PROCEEDINGS OF SCIENCE

Open Data at ATLAS: Bringing TeV collisions to the World

Mariana Vivas Albornoz* **on behalf of the ATLAS Collaboration**

University of Massachusetts, Amherst,
Amherst, MA 01003, USA

E-mail: mariana.vivas.albornoz@cern.ch

The ATLAS Collaboration is committed to advancing particle physics education and inspiring the next generation of physicists. Through the open data for education initiative, the aim is to demystify science, promote high-energy physics to a broader audience, and build trust in our methodologies. Launched in 2016, this project currently includes two datasets from proton–proton collisions at the Large Hadron Collider: a fraction of both 2012 run with $\sqrt{s}$ = 8 TeV and 2016 run with $\sqrt{s}$ = 13 TeV. To facilitate the use of these data, a range of documentation and exercises has been developed, from interactive tools for data analysis to programming tools with a complete analysis workflow. This year, the whole data from 2015 and 2016 runs will be released, bringing the total integrated luminosity to 36 fb^{-1}, enabling new analyses and enhancing existing examples.

ATL-OREACH-PROC-2024-004
24 September 2024

42nd International Conference on High Energy Physics (ICHEP2024)
18-24 July 2024
Prague, Czech Republic

*Speaker

1. Why Open Data for Education?

The ATLAS Collaboration, as one of the largest experiments in particle physics, is committed to advancing education and fostering the development of future scientists. Through the release of open data, we have identified means to achieve both of these objectives: providing resources for educational purposes, while attracting and nurturing the next generation of physicists.

The open data for education initiative is designed to:

- **Promote science and increase awareness**: Releasing open data makes scientific research available to the public. It helps demystify complex scientific concepts and encourages a broader interest in science and technology.

- **Introduce people to particle physics**: Providing open access to the data allows students, teachers and enthusiasts to learn about particle physics.

- **Build public trust in science and methodology**: Transparency through open data fosters trust in scientific research by allowing the public to see the data and methodologies behind important discoveries.

2. ATLAS Open Data for Education

The ATLAS Open Data project started in 2016 with the first release of 8 TeV data for education [1]. It continued with a 13 TeV dataset [2] in 2020, which also enriched the project with hands-on material, open tools, and documentation. All of this is hosted at the ATLAS Open Data website [3].

2.1 The Data

In 2016 two distinct datasets from the 8 TeV proton–proton collision data were published, recorded in 2012 at the Large Hadron Collider (LHC). These datasets were released in XML and ROOT ntuple format, with an integrated luminosity of 2 fb^{-1} provided in XML format and an additional 1 fb^{-1} in ROOT ntuple format. With nearly 15 million events, this dataset offers a comprehensive view of the ATLAS experiment's 2012 data-taking period, and can be found in the CERN Open Data portal [4].

For the following release in 2020, the data were collected by the ATLAS detector at the LHC at 13 TeV during the year 2016 and correspond to an integrated luminosity of 10 fb^{-1}. The proton–proton collision data are accompanied by a set of Monte Carlo simulated samples describing several processes which are used to model the expected distributions of different signal and background events.

The released samples are provided in a simplified data format, reducing the information content of the original data analysis format used within the experiment. The resulting format is a simplified TTree tuple (or ROOT ntuple) with more than 80 branches, providing a more detailed view compared to the previously released 8 TeV samples, which have 45 branches.

2.2 Example Analyses and Open Tools

We recognize the need to have different materials for different types of users. That's why three different user paths were created. These can help people interact with the material. A "Quick start" guide, for the users visiting the website for the first time or those who do not want to spend time configuring tools; a "Deep Dive" guide, for students who want to use multiple exercises and tools or teachers who want to use the data in a class; and an "Online Data Analyser" guide, for the users who prefer a no-code approach to the analysis.

These paths explain to the users how to navigate the three main learning tools offered: the Jupyter notebooks, the analysis frameworks, and the interactive histogram analyser.

The interactive histogram analyser provides a tool for performing analysis without coding. Users can interact with histograms representing different variables, apply cuts, and immediately see how these adjustments affect the significance of the signal.

For those interested in coding, the most accessible analyses are available in Python Jupyter notebooks. These notebooks include example analyses within the Standard Model, such as searches for the Higgs boson in the ZZ and $\gamma\gamma$ channels, as well as Beyond the Standard Model searches. Additionally, the notebooks demonstrate various approaches to using machine learning as a tool in high-energy physics for separating signal from background.

Analysis frameworks in different formats are also offered. This includes frameworks using Python Uproot, Python PyROOT, and C++. The frameworks provide the code needed to analyze data and generate plots. Examples include W-boson production in the single-lepton final state, t-channel single-top-quark production in the single-lepton final state, and searches for Supersymmetric particles in the two-lepton final state, among others.

3. Future Plans

The next release will include a subset of the data from the ATLAS experiment's 2015 and 2016 proton–proton collisions, with an integrated luminosity of 36 fb^{-1}. This release will enable new analyses, such as the discovery of the Higgs boson in the $b\bar{b}$ decay channel.

With this new release, the goal is to add new analyses, statistically improve existing public analyses, and provide more comprehensive documentation and tools to make the data more accessible to users.

References

[1] ATLAS collaboration, *Review of ATLAS Open Data 8 TeV datasets, tools and activities*, https://cds.cern.ch/record/2624572, 2018.

[2] ATLAS collaboration, *Review of the 13 TeV ATLAS Open Data release*, https://cds.cern.ch/record/2707171, 2020.

[3] ATLAS Collaboration, "ATLAS Open Data." https://opendata.atlas.cern. Last accessed 29 August 2024.

[4] ATLAS Collaboration, "ATLAS samples for 2016 open data release." DOI. Last accessed 29 August 2024.