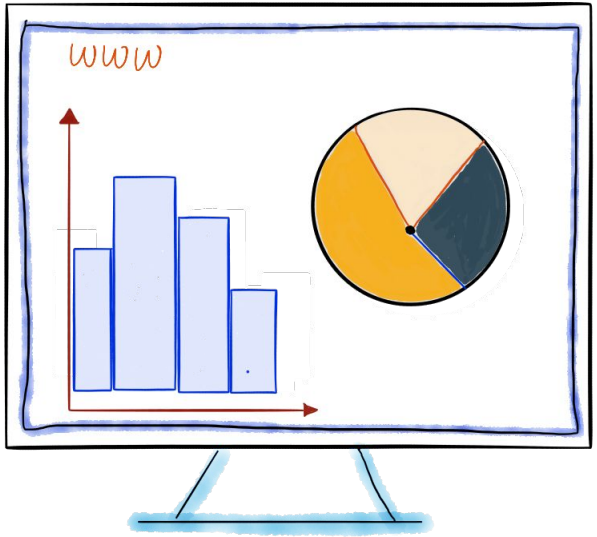
ATLAS *Key* in Open Data



Arturo Sánchez

University of Udine, ICTP and INFN

CHEP 2018, Sofia - July 12th, 2018

ATLAS OpenData and OpenKey: using low tech computational tools for students training in High Energy Physics

One of the significant challenges in High Energy Physics development is the fact that many potential -and precious- students and young researchers live in countries where internet access and computational infrastructure are poor compared to institutions already participating.
To accelerate the process, the ATLAS Open Data project releases useful and meaningful data and tools using standard and easy-to-deploy computational means, such as custom and light Linux Virtual Machines, open source technologies, web and desktop applications. The ATLAS Open Key, a simple USB pen, allows transporting all those resources around the globe. As simple as it sounds, this approach is helping to train students that are now PhD candidates and to integrate HEP educational programs at Master level in universities where it did not exist before.
The software tools and resources used will be presented, as well as results and stories, ideas and next steps of the ATLAS Open Data project.

ATLAS Open Data+Key

Why?... because it is part of our duties as collaboration, and to boost diversity in HEP

A Talent Training and Acquisition program

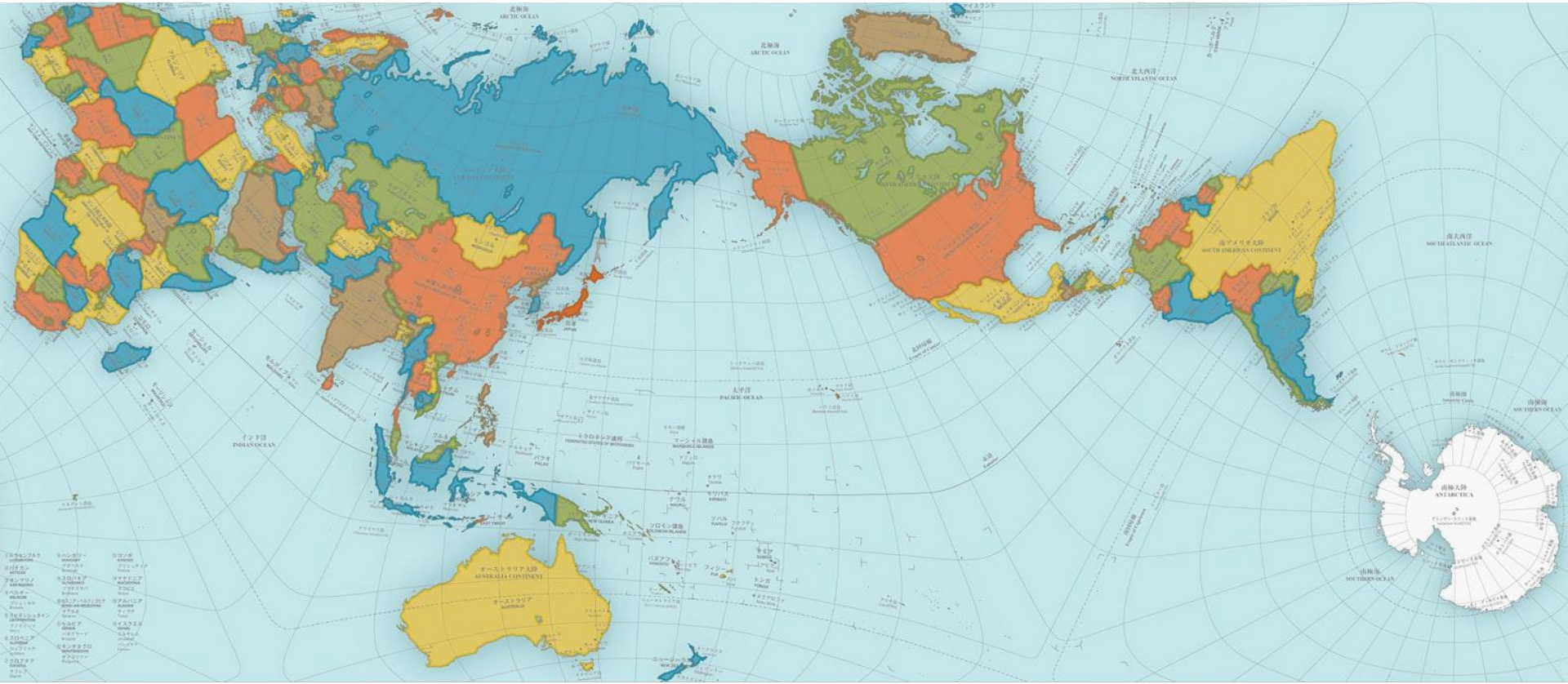
- Why do we need to reach them?

The ATLAS Collaboration, as part of the HEP community, has the responsibility and the need **to**

CERN

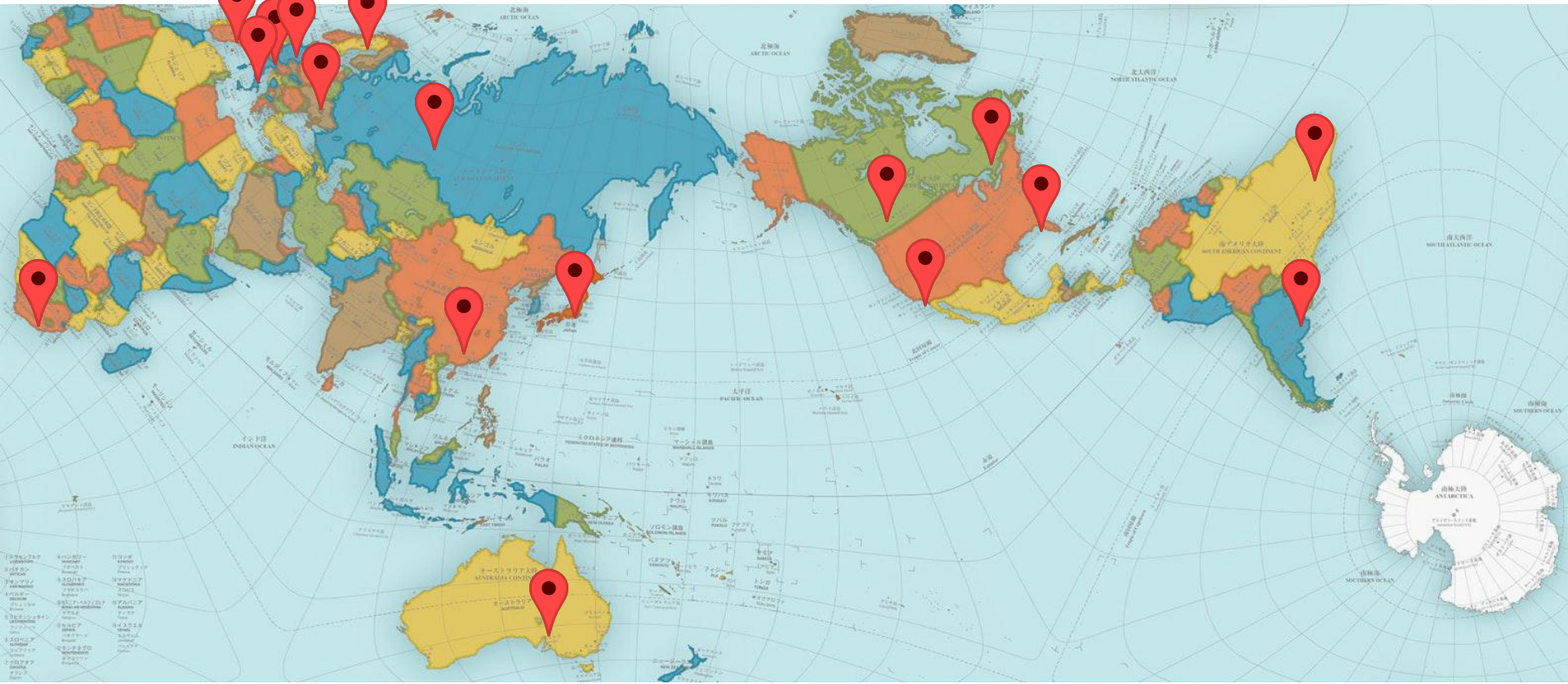
ATLAS EXPERIMENT

transfer knowledge and training **to**



Source → http://www.authagraph.com/category/about

its members institutions **and**



everybody else!

