



ATLAS Open Data

António Jacques Costa,
on behalf of the ATLAS Collaboration

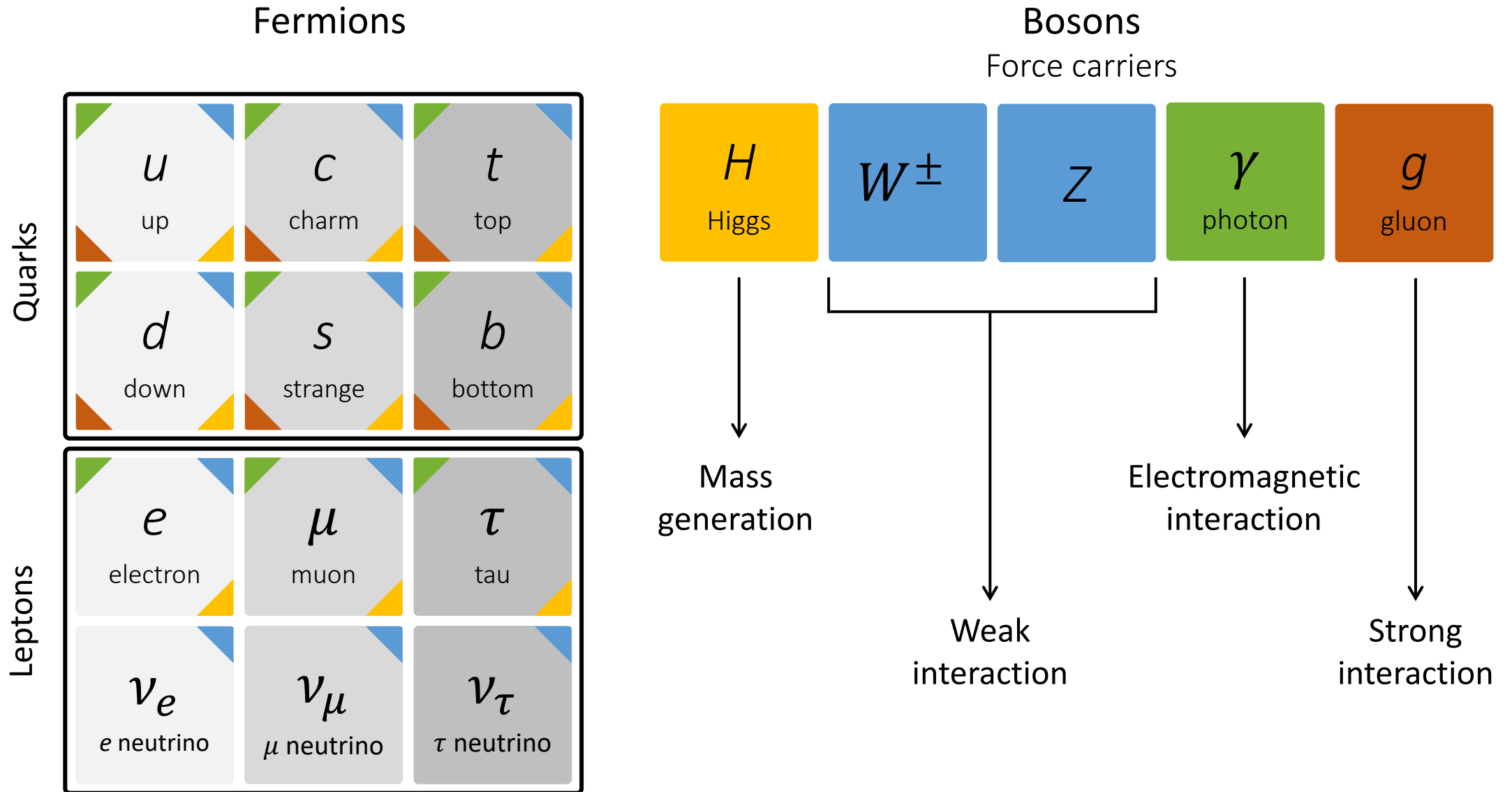
IOP Open Data Workshop

for Higher Education

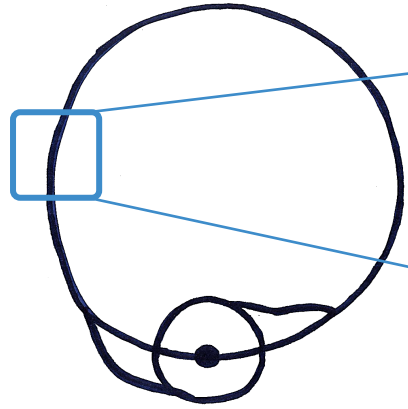
14th May 2026



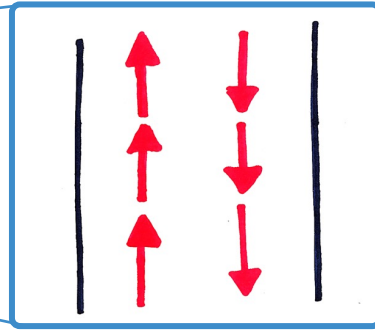
