THE ATLAS OPEN DATA PROJECT

Caterina Doglioni, on behalf of the ATLAS Collaboration

LUNDS UNIVERSITET

The **ATLAS Open Data project** is a collection of high-level LHC collision data and tools to be used for education, training, outreach and citizen science. These resources are meant to be platform-independent, and are available on the web and on digital supports, such as USB drives. The project includes a community hub where users can interact and actively participate in discussions.

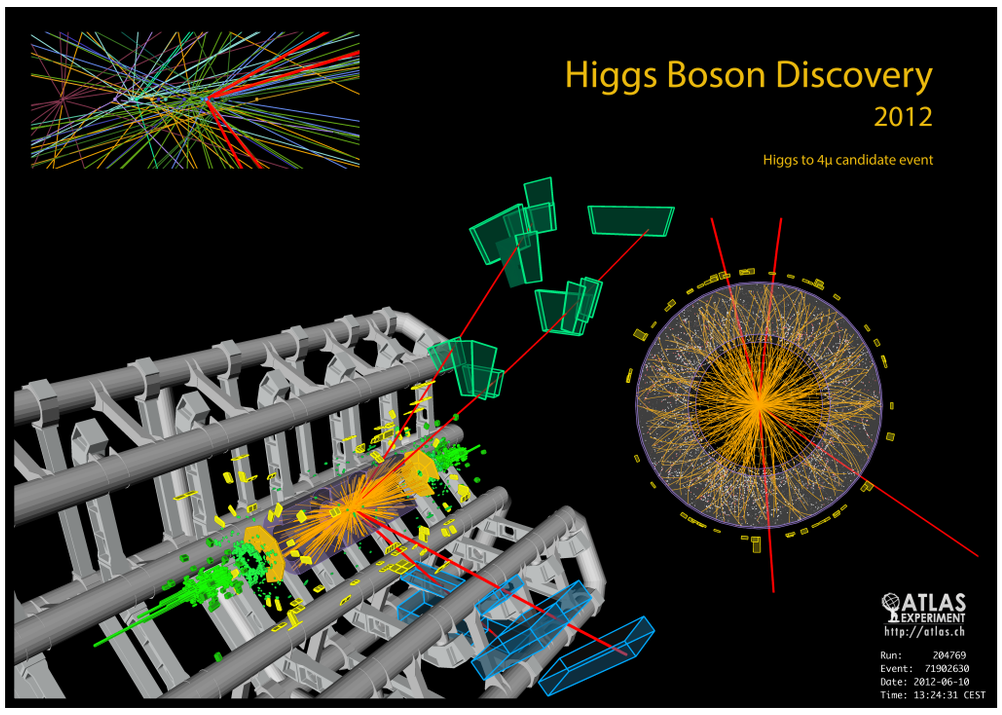
The data

Which data are available

The ATLAS data available within the Open Data project corresponds to approximately 1/fb of LHC proton-proton collision data recorded in 2012, called the **ATLAS Open Data 2016 dataset**. It corresponds to approximately 100 trillion proton-proton collisions. This is part of the data that allowed ATLAS to **discover the Higgs boson**. For this reason, this fraction of the 2012 data has an important scientific, educational and historic value.
Simulated data is also made available for various Standard Model and new physics signal models, for comparison with LHC data.

How data are made available

One of the goals of the ATLAS Open Data project is to allow users to easily gain insight into the process of doing a physics analysis. For this reason, **data are provided in a simplified format containing high-level physics objects** (e.g. electrons, muons, jets...), reconstructed and calibrated with algorithms employed by the ATLAS detector at the time of the data release. Wherever necessary, **simplicity is favoured over precision**. The data events are encoded in a **ROOT *ntuple***, containing a collection of each of the physics objects. A **loose preselection** on the quality of the events and of the objects is applied to decrease the processing time on events that would not pass the analysis selection.



Event display of a candidate Higgs boson decaying to four muons

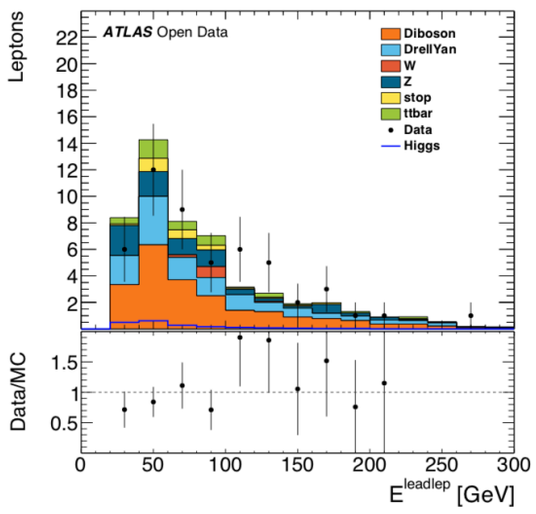
The tools

Example analyses

A number of simplified analyses are provided alongside the ATLAS Open Data dataset:

- **High-statistics Standard Model analyses**, allowing users to measure properties of the SM particles (e.g. the mass of the Z boson), or to confirm agreement between data and simulation. These analyses select the processes $W \to l\nu$, $Z \to ll$, $tt \to l\nu jjj$.
- **Low-statistics Standard Model analyses**, showcasing the difficulties encountered by searches for known, rare processes over larger backgrounds, such as the Higgs boson or diboson production. These analyses select *WZ, ZZ* and Higgs events.
- **A search for a new physics process**, allowing users to search for a hypothetical *Z'* signal beyond the Standard Model, decaying into top-antitop pairs.

These analyses produce data/MC comparison plots, as shown on the right figure for the $H \to WW$ analysis.



Plot produced by the $H \to WW$ analysis after event selection: energy of the leading lepton in the event.

Software tools

The pictures below show a selection of the tools used to build and share the data and the example analyses.

python
ROOT Data Analysis Framework
jupyter
CernVM Software Appliance
opendata CERN

Analyses and docs

The ATLAS Open Data dataset is complemented by analysis tools that can be run on:

- Stand-alone **virtual machines** containing both data and analysis tools that can be installed from the web or from a USB drive
- **ROOTBooks** = Jupyter+ROOT for executing analysis on the Cloud or in the provided VMs **on the Open Data website** with histogram visualization tools.

Documentation is provided via **GitBooks**. It comprises a general introduction and specific instructions for each one of the analyses. **Video guides** are also available.

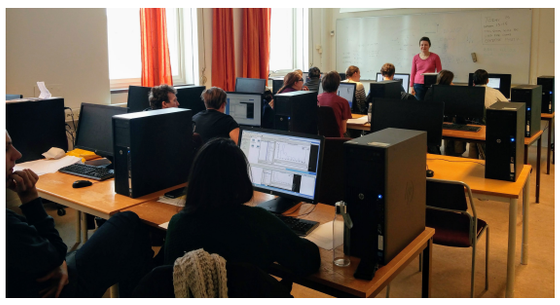
The users

Examples

Universities from all over the world use the ATLAS Open Data for **teaching** students at the basic and advanced level the essentials of data analysis in particle physics. Among the universities and institutes using open data are Maastricht (Belgium), Montreal (Canada), UIS (Colombia), Athens (Greece), TU Dresden (Germany), KTH and Lund (see detail on the right, Sweden), Oslo (Norway), LIP (Portugal), CERN (Switzerland), Birmingham (UK), U of Michigan and California State (US), UCV and USB (Venezuela).
The ATLAS Open Data is used for **Masterclasses**, days in which high school students interact with researchers for a day, do hands-on data analysis, and share their results with other institutes in a CERN videoconference.
ATLAS Open Data is also used **for e-courses and student theses** by CEVALE2VE (Centro Virtual de Altos Estudios de Altas Energias), which builds collaborative networks with and between Latin American institutions.

Advanced Particle Physics course, Lund University

Modern experimental particle physics at Lund University is a course for students that have already a basic knowledge of particle physics, usually taken by those who are currently or will soon be doing their Master's or Bachelor's project with the ATLAS group. Most of the students are already **familiar with basic programming** (Python).The aim of the course is to bring the students up-to-date with contemporary particle physics and the status of the standard model from the experimentalist's point of view. It also includes a basic introduction to the statistical methods used. An important part of the course is the reconstruction and identification of particles as well as analysis strategies.
This is where **we use the ATLAS Open Data for hands-on data analysis.**



2018 ATLAS Open Data Lund lecture

- The students are divided in **groups** of 2 or 3 students
- Each group **chooses one of the Open Data analyses** and reads additional material to prepare their presentation
- Each group gives a **15' presentation** about their analysis, in presence of ATLAS physicists, with a Q&A session
- Each group goes through the **introductory GitBook**
- The groups receive a **guide-sheet** to guide them towards specific tasks of each analysis, e.g. superimpose signal and background for the beyond the SM analysis
- The students **discuss** among themselves and with the teachers

References: ATLAS Open Data Portal – http://opendata.atlas.cern, IPPOG Masterclasses - http://physicsmasterclasses.org
Explore LHC data on new ATLAS educational platform - https://atlas.cern/updates/atlas-news/explore-lhc-data-new-atlas-educational-platform
Higgs event display - https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2012-092/
Review of the ATLAS Open Data Dataset - https://cds.cern.ch/record/2203649/files/ATL-OREACH-PUB-2016-001.pdf

ATLAS EXPERIMENT