Source → http://www.authagraph.com/category/about

https://atlas.cern

The ATLAS Experiment has a wide range of educational resources available for students and teachers. Explore the categories below to find out how you can engage with ATLAS.

Primary School Students

Resources for students aged 5-12

Secondary School Students

Educational initiatives for students aged 12-18

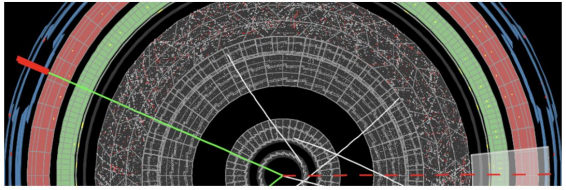
University Students

Opportunities for Undergraduate, Masters and PhD students

Teachers

Ideas for physics and science educators

Citizen Science



Resources & Multimedia

The ATLAS Collaboration and the broad set of HEP experiments rely on public funding to support their programs.
Even more important, their **continuity depends on a constant integration and replacement of human power.**

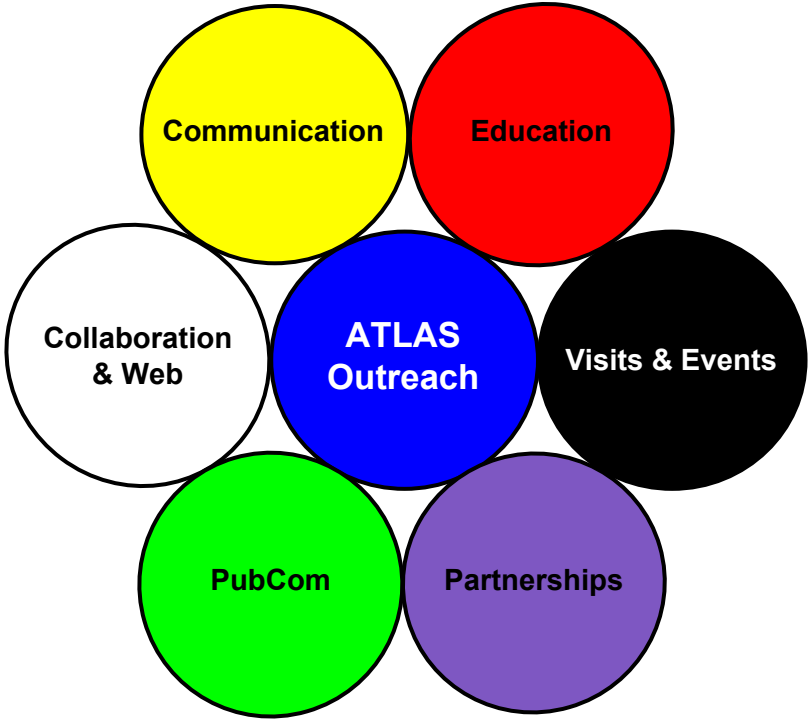
Humans -**students, professionals**- that need to be trained by other senior members of those collaborations.

We can call it **Transfer of Knowledge**, and it goes in both ways: to retribute to society and to keep running the -even more- large scientific endeavours worldwide.

No single nation or entity can do this job alone, nor it should do so. **When knowledge is everybody's property, progress is possible and equally distributed.**

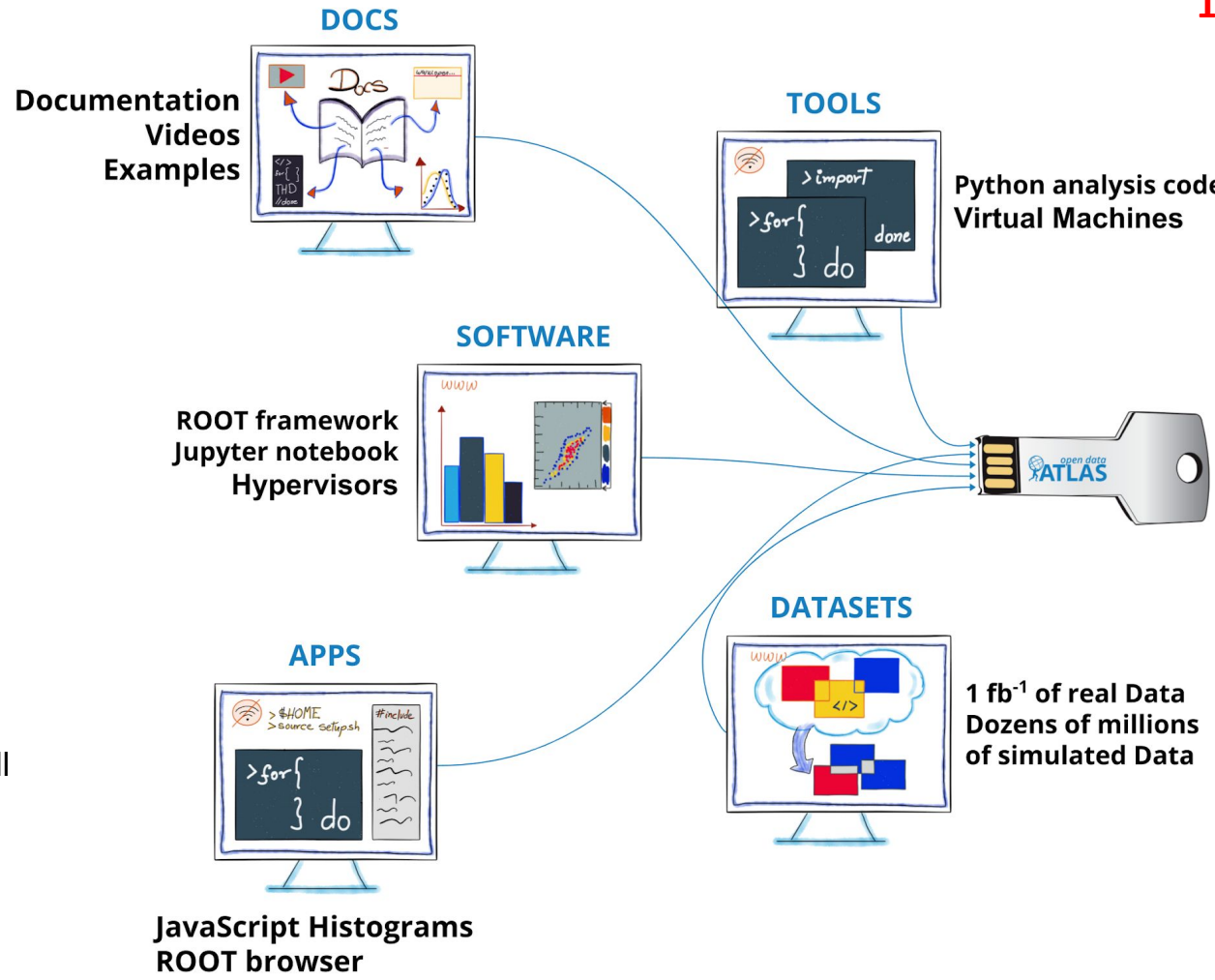
ATLAS Outreach Group tries to achieve that goal.

Today's presentation is about one of those **efforts that strongly depends on the computer tools developed for the Open Source and HEP** worldwide



We fill USB keys with all the educational resources and give them away to instructors & professors to

- Perform workshops -enhanced “masterclasses”- for advanced students
- Create and run laboratory’s courses and *e*-courses in HEP
- Promote HEP into their institutions and with policymakers
- To dedicated students and small libraries who share/copy the resources between them, as we are used to do with books.