The Standard Model of Particle Physics



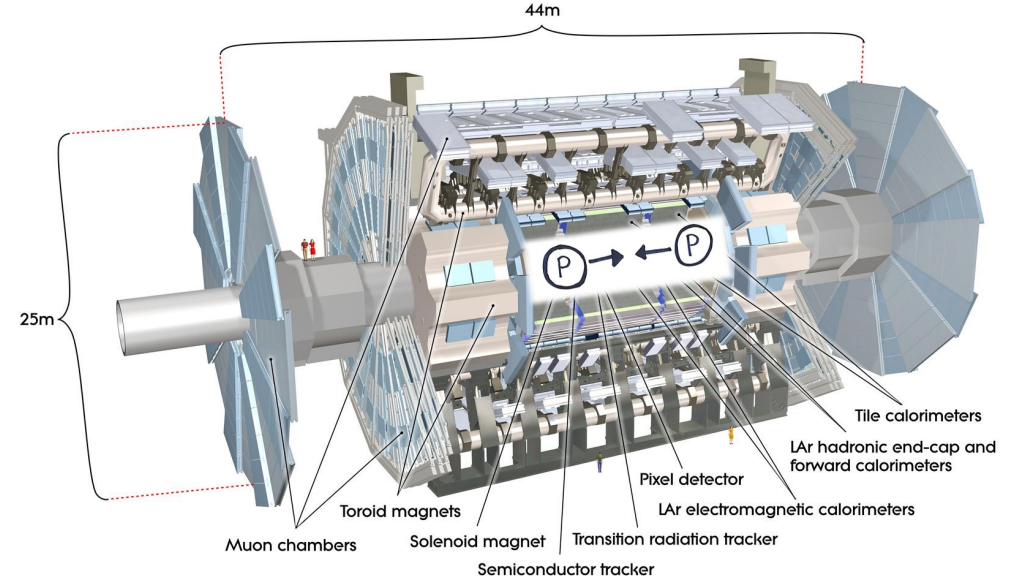
The ATLAS Experiment at CERN



Large Hadron Collider @ CERN



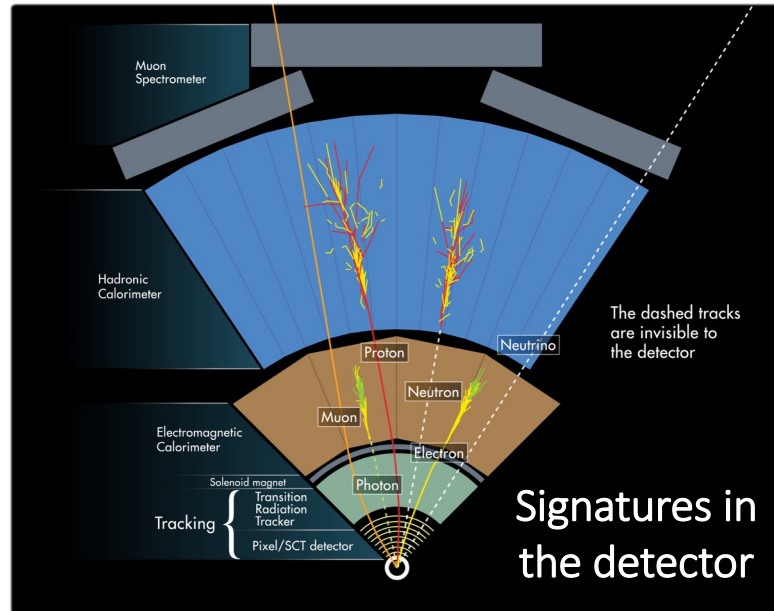
Protons travelling at almost speed of light



