah Abah (Nigeria)
  - 2018: Ya-Feng Lo (Taiwan)
  - 2019: Yixin Wang (China) & Shodruz Umedov (Tajikistan)
- High School and ICTP Students @CERN
  - 2017: Tim Hebenstreit (Germany)
  - 2018: Amel Alhassan (Sudan) – ICTP
  - 2019: Ander Harris, William Dawson-Holgate & Kip Parker (UK)

ATLAS Open Data

- Substantial testing of tools and frameworks at various workshops helped to:
  - See what worked and what did not
  - Check whether the level was appropriate
  - Validate whether tools were easy to use and facilities able to run them
  - Understand new needs

- Held workshops
  - ICTP workshops: CODATA-RDA Research Data Science Summer School in 2017, 2018 and 2019 Reaching ~50 students
  - Multiple workshops in Latin America under the PWF-ICTP program in 2016, 2017, 2018, 2019, 2020 (January) ~1000 students
  - Multiple workshops at CERN in 2016, 2017, 2018, 2019

ATLAS Open Data

Master students did their theses to enhance or develop a new physics analysis, or develop extra studies: (examples)

- Iskya Garcia (Venezuela)
- Maria Di Domenico (Venezuela)
- Meirin Evans (UK)
- Stanislav Biryukov (UK)
- Francois de Tournemire (UK)
- Aodhan Burke and Jack Harrison (UK)
- Arturo Prieto Tirado (Spain)

PhD students did their qualification task on the development of the data format

- Even Simonsen Haland (Norway)
- Meirin Evans (UK)

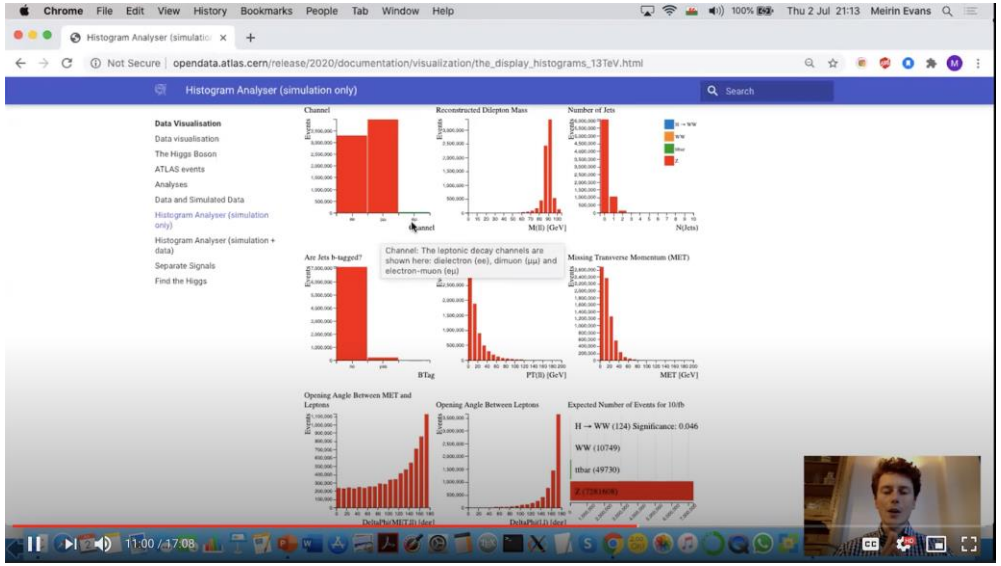
Great way for students to learn about particle physics and analysis techniques without needing to learn the full ATLAS software!

ATLAS Open Data

ATLAS Open Data being used for educational purposes all over the world!!

Currently working on various video tutorials to assist non ATLAS members to learn how to use our Open Data

- **Multimedia vital for online learning**
- **Supports instructors to deliver a course or a workshop.**
- **Reach out to more people, bring ATLAS physics to their homes, and hopefully inspiring young into physics!**



THANK YOU!