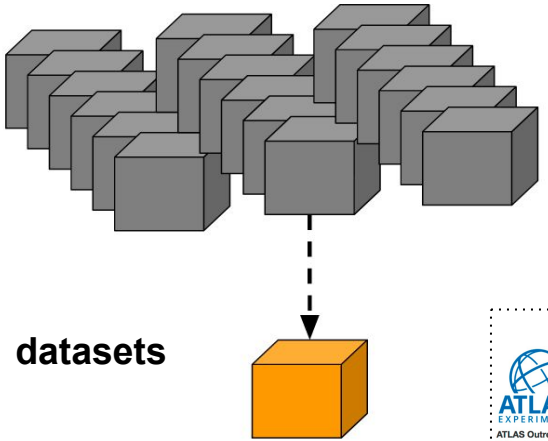
ATLAS Open Data

How?…

the ATLAS members, try to “be” there

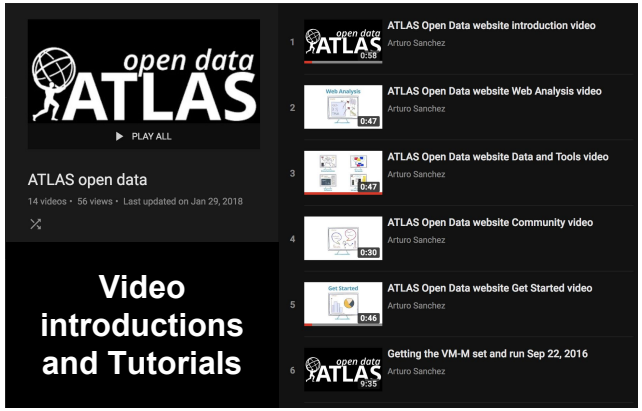
How are we trying to get more people involved?

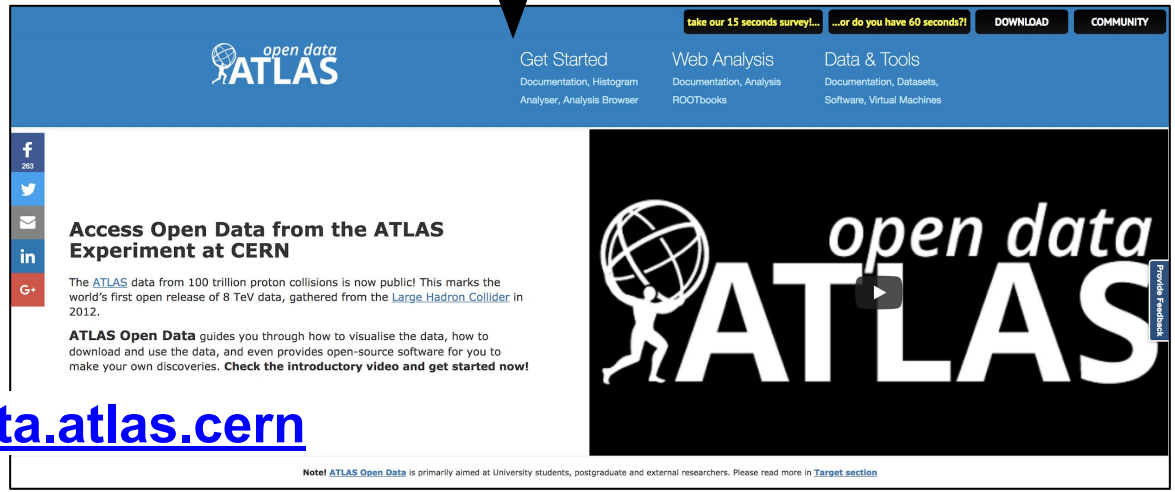
- The Data and the Tools
- Keep data useful and not *just big*
- The Virtual Machine’ approach
- Going for the Apps’ development



datasets

The ATLAS Open Data project ([more here](http://opendata.atlas.cern)) aims to release real and simulated **Data**, together with Open Source **software resources** to analyse those samples.
As well as **documentation** in several useful formats!





http://opendata.atlas.cern

Many resources are in constant integration and updates

We deploy the resources on the Internet.
In a nutshell, they are a series of

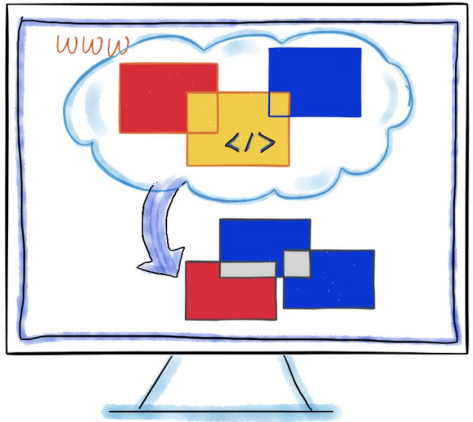
- Data samples in ROOT n-tuple format
- Software and Jupyter Notebooks in Python and C++ to analyse the samples and produce physics analysis
- JavaScript (JS) applications to produce cut-and-count analysis
- Virtual Machines with several Linux-based OS and ROOT CERN analysis framework
- **GitHub & GitLab** repositories
- GitBooks to document the several possible activities that can be performed

Data & Tools Repository

Here you have in a single place all the necessary pieces in order to start you physics analysis in a more complete way. Look into the data like an ATLAS particle physicist!

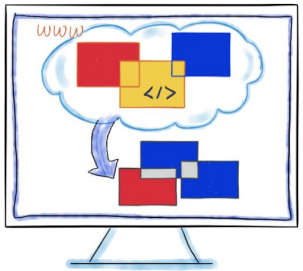
In this section, you can find where to download:

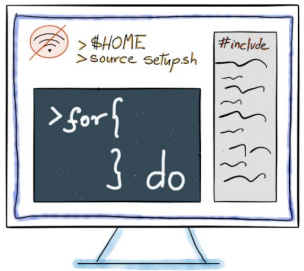
- The complete collection of available datasets
- The different analysis software
- The virtual machines to perform physics searches

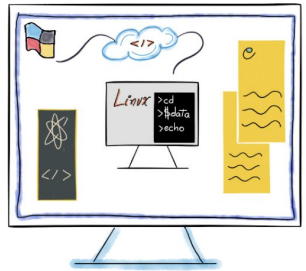


Downloads

Datasets	Software	Virtual Machines

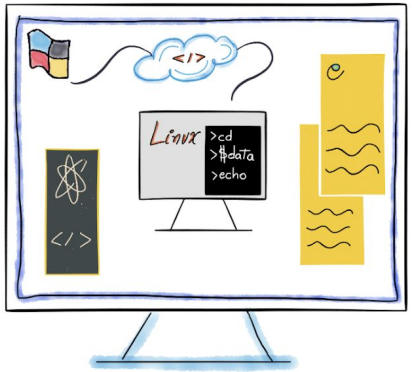






Data & Tools - Virtual Machines

Download a Virtual Machine, which includes useful software, samples and tools for your analysis. Follow the instructions to run a Virtual Machine (**VM**) on any operating system.



Documentation
Datasets
Software
Virtual Machines ✓

We design and develop VM images that can be run in any OS *via* a hypervisor like VirtualBox

VM Version **S** | VM Version **S+** | VM Version **M** | VM Version **L** | VM Version **N** | VM Version **μ**