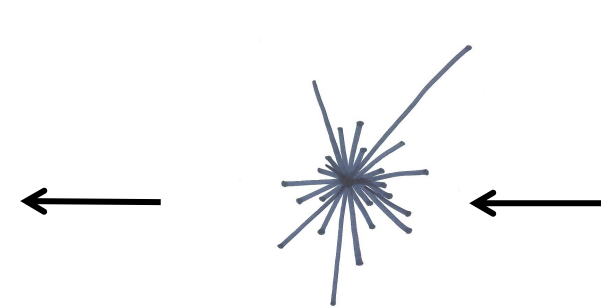
Protons collide inside detector (e.g. ATLAS experiment)



Data collection of collision events and respective particles



Signatures in the detector



Different particles are created in the collisions

Open Data Releases

Accessible

Transferable skills

Usable and versatile

Wide reach

For Research

Released

Proton-Proton Collisions

36 fb⁻¹ proton-proton collisions
7B data, 2B simulation events, 373 MC sets [\[link\]](#)

Released

Heavy Ion Minimum Bias

486 μb⁻¹ lead-lead collisions
221M data + 100k simulation events [\[link\]](#)

Released

Event Generation Data

>12B simulated events, >6000 datasets
13 TeV and 13.6 TeV [\[link\]](#)

Coming
soon

Heavy Ion Hard Probes

New format, being prepared now

For Education and Outreach

Released

8 TeV (2016)

3 fb⁻¹ proton-proton collisions
60M events, 42 simulation datasets [\[link\]](#)

Released

13 TeV (2020)

10 fb⁻¹ proton-proton collisions
940M events, 228 datasets [\[link\]](#)

Released

13 TeV (2025 Beta)

36 fb⁻¹ proton-proton collisions
9.8B events, 373 simulation datasets [\[link\]](#)

All data are released under a [CC0 license](#) (free use) with a request for citation

ATLAS Open Data website

ATLAS Open Data

High Energy Physics data for everyone.

For Education 🎓

To provide data and tools to high school, undergraduate and graduate students, as well as teachers and lecturers, to help educate them and exercise in physics analysis techniques used in experimental particle physics.

For Research 🔬

To provide researchers with high-quality data recorded by the ATLAS detector, enabling them to conduct state of the art analyses in particle physics.

[Get Started](#)

<https://opendata.atlas.cern>

Get Started

In this section you will find different suggested paths to get involved with ATLAS Open Data. This are just suggestions on what we think it will be more usefull to check in each case. However, feel free to check the website freely.

Quick start 🚀

The quickest way to start learning with ATLAS Open Data.

Deep Dive 🧐

For extended use. Let's dive into what ATLAS has to offer!

Researchers Toolkit 🔬

Detailed information and resources for researchers

Web apps for Education 📱

Simple web-based applications for education

- Goal to have **open data** that's **accessible to a wide variety of audiences**
- Learners can move to **more or less complex setups** painlessly
- If you want **more complexity** than the Outreach and Education Open Data allow, try the **Open Data for Research!**
- If **O&E Open Data** is **too complex**, try the **web apps!**

Documentation

The screenshot shows the ATLAS Open Data website. The top navigation bar includes 'ATLAS Open Data', 'Get Started', 'Data', 'Tutorials', 'Python API', 'Documentation', 'About us', and 'Help'. A search bar is located on the right. The left sidebar lists various topics, with 'The Standard Model and Beyond' selected. The main content area features the title 'The Standard Model of Particle Physics and Beyond' and a detailed introduction to the Standard Model (SM) of particle physics. It explains that the SM is a theoretical framework describing fundamental particles and their interactions, excluding gravity. It mentions the success of the SM in predicting experimental outcomes and its status as an incomplete theory, encouraging ongoing research into physics beyond the Standard Model. A link to a 'Standard Model cheat sheet' is provided. Below the introduction, there is a section titled 'Constituents of the Standard Model' with sub-sections for 'Fermions' and 'Bosons'. The 'Fermions' section describes them as particles with spin-1/2, categorized into quarks and leptons. It lists quarks (up, down, charm, strange, top, bottom) and leptons (electrons, muons, tau, and their respective neutrinos).