VM Version S+

http://opendata.atlas.cern/extendedanalysis/vm-toolbox.php

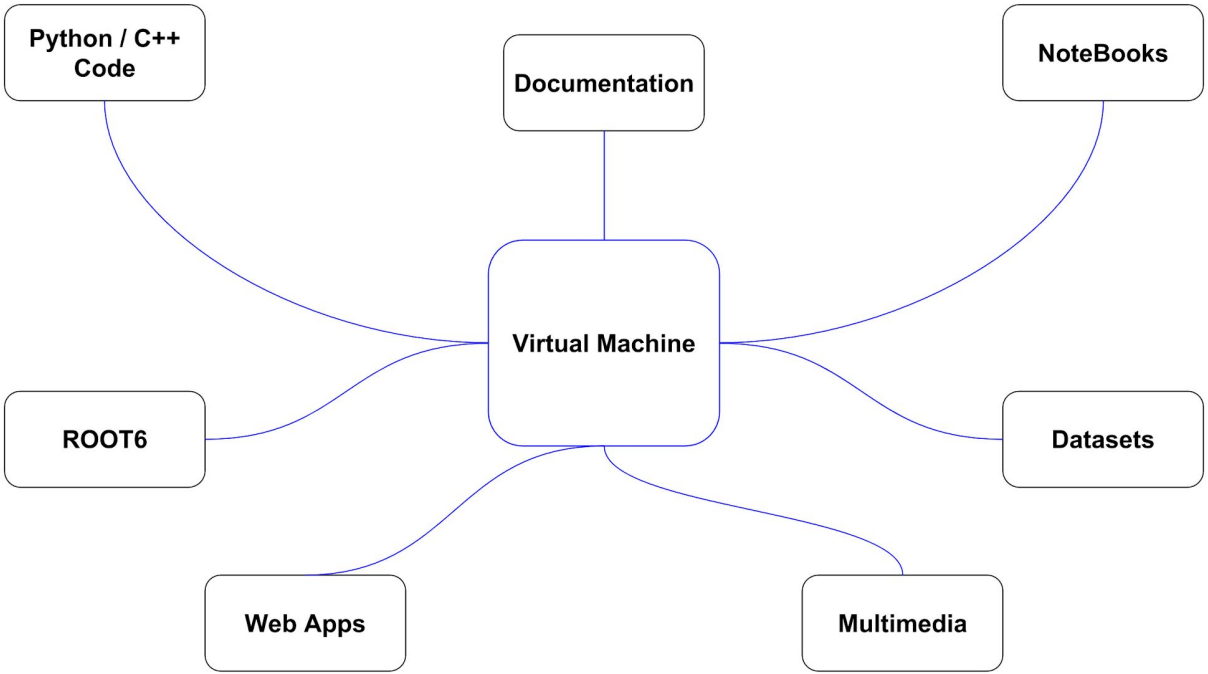
File type	Name	Last modified	Size
	VM_ATLAS-OpenData-ROOT6-Xubuntu-15.04_May_2017-size_S.ova	28-May-2017 03:30	3,1Gb

Description of 28-May-2017 VM: VM-S+ for "*small + ROOTbooks*", it has a total size of ~3.1GB. This contains a complete Xubuntu OS 15.04, important software dependencies, the version of **ROOT6** v6-07-06. It does not contain datasets inside.

Simple specifications aim to guide the user to select the right VM for her/his task(s)

Quick view of the anatomy of the ATLAS Open Data VM

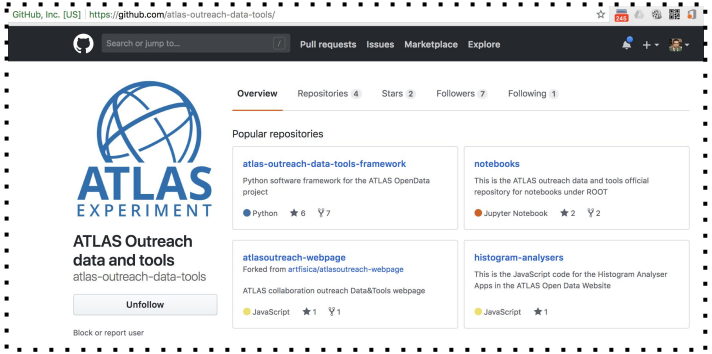
- Based in a Linux-kind OS with standard graphical UI
- The OS in enhanced with *all* ROOT's needed libraries and dependencies
- ROOT5 or ROOT6 analysis framework and IPython
- The Open **Datasets and Software** analysis frameworks
- Jupyter-notebook technology and Examples Notebooks
- Documentation in form of PDFs and Video tutorials.



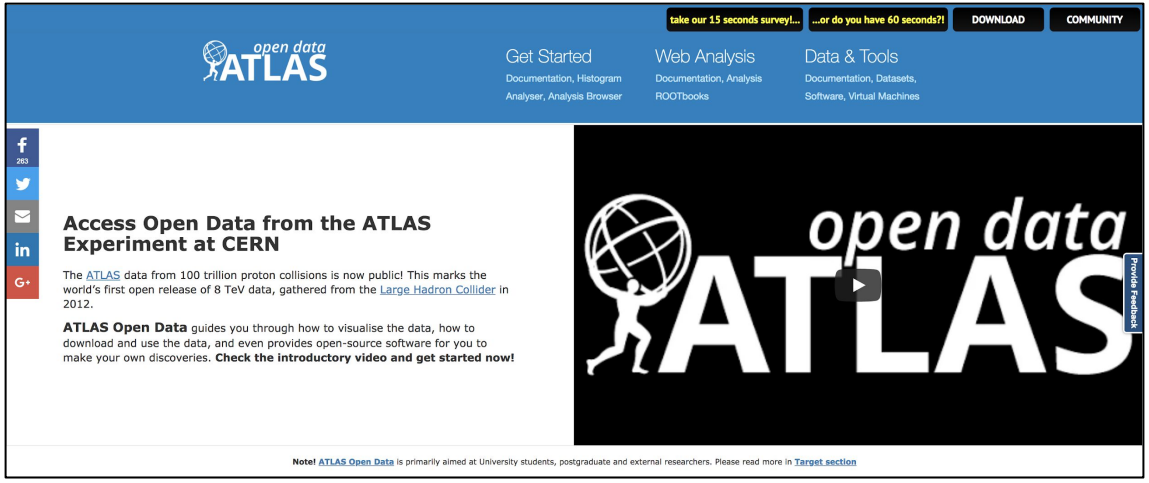
The software provided in the ATLAS Open Data is

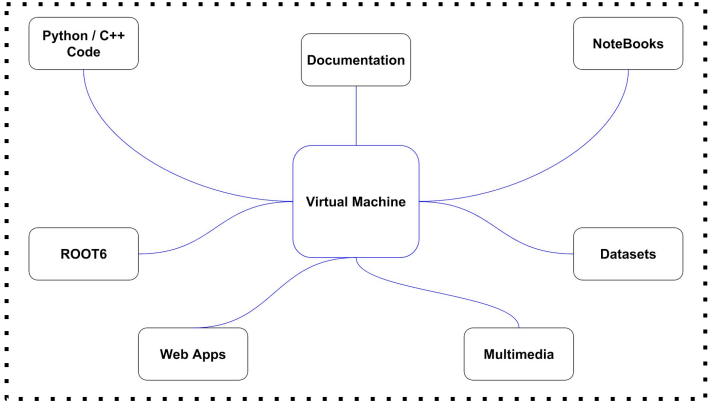
- The preparation of the enhanced **OS with *all* ROOT**'s needed libraries and dependencies
- The **analysis frameworks written in Python and C++** for Particle Physics educational analysis
- **Jupyter Notebooks using the ROOT kernel** with concrete analysis examples
- The **JS web-based applications** to run out of the box.

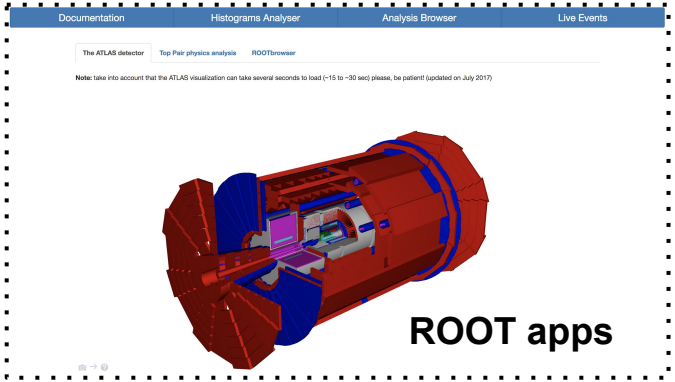
https://github.com/atlas-outreach-data-tools

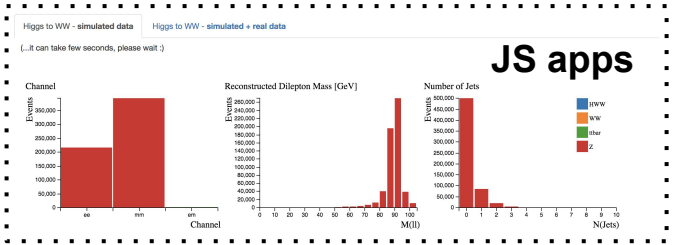


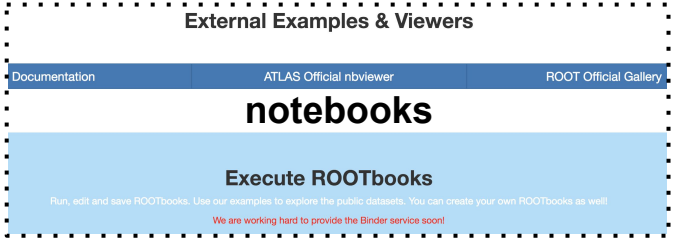
software











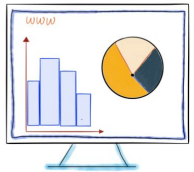
Simple and useful Web-based Apps that

- Allow to explore the ATLAS detector, thanks to ROOT.JS
- Allow to "touch" the histograms and perform simple analysis
- Interact with **Jupyter Notebooks using the ROOT kernel** with concrete analysis examples

http://universidad.ch/ATLAS/outreach/presentations/March_28_2018/_book

ATLAS Open Data *apps*

looking into the interactivity of the website to the desktop



Arturo Sánchez, ATLAS Open Data Project University of Udine, ICTP and INFN
WLCG & HSF workshop, Naples, Italy | March 28th2018

Other ideas like the use of the Electron.JS apps' developing framework to →

→ get the documentation, JS apps and media in a self-contained application for offline use

Plain Text !

JavaScript web-apps

Multimedia

Jupyter Notebooks

GitBooks

WebGL Apps

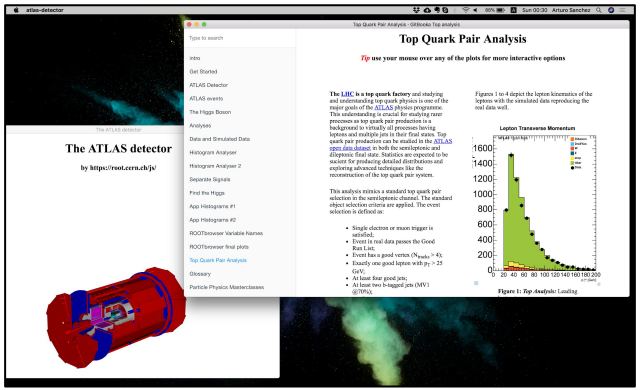
Electron JS App Development Framework

+ HTML + CSS + *time*

Ubuntu Apps

Windows Apps

Mac OS Apps



Apps that can be distributed on the web portal and with the use of the Key !

Electron JS App Development Framework

+ HTML + CSS + *time*

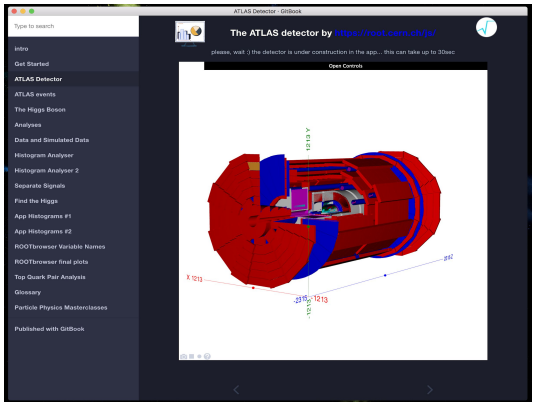
Ubuntu Apps

Windows Apps

Mac OS Apps

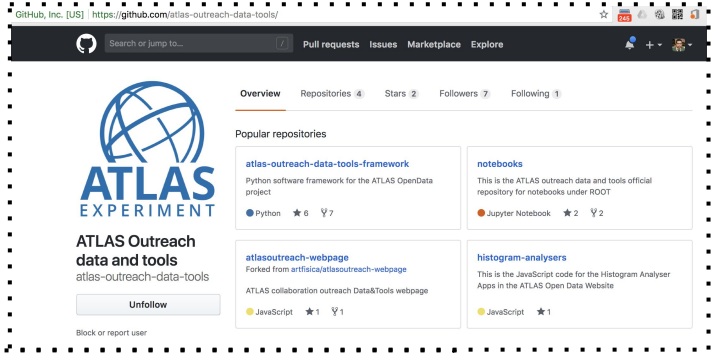
ATLAS Open Data Web platform

ATLAS Open Key



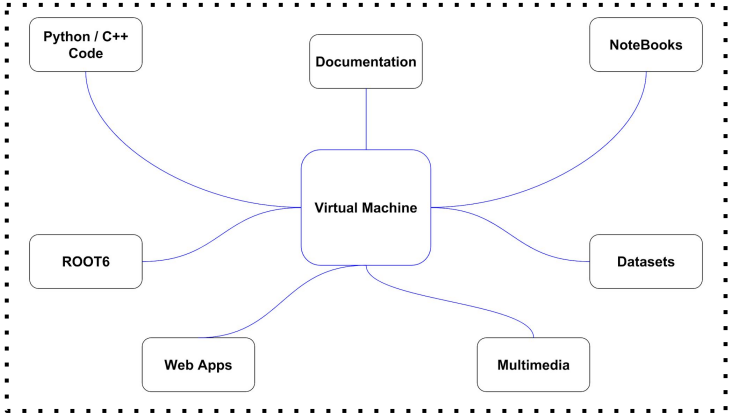
The software provided in the ATLAS Open Key is

- The preparation of the enhanced **OS with *all* ROOT**'s needed libraries and dependencies
- The **analysis frameworks written in Python and C++** for Particle Physics educational analysis
- **Jupyter Notebooks using the ROOT kernel** with concrete analysis examples
- The **JS web-based applications** to run out of the box.



software

https://gitlab.cern.ch/arturos/outreach

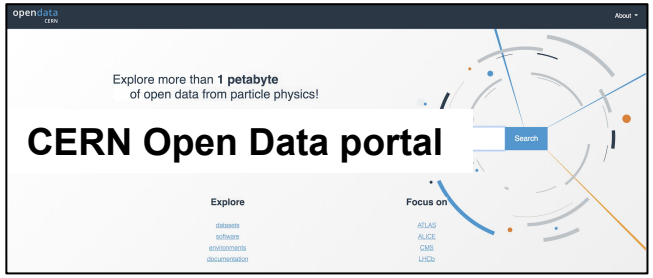


ATLAS Open Key

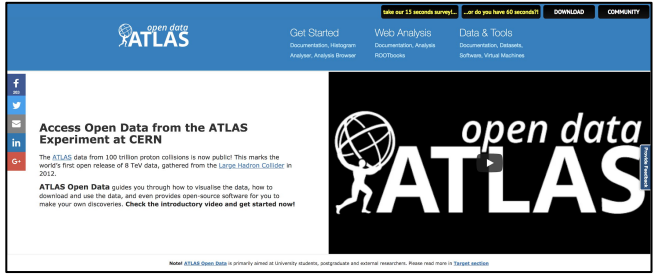
And…
…now?

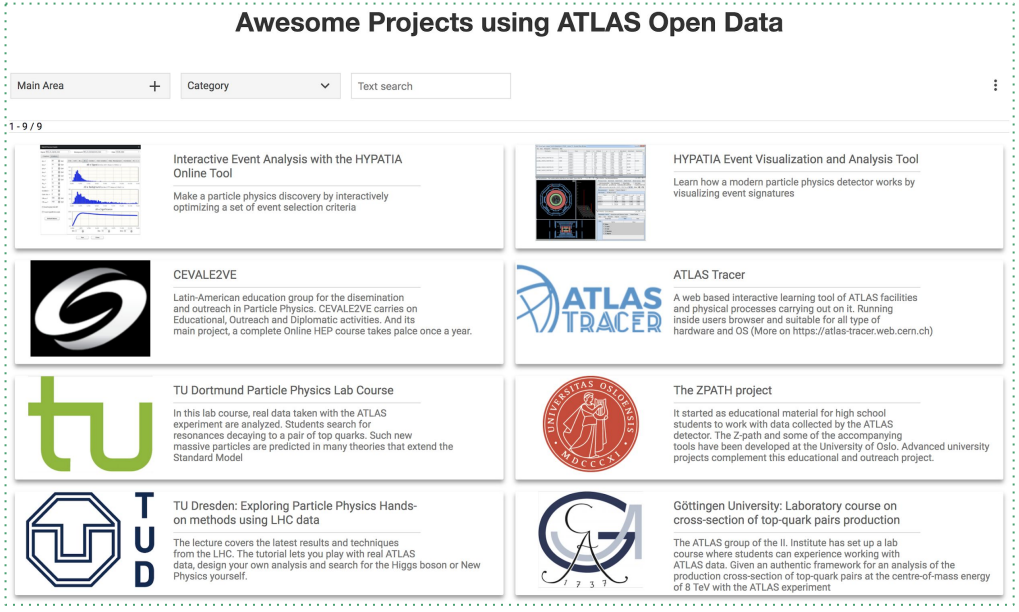
The Open Key project

- How does it works?
- Results and cases
- Plans



On top of the core teams platforms, extra means and projects were developed by the ATLAS and HEP community to reach and hub Groups and Resources