ATLAS Open Data

Get Started Data Tutorials Python API Documentation About us Help

Search

ATLAS Open Data

The Large Hadron Collider

The ATLAS Experiment

The Standard Model and Beyond

Introduction to Heavy-Ion Physics

Data Formats

Data Collection

Simulated Data

Physics Objects

Setting Uncertainties

Example Analysis

Citing ATLAS

Support and Resources

The Standard Model and Beyond

The Standard Model of Particle Physics and Beyond

The Standard Model (SM) of particle physics is a theoretical framework that describes the fundamental particles and their interactions, excluding gravity. It summarizes our current understanding of quantum mechanics and field theory, combining the electroweak interaction and quantum chromodynamics (QCD) into a coherent model. Despite its success in predicting experimental outcomes, the Standard Model is considered an incomplete theory, encouraging ongoing research into physics beyond the Standard Model. For a simpler description of the Standard Model, check this [Standard Model cheat sheet](#).

Constituents of the Standard Model

Fermions

Fermions—particles with spin-1/2—, are the building blocks of matter, categorized into quarks and leptons.

- Quarks** possess fractional electric charges and participate in all four fundamental forces, notably the strong interaction, mediated by gluons. They are confined within hadrons due to color confinement and are found in six flavors: up (u), down (d), charm (c), strange (s), top (t), and bottom (b), each with corresponding anti-particles.
- Leptons**, carrying integer electric charges, interact through the electromagnetic and weak forces but not the strong force. The lepton family includes electrons (e), muons (μ), tau (τ), and their respective neutrinos (ν_e, ν_μ, ν_τ), which are electrically neutral and interact primarily via the weak force.

Bosons

The screenshot shows the 'Video Tutorials for Open Data' page. The top navigation bar includes 'Video Tutorials' and 'Video Tutorials for Open Data'. The main content area features the title 'Video Tutorials for Open Data' and a paragraph explaining that there are multiple video tutorials available on YouTube that can be used as a starting point to learn more about how to set-up the software required to do analysis with ATLAS Open Data and how to analyse the data. It provides an example of learning how to rediscover the Higgs boson and how to create histograms with PyROOT. Below this, it lists the available video tutorials: 'Getting Started with ATLAS Open Data'. A video thumbnail is shown with the text 'OPEN DATA HOW TO GET STARTED'. Below the thumbnail, it states that this ATLAS Open Data tutorial will teach you how to start your journey with ATLAS Open data. It lists what you'll learn in this video: how to navigate the website, running notebooks online and in a docker container, and different levels of complexity in which you can access our material. At the bottom, there is a link to 'Find the Higgs boson with your mouse'.

Video Tutorials Video Tutorials for Open Data

Video Tutorials for Open Data

We have multiple video tutorials available on YouTube that you can use as a starting point to learn more about how to set-up the software required to do analysis with ATLAS Open Data and how to analyse the data.

For example you can learn how to rediscover the Higgs boson! But you can also learn for example on how to create histograms with PyROOT!

The available video tutorials are:

[Getting Started with ATLAS Open Data](#)

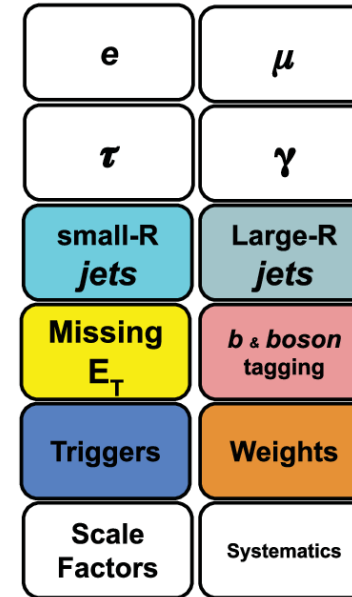
This ATLAS Open Data tutorial will teach you how to start your journey with ATLAS Open data. What you'll learn in this video:

- how to navigate the website
- running notebooks online and in a docker container
- different levels of complexity in which you can access our material

[Find the Higgs boson with your mouse](#)

Education datasets content

- Datasets contain **simplified information** with respect to the internal ATLAS collected and simulated data
- **Facilitated understanding** and addressing education goal
- Different datasets/selections on physics objects allow coverage of **simple to more complex uses**