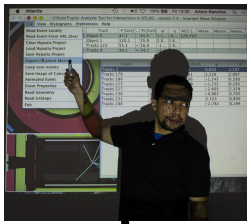
Multiple Workshops and hands-on sessions help us to get feedback and learn how to (and how not to) produce and use the resources

https://indico.cern.ch/event/647642

And some attention from local media

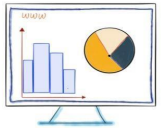
https://www.mppeuct.gob.ve/actualidad/noticias/fisicos-aprenden-trabajar-con-datos-gran-escala

http://elestimulo.com/blog/dos-venezolanos-a-la-caza-de-la-materia-oscura



PWF

open data ATLAS Tutorial @CERN



Learn about the ATLAS OpenData project

http://opendata.atlas.cern

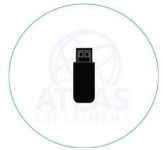
and how it's used by the Outreach community all over the World

Get and Use Public Data & Tools

- Datasets
- Web Apps
- Analysis Code
- Virtual Machines
- Docs & Multimedia
- **and...**

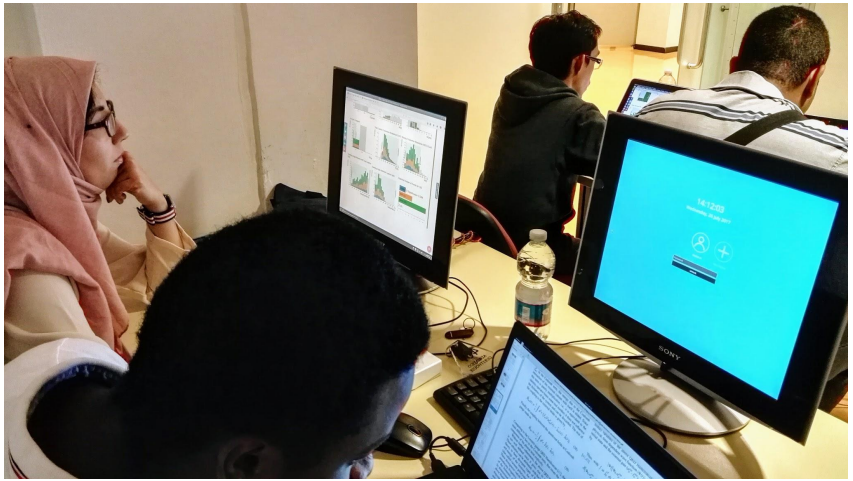
Agenda?
https://indico.cern.ch/e/atlasopendata
To whom?
Everybody @CERN
When?
Friday June 23rd. 14h-15h
Where?
40-R-C10

Questions?
atlas-outreach-data-and-tools@cern.ch

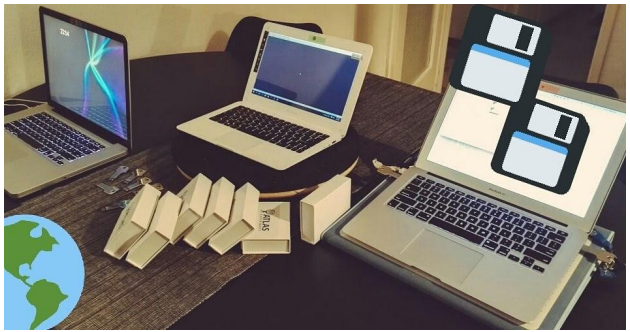


...the Key

Multiple collaborations (ICTP, PWF, CEVALE2VE...), workshops and hands-on sessions help us to get feedback and learn how to (and how not to) produce and use the resources



https://www.ictp.it/physics-without-frontiers/current-country-projects.aspx

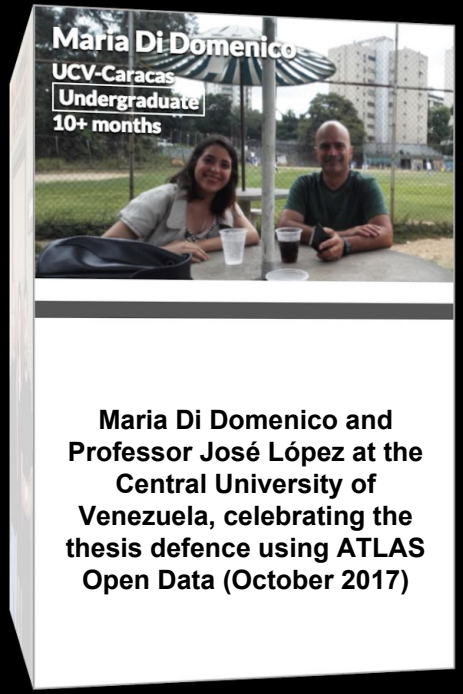


https://www.instagram.com/p/BW3a9PtAoP0

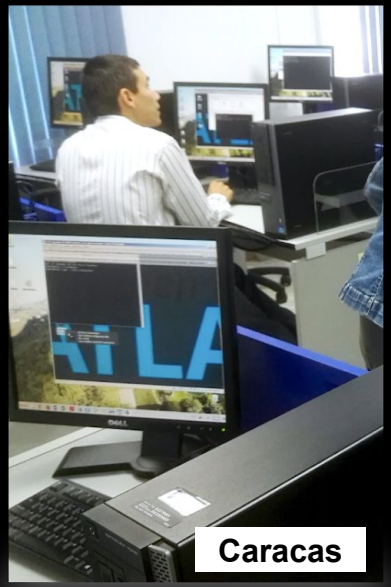
The VMs & the Jupyter Notebook -or ROOTbooks- have been valuable resources to train, teach and explore particle physics…
...and to create a *small* community of developers and instructors around them!

ATLAS Open Data ROOTbooks are a reality thanks to the effort of many people like you

step over the picture to know more about some of them!



Maria Di Domenico and Professor José López at the Central University of Venezuela, celebrating the thesis defence using ATLAS Open Data (October 2017)



http://opendata.atlas.cern/webanalysis/ROOTbooks.php

Our first "version" of the ATLAS Open Key helped to the development of two theses in HEP.

Since 2016 we have been testing and learning how this simple idea can be improved

http://cds.cern.ch/record/2293251

Thesis

Report number	CERN-THESIS-2017-239
Title	**Reconstrucción de masas invariantes de bosones del Modelo Estándar usando datos públicos de ATLAS Open Data**
Translation of title	Reconstruction of the invariant masses of bosons of the Standard Model using public data from ATLAS Open Data
Author(s)	Di Domenico Franco, Maria Rosaria (Caracas, U. Central) ; López Rodríguez, José Antonio (dir.) (Caracas, U. Central) ; Sánchez Pineda, Arturo (dir.) (Caracas, U. Central)
Publication	Caracas, Venezuela : Universidad Central de Venezuela, 2017-10-04. - 207 p.
Thesis note	Licenciate : Caracas, U. Central : 2017-09-01
Note	Presented 04 Oct 2017
Subject category	Particle Physics - Experiment ; Education and Outreach
Accelerator/Facility, Experiment	CERN LHC ; ATLAS
Abstract	In this thesis were studied and produced a set of analysis with the end of reconstructing the invariant masses of the W, Z and Higgs bosons. Together with other physical variables, following the rules of the Standard Model. These analyses are produced using proton-proton collision data collected by ATLAS Experiment in 2012, at a centre of mass energy of $8\ TeV$ corresponding to an integrated luminosity of $1\ fb^{-1}$. This data is available in the http://OpenData.ATLAS.cern portal. All this was made using a cloud computing platform called SWAN, also developed and localized at CERN. SWAN is based on Jupyter notebooks and ROOT. With the use of SWAN, several physical analysis were made, from the initial production to the final product: a set of self-explained notebooks showing the physics and the programming elements that are needed to do such reconstructions on analysis of cut-and-count type. Finally, these notebooks will be used as educative resources for teaching and outreach of HEP at the university level.