Around 90 variables

How to run

- Big diversity of approaches and possible target audiences
 - Different environments, in which to use analysis code, provided to suit different needs

Online cloud services

<p>SWAN</p> 	<p>Binder</p> 	<p>Google Colab</p> 
<p><u>ESCAPE</u> Virtual Research Environment</p> 	<p><u>EXPLORE</u>, LHC Open Data Analysis Platform</p> 	

Offline services

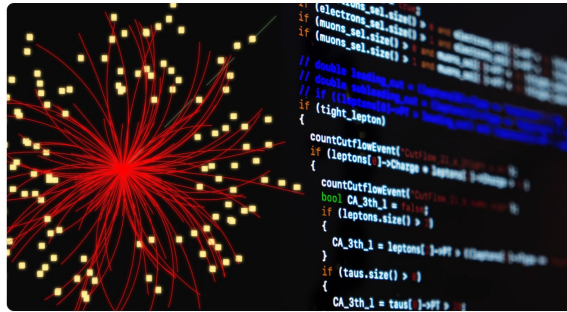


docker

Classroom Application

- Light-weight introduction to Open Data
- Fitted for high-school classes

Welcome to ATLAS Open Data in the Classroom



Please select your language to continue:

Select Language

English

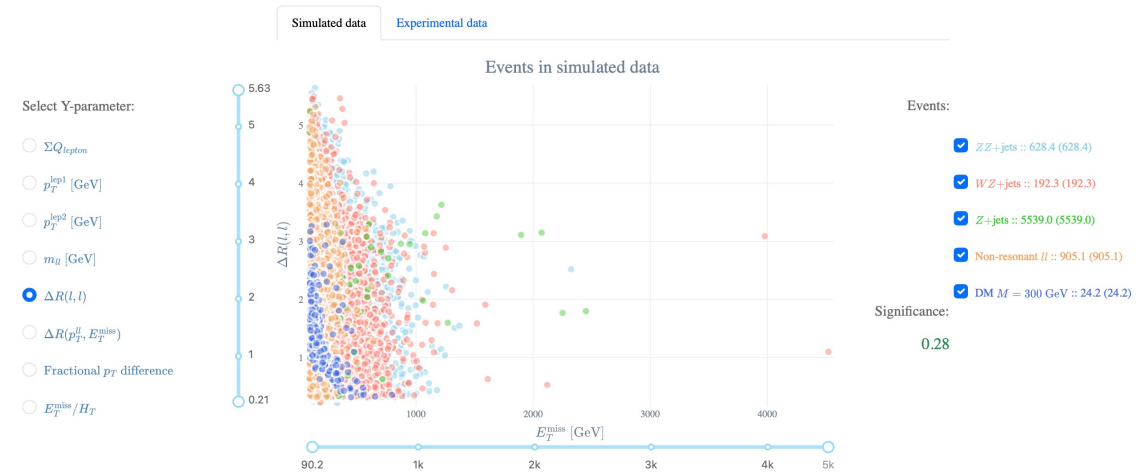
English

Spanish

Italian

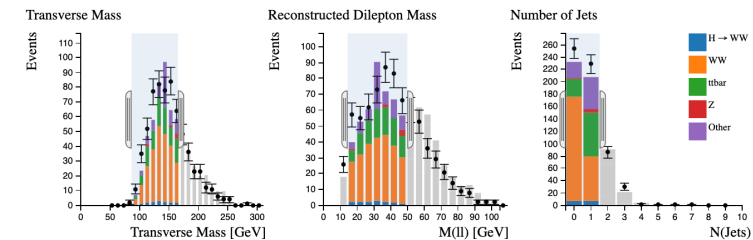
Machine Learning (ML) on the web

- Visual and interactive introduction to ML using Open Data
- No code required



Online histogram analyser

- Web-based histogram visualisation
- Direct, basic entry-level, no code hands-on experience
- Documentation to provide context



Education notebooks and frameworks

- Notebooks and frameworks accessible via [website](#) and Github ([Notebooks](#), [C++ framework](#))
- Provided for different analyses and programming languages, along with documentation, respective description and difficulty associated
- Python support package released, to access samples and setup analysis environment: [atlasopenmagic](#)
- Notebooks on [Introduction to HEP concepts](#)

Concepts

These notebooks introduce a variety of concepts in High Energy Particle physics. They are intended to provide both a conceptual introduction and some real examples of how to study the concepts with the ATLAS Open Data.