http://cds.cern.ch/record/2291838

Thesis

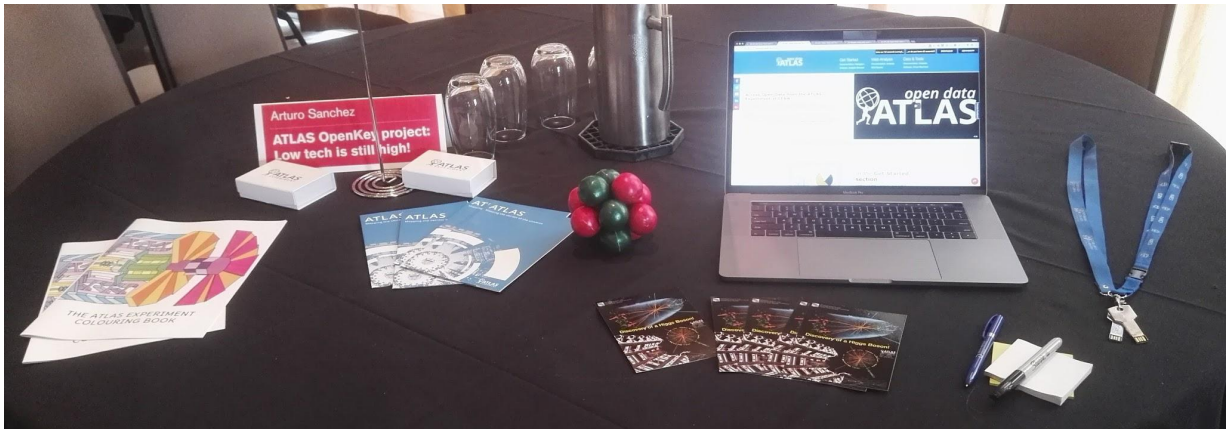
Report number	CERN-THESIS-2017-217
Title	**Perspectivas y Evaluación de producción de Materia Oscura en asociación con un quark liviano, un quark pesado (quark b) o un boson electro débil en colisiones de particular a energías de colisión de $\sqrt{(s)} = 8$ TeV**
Translation of title	Perspectives and Evaluation of Dark Matter production in association with a light quark, a heavy quark (b-quark) or an electroweak boson in particle colliders at a centre-of-mass energy of $\sqrt{(s)} = 8$ TeV
Author(s)	Garcia, Iskya (Caracas, U. Central) ; Sanchez Pineda, Arturo (dir.) (ICTP, Trieste)
Publication	2017. - 130 p.
Thesis note	Master : Caracas, U. Central : 2017-11-07
Note	Presented 14 Jul 2017
Subject category	Particle Physics - Experiment ; Education and Outreach
Accelerator/Facility, Experiment	CERN LHC ; ATLAS
Abstract	The Weakly Interacting Massive Particles, or WIMPs, are the only major Dark Matter (DM) candidate particles currently accessible at the LHC. It is expected that the production of a pair of these particles in association with particles belonging to the Standard Model (SM) will be evidenced in final states with the presence of large quantities of Missing Transverse Moment (E_T^{miss}). This thesis summarises a model independent search for DM, in several final states that contain E_T^{miss}, called: Mono-jet, Mono-b and Mono-Z. Where the designation "Mono" refers to the presence of a single jet or boson which is used as a guide to discern between DM production and other final states with E_T^{miss} but corresponding to only SM particle decays. The analysis were performed using open data published by ATLAS in 2016, collected during 2012, and that corresponds to an integrated total luminosity of 1 fb1. Where a pair of protons were collided with center-of-mass energy of $\sqrt{(s)} = 8$ TeV. No excess events were observed over predictions of the SM in any of the final states studied, as is evidenced by p-values calculations. Discarding the presence of events coming from new physics, coinciding with official results of the ATLAS Collaboration for such DM searches at $\sqrt{(s)} = 8$ TeV. All this using tools and techniques of high energy physics that is currently in used at the LHC collaborations.

Summary

ATLAS Open Data and its Key help to perform Knowledge Transfer:
to retribute to society and to keep running the -even more- large scientific endeavours worldwide.

http://opendata.atlas.cern/community/contact.php

Thanks!



CreativeCommons Global Summit 2017 & 2018: http://sched.co/E8cU

backup

ATLAS Open Key

Why? How? Who? Where? And…
…now?

A Talent Training & Acquisition program

- Why we need to reach them?

Where are our target audiences?

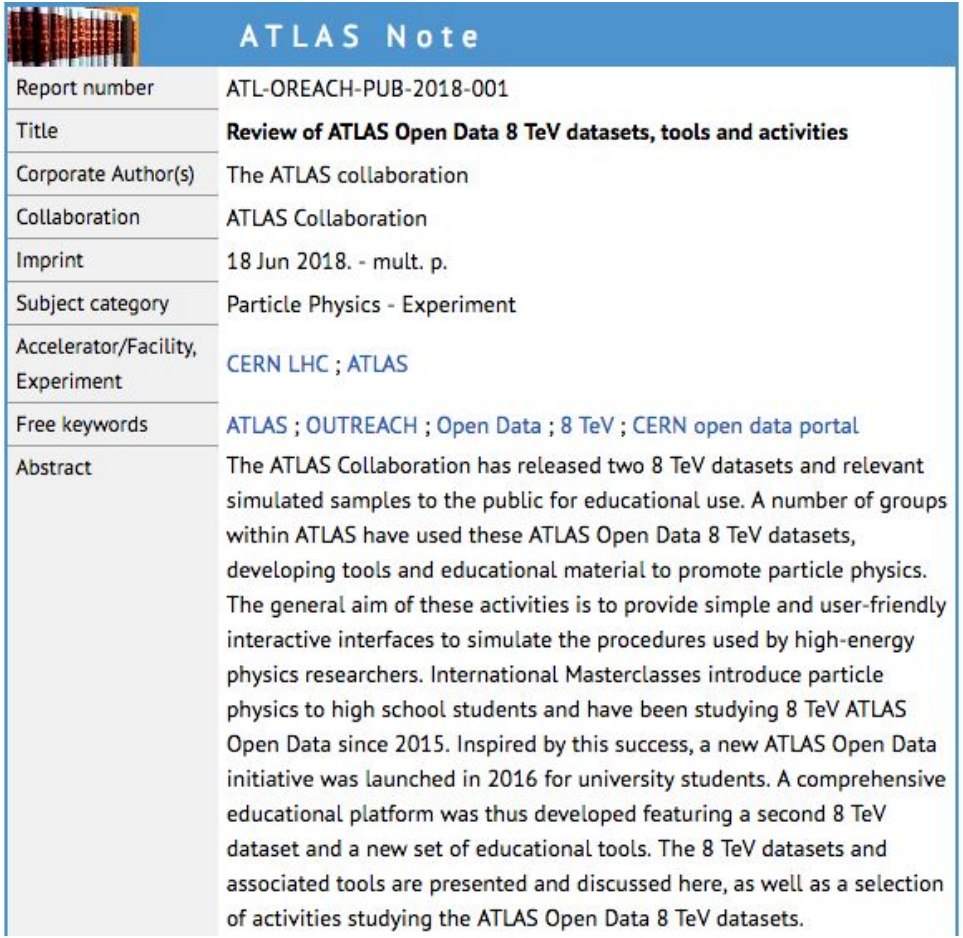
- Worldwide

How are we trying to get more people involved?

- The Data and the Tools
- Keep data useful and not *just big*
- The Virtual Machine' approach
- Going for the Apps' development

The Open Key project

- How it works?
- Results and cases
- Plans



ATLAS Note

Report number	ATL-OREACH-PUB-2018-001
Title	**Review of ATLAS Open Data 8 TeV datasets, tools and activities**
Corporate Author(s)	The ATLAS collaboration
Collaboration	ATLAS Collaboration
Imprint	18 Jun 2018. - mult. p.
Subject category	Particle Physics - Experiment
Accelerator/Facility, Experiment	CERN LHC ; ATLAS
Free keywords	ATLAS ; OUTREACH ; Open Data ; 8 TeV ; CERN open data portal
Abstract	The ATLAS Collaboration has released two 8 TeV datasets and relevant simulated samples to the public for educational use. A number of groups within ATLAS have used these ATLAS Open Data 8 TeV datasets, developing tools and educational material to promote particle physics. The general aim of these activities is to provide simple and user-friendly interactive interfaces to simulate the procedures used by high-energy physics researchers. International Masterclasses introduce particle physics to high school students and have been studying 8 TeV ATLAS Open Data since 2015. Inspired by this success, a new ATLAS Open Data initiative was launched in 2016 for university students. A comprehensive educational platform was thus developed featuring a second 8 TeV dataset and a new set of educational tools. The 8 TeV datasets and associated tools are presented and discussed here, as well as a selection of activities studying the ATLAS Open Data 8 TeV datasets.

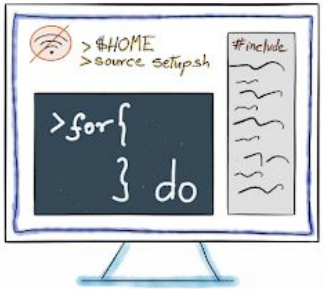
Review of ATLAS Open Data 8 TeV datasets, tools and activities

ATL-OREACH-PUB-2018-001
https://cds.cern.ch/record/2624572

Projects Ongoing

open data ATLAS

The ATLAS Open Data C++ Analysis Framework's project

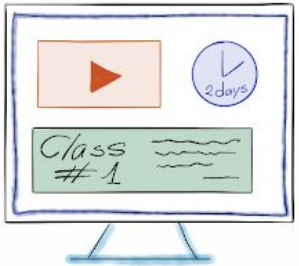


20.06.2018

[Internship Proposal #1 (2018)]

open data ATLAS

The ATLAS Open Data WorkshopS' project

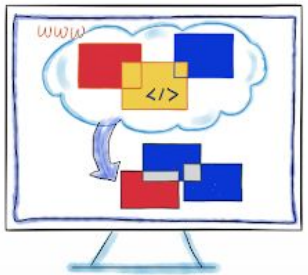


26.05.2018

[Internship Proposal #2 (2018)]

open data ATLAS

The ATLAS Open Data 13 TeV Data&MC project



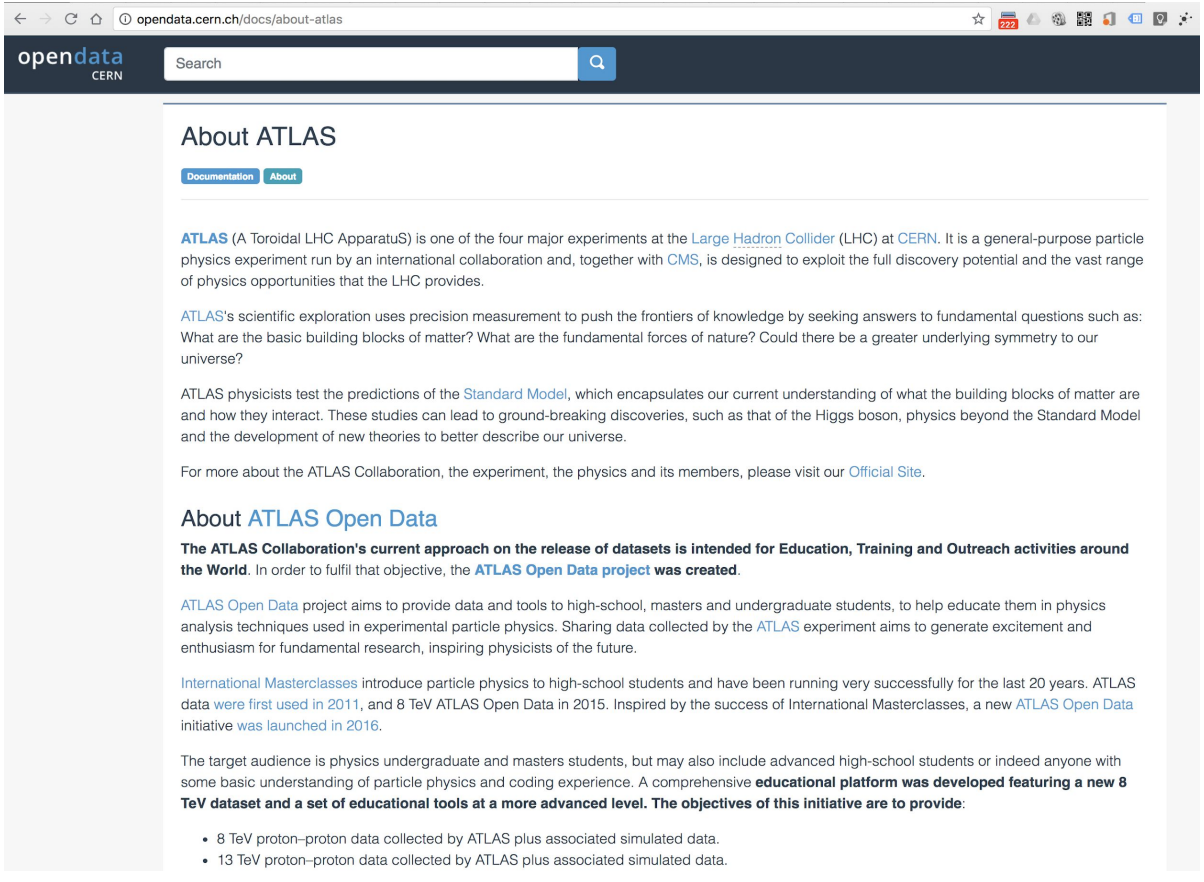
10.06.2018

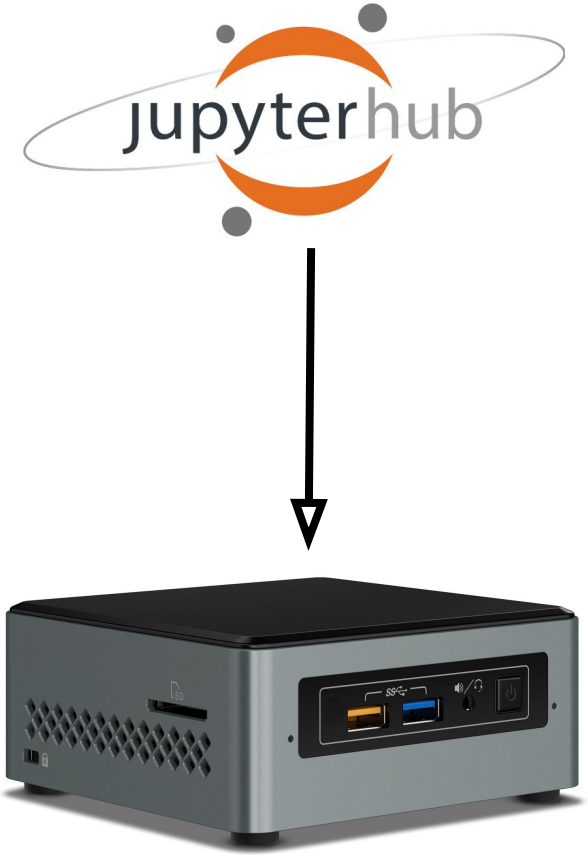
[Internship Proposal #3 (2018)]

http://opendata.cern.ch/docs/about-atlas

ATLAS Open Data is Updated in CERN Open Data

http://opendata.atlas.cern (ATLAS Open Data platform) , http://opendata.cern.ch (CERN Open Data)

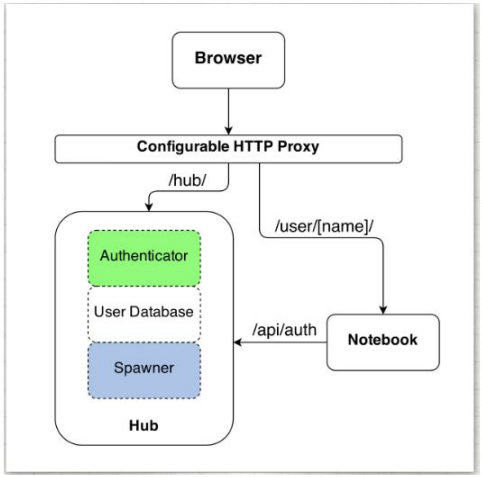




JupyterHub, a multi-user **Hub**, spawns, manages, and proxies multiple instances of the single-user Jupyter notebook server. JupyterHub can be used to serve notebooks to a class of students, a corporate data science group, or a scientific research group.

Three subsystems make up JupyterHub:

- a multi-user **Hub** (tornado process)
- a **configurable http proxy** (node-http-proxy)
- multiple **single-user Jupyter notebook servers** (Python/IPython/tornado)



JupyterHub performs the following functions:

- The Hub launches a proxy
- The proxy forwards all requests to the Hub by default
- The Hub handles user login and spawns single-user servers on demand
- The Hub configures the proxy to forward URL prefixes to the single-user notebook servers

https://jupyterhub.readthedocs.io/en/stable