Jupyter Notebooks

Uproot

Accessing Metadata **NEW**

This notebook introduces the `atlasopenmagic` package, which is used throughout these notebooks for environment setup and data access. It introduces the concept of metadata, explains what metadata are available for the samples that have been provided, and introduces the use of a handful of convenient search functions to identify samples that might be of interest based on their metadata.

[Launch Binder](#) [Open in Colab](#)

Detector Acceptance and Efficiency **NEW**

This notebook introduces the concepts of detector acceptance and efficiencies, used to understand the response of the detector to different types of events. Using the 2025 release of ATLAS Open Data, we guide you through a practical example of how to calculate acceptance and efficiency, and show you how these concepts are used in real physics analyses.

[Launch Binder](#) [Open in Colab](#)

[Introduction to HEP concepts](#)

Standard Model

These notebooks dive into the world of Standard Model searches, exploring the fundamental particles and forces that constitute the universe as described by the Standard Model of particle physics. Through these analyses, we aim to test the predictions of the Standard Model, enhancing our understanding of the universe.

Jupyter Notebooks

Uproot

Higgs to ZZ **NEW**

This notebook uses the 2025 release of the ATLAS Open Data to show you the steps to rediscover the Higgs boson yourself! You will discover the Higgs boson decaying into a pair of Z bosons, which are in turn decaying into a lepton-antilepton pair each.

[Launch Binder](#) [Open in Colab](#)

Higgs to $\gamma\gamma$ analysis **NEW**

This notebook uses the 2025 release of ATLAS Open Data, with 36.1 fb^{-1} , to show you the steps to rediscover the Higgs boson yourself! You will discover the Higgs boson decaying into two photons.

[Launch Binder](#) [Open in Colab](#)

[Analysis notebooks](#)

```
import atlasopenmagic as atom
atom.available_releases()
atom.set_release('2025e-13tev-beta')
```

Available releases:

```
=====
2016e-8tev      2016 Open Data for education release of 8 TeV proton-proton collisions (https://opendata.cern.ch/record/3860).
2020e-13tev    2020 Open Data for education release of 13 TeV proton-proton collisions (https://cern.ch/2r7xt).
2024r-pp       2024 Open Data for research release for proton-proton collisions (https://opendata.cern.ch/record/80020).
2024r-hi       2024 Open Data for research release for heavy-ion collisions (https://opendata.cern.ch/record/80035).
2025e-13tev-beta 2025 Open Data for education and outreach beta release for 13 TeV proton-proton collisions(https://opendata.cern.ch/record/93910).
2025r-evgen    2025 Open Data for research release for event generation (https://opendata.cern.ch/record/160000).
Fetching and caching all metadata for release: 2025e-13tev-beta...
Successfully cached 374 datasets.
Active release: 2025e-13tev-beta. (Datasets path: REMOTE)
```

```
# Select the skim to use for the analysis
skim = "GamGam"
```

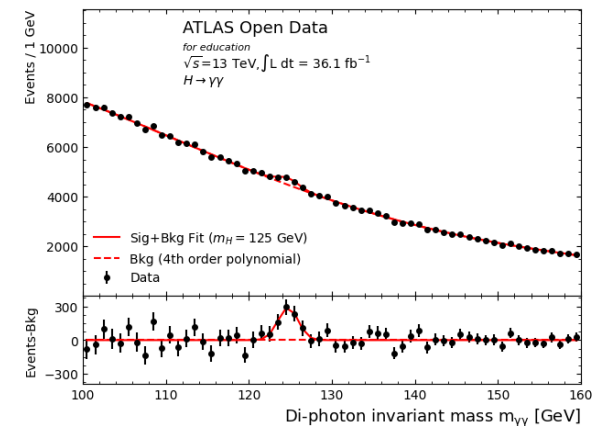
```
# Let's get the list of files to go through
# Notice that we use "cache" so that the files are downloaded locally and not streamed
files_list = atom.get_urls('data', skim, protocol='https', cache=True)
```

```
# Perform the cuts for each data entry in the tree and calculate the invariant mass
for data in tree.iterate(variables, library="ak", entry_stop=numevents*fraction):
```

```
    photon_isTightID = data['photon_isTightID']
    data = data[cut_photon_reconstruction(photon_isTightID)]
```

```
    photon_pt = data['photon_pt']
    data = data[cut_photon_pt(photon_pt)]
```

```
    data = data[cut_isolation_pt(data['photon_ptcone20'], data['photon_pt'])]
```



Hearing from the community

- [External projects](#)

Examples from institutes/individuals across the world of Open Data being put into use for education and research (courses, workshops, activities)

- [Community contributions](#)

- Research focused developed material

Projects


Explore various projects and initiatives based on ATLAS Open Data.

Filter by programming language:


Choose options




Analysis Grand Challenge of the PHYSLITE data format
Responsible: [Alexander Held](#) (University of Wisconsin Madison)
Language: English
Programming language: Python
Difficulty: Advanced Undergrad
Length: 3 hours
[View Source](#)




Discovery of the Higgs Boson
Responsible: [Martin Sever](#) (University of Melbourne)
Language: English
Programming language: Python, ROOT
Difficulty: Advanced Undergrad
Length: 18 hours
[View Source](#)



HEP Data Analysis Tutorial with the Scientific Python Ecosystem
Responsible: [Vangelis Kourlitis](#) (Aristotle University of Thessaloniki)
Language: English
Programming language: Python
Difficulty: Masters level
Length: 10 hours
[View Source](#)



Introduction to Machine Learning for Physicists
Responsible: [Ethan Simpson](#) (University of Manchester)
Language: English
Programming language: Python
Difficulty: Early grad
Length: 3 hours
[View Source](#)



Final Project ZBoson and Search for New Resonances with ATLAS
Responsible: [Marjorie Shapiro](#) (UC Berkeley)
Language: English
Programming language: Python
Difficulty: 3rd year undergrad level
Length: 4 hours
[View Source](#)


Community Contributions

Here we gather various projects and analyses created using our open data for research. We believe in the power of collaboration and the insights that can emerge from diverse perspectives. If you've used our open data for something cool, we would love to hear about it! Please share your work with us through the [contact us](#) form. Your contributions can inspire others and help to show the potential of open data.


A full list of academic uses of ATLAS Open Data can be found [on INSPIRE-HEP](#).

We check projects before posting them here, but do not perform a detailed validation. In case you find any issues, you are always welcome to get in touch with the authors.

Notebooks



HOW THE SCIENTIFIC PYTHON ECOSYSTEM HELPS ANSWERING FUNDAMENTAL QUESTIONS OF THE UNIVERSE
By [Vangelis Kourlitis](#)
[launch](#) [binder](#)



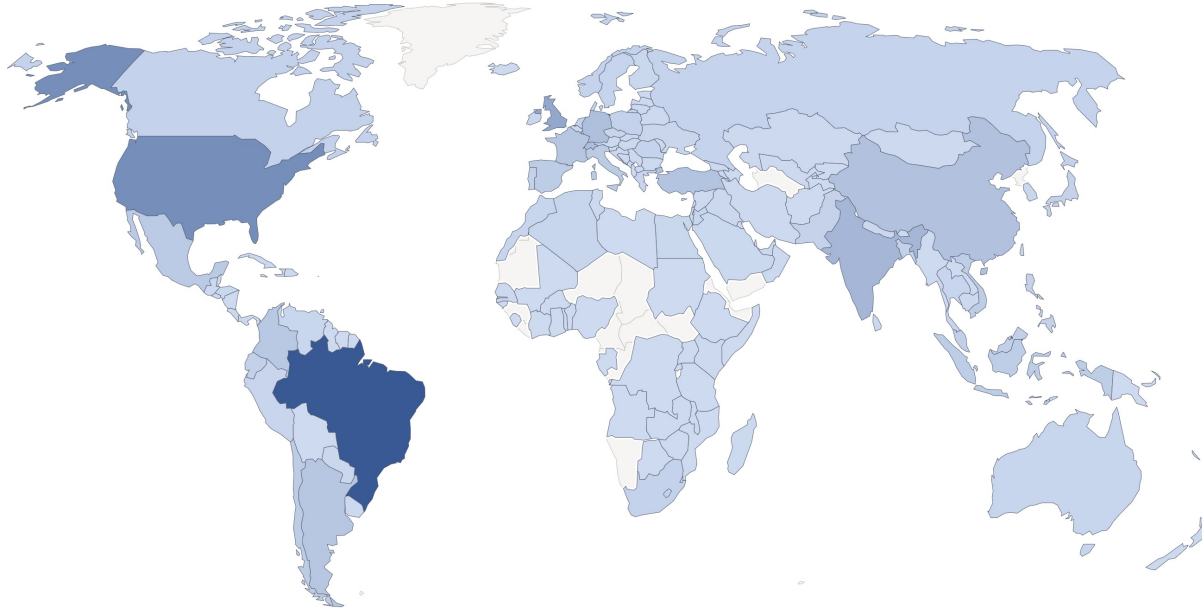
HIGGS DIPHOTON SIGNAL-BACKGROUND CLASSIFICATION
By [Berna Dülger](#) [launch](#) [binder](#)

Impact evaluation

- Monitoring available for about a year now
- Collected in accordance with privacy rules
- Complemented by feedback forms and open data related projects talk invitations to our meetings

Visitor Map

51,204 visits



Web-browser language



Language	Visits
Unknown	24,098
English	18,454
Chinese	1,960
Spanish	1,710
German	658
Turkish	598
French	588
Italian	504
Portuguese	476
Japanese	443

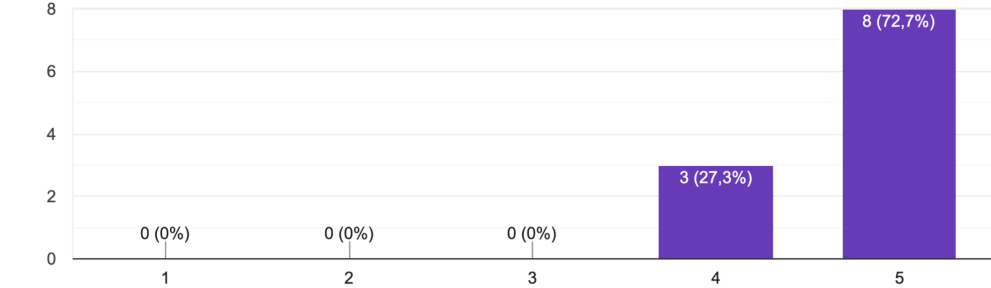
First ATLAS Open Data Tutorial

- [First tutorial](#), open to all people, at the end of November 2025 at CERN
- Great opportunity to improve our documentation across the board
- Trying to serve all the audiences at once was very tricky
- Expecting to hold some online tutorials this year for specific audiences



Overall satisfaction with the ATLAS Open Data Tutorial

11 responses



How would you rate the following aspects of the tutorial?



ATLAS Open Data

- Aims to improve **scientific literacy**, as well foster transfer of highly valuable **skills**, while **reducing accessibility barriers**
- ATLAS Open Data [website](#) is the starting point to anyone interested in using the data, and where all material can be found
- Education and Research Open Data releases provided to the **general public**, with **lots of material** for a wide **variety of audiences**
 - Being used in **schools, universities**, in **public events** and by **interested individuals around the world**
- Communication
 - **Notebook** and **web applications feedback** collected via quick form
 - **Announcements** of new releases and posts about big events in **social media**
 - Collection of **external contributions** derived from Open Data
 - **Tutorials**
- **Interdisciplinary growing potential**
 - **Open data** material provides a great **testing ground** for many things, including **Machine-Learning** developments

Backup

Open Data for Education and Outreach 2025 (beta) release

- Education and Outreach flat ntuples obtained from 36 fb⁻¹ PHYSLITE Open Data for Research samples of data recorded in 2015 and 2016
- [PhysLiteToOpenData](#) derivation framework + [skimming code](#) are public
 - Filtering for the **most relevant physics objects** and uncertainties
 - Allows users to **customise level of information** in ntuples
 - With **Docker** image and local development usage options
 - Repository registered in [Zenodo](#) with a DOI so that it can be cited with a persistent identifier
- Open Data for E&O Release
 - **Unskimmed samples** with very basic cuts applied, such as event cleaning, good runs list, overlap removal
 - **Skimmed datasets** ($\gamma\gamma, \mu\mu, bb, 3\ell, \ell + ME_T, \dots$) varying in size from ~1.5 GB to ~350 GB
- **More data** than last E&O release (36 fb⁻¹ vs 10 fb⁻¹), using latest ATLAS tools and setups, and with **new features** (truth particle collections, more triggers, variables, etc)

